

## The use of high-throughput phenotyping in genomic selection context

Reyna Persa<sup>1</sup>, Pedro Cesar de Oliveira Ribeiro<sup>1,2</sup> and Diego Jarquin<sup>1\*</sup>


Crop Breeding and Applied Biotechnology  
21(S): e385921S6, 2021  
Brazilian Society of Plant Breeding.  
Printed in Brazil  
<http://dx.doi.org/10.1590/1984-70332021v21S619>

**Abstract:** *One of the biggest challenges that breeders face is the development of improved cultivars in changing climate conditions posing extra challenges to their labor. On the other hand, the availability of data generated with automated systems offers an opportunity to characterize genetically and phenotypically genotypes with high detail. Modern sequencing technologies delivering hundreds of thousands of molecular markers, offered the opportunity of selecting genotypes without the need of observing these in fields and this methodology was coined as Genomic Selection (GS). More recently, sophisticated automated phenotyping platforms depending on sensors able to measure a large number of plant features were also developed and have shown potential in plant breeding applications. These modern phenotyping systems that attempt to efficiently deliver phenotypic information on secondary traits are also known as high-throughput phenotyping platforms (HTPPs). The integration of HTPP with GS models opened a new research front to improve the efficiency of the selection methods based on genomic data only, specially of those traits depending on a large number of genes with small effects (complex traits). However, there are still remaining some issues to solve for developing a robust methodology able to combine in an efficient and informed way these two high dimensional data types. In this document, we provide an overview of the statistical analysis of the data derived of the HTPPs for improving the predictive ability of conventional GS models. First, we provide a brief introduction showing the utility of genomic data in plant breeding applications. Then, we provide an overview of the field-based HTPPs considering the light detection and ranging, and the unmanned aerial vehicles and how the image data derived from these platforms can be used to accelerate genetic gains. After that, we discuss about the extension of the conventional GS models to allow the incorporation of data derived of the HTPPs as main effects and also in interaction with environmental factors. The availability of several sources of information have opened a venue to investigate besides the univariate or single trait model, models based on multiple traits and also models that consider multiple time measures allowing longitudinal GS studies. Finally, we provide some conclusions as well as we mention some the current issues that do not allow to fully exploit the potential of HTPPs in plant breeding applications.*

**Keywords:** *Development of improved cultivars, Next-Generation Sequencing (NGS), Genomic Selection (GS), sensor types, categories platforms, high-throughput phenotyping platforms (HTPPs).*



**\*Corresponding author:**

E-mail: [diego.jarquin@gmail.com](mailto:diego.jarquin@gmail.com)  
 ORCID: 0000-0002-5098-2060

**Received:** 11 June 2021

**Accepted:** 12 June 2021

**Published:** 06 July 2021

<sup>1</sup> University of Nebraska – Lincoln, Lincoln NE, 68583, USA

<sup>2</sup> Departamento de Biologia, Universidade Federal de Viçosa, 36570-000, Viçosa, MG, Brazil

## INTRODUCTION

Currently, the world faces an increased demand in the production of food for human consumption for a growing population, thus there is also increased demand for land and water resources to establish plant cultivars (Fróna et al. 2019). At the same time, climate change is one of the biggest challenges to produce food in a sustainable way forcing plant breeders to come across with solutions for developing new improved cultivars in shorter periods of time. Some potential solutions may consider to reduce the breeding cycles for releasing cultivars, increasing adaptability to new environments, and the resistance to a wide range of biotic and abiotic stresses (He and Li. 2020). The availability of cheaper and faster sequencing methodologies have encouraged breeders, farmers, and agronomists to adopt and leverage genomic data to increase crop yield and genetic gains (Li et al. 2021). For example, the next-generation sequencing (NGS) technology has shown great progress in functional genomics (Huang et al. 2012) and it has allowed generating genome-wide DNA polymorphism data for a large number of experiments in an efficient way (Mardis 2008).

Among the applications of genomic data in breeding applications, breeders have implemented molecular marker-assisted selection (MAS) for cultivar development. This method uses a reduced set of molecular markers for selecting the best candidate genotypes and it has shown to be more efficient for selection in plant breeding than the pedigree- and phenotypic-based selection (Mohan et al. 1997). However, marker-assisted selection has some limitations when some traits are controlled by a large number of genes with small effects, i.e., quantitative traits (Bernardo 2008). Employing sequencing techniques such as genotyping by sequencing has led to the discovery of a large number of markers providing a cost-efficient tool for breeders (Elshire et al. 2011). To overcome the limitations imposed by MAS, genomic selection (GS) emerged as method that jointly uses information of all molecular markers for selection purposes.

The concept of GS was first introduced by Bernado in 1994, and later Meuwissen et al. (2001) introduced a series of models to solve the challenges posed by considering a very large number of makers ( $p$ ) covariates compared with the total number of observations ( $n$ ) for model fitting ( $p > n$ ). GS is a method that employs genomic prediction models (GP) helping to reduce phenotyping costs using predictions as surrogates of phenotypes. Thus, potentially GS could help to speeding up the selection process by skipping the growing season if accurate predictions of untested genotypes are provided (Krchov and Bernardo 2015). This methodology requires genomic and phenotypic information for model calibration and it is used to screen candidate genotypes using their marker profiles only.

Hence, GS has the advantage of increasing genetic gains by reducing breeding cycles (Rincent et al. 2017). Other authors have combined different sources of information in the prediction models. For example, marker and pedigree information have been integrated in prediction models (de los Campos et al. 2009, Crossa et al. 2010, Christensen et al. 2012) showing marginal improvements in predictive ability. In addition, the inclusion of environmental covariates, and the genotype by environment interaction ( $G \times E$ ) (Burgueno et al. 2012, Jarquin et al. 2014, Perez-Rodriguez et al. 2017) have shown significant improvements (~25%-50%) with respect to conventional prediction models based on maker data only.

Since plant breeders are interested in releasing stable cultivars for a wide range of environmental conditions, multi-environment trials should be conducted which in many cases result expensive and infeasible. The evaluation of multiple plant breeding locations on large geographic locations is necessary for assessing the adaptability of new candidate phenotypes (Chapman et al. 2014). The main advantage of GS is that it does not require phenotypic information from the validation set to aid in the selection process; however, initially the model training requires both, phenotypic and marker data. Hence, precise phenotypic information is very crucial (White et al. 2012, Cobb et al. 2013, Crain et al. 2018) for obtaining accurate predictions.

## SENSOR TYPES, CATEGORIES PLATFORMS, AND HIGH-THROUGHPUT PHENOTYPING IN PLANT BREEDING

Breeding programs have the necessity of monitoring a large number of genotypes in fields to select the best candidates to be advanced to next breeding cycles. Recently, technological advancements based on remote sensing have been used to develop High Throughput Phenotyping (HTP) platforms with applications in plant breeding programs. These platforms have revolutionized the way for collecting phenotypic information in a faster and efficient way helping breeders to reduce

manual labor. These tools have been used for phenotyping physiological traits on large numbers of breeding lines by capturing thousands of plant images every day with a high degree of automatization (Fahlgren et al. 2015). Thus, the advantages derived of using HTP platforms currently are helping breeders to decrease the time for phenotyping lines in fields, accelerate plant breeding while reducing phenotyping costs (Krause et al. 2018).

HTP platforms can examine, quantify, and precisely assess phenotypes in agricultural experiments by using sensors and data collection systems employing low or no physical human intervention known as labor non-destructive (Li et al. 2021). Thus, this methodology is able to produce useful data for selecting new lines and for developing prediction models to predict new traits. The generated data may be used as the primary trait response or as a covariate associated with the principal trait response, usually grain yield.

### Type of sensors

Several HTP platforms have been developed and have shown to be efficient tools by providing phenotypic information through improved spatial and temporal resolution (Yue et al. 2019). These, HTP platforms are equipped with sensors enabling the efficient collection of phenotypic data. The principal function of the sensors is to interpret plants as optical objects, and every part of the plant exhibits a distinctive spectral signature occurring from wavelength-specific properties of absorbance, reflectance, and transmittance from the vegetation surface (Liew et al. 2008).

The most common sensors used in breeding applications include: 1) Red-green-blue (RGB) sensor, the wavelength is approximately 400 to 700 nm and it possess sensing visible wavelengths, the RGB band sensor is able to capture images of morphological features of plants (Kim et al. 2018, Kim et al. 2020); 2) Near-infrared sensor (NIR), with a wavelength that varies between 700 to 1400 nm, the sensing highest reflectance of plant green area is ranged between 700 to 1300 nm, this sensor possesses a high correlation on vegetation status (Bei et al. 2011, Kim et al. 2020); 3) Fluorescence sensor, with a wavelength that varies between 180 to 800 nm, its sensing fluorescence is emitted by short-wave light absorption, and it is able to detect drought-related traits (Chaerle et al. 2006, Kim et al. 2020); 4) Hyperspectral sensor, sensing thousands of bands per pixel and generally these sensors are used for indoor purposes (Hamada et al. 2007, Kim et al. 2020); 5) Thermal sensor, with a wavelength of 700 to  $10^6$  nm and sensing emitted radiation of object that expands with the detection object temperature, these sensors are the most appropriate to phenotype drought-related traits (Baluja et al. 2012, Kim et al. 2020); and 6) Tridimensional 3D and laser-imaging detection and ranging (LiDAR) sensor this remoting sensing technology possess a wavelength that ranges between 250 to 2,000 nm and the surface scan of the main object and distance measurement is analyzing by the reflected light, this sensor can be used with aerial platforms, and ground vehicles (Lin 2015, Kim et al. 2020).

### Categories of HTP platforms

The HTP platforms have been classified into three categories according to the type of infrastructure used: greenhouse, ground-based, and aerial systems (Singh et al. 2019). The HTP phenotyping platforms implemented in greenhouses allow, to some degree, to control and manipulate the environmental conditions such as temperature, humidity, photoperiod, light intensity, and nutrients factors. Thus, depending on the mechanical structure and sensors, movement between the sensors and plants an indoor HTP plant phenotyping platform can be classified as a benchtop-type or a conveyor-type. These platforms (benchtop-type and conveyor-type) contain cameras, complement light sources, automated watering, and weighing machines, precisely to capture plant phenotypic data (Li et al. 2021).

The principal function of the conveyor-type is based on the “plant-to-sensor” mode; in this context, potted plants are carried in an imaging room with cameras through an automatic door on the conveyor operated by a computer for automatic imaging of the plants. This kind of platform has been implemented in crops such as barley, sorghum, rice, maize, and chickpea (Li et al. 2021). Some of the traits that have been evaluated with the conveyor-type tool are biomass, plant height, width, compactness, drought stress, biomass, leaf content, water content, nitrogen content, and salt tolerance (Chen et al. 2014, Neumann et al. 2015, Neilson et al. 2015, Atieno et al. 2017, Miao et al. 2020).

On the other hand, the indoor benchtop-type measures phenotypic traits vulnerable to environmental fluctuations, and the imaging head of this platform includes several sensors, and it is operated by a computer-controlled mechanical arm that automatically localizes where the plant is located and collects phenotypic data in real-time; usually, this

dispositive is used for small species with delicate stems (Li et al. 2021). Some studies have been reported in *Arabidopsis thaliana*, maize, and soybean evaluating traits such as the rosette size, evaporation, radiation dosage stress, projected area, the growth rate of the ear, and silk, water efficiency, drought stress (Tisné et al. 2013, Bricchet et al. 2017, Peirone et al. 2018, Chang et al. 2020).

Regarding to the crops observed in natural field where the environmental conditions (i. e. temperature, solar radiation, humidity, rain) cannot be controlled or manipulated, plant phenotyping could be time-consuming on large number of plots when using human labor. Therefore, an alternative would consider the adoption of the new HTP methodologies such as the ground-based platforms for collecting phenotypic data. This kind of platform has attracted breeder's attention because of its importance in plant genotyping and phenotyping (Bai et al. 2019). The ground-based HTP platforms can be categorized into four types: the first one is known as pole/tower-based field HTP where the sensors are mounted directly atop a pole or tower (Li et al. 2021). This tool is able to measure canopy height, photosynthesis, and crop lodging and it has been used in maize, soybean, wheat, and sugar beet (Raesch et al. 2014, Friedli et al. 2016, Susko et al. 2018).

The second one is a mobile ground-based HTP platform, which has been manipulated and adapted in tractors, bicycles, and designed platforms for moving through the field for generating phenotypic data. This mobile ground-based tool can provide higher resolution data than aerial remote sensing and collects phenotypic data of time series (Li et al. 2021). This tool has been utilized in cereals (maize, wheat, rice) and soybean for monitoring leaf area, leaf angular, canopy transpiration, canopy temperature, canopy coverage, canopy height, and plant height (Derry et al. 2016, Bai et al. 2019, Qiu et al. 2019).

The third type of ground-based tool is a gantry-based field, which has a frame with a sensor box that collects phenotypic data information along with XYZ directions (Li et al. 2021). Some studies that utilized this tool were performed in peanut, cowpea, maize, and wheat mostly measuring traits for canopy transpiration, plant height, 3D leaf area, water stress, nitrogen stress resistance, canopy temperature, and canopy closure (Vadez et al. 2015, Virlet et al. 2017, Beauchene et al. 2019).

The fourth HTP platform ground-based tool is the cable-suspended field. It possesses a sensor system well located to obtain the phenotypic data through a cable-driven system integrated with a sensor bar (Li et al. 2021). Some studies have been developed for monitoring plant height, ground cover, canopy temperature, canopy height in soybean, maize, and wheat (Kirchgessner et al. 2016, Bai et al. 2019).

Sometimes it is not feasible to monitor cultivars using the field-based HTP platforms due to impractical situations such as tall cultivars, muddy soil conditions, and a large number of plots (Watanabe et al. 2017). For these cases, unmanned aerial systems (UAS), manned aerial platforms (MAP), and satellite platforms can be helpful. Unmanned aerial platform systems have demonstrated great potential for measuring phenotypic traits and become an alternative to ground-based phenotyping platforms. The UAS evaluates plots in a short period of time by mitigating the effects of the constantly changing weather conditions; also, this platform allows high spatial and temporal resolution measurements producing valuable information for plant breeding programs, this platform requires one operator (Haghighattalab et al. 2016). Some authors have employed this tool for monitoring carotenoid content in inbred lettuce, canopy height, and coverage in rice (Wan et al. 2020), and yield in cotton.

Another aerial phenotyping platform is known as the manned aerial platform HTP (MAP). MAP is mostly operated by aerial mapping vendors; this platform has been used in urban areas and the forestry industry mostly but also for agriculture purposes. The platform has an exceptional and high-resolution multispectral on the images and its sensors are placed on a small airplane (Kim et al. 2020), and it needs an operator to conduct the crop images during the flight. For instance, Deery et al. (2016) performed a study in wheat by monitoring canopy temperature through MAP on large plots for selection in plant breeding programs.

Recently, with the development of satellite HTP platforms, there is an increasing interest in using this technology on plant phenotyping applications. These satellites are able to show panchromatic imagery, multispectral imagery, or radio detection and ranging (RADAR) data (Li et al. 2021). In this context, Basso et al. (2016) used RapidEye multispectral data to evaluate the capability of a high spectral resolution satellite. These authors were interested in detecting within-field crop nitrogen stress via vegetation indices in wheat. Currently, the satellite platforms are continuously improving their

image resolution since these can be affected by the weather conditions. However, due to the large areas that can be explored with the satellite platforms, there is a promising future for their use after overcome limitations such as low spectral and spatial resolution and lack of quality measures (Al-Wassai and Kalyankar 2013).

In summary, many of the implementations of HTP platforms have helped breeders to accurately characterize phenotypes of large number of individuals or populations (i.e., genotypes) in the fields under different environmental conditions.

### High-throughput phenotyping in plant breeding

Some interesting studies that have leveraged data derived from high throughput platforms (HTP) in plant breeding applications are mentioned below. Baek et al. (2020) measured and analyzed the morphology (i.e., area, perimeter, width, height, thickness, circularity roundness, and solidity) and a color index of soybean seeds via imaging processing derived from a digital camera. These authors showed that the morphological phenotype data is useful for detecting individuals with specific traits. For example, seed color can be related to different traits such as functional substances, diseases, and responses to environmental conditions (Yang et al. 2010, Silva et al. 2012). Thus, the obtained data is valuable in the plant management for decreasing the labor time before planting.

Wang et al. (2018) assessed the performance of five different sensing technologies (HTP) in sorghum for measuring plant height. These five HTP platforms were: LIDAR-Lite sensor (ground-vehicle), ultrasonic sensor, Kinect camera, DSLR cameras, and digital camera (unmanned aerial vehicle). All of the data compilation and the data processing methodologies (plot-level height values assessed by the ultrasonic sensor, LL2 sensor, the Kinect camera, and the proximal imaging by four DSLR cameras) were contrasted with manual measurements to find which method was the most accurate for generating precise estimations of the real values. The authors concluded that the two best methods for measuring plant height were the sensor fusion technology by LiDAR and an array of high-resolution cameras held by a ground vehicle platform.

Watanabe et al. (2017) evaluated the performance of an unmanned aerial vehicle (UAV) in sorghum to measure plant height and the response to nitrogen availability. Plant height was also measured with a ruler (PHR) to contrast the estimations derived from unmanned aerial vehicle (PHUAV) under normal (N-P-K: 1.2-1.8-1.6 kg a<sup>-1</sup>) and low (N-P-K: 0.6-1.8-1.6 kg a<sup>-1</sup>) fertilization. An average correlation coefficient between plant height measured by PHUAV and plant height measured with a ruler (PHR) of 0.53 was obtained. The results showed that under normal fertilization the correlation coefficients between observed and predicted PHUAV varied between 0.448–0.492. These correlation coefficients were higher under low fertilization (0.629-0.675). These authors stated that UAV remoting sensing has the potential to reduce labor costs and offers the opportunity of monitoring plant growth over time.

Singh et al. (2019) performed a study using different digital elevation models from unmanned aerial systems (UAS) as an alternative to visual evaluations of lodging. Lodging is considered a complex trait in cereals such as wheat, oat, and barley and it is responsible for decreasing yield potential. Also, lodging is considered an ideal trait for performing HTP in large areas to conduct genomic studies in breeding applications. The digital and visual measures of lodging were highly correlated (0.93) in the two years (2016 and 2017) when the experiment was conducted in fields; being taller plants with early heading those that showed more lodging. A Genome-Wide Association Study (GWAS) performed on the collected data showed a significant and consistent peak at chromosome 2A, and prediction accuracies across environments varied between 0.19 to 0.55. The authors showed that UAS-based phenotyping is a useful HTP tool to perform quantitative assessment of lodging on plots, for implementing GWAS, and whole genomic prediction studies. Thus, these results showed the importance of the applications of UAS-based phenomics on a large number of plots for breeding and genetic studies.

Lately, Krause et al. (2020) studied the potential of grain yield in wheat using vegetation indices that integrate the measurement of canopy structure and photosynthetic activity depending on the amount of light reflected off of the crop canopy. These vegetation indices were collected via aerial HTP systems equipped with various cameras and sensors at the early generation (seed-limited stage to improve selection for grain yield). The obtained results showed that this methodology increased the selection at the early generation, thus it has potential for breeding applications.

## THE INTEGRATION OF HIGH-THROUGHPUT PHENOTYPING INTO GENOMIC SELECTION - RECENT STUDIES

In the last few years, some authors have published studies in different crop species employing HTP for GS purposes. For example, some of these considered different methodologies for targeting traits of interest via hyperspectral data (Sun et al. 2017, Crain et al. 2018, Lozada et al. 2020, Galán et al. 2020, Krause et al. 2020, Kismiantini et al. 2021). As mentioned, HTP platforms provide high-quality data that accurately characterize broad training populations at the different developmental stages during the growing season. Potentially, the combination of genomic selection (GS) and HTP can improve accuracy of prediction models while reducing labor costs in breeding programs (Araus et al. 2018, Moreira et al. 2020). In general, breeders consider grain yield one of the most important traits for evaluation and improvement. Under this perspective, the HTP platforms can detect/measure traits that are highly correlated with grain yield such as plant height, biomass, and other quantitative traits (Wang et al. 2018), spectral reflectance, and canopy temperature (Amani et al. 1996, Crain et al. 2018).

Sun et al. (2017) utilized longitudinal information of secondary traits, canopy temperature, and normalized difference vegetation index (NDVI) in addition to marker and pedigree data for predicting grain yield in wheat. A two-step analysis procedure was implemented for this study and three statistical models were evaluated: simple repeatability (SR), multi-trait (MT), and random regression (RR) with secondary traits; and the best linear unbiased prediction (BLUP) model was used for comparison purposes. The RR model was slightly better than the SR and MT models. This study illustrated the importance of using HTP data to improve the conventional GP models. More recently, Lozada et al. (2020) used five spectral reflectance indices (SRI) from HTP platforms and 11,089 SNPs to assess the effects of using secondary traits (fixed effects) for improving prediction accuracy to select winter wheat lines with yield potential.

The evaluated parameters were genetic and phenotypic correlations, response to selection, correlated responses, and relative efficiency of selection for different spectral traits. The spectral traits showed moderate to high phenotypic and genetic correlations resulting in a remarkable improvement in prediction accuracy for grain yield ( $p < 0.001$ ). The obtained results showed advantages of integrating HTP tools with genomic prediction for increasing genetic gains in winter wheat breeding programs. Crain et al. (2018) incorporated HTP wheat data into GS models where the wheat lines were evaluated in two different environments: drought (2014-2015) and heat (2015). The Normalized difference vegetation index (NDVI) and canopy temperature (CT) from heading until physiological maturity were measured with a portable georeferenced phenotyping vehicle (Phenocart).

The statistical models used for predicting grain yield were a univariate-single trait model (uniGS), a model using only HTP traits as predictors (HTPr), a genomic model with phenotypic covariates (GS+HTP), and a multi-trait model that incorporated grain yield, NDVI and CT for responses (multiGS). All models were assessed using three different approaches: partial least square regression, elastic net, and GBLUP. The results indicated that multiGS GBLUP showed an average improvement of 7% in prediction accuracy with respect to marker (uniGS) selection method across the three experiments, and between -5% and 20% for the different experiments. The HTPr models showed the most variable performance because for the heat 2015 experiment it returned the highest prediction while a negative correlation for the drought 2015 experiment. The GBLUP models returned the best results when HTP data was also included in the model.

These authors showed that the incorporation of HTP data helps to increase prediction accuracy of conventional GS models suggesting that collecting multiple HTP measurements has the potential to improve genetic gain by improving the predictive ability of models. Montesinos-López et al. (2017a) used data derived of a HTP platform to study the effect of predictive ability when including the interaction between hyperspectral bands B (reflectance data for vegetation indices in wheat) and environments (B×E) for predicting grain yield. Conceptually, this model term includes all first order interactions between each band and each environment.

In addition, these authors evaluated Bayesian functional models, considering the main effect of all of the bands, marker SNPs, lines and environments and also the B×E interaction and concluded that the model that includes the B×E interaction significantly outperforms the conventional GS model. Krause et al. (2019) proposed multi-kernel, multi-environment GBLUP models using genetic marker-pedigree, hyperspectral reflectance data, and environmental information, to evaluate grain yield in wheat. Two of these models were based on main effects only while a third model

also included the interaction between hyperspectral bands and environments. The highest prediction accuracy was obtained when implementing a multi-kernel model that combines marker-pedigree information with hyperspectral reflectance and their interactions with environmental factors.

A similar study was performed by Galán et al. (2020) using spectral data collected by an unmanned aerial vehicle (UAV). The integration of genotypic, hyperspectral, and phenotypic data helped to improve biomass yield prediction in hybrid rye. The phenotypic and hyperspectral information was collected in four locations and three prediction models were compared: a single-kernel model with either genomic or hyperspectral reflectance, a multi-kernel model combining both data types, and a bivariate model that also incorporates information on plant height. The results showed that across environments, the second model performed the best followed by the single kernel model. The obtained results showed good prediction accuracy for grain yield with larger training sets. The authors concluded that the integration of HTP and genomic selection have potential for predicting gain in yield in plant breeding programs.

Other studies presented by Montesinos-López et al. (2017b) and Montesinos-López et al. (2018) showed the benefits of considering functional analysis to minimize data dimensionality, allowing at the same time the production of a higher signal-to-noise ratio for each observed value. In addition, Montesinos-López et al. (2017b) showed that HTP measures obtained at different time points lead to a small increase in prediction accuracy relative to the models that use data from a single time point.

Lately, Kismiantini et al. (2021) investigated the use of generalized Poisson regression (Ridge regression, Lasso regression, and Elastic net regression) for genome-enabled prediction of phenotypes with genomic and hyperspectral images. These authors proposed to use this model because they stated that breeding phenotypic values are very noisy and the Poisson regression models permits an efficient integration of information from various sources (genotypic data, environmental data, and high-resolution images) for predicting grain yield. The models with the best performance were the elastic net regression and the Lasso regression when the interaction of genomic data, measures of high-resolution data with environmental factors were include.

Recently, Gonçalves et al. (2021) performed a study using Near-infrared spectroscopy (HTP) and genomic data for predicting fiber (FIB) and sucrose content (PC) in a sugarcane population. The data of a commercial population comprising 385 individuals genotyped for 124,307 SNPs and screened via NIR spectroscopy was analyzed. These authors compared three different ways for fitting the models: (M1) genome-based, (M2) NIR-based prediction, and (M3) the integration of SNPs markers plus NIR wavenumber variables as predictors. The different data type combinations were fitted using two regression methods: the BayesB regression and the partial least squares (PLS). The BayesB method returned the lowest predictive ability for FIB and PC when using only information of M1. The highest prediction accuracy was found with M2 for FB (0.61) and PC (0.54). M3 (a combination of SNPs markers plus NIR spectra data) returned a predictability of 0.58 for PC and showed a marginal improvement for FIB (0.62).

These authors concluded that under the M3 model, the improvements were only marginal. A Tukey test was performed to verify whether the difference was significant or not. In addition, the PLS regression results were similar to those from the BayesB. In this case, M1 had the lowest prediction accuracy for both traits. M2 presented the highest prediction accuracy for FIB (0.39) and with M3 small improvements were shown for PC (0.36). In addition, the genomic prediction model based on marker data (M1) only showed a poor predictive ability being the highest for PC (0.16). The authors stated that the results obtained with only the main effect of the markers were poor perhaps because small training sizes were considered. The combination of M1 and M2 models could have improved the prediction accuracy; however, in this case it did not occur because perhaps M3 NIR spectra captured most of the variation in both traits losing relevance the maker data. The authors concluded that using NIR spectra data intents to be a good tool for reducing time and expensive laboratory protocols via statistical prediction models.

## CONCLUSIONS

High-throughput (HPT) platforms have revolutionized the traditional methods for phenotyping cultivars reducing the breeder's labor time. Those platforms are categorized as indoors (greenhouse) and outdoors (ground-field, aerial) platforms. The indoors tools are designed for small experiments in a greenhouse where the environmental conditions (temperature, light, humidity, water stress) can be controlled and manipulated up to some extent. On the other hand, the outdoor HTP platforms are used for phenotyping a large number of plots. If the cultivars present very tall plants,

or if the ground has a higher slope difficult to cope with, or if the soil conditions are muddy, the best HTP platform to be used will be based on aerial tools.

In addition, high-throughput phenotyping and phenomics are essential tools that provide different ways to record plant development, to learn about the genetic basis of quantitative traits, and to leverage the interaction between genotypes and environment in prediction models. The use of this methodology (HTP) has the potential to change the way of how breeders select their material for advancement in breeding programs. HTP tools have shown to be successful when thousands of plots need to be phenotyped, by reducing costs, and manual labor. One of the main advantages of incorporating HTP data in prediction models is that it allows increasing the prediction accuracy helping breeders to select the lines with the best performance. In all of the studies presented here, the authors agreed that combining pedigree, marker information, environmental data, and HTP tools resulted in a significant improvement of the prediction accuracy with respect to the conventional GP models. In addition, the integration of GS and HTP data outperformed the predictive ability of the models that include only main effects of the markers.

Nowadays, more studies have shown interest in considering the integration of genomic data with HTP platforms for improving predictive ability of genotypes using data at early stages. Especially, because of the time for collecting data is short and usually there is a need to phenotype thousands of plots attracting the breeder's interest in continuing to explore and adopt these new technologies to achieve the main objectives of plant breeding programs.

However, one of the questions that remains unclear is regarding the suitability of implementing this methodology for selection purposes because in any case the genotypes need to be observed in fields in order to collect the data using the HTP platforms. For example, the breeders could wait until the harvest season to collect the actual phenotype of interest making a little impractical its application. On the other hand, one of the attractive features of GS is that it is not necessary to test candidate genotypes in fields in order to conduct the selection even before the planting season. A potential solution in this case could be to use marker data to predict the data derived of the HTP platforms then incorporate the obtained values as covariables of the prediction model. We hope these issues will be solved in the near future unleashing the full potential of HTP for aiding GS applications.

## REFERENCES

- Al-Wassai FA and Kalyankar NV (2013) Major limitations of satellite images. **Journal of Global Research in Computer Sciences** **4**: 51-59.
- Amani I, Fischer RA and Reynolds MP (1996) Canopy temperature depression association with yield of irrigated wheat cultivars in a hot climate. **Journal of Agronomy and Crop Science** **176**: 119-129.
- Araus JL, Kefauver CS, Zaman-Allah M, Olsen SM and Cairns EJ (2018) Translating high-throughput phenotyping into genetic gain. **Trends in Plant Science** **23**: 451-466.
- Atieno J, Li Y, Langridge P, Dowling K, Brien C, Berger B, Varshney RK and Sutton T (2017) Exploring genetic variation for salinity tolerance in chickpea using image-based phenotyping. **Scientific Reports** **7**: 1300.
- Baek J, Lee E, Kim N, Kim SL, Choi I, Ji H, Chung YS, Choi M-S, Moon J-K, Kim K-H (2020) High throughput phenotyping for various traits on soybean seeds using image analysis. **Sensors** **20**: 248.
- Bai G, Ge Y, Scoby D, Leavitt B, Stoerger V, Kirchgessner N, Irmak S, Graef G, Schnable J and Awada T (2019) NU-Spidercam: A large-scale, cable-driven, integrated sensing and robotic system for advanced phenotyping, remote sensing, and agronomic research. **Computers and Electronics in Agriculture** **160**: 71-81.
- Baluja J, Diago MP, Balda P, Zorer R, Meggio F, Morales F and Tardaguila J (2012) Assessment of vineyard water status variability by thermal and multispectral imagery using an Unmanned Aerial Vehicle (UAV). **Irrigation Science** **30**: 511-522.
- Basso B, Fiorentino C, Cammarano D and Schulthess U (2016) Variable rate nitrogen fertilizer response in wheat using remote sensing. **Precision Agriculture** **17**: 168-182.
- Beauchene K, Leroy F, Fournier A, Huet C, Bonnefoy M, Lorgeou J, de Solan B, Piquemal B, Thomas S and Cohan J-P (2019) Management and characterization of abiotic stress via phénofield®, a high-throughput field phenotyping platform. **Frontiers in Plant Science** **10**: 904. doi: 10.3389/fpls.2019.00904
- Bei R, Cozzolino D, Sullivan W, Cynkar W, Fuentes S, Dambergers R, Pech J and Tyerman S (2011) Non-destructive measurement of grapevine water potential using near infrared spectroscopy. **Australian Journal of Grape and Wine Research** **17**: 62-71.
- Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. **Crop Science** **34**: 20-25.
- Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science** **48**: 1649-1664.
- Brichet N, Fournier C, Turc O, Strauss O, Artzet S, Pradal C, Welcker C, Tardieu F and Cabrera-Bosquet L (2017) A robot-assisted imaging pipeline for tracking the growths of maize ear and silks in a high-throughput phenotyping platform. **Plant Methods** **13**: 96.
- Burgueno J, de los Campos G, Weigel K and Crossa J (2012) Genomic



## The use of high-throughput phenotyping in genomic selection context

- prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. **Crop Science** **52**: 707-719.
- Chaerle L, Leinonen I, Jones HG and Van Der Straeten D (2006) Monitoring and screening plant populations with combined thermal and chlorophyll fluorescence imaging. **Journal of Experimental Botany** **58**: 773-784.
- Chang S, Lee U, Hong MJ, Jo YD and Kim JB (2020) High-throughput phenotyping (HTP) data reveal dosage effect at growth stages in *Arabidopsis thaliana* irradiated by gamma rays. **Plants** **9**: 557.
- Chapman SC, Merz T, Chan A, Jackway P, Hrabar S, Dreccer MF, Holland E, Zheng B, Ling TJ and Jimenez-Berni J (2014) Pheno-copter: a low-altitude, autonomous remote-sensing robotic helicopter for high-throughput field-based phenotyping. **Agronomy** **4**: 279-301.
- Chen D, Neumann K, Friedel S, Kilian B, Chen M, Altmann T and Klukas C (2014) Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. **Plant Cell** **26**: 4636-4655.
- Christensen OF, Madsen P, Nielsen B, Ostersen T and Su G (2012) Single-step methods for genomic evaluation in pigs. **Animal** **6**: 1565-1571.
- Cobb JN, DeClerck G, Greenberg A, Clark R and McCouch S (2013) Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. **Theoretical and Applied Genetics** **126**: 867-887.
- Crain J, Mondal S, Rutkoski J, Singh RP and Poland J (2018) Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. **The Plant Genome** **11**: 170043.
- Crossa J, Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, Banziger M and Braun HJ (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics** **186**: 713-724.
- de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K and Cotes JM (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigrees. **Genetics** **182**: 1 375-385.
- Derry DM, Rebetzke GJ, Jimenez-Berni JA, James RA, Condon AG, Bovill WD, William D, Hutchinson P, Scarrow J, Davy R and Furbank RT (2016) Methodology for high-throughput field phenotyping of canopy temperature using airborne thermography. **Frontiers in Plant Science** **7**: 1808.
- Elshire RJ, Glaubit JC, Sun Q, Poland JA, Kawamoto K, Buckler ES and Mitchell SE (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. **PLoS One** **6**: e19379.
- Fahlgren NA, Gehan M and Baxter I (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. **Plant Biology** **24**: 93-99.
- Friedli M, Kirchgessner N, Grieder C, Liebisch F, Mannale M and Walter A (2016) Terrestrial 3D laser scanning to track the increase in canopy height of both monocot and dicot crop species under field conditions. **Plant Methods** **12**: 9.
- Fróna D, Szenderák J and Harangi-Rákos M (2019) The challenge of feeding the world. **Sustainability** **11**: 5816.
- Galán RJ, Bernal-Vasquez AM, Jebsen C, Piepho HP, Thorwarth P, Steffan P, Gordillo A and Miedaner T (2020) Integration of genotypic, hyperspectral, and phenotypic data to improve biomass yield prediction in hybrid rye. **Theoretical and Applied Genetics** **133**: 3001-3015.
- Gonçalves MTV, Morota G, Costa PMdA, Vidigal PMP, Barbosa MHP and Peternelli LA (2021) Near-infrared spectroscopy outperforms genomics for predicting sugarcane feedstock quality traits. **PLoS One** **16**: e0236853.
- Haghighattalab A, González Pérez L, Mondal S, Singh D, Schinostock D, Rutkoski J, Ortiz-Monasterio I, Prakash Singh R, Goodin D and Poland J (2016) Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. **Plant Methods** **12**: 35.
- Hamada Y, Stow DA, Coulter LL, Jafolla JC and Hendricks LW (2007) Detecting tamarisk species (*Tamarix* spp.) in riparian habitats of southern California using high spatial resolution hyperspectral imagery. **Remote Sensing of Environment** **109**: 237-248.
- He T and Li C (2020) Harness the power of genomic selection and the potential of germplasm in crop breeding for global food security in the era with rapid climate change. **The Crop Journal** **8**: 688-700.
- Huang W, Li L, Myers RJ and Marth TJ (2012) ART: a next-generation sequencing read simulator. **Bioinformatics** **28**: 593-594.
- Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, Piroux F, Guerreiro L, Perez P, Calus M, Burgueño 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.
- Kim DW, Yun HS, Jeong SJ, Kwon YS, Kim SG, Lee WS and Kim HJ (2018) Modeling and testing of growth status for Chinese cabbage and white radish with UAV-based RGB imagery. **Remote Sensing** **10**: 563-587.
- Kim J, Kim K-S, Kim Y and Chung YS (2020) A short review: Comparisons of high-throughput phenotyping methods for detecting drought tolerance. **Scientia Agricola** **78**: e20190300300.
- Kirchgessner N, Liebisch F, Yu Kang, Pfeifer J, Friedli M, Hund A and Walter A (2016) The ETH field phenotyping platform FIP: a cable-suspended multi-sensor system. **Functional Plant Biology** **44**: 154-168.
- Kismiantini, Montesinos-López OA, Crossa J, Setiawan EP and Wutsqa DU (2021) Prediction of count phenotypes using high-resolution images and genomic data. **G3 Genes|Genomes|Genetics** **11**: jkab035.
- Krause MR, Mondal S, Crossa J, Singh PR, Pinto F, Haghighattalab A, Shrestha S, Rutkoski J, Gore AM, Sorrells EM and Poland J (2020)

- Aerial high-throughput phenotyping enables indirect selection for grain yield at the early generation, seed-limited stages in breeding programs. **Crop Science** **60**: 3096-3114.
- Krause RM, González-Pérez L, Crossa J, Pérez-Rodríguez P, Montesinos-López O, Singh PR, Dreisigacker S, Poland J, Rutkoski J, Sorrells M, Gore MA and Mondal S (2019) Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. **G3 Genes | Genomes | Genetics** **9**: 1231-1247.
- Krchov LM and Bernardo R (2015) Relative efficiency of genome-wide selection for testcross performance of doubled haploid lines in a maize breeding program. **Crop Science** **55**: 2091-2099.
- Li D, Quan C, Song Z, Li X, Yu G, Li C and Muhammad A (2021) High-throughput plant phenotyping platform (HT3P) as a novel tool for estimating agronomic traits from the lab to the field. **Frontiers in Bioengineering Biotechnology** **8**: 623705.
- Liew OW, Chong PC, Li B and Asundi AK (2008) Signature optical cues: Emerging technologies for monitoring plant health. **Sensors** **8**: 3205-3239.
- Lin Y (2015) Lidar: an important tool for next-generation phenotyping technology of high potential for plant phenomics? **Computers and Electronics in Agriculture** **119**: 61-73.
- Lozada DN, Godoy JV, Ward BP and Carter AH (2020) Genomic prediction and indirect selection for grain yield in US pacific northwest winter wheat using spectral reflectance indices from high-throughput phenotyping. **International Journal of Molecular Sciences** **21**: 165.
- Mardis ER (2008) The impact of next-generation sequencing technology on genetics. **Trends in Genetics** **24**: 133-141.
- Meuwissen TH, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. **Genetics** **157**: 1819-1829.
- Miao C, Pages A, Xu Z, Rodene E, Yang J and Schnable JC (2020) Semantic segmentation of sorghum using hyperspectral data identifies genetic associations. **Plant Phenomics** **2020**: 4216373.
- Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR and Sasaki T (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. **Molecular Breeding** **3**: 87-103.
- Montesinos-López A, Montesinos-López OA, Cuevas J, Mata-López WA and Burgueño J (2017b) Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyperspectral reflectance. **Plant Methods** **13**: 62.
- Montesinos-López A, Montesinos-López OA, de los Campos G, Crossa J, Burgueño J and Luna-Vazquez FJ (2018) Bayesian functional regression as an alternative statistical analysis of high-throughput phenotyping data of modern agriculture. **Plant Methods** **14**: 46.
- Montesinos-López OA, Montesinos-López A, Crossa J, los Campos G, Alvarado G, Suchismita M, Rutkoski J, González-Pérez L and Burgueño J (2017a) Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. **Plant Methods** **13**: 4.
- Moreira FF, Oliveira HR, Volenec JJ, Rainey KM and Brito LF (2020) Integrating high-throughput phenotyping and statistical genomic methods to genetically improve longitudinal traits in crops. **Frontiers in Plant Science** **11**: 681.
- Neilson E, Edwards AM, Blomstedt C, Berger B, Møller BL and Gleadow RM (2015) Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. **Journal of Experimental Botany** **66**: 1817-1832.
- Neumann K, Klukas C, Friedel S, Rischbeck P, Chen D, Entzian A, Stein N, Graner A and Kilian B (2015) Dissecting spatiotemporal biomass accumulation in barley under different water regimes using high-throughput image analysis. **Plant Cell & Environment** **38**: 1980-1996.
- Peirone LS, Pereyra Irujo GA, Bolton A, Erreguerena I and Aguirrezábal LA (2018) Assessing the efficiency of phenotyping early traits in a greenhouse automated platform for predicting drought tolerance of soybean in the field. **Frontiers in Plant Science** **9**: 587.
- Perez-Rodríguez P, Crossa J, Rutkoski J, Poland J, Singh R, Legarra A, Autrique E, de los Campo G, Burgueno J and Dreisigacker S (2017) Single-step genomic and pedigree genotype × environment interaction models for predicting wheat lines in international environments. **The Plant Genome** **10**(2): 1-15.
- Qiu Q, Sun N, Bai H, Wang N, Fan Z, Wang Y, Meng Z, Li B and Cong Y (2019) Field-based high-throughput phenotyping for maize plant using 3D LiDAR point cloud generated with a "Phenomobile". **Frontiers in Plant Science** **10**: 554.
- Raesch A, Muller O, Pieruschka R and Rascher U (2014) Field observations with Laser-Induced Fluorescence Transient (LIFT) method in barley and sugar beet. **Agriculture** **4**: 159-169.
- Rincen R, Charcosset A and Moreau L (2017) Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured populations. **Theoretical and Applied Genetics** **130**: 2231-2247.
- Silva CA, Abreu ADFB, Ramalho MAP and Maia LGS (2012) Chemical composition as related to seed color of common bean. **Crop Breeding and Applied Biotechnology** **12**: 132-137.
- Singh D, Xu W, Uttam K, Liangliang G, Muhammad N, Muhammad I, Singh RP and Poland JA (2019) High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. **Frontiers in Plant Science** **10**: 394.
- Sun J, Rutkoski JE, Poland JA, Crossa J, Jannink J and Sorrells ME (2017) Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. **The Plant Genome** **10**(2): 1-12.
- Susko AQ, Gilbertson F, Heuschele DJ, Smith K and Marchetto P (2018) An automatable, field camera track system for phenotyping crop lodging and crop movement. **HardwareX** **4**: e00029.
- Tisné S, Serrand Y, Bach L, Gilbault E, Ben Ameer R, Balasse H, Voisin R, Bouchez D, Durand-Tardif M, Guerche P, Chareyron G, Da Rugna J,

## The use of high-throughput phenotyping in genomic selection context

- Camilleri C, Loudet O (2013) Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity. **The Plant Journal** **74**: 534-544.
- Vadez V, Kholova J, Hummel G, Zhokhavets U, Gupta SK and Hash CT (2015) LeasyScan: a novel concept combining 3D imaging and lysimetry for high-throughput phenotyping of traits controlling plant water budget. **Journal of Experimental Botany** **66**: 5581-5593.
- Virlet N, Sabermanesh K, Sadeghi-Tehran P and Hawkesford MJ (2017) Field scanalyzer: an automated robotic field phenotyping platform for detailed crop monitoring. **Functional Plant Biology** **44**: 143-153.
- Wan L, Cen H, Zhu J, Zhang J, Zhu Y, Sun D, Du X, Zhai L, Weng H, Li Y, Li X, Bao Y, Shou J and He Y (2020) Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer – a case study of small farmlands in the South of China. **Agricultural and Forest Meteorology** **291**: 108096.
- Wang X, Singh D, Marla S, Morris G and Poland J (2018) Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. **Plant Methods** **14**: 53.
- Watanabe K, Guo W, Arai K, Takanashi H, Kajiya-Kanegae H, Kobayashi M, Yano K, Tokunaga T, Fujiwara T, Tsutsumi N and Iwata H (2017) High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. **Frontiers in Plant Science** **8**: 421.
- White J, Andrade-Sanchez P, Gore MA, Bronson KF, Coffelt TA, Conley MM, Feldmann KA, French AN, Heun JT, Hunsaker DJ, Jenks MA, Kimball BA, Roth RL, Strand RJ, Robbins KT, Wall GW and Wang G (2012) Field-based phenomics for plant genetics research. **Field Crops Research** **133**: 101-112.
- Yang K, Jeong N, Moon JK, Lee YH, Lee SH, Kim HM, Hwang CH, Back K, Palmer RG and Jeong SC (2010) Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean. **Journal of Heredity** **101**: 757-768.
- Yue JB, Yang GJ, Tian QJ, Feng HK, Xu KJ and Zhou CQ (2019) Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. **ISPRS Journal Photogrammetry and Remote Sensing** **150**: 226-244.