University of São Paulo
"Luiz de Queiroz" College of Agriculture

From pixel to knowledge: how high-throughput phenotyping helps to dissect the genetic architecture and improves predictive ability in maize under inoculation with plant growth-promoting bacteria

Rafael Massahiro Yassue

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2022

Rafael Massahiro Yassue
Agronomist

From pixel to knowledge: how high-throughput phenotyping helps to dissect the genetic architecture and improves predictive ability in maize under inoculation with plant growth-promoting bacteria

Advisor:
Prof. Dr. ROBERTO FRITSCHE NETO

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2022

Dados Internacionais de Catalogação na Publicação

DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Yassue, Rafael Massahiro

From pixel to knowledge: how high-throughput phenotyping helps to dissect the genetic architecture and improves predictive ability in maize under inoculation with plant growth-promoting bacteria / Rafael Massahiro Yassue. - - Piracicaba, 2022.

117 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Fenômica 2. GWAS 3. Predição genômica 4. Aprendizagem de máquina 5. Multispectral 6. Shovelomics I. Título

DEDICATÓRIA

À agricultura Brasileira.

Dedico

AGRADECIMENTOS

À minha família, aos meus pais e aos meus irmãos pelo apoio e amor durante toda minha vida.

À minha noiva, Diandra, por todo incentivo, carinho e companheirismo.

Ao meu orientador, Professor Dr. Roberto Fritsche Neto, por acreditar, incentivar e ensinar durante toda jornada do doutorado.

Aos Professores Dr. Gota Morota e Dr. James Chen por me receberem durante meu intercâmbio. Agradeço a oportunidade e pelos ensinamentos.

À Escola Superior de Agricultura "Luiz de Queiroz" e a Virginia Polytechnic Institute and State University pela minha formação acadêmica.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNpq) pelo apoio financeiro durante o período do doutorado.

Aos meus amigos do Laboratório de Melhoramento de Plantas Alógamas e do The Lab of Animal Data Sciences pelas discussões, ensinamentos, suporte durante as horas dificeis e pelos bons momentos compartilhados.

Aos amigos do GVENCK que proporcionaram ótimos momentos juntos.

CONTENTS

RESUMO

Do pixel a informação: como fenotipagem de alto rendimento pode avaliar a arquitetura genética e melhor a capacidade preditiva em milho sob inoculação de bactérias promotoras de crescimento de plantas?

Bacterias promotoras de crescimento de plantas (BPCP) podem ter um papel crucial no futuro da agricultura devido a sua capacidade de promover o crescimento de plantas, sem causar nenhum tipo de dano ambiental. Além disso, BPCP possuem a capacidade de aumentar a resiliência do seu hospedeiro contra estresses bióticos e abióticos, além de promover o aumento da absorção de nutrientes. No entanto, poucos trabalhos estudaram a arquitetura genética da resposta ao BPCP. Outro campo emergente é a fenotipagem de alto rendimento (FAR) que pode ser usada para melhorar a avaliação dos novos fenótipos e ser integrada em estudos de arquitetura genética. Com base nisso, estudamos a arquitetura genética da resposta as BPCP usando um painel público de associação de milho tropical contendo 360 linhagens genotipadas usando a metodologia *genotype-by-sequence* com um total de 13.826 polimorfismos de nucleotídeo único (SNPs). Para as avaliações foram utilizadas câmeras RGB, multi e hiperespectral, além dos fenótipos tradicionais. Além disso, desenvolvemos uma plataforma de FAR de baixo custo para experimentos em casa de vegetações. No trabalho são discutidos vários modelos single, multi-trait e de aprendizado de máquina, e suas aplicações no contexto de estudos genéticos. Coletivamente, nossos resultados revelam a utilidade do BPCP no aumento da resiliência das plantas e as aplicações dos fenótipos FAR em estudos genéticos para dissecar a arquitetura genética e melhorar a acurácia em modelos preditivos.

Palavras-chave: Fenômica, GWAS, Predição genômica, Aprendizado de máquina, Hiperspectral, Multispectral, Shovelomics

ABSTRACT

From pixel to information: how high-throughput phenotyping can assess the genetic architecture and improve predictive ability in maize under inoculation with plant growth-promoting bacteria?

Plant growth-promoting bacteria (PGPB) may play an important role in the agriculture in the future due to the ability of these bacteria in promote growth without causing any type of environmental damage. Besides, they can increase the plant resilience against biotic and abiotic stress and improve nutrient uptake. Nevertheless, only a few works have studied the genetic architecture of the response to PGPB. Another emerging field is the high-throughput phenotyping (HTP) which can be used to improve the assessment of the new phenotypes and be integrated in genetics studies. Based on this, we study the genetic architect of the response to PGPB using a public tropical association panel containing 360 inbreeds lines genotyped using genotype-by-sequence methodology with 13,826 single-nucleotide polymorphisms using RGB, multi, and hyperspectral cameras, besides the traditional phenotypes. Also, we develop a low-cost HTP platform for greenhouses experiments. In addition, several single-trait, multi-trait, machine learning models and its application in the context of genetics studies is discussed. Collectively, our results reveal the usefulness of PGPB in increase plant resilience and the applications of HTP phenotypes in genetics studies to dissect the genetic architecture and improve the accuracy in predictive models.

Keywords: Phenomics, GWAS, Genomic prediction, Machine learning, Hyperspectral, Multispectral images, Shovelomics

1. INTRODUCTION

Producing more food while consuming fewer natural resources is the next-generation challenge. Currently, one of the biggest gaps in agriculture is the need of adding external input for nutrients, such as nitrogen, phosphorus, and potassium. Several studies have shown the potential of using plant growth-promoting bacteria (PGPB) to increase plant biomass and yield while improving nutrient update and fixing nitrogen. Even though the potential of PGPB in the agriculture, its applications remain restricted to Fabaceae species such as soybean. The strong association between soybean and some PGPB species is likely due to the co-evolution between these species turning them highly co-dependent. In maize, studies have shown that some landraces can fix between 29-82% of the plant nitrogen through diazotrophic microbiota revealing the potential of this technology in other species (Van Deynze et al. 2018).

One of the many possible causes for not the wide use of PGPB in agriculture for all crops is the highest intention between PGPB, environmental, and the host genotype impacting the reproducibility of the results. Even though, these highly complex interactions open doors for plant breeding studies to leverage the host-dependent performance and improve specific new cultivars that may have a higher and more stable response to PGPB (Wintermans et al. 2016).

In order to develop a plant breeding strategy, two major concerns must be addressed in order to optime genetic gains. It is necessary to develop phenotyping technics appropriate for PGPB evaluation and study the genetic architecture of the response to the PGPB. In both cases, high-throughput phenotyping (HTP) is a valuable tool that the breeder can explore (Araus et al., 2018). Even though the advance in the use of cameras and sensors in agriculture, there is still a lack of low-cost alternatives (Du et al., 2021). The most common HTP platform is the unmanned aerial vehicle (UAV)-based approach likely due to the lower cost of these technologies and consolidated image processing pipeline. For greenhouse experiments, usually, the available platforms are commercial, protected by patents, and high cost making it unfeasible for small plant breeding companies and research institutions.

The nature of the HTP data makes challenging to incorporate into genetics studies because traditionally, genetic quantitative studies have focused only on one or few traits. In addition, the ability of the cameras and sensors to quickly evaluate and

non-destructively the plants make it possible to assess the plant development across time. Hundreds to thousands of phenotypes across the plant's development make it challenging for the breeder to make inferences about these new phenotypes. Nevertheless, these phenotypes could be valuable information with the potential to be used in genetic architecture studies and improve the prediction of target traits (Feng et al., 2017, and Krause et al. 2019).

Based on the above, the goals of this work are: i) study the genetic architecture of the response to the plant growth-promoting bacteria in a public tropical maize germplasm, ii) develop a low-cost high-throughput phenotyping platform for green-house experiments and image processing pipeline, iii) Compare the genomic and phenomic prediction using hyperspectral camera and several models in the context of PGPB inoculation studies, iv) Integrate hyperspectral image data in genetic architecture studies in maize under inoculation of PGPB.

References

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. Trends in Plant Science, 23(5), 451- 466.

Du, J., Fan, J., Wang, C., Lu, X., Zhang, Y., Wen, W., Liao, S., Yang, X., Guo, X., & Zhao, C. (2021). Greenhouse-based vegetable high-throughput phenotyping platform and trait evaluation for large-scale lettuces. Computers and Electronics in Agriculture, 186, 106193.

Feng, H., Guo, Z., Yang, W., Huang, C., Chen, G., Fang, W., Xiong, X., Zhang, H., Wang, G., Xiong, L., and Liu, Q. (2017). An integrated hyperspectral imaging and genome-wide association analysis platform provides spectral and genetic insights into the natural variation in rice. Scientific Reports, 7(1)

Krause, M. R., Gonzalez-P erez, L., Crossa, J., Perez-Rodr 谋 guez, P., Montesinos-Lopez, O., Singh, R. P., Dreisigacker, S., Poland, J., Rutkoski, J., Sorrells, M., Gore, M. A., and Mondal, S. (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. G3 Genes|Genomes|Genetics, 9(4):1231-1247

Van Deynze A, Zamora P, Delaux P-M et al (2018) Nitrogen fixation in a landrace of maize is supported by a mucilage-associated diazotrophic microbiota. PLOS Biology 16:e2006352

Wintermans PCA, Bakker PAHM, Pieterse CMJ (2016) Natural genetic variation in Arabidopsis for responsiveness to plant growth-promoting rhizobacteria. Plant Mol Biol 90:62-634

2. ON THE GENETIC ARCHITECTURE IN A PUBLIC TROPICAL MAIZE PANEL OF THE SYMBIOSIS BETWEEN CORN AND PLANT GROWTH-PROMOTING BACTERIA AIMING TO IMPROVE PLANT RESILIENCE

Rafael Massahiro Yassue[1], Humberto Fanelli Carvalho[1], Raysa Gevartosky[1], Felipe Sabadin[1], Pedro Henrique Souza[1], Maria Leticia Bonatelli[1], João Lúcio Azevedo[1], Maria Carolina Quecine[1], Roberto Fritsche-Neto[1]

[1]Department of Genetics, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil

Abstract

Exploring the symbiosis between plants and plant-growth-promoting bacteria (PGPB) is a new challenge for sustainable agriculture. Even though many works have reported the beneficial effects of PGPB in increasing plant resilience for several stresses, its potential is not yet widely explored. One of the many reasons is the differential symbiosis performance depending on the host genotype. This opens doors to plant breeding programs to explore the genetic variability and develop new cultivars with higher responses to PGPB interaction and, therefore, have higher resilience to stress. Hence, we aimed to study the genetic architecture of the symbiosis between PGPB and tropical maize germplasm, using a public association panel and its impact on plant resilience. Our findings reveal that the synthetic PGPB population can modulate and impact root architecture traits, improve resilience to nitrogen stress, and 37 regions were significant for controlling the symbiosis between PGPB and tropical maize. In addition, we found two overlapping SNPs in the GWAS analysis indicating strong candidates for further investigations. Furthermore, genomic prediction analysis with genomic relationship matrix computed using only significant SNPs obtained from GWAS analysis substantially increased the predictive ability for several traits endorsing the importance of these genomic regions for the response of PGPB. Finally, the public tropical panel reveals a significant genetic variability to the symbiosis with the PGPB and can be a source of alleles to improve plant resilience.

KEYWORDS: Shovelomics; Root architecture; GWAS; Symbiosis interaction; Genomic prediction

DECLARATIONS

References

Yassue, R.M., Carvalho, H.F., Gevartosky, R. *et al*. On the genetic architecture in a public tropical maize panel of the symbiosis between corn and plant growth-promoting bacteria aiming to improve plant resilience. *Mol Breeding* **41,** 63 (2021). https://doi.org/10.1007/s11032-021-01257-6

Funding and acknowledgment

This study was financed in part by São Paulo Research Foundation (FAPESP, Process: 19/04697-2; 17/24327-0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Conflict of interest

The authors declare no conflict of interest.

Availability of data and material

Genomic data: https://data.mendeley.com/datasets/5gvznd2b3n

Code availability

Code for image processing: https://github.com/RafaelYassue/Root-phenotyping

Ethics approval

Not applicable because this article does not contain any studies with human or animal subjects.

Consent to participate

Not applicable because this article does not contain any studies with human or animal subjects.

Consent for publication

All authors have approved the manuscript and agreed with submission to the Molecular Breeding journal.

2.1. Introduction

Due to the growing need for food (Ray et al. 2013) and environmental pressure (Qi et al. 2018), new approaches that increase production in a sustainable way are required (Gaffney et al. 2019). Tropical agriculture will have to rise to meet the food demand in tropical developing nations (Laurance et al. 2014). Also, studies estimate that it will be necessary to increase the use of fertilizers, with emphasis on N, P_2O_5, and K_2O (Pradhan et al. 2015). Furthermore, multidisciplinary research that aims to increase production, sustainability, and plant resilience is justified.

The use of plant growth-promoting bacteria (PGPB) has been a promising field of interest due to the ability to increase production (Martins et al. 2018) and the resilience of the host caused by direct and indirect mechanisms. The most common mechanism is related to biofertilization, which consists of nutrient uptake and hormones production as well as the ones related to the improvement of plant defense (reviewed by (Santoyo et al. 2016; Vejan et al. 2016)). Many studies have proved the benefits of PGPB reducing the plant abiotic stresses caused by salinity (Rojas-Tapias et al. 2012), heavy metals (Gamalero et al. 2009), and drought (Sandhya et al. 2010). However, the use of PGPB as inoculants in agriculture is still incipient.

One of the main challenges of an inoculant with PGPB is the reproducibility of its results in the field (Bashan et al. 2014). Studies have shown that PGPB is influenced by several factors, such as soil type (Egamberdiyeva 2007), nitrogen fertilization (Rodríguez-Blanco et al. 2015), microbe-microbe interaction (Gaiero et al. 2013), and plant-PGPB interactions (Wintermans et al. 2016). In this sense, during the development of a new PGPB inoculant, both environmental and genetic factors must be considered (Lemanceau et al. 2017).

In recent years, the genotyping cost has decreased substantially, and its benefits have become consolidated. Currently, one of the gaps for greater gains using genotyping is the poor ability to phenotyping. High-throughput phenotyping (HTP) is a suitable alternative to increasing phenotyping power (Araus and Cairns 2014). One of the benefits of HTP is measuring secondary traits with direct or indirect effects on primary characteristics (Qiao et al. 2019). For roots, shovelomic analysis has been used to study the root system architecture traits and their impact on the final phenotype (Trachsel et al. 2011).

Root architecture is responsible for the nutrients and water uptake and its evaluations and selection have been the focus of plant breeding for many years (York et al. 2015). In maize, root traits have been identified as key phenotypes to overcome stress in specific environments (Mi et al. 2010; Lynch 2013; Adebayo et al. 2020). Also, in PGPB studies, the root and its architecture play a significant role in the direct interaction with the soil microbiome (Compant et al. 2010). Studies have also revealed that PGPB can modulate root architecture (Gutiérrez-Luna et al. 2010), and the host genotype can influence this response (Wintermans et al. 2016).

Modern plant breeding may have caused a bottleneck in genetic diversity for the symbiosis with PGPB due to the lower response in modern varieties (Valente et al. 2020). In maize, landraces reveal the ability to fix 29%-82% of the plant nitrogen through interaction with PGPB (Van Deynze et al. 2018). Additive and dominance effects have been reported (Vidotti et al. 2019b; Wagner et al. 2020), providing insights into the importance of the host genotype in the impact of PGPB response. GWAS studies have been employed to discover candidate genes to the response of PGPB in *Arabidopsis* (Wintermans et al. 2016; Proietti et al. 2018; Cotta et al. 2020) and maize (Vidotti et al. 2019a).

Based on the genetic variability for the response of PGPB inoculation in many crops, the response of maize to the symbiosis of PGPB may be improved with the support of plant breeding (Kroll et al. 2017; Wei and Jousset 2017) and contribute to more sustainable food production through the increase of yield, sustainability, and plant resilience. Therefore, we aimed to study the symbiosis's genetic architecture between the tropical maize germplasm and a synthetic population of plant growth-promoting bacteria. Also, we present a pipeline for shovelomics evaluation and analysis. Hence, this information may contribute to plant breeding programs, focusing on new strategies to produce more resilient crops.

2.2 Materials and Methods

Public tropical maize panel

Our tropical maize germplasm panel contains 360 inbred lines used to analyze the symbiosis's genetic architecture between maize and PGPB. Among them, 179 inbred lines are from ESALQ-USP (Luiz de Queiroz College of Agriculture-University of Sao Paulo) and 181 from IAPAR (Instituto de Desenvolvimento Rural do

Paraná). The genomic and phenotypic information about this panel is available on the Mendeley platform (https://data.mendeley.com/datasets/5gvznd2b3n).

Bacterial strain and inoculum

The PGPB *Bacillus thuringiensis* RZ2MS9 and *Delftia* sp. RZ4MS18 was isolated from *Paullinia cupana* (Batista et al. 2018, Batista et al. 2021), *Pantoea agglomerans* 33.1 was isolated from *Eucalyptus grandis* (Quecine et al. 2012), and *Azospirillum brasilense* Ab-v5 is a commercial inoculant (Hungria et al. 2010). They were selected based on previous studies that have reported them as potential inoculants (Batista et al. 2018, Quecine et al. 2012, Hungria et al. 2010) to compose the synthetic population. *In vivo* trials revealed that these PGPB did not have antagonistic effects among each other and when co-inoculated, promoted growth in maize (unpublished data).

The synthetic population inoculum was prepared by growing each bacterium individually in Luria-Bertani (LB) medium at 28°C with 150 rpm agitation for 24h. The concentration of each bacterium was measured in a spectrophotometer. The synthetic population was composed of each bacteria's adjusted culture medium containing approximately 10^8 colony-forming units/mL. The treatment without PGPB consisted of preparing the inoculum with liquid LB only. Each plot containing three seeds was individually inoculated with 1 ml of the respective treatments, agitated, and sown afterward.

Greenhouse experiment

The experiments were carried out under greenhouse conditions at Luiz de Queiroz College of Agriculture (ESALQ/USP), Brazil (22˚42'39 "S; 47˚38'09 "W, altitude 540 m). The 360 inbred lines were evaluated in two experiments: with (B+) and without (B-) PGPB inoculation. Each experiment was conducted in an augmented blocks design with two replicates across time, each one consisting of six blocks with 60 inbred lines and three common checks. Each replicate of B+ and B- experiments were installed together in a greenhouse. The treatments B+ and B- were physically separated due to the ability of the PGPB to migrate from one to another (Chi et al. 2005; Ji et al. 2010). Furthermore, we calculated the difference between the treatments B+ and B- to compound the Delta. Lastly, we performed analyzes

considering the values from B+ and B- individually and the Delta value as a response to the inoculation.

In order to evaluate the resilience to N stress and identify genomic regions responsible for the symbiosis between PGPB and maize, we tested the genotypes with and without inoculation with PGPB (B+ and B-, respectively) in low nitrogen conditions similar to (Vidotti et al. 2019a). The low nitrogen condition consisted of no external nitrogen input, and all the nitrogen available to the plants was due to the natural soil organic matter or fixed from PGPB.

The maize plants were grown in 3-L plastic pots containing soil. Chemical and physical soil analysis is available in supplementary table 1. The planting fertilization was done according to soil nutrient content and crop demands provided by Soil-app (Matias et al. 2021). It consisted of potassium chloride, simple superphosphate, and limestone soil conditioner inputs added and mixed into the soil. Each plot was sown with three seeds, and after germination, the seedlings were thinning to only one. During the experiments, temperature, radiation, and humidity were monitored and are available in the supplemental figure 1. Twice a week, 200 ml of a complementary fertilizer without nitrogen and adapted from Hoagland and Snyder (1933) were applied in each plot. Irrigation and other cultural practices were carried out according to the needs of the crop.

Evaluations began when most of the plants were in the V6 growth stage (six expanded leaves). Plant height (PH, cm) was measured from the soil to the last expanded leaf's collar, and the number of expanded leaves was counted (NL). Afterward, the plants were cropped at the soil base, and the stalk diameter was measured using a digital caliper (SD, mm). Finally, the harvested shoot (leaves and stalk) was dried in a forced draft oven at 60°C for 72h to obtain shoot dry mass (SDM, g).

The roots were carefully washed with water, and each root was stored in plastic pots with a 25% ethanol solution for preservation. First, root images were taken to calculate root angle (RA, degree, °) and convex hull area (CHA, cm^2) using a Nikon CoolPix S8100 camera attached to a platform with a fixed height and position. For RA, the images were cropped in order to consider the first 10 centimeters representing the topsoil. These images were processed using a Python script that is available on GitHub (https://github.com/RafaelYassue/Root-phenotyping). Then, new root images were acquired by an Epson LA2400 scanner and processed using the WinRHIZO (Reagent

Instruments Inc., Quebec, Canada) to obtain lateral and axial root volume (LRV, ARV, cm^3, respectively), root length (RL, cm), and root average diameter (RAD, mm). Representative images of root phenotypes are available in Supplementary figure 5. Next, the roots were dried out to determine the root dry mass (RDM, g). Furthermore, the ratio of shoot/root (RSR, $g\ g^{-1}$) was obtained by dividing the SDM by the RDM.

Phenotypic analysis

To test the interaction between genotype and treatment (B+ and B-) we used the followed full model:

$$\mathbf{y} = \mathbf{X}_1\mathbf{t} + \mathbf{X}_2\mathbf{NL} + \mathbf{Z}_1\mathbf{r} + \mathbf{Z}_2\mathbf{b} + \mathbf{Z}_3\mathbf{g} + \mathbf{Z}_4\mathbf{gt} + \varepsilon \quad \text{Eq. 1}$$

where $\mathbf{y}$ refers to the phenotypic observation vector, $\mathbf{X}_1$ and $\mathbf{X}_2$ are the incidence matrix for the fixed effect, $\mathbf{Z}_1$, $\mathbf{Z}_2$,$\mathbf{Z}_3$, and $\mathbf{Z}_4$ are the incidence matrices for the random effects. $\mathbf{t}$ is the fixed effects of treatment (B+ and B-); $\mathbf{r}$ is random effects of replicates, where $\mathbf{r} \sim N(\mathbf{o}, \mathbf{I}\sigma^2_r)$; $\mathbf{b}$ is random effects of the block within replicates, where $\mathbf{b} \sim N(\mathbf{o}, \mathbf{I}\sigma^2_b)$; $\mathbf{g}$ is the vector of random effects of genotype values, where $\mathbf{g} \sim N(\mathbf{o}, \mathbf{I}\sigma^2_g)$; $\mathbf{gt}$ is the vector of random effects of the interaction between genotype and treatment, where $\mathbf{gt} \sim N(\mathbf{o}, \mathbf{I}\sigma^2_{gt})$; and ε is the random residual effects, where $\varepsilon \sim N(\mathbf{o}, \mathbf{I}\sigma^2_e)$. The random effects were tested using the likelihood ratio test (LRT) and the fixed effects using the Wald test. In order to correct the germination and seed vigor differences, the number of leaves ($\mathbf{NL}$) was used as a covariable. Spatial analysis and unstructured residual effects were tested, and a good fit of the model was not reached.

After, we obtained the entry-mean heritabilities for each combination of trait and treatment using the reduced model for each treatment (B+, B-, and Delta):

$$\mathbf{y} = \mathbf{X}_1\mathbf{NL} + \mathbf{Z}_1\mathbf{g} + \mathbf{Z}_2\mathbf{r} + \mathbf{Z}_3\mathbf{b} + \varepsilon \quad \text{Eq. 2}$$

The heritabilities using the full and reduced model (B+, B-, Delta) were calculated at the entry-mean level with the variance components from Eq. 1 and Eq. 2 using the following equations, respectively:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gt}^2}{t} + \frac{\sigma_t^2}{tr}} \quad \text{Eq. 3}$$

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{r}} \quad \text{Eq. 4}$$

In which h^2 refers to the mean-entry heritability, σ^2_g, σ^2_{gt}, and σ^2_e are the variance components due to genotype, the interaction between genotype x treatment, and residual effects, respectively, and r and t is the number of replicates (r = 2) and treatment (t = 2), respectively. The analysis was performed using ASReml-R 4.0 (Butler et al. 2017).

Principal component analysis and Pearson correlation were performed to understand the association between traits. The interaction plot was performed using the R package Raincloud plots (Allen et al. 2021). The Levene's test was used to assess the equality of variances for treatment effects. To simplify a three-way interaction for genotype x treatment x traits, we used a reduced model considering treatment effects as a standardized mean of each treatment (B+, B-, Delta) for each trait. Then, a treatment by trait biplot (TT) was used to study the associations of traits across treatment. A two-way table containing the standardized mean of each trait in each treatment was used to obtain the singular value decomposition and the biplot using the R package metan (Olivoto and Lúcio 2020). TT biplot analysis was adapted from genotype by trait biplot, and a detailed explanation about the analysis can be found in Yan et al (2018).

Genotypic data

All 360 tropical inbred lines were genotyped using a genotyping-by-sequencing (GBS) method following the two enzymes (*Pst*I and *Mse*I) protocol (Sim et al. 2012; Poland et al. 2012). For that, the DNA of tropical maize lines was extracted from young and healthy leaves using the CTAB protocol (Doyle and Doyle 1987). Individual genomic DNA samples were digested by restriction enzymes, and samples were included in a sequencing plate. The sequencing was performed on the Illumina NextSeq 500 platform (Illumina Inc., San Diego, CA, United States). Sequence data were aligned against the B73 (B73-RefGen_v4) maize reference genome, and the SNP calling was performed using the software TASSEL 5.0 (Bradbury et al. 2007) under default parameters values. The SNP dataset was filtered, and markers with call rate < 90%, non-biallelic, minor allele frequency (MAF) lower than 5%, and heterozygous loci on at least one individual were removed from the dataset, and remaining missing data

were imputed by Beagle 5.0 algorithm (Browning et al. 2018). In addition, markers with pairwise linkage disequilibrium (LD) higher than $r^2 > 0.99$ were removed using the SNPRelate package (Zheng et al. 2012). Finally, a total of 13,826 SNPs were considered for the genomic analyses. The coverage and depth of the genotyping-by-sequencing data are shown in supplementary figure 6.

Population structure and LD decay

The principal components analysis (PCA) regarding the population structure was calculated based on the additive genomic relationship matrix (VanRaden 2008). To identify the most likely number of groups (K) in our panel, we estimated the optimal K-value based on the inferred number of groups producing the lowest cross-validation error using the software ADMIXTURE (Alexander and Lange 2011). We ran the software assuming 2–50 subpopulations using default parameters. We estimated the LD (r^2) between all SNP within a distance lower than 1 Mbp in the same chromosome, and r^2 values were plotted against base-pair distance to obtain the LD decay by chromosome. A second-degree locally weighted scatterplot smoothing (LOESS) function was used to draw a trend line for detecting the LD decay with a threshold of 0.1 (Esteras et al. 2013, Zhu et al. 2008).

GWAS analyses

The GWAS analyses were carried out for each combination of trait and inoculation (B+ and B-) and the difference between B+ and B- (Delta) using the FarmCPU R package (Liu et al. 2016). We tested models containing 0 to 6 principal components to correct the population structure effect, and the best model fit was based on QQplot. The Manhattan and QQ plots were obtained using the CMplot R package (Yin et al. 2020). We used the *FarmCPU.P.Threshold* function with 100 permutations to obtain a *p* threshold for each trait. The significant SNPs were annotated using a windows range upstream and downstream, based on the LD decay of respective chromosomes (Supplementary figure 8). The MaizeMine V1.3. was used to obtain the genes on that window (Shamimuzzaman et al. 2020).

Genomic prediction

Genomic predictions (GP) analyses were conducted for each combination of trait x treatment (B+, B-, and Delta). Due to the non-significant effects of the interaction, we did not consider interaction models for the GP analysis. For GP, we used the following three models:

$\hat{\mathbf{g}} = 1u + \mathbf{Za} + \varepsilon$ GBLUP

$\hat{\mathbf{g}} = 1u + \mathbf{Za} + \mathbf{SNPs} + \varepsilon$ GBLUP_MAS

$\hat{\mathbf{g}} = 1u + \mathbf{Za} + \varepsilon$ GBLUP_GMS

where ĝ is the vector of the adjusted mean for genotype for each treatment (B+, B-, and Delta) from Eq. 2, considering genotype as a fixed effect. Z is the incidence matrix of random effects of genotypes, and a is the vector of additive effects, where a ~ N(0,G σ2a). ε is the vector of random residuals with ε~ N (0, I σ2ε). G is the additive genomic relationship matrix (VanRaden 2008).

For the GBLUP_MAS model, SNPs are the fixed effects of the significant SNPs obtained from the GWAS analysis for each treatment (B+ and B-), except for Delta that the matrix was composed of all significant SNPs from B+, B-, and Delta GWAS analysis. For the model GBLUP_GMS (GWAS Marker Selection GBLUP), genomic additive matrix (G) was obtained using only the significant SNPs from the GWAS analysis for each treatment (B+ and B-), except for Delta that we used all significant SNPs. A similar approach for marker selection based on GWAS for genomic prediction is described by Jeong et al. (2020). The covered of the significant SNPs from GWAS analysis of each treatment (B+, B-, and Delta) were obtained using a windows range upstream and downstream, based on LD decay of respective chromosomes.

In order to evaluate the model performance, we used the CV-α cross-validation with 5 folds and 4 replications (Yassue et al. 2021). The predictive ability of each model was calculated by the Pearson correlation between the predicted and observed values from the validation set.

2.3 Results

Exploratory analysis, significance, and heritability

The phenotypic data revealed that the synthetic population of PGPB did promote growth in maize for most traits (Figure 1 A). For SDM and RDM, the inoculation with PGPB promoted an increase of 12.78% and 20.65%, respectively. Regarding the root's architecture traits, they were also influenced by the inoculation. The plants inoculated with PGPB (B+) tended to have a higher mean for most root traits, except RA and RSR. Conversely, for RA, RAD, and RSR, the treatment without PGPB (B-) tended to have a higher phenotypic variation. TT biplot analysis revealed the association between traits and treatments. RAD, RDSM, RSR, RA, and SDM showed higher variance than other traits. On the other hand, PH, CHA, SD, and RL had smaller discrepancies across treatments. RDM and SD were associated with the Delta treatment, while SDM was associated with B+ (Figure 1 B.).

The traits most correlated with SDM were RDM (0.80), ARV (0.71), RL (0.64), LRV (0.56), SD (0.58), PH (0.50), and CHA (0.40). The root traits were also highly correlated (RDM x RL, LRV x RL, LRV x RDM, ARV x RDM). Overall, the correlation between traits for the treatment B+ and B- were similar. Although, for RAD x CHA, RDM x PH, SD x RAD, SDM x RAD, treatment B- had higher values, except for RL x CHA (Supplementary Figure 3). According to PCA analysis (Supplementary Figure 4), we observed that most of the traits were positively associated with bacteria inoculation, except for RA and RSR. Also, it was possible to cluster the genotypes by treatment (B+ and B-).

Significant effects for treatment and genotypes were observed for most traits. On the other hand, the interaction between genotype and treatment was not observed (Supplementary Table 2). Nevertheless, Levene's test reveals a difference between the variances in the treatment B+ and B- for all traits except for CHA, SDM, and PH. Furthermore, the heritabilities varied between treatment and among traits. PH and SD had the higher heritabilities, meanwhile, root-traits and Delta treatment heritabilities tended to be lower (Table 1).

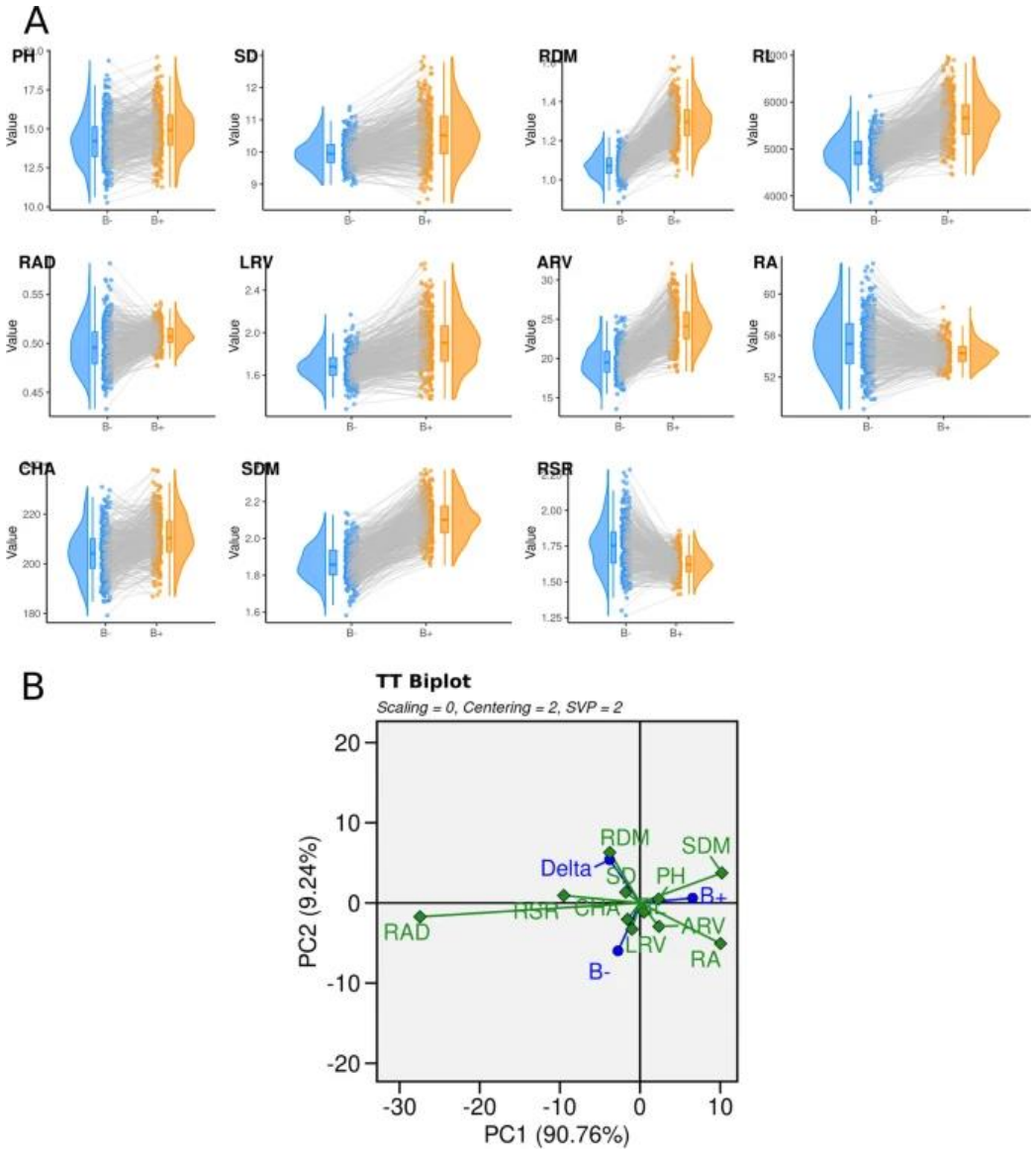

Figure 1. A. Interaction plot considering 360 genotypes with and without inoculation with PGPB (B+, B-, respectively). **B.** Treatment by trait analysis considering eleven traits and three Treatments (B+, B-, and Delta). PH: plant height, SD: stalk diameter, RDM: root dry mass, RL: root length, RAD (root average diameter), LRV: lateral root volume, ARV: axial root volume, RA: root Angle, CHA: convex hull area, SDM: shoot dry mass, RSR: ratio of shoot/root.

Table 1. Entry-mean based heritability considering the full model (Eq. 1) and the reduced model for the treatments B+, B-, and Delta (Eq. 2)

Trait[a]	PH	SD	RDM	RL	RAD	LRV	ARV	RA	CHA	SDM	RSR
Full model	0.77	0.67	0.49	0.44	0.43	0.61	0.44	0.34	0.47	0.43	0.51
B+	0.65	0.63	0.35	0.39	0.21	0.54	0.39	0.12	0.32	0.23	0.25
B-	0.63	0.38	0.23	0.31	0.42	0.32	0.34	0.29	0.31	0.20	0.44
Delta	0.08	0.03	0.01	0.13	0.14	0.00	0.14	0.03	0.08	0.00	0.11

[a]PH: plant height, SD: stalk diameter, RDM: root dry mass, RL: root length, RAD (root average diameter), LRV: lateral root volume, ARV: axial root volume, RA: root angle, CHA: convex hull area, SDM: shoot dry mass, RSR: ratio of shoot/root

Population structure and LD decay

The distribution of r^2 declined as the physical distance increased. The LD decay showed different values across chromosomes (Supplementary Figure 8) and ranged from ~175 kbp to ~200 kbp, considering the r^2 cutoff of 0.10. The first three principal components from PCA analysis showed that the origins of the genotypes (ESALQ and IAPAR) did not form a prominent group. The ADMIXTURE analysis revealed that the population probably presents 18 groups (Supplementary Figure 7).

GWAS

A total of 13,826 SNPs were used for GWAS analysis for all combinations of traits and treatment (Suppl. fig. 9-11). The best good fitness of each model depended on the trait and inoculation. The proportion of the explained variance by significant SNPs ranged from 0.025 to 0.158 (Table 2). There were 30 significant SNPs to treatment B+, 27 for treatment B-, and 8 for Delta. Root-traits had a higher number of significant SNPs. No significant SNPs were found for RDM and ARV for B+ treatment, and RDM and RA for B-. Contrastingly, for Delta, we found significant SNPs only for PH, SD, and RDM.

Significant SNPs were found in all chromosomes (Fig. 3 B). The SNP CM000784.4_172073449 was significant for the traits RL, SD, and LRV in the treatment B+. For B- and Delta, no overlapping SNPs were found for different traits. On the other hand, the SNP CM007648.1_169026437 was significant for SD in the

treatment B+ and Delta. Based on the LD decay of each chromosome, we identified the genes in linkage disequilibrium with the significant SNPs and found 53 overlapping genes associated simultaneously for B+ and Delta (Fig. 3 A). Meanwhile, no candidate genes were shared between B+ and B-, and B- and Delta. For B+ treatment, 55 overlapping genes were found for three different traits.

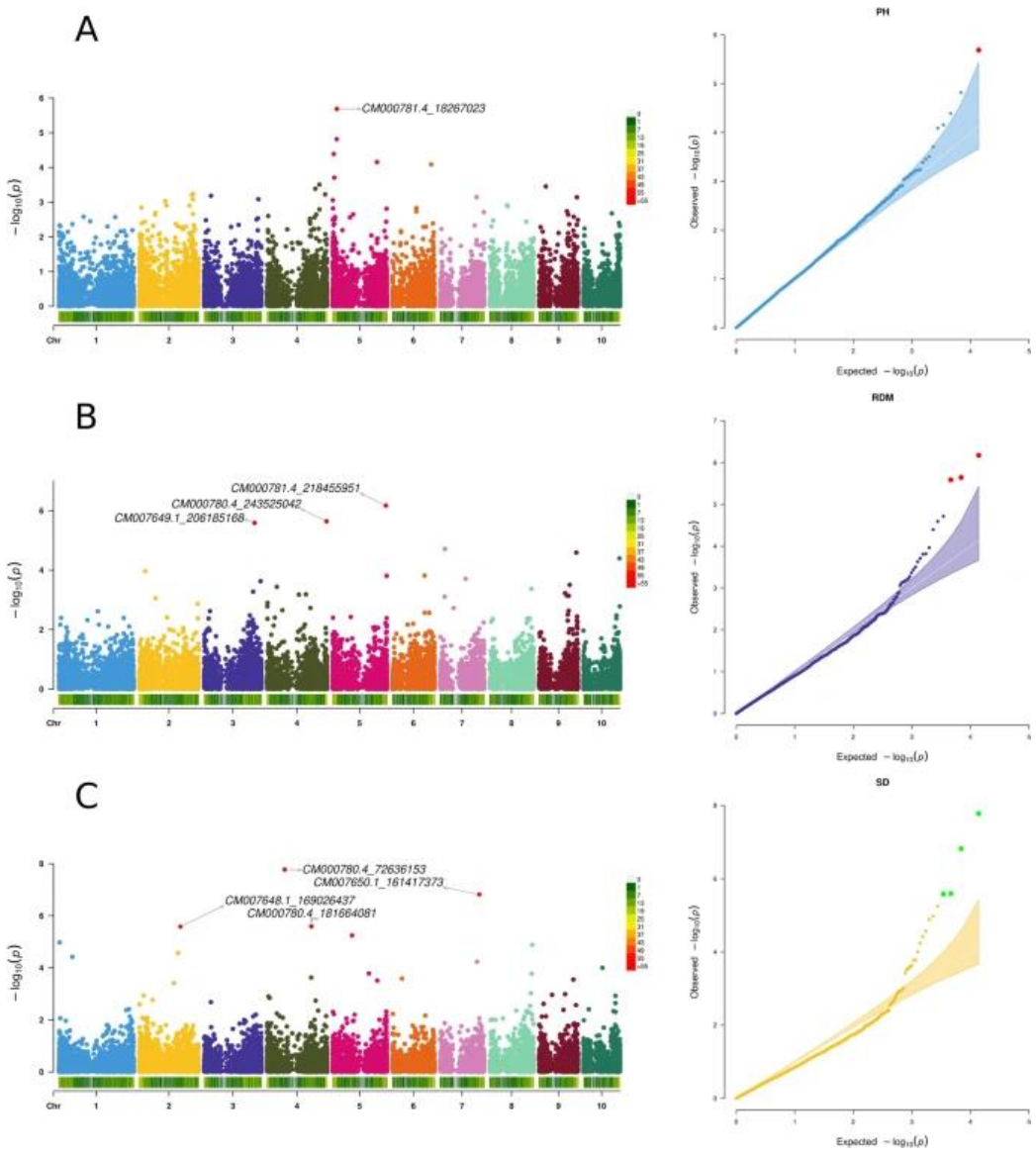

Figure 2. Manhattan and QQplot for the Delta treatment for (A) PH, (B) RDM, and (C) SD. The highlight SNPs were significant under the permutation test threshold. PH: plant height, SD: stalk diameter, SDM: shoot dry mass

Table 2. Trait, treatment, Marker, chromosome (Chrm), physical position (Posi**)**, minor allele frequency (MAF), -$\log_{10}$(p.values), number of genes, and % of the explained variance for each significant SNP from GWAS analysis

Trait[a]	Treat	Marker	Chrm	Posi	-$\log_{10}$(p.value)	MAF	NG	r^{2b}	Total[c]
PH	B+	CM000780.4_165053545	4	165053545	5.516	0.211	31	0.015	
PH	B+	CM000780.4_241977167	4	241977167	6.809	0.236	39	0.020	
PH	B+	CM000781.4_4105099	5	4105099	5.973	0.314	33	0.012	0.1176
PH	B+	CM000786.4_145397497	10	145397497	5.403	0.183	64	0.011	
PH	B+	CM007649.1_192589545	3	192589545	6.388	0.193	31	0.015	
PH	B+	CM007650.1_99946782	7	99946782	7.789	0.153	3	0.045	
RA	B+	CM007648.1_230920688	2	230920688	10.085	0.207	21	0.043	0.043
CHA	B+	CM000781.4_59021533	5	59021533	7.827	0.132	10	0.033	0.0646
CHA	B+	CM007647.1_27743555	1	27743555	6.608	0.225	18	0.032	
RAD	B+	CM000780.4_23997789	4	23997789	7.442	0.172	16	0.025	0.067
RAD	B+	CM007647.1_215419190	1	215419190	6.934	0.143	21	0.042	
SD	B+	CM000780.4_235759685	4	235759685	7.666	0.286	19	0.027	
SD	B+	CM000782.4_2916367	6	2916367	6.052	0.165	25	0.017	
SD	B+	CM007648.1_38816460	2	38816460	6.755	0.264	30	0.030	0.1584
SD	B+	CM007648.1_169026437	2	169026437	8.497	0.364	16	0.027	
SD	B+	CM007649.1_31624985	3	31624985	7.306	0.081	22	0.031	
SD	B+	CM007650.1_41779668	7	41779668	8.555	0.375	17	0.026	
RL	B+	CM000784.4_172073449	8	172073449	5.652	0.207	55	0.018	0.018
SDM	B+	CM000780.4_181569268	4	181569268	7.142	0.339	48	0.026	
SDM	B+	CM000784.4_172073449	8	172073449	5.325	0.207	55	0.017	0.0866
SDM	B+	CM007647.1_40109546	1	40109546	5.870	0.483	34	0.026	
SDM	B+	CM007648.1_206189255	2	206189255	5.323	0.396	26	0.017	
LRV	B+	CM000780.4_130864198	4	130864198	6.780	0.165	8	0.022	
LRV	B+	CM000784.4_172073449	8	172073449	9.494	0.207	55	0.034	0.0796
LRV	B+	CM007648.1_205666059	2	205666059	6.141	0.196	21	0.024	
RSR	B+	CM000782.4_92896441	6	92896441	9.767	0.083	8	0.033	0.13
RSR	B+	CM000785.4_154128247	9	154128247	7.560	0.385	43	0.020	

RSR	B+	CM007647.1_230383010	1	230383010	5.884	0.064	32	0.036	
RSR	B+	CM007648.1_82524308	2	82524308	6.088	0.288	12	0.019	
RSR	B+	CM007649.1_65858532	3	65858532	6.157	0.276	13	0.019	

[a]PH: plant height, SD: stalk diameter, RL: root length, RAD: root average diameter, LRV: lateral root volume, RA: root angle, CHA: convex hull area, SDM: shoot dry mass, RSR: ratio of shoot/root
[b] the proportion of phenotypic variance explained by SNP
[c] Sum of the proportion of phenotypic variance explained by SNPs.

Table 2 (continued)**.** Trait, treatment, Marker, chromosome (Chrm), physical position (Posi**)**, minor allele frequency (MAF), -$\log_{10}$(p.values), number of genes, and % of the explained variance for each significant SNP from GWAS analysis.

Trait[a]	Treat	Marker	Chrm	Posi	p.value	MAF	NG	r²b	Totalc
PH	B-	CM000784.4_130925210	8	0.018	5.653	0.186	12	0.018	0.0792
PH	B-	CM007647.1_115455586	1	0.029	9.931	0.118	8	0.029	
PH	B-	CM007647.1_297746433	1	0.013	6.310	0.061	37	0.013	
PH	B-	CM007649.1_57358057	3	0.019	6.887	0.319	14	0.019	
CHA	B-	CM007647.1_29786831	1	0.025	7.069	0.368	19	0.025	0.025
RAD	B-	CM007647.1_299457743	8	0.028	9.082	0.050	49	0.028	0.107
RAD	B-	CM000784.4_173246547	9	0.016	5.632	0.306	26	0.016	
RAD	B-	CM000785.4_148920255	9	0.024	6.834	0.124	34	0.024	
RAD	B-	CM000785.4_149343432	1	0.183	6.122	0.163	35	0.183	
RAD	B-	CM007647.1_190289865	1	0.020	6.001	0.361	29	0.020	
SD	B-	CM000782.4_71684361	6	0.042	8.243	0.228	9	0.042	0.1105
SD	B-	CM000785.4_45942495	9	0.031	9.280	0.369	14	0.031	
SD	B-	CM000785.4_142887012	9	0.018	6.567	0.179	22	0.018	
SD	B-	CM007649.1_13703357	3	0.020	5.947	0.436	22	0.020	
RL	B-	CM000781.4_186664447	5	0.036	9.646	0.188	27	0.036	0.036
SDM	B-	CM000780.4_18077503	4	0.021	6.120	0.344	9	0.021	0.1114
SDM	B-	CM007648.1_199939708	2	0.027	7.876	0.283	22	0.027	
SDM	B-	CM007650.1_12165480	7	0.039	8.585	0.065	17	0.039	
SDM	B-	CM007650.1_159938550	7	0.025	7.202	0.418	30	0.025	
ARV	B-	CM000784.4_176063412	8	0.029	6.230	0.138	49	0.029	0.1043
ARV	B-	CM000786.4_143378724	10	0.047	9.735	0.167	42	0.047	
ARV	B-	CM007649.1_149606256	3	0.028	6.080	0.051	23	0.028	
LRV	B-	CM007647.1_96702974	1	0.018	6.372	0.056	11	0.018	0.036
LRV	B-	CM007648.1_34931191	2	0.018	6.305	0.264	23	0.018	
RSR	B-	CM000780.4_238721378	4	0.011	5.663	0.050	15	0.011	0.051
RSR	B-	CM000784.4_109465968	8	0.022	7.059	0.297	15	0.022	
RSR	B-	CM007647.1_16710073	1	0.018	6.344	0.254	25	0.018	

Table 2 (continued)**.** Trait, treatment, Marker, chromosome (C), physical position (Posi**)**, minor allele frequency (MAF), -$\log_{10}$(p.values), number of genes, and % of the explained variance for each significant SNP from GWAS analysis

Trai	Trea	Marker	C	Posi	-$\log_{10}$(p.value)	MAF	N	r^{2b}	Total[c]
PH	Delta	CM000781.4_18267023	5	18267023	5.689	0.293	27	0.039	0.039
SD	Delta	CM000780.4_72636153	4	72636153	7.776	0.302	8	0.03	
SD	Delta	CM000780.4_181664081	4	181664081	5.592	0.286	39	0.015	
SD	Delta	CM007648.1_169026437	2	169026437	5.581	0.363	16	0.017	0.087
SD	Delta	CM007650.1_161417373	7	161417373	6.819	0.307	23	0.025	
RDM	Delta	CM000780.4_243525042	4	243525042	5.643	0.153	57	0.021	
RDM	Delta	CM000781.4_218455951	5	218455951	6.173	0.362	49	0.023	0.066
RDM	Delta	CM007649.1_206185168	3	206185168	5.586	0.052	26	0.022	

[a]PH: plant height, SD: stalk diameter, SDM: shoot dry mass
[b] the proportion of phenotypic variance explained by SNP.
[c] Sum of the proportion of phenotypic variance explained by SNPs.

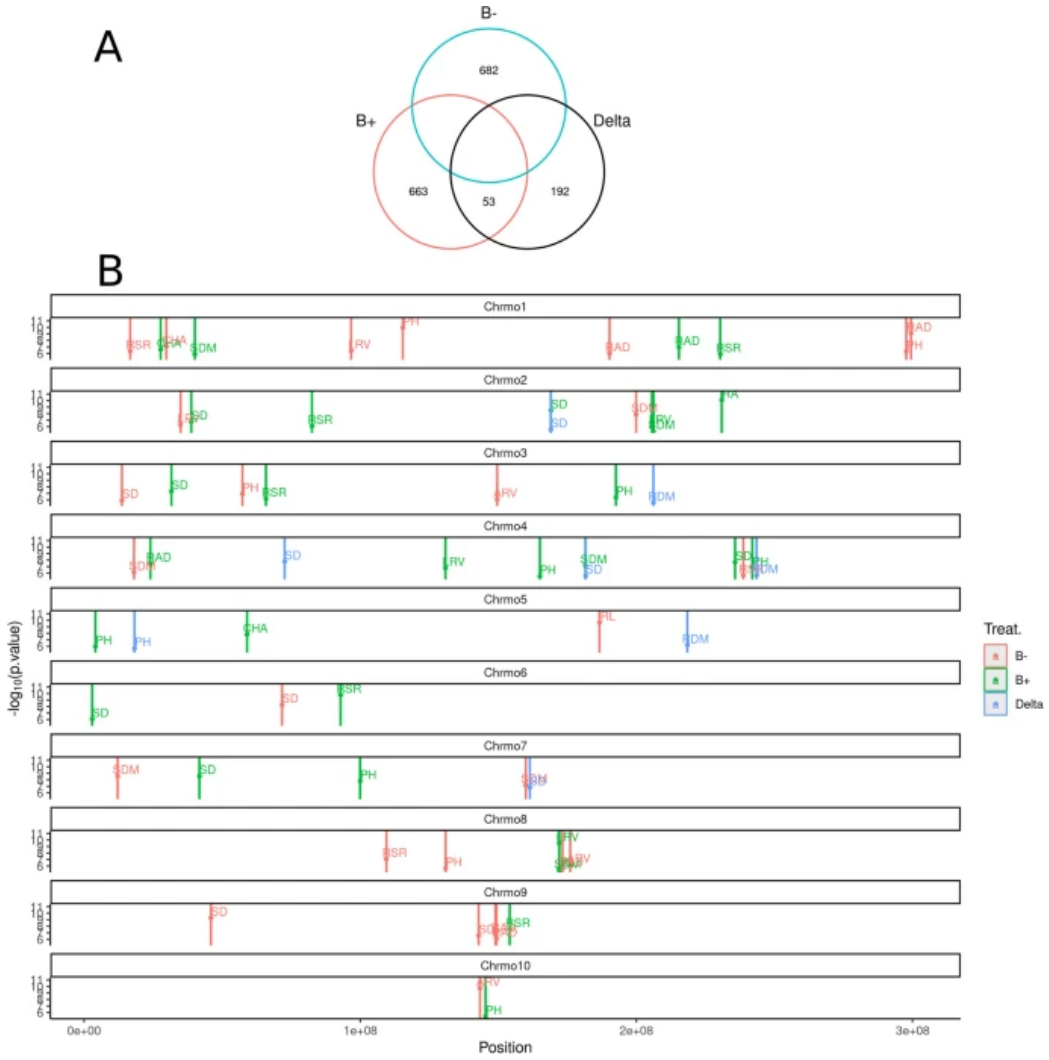

Figure 3. Venn diagrams (**A**) for overlapping genes for the treatments B+, B-, and Delta, and **(B)** Distribution of the significant SNPs from the GWAS analysis across the chromosomes for eleven traits and treatments (B+, B-, Delta). The bar between the significant SNP indicates the coverage upstream and downstream, based on the LD decay of the respective chromosomes. PH: plant height, SD: stalk diameter, RDM: root dry mass, RL: root length, RAD: root average diameter, LRV: lateral root volume, ARV: axial root volume, RA: root angle, CHA: convex hull area, SDM: shoot dry mass, RSR: ratio of shoot/root.

Genomic Prediction

Three genomic prediction models were used in order to evaluate the predictive ability for several traits under different treatments. The first model consisted of using a GBLUP model, the second a GBLUP plus the significant SNPs obtained in the GWAS analysis (GBLUP_MAS) as fixed effects. The latter used the genomic additive matrix only the significant SNPs from GWAS (GBLUP_GMS). For the treatment B+, B-, and Delta, the coverage of the significant SNPs from GWAS analysis based on the LD decay was 10.93 Mbp, 10.66 Mbp, and 24.39 Mbp, respectively.

The predictive abilities varied across the traits and treatment (B+, B-, and Delta). For all traits, the treatment Delta was the most difficult to predict; conversely, B+ and B- had similar performance. A similar performance was also observed between GBLUP and GBLUP_MAS. On the other hand, the use of GBLUP_GMS increased predictive ability for several traits and treatments.

Roots-related traits tended to have lower predictive abilities as well as the Delta treatment. For the treatment Delta and LRV, RA, RAD, RSR, the predictive abilities were near zero, regardless of the genomic prediction model. However, GBLUP_GMS increased predictive abilities for ARV, CHA, PH, RDM, RL, SD, and SDM in the Delta treatment.

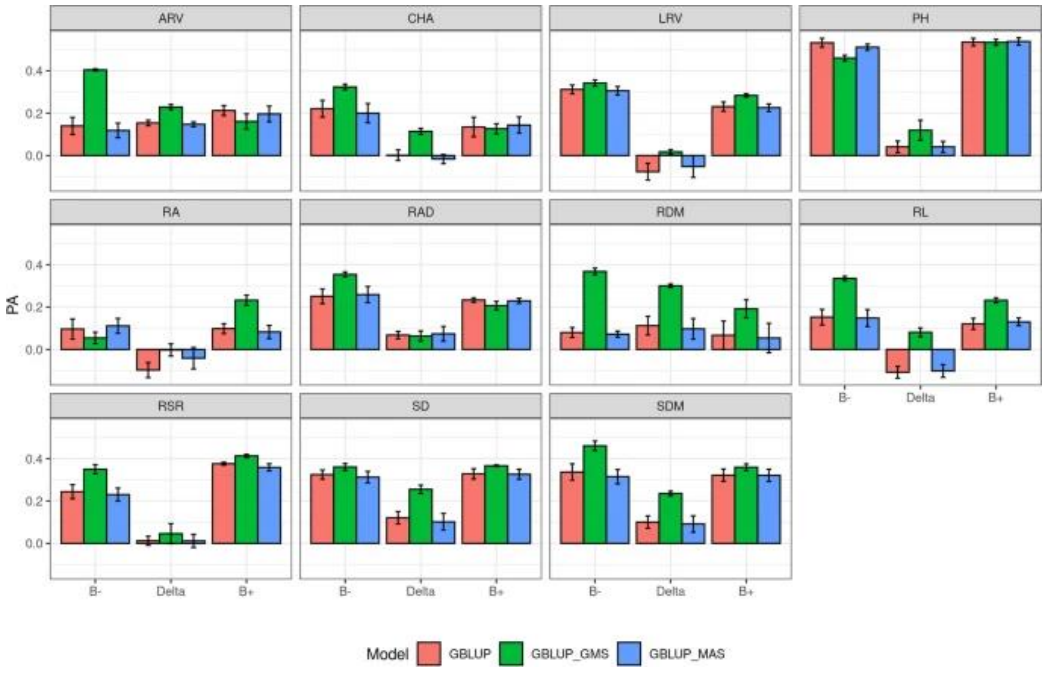

Figure 4. Predictive abilities for eleven traits under three treatments (B+, Delta, and B-) and three genomic prediction models (GBLUP, GBLUP_MAS, and GBLUP_GMS). PH: plant height, SD: stalk diameter, RDM: root dry mass, RL: root length, RAD (root average diameter), LRV: lateral root volume, ARV: axial root volume, RA: root angle, CHA: convex hull area, SDM: shoot dry mass, RSR: ratio of shoot/root.

2.4 Discussion

One of the main challenges in the studies with PGPB is to test the potential biostimulant candidates in field conditions due to its high interaction between host genotype and environmental factors (Rouphael et al. 2018). Our study employed greenhouse conditions, considering the trade-off between the number of genotypes tested and the real environmental conditions (Rouphael et al. 2018). In addition,

early trials may be used to evaluate the final performance of the genotypes (Strigens et al. 2012) and assess the population's genetic variance (Wang et al. 2016). Also, many other studies have considered early plant development as a strategy to select for stress tolerance (Grieder et al. 2014; Obeidat et al. 2018).

Recent studies have shown that the host genotype influences the symbiosis with PGPB, suggesting a host's genetic control (Wintermans et al. 2016; Proietti et al. 2018; Vidotti et al. 2019a). Moreover, the host heterosis plays an important role in shaping bacterial and fungal rhizosphere community composition (Wagner et al. 2020). To the best of our knowledge, this is the first study to evaluate the symbiosis between a synthetic population of PGPB and a public tropical maize association panel. Our results revealed that the synthetic population of PGPB showed biostimulant effects in the panel. Also, the PGPB impacted root architecture, and for most traits, influenced the phenotypic variation (Vidotti et al. 2019b). The PGPB benefits in roots have been associated due to the ability of the PGPB to produce plant hormones, such as indole-3-acetic acid (IAA) (Remans et al. 2008), ethylene (Barnawal et al. 2012), abscisic acid (Belimov et al. 2014), gibberellin (Khan et al. 2014), and cytokinins (Liu et al. 2013; Khan et al. 2014).

The strong correlation between root traits and PH, SD, and SDM confirms its importance for absorption and nutrient supply to biomass synthesis. Also, the correlation between root architecture traits revealed a mutual association between them, although GWAS analysis revealed that probably different genomic regions control them. Differential correlation between traits for treatments B+ and B-, and phenotypic variance may indicate that the PGPB can modulate the plant and root architecture. The association between RDM and SD and Delta treatment reveals the importance of these traits for the response to PGPB. Hence, seven significant SNPs were found for RDM and SD in the Delta treatment.

The effects of genotype and treatments were significant for most of the traits revealing that the PGPB promoted growth in the population, which has genetic variability. On the other hand, we didn't find significant effects for the interaction between genotype and treatment. The lack of interaction between genotype and treatment may be due to several factors such as small or absence of phenotypic plasticity in the population, early evaluations, and/or nitrogen-limited conditions limited growth in the B- treatment, making it difficult to assess the interaction.

Although, we found different SNPs in linkage disequilibrium for the treatment B+ and B- for several traits. The differential association between SNPs and treatments may suggest that other genomic regions are responsible for growth in the presence or absence of PGPB. Furthermore, SNPs associated with the Delta treatment highlight the genetic basis for the response of PGPB. At the same time, the GWAS results reported here for B+, B-, and Delta must be interpreted with caution since the traits studied are expected to be polygenic with small effects, and the reproducibility of the GWAS is expected to be low in this cases (Bian et al. 2014). Also, B+ and B- differences may be functional and not directly related to the bacteria.

The small heritability for Delta treatment reveals the challenges to evaluate the response to the PGPB. Nevertheless, we found eight significant SNPs associated for three different traits. Overlapping SNPs for the treatments B+ and Delta and others for B+ may indicate the presence of pleiotropic effects or linkage disequilibrium. Unfortunately, the magnitude of the LD decay in our population limits the interpretation of our analysis due to the size of the cover of each SNPs, making it difficult to find possible candidate genes for the trait. However, the overlapping genes in the treatments B+ and Delta and from the different traits for B+ can be used as possible candidate genes for further investigation of candidate genes for symbiosis between plant growth-promoting bacteria and maize.

Several studies have been using GWAS and genomic prediction in order to understand the genetic architecture of many traits (Wallace et al. 2016; Galli et al. 2020). Our results revealed that it is possible to increase the predictive ability of several traits when using a subset of SNPs representing important genomic regions, even if the trait heritability is low. The increase in predictive ability, especially in the Delta treatment, when considering only a small part of the genome to compute the GRM, may endorse the importance of these genomic regions for the response to PGPB. Also, these genomic regions can be explored via plant breeding for selection.

Root-traits information seems to be a key to breeding for resilience (Lombardi et al. 2021). Besides the difficulty of phenotyping the roots, a large amount of information generated and how to use this information in the decision-making process is still not fully comprehended. The evaluation of roots is usually laborious due to the need to wash the roots and evaluating them using visual scores (Trachsel et al. 2011) or image analysis (Seethepalli et al. 2021). In our work, we presented a shovelomics pipeline in order to evaluate, analyze, and apply the root

traits in the genetic architecture studies and their possible application into plant breeding programs. The predictive ability of RDM, SDM, and most of the traits in the Delta treatment substantially increased when we used the GRM with only the SNPs in disequilibrium with the above and underground traits highlighting the importance of phenotyping these traits (Yonis et al. 2020).

Our results corroborate the hypothesis that multiple genes with small effects are responsible for the response to the PGPB (Cotta et al. 2020). Furthermore, the genetic basis of the response to the PGPB can be used for plant breeding programs to maximize the symbiosis between tropical maize and PGPB and increase plant resilience against biotic and abiotic stress. Also, we suggest that further studies should be conducted in order to validate the SNPs and the genes responsible for the interaction between tropical maize and PGPB.

2.5 Conclusion

Despite the limitations, our study contributed to understanding the role of the host genotype in the symbiosis with PGPB. In tropical maize, it is controlled by many genes and has a quantitative inheritance. Furthermore, our tropical maize germplasm showed a significant genetic variability to the symbiosis with PGPB, being a good source of alleles for plant breeding programs to develop more resilient genotypes for tropical agriculture.

References

Adebayo AR, Kutu FR, Sebetha ET (2020) Data on root system architecture of water efficient maize as affected by different nitrogen fertilizer rates and plant density. Data in Brief 30:105561

Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:246

Allen M, Poggiali D, Whitaker K, et al (2021) Raincloud plots: a multi-platform tool for robust data visualization. Wellcome Open Res 4:63

Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci 19:52–61

Barnawal D, Bharti N, Maji D, et al (2012) 1-Aminocyclopropane-1-carboxylic acid (ACC) deaminase-containing rhizobacteria protect Ocimum sanctum plants during waterlogging stress via reduced ethylene generation. Plant Physiol Biochem 58:227–235

Bashan Y, de-Bashan LE, Prabhu SR, Hernandez J-P (2014) Advances in plant growth-promoting bacterial inoculant technology: formulations and practical perspectives (1998–2013). Plant and Soil 378:1–33

Bashan Y, Levanony H (1990) Current status of Azospirillum inoculation technology: Azospirillum as a challenge for agriculture. Canadian Journal of Microbiology 36:591–608

Batista BD, Lacava PT, Ferrari A, et al (2018) Screening of tropically derived, multi-trait plant growth-promoting rhizobacteria and evaluation of corn and soybean colonization ability. Microbiol Res 206:33-42

Batista BD, Dourado MN, Figueredo EF et al (2021) The auxin-producing Bacillus thuringiensis RZ2MS9 promotes the growth and modifies the root architecture of tomato (Solanum lycopersicum cv. Micro-Tom). Arch Microbiol *in press*.

Belimov AA, Dodd IC, Safronova VI, et al (2014) Abscisic acid metabolizing rhizobacteria decrease ABA concentrations in planta and alter plant growth. Plant Physiol Biochem 74:84–91

Bian Y, Yang Q, Balint-Kurti PJ, Wisser RJ, Holland JB (2014) Limits on the reproducibility of marker associations with southern leaf blight resistance in the maize nested association mapping population. BMC genomics. 15:1-5

Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Browning BL, Zhou Y, Browning SR (2018) A One-Penny Imputed Genome from Next-Generation Reference Panels. Am J Hum Genet 103:338–348

Butler DG, Cullis BR, Gilmour AR, et al (2017) ASReml-R Reference Manual Version 4. VSN International Ltd

Chi F, Shen S-H, Cheng H-P, et al (2005) Ascending migration of endophytic rhizobia, from roots to leaves, inside rice plants and assessment of benefits to rice growth physiology. Appl Environ Microbiol 71:7271–7278

Compant S, Clément C, Sessitsch A (2010) Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. Soil Biology and Biochemistry 42:669–678

Cotta MS, do Amaral FP, Cruz LM, et al (2020) Genome-wide Association Studies Reveal Important Candidate Genes for the Bacillus pumilus TUAT-1-Arabidopsis thaliana Interaction. bioRxiv

Creus CM, Graziano M, Casanovas EM, et al (2005) Nitric oxide is involved in the

Azospirillum brasilense-induced lateral root formation in tomato. Planta 221:297–303

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue

Egamberdiyeva D (2007) The effect of plant growth promoting bacteria on growth and nutrient uptake of maize in two different soils. Applied Soil Ecology 36:184–189

Esteras C, Formisano G, Roig C, Díaz A, Blanca J, Garcia-Mas J, Gómez-Guillamón M L, López-Sesé AI, Lázaro A, Monforte AJ, Picó B (2013). SNP genotyping in melons: genetic variation, population structure, and linkage disequilibrium. Theoretical and Applied Genetics, 126(5), 1285–1303

Gaffney J, Bing J, Byrne PF, et al (2019) Science-based intensive agriculture: Sustainability, food security, and the role of technology. Global Food Security 23:236–244

Gaiero JR, McCall CA, Thompson KA, et al (2013) Inside the root microbiome: bacterial root endophytes and plant growth promotion. Am J Bot 100:1738–1750

Galli G, Alves FC, Morosini JS, Fritsche-Neto R (2020) On the usefulness of parental lines GWAS for predicting low heritability traits in tropical maize hybrids. PLoS One 15:e0228724

Gamalero E, Lingua G, Berta G, Glick BR (2009) Beneficial role of plant growth promoting bacteria and arbuscular mycorrhizal fungi on plant responses to heavy metal stress. Can J Microbiol 55:501–514

Grieder C, Trachsel S, Hund A (2014) Early vertical distribution of roots and its association with drought tolerance in tropical maize. Plant and Soil 377:295–308

Gutiérrez-Luna FM, López-Bucio J, Altamirano-Hernández J, et al (2010) Plant growth-promoting rhizobacteria modulate root-system architecture in Arabidopsis thaliana through volatile organic compound emission. Symbiosis 51:75–83

Hoagland DR, Snyder WC (1933) Nutrition of strawberry plant under controlled conditions : (a) effects of deficiencies of boron and certain other elements : (b) susceptibility to injury from sodium salts. Proceedings of the American Society for Horticultural Science 30:288

Hungria M, Campo RJ, Souza EM, Pedrosa FO (2010) Inoculation with selected strains of *Azospirillum brasilense* and *A. lipoferum* improves yields of maize and wheat in Brazil. Plant Soil 331:413-425

Jeong S, Kim J-Y, Kim N (2020) GMStool: GWAS-based marker selection tool for genomic prediction from genomic data. Sci Rep 10:19653

Ji X, Lu G, Gai Y, et al (2010) Colonization of Morus alba L. by the plant-growth-promoting and antagonistic bacterium Burkholderia cepacia strain Lu10-1. BMC Microbiology 10:243

Kant S, Bi Y-M, Rothstein SJ (2011) Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. J Exp Bot 62:1499–1509

Khan AL, Waqas M, Kang S-M, et al (2014) Bacterial endophyte Sphingomonas sp. LK11 produces gibberellins and IAA and promotes tomato plant growth. J Microbiol 52:689–695

Khan MIR, Trivellini A, Fatma M, et al (2015) Role of ethylene in responses of plants to nitrogen availability. Front Plant Sci 6:927

Kroll S, Agler MT, Kemen E (2017) Genomic dissection of host–microbe and microbe–microbe interactions for advanced plant breeding. Current Opinion in Plant Biology 36:71–78

Laurance WF, Sayer J, Cassman KG (2014) Agricultural expansion and its impacts on tropical nature. Trends Ecol Evol 29:107–116

Lemanceau P, Blouin M, Muller D, Moënne-Loccoz Y (2017) Let the Core Microbiota Be Functional. Trends Plant Sci 22:583–595

Liu F, Xing S, Ma H, et al (2013) Cytokinin-producing, plant growth-promoting rhizobacteria that confer resistance to drought stress in Platycladus orientalis container seedlings. Appl Microbiol Biotechnol 97:9155–9164

Liu X, Huang M, Fan B, et al (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet 12:e1005767

Lombardi M, De Gara L, Loreto F (2021) Determinants of Root System Architecture For Future-Ready, Stress-Resilient Crops. Physiol Plant. https://doi.org/10.1111/ppl.13439

Lynch JP (2013) Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. Annals of Botany 112:347–357

Martins MR, Jantalia CP, Reis VM, et al (2018) Impact of plant growth-promoting bacteria on grain yield, protein content, and urea-15 N recovery by maize in a Cerrado Oxisol. Plant and Soil 422:239–250

Matias FI, Sabadin JFG, Moreira LA, et al (2021) Soil-app: a tool for soil analysis interpretation. Sci Agric 78.: https://doi.org/10.1590/1678-992x-2019-0113

Ma W, Li J, Qu B, et al (2014) Auxin biosynthetic gene TAR2 is involved in low nitrogen-mediated reprogramming of root architecture in Arabidopsis. Plant J 78:70–79

Mi G, Chen F, Wu Q, et al (2010) Ideotype root architecture for efficient nitrogen acquisition by maize in intensive cropping systems. Sci China Life Sci 53:1369–1373

Molina-Favero C, Creus CM, Lanteri ML, et al (2007) Nitric Oxide and Plant Growth Promoting Rhizobacteria: Common Features Influencing Root Growth and Development. Advances in Botanical Research 1–33

Obeidat W, Avila L, Earl H, Lukens L (2018) Leaf Spectral Reflectance of Maize Seedlings and Its Relationship to Cold Tolerance. Crop Science 58:2569–2580

Olivoto T, Lúcio AD (2020) metan: An R package for multi-environment trial analysis. Methods in Ecology and Evolution 11:783–789

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253

Pradhan P, Fischer G, van Velthuizen H, et al (2015) Closing Yield Gaps: How Sustainable Can We Be? PLoS One 10:e0129487

Proietti S, Caarls L, Coolen S, et al (2018) Genome-wide association study reveals novel players in defense hormone crosstalk in Arabidopsis. Plant Cell Environ 41:2342–2356

Qiao S, Fang Y, Wu A, et al (2019) Dissecting root trait variability in maize genotypes using the semi-hydroponic phenotyping platform. Plant and Soil 439:75–90

Qi X, Fu Y, Wang RY, et al (2018) Improving the sustainability of agricultural land use: An integrated framework for the conflict between food security and environmental deterioration. Applied Geography 90:214–223

Quecine MC, Araujo WL, Rossetto PB, et al (2012) Sugarcane Growth Promotion by the Endophytic Bacterium Pantoea agglomerans 33.1. Appl Environ Microb 78:7511-7518

Ray DK, Mueller ND, West PC, Foley JA (2013) Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS One 8:e66428

Remans R, Beebe S, Blair M, et al (2008) Physiological and genetic analysis of root responsiveness to auxin-producing plant growth-promoting bacteria in common bean (Phaseolus vulgaris L.). Plant and Soil 302:149–161

Rodríguez-Blanco A, Sicardi M, Frioni L (2015) Plant genotype and nitrogen fertilization effects on abundance and diversity of diazotrophic bacteria associated with maize (Zea mays L.). Biology and Fertility of Soils 51:391–402

Rojas-Tapias D, Moreno-Galván A, Pardo-Díaz S, et al (2012) Effect of inoculation with plant growth-promoting bacteria (PGPB) on amelioration of saline stress in maize (Zea mays). Applied Soil Ecology 61:264–272

Rosier A, Medeiros FHV, Bais HP (2018) Defining plant growth promoting rhizobacteria molecular and biochemical networks in beneficial plant-microbe interactions. Plant and Soil 428:35–55

Rouphael Y, Spíchal L, Panzarová K, et al (2018) High-Throughput Plant Phenotyping for Developing Novel Biostimulants: From Lab to Field or From Field to Lab? Front Plant Sci 9.: https://doi.org/10.3389/fpls.2018.01197

Sandhya V, Ali SZ, Grover M, et al (2010) Effect of plant growth promoting Pseudomonas spp. on compatible solutes, antioxidant status and plant growth of maize under drought stress. Plant Growth Regulation 62:21–30

Santoyo G, Moreno-Hagelsieb G, del Carmen Orozco-Mosqueda M, Glick BR (2016) Plant growth-promoting bacterial endophytes. Microbiological Research 183:92–99

Seethepalli A, Dhakal K, Griffiths M, (2021) et al RhizoVision Explorer: Open-source software for root image analysis and measurement standardization

Shamimuzzaman M, Gardiner JM, Walsh AT, et al (2020) MaizeMine: A Data Mining Warehouse for the Maize Genetics and Genomics Database. Front Plant Sci 11:592730

Sim S-C, Durstewitz G, Plieske J, et al (2012) Development of a Large SNP Genotyping Array and Generation of High-Density Genetic Maps in Tomato. PLoS ONE 7:e40563

Strigens A, Grieder C, Haussmann BIG, Melchinger AE (2012) Genetic Variation among Inbred Lines and Testcrosses of Maize for Early Growth Parameters and Their Relationship to Final Dry Matter Yield. Crop Science 52:1084–1092

Trachsel S, Kaeppler SM, Brown KM, Lynch JP (2011) Shovelomics: high throughput phenotyping of maize (Zea mays L.) root architecture in the field. Plant and Soil 341:75–87

Valente J, Gerin F, Le Gouis J, et al (2020) Ancient wheat varieties have a higher ability to interact with plant growth-promoting rhizobacteria. Plant, Cell & Environment 43:246–260

Van Deynze A, Zamora P, Delaux P-M, et al (2018) Nitrogen fixation in a landrace of maize is supported by a mucilage-associated diazotrophic microbiota. PLOS Biology 16:e2006352

VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science 91:4414–4423

Vejan P, Abdullah R, Khadiran T, et al (2016) Role of Plant Growth Promoting Rhizobacteria in Agricultural Sustainability—A Review. Molecules 21:573

Vidotti MS, Lyra DH, Morosini JS, et al (2019a) Additive and heterozygous (dis)advantage GWAS models reveal candidate genes involved in the genotypic variation of maize hybrids to Azospirillum brasilense. PLoS One 14:e0222788

Vidotti MS, Matias FI, Alves FC, et al (2019b) Maize responsiveness to Azospirillum brasilense: Insights into genetic control, heterosis and genomic prediction. PLoS One 14:e0217571

Wagner MR, Roberts JH, Balint-Kurti P, Holland JB (2020) Heterosis of leaf and rhizosphere microbiomes in field-grown maize. New Phytologist 228:1055–1069

Wallace JG, Zhang X, Beyene Y, et al (2016) Genome-wide Association for Plant Height and Flowering Time across 15 Tropical Maize Populations under Managed Drought Stress and Well-Watered Conditions in Sub-Saharan Africa. Crop Science 56:2365–2378

Wang X, Wang H, Liu S, et al (2016) Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. Nat Genet 48:1233–1241

Wei Z, Jousset A (2017) Plant Breeding Goes Microbial. Trends Plant Sci 22:555–558

Wintermans PCA, Bakker PAHM, Pieterse CMJ (2016) Natural genetic variation in Arabidopsis for responsiveness to plant growth-promoting rhizobacteria. Plant Mol Biol 90:623–634

Yan W, Frégeau-Reid J (2018) Genotype by yield* trait (GYT) biplot: a novel approach for genotype selection based on multiple traits. Scientific reports 8:1-10

Yassue RM, Sabadin JFG, Galli G, Alves FC (2021) CV-α: designing validations sets to increase the precision and enable multiple comparison tests in genomic prediction. Euphytica 217:106

Yin L, Zhang H, Tang Z, et al (2020) rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated tool for Genome-Wide Association Study. 2020.08.20.258491

Yonis BO, Pino Del Carpio D, Wolfe M, et al (2020) Improving root characterisation for genomic prediction in cassava. Sci Rep 10:8003

York LM, Galindo-Castañeda T, Schussler JR, Lynch JP (2015) Evolution of US maize (Zea mays L.) root architectural and anatomical phenes over the past 100 years corresponds to increased tolerance of nitrogen stress. J Exp Bot 66:2347–2358

Zheng X, Levine D, Shen J, et al (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and Prospects of Association Mapping in Plants. The Plant Genome, 1:1-16

3. A LOW-COST GREENHOUSE-BASED HIGH-THROUGHPUT PHENOTYPING PLATFORM FOR GENETIC STUDIES: A CASE STUDY IN MAIZE UNDER INOCULATION WITH PLANT GROWTH PROMOTING BACTERIA

Rafael Massahiro Yassue[1], Giovanni Galli[1], Ronaldo Borsato Junior[1], Hao Cheng[2], Gota Morota[3,4,*], and Roberto Fritsche-Neto[1,5,*]

[1]Department of Genetics, Luiz de Queiroz College of Agriculture, University of São Paulo, São Paulo, Brazil

[2]Department of Animal Science, University of California, Davis, USA

[3]Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, USA

[4]Center for Advanced Innovation in Agriculture, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 USA

[5]Quantitative Genetics and Biometrics Cluster, International Rice Research Institute, Los Baños, Philippines

Running title: A low-cost high-throughput maize phenotyping platform to study plant growth-promoting bacteria

Keywords: biostimulant, genomic correlation, genomic heritability, image-derived phenotyping, maize

Abbreviations: canopy coverage (CC); canopy volume (CV); genomic best linear unbiased prediction (GBLUP); ground control points (GCP); growing degree days (GDD); high-throughput phenotyping (HTP); high-throughput phenotyping plant height (PH_{HTP}); Luria-Bertani medium (LB); normalized difference vegetation index (NDVI); number of fully expanded leaves (NL); plant growth-promoting bacteria (PGPB); plant height (PH); real-time kinematic (RTK); shoot dry mass (SDM); single-nucleotide polymorphism (SNP); unoccupied aerial vehicle (UAV).

Core ideas

- A low-cost greenhouse-based HTP platform was developed.

- Image-derived phenotypes presented moderate to high genomic heritabilities and correlations.
- Plant growth-promoting bacteria can improve plant resilience under nitrogen-limited conditions.

References

Yassue, Rafael Massahiro, et al. "A low-cost greenhouse-based high-throughput phenotyping platform for genetic studies: A case study in maize under inoculation with plant growth-promoting bacteria." The Plant Phenome Journal 5.1 (2022):e20043.https://doi.org/10.1002/ppj2.20043

Abstract

Greenhouse-based high-throughput phenotyping (HTP) presents a useful approach for studying novel plant growth-promoting bacteria (PGPB). Despite the potential of this approach to leverage genetic variability for breeding new maize cultivars exhibiting highly stable symbiosis with PGPB, greenhouse-based HTP platforms are not yet widely used because they are highly expensive; hence, it is challenging to perform HTP studies under a limited budget. In this study, we built a low-cost greenhouse-based HTP platform to collect growth-related image-derived phenotypes. We assessed 360 inbred maize lines with or without PGPB inoculation under nitrogen-limited conditions. Plant height, canopy coverage, and canopy volume obtained from photogrammetry were evaluated five times during early maize development. A plant biomass index was constructed as a function of plant height and canopy coverage. Inoculation with PGPB promoted plant growth in early developmental stages. Phenotypic correlations between the image-derived phenotypes and manual measurements were at least 0.47 in the later stages of plant development. The genomic heritability estimates of the image-derived phenotypes ranged from 0.23 to 0.54. Moderate-to-strong genomic correlations between the plant biomass index and shoot dry mass (0.24–0.47) and between HTP-based plant height and manually measured plant height (0.55–0.68) across the developmental stages showed the utility of our HTP platform. Collectively, our results demonstrate the usefulness of the low-cost HTP platform for large-scale genetic and management studies to capture plant growth.

3.1 Introduction

Recent studies have reported the benefit of using plant growth-promoting bacteria (PGPB) to increase yield and resilience against biotic and abiotic stresses (Arif et al., 2020; Kumar et al., 2020) through various molecular mechanisms, including nitrogen

fixation and phytohormone production (Compant et al., 2010; Manoj et al., 2020). Importantly, Wintermans et al. (2016) and Vidotti et al. (2019a) found a differential response of genotypes under PGPB inoculation, suggesting that the response has a genetic basis. These findings opened frontiers for a plant breeding program to breed new cultivars having highly stable PGPB responses (Vidotti et al., 2019b). However, the difficulty of monitoring a large number of lines across phenological growth stages under different inoculation conditions constrains our ability to analyze the genetics of dynamic PGPB responses.

With the advancement in genotyping technologies, phenotyping is considered a new bottleneck in plant breeding (Furbank and Tester, 2011; Araus et al., 2018). Image-derived high-throughput phenotyping (HTP) presents a new avenue for automatic characterization of plants, owing to its capacity to generate difficult-to-measure phenotypes over time using advanced sensors and cameras (Araus and Cairns, 2014; Campbell et al., 2019; Mazis et al., 2020). Greenhouse-based HTP platforms have been developed to evaluate a number of plant responses, such as morphological (Brichet et al., 2017), disease (Thomas et al., 2018), and physiological (Wang et al., 2018) under microbial inoculants (Chai et al., 2021) and biotic and abiotic stresses (Araus and Cairns, 2014; Campbell et al., 2018). Therefore, leveraging HTP to evaluate hundreds or thousands of genotypes non-destructively under different managements (Araus et al., 2018; Rouphael et al., 2018) is a promising approach to study the interaction between plant genotypes and PGPB. The choice of the HTP platform largely depends on the trade-off between the precision of phenotypes, the number of managements it can evaluate, and cost.

One major factor limiting the wide deployment of image-derived HTP in plant breeding programs is the high cost of setting up an HTP platform, especially for small breeding programs or research institutions. In field trials, an unoccupied aerial vehicle (UAV) is the commonly used cost-effective HTP technology to collect high-throughput data (Xie and Yang, 2020; Araus et al., 2018). In greenhouses, conveyor (plant-to-sensor) and benchtop (sensor-to-plant) type systems are often used for automated HTP platforms (Li et al., 2021). The conveyor type automatically transports potted plants into an imaging room. In contrast, the benchtop type is equipped with a computer-controlled mechanical arm that can automatically locate the position of a plant for phenotyping. Although both conveyor and benchtop systems support diverse

cameras, their installation costs are expensive and may require modification of the existing greenhouse facilities. When there are budget constraints, researchers are motivated to build self-developed HTP platforms because large-scale greenhouse-based HTP platforms are produced mainly by commercial companies (Czedik-Eysenberg et al., 2018), which are forbiddingly expensive.

Several efforts have been made to develop a novel low-cost custom greenhouse HTP platform (Zhou et al., 2018; Du et al., 2021). The most common approach is to use a sliding track or cable railing system to move the imaging system that consists of the camera in the x and y axes. The images are processed using image stitching or photogrammetry techniques to obtain 2-D or 3-D phenotypes. However, this type of HTP platform is yet to be widely adopted in genetic and management studies because the number of genotypes or manage- ments that can be accommodated is limited. Therefore, the objective of this study was to build a low-cost non-commercial sensor-to-plant greenhouse-based HTP platform using a multispectral camera that has the capacity to accommodate hundreds of maize lines and develop an image-processing pipeline to obtain growth-related image-derived phenotypes. We assessed the utility of the image-derived phenotypes by evaluating 360 genotypes under different PGPB management in the early stages of maize development.

3.2 Materials and Methods

Low-cost high-throughput phenotyping platform

A low-cost greenhouse HTP platform was built, wherein the camera was positioned in a way that it obtained images from directly above the plants. The system was built in a conventional greenhouse with dimensions of 3.5 × 11 × 6 m height, width, and length, respectively. A cooling wall and ventilation were used to maintain the desired temperature, and additional luminosity was supplied using LED lamps.

The image capture system was inspired by the UAV flight plans. It consists of two fixed parallel tracks (9 m) and one mobile perpendicular track (5 m). They were positioned 2.5 m above the ground. The two parallel tracks were fixed to the greenhouse roof, as well as two support tracks to ensure stability and alignment. The parallel tracks move the perpendicular track along the x-axis, whereas the perpendicular track moves the sensors along the y-axis. Each track contained an individual 96-watt electric motor. These electric motors were controlled to achieve the

desired overlap using a remote control system (Figure 1). The speed of the tracks was set at 0.16 m/s. The two electric motors were synchronized to keep them aligned. The speed of the tracks was set at 0.16 m/s. A medium-density fiberboard platform (20 cm x 20 cm) was designed to accommodate the multispectral camera, light sensor, and battery. The fiberboard platform was attached to the y-axis mobile track. Four ground control points (GCP) geo-referenced with real-time kinematic (RTK) were used to assemble the orthomosaics. Top-view image data collection was performed between 12:00 and 13:00 with an overlap of 80 % frontal and 70% lateral views. The multispectral camera used was a Parrot Sequoia (Parrot SA, Paris, France), including green (550 nm), red (660 nm), red-edge (735 nm), and near-infrared (790 nm).

Image processing and data extraction Multispectral images were processed by assembling orthomosaics and the dense point cloud using Agisoft Metashape software (Agisoft LLC, St. Petersburg, Russia). The images were imported, aligned, and optimized using GCP. This was followed by the calculation of the dense point clouds and the stitching of orthomosaics.

The orthomosaics were analyzed using QGIS software (QGIS Development Team, 2021) to obtain a shapefile for each plot. The plots were manually identified, and a geometry point was assigned at the center of the plant. Then, a round positive buffer of 0.10 m was drawn for each plot. The shapefile of each plot was manually adjusted to reduce overlaps across plants. We applied image segmentation to the orthomosaics using the normalized difference vegetation index (NDVI) (Rouse et al., 1974) with a threshold of 0.35 to separate canopy vegetation from the background. The reflectance of each plot was calculated as the mean of each wavelength (green, red, red-edge, and near-infrared) using the R package FIELDimageR (Matias et al., 2020). The NDVI was calculated using the following formula: NDVI = (NIR -RED)/ (NIR + RED), where NIR and RED are the reflectances at the near-infrared and red wavelengths, respectively. Canopy coverage (CC) was calculated from the sum of the pixels in the canopy vegetation and transformed to cm^2 based on the resolution of the orthomosaics (mm $pixel^{-1}$).

Dense cloud points were used to estimate plant height (PH_{HTP}) and canopy volume (CV). Each point from the dense cloud point was composed of GPS coordinates (latitude, longitude, and altitude in the universal transverse mercator).

The dense cloud point data were processed using the R package lidR (Roussel and Auty, 2021). A round positive buffer of 0.01 m was generated at the center of each plant to obtain the corresponding points of each plot. PH_{HTP} was constructed from the difference between the 90 percentile of the top of the point cloud altitude and the pot altitude before plant germination (0 leaves) (Figure 2) (Galli et al., 2021). The image-derived plant biomass index, f(biomass), was derived from the product of PH_{HTP} (cm) and CC (cm^2) (Li et al., 2020) as f(biomass) = $PH_{HTP} \times CC$. For CV, the dense cloud points were filtered by colors using the "Select Points by Color" function in the Agisoft Metashape software to remove the background. Plants were then reconstructed from the point cloud data, and the CV was estimated using the α-shape algorithm (Lafarge and Pateiro-Lopez, 2020). The algorithm requires an α value that controls the tightness of the 3-D reconstruction of the points. The optimal value of α that yielded the greatest correlation with manual measurements was 0.01 (Moreno et al., 2020).

Plant growth-promoting bacteria experiment

A tropical maize association panel containing 360 inbred lines was used to study the response to PGPB. Of these, 179 inbred lines were from the Luiz de Queiroz College of Agriculture-University of Sao Paulo (ESALQ-USP) and 181 were from the Instituto de Desenvolvimento Rural do Paraná (IAPAR).

The inbred lines were evaluated under two managements: with (B+) and without (B-) PGPB inoculation under nitrogen stress. The B+ management consisted of a synthetic population of four PGPB. *Bacillus thuringiensis*, *Delftia* sp. RZ4MS18 (Batista et al., 2018, 2021), *Pantoea agglomerans* (Quecine et al., 2012), and *Azospirillum brasilense* Ab-v5 (Hungria et al., 2010) were selected based on a preliminary experiment that showed their ability to promote growth when co-inoculated. Each species was grown individually in Luria-Bertani (LB) medium at 28°C with agitation at 150 rpm for 24 h. The synthetic population was composed of an adjusted volume of each bacterial culture medium containing approximately 10^8 colony-forming units/mL. The B- management consisted of an inoculum with liquid LB only. Each plot containing three seeds was individually inoculated with 1 ml of the respective management, agitated, and sown afterwards. Each line was replicated twice across time, and each replication was composed of an augmented block design with six blocks and three common checks. Each experimental unit consisted of one pot

containing one plant. The managements B+ and B- were evaluated simultaneously. A pot had 20 centimeters of diameter with three liters of capacity. The pots were separated from each other at a distance of approximately 2 cm. To facilitate fertilization, irrigation, and weed removal, a space of approximately 50 cm for every six rows of pots was created. Detailed information about the experimental design and the PGPB inoculation is available in Yassue et al. (2021).

A total of 13,826 single-nucleotide polymorphisms (SNPs) were available for the maize inbred lines using a genotyping-by-sequencing method following the two-enzymes (PstI and MseI) protocol (Sim et al., 2012; Poland et al., 2012). DNA was extracted using the cetyltrimethylammonium bromide method (Doyle and Doyle, 1987). SNP calling was performed using the TASSEL 5.0 software (Bradbury et al., 2007) with B73 (B73-RefGen v4) as the reference genome. The SNP markers were filtered if the call rate was less than 90%, non-biallelic, and the minor allele frequency was less than 5%. Missing marker codes were imputed using the Beagle 5.0 software (Browning et al., 2018). Markers with pairwise linkage disequilibrium higher than 0.99, were removed using the SNPRelate R package (Zheng et al., 2012).

Manually measured and high-throughput phenotypes

The experiments were performed at ESALQ-USP in Brazil (22°42'39 "S; 47°38'09 "W, altitude 540 m). The final evaluation was conducted when most plants had developed six fully expanded leaves, approximately 33 days after sowing. The growth-related manually measured traits that were evaluated were plant height (PH) and shoot dry mass (SDM). Plant height was measured from the soil to the last expanded leaf's ligule, and SDM was obtained from the dry mass of the leaves and stalk.

The image-derived phenotypes were collected over time to capture plant growth, as previously described. For each replication, measurements were made at six time points defined by the number of expanded leaves: 0 (before germination), 2, 3, 4, 5, and 6 (Hanway, 1966). Since the genotypes presented expected inconsistencies in growth stages, the number of expanded leaves was determined as the mode of the population at a given time. A time point before the germination step was used to obtain the PH_{HTP}. Heat accumulation was calculated from the growing degree days (GDD) based on the formula: $GDD = T_i - T_{base}$, where T_i is the daily mean air temperature and T_{base} is the base temperature of 10°C. Mean air temperature was calculated using the following

formula: $T_i = (t_{max} - t_{min})/2$, where T_{max} and T_{min} are the maximum and minimum temperatures, respectively, of day i (Gilmore and Rogers, 1958). The R package pollen (Nowosad, 2019) was used to calculate GDD. Phenotypic correlations were estimated using Pearson correlations between the image-derived phenotypes (PH_{HTP}, CC, f(biomass), and CV) and manually measured phenotypes (PH and SDM).

Likelihood-ratio and Wald tests

The following model was used to test the effects of genotype, management (B+ and B-), andtheir interaction.

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_1\mathbf{r} + \mathbf{X}_2\mathbf{b} + \mathbf{X}_3\mathbf{m} + \mathbf{Z}_1\mathbf{g} + \mathbf{Z}_2\mathbf{gm} + \boldsymbol{E}$$

where $\mathbf{y}$ is the vector of phenotypes; $\mathbf{1}$ is the vector of ones; $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ are the incidence matrices for the fixed effects; $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the incidence matrices for the random effects; μ is the overall mean; $\mathbf{r}$, $\mathbf{b}$, and $\mathbf{m}$ are the fixed effects for replication, block within replication, and management (B+ and B-), respectively; $\mathbf{g} \sim N(0, \mathbf{G}\sigma_g^2)$ is the vector of random effect of genotype; $\mathbf{gm} \sim N(0, \mathbf{G} \otimes \mathbf{I}\sigma^2_{gm})$ is the vector of random effects of the interaction between genotype and management; and $\boldsymbol{E} \sim N(0, \mathbf{I}\sigma_e^2)$ is the random residual effect. Here $\mathbf{G}$ is the additive genomic relationship matrix (VanRaden, 2008); $\mathbf{I}$ is the identity matrix; σ^2_g is the additive genomic variance; σ^2_{gm} is the genotype-by-management interaction variance; and σ^2_e is the residual variance. The significance of random and fixed effects was assessed using the Wald and likelihood-ratio tests, respectively. The analysis was performed using the R package ASReml-R (Butler et al., 2017).

Conventional heritability

The conventional heritability (repeatability) was calculated for image-derived and manually measured phenotypes using the model described earlier, but the management (**m**) and genotype-by-management interaction terms (**gm**) were dropped. It was assumed that $\mathbf{g} \sim N(0, I\sigma^2_g)$ and $\boldsymbol{E} \sim N(0, I\sigma^2)$. The conventional heritability was estimated using the following formula:

$$h_c^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\epsilon^2}{n_r}},$$

where n_r is the number of replications (2).

Bayesian genomic best linear unbiased prediction

Univariate and bivariate Bayesian genomic best linear unbiased prediction (GBLUP) models were used to estimate genomic heritability and genomic correlation separately for B+ and B-. These Bayesian models were the same as those used for the Wald and likelihood-ratio tests, but the management (**m**) and genotype-by-management interaction terms (**gm**) were dropped. For the univariate model, a flat prior was assigned to **r** and **b**. The variance components, σ^2_g and σ^2_e, were drawn from a scaled inverse χ^2 distribution. For the bivariate model, **y** is the vector of phenotypes of two responses; $\mathbf{g} \sim N(0, \Sigma_g \otimes \mathbf{G})$ is the vector of genotypes; $\boldsymbol{E} \sim N(0, \Sigma_e \otimes \mathbf{I})$ is the residual; $\otimes$ is the Kronecker product; and Σ_g and Σ_e are the variance-covariance matrices for additive genomic and residual effects taking the forms

$$\Sigma_g = \begin{bmatrix} \sigma^2_{g_1} & \sigma_{g_{12}} \\ \sigma_{g_{21}} & \sigma^2_{g_2} \end{bmatrix}, \quad \Sigma_\epsilon = \begin{bmatrix} \sigma^2_{\epsilon_1} & \sigma_{\epsilon_{12}} \\ \sigma_{\epsilon_{21}} & \sigma^2_{\epsilon_2} \end{bmatrix},$$

where subscripts 1 and 2 refer to the first and second phenotypes. An inverse Wishart distribution was assigned to Σg and Σe with degrees of freedom ν = 4 and scale matrix S such that the prior means of Σg and Σe equal half of the phenotypic variance. All the Bayesian GBLUP models were fitted using 60,000 Markov chain Monte Carlo samples, 10,000 burn-in, and a thinning rate of 60 implemented in JWAS software (Cheng et al., 2018a,b). Model convergence was assessed using trace plots of the posterior distributions of the variance components.

Genomic heritability and genomic correlation

The variance components obtained from the univariate Bayesian GBLUP were used to estimate genomic heritability using the following formula: $h^2 = \sigma^2_g/(\sigma^2_g+\sigma^2_g/n)$. The estimates of genomic correlation were obtained from the estimated variance-covariance matrix in the bivariate Bayesian GBLUP model.

Data availability

The genotype and image data are available at https://doi.org/10.17632/5gvznd2b3n.3 and https://doi.org/10.17632/frsfpgnsyz.1, respectively.

3.3 Results

Image processing and data extraction

A total of 756 plots (plants) in each replication across time were evaluated during plant development. Each collection of images took approximately 10 min. The ground resolution of the orthomosaics was approximately 2.30 mm pix^{-1}, and the GCP error was approximately 4 cm (Table 1). Despite the difference between days after sowing, accumulated GDD were similar between replication 1 and replication 2. Additionally, the ground resolution of the orthomosaic values and GCP errors were consistent across different numbers of leaves.

Plant growth-promoting bacteria experiment

Statistical hypothesis testing and phenotypic correlation

The summary statistics of manually measured and image-derived phenotypes are shown in Supplementary Figures S1 and S2. The management and genotype effects were statisti cally significant for all image-derived and manually measured phenotypes across the different stages of maize development (Supplementary Tables S1-S4). This suggests that the presence of PGPB and genetic diversity significantly affect plant development and growth for PH_{HTP}, CC, f(biomass), and CV. However, the genotype-by-management interaction was not statistically significant. Similarly, the main effects of management and genotype were consistently significant, but the genotype-by-management interaction was absent for manually measured PH and SDM (Supplementary Table S5). Figure 3 shows the growth patterns of the image-derived phenotypes with (B+) or without (B-) PGPB inoculation. The B+ management produced higher mean values than the B- management for all image-derived and manually measured phenotypes, suggesting that PGPB inoculation promotes plant growth in early developmental stages as expected. Moderate phenotypic correlations were observed between the HTP and manually measured phenotypes (Table 2 and Supplementary Figure S3). Phenotypic correlations between PH_{HTP} and PH ranged from 0.23 to 0.64 (B+) and 0.36 to 0.57 (B-). Image-derived phenotypes CC, f(biomass), and

CV were equally correlated with SDM. The later growth stages tended to show higher phenotypic correlations (4, 5, and 6 leaves). Overall, B+ and B- showed a similar pattern of phenotypic correlations.

Heritability

Estimates of conventional and genomic heritability varied across image-derived phenotypes and stages of maize development (Table 3 and Supplementary Table S6). Earlier developmental stages tended to show higher estimates of conventional and genomic heritability. Among image-derived phenotypes, PH_{HTP} showed the highest estimates of genomic heritability ranging from 0.35 to 0.54 (B+) and 0.34 to 0.48 (B-). In contrast, CV showed the lowest genomic heritability estimates, particularly when the number of leaves was five. The conventional heritability estimates for manually measured PH were 0.59 (B+) and 0.57 (B-), while those of SDM were 0.16 for both managements. The genomic heritability estimates of manually measured PH were 0.61 (B+) and 0.57 (B-), while those of SDM were 0.30 (B+) and 0.28 (B-). Overall, the difference in the heritability estimates between B+ and B- was small.

Genomic correlation

The genomic correlations between image-derived and manually measured phenotypes showed a similar tendency to those of phenotypic correlations (Table 2). High genomic correlations were observed between PH_{HTP} and PH in the later stages of maize development for both B+ and B-. The image-derived f(biomass) showed the strongest genomic correlations with SDM, followed by CC. No differences in genomic correlations were observed between B+ and B-. Additionally, moderate-to-strong phenotypic and genomic correlations were observed across the developmental stages for each image-derived phenotype (Figure 4). As expected, measurements made at growth stages with adjacent number of leaves showed higher correlations.

3.4 Discussion

A greenhouse HTP platform was developed to evaluate the influence of PGPB on plant growth using image-derived phenotypes. HTP platforms play an important role in plant breeding programs, genetics, and management studies because they allow the evaluation of plant growth and development in a non-

destructive, time-efficient, and less laborious manner. The image-capture system and processing were designed to be similar to those of UAVs. The roof structure of the greenhouse was used to attach the tracks to save costs and enable easy installation without restructuring the greenhouse itself. The total cost to develop our greenhouse-based HTP system was approximately US$5,000. Our expenses were higher than those of a recently developed HTP system for soybean (Zhou et al., 2018). However, the size of the HTP platform developed in this study is larger and can accommodate more genotypes. In terms of cost per m^2, our HTP platform is still cost-efficient because the cost associated with our HTP system was $75 per m^2, whereas that of Zhou et al. (2018) was $40 per m^2. Although the orthomosaics and the dense point clouds were constructed using the proprietary Agisoft Metashape software, one of the open-source alternatives is WebOD (Vacca, 2020).

The image overlap during the capture was controlled by the opening angle of the camera, speed of the track, and the y-axis distance, so that different cameras can be easily utilized by adjusting these factors. The coordinate system used for GCP was a universal transverse mercator obtained from RTK GPS, which may not always work indoors because of the greenhouse roof. An alternative option is to use a local coordinate system.

The image analysis pipeline consisted of aligning the images, obtaining dense cloud points, and mosaicking (Figure 2). The most laborious steps were to manually identify each plot and adjust its shapefile to avoid overlapping plots. Several approaches have been proposed to automate the plot identification step, such as the fieldShape function in the FIELDimageR R package (Matias et al., 2020) or a negative buffer area (Galli et al., 2020). However, these methods did not produce adequate results in our case, probably because of leaf over-lapping (Ahmed et al., 2019). An alternative emerging approach is to implement semantic segmentation and object detection based on deep learning (Xie et al., 2017; Zou et al., 2020).

The effects of genotype and management were significant and consistent between the image-derived and manually measured phenotypes. This suggests that image-derived phenotypes can be used to assess the differences within genotypes or managements. Additionally, the image-derived phenotypes were capable of capturing plant growth at different stages of plant development. The image-derived genomic heritability estimates tended to be lower than those of manually measured phenotypes and decreased as the plants developed. This was likely due to the difficulty in

accurately phenotyping taller plants. The magnitude of the genomic correlations and genomic heritabilities were similar between management groups B+ and B-. This was expected because the genotype-by-management interaction term was not significant. Our HTP platform was able to consistently capture genetic variability within each management.

No significant interaction between genotype and management for both HTP and manually measured phenotypes may also indicate the absence of phenotypic plasticity for PGPB responses in our population. Our findings agree with those of Vidotti et al. (2019a) and Vidotti et al. (2019b), who did not find significant genotype and management interactions in hybrid maize using different genotypes and PGPB from this study. This might be because both managements were tested under nitrogen-limited conditions or the experiment only covered the early developmental stages. For example, Guo et al. (2020) reported that low nutrients in optimal irrigated growth conditions might contribute to the absence of genotype-by-water availability interaction in wheat. On the other hand, the significant management suggests that PGPB can promote plant growth. Nevertheless, further studies are needed to vary nitrogen levels, assess PGPB responses at the later stages of development, and validate our results in field trials.

Moderate-to-strong phenotypic and genomic correlations between PH_{HTP} and PH revealed that image-derived PH_{HTP} can be a good predictor for manually measured PH. Similarly, a moderate genomic correlation between f(biomass) and SDM suggests that f(biomass) can be used as a secondary or correlated phenotype for SDM in genomic predictions (Rutkoski et al., 2016). We also investigated the utility of spectral indices (e.g., NDVI) as a proxy for SDM. However, the phenotypic correlation between NDVI and SDM was low (average was 0.13). A potential reason for this might be the difficulty in accurately calibrating images using a calibrated reflectance panel or a sunshine (light) sensor. The reduction of sunlight inside the greenhouse due to the polyethylene roof may have limited the calibration accuracy. Unlike Li et al. (2020), this was the main reason why we did not include NDVI to calculate f(biomass).

The architecture of maize plants makes image-derived phenotyping harder because stalks and leaves grow beyond their pots and interfere with neighboring pots. This can be minimized by increasing the distance between the pots and distributing them equidistantly if a larger greenhouse is available. Another limiting factor that

may reduce the correlation between PH_{HTP} and PH is related to plant morphology. For instance, during maize growth, the leaf development stage directly impacts plant height projection. Alternatively, we can measure $\mathbf{PH}_{HTP}$ at the leaf ligule of the last fully expanded leaf. However, locating the leaf ligule in the HTP platform is a challenging task because $\mathbf{PH}_{HTP}$ is based on plant height projection (Figure S4).

There are several greenhouse-based HTP platforms available that differ in terms of precision, resolution, and applications (Li et al., 2021). The advantage of our HTP platform is its low cost compared to commercial platforms, while having the capacity to phenotype many lines. Despite the fact that our image-derived phenotypes were slightly less correlated with manually measured phenotypes than other related studies found in the literature (Campbell et al., 2015; Zhou et al., 2018; Volpato et al., 2021), our results confirm that image-derived phenotypes can provide valuable information for capturing temporal PGPB responses in maize. Further research is warranted to evaluate the utility of image-derived phenotypes to study PGPB responses in longitudinal genomic predictions and genome-wide association studies (Campbell et al., 2019; Baba et al., 2020; Anderson et al., 2020).

3.5 Conclusions

We developed a low-cost high-throughput phenotyping platform capable of capturing plant growth across developmental stages. This platform was used to study the symbiosis between PGPB and maize. We found a moderate-to-strong phenotypic and genomic correlation between the image-derived and manually measured phenotypes, where PGPB promoted growth in early developmental stages in the population. The findings reported in this study will help small plant breeding programs or public research institutions to integrate phenomics, genetic, and management studies under a limited budget.

Acknowledgments

This study was supported in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Grant #17/24327-0 and #19/04697-2 from São Paulo Research Foundation (FAPESP).

Author contributions

Rafael Massahiro Yassue: Conceptualization; Data curation, Formal analysis, Investigation; Methodology; Visualization; Writing-original draft; Writing-review & editing. Giovanni Galli: Investigation; Methodology; Writing-review & editing. Ronaldo Borsato Junior: Investigation; Writing-review & editing. Hao Cheng: Software, Writing-review & editing. Gota Morota: Conceptualization; Methodology; Supervision; Writing-original draft; Writing-review & editing. Roberto Fritsche-Neto: Conceptualization; Funding acquisition; Supervision; Writing-review & editing.

Conflict of interest

None declared.

References

Ahmed, I., Eramian, M., Ovsyannikov, I., van der Kamp, W., Nielsen, K., Duddu, H. S., Rumali, A., Shirtliffe, S., and Bett, K. (2019). Automatic detection and segmentation of lentil crop breeding plots from multi-spectral images captured by UAV-mounted camera. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE.

Anderson, S. L., Murray, S. C., Chen, Y., Malambo, L., Chang, A., Popescu, S., Cope, D., and Jung, J. (2020). Unoccupied aerial system enabled functional modeling of maize height reveals dynamic expression of loci. *Plant Direct*, 4(5).

Araus, J. L. and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends in Plant Science*, 19(1):52–61.

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends in Plant Science*, 23(5):451–466.

Arif, I., Batool, M., and Schenk, P. M. (2020). Plant microbiome engineering: Expected benefits for improved crop growth and resilience. *Trends in Biotechnology*, 38(12):1385–1396.

Baba, T., Momen, M., Campbell, M. T., Walia, H., and Morota, G. (2020). Multi-trait random regression models increase genomic prediction accuracy for a temporal physiological trait derived from high-throughput phenotyping. *PloS One*, 15(2):e0228118.

Batista, B. D., Dourado, M. N., Figueredo, E. F., Hortencio, R. O., Marques, J. P. R., Piotto, F. A., Bonatelli, M. L., Settles, M. L., Azevedo, J. L., and Quecine, M. C. (2021). The auxin-producing *Bacillus thuringiensis* RZ2ms9 promotes the growth and modifies the root architecture of tomato (*Solanum lycopersicum* cv. micro-tom). *Archives of Microbiology*.

Batista, B. D., Lacava, P. T., Ferrari, A., Teixeira-Silva, N. S., Bonatelli, M. L., Tsui, S., Mondin, M., Kitajima, E. W., Pereira, J. O., Azevedo, J. L., and Quecine, M. C. (2018). Screening of tropically derived, multi-trait plant growth- promoting rhizobacteria and evaluation of corn and soybean colonization ability. *Microbiological Research*, 206:33–42.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19):2633–2635.

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(1).

Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, 103(3):338–348.

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., and Thompson, R. (2017). *ASReml-R Reference Manual Version 4*.

Campbell, M., Momen, M., Walia, H., and Morota, G. (2019). Leveraging breeding values obtained from random regression models for genetic inference of longitudinal traits. *The Plant Genome*, 12(2):180075.

Campbell, M. T., Knecht, A. C., Berger, B., Brien, C. J., Wang, D., and Walia, H. (2015). Integrating image-based phenomics and association analysis to dissect the genetic archi- tecture of temporal salinity responses in rice. *Plant Physiology*, 168(4):1476–1489.

Campbell, Z. C., Acosta-Gamboa, L. M., Nepal, N., and Lorence, A. (2018). Engineering plants for tomorrow: how high-throughput phenotyping is contributing to the development of better crops. *Phytochemistry Reviews*, 17(6):1329–1343.

Chai, Y. N., Ge, Y., Stoerger, V., and Schachtman, D. P. (2021). High-resolution phenotyping of sorghum genotypic and phenotypic responses to low nitrogen and synthetic microbial communities. *Plant, Cell & Environment*, 44(5):1611–1626.

Cheng, H., Fernando, R., and Garrick, D. (2018a). JWAS: Julia implementation of whole-genome analysis software. In *Proceedings of the world congress on genetics applied to livestock production*.

Cheng, H., Kizilkaya, K., Zeng, J., Garrick, D., and Fernando, R. (2018b). Genomic prediction from multiple-trait bayesian regression methods using mixture priors. *Genetics*, 209(1):89–103.

Compant, S., Clément, C., and Sessitsch, A. (2010). Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. *Soil Biology and Biochemistry*, 42(5):669–678.

Czedik-Eysenberg, A., Seitner, S., Güldener, U., Koemeda, S., Jez, J., Colombini, M., and Djamei, A. (2018). The 'phenobox', a flexible, automated, open-source plant phenotyping solution. *New Phytologist*, 219(2):808–823.

Doyle, J. and Doyle, J. (1987). A rapid dna isolation procedure for small quantities of fresh leaf tissue. *PHYTOCHEMICAL BULLETIN*, 17(RESEARCH).

Du, J., Fan, J., Wang, C., Lu, X., Zhang, Y., Wen, W., Liao, S., Yang, X., Guo, X., and Zhao, C. (2021). Greenhouse-based vegetable high-throughput phenotyping platform and trait evaluation for large-scale lettuces. *Computers and Electronics in Agriculture*, 186:106193.

Furbank, R. T. and Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, 16(12):635–644.

Galli, G., Horne, D. W., Collins, S. D., Jung, J., Chang, A., Fritsche-Neto, R., and Rooney, W. L. (2020). Optimization of UAS-based high-throughput phenotyping to estimate plant health and grain yield in sorghum. *The Plant Phenome Journal*, 3(1).

Galli, G., Sabadin, F., Costa-Neto, G. M. F., and Fritsche-Neto, R. (2021). A novel way to validate UAS-based high-throughput phenotyping protocols using in silico experiments for plant breeding purposes. *Theoretical and Applied Genetics*, 134(2):715–730.

Gilmore, E. C. and Rogers, J. S. (1958). Heat units as a method of measuring maturity in corn 1. *Agronomy Journal*, 50(10):611–615.

Guo, X., Svane, S. F., Füchtbauer, W. S., Andersen, J. R., Jensen, J., and Thorup-Kristensen, K. (2020). Genomic prediction of yield and root development in wheat under changing water availability. *Plant Methods*, 16(1).

Hanway, J. J. (1966). *How a corn plant develops*. Iowa State University.

Hungria, M., Campo, R. J., Souza, E. M., and Pedrosa, F. O. (2010). Inoculation with selected strains of azospirillum brasilense and a. lipoferum improves yields of maize and wheat in Brazil. *Plant and Soil*, 331(1-2):413–425.

Kumar, A., Singh, S., Gaurav, A. K., Srivastava, S., and Verma, J. P. (2020). Plant growth-promoting bacteria: Biological tools for the mitigation of salinity stress in plants. *Frontiers in Microbiology*, 11.

Lafarge, T. and Pateiro-Lopez, B. (2020). *alphashape3d: Implementation of the 3D Alpha-Shape for the Reconstruction of 3D Sets from a Point Cloud*. R package version 1.3.1.

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 and Biotechnology*, 8:1533.

Li, F., Piasecki, C., Millwood, R. J., Wolfe, B., Mazarei, M., and Stewart Jr, C. N. (2020). High-throughput switchgrass phenotyping and biomass modeling by uav. *Frontiers in Plant Science*, 11:1532.

Manoj, S. R., Karthik, C., Kadirvelu, K., Arulselvi, P. I., Shanmugasundaram, T., Bruno, B., and Rajkumar, M. (2020). Understanding the molecular mechanisms for the enhanced phytoremediation of heavy metals through plant growth promoting rhizobacteria: A re- view. *Journal of Environmental Management*, 254:109779.

Matias, F. I., Caraza-Harter, M. V., and Endelman, J. B. (2020). FIELDimageR: An R package to analyze orthomosaic images from agricultural field trials. *The Plant Phenome Journal*, 3(1).

Mazis, A., Choudhury, S. D., Morgan, P. B., Stoerger, V., Hiller, J., Ge, Y., and Awada, T. (2020). Application of high-throughput plant phenotyping for assessing biophysical traits and drought response in two oak species under controlled environment. *Forest Ecology and Management*, 465:118101.

Moreno, H., Rueda-Ayala, V., Ribeiro, A., Bengochea-Guevara, J., Lopez, J., Peteinatos, G., Valero, C., and Andújar, D. (2020). Evaluation of vineyard cropping systems using on-board RGB-depth perception. *Sensors*, 20(23):6912.

Nowosad, J. (2019). *pollen: Analysis of Aerobiological Data*. R package version 0.71. Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J.-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, 7(2):e32253.

QGIS Development Team (2021). *QGIS Geographic Information System*. QGIS Association. Quecine, M. C., Araújo, W. L., Rossetto, P. B., Ferreira, A., Tsui, S., Lacava, P. T., Mondin, M., Azevedo, J. L., and Pizzirani-Kleiner, A. A. (2012). Sugarcane growth promotion by the endophytic bacterium pantoea agglomerans 33.1. *Applied and Environmental Microbiology*, 78(21):7511–7518.

Rouphael, Y., Spíchal, L., Panzarová, K., Casa, R., and Colla, G. (2018). High-throughput plant phenotyping for developing novel biostimulants: From lab to field or from field to lab? *Frontiers in Plant Science*, 9.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., et al. (1974). Monitoring vegetation systems in the great plains with erts. *NASA special publication*, 351(1974):309.

Roussel, J.-R. and Auty, D. (2021). *Airborne LiDAR Data Manipulation and Visualization for Forestry Applications*. R package version 3.1.4.

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., Crossa, J., Reynolds, M., and Singh, R. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3: Genes, Genomes, Genetics*, 6(9):2799–2808.

Sim, S.-C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M. W., Deynze, A. V., Hamilton, J. P., Buell, C. R., Causse, M., Wijeratne, S., and Francis, D. M. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS ONE*, 7(7):e40563.

Thomas, S., Behmann, J., Steier, A., Kraska, T., Muller, O., Rascher, U., and Mahlein, A.-K. (2018). Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. *Plant Methods*, 14(1).

Vacca, G. (2020). Web open drone map (webodm) a software open source to photogrammetry process. In *Fig Working Week Smart surveyors for land and water management*. Amsterdam.

VanRaden, P. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11):4414–4423.

Vidotti, M. S., Lyra, D. H., Morosini, J. S., Granato, Í. S. C., Quecine, M. C., de Azevedo, J. L., and Fritsche-Neto, R. (2019a). Additive and heterozygous (dis)advantage GWAS models reveal candidate genes involved in the genotypic variation of maize hybrids to Azospirillum brasilense. *PLOS ONE*, 14(9):e0222788.

Vidotti, M. S., Matias, F. I., Alves, F. C., Pérez-Rodríguez, P., Beltran, G. A., Burgueño, J., Crossa, J., and Fritsche-Neto, R. (2019b). Maize responsiveness to Azospirillum brasilense: Insights into genetic control, heterosis and genomic prediction. *PLOS ONE*, 14(6):e0217571.

Volpato, L., Pinto, F., González-Pérez, L., Thompson, I. G., Borém, A., Reynolds, M., Gérard, B., Molero, G., and Rodrigues, F. A. (2021). High throughput field phenotyping for plant height using UAV-based RGB imagery in wheat breeding lines: Feasibility and validation. *Frontiers in Plant Science*, 12.

Wang, H., Qian, X., Zhang, L., Xu, S., Li, H., Xia, X., Dai, L., Xu, L., Yu, J., and Liu, X. (2018). A method of high throughput monitoring crop physiology using chlorophyll fluorescence and multispectral imaging. *Frontiers in Plant Science*, 9.

Wintermans, P. C. A., Bakker, P. A. H. M., and Pieterse, C. M. J. (2016). Natural genetic variation in Arabidopsis for responsiveness to plant growth-promoting rhizobacteria. *Plant Molecular Biology*, 90(6):623–634.

Xie, C., Wang, J., Zhang, Z., Zhou, Y., Xie, L., and Yuille, A. (2017). Adversarial examples for semantic segmentation and object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE.

Xie, C. and Yang, C. (2020). A review on plant high-throughput phenotyping traits using UAV-based sensors. *Computers and Electronics in Agriculture*, 178:105731.

Yassue, R. M., Carvalho, H. F., Gevartosky, R., Sabadin, F., Souza, P. H., Bonatelli, M. L., Azevedo, J. L., Quecine, M. C., and Fritsche-Neto, R. (2021). On the genetic architecture in a public tropical maize panel of the symbiosis between corn and plant growth-promoting bacteria aiming to improve plant resilience. *Molecular Breeding*, 41(10).

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24):3326–3328.

Zhou, J., Chen, H., Zhou, J., Fu, X., Ye, H., and Nguyen, H. T. (2018). Development of an automated phenotyping platform for quantifying soybean dynamic responses to salinity stress in greenhouse environment. *Computers and Electronics in Agriculture*, 151:319–330.

Zou, H., Lu, H., Li, Y., Liu, L., and Cao, Z. (2020). Maize tassels detection: a benchmark of the state of the art. *Plant Methods*, 16(1).

Appendix

Table 1: Replication (Rep), number of fully expanded leaves (NL), days after sowing (DAS), ground resolution of orthomosaic (GRO), ground control points (GCP) error, and accumulated growing degree days (GDD) across five evaluations during maize growth development.

Rep	NL	DAS	GRO (mm pixel^{-1})	GCP error (m)	GDD (°C)[a]
1	2	11	2.32	0.04	169.6
1	3	15	2.29	0.06	229.3
1	4	18	2.27	0.05	268.5
1	5	22	2.25	0.05	320.4
1	6	27	2.25	0.04	394.8
2	2	14	2.34	0.03	172.1
2	3	21	2.35	0.03	250.1
2	4	27	2.31	0.03	310.6
2	5	30	2.31	0.03	338.8
2	6	37	2.29	0.03	410.3

[a] The base temperature used for GDD estimation was 10°C

Table 2: Phenotypic (r_p) and genomic (r_g) correlations between high-throughput phenotyp- ing and manually measured phenotypes across maize development with (B+) or without (B-) plant growth-promoting bacteria inoculation. PH_{HTP}: image-derived plant height; PH: manually measured plant height; CC: canopy coverage; SDM: shoot dry mass; f(biomass): plant biomass index; CV: canopy volume; and NL: number of fully expanded leaves.

	PH_{HTP}:PH		CC:SDM		f(biomass):SDM		CV:SDM	
NL	r_p	r_g	r_p	r_g	r_p	r_g	r_p	r_g
				B+				
2	0.29	0.55	0.35	0.20	0.35	0.24	0.42	0.14
3	0.23	0.59	0.45	0.30	0.39	0.39	0.62	0.29
4	0.61	0.64	0.51	0.35	0.62	0.47	0.38	0.36
5	0.64	0.67	0.47	0.30	0.64	0.43	0.62	0.32
6	0.54	0.66	0.53	0.32	0.60	0.42	0.65	0.31
				B-				
2	0.38	0.60	0.35	0.35	0.38	0.41	0.29	0.21
3	0.36	0.67	0.51	0.41	0.48	0.46	0.49	0.29
4	0.53	0.68	0.59	0.47	0.59	0.43	0.40	0.36
5	0.57	0.63	0.62	0.45	0.66	0.47	0.50	0.32
6	0.56	0.67	0.63	0.32	0.63	0.44	0.61	0.31

Table 3: Genomic heritability estimates of image-derived phenotypes across maize devel- opment with (B+) or without (B-) plant growth-promoting bacteria inoculation. PH_{HTP}: image-derived plant height; PH: manually measured plant height; CC: canopy coverage; SDM: shoot dry mass; f(biomass): plant biomass index; CV: canopy volume; and NL: num- ber of fully expanded leaves.

NL	PH_{HTP}		CC		f(biomass)		CV	
	B+	B-	B+	B-	B+	B-	B+	B-
2	0.54	0.48	0.46	0.43	0.46	0.45	0.31	0.29
3	0.36	0.44	0.33	0.37	0.32	0.40	0.23	0.33
4	0.43	0.36	0.35	0.36	0.36	0.34	0.23	0.22
5	0.41	0.44	0.23	0.22	0.25	0.24	0.21	0.21
6	0.35	0.34	0.23	0.23	0.24	0.25	0.27	0.25

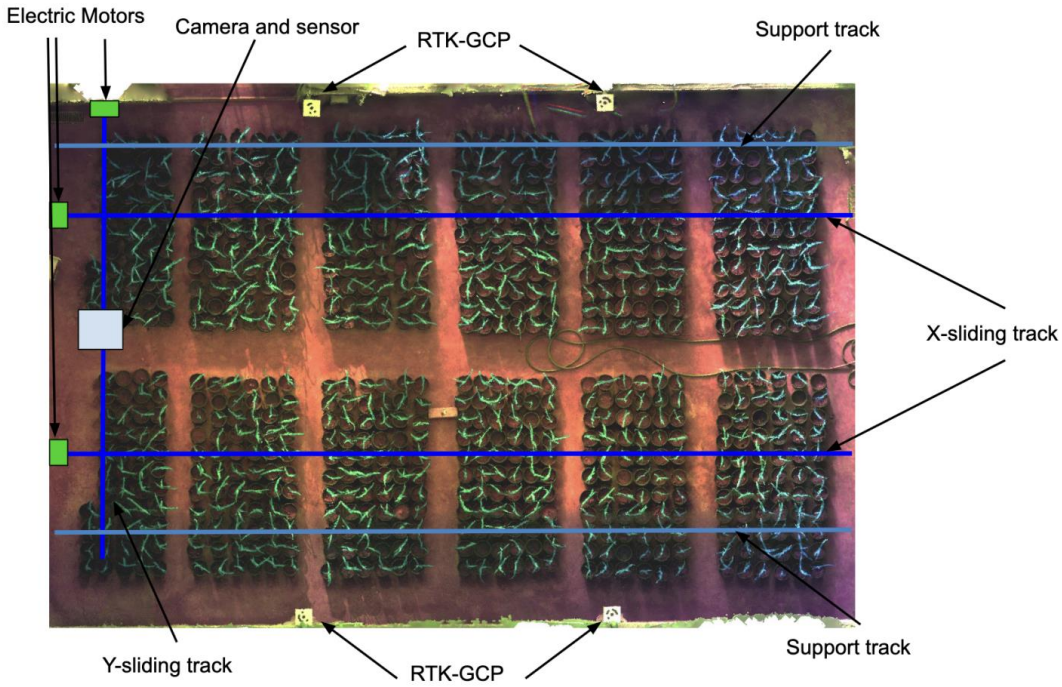

Figure 1: Summary of the image acquisition using a low-cost high-throughput phenotyping platform for greenhouse experiments. The dark and light blue lines indicate the y and x sliding tracks and support tracks, respectively. The green boxes indicate the positions of the electric motors. The blue square is the mobile platform that contains the multispectral camera and the sensor. Real time kinematic ground control points (RTK-GCP) were usedto assemble the orthomosaics.

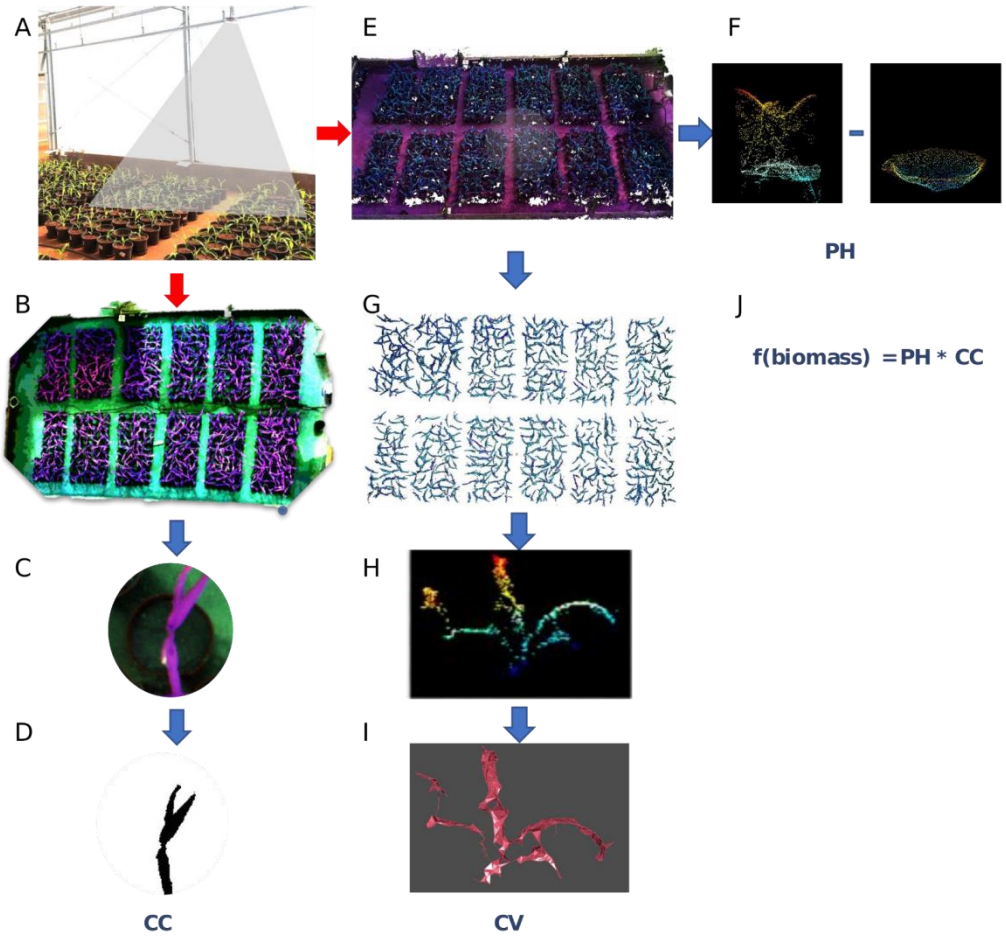

Figure 2: Summary of the multispectral image processing. (A) image acquisition; (B) mo- saicking; (C) plot clip; (D) canopy coverage (CC), (E) dense cloud point; (F) plant height (PH) was calculated from the difference between the dense cloud point with the plants and the dense cloud point with only the pot; (G) dense cloud point after applying the filter to remove the background; (H) dense cloud point for each plot; (I) 3-D reconstruction of the dense cloud point to obtain canopy volume (CV); and (J) f(biomass) (plant biomass index) was obtained from PH and CC.

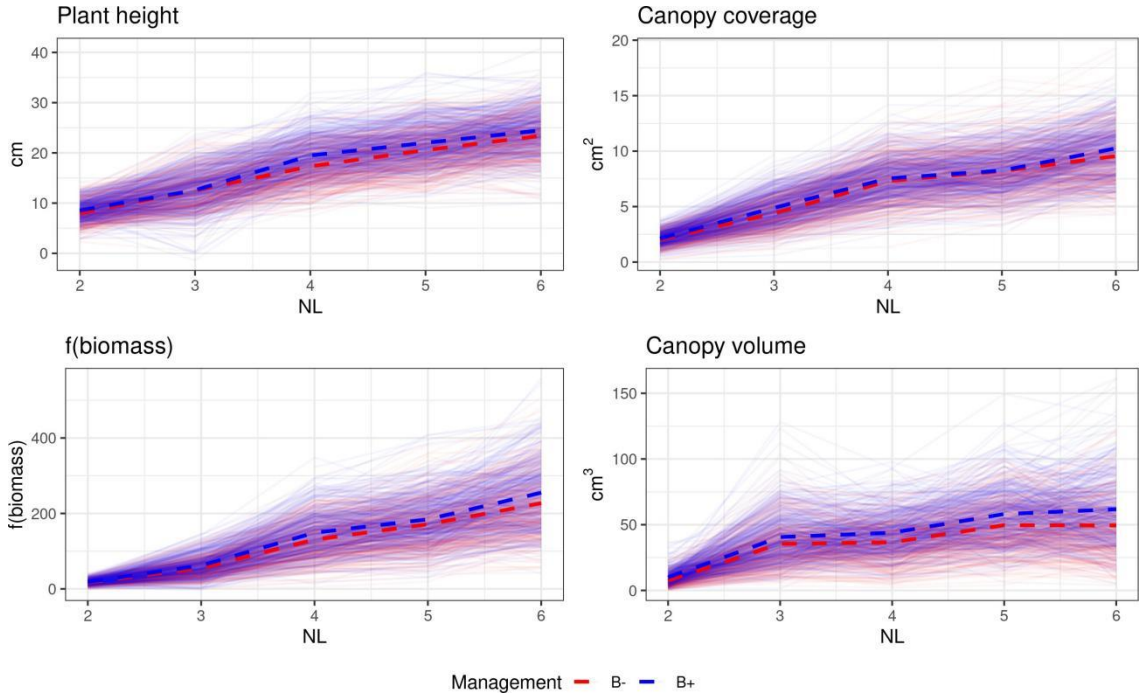

Figure 3: Growth patterns of genotypes across maize development with (B+) or without (B-) plant growth-promoting bacteria inoculation. The blue and red dashed lines represent the means of B+ and B- managements, respectively. Each thin colored line represents the mean of a genotype.

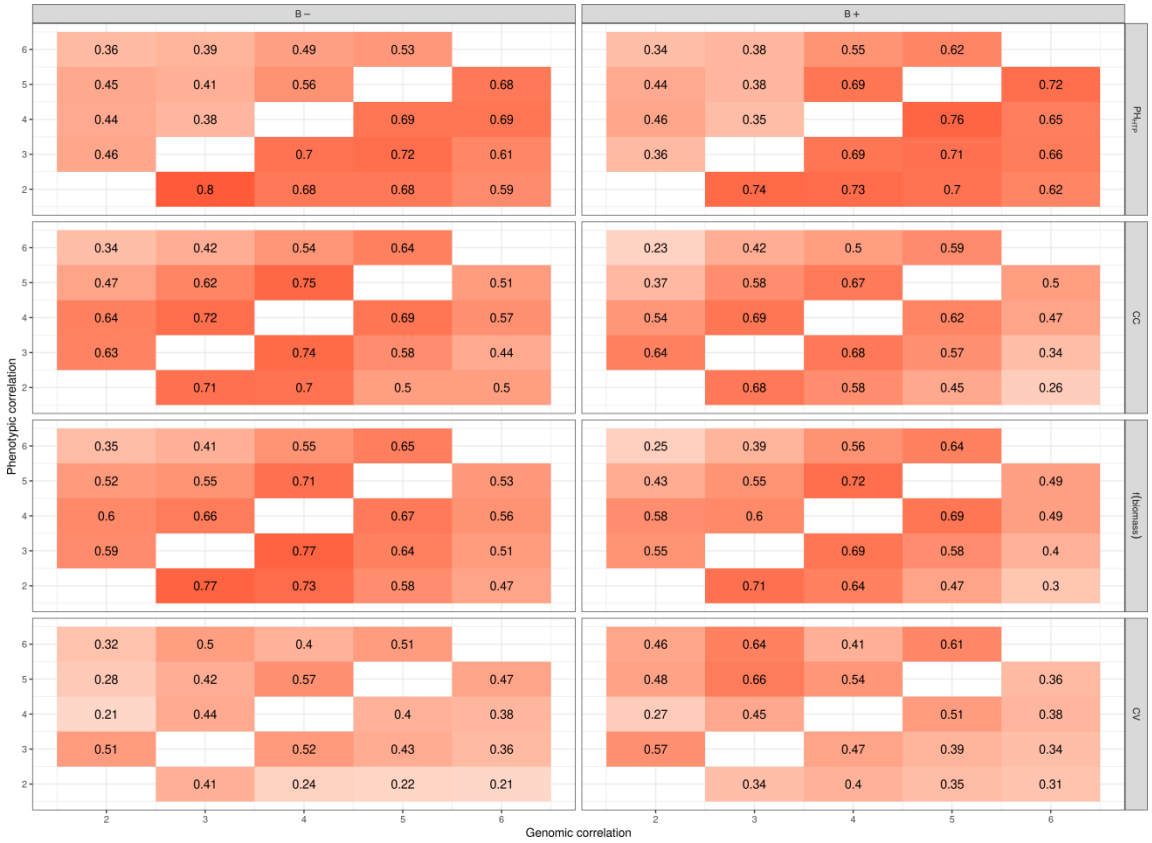

Figure 4: Graphical display of phenotypic and genomic correlations for image-derived pheno- types across maize development (number of leaves varied from 2 to 6). The upper and lower diagonal elements show phenotypic and genomic correlations, respectively. PH_{HTP}: plant height from high-throughput phenotyping; CC: canopy coverage, f(biomass): plant biomass index; and CV: canopy volume.

4. CLASSIFICATION OF PLANT GROWTH-PROMOTING BACTERIA INOCULATION STATUS AND PREDICTION OF GROWTH-RELATED TRAITS IN TROPICAL MAIZE USING HYPERSPECTRAL IMAGE AND GENOMIC DATA

Rafael Massahiro Yassue[1,2], Giovanni Galli[1], Roberto Fritsche-Neto[1,3,*], and Gota Morota[2,4,*]

[1]Department of Genetics, 'Luiz de Queiroz' College of Agriculture, University of São Paulo, São Paulo, Brazil

[2]Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, USA

[3]Quantitative Genetics and Biometrics Cluster, International Rice Research Institute, Los Baños, Philippines

[4]Center for Advanced Innovation in Agriculture, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 USA

Running title: Phenomic and genomic prediction in tropical maize

Keywords: automated machine learning, complex trait prediction, hyperspectral phenotyping, phenomic prediction.

Core ideas

- Hyperspectral reflectance data can classify plant growth-promoting bacteria inoculation status
- Phenomic prediction performs better than genomic prediction depending on the target phenotype
- AutoML is a promising approach for automating hyperparameter tuning for classification and prediction

* Corresponding authors:

E-mail:r.fritscheneto@irri.org and morota@vt.edu

ORCID: 0000-0002-7424-2227 (RMY), 0000-0002-3400-7978 (GG), 000-0003-4310-0047 (RFN), and 0000-0002-3567-6911 (GM)

Abbreviations: automated machine learning (AutoML), Bayesian ridge regression (BayesRR), best linear unbiased estimators (BLUE), least absolute shrinkage and selection operator (LASSO), ordinary least squares (OLS), partial least squares (PLS), partial least squares discriminant analysis (PLS-DA), partial least squares regression (PLS-R), plant growth-promoting bacteria (PGPB), plant height (PH), shoot dry mass (SDM), single-nucleotide polymorphism (SNP), stalk diameter (SD), with plant growth-promoting bacteria inoculation (B+), and without plant growth-promoting bacteria inoculation (B-).

Abstract

Recent technological advances in high-throughput phenotyping have created new opportunities for the prediction of complex traits. In particular, phenomic prediction using hyperspectral reflectance could capture various signals that affect phenotypes genomic prediction might not explain. A total of 360 inbred maize lines with or without plant growth-promoting bacterial inoculation management under nitrogen stress were evaluated using 150 spectral wavelengths ranging from 386 to 1021 nm and 13,826 single-nucleotide polymorphisms. Six prediction models were explored to assess the predictive ability of hyperspectral and genomic data for inoculation status and plant growth-related traits. The best models for hyperspectral prediction were partial least squares and automated machine learning. The Bayesian ridge regression and BayesB were the best performers for genomic prediction. Overall, hyperspectral prediction showed greater predictive ability for shoot dry mass and stalk diameter, whereas genomic prediction was better for plant height. The prediction models that simultaneously accommodated both hyperspectral and genomic data resulted in a predictive ability as high as that of phenomics or genomics alone. Our results highlight the usefulness of hyperspectral-based phenotyping for management and phenomic prediction studies.

4.1 Introduction

Addressing the growing food demand by increasing sustainable food production is critical in agriculture. The use of plant growth-promoting bacteria (PGPB) as

inoculants to increase plant productivity and resilience to biotic and abiotic stresses has recently gained traction (Vejan et al., 2016; Majeed et al., 2018). However, an effective assessment of PGPB responses is difficult because the interaction between the PGPB × host genotype × environment is complex (Wintermans et al., 2016; Xiao et al., 2017). A method that can easily and accurately phenotype and predict plant growth-related traits under PGPB inoculation is needed (Rouphael et al., 2018; Susič et al., 2020).

Whole-genome molecular markers have been widely used for complex trait prediction (Meuwissen et al., 2001; Crossa et al., 2013; Fritsche-Neto et al., 2021). Genomic prediction performance mainly relies on the genetic relationship between individuals in the reference and target populations and linkage disequilibrium between genetic markers and quantitative trait loci. Prediction performance becomes suboptimal when the aforementioned relationship is weak or markers do not sufficiently capture quantitative trait loci signals (Habier et al., 2007; Windhausen et al., 2012; Sallam et al., 2015).

Hyperspectral cameras, sensors capable of capturing images in a wide spectrum of wave-lengths, have recently been added to the realm of phenotyping tools available for plant genetics and breeding applications. It has been reported that the hyperspectral signatures of plant canopies are associated with plant nutrient status (Cilia et al., 2014; Mahajan et al.,2016), plant growth-related traits (Yang and Chen, 2004; Kaur et al., 2015), plant biomass (Jia et al., 2019; Ma et al., 2020), plant health (Lowe et al., 2017; Thomas et al., 2017), genotype discrimination (Chivasa et al., 2019), leaf water content (Ge et al., 2016), and soil microbial community composition (Carvalho et al., 2016). In particular, high-throughput phenotyping data-driven complex trait prediction, which is also known as phenomic prediction, is an active research topic for categorical and continuous phenotypes (Edlich-Muth et al., 2016; Rincent et al., 2018; Krause et al., 2019; Cuevas et al., 2019; Shu et al., 2021; Wang et al., 2021). Phenomic prediction is expected to capture the molecular composition of a plant, such as biochemical or physiological signals (endophenotypes), influencing phenotypes that genomic prediction may not directly explain (Rincent et al., 2018). Hyperpectral reflectance data can be used to evaluate plant growth- or stress-related phenotypes in response to PGPB inoculation.

Several statistical learning models have been applied to phenomic prediction using hyperspectral image data. These include Bayesian ridge regression (BayesRR), least absolute shrinkage and selection operator (LASSO), partial least squares (PLS), BayesB, and neural networks (Montesinos-López et al., 2017b; Nigon et al., 2020; Qun'ou et al., 2021; Yoosefzadeh-Najafabadi et al., 2021). The use of the entire spectrum simultaneously, rather than selecting a small set of known wavelengths (e.g., spectra indices), can be beneficial for complex trait prediction (Aguate et al., 2017; Montesinos-López et al., 2017b,a). In general, fitting a machine learning model with good accuracy requires knowledge of the model structure to tune the hyperparameters. However, optimal hyperparameters are difficult to determine and often tuned manually using naive approaches (van Rijn and Hutter, 2018; Yang and Shami, 2020). As an alternative, automated machine learning (AutoML) has been proposed to reduce the need for human interference during the hyperparameter tuning process so that the application of machine learning becomes more automated and precise. (Feurer et al., 2015; Jin et al., 2019). Despite its potential, the use of AutoML for hyperspectral-based phenomic prediction of complex traits has not yet been fully explored.

Therefore, the objectives of this study were to 1) evaluate the utility of hyperspectral image data to classify PGPB inoculation status, and 2) compare the predictive ability of genomic prediction, hyperspectral prediction, and their combination for growth-related phenotypes under PGPB management.

4.2 Materials and Methods

Plant growth-promoting bacteria experiment

A tropical maize association panel containing 360 inbred lines was used to study responses to PGPB. The inbred lines were evaluated with (B+) and without (B-) PGPB inoculation under nitrogen stress. The B+ management system consisted of a synthetic population of four PGPB, composed by *Bacillus thuringiensis* RZ2MS9, *Delftia* sp. RZ4MS18 (Batista et al., 2018, 2021), *Pantoea agglomerans* 33.1 (Quecine et al., 2012), and *Azospirillum brasilense* Ab-v5 (Hungria et al., 2010). The B- management included an inoculum with only liquid Luria-Bertani medium. Fertilization, irrigation, and other cultural practices were conducted according to the crop needs, except for nitrogen, which was not supplied. The phenotyping was performed when most of the plants had six expanded leaves. The manually measured

phenotypes were plant height (PH), stalk diameter (SD), and shoot dry mass (SDM). Further information on the experimental design is available in Yassue et al. (2021a,b).

Genomic data

A genotyping-by-sequencing method followed by the two-enzymes (PstI and MseI) protocol (Sim et al., 2012; Poland et al., 2012) was used to generate a total of 13,826 single-nucleotide polymorphisms (SNPs). The cetyltrimethylammonium bromide method was used to extract DNA (Doyle and Doyle, 1987). TASSEL 5.0 (Bradbury et al., 2007) was used to perform SNP calling with B73 (B73-RefGen v4) as the reference genome. The SNP markers were removed if the call rate was less than 90%, non-biallelic, or the minor allele frequency was less than 5%. Missing marker codes were imputed using Beagle 5.0 (Browning et al., 2018). Markers with pairwise linkage disequilibrium higher than 0.99 were removed using the SNPRelate R package (Zheng et al., 2012).

Hyperspectral imaging and processing

Hyperspectral images of the maize lines grown under B+ and B – management conditions were obtained using a benchtop system of Pika L. camera (Resonon, Bozeman, MT, USA). The last expanded leaf was cut at the base of the stalk and immediately taken to the laboratory for imaging. The leaves were refrigerated using a cooler and gel refrigerant packs. Images were collected inside the dark room with a light supply to control for light variation. Radiometric calibration was performed using white and black tile panels approved by the camera system manufacturer. Each image cube (height, length, and band), consisted of 150 bands, varied from 386 to 1021 nm. The region of interest for each cube was the middle of the leaf. Once images were collected, the image processing step included applying a mask to remove the background and calculating the mean value of the reflectance from each pixel that referred to plant tissue. Hyperspectral image processing was performed using the Spectral Python (SPy) package. A summary of the data acquisition and processing is shown in Figure 1.

Prediction model

First, the utility of hyperspectral reflectance data to classify the PGPB inoculation status (B+ or B-) was evaluated. The purpose was to investigate whether the differences in genotypes' biochemical or physiological signals between the two inoculation status are reflected at the hyperspectral level. If so, we expect to see a reasonable classification accuracy of the PGPB status using hyperspectral data. In this scenario, a classification accuracy of 0.5 is expected when genomic data are used as predictors because genomics is irrelevant to the presence or absence of inoculation. Second, the predictive abilities of genomic prediction, hyperspectral prediction, and their combination were compared for growth-related phenotypes, including PH, SD, and SDM.

The best linear unbiased estimators (BLUE) of genotypes were obtained using the fol-lowing equation:

$$\mathbf{y} = \mu + \mathbf{X}_1\mathbf{r} + \mathbf{X}_2\mathbf{b} + \mathbf{X}_3\mathbf{g} + \boldsymbol{E},$$

where $\mathbf{y}$ is the vector of phenotypes (PH, SD, SDM, and hyperspectral reflectance); $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ are the incidence matrices for the fixed effects; μ is the intercept; $\mathbf{r}$, $\mathbf{b}$, and $\mathbf{g}$ are the fixed effects for replication, block within replication, and genotypes, respectively; and $\boldsymbol{E} \sim N(0, \mathbf{I}\sigma^2_e)$ is the random residual effect, where $\mathbf{I}$ is the identity matrix and σ^2_E is the residual variance. The analysis was performed using the R package ASReml-R (Butler et al., 2017).

The following classification and prediction models were used, which are summarized in Table 1.

Logistic regression and ordinary least squares: Logistic regression and ordinary least squares (OLS) were used to classify the inoculation status and predict growth-related traits, respectively, using the hyperspectral reflectance data. These models were not used for ge- nomic prediction because the number of predictors was greater than the number of samples (maize lines).

Partial least squares: Partial least squares discriminant analysis (PLS-DA) and partial least squares regression (PLS-R) models (Wold et al., 2001) were fit using the caret R package (Kuhn, 2008). Partial least squares identifies latent variables that maximize the covariance between the predictors and responses while minimizing the

error. The optimal number of the latent variables was estimated in the inner training set using k-fold cross-validation, with k set to four.

Least absolute shrinkage and selection operator: Least absolute shrinkage and selection operator was fit using the glmnet R package for linear regression as well as for generalized linear models (Friedman et al., 2010). The tuning parameter lambda was estimated by k-fold cross-validation with k set to four and was chosen according to the minimum mean cross-validation error.

BayesRR and BayesB: BayesRR and BayesB were fit for classification and prediction by treating the covariates as random (Meuwissen et al., 2001). For BayesRR, marker effects were sampled from a univariate normal distribution with a null mean and marker variance σ^2_a, having a scaled inverse χ^2 prior with scale parameter S_α and $\nu_\alpha = 4$ degrees of freedom.

BayesB assumes that a priori, the marker effects have identical and independent mixture distributions, where each has a point mass at zero with a probability π and a scaled t distribution with a probability 1-π having a null mean and marker variance σ^2_α, scale parameter S_α, and $\nu_\alpha = 4$ degrees of freedom. The mixture parameter, π, was set to 0.99. For both BayesRR and BayesB, a flat prior was assigned to the intercept. The scale parameter was chosen such that the prior mean of σ^2_a equals half of the phenotypic variance. All the Bayesian models were fitted using 60,000 Markov chain Monte Carlo samples, 6,000 burn-ins, and a thinning rate of 60 implemented in JWAS (Cheng et al., 2018). Model convergence was assessed using trace plots of the posterior means of the parameters.

Automated machine learning: Auto-sklearn automated machine learning aims to tune hyperparameters automatically by leveraging meta-learning, Bayesian optimization, and ensemble learning. The automated machine learning algorithm uses a meta-learning pro- cess that is quick but roughly explores the entire machine learning configuration space, which is then followed by Bayesian optimization to fine-tune the hyperparameters for per- formance. Finally, the ensemble process combines several machine learning models with different weights to increase predictive ability.

The time limits for searching for an appropri- ate model and a single call were set to 900 s and 30 s, respectively. The AutoSklearnClassifier and AutoSklearnRegressor functions from AutoSklearn 0.14.2 (Feurer et al., 2015) were used for classification and prediction, respectively.

Multi-omic prediction: Multi-omic data integration may increase prediction performance relative to single-omic data (Krause et al., 2019; Galán et al., 2020; Guo et al., 2020). Multi-omic prediction was performed by combining hyperspectral and genomic data using BayesRR and BayesB for growth-related phenotypes by setting different priors for each omic covariate. The framework closely followed that of Gonçalves et al. (2021) and Baba et al. (2021). The mixture parameter π was set to 0.99 for hyperspectral and genomic terms in multi-omics BayesB.

Predictive performance

We used repeated random sub-sampling cross-validation to evaluate model performance. We split the population into training (80%) and test (20 %) sets, while maintaining balanced classes for inoculation categories (B+ and B-) and genotypes (Figure 2). The B+ and B- management conditions were jointly used for the classification models. Five genotypes were removed from the analysis to maintain a balanced structure because they were present only in one management. The training set was further split into inner training and validation sets for the models that required hyperparameter tuning. The inner training set was used for fine-tuning hyperparameters. The final model performance was evaluated in an independent test set that was not used in model training (Figure 2A).

For growth-related phenotypes, predictions were performed separately for each management condition (B+ and B-) (Figure 2B). The hyperparameters were tuned in the inner training set, similar to the classification tasks. We did not consider the interaction effect between the genotype and management condition because of the lack of such an effect (Yassue et al., 2021a,b). The predictive ability of the model was assessed using the Pearson correlation between the predictive values and BLUE of the genotypes.

4.3 Results

Correlation between growth-related phenotypes and hyperspectral bands

The hyperspectral signature and principal component biplot of the maize genotypes are shown in (Figure 3). No apparent visual patterns or clusters distinguished between the management conditions across the 150 hyperspectral bands. High hyperspectral variability was observed at the green and near-infrared wavelengths. The correlation matrix of the hyperspectral reflectance data showed that nearby bands had a higher correlation. As the distance between bands increased, the correlation decreased (Figure S1).

The correlation between the single-band reflectance and growth-related phenotypes varied between -0.1 to 0.40, depending on the target phenotype and management condition (Figure 4). Overall, SDM was the most correlated trait with band reflectance values. Additionally, the correlation with band reflectance was higher for the inoculated samples (B+). Most of the bands were associated with the target phenotypes and the correlation peaks occurred between blue and green for PM and between RedEdge and near-infrared for SD and SDM.

Classification of inoculation status

The accuracy of classifying the inoculation status (B+ and B-) using hyperspectral reflectance data is shown in Figure 5. AutoML and PLS were the best classification models, followed by OLS and LASSO. The accuracy achieved by AutoML and PLS was higher than 0.8, demonstrating that the hyperspectral profiles of the B + and B- inoculation status were distinct. However, BayesRR and BayesB did not perform well, and their performance was worse than that of OLS.

Prediction of growth-related phenotypes

The performance of hyperspecral-driven phenomic prediction and SNP-driven genomic prediction is shown in Figure 6. We obtained the highest and lowest predictions for PH and SDM, respectively, using phenotypic prediction. Furthermore, we observed higher predictions for B+ than for B- in PH, whereas the opposite was observed for SD. The predictions were higher for B+, except when BayesB was used.

AutoML, PLS, and LASSO performed equally well in predicting SD and SDM. In contrast, no notable differences were observed in PH. In genomic prediction, the best prediction was obtained for PH. The B+ management conditions were more predictable than B-. Overall, the predictive performance was not sensitive to the choice of the genomic prediction model. The multi-omics models improved the prediction correlations for all phenotypes compared with the single omics model; however, this gain was incremental.

4.4 Discussion

Phenotyping PGPB response

Recent studies have found that inoculation with PGPB can modify plant structure and increase plant biomass and resilience to nitrogen stress (Yassue et al., 2021a,b). This study evaluated PGPB responses at the level of hyperspectral reflectance data. The ability to classify the inoculation status using phenomics showed that the hyperspectral camera could capture signals unique to each management condition. Previous studies have reported that the PGPB species used in this study are capable of producing indole acetic acid, fixing nitrogen, and promoting phosphate solubilization (Quecine et al., 2012; Batista et al., 2018, 2021). A field trial study reported that *Azospirillum brasilense* can increase nitrogen, potassium, boron in the leaves of maize (Hungria et al., 2010). However, indole acetic acid production or nutrient status was not evaluated in the present study, and the interpretation of the hyperspectral signature is limited.

The results of our study align with those of Carvalho et al. (2016), who showed that leaf hyperspectral patterns in winter wheat have the potential to detect changes in soil microbial communities. Moreover, Susič et al. (2020) reported successful classification of PGPB inoculation status for nematicidal effects in tomatoes using hyperspectral image data. They found that the hyperspectral signature can be used to assess plant stress after inoculation with PGPB. The minor difference in hyperspectral reflectance curves between the B+ and B- management conditions, in addition to the lack of clustering in PCA, suggests that most of the bands probably contributed to inoculation status classification. Further studies should be conducted to evaluate the structural, morphological, and chemical differences between B+ and B- management conditions.

Model performance

The statistical modeling of high-throughput phenotyping data in quantitative genetics is becoming increasingly important (Morota et al., 2022). We evaluated the utility of hyperspectral-based phenomic prediction to classify PGPB inoculation status and compared the predictive ability of genomic prediction, hyperspectral prediction, and data integration for growth-related phenotypes using statistical prediction models in tropical maize. Generally, PLS and AutoML were competitive in many scenarios and the performance of PLS in our study agreed with previous work that used smaller datasets (Fu et al., 2019; Galli et al., 2020; Shu et al., 2021). Montesinos-López et al. (2017b) also reported that PLS-R performed better than BayesB for predicting wheat yield using hyperspectral reflectance.

AutoML performed equally well or better than PLS for phenomic classification, suggesting the usefulness of hyperparameter tuning and ensemble learning in machine learning (Figure 5). We obtained better predictions in the B+ management conditions for PH, SDM, and for SD in the B- management conditions. This could be explained by the extent of correlation between growth-related phenotypes and hyperspectral reflectance (Figure 3).

For genomic prediction, the higher predictive ability for PH was probably due to its higher heritability in comparison to SD and SDM (Yassue et al., 2021a,b). Similar to the phenomic prediction, AutoML yielded relatively good predictions. In addition to AutoML, BayesRR and BayesB were competitive and stable across the three phenotypes investigated. This was expected because these models are well accepted in the genomic prediction literature. However, BayesRR and BayesB did not perform well with hyperspectral reflectance data. Simple OLS outperformed BayesRR and BayesB in some cases, suggesting that shrinkage or variable selection is not necessarily beneficial when the number of predictors is lower than the number of samples (Whittaker et al., 2000). The strength of OLS is that the estimated effect has the property of the best linear unbiased estimator, and its expectation is equal to the true effect if the Gauss–Markov theorem is satisfied (Searle and Gruber, 2016). This property appears to be useful for hyperspectral prediction.

Phenomic vs genomic prediction

The hyperspectral signature of plants is considered an endophenotype because it can capture the expression of genotypes under specific conditions and is useful in predicting complex traits (Rincent et al., 2018). We found that phenomic prediction was more predictive than genomic prediction for SD and SDM. In contrast, genomic prediction was better than phenomic prediction for PH. Although the multi-omic model produced the highest predictive correlations for all three phenotypes by integrating genomic and hyperspectral information, there was no noticeable enhancement over the best single-omic prediction. Our results are consistent with those of previous studies (Xu et al., 2017; Goncalves et al., 2021) reporting for different species or omic data combinations.

Overall, our results showed that the hyper-spectral signature of genotypes is a valuable resource for complex trait prediction. Further studies are needed to improve hyperspectral image acquisition for greenhouse experiments. The genetic gain equation, also known as the breeder's equation, is R_t = irσaL/t where R_t is the response to selection by time, i is the selection intensity, r is the selection accuracy, σ_a is the square root of additive genetic variation, and L is the generation interval (Cobb et al., 2019). Phenomic prediction has the potential to alter i and r because it can increase the selection intensity and accuracy by phenotyping a larger number of genotypes. Conversely, genomic prediction may also increase accuracy and reduce the generation interval. Because the hyperspectral signature is an endophenotype, it can be influenced by environmental effects, unlike genomic information, which is specific to the individual. Hence, the use of phenomic or genomic prediction models depends on the goal of selection (Hickey et al., 2017).

The disadvantage of using a benchtop camera is the need to bring plants to an imaging room. In this study, the region of interest was the middle portion of the leaf, which required manual collection of maize leaves. This could limit the applicability hyperspectral imaging in plant breeding or genetics program pipelines because of the laborious data collection time. The use of hyperspectral cameras in a low-cost, high-throughput phenotyping platform, such as that reported by Yassue et al. (2021b), may ease the application of phenomic prediction that uses hyperspectral reflectance data.

4.5 Conclusions

We found that hyperspectral reflectance data were useful predictors for classifying PGPB inoculation status and predicting growth-related phenotypes in tropical maize. Phenomic prediction showed better performance than genomic prediction for SD and SDM. AutoML is a promising approach for classification and prediction tasks that mitigate manual hyper-parameter tuning. The integration of hyperspectral and genomic data resulted in predictive performance as high as that of the best single-omic model. Overall, our results highlight the usefulness of hyperspectral imaging in management and phenomic prediction studies.

Acknowledgments

The authors acknowledge Pedro Takao Yamamoto and Fernando Henrique Iost Filho for their support in collecting the hyperspectral images.

Funding

This study was supported in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Grant #17/24327-0, #19/04697-2, and #2017/19407-4 from São Paulo Research Foundation (FAPESP), International Business Machines Corporation (IBM, Brasil), and Virginia Polytechnic Institute and State University.

Author contributions

Rafael Massahiro Yassue: Conceptualization; Data curation, Formal analysis, Investigation; Methodology; Visualization; Writing-original draft; Writing-review & editing. Giovanni Galli: Investigation; Methodology; Writing-review & editing. Roberto Fritsche-Neto: Conceptualization; Funding acquisition; Supervision; Writing-review & editing. Gota Morota: Conceptualization; Methodology; Funding acquisition; Supervision; Writing-original draft; Writing-review & editing.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Aguate, F. M., Trachsel, S., González-Pérez, L., Burgueño, J., Crossa, J., Balzarini, M., Gouache, D., Bogard, M., and De los Campos, G. (2017). Use of hyperspectral image data outperforms vegetation indices in prediction of maize yield. *Crop Science*.

Baba, T., Pegolo, S., Mota, L. F., Peñagaricano, F., Bittante, G., Cecchinato, A., and Morota, G. (2021). Integrating genomic and infrared spectral data improves the prediction of milk protein composition in dairy cattle. *Genetics Selection Evolution*, 53(1):1–14.

Batista, B. D., Dourado, M. N., Figueredo, E. F., Hortencio, R. O., Marques, J. P. R., Piotto, F. A., Bonatelli, M. L., Settles, M. L., Azevedo, J. L., and Quecine, M. C. (2021). The auxin-producing bacillus thuringiensis RZ2ms9 promotes the growth and modifies the root architecture of tomato (solanum lycopersicum cv. micro-tom). *Archives of Microbiology*.

Batista, B. D., Lacava, P. T., Ferrari, A., Teixeira-Silva, N. S., Bonatelli, M. L., Tsui, S., Mondin, M., Kitajima, E. W., Pereira, J. O., Azevedo, J. L., and Quecine, M. C. (2018). Screening of tropically derived, multi-trait plant growth- promoting rhizobacteria and evaluation of corn and soybean colonization ability. *Microbiological Research*, 206:33–42.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19):2633–2635.

Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, 103(3):338–348.

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., and Thompson, R. (2017). *ASReml-R Reference Manual Version 4*.

Carvalho, S., van der Putten, W. H., and Hol, W. H. G. (2016). The potential of hyperspectral patterns of winter wheat to detect changes in soil microbial community composition. *Frontiers in Plant Science*, 7.

Cheng, H., Fernando, R., and Garrick, D. (2018). JWAS: Julia implementation of whole-genome analysis software. In *Proceedings of the world congress on genetics applied to livestock production*.

Chivasa, W., Mutanga, O., and Biradar, C. (2019). Phenology-based discrimination of maize (zea mays l.) varieties using multitemporal hyperspectral data. *Journal of Applied Remote Sensing*, 13(01):1.

Cilia, C., Panigada, C., Rossini, M., Meroni, M., Busetto, L., Amaducci, S., Boschetti, M., Picchi, V., and Colombo, R. (2014). Nitrogen status assessment for variable rate fertilization in maize through hyperspectral imagery. *Remote Sensing*, 6(7):6549–6565.

Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., Hagen, T., Quinn, M., and Ng, E. H. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theoretical and Applied Genetics*, 132(3):627–645.

Crossa, J., Pérez, P., Hickey, J., Burgueño, J., Ornella, L., Cerón-Rojas, J., Zhang, X., Dreisigacker, S., Babu, R., Li, Y., Bonnett, D., and Mathews, K. (2013). Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity*, 112(1):48–60.

Cuevas, J., Montesinos-López, O., Juliana, P., Guzmán, C., Pérez-Rodríguez, P., González-Bucio, J., Burgueño, J., Montesinos-López, A., and Crossa, J. (2019). Deep kernel for genomic and near infrared predictions in multi-environment breeding trials. *G3: Genes, Genomes, Genetics*, 9(9):2913–2924.

Doyle, J. and Doyle, J. (1987). A rapid dna isolation procedure for small quantities of fresh leaf tissue. *PHYTOCHEMICAL BULLETIN*, 17(RESEARCH). Edlich-Muth, C., Muraya, M. M., Altmann, T., and Selbig, J. (2016). Phenomic prediction of maize hybrids. *Biosystems*, 146:102–109.

Feurer, M., Klein, A., Eggensperger, Katharina Springenberg, J., Blum, M., and Hutter, F. (2015). Efficient and robust automated machine learning. In *Advances in Neural Information Processing Systems 28 (2015)*, pages 2962–2970.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1).

Fritsche-Neto, R., Galli, G., Borges, K. L. R., Costa-Neto, G., Alves, F. C., Sabadin, F., Lyra, D. H., Morais, P. P. P., de Andrade, L. R. B., Granato, I., and Crossa, J. (2021). Optimizing genomic-enabled prediction in small-scale maize hybrid breeding programs: A roadmap review. *Frontiers in Plant Science*, 12.

Fu, P., Meacham-Hensold, K., Guan, K., and Bernacchi, C. J. (2019). Hyperspectral leaf reflectance as proxy for photosynthetic capacities: An ensemble approach based on multiple machine learning algorithms. *Frontiers in Plant Science*, 10.

Galán, R. J., Bernal-Vasquez, A.-M., Jebsen, C., Piepho, H.-P., 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(11):3001–3015.

Galli, G., Horne, D. W., Collins, S. D., Jung, J., Chang, A., Fritsche-Neto, R., and Rooney, W. L. (2020). Optimization of UAS-based high-throughput phenotyping to estimate plant health and grain yield in sorghum. *The Plant Phenome Journal*, 3(1).

Ge, Y., Bai, G., Stoerger, V., and Schnable, J. C. (2016). Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *Computers and Electronics in Agriculture*, 127:625–632.

Gonçalves, M. T. V., Morota, G., Costa, P. M. d. A., Vidigal, P. M. P., Barbosa, M. H. P., and Peternelli, L. A. (2021). Near-infrared spectroscopy outperforms genomics for predicting sugarcane feedstock quality traits. *PloS one*, 16(3):e0236853.

Guo, J., Pradhan, S., Shahi, D., Khan, J., Mcbreen, J., Bai, G., Murphy, J. P., and Babar, M. A. (2020). Increased prediction accuracy using combined genomic information and physiological traits in a soft wheat panel evaluated in multi-environments. *Scientific reports*, 10(1):1–12.

Habier, D., Fernando, R. L., and Dekkers, J. C. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177(4):2389–2397.

Hickey, J. M., , Chiurugwi, T., Mackay, I., and Powell, W. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature Genetics*, 49(9):1297–1303.

Hungria, M., Campo, R. J., Souza, E. M., and Pedrosa, F. O. (2010). Inoculation with selected strains of azospirillum brasilense and a. lipoferum improves yields of maize and wheat in brazil. *Plant and Soil*, 331(1-2):413–425.

Jia, M., Li, W., Wang, K., Zhou, C., Cheng, T., Tian, Y., Zhu, Y., Cao, W., and Yao, X. (2019). A newly developed method to extract the optimal hyperspectral feature for monitoring leaf biomass in wheat. *Computers and Electronics in Agriculture*, 165:104942.

Jin, H., Song, Q., and Hu, X. (2019). Auto-keras: An efficient neural architecture search system. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1946–1956. ACM.

Kaur, R., Singh, B., Singh, M., and Thind, S. K. (2015). Hyperspectral indices, correlation and regression models for estimating growth parameters of wheat genotypes. *Journal of the Indian Society of Remote Sensing*, 43(3):551–558.

Krause, M. R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R. P., Dreisigacker, S., Poland, J., Rutkoski, J., Sorrells, M., Gore, M. A., and Mondal, S. (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3 Genes|Genomes|Genetics*, 9(4):1231–1247.

Kuhn, M. (2008). Building predictive models in r using the caret package. *Journal of Statistical Software, Articles*, 28(5):1–26.

Lowe, A., Harrison, N., and French, A. P. (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods*, 13(1).

Ma, D., Maki, H., Neeno, S., Zhang, L., Wang, L., and Jin, J. (2020). Application of non-linear partial least squares analysis on prediction of biomass of maize plants using hyperspectral images. *Biosystems Engineering*, 200:40–54.

Mahajan, G. R., Pandey, R. N., Sahoo, R. N., Gupta, V. K., Datta, S. C., and Kumar, D. (2016). Monitoring nitrogen, phosphorus and sulphur in hybrid rice (oryza sativa l.) using hyperspectral remote sensing. *Precision Agriculture*, 18(5):736–761.

Majeed, A., Muhammad, Z., and Ahmad, H. (2018). Plant growth promoting bacteria: role in soil improvement, abiotic and biotic stress management of crops. *Plant Cell Reports*, 37(12):1599–1609.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4):1819–1829.

Montesinos-López, A., Montesinos-López, O. A., Cuevas, J., Mata-López, W. A., Burgueño, J., Mondal, S., Huerta, J., Singh, R., Autrique, E., González-Pérez, L., et al. (2017a). Genomic bayesian functional regression models with interactions for predicting wheat grain yield using hyper-spectral image data. *Plant Methods*, 13(1):1–29.

Montesinos-López, O. A., Montesinos-López, A., Crossa, J., de los Campos, G., Alvarado, G., Suchismita, M., Rutkoski, J., González-Pérez, L., and Burgueño, J. (2017b). Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods*, 13(1).

Morota, G., Jarquin, D., Campbell, M. T., and Iwata, H. (2022). Statistical methods for the quantitative genetic analysis of high-throughput phenotyping data. In *High-Throughput Plant Phenotyping: Methods and Protocols*, pages 237–274. Springer.

Nigon, T., Yang, C., Paiao, G. D., Mulla, D., Knight, J., and Fernández, F. (2020). Prediction of early season nitrogen uptake in maize using high-resolution aerial hyperspectral imagery. *Remote Sensing*, 12(8):1234.

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J.-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, 7(2):e32253.

Quecine, M. C., Araújo, W. L., Rossetto, P. B., Ferreira, A., Tsui, S., Lacava, P. T., Mondin, M., Azevedo, J. L., and Pizzirani-Kleiner, A. A. (2012). Sugarcane growth promotion by the endophytic bacterium pantoea agglomerans 33.1. *Applied and Environmental Microbiology*, 78(21):7511–7518.

Qun'ou, J., Lidan, X., Siyang, S., Meilin, W., and Huijie, X. (2021). Retrieval model for total nitrogen concentration based on UAV hyper spectral remote sensing data and machine learning algorithms – a case study in the miyun reservoir, china. *Ecological Indicators*, 124:107356.

Rincent, R., Charpentier, J.-P., Faivre-Rampant, P., Paux, E., Gouis, J. L., Bastien, C., and Segura, V. (2018). Phenomic selection is a low-cost and high-throughput method based on indirect predictions: Proof of concept on wheat and poplar. *G3 Genes|Genomes|Genetics*, 8(12):3961–3972.

Rouphael, Y., Spíchal, L., Panzarová, K., Casa, R., and Colla, G. (2018). High-throughput plant phenotyping for developing novel biostimulants: From lab to field or from field to lab? *Frontiers in Plant Science*, 9.

Sallam, A. H., Endelman, J. B., Jannink, J.-L., and Smith, K. P. (2015). Assessing genomic selection prediction accuracy in a dynamic barley breeding population. *The Plant Genome*, 8(1).

Searle, S. R. and Gruber, M. H. (2016). *Linear models*. John Wiley & Sons.

Shu, M., Shen, M., Zuo, J., Yin, P., Wang, M., Xie, Z., Tang, J., Wang, R., Li, B., Yang, X., and Ma, Y. (2021). The application of UAV-based hyperspectral imaging to estimate crop traits in maize inbred lines. *Plant Phenomics*, 2021:1–14.

Sim, S.-C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M. W., Deynze, A. V., Hamilton, J. P., Buell, C. R., Causse, M., Wijeratne, S., and Francis, D. M. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS ONE*, 7(7):e40563.

Susič, N., Žibrat, U., Sinkovič, L., Vončina, A., Razinger, J., Knapič, M., Sedlar, A., Širca, S., and Stare, B. G. (2020). From genome to field—observation of the multimodal nematicidal and plant growth-promoting effects of bacillus firmus i-1582 on tomatoes using hyperspectral remote sensing. *Plants*, 9(5):592.

Thomas, S., Kuska, M. T., Bohnenkamp, D., Brugger, A., Alisaac, E., Wahabzada, M., Behmann, J., and Mahlein, A.-K. (2017). Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective. *Journal of Plant Diseases and Protection*, 125(1):5–20.

van Rijn, J. N. and Hutter, F. (2018). Hyperparameter importance across datasets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM.

Vejan, P., Abdullah, R., Khadiran, T., Ismail, S., and Boyce, A. N. (2016). Role of plant growth promoting rhizobacteria in agricultural sustainability—a review. *Molecules*, 21(5):573.

Wang, J., Wu, B., Kohnen, M. V., Lin, D., Yang, C., Wang, X., Qiang, A., Liu, W., Kang, J., Li, H., Shen, J., Yao, T., Su, J., Li, B., and Gu, L. (2021). Classification of rice yield using UAV-based hyperspectral imagery and lodging feature. *Plant Phenomics*, 2021:1–14.

Whittaker, J. C., Thompson, R., and Denham, M. C. (2000). Marker-assisted selection using ridge regression. *Genetics Research*, 75(2):249–252.

Windhausen, V. S., Atlin, G. N., Hickey, J. M., Crossa, J., Jannink, J.-L., Sorrells, M. E., Raman, B., Cairns, J. E., Tarekegne, A., Semagn, K., Beyene, Y., Grudloyma, P., Technow, F., Riedelsheimer, C., and Melchinger, A. E. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3 Genes|Genomes|Genetics*, 2(11):1427–1436.

Wintermans, P. C. A., Bakker, P. A. H. M., and Pieterse, C. M. J. (2016). Natural genetic variation in arabidopsis for responsiveness to plant growth-promoting rhizobacteria. *Plant Molecular Biology*, 90(6):623–634.

Wold, S., Sjöström, M., and Eriksson, L. (2001). Pls-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2):109–130.

Xiao, X., Fan, M., Wang, E., Chen, W., and Wei, G. (2017). Interactions of plant growth-promoting rhizobacteria and soil factors in two leguminous plants. *Applied Microbiology and Biotechnology*, 101(23-24):8485–8497.

Xu, Y., Xu, C., and Xu, S. (2017). Prediction and association mapping of agronomic traits in maize using multiple omic data. *Heredity*, 119(3):174–184.

Yang, C.-M. and Chen, R.-K. (2004). Modeling rice growth with hyperspectral reflectance data. *Crop Science*, 44(4):1283–1290.

Yang, L. and Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415:295–316.

Yassue, R. M., Carvalho, H. F., Gevartosky, R., Sabadin, F., Souza, P. H., Bonatelli, M. L., Azevedo, J. L., Quecine, M. C., and Fritsche-Neto, R. (2021a). On the genetic architecture in a public tropical maize panel of the symbiosis between corn and plant growth-promoting bacteria aiming to improve plant resilience. *Molecular Breeding*, 41(10).

Yassue, R. M., Galli, G., Junior, R. B., Cheng, H., Morota, G., and Fritsche-Neto, R. (2021b). A low-cost greenhouse-based high-throughput phenotyping platform for genetic studies: a case study in maize under inoculation with plant growth-promoting bacteria. *bioRxiv (Preprint)*.

Yoosefzadeh-Najafabadi, M., Earl, H. J., Tulpan, D., Sulik, J., and Eskandari, M. (2021). Application of machine learning algorithms in plant breeding: Predicting yield from hyperspectral reflectance in soybean. *Frontiers in Plant Science*, 11:624273.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24):3326–3328.

Appendix

Table 1: A list of models and covariates included in the analysis.

	Classification	Prediction		
Model	Hyperspectral image	Hyperspectral image	Genomics	Integration
LR[1]	✓			
OLS[2]		✓		
PLS-DA[3]	✓			
PLS-R[4]		✓	✓	
LASSO[5]	✓	✓	✓	
BayesRR[6]	✓	✓	✓	✓
BayesB	✓	✓	✓	✓
AutoML[7]	✓	✓	✓	

1 LR: logistic regresion
2 OLS: ordinary least squares
3 PLS-DA: partial least squares discriminant analysis
4 PLS-R: partial least squares regression
5 LASSO: least absolute shrinkage and selection operator
6 BayesRR: Bayesian ridge regression
7 Automated machine learning

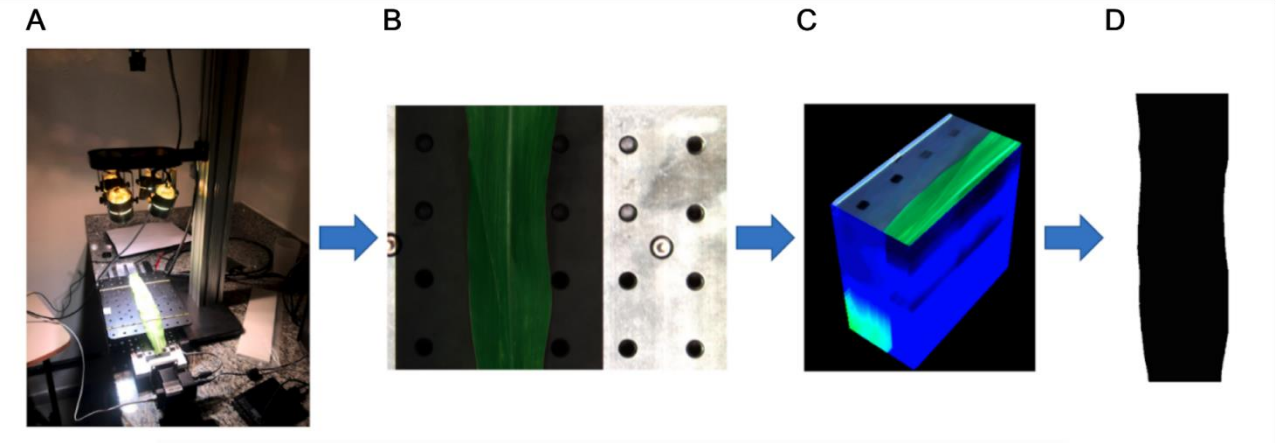

Figure 1: Summary of data acquisition and processing. A) the hyperspectral benchtop camera was used for data collection; B) region of interest of the last completed expanded leaf; C) cube image; and D) image mask.

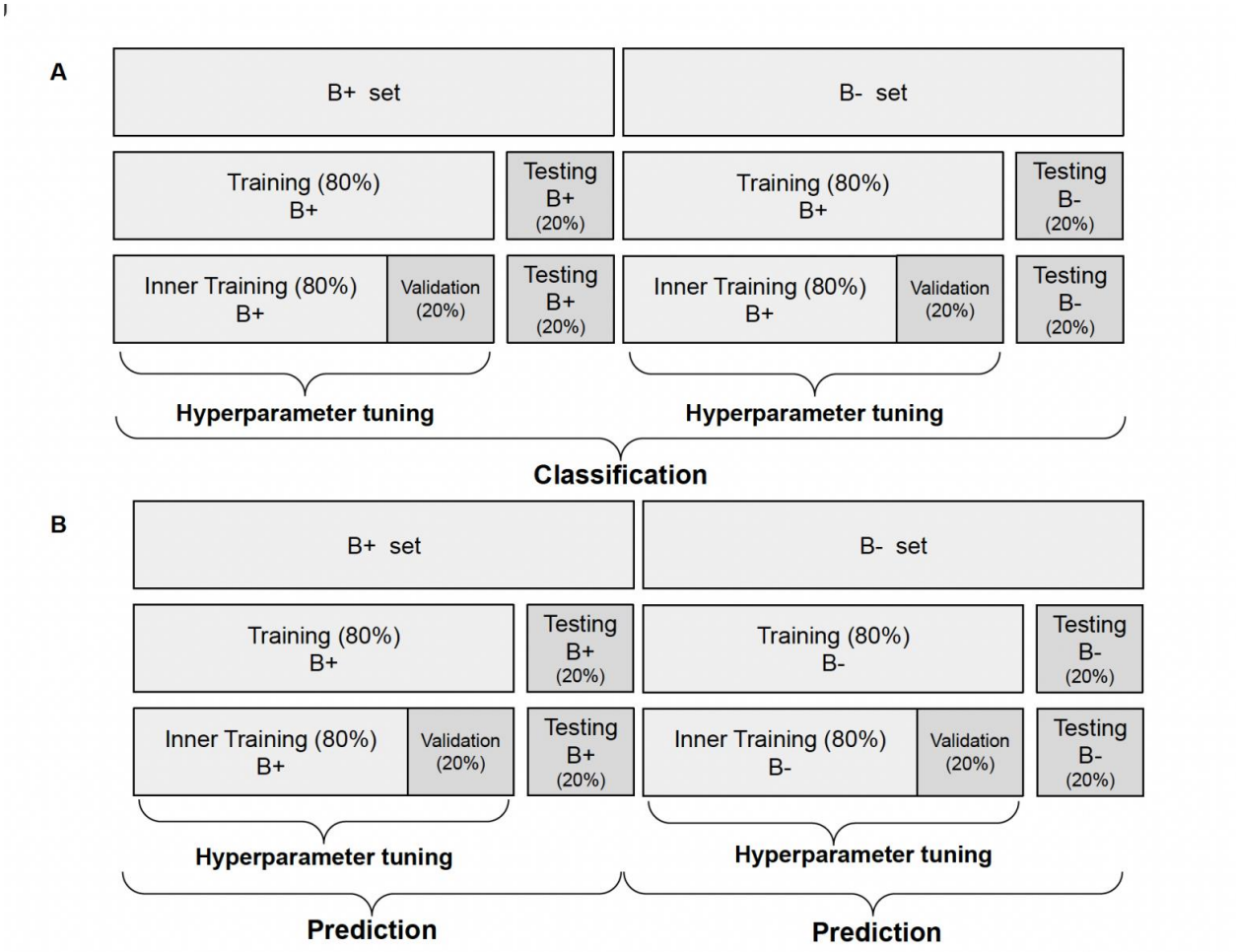

Figure 2: Graphic representation of a cross-validation design using the hyperspectral image and genomic data. The data were divided into training and testing sets. For PLS, LASSO, and Auto-sklearn, the training set was split into inner training and validation sets to perform hyperparameter tuning. This process was repeated 20 times using repeated random subsam- pling cross-validation. A) Classification was performed jointly using with (B+) and without (B-) plant growth-promoting bacteria inoculation conditions. B) Prediction was performed separately for each management condition.

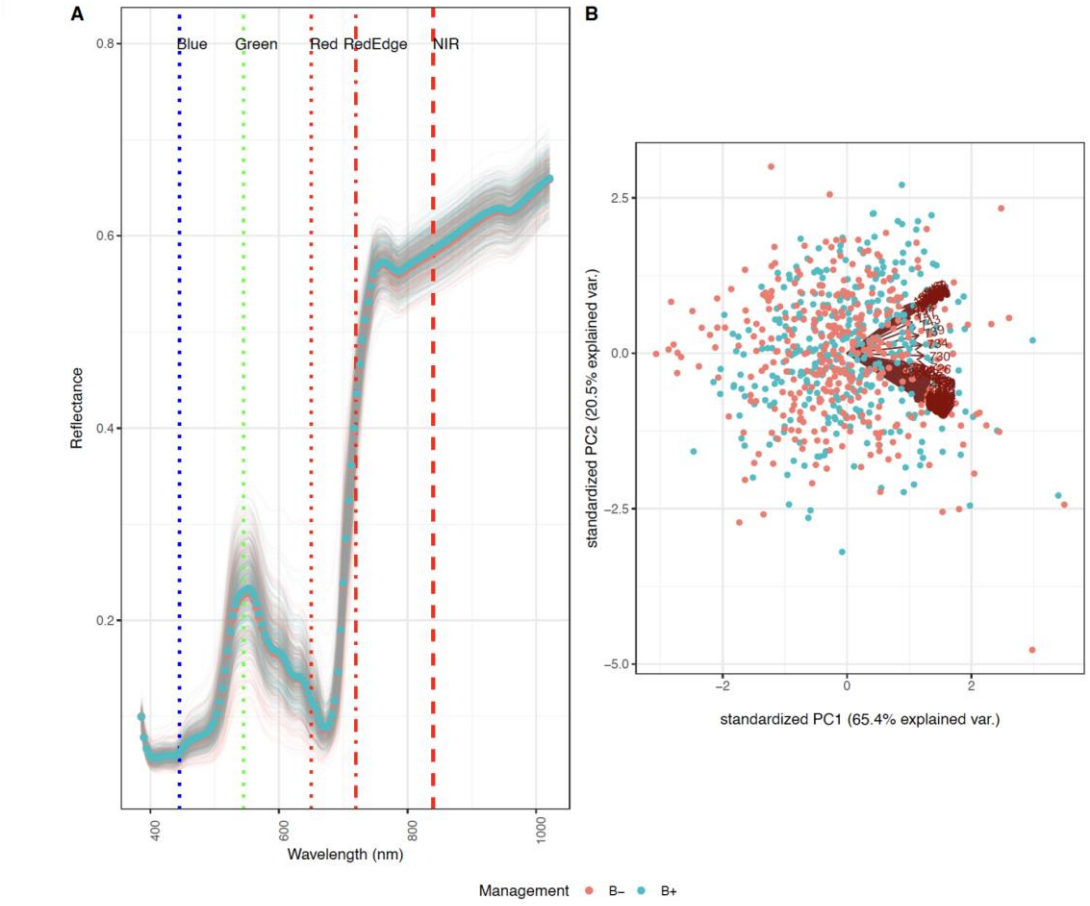

Figure 3: A: Spectral signatures of the maize genotypes under with (B+) and without (B-) plant growth-promoting bacteria mangement conditions. B: Principal component biplot of the 360 maize genotypes and 150 spectral wavelengths under with (B+) and without (B-) plant growth-promoting bacteria. Each point and arrow represent a genotype and a spectral wavelength, respectively.

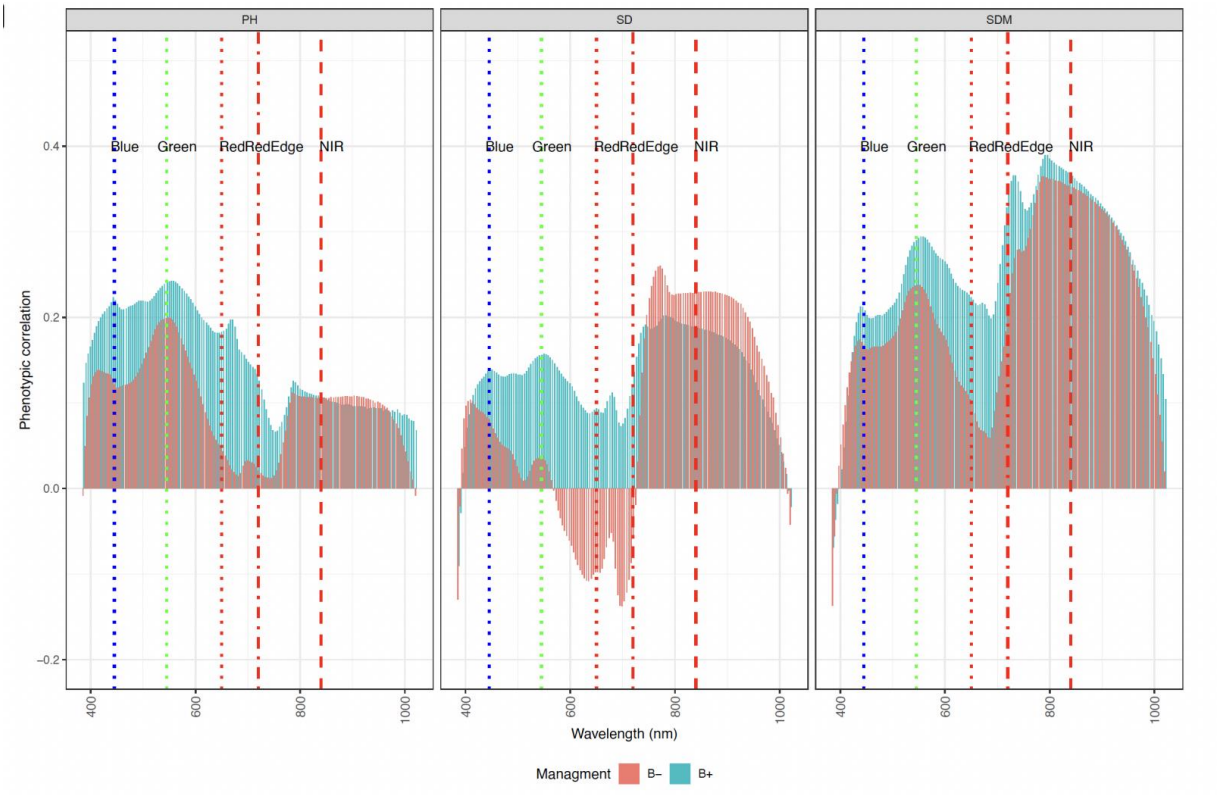

Figure 4: Correlations between manually measured growth-related phenotypes and hyper- spectral reflectance values under the two management conditions (B+ and B-). PH: plant height; SD: stalk diameter; and SDM: shoot dry mass.

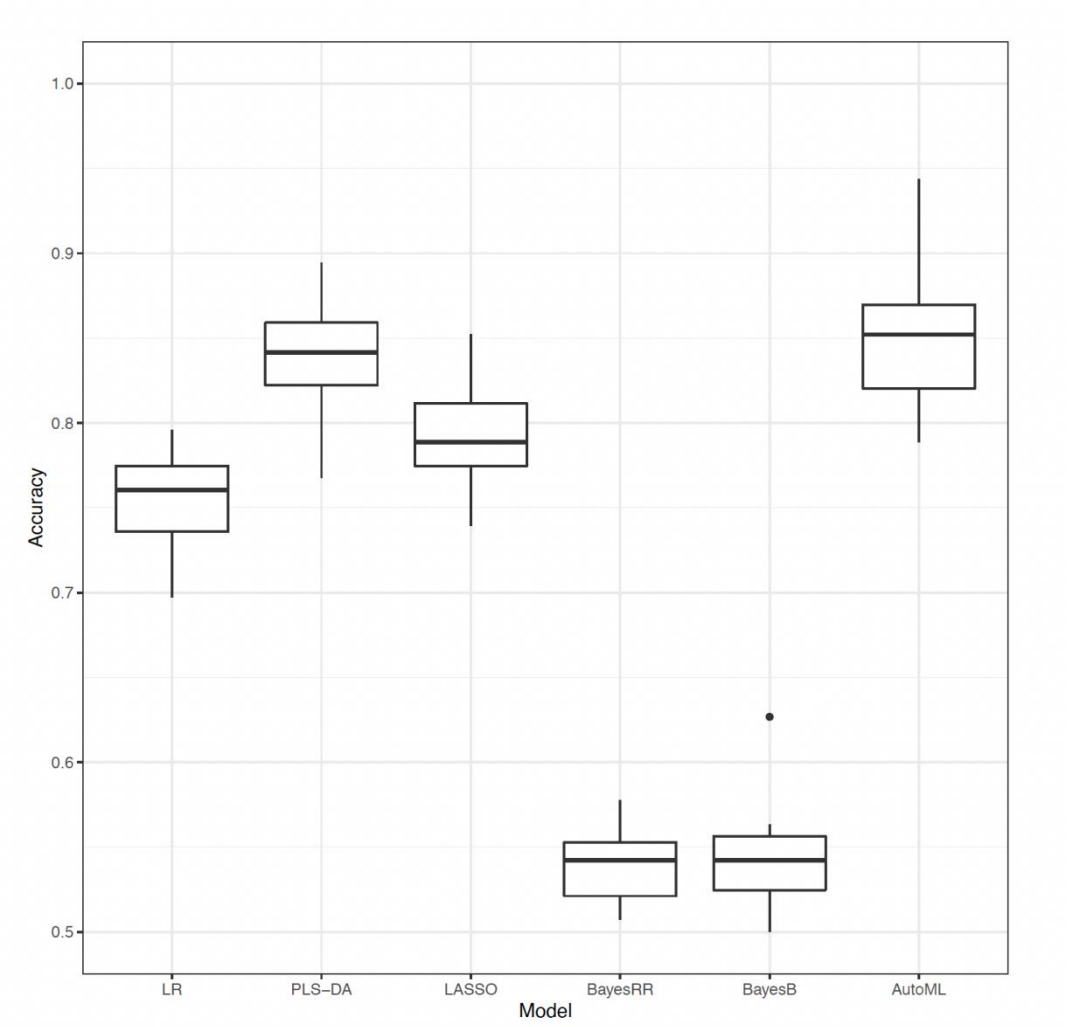

Figure 5: Classification accuracy of inoculation status (B+ and B-) using 150 hyperspectral bands. OLS: ordinary least squares; PLS-DA: partial least squares discriminant analysis; LASSO: least absolute shrinkage and selection operator; BayesRR: Bayesian ridge regression; and AutoML: automated machine learning.

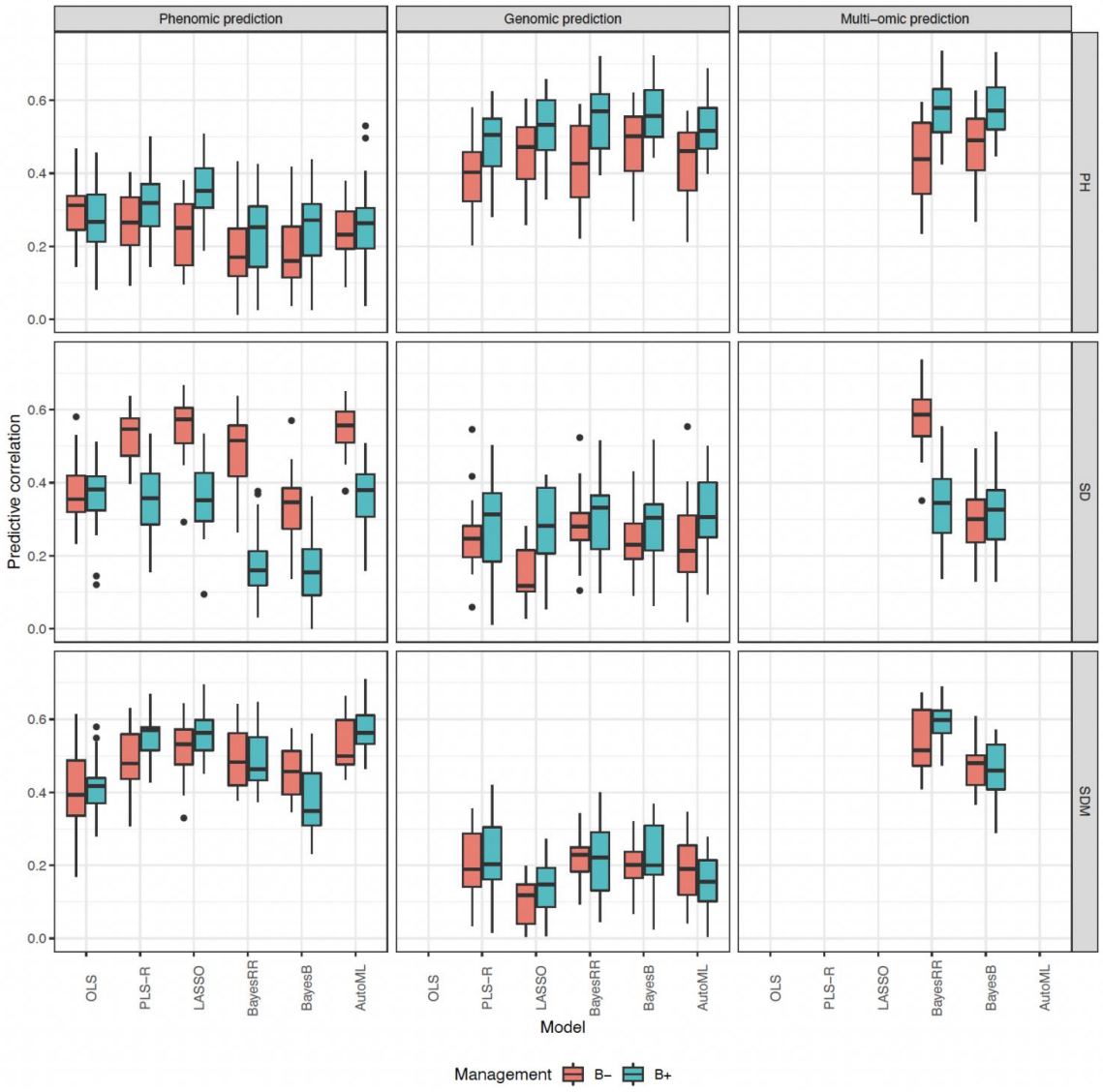

Figure 6: Predictive ability of plant height (PH), stalk diameter (SD), and shoot dry mass (SDM) using phenomic prediction, genomic prediction, and multi-omic prediction models in each management condition (B+ and B-). OLS: ordinary least squares; PLS-R: partial least squares regression; LASSO: least absolute shrinkage and selection operator; BayesRR: Bayesian ridge regression; and AutoML: automated machine learning.

5. GENOME-WIDE ASSOCIATION ANALYSIS OF HYPERSPECTRAL REFLECTANCE DATA TO DISSECT THE GROWTH-RELATED TRAIT GENETIC ARCHITECTURE IN MAIZE UNDER INOCULATION WITH PLANT GROWTH-PROMOTING BACTERIA

Rafael Massahiro Yassue[1], Giovanni Galli[1], Chunpeng James Chen[2,3], Roberto Fritsche-Neto[1,4,*], and Gota Morota[2,3,*]

[1]Department of Genetics, 'Luiz de Queiroz' College of Agriculture, University of São Paulo, São Paulo, Brazil

[2]School of Animal Sciences, Virginia Polytechnic Institute and State University, Blacksburg, USA

[3]Center for Advanced Innovation in Agriculture, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 USA

[4]Quantitative Genetics and Biometrics Cluster, International Rice Research Institute, Los Baños, Philippines

Running title: Integrating hyperspectral data into genome-wide association analysis

Keywords: genome-wide association analysis, growth trait, hyperspectral wavelength, multi-phenotype,

Core ideas:

• Hyperspectral reflectance data captures the maize genetic variability in response to plant growth-promoting bacteria.

• Eight SNPs were detected influencing both the manually measured and hyperspectral phenotypes

• Shiny application developed interactively visualizes multi-phenotype Manhattan plots

* Corresponding author

E-mail:r.fritscheneto@irri.org and morota@vt.edu

ORCID: 0000-0002-7424-2227 (RMY), 0000-0002-3400-7978 (GG), 0000-0002-2018-0702 (CJC), 000-0003-4310-0047 (RFN), and 0000-0002-3567-6911 (GM)

Email addresses: rafael.yassue@usp.br (RMY), giovannigalli@alumni.usp.br (GG), niche@vt.edu(CJC), r.fritscheneto@irri.org (RFN), and morota@vt.edu (GM)

Abbreviations: genome-wide association (GWA), plant growth-promoting bacteria (PGPB), plant height (PH), shoot dry mass (SDM). single-nucleotide polymorphisms (SNPs), stalk diameter (SD), with plant growth-promoting bacterial inoculation (B+), and without plant growth-promoting bacterial inoculation (B-),

Abstract

Plant growth-promoting bacteria (PGPB) has potential to increase yield and resilience against biotic and abiotic stresses in crops. Integrating hyperspectral reflectance data into growth-related traits may shed some light on their genetic basis because hyperspectral data can assess the biochemical or physiological attributions of plants. The objective of the current study was to integrate hyperspectral image data into genome-wide association analysis to study maize growth-related traits under PGPB inoculation. A total of 360 inbred maize lines with 13,826 single nucleotide polymorphisms (SNPs) were evaluated under two management conditions: with (B+) and without (B-) PGPB inoculation. Their 150 hyperspectral wavelengths reflectance ranging from 386 to 1021 nm and 131 hyperspectral indices were used in the analysis. Plant height, stalk diameter, and shoot dry mass were manually measured. Overall, the hyperspectral signatures had similar or higher genomic heritability estimates than those of manually measured phenotypes and they were genetically correlated with the manually measured phenotype. Several hyperspectral reflectance values and spectral indices were identified by genome-wide association analysis as potential markers for growth-related traits under PGPG inoculation. A total of eight SNPs were detected impacting both the manually measured and hyperspectral phenotypes. The hyperspectral phenotypes were associated with genes that have been previously reported as candidates for nitrogen uptake efficiency, tolerance to abiotic stress, and kernel size-related. Taken together, our results show the usefulness of hyperspectral-based phenotyping for studying maize growth-related traits in response to PGPB inoculation. In addition, a Shiny web application was developed to interactively explore multi-phenotype genome-wide association results.

5.1 Introduction

Increasing food production in a sustainable manner is a necessity due to rising food demand, especially in growing tropical developing nations (Laurance et al., 2014). Recent studies found that plant growth-promoting bacteria (PGPB) are a suitable alternative for increasing plant resilience against biotic and abiotic stress with potential to increase food production (Compant et al., 2005; Batista et al., 2018;

Yassue et al., 2021). Plant growth-promoting bacteria have the potential to promote morphological (Mantelin, 2003) and functional (Di Benedetto et al., 2017) changes in plants. Some of the reported effects are an increase in nutrients uptakes, such as nitrogen, phosphate, potassium, and iron (Egam-berdiyeva, 2007; Pii et al., 2015), and the activation of the response against diseases and abiotic stress (Olanrewaju et al., 2017; Singh et al., 2018).

Hyperspectral image data have been increasingly being applied for plant genetics and managements studies because of their associations with target phenotypes, such as water content (Ge et al., 2016), nutrient plant status (Mahajan et al., 2014; Nigon et al., 2021), disease symptoms (Thomas et al., 2017), yield (Yang et al., 2021), and plant biomass (Krause et al., 2019) because hyperspectral data can assess the biochemical or physiological attributions of plants. For example, hyperspectral phenotypes that are genetically correlated with the target phenotype can potentially aid genomic prediction (Krause et al., 2019; Sandhuet al., 2021). In addition, hyperspectral phenotypes can be used to perform genetic inference studies, such as genome-wide association (GWA), heritability, and genetic correlation analyses, to investigate the association between hyperspectral bands and genome (Feng et al.,2017; Sun et al., 2019; Barnaby et al., 2020; Wu et al., 2021).

Bayesian whole-genome regression models are advantageous for GWA studies because they implicitly account for population structure and the multiple-testing problem of classical single-marker linear mixed models by fitting all markers simultaneously (Fernando et al., 2017; Wolc and Dekkers, 2022). Despite the increasing use of high-throughput phenotyping in collecting hundreds to thousands of phenotypes, still there are limited number of whole-genome regression studies that integrate hyperspectral data into genetic inference research. How hyperspectral wavelengths data are associated with PGPB responses in maize remains elusive because it is particularly difficult to interpret the changes in a hyperspectral reflectance pattern into the plant biological process. Therefore, the objectives of this study were 1) investigate whether hyperspectral signatures are under geomic control, 2) determine if variation in hyperspectral wavelengths correlates with growth-related traits at the genomic level, and 3) identify specific genomic regions associated with hyperspectral wavelengths that can be used to identify corresponding candidate genes to study maize growth-related traits under PGPB inoculation in maize. We employed Bayesian whole-genome GWA

methods to identify possible candidate genes influencing growth-related traits and hyperspectral rebands. In addition, a Shiny web application was developed to interactively explore multi-phenotype GWA results.

5.2 Materials and Methods

Plant growth-promoting bacteria experiment

A tropical maize association panel containing 360 inbred lines was used to study the genetic basis of PGPB response. The inbred lines were evaluated with (B+) and without (B-) PGPB inoculation under nitrogen stress. The PGPB used in this study were *Bacillus thuringiensis* , *Delftia* sp. RZ4MS18 (Batista et al., 2018) *Pantoea agglomerans* (Quecine et al., 2012), and *Azospirillum brasilense* (Hungria et al., 2010), and they were co-inoculated in the maize seeds. The B- management included an inoculum with only liquid Luria-Bertani medium. Irrigation, weed control, and fertilizer except for nitrogen, were carried out according to the need of the crop. The plants were evaluated when most of the plants had six expanded leaves, approximately 33 days after sowing. The manually measured phenotypes were plant height (PH), stalk diameter (SD), and shoot dry mass (SDM). Further information about the experiment design can be found in Yassue et al (2021, 2022b).

Genomic data

The 360 inbred lines were genotyped using the genotyping-by-sequencing method followed by the two-enzymes (PstI and MseI) protocol (Sim et al., 2012; Poland et al., 2012). Deoxyribonucleic acid was extracted from the leaves using the cetyltrimethylammonium bromide method (Doyle and Doyle, 1987). Single nucleotide polymorphism (SNP) calling was performed using TASSEL 5.0 (Bradbury et al., 2007) based on B73 (B73-RefGen v4) as the reference genome. The SNP markers were removed if the call rate was less than 90%, non-biallelic, or the minor allele frequency was less than 5%. Missing marker codes were imputed using Beagle 5.0 (Browning et al., 2018). Markers with pairwise linkage disequilibrium higher than 0.99 were removed using the SNPRelate R package (Zheng et al., 2012). A total of 13,826 SNPs were obtained after the quality control process. Detailed information regarding the population genomics is available in Yassue et al. (2021).

Hyperspectral imaging and processing

Hyperspectral images of each leaf for plant grown under management conditions B+ and B- were collected using a benchtop system of Pika L. camera (Resonon, Bozeman, MT, USA). The middle portion of the last completed expanded leaf was used as a region of interest for the hyperspectral imaging. A dark room with additional light supply was used to minimize light variation. Radiometric calibration was performed according to the camera system manufacturer. For each plant, a hyperspectral cube image contained 150 wavelength bands in the range of 386 to 1021 nm. Image processing based on the Spectral Python module was performed applying a mask to remove the background from the image and the mean of reflectance of each pixel was used for further analysis. A summary of hyperspectral imaging and processing can be found in Yassue et al. (2022a). In addition, a total 131 hyperspectral indices were calculated based on the mean reflectance value for each wavelength using the R package hsdar (Lehnert et al., 2019). These hyperspectral indices have been reported to be associated with a variety of phenotypes, such as nutrient and chlorophyll content, pigments, photosynthesis, and water content (Zarco-Tejada et al., 2004, 2005; Ranjan et al., 2012; Gitelson et al., 2014). A summary of hyperspectral indices and their correlations are available in the supplementary material Table S1-S4 and Figure S1, respectively.

Univariate BayesC

BayesC (Kizilkaya et al., 2010; Habier et al., 2011), which is one of the Bayesian whole- genome regression models, was used to estimate the markers effect and variance components for each combination of phenotype and management (B+ and B-).

$$y_{kli} = \mu + r_k + b_l + \sum_{j=1}^{m} w_{ij}\alpha_j + \epsilon_{kli},$$

where y_{kli} is the vector of phenotypes (manually measured or hyperspectral wavelength) in the kth repetition, lth block within repetition, and ith genotype; μ is the overall mean; r and b are the fixed effects for replication and block within replication, respectively; w_{ij} is the incidence matrices of marker covariates for each SNP coded as 0, 1, or 2; and α_j is the$j$*th* marker effect. The prior of α_j was:

$$\alpha_j|\pi,\sigma_\alpha^2 = \begin{cases} 0 & \text{with probability of } \pi \\ \sim| N(0,\sigma_\alpha^2) & \text{with probability } (1-\pi) \end{cases}$$

where σ^2 is the common marker genetic variance, π is a mixture proportion set to 0.99, and ϵ is the vector of residual term. A Gaussian prior $N(0, \sigma^2)$ was assigned to the vector of residuals and a flat prior was assigned to μ, r, and b. The scaled inverse χ^2 distribution was assigned to σ^2_α and σ^2_ϵ by setting the degrees of freedom to 4 and choosing the scale parameter such that the prior mean of the variance equals half of the phenotypic variance. The variance components obtained from BayesC were used to estimate heritability, where $h^2 = \sigma^2_g/(\sigma^2_g+\sigma^2_{e/n})$, where h_g is the genomic heritability, σ^2_g and σ^2_e are the additive genomic and residual variances, respectively, and n is the number of replication (2).

Bivariate BayesC

Bivariate BayesC was used to estimate the genetic correlation between manually measured and hyperspectral traits. The model description follows that of univariate BayesC with some modification. Here, **y** is the vector of manually measured and hyperspectral phenotypes and

the marker effect of trait k for locus j followed

$$\alpha_{jk}|\pi_k,\sigma^2_{\alpha_k} = \begin{cases} 0 & \text{with probability of } \pi_k \\ \sim N(0,\sigma^2_{\alpha_k}) & \text{with probability } (1-\pi_k). \end{cases}$$

The jth marker effect can be reparameterized as $\boldsymbol{\alpha_j} = \boldsymbol{D_j\beta_j}$, where $\boldsymbol{D_j}$ is a diagonal matrix with elements $diag(\mathbf{D}_j)=\boldsymbol{\delta}_j=(\delta_{j1}, \delta_{j2})$ indicating whether the jth marker effect for trait k is zero or non-zero, and $\boldsymbol{\beta_j}$ follows a multivariate normal distribution with null mean and covariance matrix $\Sigma\alpha = \begin{bmatrix} \sigma^2_{\alpha_1} & \sigma_{\alpha_{12}} \\ \sigma_{\alpha_{21}} & \sigma^2_{\alpha_2} \end{bmatrix}$, where α_1, α_2, and α_{12} (α_{21}) are marker genetic variance for trait1, marker genetic variance for trait2, and marker covariance between traits 1 and 2, respectively, and the residuals were

assumed independently and identically distributed multivariate normal vectors with null mean and covariance matrix $\boldsymbol{\Sigma}_{\epsilon}$ (Cheng et al., 2018b). The covariance matrices, $\boldsymbol{\Sigma}_{\alpha}$ and $\boldsymbol{\Sigma}_{\epsilon}$, were assigned an inverse Wishart prior distribution with $\mathbf{W}^{-1}(S_{\alpha}, \nu_{\alpha})$ and $\mathbf{W}^{-1}(S_{\epsilon}, \nu_{\epsilon})$, respectively. We assumed all possible combinations for $\boldsymbol{\delta}_j$, namely, (0,0), (0,1), (1,0), and (1,1) having nonzero probability.

Bayesian GWA analysis

Aforementioned BayesC was used to perform GWA analysis of hyperspectral reflectance values and manually measured phenotypes. Candidate markers were chosen according to their posterior inclusion probability. A posterior inclusion probabilityis the posterior probability of a given marker being included in the model (Fernando and Garrick, 2013). Following an earlier work (Fan et al., 2011), the posterior inclusion probability threshold of 0.10 was used for manually measured phenotypes, while 0.50 was used for hyperspectral GWAS to be more conservative. All the Bayesian analyses were fit using 60,000 Markov chain Monte Carlo samples, 6,000 burn-in, and a thinning rate of 60 implemented in JWAS (Cheng et al., 2018a). Model convergence was assessed using trace plots of the posterior means of the parameters. For each selected SNP associated with hyperspectral phenotypes, genes within an interval of 50 kilobase pair (kbp) upstream and downstream of the SNP were explored using the MaizeMine V1.3 server (Shamimuzzaman et al., 2020). One challenge was a large number of Manhattan plots that potentially need to be generated because there were three manually measured phenotypes, 150 hyperspectral wavelengths, and 131 hyperspectral indices. Instead of including all the Manhattan plots in the supplementary file, we developed a Shiny application using the R package shiny (Chang et al., 2021) that provides functions for constructing interactive web applications. This application allows the user to interactively explore all possible genome to phenome association combinations.

5.3 Results

Estimates of genomic heritability and correlation

The genomic heritability estimates of manually measured PH, SD, and SDM in B+ (B-) condition were 0.61 (0.57), 0.60 (0.39), 0.30 (0.28), respectively. Plant height had the highest genomic heritability estimates, while SDM had the lowest estimates. The mean (standard deviation) of genomic heritability estimates for hyperspectral wavelength and

indices were 0.45 (0.053) and 0.41 (0.081), respectively (Figure 1). The individual hyperspectral wavelengths that explained the largest genomic heritability estimates were 645 and 649 nm (B+) and 512 and 507 nm (B-). The estimates of genomic heritability showed a somewhat decreasing tendency as wavelength increases. For hyperspectral indices, NDNI and NPQI for B- and EVI for B+ showed the highest genomic heritability estimates. Although the mean of genomic heritability estimates of hyperspectral indices was lower than that of hyperspectral wavelengths, some indices explained more genetic variance than individual wavelengths. Overall, similar genomic heritability estimates were observed for B+ and B- management for most of wavelengths or indices.

Genomic correlation estimates between hyperspectral wavelengths and manually measured traits were largely positive (Figure 2). In PH, the genomic correlation estimates ranged from -0.34 to 0.34. Positive correlations were observed for both B- and B+ in the wavelength varying from 400–700 nm, while negative correlations were observed only for B+ after 700 nm. The hyperspectral wavelengths that provided the largest genomic correlation estimates in absolute values were 578 nm (0.286) and 398 nm (0.193) for B- and B+, respectively. The extent of genomic correlation estimates was low in SD. The sign of estimates was mostly positive except for the beginning and end of wavelengths. The hyperspectral wavelength showed positive correlations throughout the entire spectra range except at the start of wavelength in SDM. In particular, higher correlations were observed between 700–1000 nm for B-. The hyperspectral wavelengths that provided the largest genomic correlation estimates with SDM were 817 nm (0.392) and 734 nm (0.235) for B- and B+, respectively.

On the other hand, genomic correlation estimates greatly varied across the hyperspectral indices. The extent of correlation estimates was lower in SD than those of PH and SDM. While the hyperspectral indices that provided the largest genomic correlation estimates with PH were D2 (0.339) and RDVI (-0.292) for B- and B+, respectively, those with SDM were RARSb (-0.357) and NPQI (-0.340) for B- and B+, respectively.

GWA analyses for growth-related and hyperspectral traits

A total of 86 SNPs were selected from BayesC analysis using the posterior inclusion probability threshold of 0.10 for PH, SD, and SDM (Figure 3 and Supplementary Tables S5-S9). Plant height showed the highest number of selected markers (21 and

24 for B+ and B-, respectively), while SDM had the lowest (5 for B+). No markers were detected for SDM in B-. No overlapping SNPs were identified across manually measured phenotypes, while only four SNPs were selected for both PGPB inoculation conditions (B- and B+) indicating that the PGPB may alter the plant growth patterns and the genomic regions controlling them. The conservative posterior inclusion probability threshold of 0.50 was used to find SNPs controlling the hyperspectral-derived phenotypes to identify candidate genes. Out of 25 detected SNPs, five SNPs were associated with at least five different hyperspectral phenotypes. The gene annotation of each selected SNP within an interval of 100 kbp showed presence of genes that have been previously reported in the literature related to growth-related phenotypes or response to abiotic stress (Table 1). The hyperspectral indices, Chlg, CRI2, CRI3, CRI4, Datt6, GMI1, PARS, SD, and SR3, were associated with the genes , nrt2.2, and Zm00001d054060 on chromosome 4. The index CRI1 was associated with the gene Zm00001d012924 on chromosome 5, and the index EVI was associated with the Zm00001d029820 and Zm00001d007843 on chromosomes 1 and 2, respectively, and 100275163 on chromosome 6. In addition, the hyperspectral wavelengths varying from 398 nm to 434 nm were associated with the gene Zm00001d012719 on chromosome 8.

The gene annotation of the overlapped SNPs between manually measured and hyperspectral phenotypes reveals that some may play importante role in nitrogen update, plant response to biotic and abiotic stress (Table 2). The genes Zm00001d052164 and Zm00001d052165 have been previous reported as putative gene involved in regulating nitrogen assimilation (Wang et al., 2020). Zm00001d006916 has been reported as a candidate gene responsible to autophagy process that may play important role in response to abiotic stress (Tang and Bassham, 2018). The LOC100216557 gene have been associated to more resistant maize plants to aphid and may responsible to the plant defense responses and stress tolerance (Srivastava et al., 2018) while 103647869 is a candidate gene to resistance to *A. flavus* infection and/or aflatoxin contamination (Liu et al., 2021). The gene 100274563 has been reported associated with ear weight per plant (Zhou et al., 2020).

Integration of GWA analyses

A total of eight SNPs were influencing both the manually measured and hyperspectral phenotypes and they were visualized in a phenome-wide association plot (Figure 4). In general, the most of SNPs were identified in the management B+. The two SNPs, CM000780.4 181569268 and CM000782.4 108027757, exhibited a strong association for a wide range of hyperspectral phenotypes along with PH and SDM in management B+. A Shiny web application was developed to interactively explore multi-phenotype genome-wide association results https://github.com/vt-ads/shiny_manhattan_plot.

5.4 Discussion

The utility of hyperspectral images as phenotypes have been gaining traction lately because they can capture the resonance of certain physicochemical compounds in plants. Growth-related traits, such as PH, SD, and SDM in maize, are known as complex traits that are controlled by many genes of small effect. Therefore, we expected that hyperspectral data can shed some light on genetic analysis and has the potential to assess the genetic variability of maize to aid the identification of candidate genes in maize. However, translating reflectance values or hyperspectral indices into a biological context, such as metabolic path, morphological, or functional change in the plant can be difficult and time-consuming. In this work, GWA analaysis of manually measured phenotypes, single-band reflectance, and hyperspectral indices were used to investigate the genetic basis of PGPB response. BayesC, which performs variable selection, was applied for GWA analysis. The posterior inclusion probability of each marker was used to identify relevant SNPs. The preference of using posterior inclusion probability instead of the window posterior probability of association (Fernando et al., 2017) was due to the low marker density and unequal distribution of SNPs across the genome in the maize population.

Estimates of genomic heritability and correlation

The hyperspectral phenotypes showed a similar range of genomic heritability estimates relative to the manually measured phenotypes. For most of hyperspectral-derived phenotypes, the heritability estimates varied from 0.30 to 0.50, indicating that the hyperspectral data are capable of capturing genetic variation. In addition, the genetic correlation between the manually measured and hyperspectral phenotypes showed that these responses were probably influenced by the same set of genes.

The relatively higher genomic correlation estimates for PH (B-) in the spectra range of 400 to 700 nm (visible spectrum) may indicate the association between plant height and leaf pigments, such as caretonoids, chlorophyll a and b, and nitrogen concentrations (Zhao et al., 2003; Ayala-Silva and Beyl, 2005). Similarly, the higher genomic correlation estimates were observed for SDM between 700–1000 nm (near infrared). The association between near infrared spectra and plant biomass has been previously reported in maize (Ma et al., 2020). The wavelengths in this range were also reported for nitrogen content in oilseed rape (Müller et al., 2008) and wheat (Hansen and Schjoerring, 2003).

Genome-wide association analysis

Overall, more SNP associations were observed with the management B+ although the genomic heritability estimaates were similar across the two managements. The GWA analysis found that SNP CM000780.4 245633076 is associated with nine hyperspectral phenotypes suggesting three candidate genes, nrt2, nrt2.2, and Zm00001d054060. These genes have been previously reported as part of NO_3^- transporter gene families and candidates for nitrate uptake along the maize primary root (Liu et al., 2009; Sorgonà et al., 2011; Wang et al., 2020).

The genes Zm00001d029820 and Zm00001d012924 have been reported as a putative for plant development and environmental stress conditions (Zhang et al., 2020; Zhu et al., 2021). The candidate gene Zm00001d007843 was previously reported for kernel size-related genes (Zhou et al., 2021) and the gene Zm00001d012719 was shown as a candidate transcription factor to mediate plant response to abiotic stress (Vendramin et al., 2020). Although we did not directly evaluate the phenotypes related to the candidates genes reported in the literature, the hyperspectral signatures may be capable of assessing these phenotypes indirectly.

The Shiny web application with interactive interface is a powerful tool for visualization and interpretation of GWA analysis for multiple phenotypes. The two-way Manhattan plots can be used to investigate associations across traits and managements. In addition, the phenome-wide association plots can be used to identify and visualize markers having joint influence across hundreds to thousands phenotypes. The Shiny application can be easily extended to other high-throughput phenotyping data, such as longitudinal, fluorescence, and thermal data.

5.5 Conclusions

Hyperspectral images were collected to study hyperspectral signatures of growth-related traits in tropical maize. The hyperspectral signatures were able to capture the genetic variance in the population and were associated with growth-related traits under PGPB inoculation. Genone-wide association analysis of hyperspectral data identified genomic regions influencing both manually measured phenotypes and hyperspectral bands. In addition, a Shiny application for multiple phenotypes GWAS was developed.

Acknowledgments

The authors acknowledge Pedro Takao Yamamoto and Fernando Henrique Iost Filho for their support in collecting the hyperspectral images.

Funding

This study was supported in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Grant #17/24327-0, #19/04697-2, and #2017/19407-4 from São Paulo Research Foundation (FAPESP), International Business Machines Corporation (IBM, Brasil), and Virginia Polytechnic Institute and State University.

References

Ayala-Silva, T. and Beyl, C. A. (2005). Changes in spectral reflectance of wheat leaves in response to specific macronutrient deficiency. *Advances in Space Research*, 35(2):305–317.

Barnaby, J. Y., Huggins, T. D., Lee, H., McClung, A. M., Pinson, S. R. M., Oh, M., Bauchan, G. R., Tarpley, L., Lee, K., Kim, M. S., and Edwards, J. D. (2020). Vis/NIR hyperspectral imaging distinguishes sub-population, production environment, and physicochemical grain properties in rice. *Scientific Reports*, 10(1).

Batista, B. D., Dourado, M. N., Figueredo, E. F., Hortencio, R. O., Marques, J. P. R., Piotto, F. A., Bonatelli, M. L., Settles, M. L., Azevedo, J. L., and Quecine, M. C. (2021). The auxin-producing bacillus thuringiensis RZ2ms9 promotes the growth and modifies the root architecture of tomato (solanum lycopersicum cv. micro-tom). *Archives of Microbiology*.

Batista, B. D., Lacava, P. T., Ferrari, A., Teixeira-Silva, N. S., Bonatelli, M. L., Tsui, S., Mondin, M., Kitajima, E. W., Pereira, J. O., Azevedo, J. L., and Quecine, M. C. (2018). Screening of tropically derived, multi-trait plant growth- promoting rhizobacteria and evaluation of corn and soybean colonization ability. *Microbiological Research*, 206:33–42.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19):2633–2635.

Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, 103(3):338–348.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. (2021). *shiny: Web Application Framework for R*. R package version 1.7.1.

Cheng, H., Fernando, R., and Garrick, D. (2018a). JWAS: Julia implementation of whole-genome analysis software. In *Proceedings of the world congress on genetics applied livestock production*.

Cheng, H., Kizilkaya, K., Zeng, J., Garrick, D., and Fernando, R. (2018b). Genomic prediction from multiple-trait Bayesian regression methods using mixture priors. *Genetics*, 209(1):89–103.

Compant, S., Duffy, B., Nowak, J., Clément, C., and Barka, E. A. (2005). Use of plant growth-promoting bacteria for biocontrol of plant diseases: principles, mechanisms of action, and future prospects. *Applied and environmental microbiology*, 71(9):4951–4959.

Di Benedetto, N. A., Corbo, M. R., Campaniello, D., Cataldi, M. P., Bevilacqua, A., Sinigaglia, M., and Flagella, Z. (2017). The role of plant growth promoting bacteria in im- proving nitrogen use efficiency for sustainable crop production: a focus on wheat. *AIMS microbiology*, 3(3):413.

Doyle, J. and Doyle, J. (1987). A rapid dna isolation procedure for small quantities of fresh leaf tissue. *PHYTOCHEMICAL BULLETIN*, 17(RESEARCH).

Egamberdiyeva, D. (2007). The effect of plant growth promoting bacteria on growth and nutrient uptake of maize in two different soils. *Applied Soil Ecology*, 36(2-3):184–189.

Fan, B., Onteru, S. K., Du, Z.-Q., Garrick, D. J., Stalder, K. J., and Rothschild, M. F. (2011). Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. *PloS one*, 6(2):e14726.

Feng, H., Guo, Z., Yang, W., Huang, C., Chen, G., Fang, W., Xiong, X., Zhang, H., Wang, G., Xiong, L., and Liu, Q. (2017). An integrated hyperspectral imaging and genome-wide association analysis platform provides spectral and genetic insights into the natural variation in rice. *Scientific Reports*, 7(1).

Fernando, R., Toosi, A., Wolc, A., Garrick, D., and Dekkers, J. (2017). Application of whole genome prediction methods for genome-wide association studies: a bayesian approach. *Journal of Agricultural, Biological and Environmental Statistics*, 22(2):172–193.

Fernando, R. L. and Garrick, D. (2013). Bayesian methods applied to gwas. In *Genome-wide association studies and genomic prediction*, pages 237–274. Springer.

Ge, Y., Bai, G., Stoerger, V., and Schnable, J. C. (2016). Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *Computers and Electronics in Agriculture*, 127:625–632.

Gitelson, A. A., Peng, Y., and Huemmrich, K. F. (2014). Relationship between fraction of radiation absorbed by photosynthesizing maize and soybean canopies and NDVI from remotely sensed data taken at close range and from MODIS 250m resolution data. *Remote Sensing of Environment*, 147:108–120.

Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC bioinformatics*, 12(1):1–12.

Hansen, P. and Schjoerring, J. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86(4):542–553.

Hungria, M., Campo, R. J., Souza, E. M., and Pedrosa, F. O. (2010). Inoculation with selected strains of azospirillum brasilense and a. lipoferum improves yields of maize and wheat in brazil. *Plant and Soil*, 331(1-2):413–425.

Kizilkaya, K., Fernando, R., and Garrick, D. (2010). Genomic prediction of simulated multi-breed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *Journal of animal science*, 88(2):544–551.

Krause, M. R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R. P., Dreisigacker, S., Poland, J., Rutkoski, J., Sorrells, M., Gore, M. A., and

Mondal, S. (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3 Genes|Genomes|Genetics*, 9(4):1231–1247.

Laurance, W. F., Sayer, J., and Cassman, K. G. (2014). Agricultural expansion and its impacts on tropical nature. *Trends in Ecology & Evolution*, 29(2):107–116.

Lehnert, L. W., Meyer, H., Obermeier, W. A., Silva, B., Regeling, B., Thies, B., and Bendix, J. (2019). Hyperspectral data analysis in R: The hsdar package. *Journal of Statistical Software*, 89(12):1–23.

Liu, H., Wu, H., Wang, Y., Wang, H., Chen, S., and Yin, Z. (2021). Comparative transcriptome profiling and co-expression network analysis uncover the key genes associated withearly-stage resistance to aspergillus flavus in maize. *BMC Plant Biology*, 21(1).

Liu, J., Chen, F., Olokhnuud, C., Glass, A. D. M., Tong, Y., Zhang, F., and Mi, G. (2009). Root size and nitrogen-uptake activity in two maize (zea mays) inbred lines differing in nitrogen-use efficiency. *Journal of Plant Nutrition and Soil Science*, 172(2):230–236.

Ma, D., Maki, H., Neeno, S., Zhang, L., Wang, L., and Jin, J. (2020). Application of non-linear partial least squares analysis on prediction of biomass of maize plants using hyperspectral images. *Biosystems Engineering*, 200:40–54.

Mahajan, G. R., Sahoo, R. N., Pandey, R. N., Gupta, V. K., and Kumar, D. (2014). Using hyperspectral remote sensing techniques to monitor nitrogen, phosphorus, sulphur and potassium in wheat (triticum aestivum l.). *Precision Agriculture*, 15(5):499–522.

Mantelin, S. (2003). Plant growth-promoting bacteria and nitrate availability: impacts on root development and nitrate uptake. *Journal of Experimental Botany*, 55(394):27–34.

Müller, K., Böttcher, U., Meyer-Schatz, F., and Kage, H. (2008). Analysis of vegetation indices derived from hyperspectral reflection measurements for estimating crop canopy parameters of oilseed rape (brassica napus l.). *Biosystems Engineering*, 101(2):172–182.

Nigon, T., Paiao, G. D., Mulla, D. J., Fernández, F. G., and Yang, C. (2021). The influence of aerial hyperspectral image processing workflow on nitrogen uptake prediction accuracy in maize. *Remote Sensing*, 14(1):132.

Olanrewaju, O. S., Glick, B. R., and Babalola, O. O. (2017). Mechanisms of action of plant growth promoting bacteria. *World Journal of Microbiology and Biotechnology*, 33(11).

Pii, Y., Mimmo, T., Tomasi, N., Terzano, R., Cesco, S., and Crecchio, C. (2015). Microbial interactions in the rhizosphere: beneficial influences of plant growth-promoting rhizobacteria on nutrient acquisition process. a review. *Biology and Fertility of Soils*, 51(4):403–415.

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J.-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, 7(2):e32253.

Quecine, M. C., Araújo, W. L., Rossetto, P. B., Ferreira, A., Tsui, S., Lacava, P. T., Mondin, M., Azevedo, J. L., and Pizzirani-Kleiner, A. A. (2012). Sugarcane growth promotion by the endophytic bacterium pantoea agglomerans 33.1. *Applied and Environmental Micro-biology*, 78(21):7511–7518.

Ranjan, R., Chopra, U. K., Sahoo, R. N., Singh, A. K., and Pradhan, S. (2012). Assessment of plant nitrogen stress in wheat (triticum aestivum l.) through hyperspectral indices. *International Journal of Remote Sensing*, 33(20):6342–6360.

Sandhu, K., Patil, S. S., Pumphrey, M., and Carter, A. (2021). Multitrait machine- and deep-learning models for genomic selection using spectral information in a wheat breeding program. *The Plant Genome*, 14(3).

Shamimuzzaman, M., Gardiner, J. M., Walsh, A. T., Triant, D. A., Tourneau, J. J. L., Tayal, A., Unni, D. R., Nguyen, H. N., Portwood, J. L., Cannon, E. K. S., Andorf, C. M., and Elsik, C. G. (2020). MaizeMine: A data mining warehouse for the maize genetics and genomics database. *Frontiers in Plant Science*, 11.

Sim, S.-C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M. W., Deynze, A. V., Hamilton, J. P., Buell, C. R., Causse, M., Wijeratne, S., and Francis, D. M. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS ONE*, 7(7):e40563.

Singh, V. K., Singh, A. K., Singh, P. P., and Kumar, A. (2018). Interaction of plant growth promoting bacteria with tomato under abiotic stress: A review. *Agriculture, Ecosystems & Environment*, 267:129–140.

Sorgonà, A., Lupini, A., Mercati, F., di Dio, L., Sunseri, F., and Abenavoli, M. R. (2011). Nitrate uptake along the maize primary root: an integrated physiological and molecular approach. *Plant, Cell & Environment*, 34(7):1127–1140.

Srivastava, R., Li, Z., Russo, G., Tang, J., Bi, R., Muppirala, U., Chudalayandi, S., Severin, A., He, M., Vaitkevicius, S. I., Lawrence-Dill, C. J., Liu, P., Stapleton, A. E., Bassham, D. C., Brandizzi, F., and Howell, S. H. (2018). Response to persistent ER stress in plants: A multiphasic process that transitions cells from prosurvival activities to cell death. *The Plant Cell*, 30(6):1220–1242.

Sun, D., Cen, H., Weng, H., Wan, L., Abdalla, A., El-Manawy, A. I., Zhu, Y., Zhao, N., Fu,H., Tang, J., Li, X., Zheng, H., Shu, Q., Liu, F., and He, Y. (2019). Using hyperspectral analysis as a potential high throughput phenotyping tool in GWAS for protein content of rice quality. *Plant Methods*, 15(1).

Tang, J. and Bassham, D. C. (2018). Autophagy in crop plants: what's new beyond arabidopsis? *Open Biology*, 8(12):180162.

Thomas, S., Kuska, M. T., Bohnenkamp, D., Brugger, A., Alisaac, E., Wahabzada, M., Behmann, J., and Mahlein, A.-K. (2017). Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective. *Journal of Plant Diseases and Protection*, 125(1):5–20.

Vendramin, S., Huang, J., Crisp, P. A., Madzima, T. F., and McGinnis, K. M. (2020). Epigenetic regulation of ABA-induced transcriptional responses in maize. *G3 Genes|Genomes|Genetics*, 10(5):1727–1743.

Wang, Y., Xu, J., Ge, M., Ning, L., Hu, M., and Zhao, H. (2020). High-resolution profile of transcriptomes reveals a role of alternative splicing for modulating response to nitrogen in maize. *BMC Genomics*, 21(1).

Wolc, A. and Dekkers, J. (2022). Application of Bayesian genomic prediction methods to genome-wide association analyses. *Genetics Selection Evolution*, 54(1):1–12.

Wu, X., Feng, H., Wu, D., Yan, S., Zhang, P., Wang, W., Zhang, J., Ye, J., Dai, G., Fan, Y., Li, W., Song, B., Geng, Z., Yang, W., Chen, G., Qin, F., Terzaghi, W., Stitzer,M., Li, L., Xiong, L., Yan, J., Buckler, E., Yang, W., and Dai, M. (2021). Using high- throughput multiple optical phenotyping to decipher the genetic architecture of maize drought tolerance. *Genome Biology*, 22(1).

Yang, W., Nigon, T., Hao, Z., Paiao, G. D., Fernández, F. G., Mulla, D., and Yang, C. (2021). Estimation of corn yield based on hyperspectral imagery and convolutional neural network. *Computers and Electronics in Agriculture*, 184:106092.

Yassue, R. M., Carvalho, H. F., Gevartosky, R., Sabadin, F., Souza, P. H., Bonatelli, M. L., Azevedo, J. L., Quecine, M. C., and Fritsche-Neto, R. (2021). On the genetic architecture in a public tropical maize panel of the symbiosis between corn and plant growth-promoting bacteria aiming to improve plant resilience. *Molecular Breeding*, 41(10).

Yassue, R. M., Galli, G., Fritsche-Neto, R., and Morota, G. (2022a). Classification of plant growth-promoting bacteria inoculation status and prediction of growth-related traits in tropical maize using hyperspectral image and genomic data. *bioRxiv (Preprint)*.

Yassue, R. M., Galli, G., Junior, R. B., Cheng, H., Morota, G., and Fritsche-Neto, R. (2022b). A low-cost greenhouse-based high-throughput phenotyping platform for genetic studies: a case study in maize under inoculation with plant growth-promoting bacteria. *The Plant Phenome Journal*, 5:e20043.

Zarco-Tejada, P., Miller, J., Morales, A., Berjón, A., and Agüera, J. (2004). Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops. *Remote Sensing of Environment*, 90(4):463–476.

Zarco-Tejada, P. J., Ustin, S. L., and Whiting, M. L. (2005). Temporal and spatial relationships between within-field yield variability in cotton and high-spatial hyperspectral remote sensing imagery. *Agronomy Journal*, 97(3):641–653.

Zhang, Z., Qu, J., Li, F., Li, S., Xu, S., Zhang, R., Xue, J., and Guo, D. (2020). Genome-wide evolutionary characterization and expression analysis of SIAMESE-RELATED family genes in maize. *BMC Evolutionary Biology*, 20(1).

Zhao, D., Reddy, K. R., Kakani, V., Read, J., and Carter, G. (2003). Corn (zea mays l.) growth, leaf pigment concentration, photosynthesis and leaf hyperspectral reflectance properties as affected by nitrogen supply. *Plant and Soil*, 257(1):205–218.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24):3326–3328.

Zhou, Q., Fu, Z., Liu, H., Wang, J., Guo, Z., Zhang, X., Tian, R., Liu, Y., Qu, J., Li, W., Yan, J., and Tang, J. (2021). Mining novel kernel size-related genes by pQTL mapping and multi-omics integrative analysis in developing maize kernels. *Plant Biotechnology Journal*, 19(8):1489–1491.

Zhou, Z., Li, G., Tan, S., Li, D., Weiß, T. M., Wang, X., Chen, S., Würschum, T., and Liu, W. (2020). A QTL atlas for grain yield and its component traits in maize (zea mays). *Plant Breeding*, 139(3):562–574.

Zhu, J., Zhou, Y., Li, J., and Li, H. (2021). Genome-wide investigation of the phospholipase c gene family in zea mays. *Frontiers in Genetics*, 11.

Appendix

Table 1: Selected single nucleotide polymorphisms markers based on BayesC using the posterior inclusion probability threshold of 0.50 for 281 hyperspectral phenotypes under with (B+) or without (B-) plant growth-promoting bacteria inoculation.

Management	Chr[1]	Marker ID[2]	Phenotypes	PIP[3]	NG[4]	Candidate genes
B-	4	CM000780.4 164335549	1	0.508	2	
B-	8	CM000784.4 174418506	5	0.551	9	
B+	1	CM007647.1 69298773	1	0.981	6	
B+	1	CM007647.1 80258452	1	0.820	3	
B+	1	CM007647.1 88464412	1	0.661	7	Zm00001d029820
B+	1	CM007647.1 113763491	1	0.894	5	
B+	1	CM007647.1 173938755	1	0.966	5	
B+	2	CM007648.1 181732219	1	0.549	8	
B+	2	CM007648.1 240687544	1	0.851	4	Zm00001d007843
B+	2	CM007648.1 214786010	14	0.602	9	
B+	4	CM000780.4 245633076	9	0.585	22	nrt2, nrt2.2 Zm00001d054060
B+	4	CM000780.4 206259028	1	0.559	8	
B+	5	CM000781.4 2005858	1	0.581	15	Zm00001d012924
B+	5	CM000781.4 213988333	1	0.604	7	
B+	6	CM000782.4 108027757	5	0.519	10	
B+	6	CM000782.4 9906627	1	0.633	12	100275163, 100192849
B+	7	CM007650.1 22239405	1	0.570	2	
B+	8	CM000784.4 177771765	1	0.826	13	
B+	8	CM000784.4 179214567	10	0.593	11	Zm00001d012719
B+	9	CM000785.4 127613348	1	0.510	3	
B+	10	CM000786.4 88567864	1	0.954	6	
B+	10	CM000786.4 118526736	1	0.569	3	
B+	10	CM000786.4 122822696	1	0.969	12	
B+	10	CM000786.4 137486546	1	0.996	9	
B+	10	CM000786.4 145821750	1	0.544	6	

[1] Chromosome number
[2] Each marker ID is comprised of chromosome ID and marker location that are separatedby the underscore sign.
[3] Average of posterior inclusion probability for the selected phenotypes
[4] Number of genes within the gene interval

Table 2: List of candidate genes influencing both manually measured phenotypes and hyper- spectral phenotypes. Selected single nucleotide polymorphisms markers and their candidate genes influencing both manually measured phenotypes and hyperspectral phenotypes under with (B+) or without (B-) plant growth-promoting bacteria inoculation.

Management	Chr[1]	Marker ID[2]	Phenotypes	NG[3]	Candidate genes
B-	9	CM000785.4 28676437	PH	8	
B+	2	CM007648.1 219293526	SD	11	103647869
B+	4	CM000780.4 181569268	SDM	19	Zm00001d052164, Zm00001d052165
B+	5	CM000781.4 75194167	PH	9	
B+	6	CM000782.4 108027757	PH	10	
B+	7	CM007650.1 41779617	SD	12	Zm00001d006916
B+	8	CM000784.4 167863049	SD	10	LOC100216557
B+	9	CM000785.4 18547071	SD	10	100274563

[1] Chromosome number
[2] Each marker ID is comprised of chromosome ID and marker location that are separated by the underscore sign.
[3] Number of genes within the gene interval

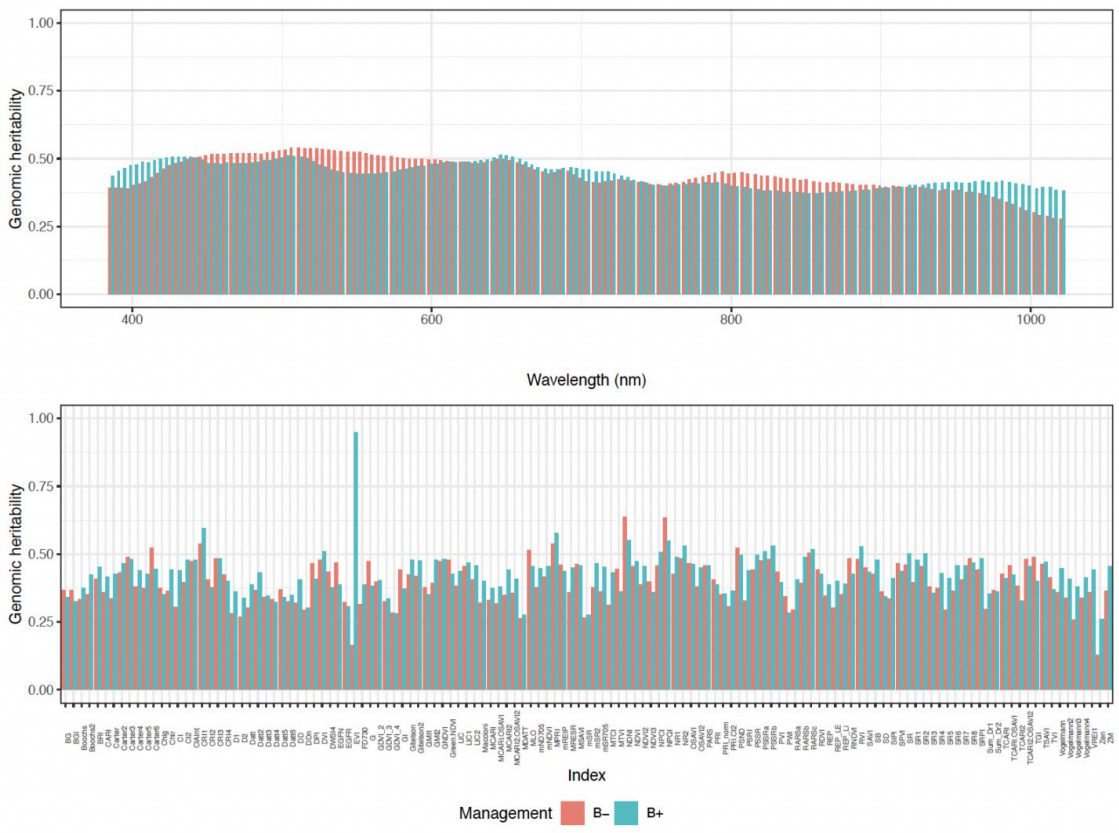

Figure 1. Genomic heritability for 150 hyperspectral reflectance values and 131 hyperspectral indices without (B-) and with (B+) plant growth-promoting bacteria inoculation.

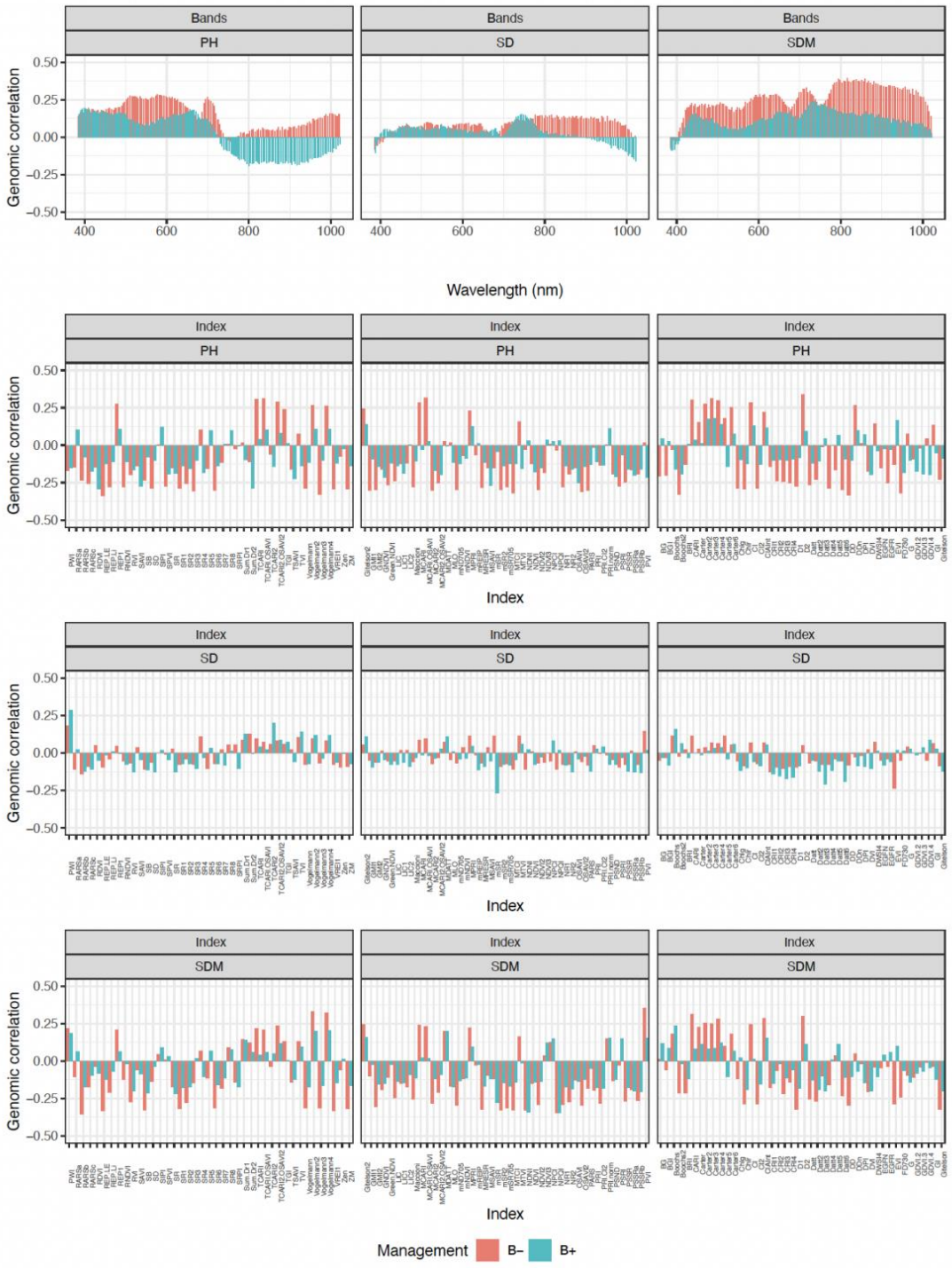

Figure 2. Genomic correlations between manually measured phenotypes and hyperspectral reflectance values and hyperspectral indices under without (B-) and with (B+) plant growth-promoting bacteria inoculation.

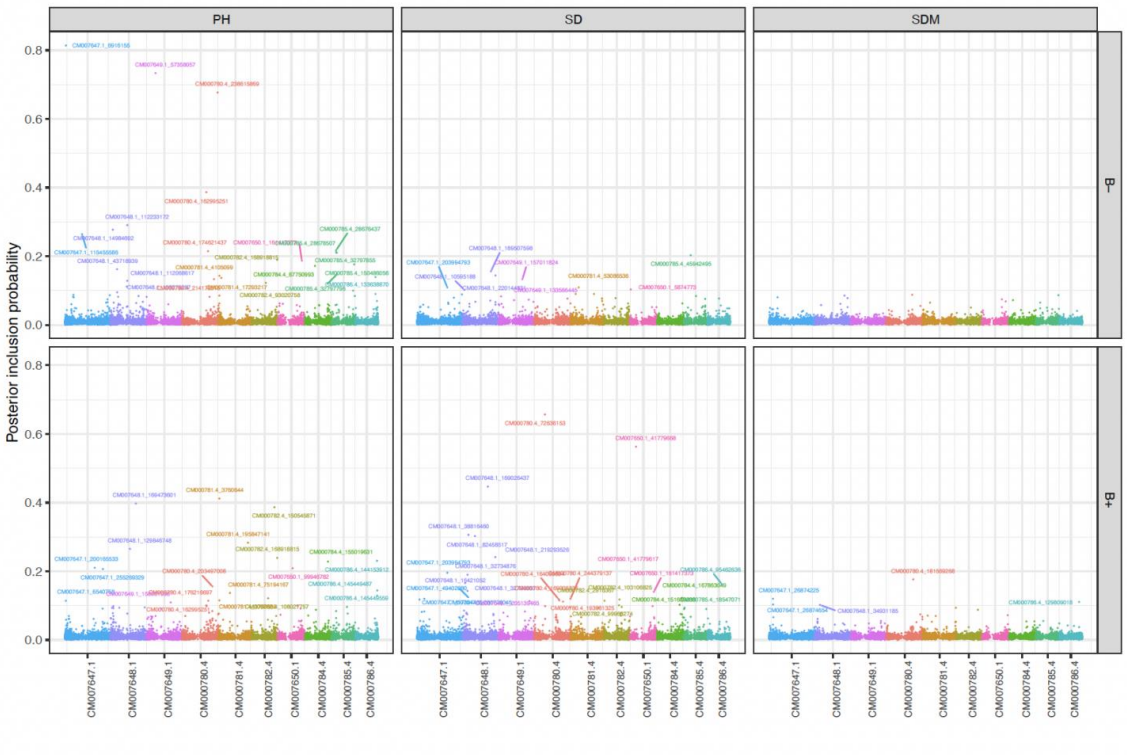

Figure 3: Genome-wide association analysis of manually measured phenotypes without (B-) and with (B+) plant growth-promoting bacteria inoculation. Plant height (PH); stalk diameter (SD) and shoot dry mass (SDM).

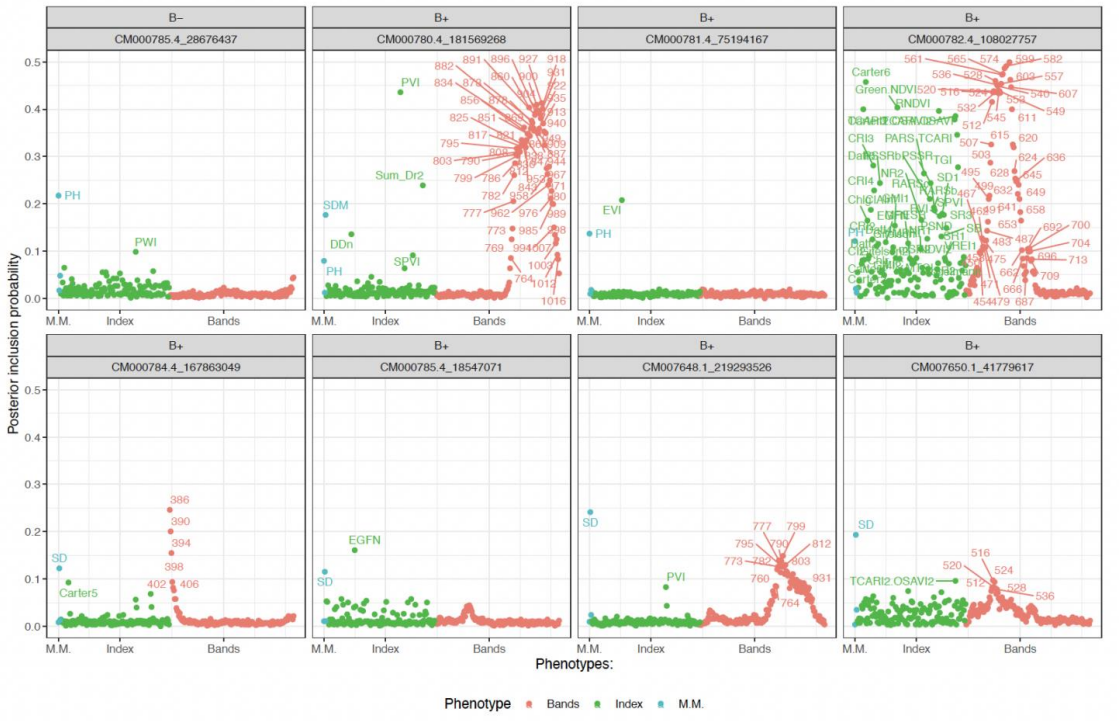

Figure 4: Phenome-wide association analysis plot of selected eight markers having influence on both manually measured and hyperspectral phenotypes without (B-) and with (B+) plant growth-promoting bacteria inoculation. Plant height (PH); stalk diameter (SD) and shoot dry mass (SDM). The abbreviations of hyperspectral indices are defined in Tables S1-S4.