Integrating hyperspectral and multiple phenotype genome-wide association analysis to dissect the genetic architecture of growth related phenotypes in maize

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

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

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

[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

Running title:

Keywords: PheWAS, Multiple phenotypes GWAS, pleiotropy

* Corresponding author
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)

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

Abstract

Introduction

Increase food production in a sustainable manner is a necessity for the next years due the constant increase in the food demand, specially in growing tropical developing nations (Laurance et al., 2014). Several studies reveal that the plant growth-promoting bacteria (PGPB) is a suitable alternative for increase plant resilience against biotic and abiotic stress and with potential to increase food production (Compant et al., 2005; Batista et al., 2018; Yassue et al., 2021a). PGPB has the potential to promote morphological (Mantelin, 2003) and functional (Benedetto et al., 2017) changes in the plant. Some of the reported effects are increase in nutrients uptake such as nitrogen, phosphate, potassium, and iron (Egamberdiyeva, 2007; Pii et al., 2015), and the activation of the response against diseases (Compant et al., 2005; Olanrewaju et al., 2017).

Hyperpsectral image data are valuable resources for genetics and managements studies due to their association with the target phenotypes. Several studies have reported that hyperspectral phenotypes may be associated with water content (Ge et al., 2016), nutrient plant status (Mahajan et al., 2014; Nigon et al., 2021), disease symptoms (Thomas et al., 2017), yield and biomass (Krause et al., 2019), and many others target phenotypes. Also, hyperspectral phenotypes has the potential to assess the genetic variability (Feng et al., 2017) and being an alternative phenotype for targets traits that are hard to measured it (Sun et al., 2019).

Genetic studies such as GWAS, heritability, and genetic correlation can be used to investigate the association between hyperspectral and target phenotypes (Feng et al., 2017; Sun et al., 2019; Herzig et al., 2019; Wu et al., 2021). One limitation of this approach is handle with hundreds to thousands of phenotypes. As alternative, several Genome wide association analysis (GWAS) methodologies for multiple phenotypes have been developed . Denny et al. (2010) proposed the use of PHEWAS analysis where association of each markers was

tested across several phenotypes in order to discovery pleiotropic effects for many diseases Safarova et al. (2019). Also, PHEGWAS have been proposed in order to visualize and interpret multi-phenotypes GWAS (George et al., 2019). Recently, Liang et al. (2020) proposed a new way to perform GWAS performing multi-marker and multi-trait tests called GPWAS. Nevertheless, there only few works that employ the use of multiple phenotypes GWAS for hyperspectral data.

One of the most common measure to assess the accuracy of high-throughput phenotyping (HTP) is the Pearson correlation between the target and HTP phenotype. Nevertheless, the phenotypic correlation take into account the phenotypic covariance, which can be decomposed as follows: $cov_p, = cov_a + cov_d + cov_i + cov_e$, where they are the phenotypic, genetic additive, dominante, epistatic and environmental covariances, respectively (Hill, 2013). Another approach is the use of genetic correlation, which takes into account only the genetic additive covariance and can be explained by pleotropic and linkage disequilibrium between locis that control both phenotypes (van Rheenen et al., 2019). The limitation of using genetic correlation is that the linkage disequilibrium tends to change during the process of selection. As alternative is the identification of HTP phenotypes that share pleiotropy genes with the target phenotype can be used for a more reliable selection.

Even tough the advance of statistical models and computer power, multiple phenotypes GWAS stills remains a challenge. Multivariate models are reliable alternatives for identify genetic association between traits (Feng et al., 2017) and even find pleiotropic candidates genes (Fernandes et al., 2021) but they are limited for only a few phenotypes at time. Increasing the number of phenotype will substantially increases computing resources needed. Phenome-wide association scans (PheWAS) is an alternative to identify which phenotypes are associated with a given genetic variant (Denny et al., 2010). PheWAS was first applied for electronic medical record (EMR) phenotypes to associate single nucleotide polymorphisms (SNPs) markers to disease. Recently, PheWAS have been applied for a wide type of phe-

notypes but limited to diseases and human-related phenotypes (Pendergrass et al., 2011; Millard et al., 2015; Verma et al., 2018; Gutiérrez-Sacristán et al., 2021).

Therefore, the goal of our work is to integrate the data from phenomics into the GWAS framework. For this, we employed Bayesian GWAS, PheWAS scan, and gene annotation in order to identify possible candidate genes to pleiotropic effects in order to study the interaction between maize and PGPB.

Materials and Methods

Plant growth-promoting bacteria experiment

A tropical maize association panel containing 360 inbred lines was used to study the genetic architecture of the response to PGPB. All the lines were evaluated under two managements conditions, with (B+) and without (B-) PGPB inoculation under nitrogen limitation conditions. The PGPB species used were *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) and they were co-inoculated in the maize seeds. The B- management consists of inoculum with only liquid Luria-Bertani medium. The irrigation, weed control, and fertilizitation except for nitrogen were carried out according to the need of the crop. The evaluations of the plants occurred when most of then 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 find in Yassue et al. (2021b,a).

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). The DNA were extract from the leaves using the cetyltrimethylammonium bromide method (Doyle and Doyle, 1987). The SNP calling were performed using the software TASSEL 5.0 (Bradbury et al., 2007) using the 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 single-nucleotide polymorphisms

(SNPs) were obtained after the quality control process. A detailed information regarding the population genomic data are available on (Yassue et al., 2021a).

Hyperspectral imaging and processing

The hyperspectral images for each leaf for each plant in the management B+ and B- were taken using a benchtop system of Pika L. camera (Resonon, Bozeman, MT, USA). The middle portion of the last completed expanded leaf were used as region of interest for each hyperspectral image. A dark room and additional light supply were used in order to reduce light variation. The radiometric calibration was performed according to the manufacture instruction. For each plant, a hyperspectral cube image containing 150 bands with wavelength varying from 386 to 1021 nm. The image processing was performed applyng a mask to remove the background from the image and the mean of reflectance of each pixel was used for further analysis. The processing was performed using the Spectral Python (SPy) package. A summary of hyperspectral imaging and processing can be find in. Using the mean reflectance for each wavelength, we calculated 131 hyperspectral index were using the package hsdar (Lehnert et al., 2019). The summary for each hyperspectral index are available in the supplementary material S04.

Single trait model

In order to estimate the markers effect and variance components, the following model were used for each combination of phenotype and management (B+ and B-):

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_1\mathbf{r} + \mathbf{X}_2\mathbf{b} + \mathbf{Z}_1\mathbf{g} + \boldsymbol{\epsilon}$$

where $\mathbf{y}$ is the vector of phenotypes (manually measured or hyperspectral); $\mathbf{1}$ is the vector of ones; $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ are the incidence matrices for the fixed effects; $\mathbf{Z}_1$ is the incidence

matrices for the random effects; μ is the overall mean; $\mathbf{r}$, $\mathbf{b}$, are the fixed effects for replication, block within replication, and management (B+ and B-), respectively; $\mathbf{g} \sim \mathrm{N}\ (0,\ \mathbf{M}\sigma_A^2)$ is the vector of random effect of genotype; and $\boldsymbol{\epsilon} \sim \mathrm{N}\ (0,\ \mathbf{I}\sigma_\epsilon^2)$ is the random residual effect. $\mathbf{I}$ is the identity matrix; σ_g^2 is the additive genomic variance; and σ_ϵ^2 is the residual variance.

The fixed effect of μ, $\mathbf{r}$, and $\mathbf{b}$ was assigned as a flat prior. The σ_ϵ^2 were assumed to be independently and identically distributed multivariate normal vectors with null mean and covariance matrix $\mathbf{R}$, which is assigned an inverse Wishart prior distribution, $\mathbf{W}^{-1}(S_e, \nu_e)$, with degrees of freedom $\nu_e = 4$ and scale matrix $\mathbf{S}_e$ such that the prior mean of $\mathbf{R}$ equals half of the phenotypic variance and has a scaled inverted chi-square distribution.

All the Bayesian analysis were fitted using 60,000 Markov chain Monte Carlo samples, 6,000 burn-in, 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.

Bivariate model

In order to estimate the genetic correlation between manually measured and hyperspectral traits, we performed bivariate BayesC analysis using the same model described above but the vector $\mathbf{y}$ is a vector containing the a manually measured and a hyperspectral phenotypes; $\mathbf{g} \sim \mathrm{N}\ (0,\ \Sigma_g \otimes \mathbf{M})$ and $\boldsymbol{\epsilon} \sim \mathrm{N}\ (0,\ \Sigma_\epsilon \otimes \mathbf{I})$ the residual; $\otimes$ is the Kronecker product;and Σ_g and Σ_ϵ are the variance-covariance matrices for additive genomic and residual effects taking the forms of

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

where subscripts 1 and 2 refer to the first and second phenotypes. An inverse Wishart distribution was assigned to Σ_g and Σ_ϵ with degrees of freedom $\nu = 4$ and scale matrix S

such that the prior means of Σ_g and Σ_ϵ equal half of the phenotypic variance.

Heritability and genetic correlation

The variance components obtained from the univariate Bayesian BayesC were used to estimate genomic heritability using the following formula.

$$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{n_r}}.$$

The estimates of genomic correlation were obtained from the estimated variance-covariance matrix in the bivariate BayesC bivariate model between all the manually measured and hyperspectral phenotypes

Bayesian GWAS and PheWAS scan

Bayesian genome-wide association analysis are usually performed computing model frequency of each marker, using the posterior probability of association of the genomic window (WPPA) or based on the marker effects (Fan et al., 2011; Fernando and Garrick, 2013; Peters et al., 2021). The preference of using model frequency instead of windows based approach is due the low market density in the population used in the study (Fan et al., 2011). Compute the model frequency for each marker (the probability the marker is included in the model) using samples of marker effects stored in *markereffects* file.

Traditionally, PheWAS analysis consists in linear regression between the genetic variants (SNPs data) and multiple phenotypes (Denny et al., 2010). The most limitation in the analysis is the lack of correction to population structure and take account to multiple testing corrections for the p values. In order to partially address this problem, we use a bayesian GWAS approach that can be easily expanded for others models. The PheWAS "Manhattan" plots were used to visualize possible association between SNP and several phenotypes (Carroll

et al., 2014). The PheWAS "Manhattan" plots contaminating the model frequency for all phenotypes for each candidate SNP were plotted using the R package ggplot2 (Wickham, 2016).

For each selected SNPs obtained from the hyperspectral phenotype, the genes within an interval of 50 kilo base pair (kbp) upstream and downstream of the marker were identified using the MaizeMine V1.3 server (Shamimuzzaman et al., 2020). Then, the genes within this interval that have been previews reported as candidate to growth-related phenotypes were considered as candidate genes for the hyperspectral phenotypes.

Results

Exploratory

The correlation between single wavelength reflectance and hyperspectral index were similar across the manually measured phenotypes (PH, SD, SDM) (Figure 1). The principal components analysis reveals no clustering between management B+ and B- however showed strong association between the wavelengths and hyperspectral index (Figure S1). This tendency was supported by the correlation matrix where 3 to 4 groups of wavelengths or index had strong correlation between then (Figure S2 and S3).

Genomic heritability and genetic correlation

The genomic heritability for the manually measured phenotypes varied from 0.28 to 0.61 (Table 1). Plant height had the highest values while SDM the lowest. For the Hyperspectral phenotypes, the wavelength and indices heritability mean values of 0.45 and 0.41, respectively. Similar heritabilities were observed for B- and B+ management for both manually measured and hyperspectral phenotypes.

The genetic correlation between hyperspectral and manually measured phenotypes varied greatly across phenotypes. The highest values was for SDM and the lowest was for SD. B+ managements tend to have higher genomic correlation. No major differences were observed between Index and wavelength was observed.

PheWAS Scan

A total of 86 SNPs were selected from Bayesian GWAS analysis using a model frequency threshold of 0.10 for PH, SDM, and SD in both management (Figure 4 and Supplementary Table 1-3). For the 281 hyperspectral phenotypes, a wide range of SNPs were also obtained

and are available in the supplementary materials.

The PheWAS scan analysis for the ten markers with the highest model frequency for the manually measured phenotype revealed no patterns across phenotypes for plant height. On the other hand, for SD and SDM, a few SNPs showed model frequency above the threshold for both manually and hyperspectral phenotype 6 5.

Multiple Phenotype GWAS

A model frequency threshold of 0.50 was used for selecting SNPs for the hyperspectral phenotypes. The gene annotation gene annotation of each selected SNPs revealed that some of this markers was associated with more than one hyperspectral phenotype and they may be associated with reported genes in the literature related to growth-related phenotypes or response to abiotic stress (Table 2).

Discussion

PGPB response

The high number of selected SNPs for manually measured phenotype suggested that several of them may be a false positive. On other hand, from the 28 SNPs previous reported in the same population but using a different model, 12 has been overlapped (Yassue et al., 2021a). Only 4 SNPs were overlapped between B+ and B-. Similar results for (Yassue et al., 2021a).

Genomic heritability and genetic correlation

PheWAS scans

Even tough only a small proportion of SNPs were associated with both manually measured and hyperspectral phenotypes, the hyperspectral phenotypes has the potential to assess the genetic variability in maize and also identify genomic candidate areas. PH, SD, and SDM are growth-related phenotypes and they have mid association with hyperspectral phenotypes. Other phenotypes that may be strong related with the reflectance, such as protein contain,

Multiple Phenotype GWAS

The limitation of our work is that even though all hyperspectral phenotypes are highly correlated between them we assume they as independent. In this sense, further studies should be performed in order to deal with high-dimension covariance matrix. Nevertheless, our findings reveals an alternative approach for identification of candidates pleiotropic effects and data visualization for multiple phenotype GWAS.

The SNP CM000780.4245633076 was selected by 9 hyperspectral phenotypes and three candidate genes was found in the genomic window. The candidate genes nrt2, nrt2.2, and Zm00001d054060 have been previews reported as NO_3^- transporter gene families and candi-

date for nitrate uptake along the maize primary root (Liu et al., 2009; Sorgonà et al., 2011; Wang et al., 2020).

The gene Zm00001d029820 and Zm00001d012924 have been reported as a putative for resposne for plant development and environmental stress conditions (Zhang et al., 2020; Zhu et al., 2021). The candidate gene Zm00001d007843 was previous reported for kernel size-related genes (Zhou et al., 2021).The candidate gene Zm00001d012719 was previous reported as candidate transcription factor to mediate plant response to abiotic stress (Vendramin et al., 2020).

Nevertheless, the LD decay of the population and the low number of markers limits the interpretation regarding the gene anotation and the results must be interpreted with limitations.

Conclusion

Despite the limitation of our work, we present a framework analysis for multiple phenotype GWAS and its interpretation for hyperspectral image data. The PheWAS scan analysis is a suitable alternative for identification of candidates pleiotripic SNPs across phenotypes.

References

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.

Benedetto, N. A. D., , Corbo, M. R., Campaniello, D., Cataldi, M. P., Bevilacqua, A., Sinigaglia, M., and and, Z. F. (2017). The role of plant growth promoting bacteria in improving nitrogen use efficiency for sustainable crop production: a focus on wheat. *AIMS Microbiology*, 3(3):413–434.

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.

Carroll, R. J., Bastarache, L., and Denny, J. C. (2014). R PheWAS: data analysis and plotting tools for phenome-wide association studies in the r environment. *Bioinformatics*, 30(16):2375–2376.

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*.

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.

Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D. R., Roden, D. M., and Crawford, D. C. (2010). PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*, 26(9):1205–1210.

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).

Fernandes, S. B., Zhang, K. S., Jamann, T. M., and Lipka, A. E. (2021). How well can multivariate and univariate GWAS distinguish between true and spurious pleiotropy? *Frontiers in Genetics*, 11.

Fernando, R. L. and Garrick, D. (2013). Bayesian methods applied to GWAS. In *Genome-Wide Association Studies and Genomic Prediction*, pages 237–274. Humana Press.

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.

George, G., Gan, S., Huang, Y., Appleby, P., Nar, A. S., Venkatesan, R., Mohan, V., Palmer, C. N. A., and Doney, A. S. F. (2019). PheGWAS: a new dimension to visualize GWAS across multiple phenotypes. *Bioinformatics*, 36(8):2500–2505.

Gutiérrez-Sacristán, A., Sáez, C., Niz, C. D., Jalali, N., DeSain, T. N., Kumar, R., Zachariasse, J. M., Fox, K. P., Palmer, N., Kohane, I., and Avillach, P. (2021). Multi-PheWAS intersection approach to identify sex differences across comorbidities in 59 140 pediatric patients with autism spectrum disorder. *Journal of the American Medical Informatics Association*, 29(2):230–238.

Herzig, P., Backhaus, A., Seiffert, U., von Wirén, N., Pillen, K., and Maurer, A. (2019). Genetic dissection of grain elements predicted by hyperspectral imaging associated with yield-related traits in a wild barley NAM population. *Plant Science*, 285:151–164.

Hill, W. (2013). Genetic correlation. In *Brenner's Encyclopedia of Genetics*, pages 237–239. Elsevier.

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.

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.

Liang, Z., Qiu, Y., and Schnable, J. C. (2020). Genome–phenome wide association in maize and arabidopsis identifies a common molecular and evolutionary signature. *Molecular Plant*, 13(6):907–922.

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.

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.

Millard, L. A. C., Davies, N. M., Timpson, N. J., Tilling, K., Flach, P. A., and Smith, G. D. (2015). MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using mendelian randomization. *Scientific Reports*, 5(1).

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).

Pendergrass, S., Brown-Gentry, K., Dudek, S., Torstenson, E., Ambite, J., Avery, C., Buyske, S., Cai, C., Fesinmeyer, M., Haiman, C., Heiss, G., Hindorff, L., Hsu, C.-N., Jackson, R., Kooperberg, C., Marchand, L. L., Lin, Y., Matise, T., Moreland, L., Monroe, K., Reiner, A., Wallace, R., Wilkens, L., Crawford, D., and Ritchie, M. (2011). The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genetic Epidemiology*, 35(5):410–422.

Peters, S. O., Kızılkaya, K., Ibeagha-Awemu, E. M., Sinecen, M., and Zhao, X. (2021). Comparative accuracies of genetic values predicted for economically important milk traits, genome-wide association, and linkage disequilibrium patterns of canadian holstein cows. *Journal of Dairy Science*, 104(2):1900–1916.

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 Microbiology*, 78(21):7511–7518.

Safarova, M. S., Satterfield, B. A., Fan, X., Austin, E. E., Ye, Z., Bastarache, L., Zheng, N., Ritchie, M. D., Borthwick, K. M., Williams, M. S., Larson, E. B., Scrol, A., Jarvik, G. P., Crosslin, D. R., Leppig, K., Rasmussen-Torvik, L. J., Pendergrass, S. A., Sturm, A. C., Namjou, B., Shah, A. S., Carroll, R. J., Chung, W. K., Wei, W.-Q., Feng, Q., Stein, C. M., Roden, D. M., Manolio, T. A., Schaid, D. J., Denny, J. C., Hebbring, S. J., de Andrade, M., and Kullo, I. J. (2019). A phenome-wide association study to discover pleiotropic effects of PCSK9, APOB, and LDLR. *npj Genomic Medicine*, 4(1).

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.

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.

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).

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 Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., and Wray, N. R. (2019). Genetic correlations of polygenic disease traits: from theory to practice. *Nature Reviews Genetics*, 20(10):567–581.

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.

Verma, A., Lucas, A., Verma, S. S., Zhang, Y., Josyula, N., Khan, A., Hartzel, D. N., Lavage, D. R., Leader, J., Ritchie, M. D., and Pendergrass, S. A. (2018). PheWAS and beyond: The landscape of associations with medical diagnoses and clinical measures across 38, 662 individuals from geisinger. *The American Journal of Human Genetics*, 102(4):592–608.

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).

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

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).

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)*.

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).

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.

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.

Table

Table 1: Genomic heritability estimates of manual manually phenotypes with (B+) or without (B-) plant growth-promoting bacteria inoculation. PH: plant height; SDM: shoot dry mass; f(biomass): plant biomass index; SD: Stalk diameter.

Management	PH	SD	SDM
B-	0.61	0.39	0.28
B+	0.61	0.60	0.30

Table 2: Selected SNPs based on BayesC approach using a threshold of 0.50 for model frequency for 281 hyperspectral phenotypes. Management, Marker ID, phenotypes, model Frequency, and number of genes

Management	Marker ID[1]	Phenotypes	MF[2]	NGs[3]	Candidate genes[4]
B+	CM000780.4_245633076	9	0.585	22	nrt2, nrt2.2 Zm00001d054060
B+	CM007648.1_181732219	1	0.549	8	
B+	CM000781.4_2005858	1	0.581	15	Zm00001d012924
B+	CM007650.1_22239405	1	0.570	2	
B+	CM007647.1_69298773	1	0.981	6	
B+	CM007647.1_80258452	1	0.820	3	
B+	CM007647.1_88464412	1	0.661	7	Zm00001d029820
B+	CM007647.1_113763491	1	0.894	5	
B+	CM007647.1_173938755	1	0.966	5	
B+	CM007648.1_240687544	1	0.851	4	Zm00001d007843
B+	CM000780.4_206259028	1	0.559	8	
B+	CM000781.4_213988333	1	0.604	7	
B+	CM000782.4_9906627	1	0.633	12	100275163, 100192849
B+	CM000784.4_177771765	1	0.826	13	
B+	CM000786.4_88567864	1	0.954	6	
B+	CM000786.4_118526736	1	0.569	3	
B+	CM000786.4_122822696	1	0.969	12	
B+	CM000786.4_137486546	1	0.996	9	
B+	CM000786.4_145821750	1	0.544	6	
B+	CM007648.1_214786010	14	0.602	9	
B+	CM000785.4_127613348	1	0.510	3	
B+	CM000784.4_179214567	10	0.593	11	Zm00001d012719
B+	CM000782.4_108027757	5	0.519	10	
B-	CM000780.4_164335549	1	0.508	2	
B-	CM000784.4_174418506	5	0.551	9	

[1] SNP id where the first part in which the first part refers to the chromosome and the second to the physical position

[2] Average of the model frequency for the selected phenotypes

[3] Number of genes within the gene interval

[4] Candidate genes based on gene annotation

Figure 1: Phenotypic correlations between plant height (PH), Stalk diameter (SD), and shoot dry mass (SDM) and hyperspectral reflectance (left) and hyperspectral index (right) with and without PGPB inoculation (B+ and B-).

Figure 2: Genomic heritability for 150 wavelengths and 131 hyperspectral indices with and without inoculation with PGPB (B+ and B-)

Figure 3: Genomic correlation for 150 wavelengths and 131 hyperspectral indices with and without inoculation with PGPB (B+ and B-)

Figure 4: Manhatan plot for traditional phenotypes with and without inoculation with PGPB (B+ and B-)

Figure 5: Phenome-wise scan for association the 10 highest model frequency SNPs for plant height (PH), stalk diameter (SD), and shoot dry mass (SDM) for B+ management

Figure 6: Phenome-wise scan for association the 10 highest model frequency SNPs for plant height (PH), stalk diameter (SD), and shoot dry mass (SDM) for B- management