

INTEGRATING PHENOMICS AND GENOMICS TOWARDS  
ACCELERATING GENETIC GAIN IN SOFT WINTER WHEAT

By

Jonathan Santiago Concepcion

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Plant Breeding, Genetics, and Biotechnology – Crop and Soil Sciences – Doctor of Philosophy

2025

## ABSTRACT

Accelerating genetic gain in plant breeding demands increased selection intensity, enhanced selection accuracy, broader genetic diversity, and shortened breeding cycle. As phenomic platforms and genomic resources continue to evolve, integrating these complementary datasets offers opportunity to improve breeding pipelines. This dissertation explores multiple approaches to potentially incorporating phenomic information with genomic data, aiming to increase prediction accuracy and selection intensity, and identify genomic regions for economically important traits in soft winter wheat.

Infrared thermal imaging enabled high-resolution differentiation of *Fusarium* head blight (FHB)-resistant and -susceptible genotypes at the single-spike level. However, field-scale implementation requires careful consideration due to uncontrolled factors under field conditions, where no direct relationship between plot-level infrared thermal readings and FHB-related traits was established. Hyperspectral imaging demonstrated superior predictive ability and prediction accuracy over genomic prediction alone for deoxynivalenol (DON) content. Integrating phenomic and genomic predictions by model blending enhanced prediction accuracy and allowed clustering-based selection on predicted DON content in F<sub>4:5</sub> breeding lines. Additionally, multiple strategies for integrating UAV-derived vegetation indices (VIs) to improve genomic prediction accuracy were evaluated, with varying degrees of success depending on how UAV-derived information were used as fixed effect, as well as training set composition. Beyond enhancing prediction, phenomic data facilitated the identification of key genomic regions associated with DON content and grain yield, underscoring the potential of phenomic information as a phenotypic input in genome-wide association studies.

Collectively, these findings support the potential application phenomics-genomics integration in potentially improving wheat breeding. By enhancing selection accuracy, identifying informative genomic regions, and potentially reducing selection intervals, this work lays the foundation for a more efficient breeding pipeline. However, careful consideration is essential when implementing combined phenomic-genomic approaches to ensure robust, field-applicable results.

Copyright by  
JONATHAN SANTIAGO CONCEPCION  
2025

For all those who failed and never gave up.  
This is for you. This is for us.

## ACKNOWLEDGEMENTS

First, I want to express my deepest gratitude to my adviser, Dr. Eric Olson, for the incredible opportunity to be a part of the MSU Wheat Breeding and Genetics Team. His mentorship has been key to my growth over the past few years—encouraging me to pursue my passion in plant breeding and genetics, take the lead on major objectives, and constantly push the boundaries of my knowledge and skills. I am forever thankful for his guidance, trust, and support throughout this journey.

I would also like to thank the members of the MSU Wheat Breeding and Genetics Team—Amanda Noble, Dennis Pennington, Samantha Mitchell, Amelia Orr, Elizabeth Ross, Aaron Newberry, Sadie Finegan, and all our hardworking undergraduate research assistants (Mattie, Trent, Courtney, Chelsey, Daniel, Laken)—for their invaluable help in fieldwork, sample prep, disease phenotyping, and data collection. I couldn't have asked for a better team.

To my committee members—Dr. Addie Thompson, Dr. Gustavo de los Campos, and Dr. David Douches—thank you for your thoughtful feedback, guidance, and for challenging me to grow as a scientist. Your insights have played a major role in shaping the direction and success of this work.

I'm also grateful to Rober Goodwin, Ava Gawel, and the entire MSU RS-GIS Team for their incredible help with UAV flights and image processing. Many thanks to Dr. Yanhong Dong for the DON analyses, and to Dr. Katherine Frels, Dr. Francisco Gomez, and Dr. Leonardo Volpato for their inputs and helpful discussions.

Lastly, thank you to my family, especially my Mom, my friends in the Philippines, my friends and loved ones here in the U.S., and the MSU Filipino community—for keeping me grounded through it all.

## TABLE OF CONTENTS

CHAPTER I: INTRODUCTION.....	1
REFERENCES.....	4
CHAPTER II: GENOMIC REGIONS INFLUENCING THE HYPERSPECTRAL PHENOME OF DEOXYNIVALENOL INFECTED WHEAT.....	7
REFERENCES.....	24
CHAPTER III: BLENDED GENOMIC AND HYPERSPECTRAL IMAGING-BASED PREDICTIONS ENABLE SELECTION FOR REDUCED DEOXYNIVALENOL CONTENT IN WHEAT GRAINS.....	29
REFERENCES.....	64
CHAPTER IV: UAV-DERIVED INFORMATION AND GENOMIC DATA INTEGRATION FORENHANCE GRAIN YIELD PREDICTIONS IN SOFT WINTER WHEAT.....	69
REFERENCES.....	119
CHAPTER V: EVALUATING CLOSE-RANGE INFRARED THERMAL IMAGING AND UNSUPERVISED K-MEANS CLUSTERING FOR FUSARIUM HEAD BLIGHT RESISTANCE SELECTION.....	126
REFERENCES .....	149
CHAPTER VI: INTEGRATED PHENOMIC AND GENOMICS: PROSPECTS, CAVEATS, IMPLICATION IN GENETIC GAIN AND APPLICATION IN WHEAT BREEDING .....	151

## CHAPTER I: INTRODUCTION

Global wheat production increased from 640 million metric tons in 2010 to 760 million in 2020 (FAO, 2022). However, wheat-harvested areas have declined, from 32 million ha in North America in 2000 to 25 million in 2020 (FAO, 2022). With the global population projected to reach 9.7 billion by 2050 (United Nations, 2022), along with reducing production area, developing high-yielding wheat varieties is crucial to ensure food security.

A major limitation in wheat production is disease prevalence, particularly Fusarium head blight (FHB). Yield losses of 10–70% have been reported due to FHB epidemics (Khan et al., 2020), with economic losses of up to \$7.6 billion in the U.S. from 1993 to 2001 (Khan et al., 2020). Mycotoxins produced by FHB pathogens degrade grain quality and pose risks to human and animal health (Soni et al., 2020). Thus, breeding FHB-resistant wheat varieties is a key management strategy (Dweba et al., 2017; Steiner et al., 2017).

Developing high-yielding, disease-resistant varieties with superior agronomic traits relies on accelerating genetic gain (Poland & Rutkowski, 2016). This requires increasing selection intensity, improving heritability through precise phenotyping, maintaining genetic diversity, and reducing generation intervals (Falconer & Mackay, 1996; Araus et al., 2018). However, despite recent advances, genetic gain has stagnated due to phenotyping limitations (Araus et al., 2018).

Selection intensity depends on evaluating large numbers of breeding lines to identify superior genotypes while maintaining genetic diversity (Poland & Rutkoski, 2016). Field evaluation of complex traits like yield and disease resistance remains a bottleneck, affecting genetic gain. Accurate phenotyping is crucial, but traditional methods like visual scoring are subjective, labor-intensive, and error-prone (Zhang et al., 2020; Su et al., 2021).



To overcome phenotyping challenges, breeders have adopted high-throughput imaging technologies (Araus & Cairns, 2013). Infrared thermography assesses canopy temperature to evaluate stress responses and disease resistance (Pineda et al., 2020). While useful for detecting drought and heat stress (Singh et al., 2022; Yao et al., 2018) and diseases like stripe rust and powdery mildew (Feng et al., 2021), thermography is extremely sensitive to environmental fluctuations (Costa et al., 2013), requiring corrections for accuracy.

More sophisticated multispectral and hyperspectral imaging have been employed for assessing abiotic stress tolerance, nutrient use efficiency, and physiological traits (Araus & Cairns, 2014; Araus et al., 2018). Spectral data have been used for predicting grain yield (Lopez-Cruz et al., 2020; Tao et al., 2020; Moghimi et al., 2020) and detecting FHB resistance (Zelazny et al., 2021; Alisaac et al., 2018). However, most studies on FHB detection using spectral imaging have been conducted in controlled environments (Mahlein et al., 2019; Alisaac et al., 2018). Recent field applications have been explored (Ma et al., 2021; Zelazny et al., 2020; Zhang et al., 2020), but further research is needed to refine these methods for breeding applications.

To complement phenotyping, molecular approaches such as genomic selection enhance selection intensity and accuracy while reducing generation intervals (Heffner et al., 2009; Poland & Rutkoski, 2016). Genomic selection enables early-generation selection, minimizing the number of lines progressing to advanced trials. It has been widely applied for improving complex traits, including grain yield and FHB resistance in wheat (Rutkoski et al., 2012; Jiang et al., 2014; Mirdita et al., 2015; Arruda et al., 2015).

Despite these advances, further genetic improvement requires identifying novel genes and alleles associated with FHB resistance and yield (Poland & Rutkoski, 2016). Given the polygenic nature of these traits, stacking QTLs and beneficial alleles is essential (Kushalappa et al., 2016).

Numerous QTLs for FHB resistance has been identified (Zheng et al., 2020; Elfatih et al., 2020; Petersen et al., 2016; Ruan et al., 2020; Soni et al., 2020; Kage et al., 2017; Gadaleta et al., 2019), but the complexity of the wheat genome warrants further exploration.

QTL identification still depends on accurate phenotypic data, emphasizing the need for advanced phenotyping methods to improve gene discovery and breeding applications (Poland & Rutkoski, 2016). Hyperspectral imaging has been used in genome-wide association studies (GWAS) for sorghum (Miao et al., 2020), rice (Feng et al., 2017; Sun et al., 2019), and barley (Herzig et al., 2019; Grieco et al., 2022) but remains underexplored in wheat.

Therefore, this dissertation aims to explore the potential use of phenomic platforms to enhance prediction accuracy and increase objectivity in evaluating major economically important traits in wheat – Fusarium head blight resistance and grain yield. Further, this dissertation aims to enhance prediction accuracy by integrating phenomic information with genomic information while identifying genomic regions with potential association with these economically important quantitative traits.

## REFERENCES

- Alisaac, E. *et al.* Hyperspectral quantification of fusarium head blight: comparison of two *Fusarium* species. *European Journal of Plant Pathology* **152**, 869-884 (2018).
- Araus, J.L. & Cairns, J.E. Field high-throughput phenotyping: the new crop breeding frontier. *Trends in Plant Science*, <https://dx.doi.org/10.16/j.tplants.2013.09.008> (2013).
- Araus, J.L. *et al.* 2018. Translating high-throughput phenotyping into genetic gain. *Trends in Plant Science* **23**, 451-465 (2018).
- Arruda, M.P. *et al.* Genomic selection for predicting head blight resistance in a wheat breeding program. *The Plant Genome*, doi:10.3835/plantgenome2015.01.0003 (2015).
- Costa, J.M., Grand, O.M. & Chaves, M.M. Thermography to explore plant-environment interactions. *Journal of Experimental Botany* **64**, 3937-3949. <https://doi.org/10.1093/jxb/ert029> (2013).
- Dawei, S. *et al.* Using hyperspectral analysis as potential high throughput phenotyping tool in GWAS for protein content of rice quality. *Plant Methods* **15**, 54 (2019).
- Dweba, C.C. *et al.* Fusarium head blight of wheat: pathogenesis and control strategies. *Crop Protection* **91**, 114-122 (2017).
- Elfatih, A. *et al.* Genetic dissection of Fusarium head blight resistance in spring wheat cv. 'Glenn'. *Euphytica* **216**, 71 (2020).
- Falconer, D.S. & Mackay, T.F.C. *Introduction to Quantitative Genetics*. Upper Saddle River, NJ: Pearson Edu. 4<sup>th</sup> ed (1996).
- Feng, H. *et al.* An integrated hyperspectral imaging and genome-wide association analysis platform provides spectral and genetic insights into the natural variation in rice. *Scientific Reports* **7**, 4401 (2017).
- Feng, Z. *et al.* Monitoring wheat powdery mildew based on hyperspectral, thermal infrared, and RGB image data fusion. *Sensors* **22**, 31 (2021).
- Gadaleta, A. *et al.* Map-based cloning of QFhb.mgb-2A identifies a WAK2 gene responsible for Fusarium head blight resistance in wheat. *Scientific Report* **9**, 6929 (2019).
- Grieco, M. *et al.* Dynamics and genetic regulation of leaf nutrient concentration in barley based on hyperspectral imaging and machine learning. *Plant Science* **315**, 111123 (2022).
- Heffner, E.L., Sorells, M.E. & Jannick J-L. Genomic selection for crop improvement. *Crop Sci.* **49**, 1-12 (2009).

- Herzing, P. *et al.* 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 (2019).
- Jiang, Y. *et al.* Potential and limits to unravel the genetic architecture and predict the variation of Fusarium head blight resistance in European winter wheat (*Triticum aestivum* L.). *Heredity* **114**, 318-326 (2015).
- Khan, M.K. *et al.* Fusarium head blight in wheat: contemporary status and molecular approaches. *3 Biotech* **10**, 172 (2020).
- Kushalappa, A.C, Yogendra, K.N. & Karre S. Plant innate immune response: qualitative and quantitative resistance. *Crit. Rev. Plant. Sci.* **35**, 38-55 (2016).
- Lopez-Cruz, M. *et al.* Regularized selection indices for breeding value prediction using hyperspectral image data. *Scientific Reports*, **10**, 8195 (2020).
- Ma, H. *et al.* Using UAV-based hyperspectral imagery to detect winter wheat fusarium head blight. *Remote Sensing* **13**, 3024 (2021).
- Mahlein, A.K. *et al.* Comparison and combination of thermal, fluorescence, and hyperspectral imaging for monitoring Fusarium head blight on spikelet scale. *Sensors* **19**, 2281 (2019).
- Masri, A.A. *et al.* Impact of primary infection site of *Fusarium* species on head blight development in wheat ears evaluated by IR-thermography. *European Journal of Plant Pathology*, **147**, 855-868 (2017).
- Miao, C. *et al.* Semantic segmentations of sorghum using hyperspectral data identifies genetic associations. *Plant Phenomics* **2020**, 421637 (2020).
- Mirdita, V *et al.* Potential and limits of whole genome prediction of resistance to Fusarium head blight and Septoria tritici blotch in a vast central European elite winter wheat population. *Theor. Appl. Genet.* **128**, 2471-2481 (2015).
- Moghim, A., Yang C & Anderson J.A. Aerial hyperspectral imagery and deep neural networks for high-throughput yield phenotyping in wheat. *Computers and Electronics in Agriculture*, **172**, 105299 (2020).
- Rutkoski, J., Benson J., Jia Y., Brown-Guedira G, Jannink J-L, Sorrells M. Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *Plant Genome*, **5**, 51-61 (2012).
- Petersen, S. *et al.* Validation of Fusarium head blight resistance QTL in US winter wheat. *Crop Science*, **57**, 1-12 (2016).

- Pineda, M., Baron, M. & Perez-Bueno, M.L. Thermal imaging for plant stress detection and phenotyping. *Remote Sensing* **13**, 68. <https://dx.doi.org/10.3390/rs1310068> (2021).
- Poland, J. & J, Rutkoski. 2016. Advances and challenges in genomic selection for disease resistance. *Annu. Rev. Phytopathol.* **54**, 79-98. <https://dx.doi.org/10.1146/annurev-pytha-080615-100056> (2016).
- Ruan, Y. *et al.* Characterization of the genetic architecture of Fusarium head blight resistance in durum wheat: the complex association of resistance, flowering, and height genes. *Frontiers in Plant Science*, **11**, 592064 (2020).
- Singh, R.N. *et al.* Application of thermal and visible imaging to estimate stripe rust disease severity in wheat using supervised image classification methods. *Ecological Informatics* **71**, 101774 (2022).
- Soni, N. *et al.* Role of laccase gene in wheat NILs differing at QTL-Fhb1 for resistance against Fusarium head blight. *Plant Science* **298**, 110574 (2020).
- Steiner, B. *et al.* Breeding strategies and advances in line selection for fusarium head blight resistance in wheat. *Trop. Plant Pathol* **42**, 165-174 (2017).
- Su, W.H. *et al.* Automatic evaluation of wheat resistance to Fusarium high blight using dual mask-RGNN deep learning frameworks in computer vision. *Remote Sensing* **13**, 26 (2021).
- Tao, H. *et al.* Estimation of the yield and plant height of winter wheat using UAV-based hyperspectral images. *Sensors*. **20**, 1231 (2020).
- Yao, Z., He D. & Lei Y. Thermal imaging for early and non-destructive detection of wheat stripe rust. In ASABE Annual Conference Meeting 2018 (p.1). *American Society of Agricultural and Biological Engineers* (2018).
- Zelazny, W.R., Chrpova, J. & Hamouz P. Fusarium head blight detection of spectral measurements in a field phenotyping setting – a pre-registered study. *Biosystem Engineering* **211**, 97-113 (2021).
- Zhang, D., *et al.* Development and evaluation of a new spectral disease index to detect wheat fusarium head blight using hyperspectral imaging. *Sensors* **20**, 2260 (2020).
- Zheng, T. *et al.* Integration of meta-QTL discovery with omics: towards a molecular breeding platform for improving wheat resistance to Fusarium head blight. *The Crop Journal* **9**, 739-749 (2020).

## CHAPTER II: GENOMIC REGIONS INFLUENCING THE HYPERSPECTRAL PHENOME OF DEOXYNIVALENOL INFECTED WHEAT

*[Published: Scientific Reports, 14:19340 (2024)]*

### Abstract

The quantitative nature of fusarium head blight (FHB) resistance requires further exploration of the wheat genome to identify regions conferring resistance. In this study, we explored the application of hyperspectral imaging of Fusarium-infected wheat kernels and identified regions of the wheat genome contributing significantly to the accumulation of Deoxynivalenol (DON) mycotoxin. Strong correlations were identified between hyperspectral reflectance values for 204 wavebands in the 397 nm to 673 nm range and DON mycotoxin. Dimensionality reduction using principal components was performed for all 204 wavebands and 38 sliding windows across the range of wavebands. The first principal component (PC1) of all 204 wavebands explained 70% of the total variation in waveband reflectance values and was highly correlated with DON mycotoxin. PC1 was used as a phenotype in a genome wide association study and a large effect QTL on chromosome 2D was identified for PC1 of all wavebands as well as nearly all 38 sliding windows. The allele contributing variation in PC1 values also led to a substantial reduction in DON. The 2D polymorphism affecting DON levels localized to the exon of TraesCS2D02G524600 which is upregulated in wheat spike and rachis tissues during FHB infection. This work demonstrates the value of hyperspectral imaging as a correlated trait for investigating the genetic basis of resistance and developing wheat varieties with enhanced resistance to FHB.

## 2.1. Introduction

Fusarium head blight (FHB) results in significant grain quality and yield reductions that limit profits for wheat farmers and presents challenges in managing mycotoxins and affects wheat production worldwide. During FHB infection by the ascomycete fungus *Fusarium graminearum* Schwabe, Deoxynivalenol (DON) mycotoxin accumulates in wheat kernels (Mirocha et al., 1994). DON is harmful to both humans and animals when ingested (Foroud et al., 2019) and is tightly regulated by testing grain at the point of sale prior to entering the marketplace. Increasing the level of genetic resistance to FHB through breeding is a highly effective mechanism to minimize DON levels on individual farms and limit the amount of mycotoxin entering the grain marketplace (Mesterhazy 2014).

Type III resistance to FHB, resistance to DON accumulation, remains a challenge in FHB-improvement programs. As a quantitative trait, Type III resistance is controlled by multiple QTLs (Bai et al., 2018). Several QTLs for lower DON accumulation have been reported with da Silva et al. (2019) reporting a large effect QTL in 5A accounting for 13% phenotypic variation for DON. He et al. (2019) identified two major QTLs in 3B, and 3D and Larkin et al. (2020) identified ten significant marker trait associations across the genome. Recently, Haile et al. (2023) identified nine QTLs associated with DON accumulation using a multi-locus genome wide association study (GWAS) model.

Phenotyping DON mycotoxin in grain samples relies on GC/MS (Gas Chromatography/Mass Spectrometry) (Tacke and Casper, 1996) methods that require extensive logistics that are time consuming and labor intensive. Obtaining samples for DON analysis requires a FHB nursery with disease pressure, extensive sampling in the field followed by threshing and milling of grain samples. High throughput imaging technologies have been explored

and exploited to improve the overall process and accuracy in phenotyping DON levels in wheat kernels (Ropelewska, 2019; Jaillais et al., 2015; Cambaza et al., 2019; Shi et al., 2020). However, DON phenotyping remains a bottleneck in elucidating the genetic basis of resistance to DON accumulation during FHB infection.

Recently, hyperspectral imaging has been explored to further increase accuracy and intensity in evaluating DON content in barley (Su et al., 2021), oats (Tekle et al., 2015; Teixido-Orries et al., 2023), and wheat (Femenias et al., 2022). At a single kernel resolution, Shen et al. (2022) and Femenias et al. (2022) imaged grain samples using wavebands at the near-infrared (NIR) range to quantify DON. Mobile handheld hyperspectral cameras like the Specim IQ (Specim, Oulo, Finland) detect reflectance values at wavebands from the visible to near infrared (VIS/NIRS) regions (Behman et al., 2018) and have been used for disease detection of root rot in grapevine (Calamita et al., 2021), powdery mildew in wild rocket (Pane et al., 2021), and root and crown rot in sugar beet (Barreto et al., 2020).

Phenomics and imaging technologies have been integrated with GWAS to elucidate the genetic architecture of quantitative traits (Xiao et al., 2022). Several studies have reported the integration of phenomics and high-throughput phenotyping for GWAS in wheat (Jiang et al., 2019; Rasheed et al., 2014; Yates et al., 2019), rice (Barnaby et al. 2020; Feng et al., 2017; Sun et al., 2019), soybean (Herritt et al., 2016; Dhanapal et al., 2016; Xavier et al., 2017), and maize (Muraya et al., 2017; Gage et al., 2018; Wang et al., 2019). While the genetic basis of hyperspectral imaging-derived phenotypes has been investigated in rice (Feng et al., 2017; Barnaby et al., 2020), and soybean (Wang et al., 2021; Yoosefzadeh-Najafabadi et al., 2021), great potential exists to leverage imaging technologies to investigate the genetic basis of quantitative traits.



This study leverages hyperspectral imaging in the identification of genomic regions associated with DON accumulation in soft winter wheat adapted to the Eastern United States. DON-infected wheat kernels of diverse wheat varieties and elite breeding lines were imaged using the Specim IQ handheld hyperspectral imaging system. The hyperspectral reflectance values generated were used to 1) determine the relationship of the hyperspectral phenome with DON mycotoxin levels in wheat kernels and 2) identify genomic regions associated with mycotoxin levels and variation in the hyperspectral phenome of DON-infected kernels.

## **2.2. Materials and Methods**

### *2.2.1. Plant materials*

A set of 200 soft red and 114 soft white winter wheat genotypes (n=314), comprised of advanced breeding lines and commercial varieties (Supplementary Table 3.1) were used in this study. Genotypes MI14W0190 and Ambassador were considered FHB-resistant checks, low DON and FHB-susceptible, high DON checks, respectively. Wheat genotypes were planted in a misted and inoculated Fusarium screening nursery in East Lansing, MI (°42.69 N, °84.48W, Elevation: 264 m) in one-meter rows using a completely randomized designed with two to four replicates per genotype.

### *2.2.2. Fusarium inoculum*

*Fusarium graminearum* cultures were collected in 2020 from Huron, Ingham, Monroe, Tuscola and Sanilac counties in Michigan, USA. Initial cultures were grown by placing infected seeds in Nash-Snyder Media for 5 to 7 days at room temperature. Isolates for field inoculation were cultured in spawn bags with a 0.2-micron filter patch (Unicorn Bags, TX, USA) containing 1.5 kg corn kernels. Corn was soaked in deionized water for 24 to 48 hours and autoclaved for 90 minutes three times in succession. One culture plate of a four-to-six-day-old culture and 100 ml autoclaved

deionized water were added to each spawn bag. Cultures developed over two to three weeks and were dried in biohazard hood for 48 hours at ambient temperature. Isolates from different locations were cultured separately. After drying, *F. graminearum* grain spawn cultures from the five locations were pooled in equal proportions by weight prior to inoculation. Field inoculation was carried out five times beginning at approximately 5 weeks prior to flowering. A misting system was run throughout the nursery for 10 minutes every hour for 12 hours, 6 am to 6 pm, to promote infection and disease development.

### *2.2.3. Deoxynivalenol evaluation*

Wheat heads from the middle 0.3 meter of each row were sampled separately and harvested by hand. The heads from each row were threshed together and all seeds were retained. A subsample of 10 grams from each row was ball-milled using Restch MM 400 miller (Retsch, PA, USA) to generate flour meeting the guidelines set by the US Wheat and Barley Scab Initiative (USWBI) ([https://scabusa.org/don\\_labs\\_umn\\_testinglab\\_protocol](https://scabusa.org/don_labs_umn_testinglab_protocol)). Deoxynivalenol concentration of flour samples was determined using GC/MS at the Department of Plant Pathology, University of Minnesota.

### *2.2.4. Hyperspectral image acquisition*

FHB-infected wheat kernels from each genotype sent for DON content measurement were imaged using, a handheld, push broom hyperspectral camera, Specim IQ (Specim, Oulo, Finland). A sample of 50 to 80 wheat kernels from each replicate of each genotype were imaged. Seeds were placed against a black background side-by-side with the white reference panel. Imaging was done inside a 51 x 51 x 51-centimeter light box (Finnhomy, USA) using the attached LED light source. The hyperspectral camera was mounted on a tripod and angled 45° facing downward over the

kernels. Default Recording Mode was used to capture reflectance values from 204 wavebands from 397 nm to 1004 nm with an integration time of 30 to 40 seconds and focus set at automatic.

#### *2.2.5. Image processing and reflectance value extraction*

Hyperspectral images were processed using QGIS 3.10.2 (QGIS, 2020). Image files (.dat) were imported as raster layer. Rendering was carried out using multiband color with Band 088 (651.92 nm) as Red Band, Band 057 (560.30 nm) as Green Band, and Band 037 (501.72) as Blue Band. Color enhancements were set at Stretch to MinMax and normal blending mode. Raster calculation was carried out at 0.3 to 0.8 threshold. Raster calculated images were saved as GeoTIFF (.tif) file and converted to vector image (polygonize) using default settings. To determine region of interest (wheat kernels) and remove unnecessary features, toggle editing by selecting features was used. Vectorized images with region of interest determined were saved as ESRI Shape File (.shp). Spectral reflectance values were extracted from each ESRI shape file using “raster” package (Hijmans et al., 2023) in R v4.2.2 (R Core Team, 2021) by calculating mean reflectance values in each waveband.

#### *2.2.6. Statistical analysis*

The normality of hyperspectral reflectance data was assessed using Shapiro-Wilk Test and variance homogeneity assumption was carried out using Levene’s Test. Wavebands with p-values <0.05 failed to meet normality and homogeneity assumption. To test variation among the wheat genotypes for DON content and spectral reflectance values in all 204 wavebands, ANOVA was carried out for reflectance values at wavebands meeting normality and homogeneity assumptions, otherwise non-parametric Kruskal-Wallis Rank Test was employed following the model:

$$(1) y = G + e$$

Where  $y$  is the DON content or spectral reflectance value of each waveband,  $G$  is the fixed effect of genotype, and  $e$  is the residual. Shapiro-Wilk Test, Levene's Test, ANOVA F-Test, and Kruskal-Wallis Rank Test were carried out in R v4.2.2 (R Core Team, 2021).

Means of DON content and spectral reflectance values for two to four replicates per genotype were calculated using the “emmeans” package (Lenth et al., 2018). Pearson's Correlation Coefficient was computed between DON means and spectral reflectance values at all individual wavebands.

#### *2.2.7. Principal Component Analysis (PCA) of wavebands*

PCA was carried out using reflectance values for all wavebands to dimensionally reduce the spectral data and identify wavebands potentially associated with DON. To evaluate the contribution of waveband ranges across the hyperspectral phenome of DON infected wheat kernels, we employed a “Sliding Window” approach where the first twenty wavebands were binned and subjected to PCA. The bin was then “slid” at five-waveband intervals and PCs were generated for the next twenty wavebands to 1004 nm for a total of 38 windows (binned wavebands). The resulting PC1 waveband reflectance values from the 38 windows were then correlated to the GC/MS-derived DON content and used as predictors of DON. PCA and correlation was carried out in R v4.2.2 (R Core Team, 2021).

#### *2.2.8. DNA isolation and genotyping*

Tissue was collected from all genotypes evaluated and DNA was isolated according to Wiersma et al. (2016). Genotyping-by-sequencing libraries were prepared according to Poland et al. (2012) scaled to a 24uL volume in 384-well format. Libraries were sequenced at 384-plex on an Illumina HiSeq 4000 instrument. SNPs were called using the TASSEL 5 GBS pipeline (Glaubitz et al., 2014). Reads were aligned to the RefSeq v1.0 wheat reference genome assembly

(International Wheat Genome Sequencing Consortium) using default parameters. For the GBSSeqToTagDBPlugin and ProductionSNPCallerPluginV2 steps, the k-mer length was set to 64 base pairs and a minimum coverage of five reads was required for each k-mer. Default settings were used for all other steps. SNPs were initially called using all families and parents. SNPs were subsequently filtered for 0.85 call rate and 0.05 minor allele frequency (MAF).

#### *2.2.8. Genome wide association mapping*

Phenotypes for GWAS included: 1) GC/MS-derived DON content, 2) PC1 of all 204 wavebands and 3) PC1 of 38 waveband bins from the “Sliding Window” approach. GWAS was carried out using the Bayesian-information and Linkage Disequilibrium Iteratively Nested Keyway (BLINK) (Huang et al., 2019) model in GAPIT v3 (Genomic Association and Prediction Integrated Tool) (Lipka et al., 2012; latest version: March 12, 2022). A total of 9,961 SNPs across all 21 chromosomes remained after filtering at  $MAF < 0.05$  and 0.85 call rate. To address potential population structure, three principal components were used in GWAS models with the exception of two principal components for one phenotypic input, and four principal components for four phenotypic inputs (Supplementary Table 2.7). LD between MTAs was investigated using TASSEL 5 (Glaubitz et al., 2014).

#### *2.2.9. Candidate gene identification*

Significant SNPs identified in GWAS were assigned to high confidence gene models in IWGSC RefSeq Annotation V1.0 ([www.wheatgenome.org](http://www.wheatgenome.org)). Descriptions of putative candidate genes were derived from the public wheat expression database Triticeae Multi-omics center (<http://202.194.139.32>) (Ma et al., 2021) and Wheat Expression Browser (Ramirez-Gonzalez et al., 2021).

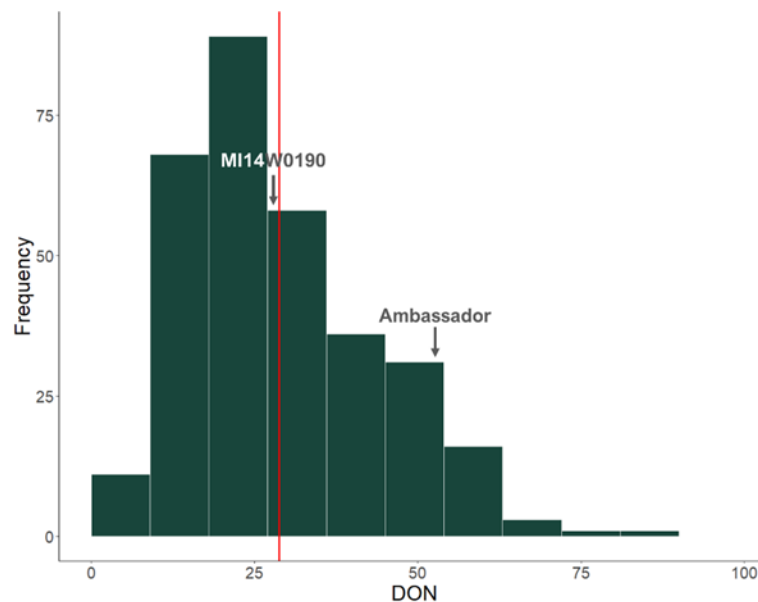
## 2.3. Results

### 2.3.1. Deoxynivalenol concentration

Wheat genotypes showed variation for DON concentration (p-value:  $3.016 \times 10^{-7}$ ) based on non-parametric Kruskal-Wallis Rank Test (Figure 2.1; Supplementary Table 2.2). Pairwise comparison of means revealed 227 genotypes (72.3%) having significantly lower DON content in comparison with susceptible check Ambassador. In comparison 248 genotypes (79.3%) were not significantly different from the resistant check, MI14W0190 and 21 genotypes (6.7%) have significantly lower DON content than MI14W0190.

### 2.3.2. Hyperspectral reflectance values and dimensionality reduction of hyperspectral phenome

Significant variation (p-value:  $< 0.05$ ) among the genotypes based on Kruskal-Wallis Rank Test and ANOVA F-Test for wavebands meeting normality and variance homogeneity assumptions, were observed in each of the 204 wavebands generated (Supplementary Table 2.2; Supplementary

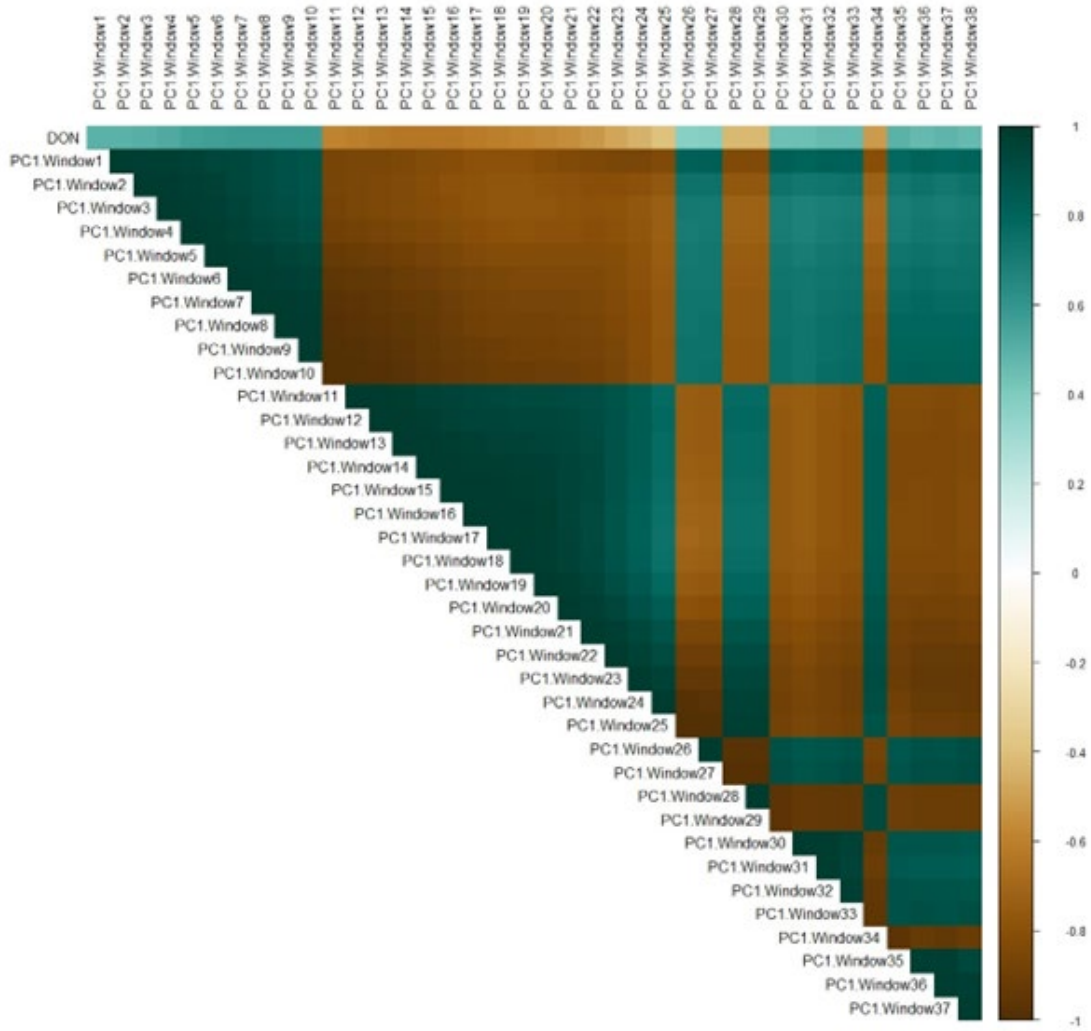


**Figure 2.1.** Phenotyping of soft winter wheat genotypes for DON accumulation. Frequency distribution of GC/MS-derived DON content of the 314 soft winter wheat genotypes (red line represents the population mean DON content).

Figure 2.1). Of the 204 wavebands evaluated, only 15 wavebands (7.35%) met the normality and variance homogeneity assumptions (Supplementary Table 2.2). Significant positive correlations were found between the 204 wavebands generated and DON content, with 97 wavebands in the 455 nm to 739 nm range demonstrating a correlation of greater than 0.5 with DON content (Supplementary Table 2.3). The 584 nm to 673 nm waveband range demonstrated the highest correlations with DON greater than 0.6 (Supplementary Table 2.3).

Principal component analysis (PCA) was performed for all 204 wavebands to dimensionally reduce the hyperspectral phenome of the DON infected wheat kernels. The first two principal components accounted for 74.0% and 6.0% of variation in hyperspectral reflectance values (Supplementary Table 2.6), respectively. Genotypes with lower values in the first principal component (PC1) demonstrated lower DON content and PC1 of all wavebands correlated with DON at 0.57. The waveband at 502 nm demonstrated the highest component loading (0.101) and was also found to be the most discriminatory among the wavebands, followed by 499 nm, 505 nm, 508 nm, and 511 nm (Supplementary Table 2.4). Wavebands in the 397 nm to 780 nm demonstrated higher component loading than most wavebands beyond 780 nm (Supplementary Table 2.4).

PCA was also carried out for binned wavebands using a sliding window approach. The first principal component for each of the 38 binned wavebands (windows) explained 51% to 99% of the variation in hyperspectral reflectance values within each bin (Supplementary Table 2.6). PC1 of each waveband bin was found to be significantly correlated with DON content (Figure 2.2, Supplementary Table 2.7). Windows 1 to 10 spanning wavebands 397 nm to 584 nm demonstrated significant positive correlations of 0.50 to 0.58 with GC/MS-derived DON content, while Windows 11 to 25 spanning wavebands 542 nm to 810 nm demonstrated a significant negative



**Figure 2.2.** Heatmap of correlations among PC1 between 38 sliding windows (binned wavebands) and DON content.

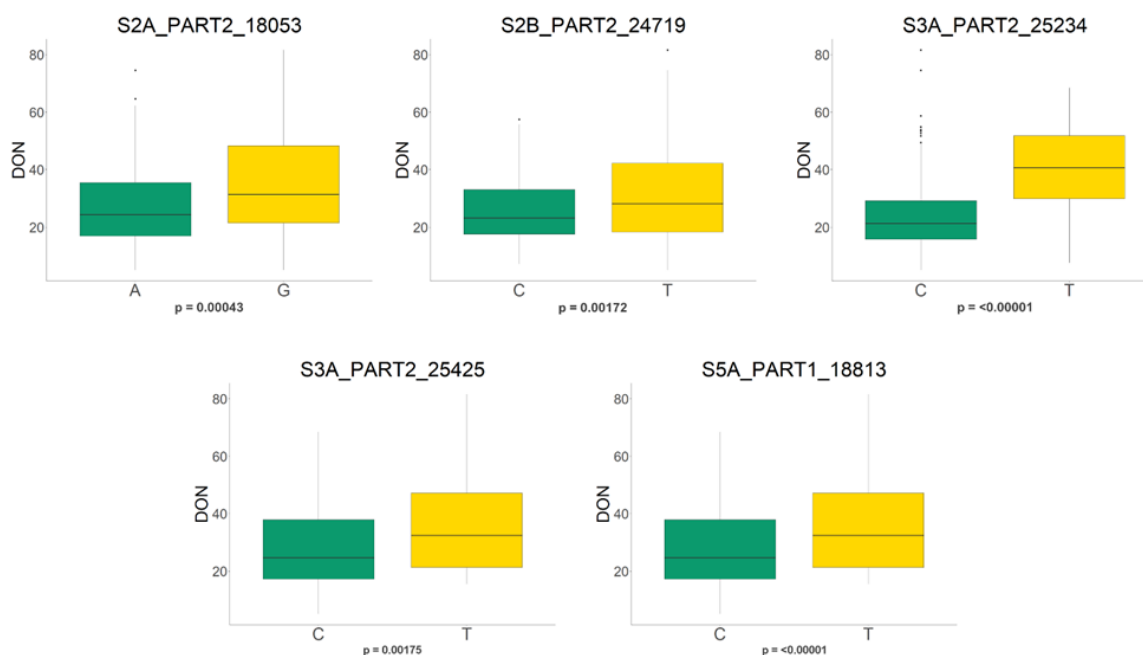
correlation from -0.40 to -0.64 (Figure 2.2; Supplementary Table 2.7). The sign of the correlation between waveband window PC1 values and DON inverted six times across the waveband spectrum (Figure 2.2). The correlation of PC1 with DON inverted from positive in Window 1 to 10 to negative from Windows 11 to 25. Correlation with DON inverted again from Windows 25 to 26, 27 to 28, 29 to 30 and 33 to 34.



**Table 2.1.** Marker-trait associations identified using GC/MS-derived DON content as phenotypic input influencing DON accumulation.

Significant SNP (MTA)	Chromosome	Position (mb)	Allele*	p.value	Alleles Effect	PVE (%)
S2A_PART2_18053	2A	642.9	<b>A/G</b>	<b>6.11 x 10<sup>-8</sup></b>	3.77	3.85
S2B_PART2_24719	2B	700.4	<b>T/C</b>	<b>6.68 x 10<sup>-7</sup></b>	2.47	1.39
S3A_PART2_25234	3A	706.4	<b>C/T</b>	<b>6.32 x 10<sup>-7</sup></b>	3.64	2.65
S3A_PART2_25425	3A	708.4	<b>C/T</b>	<b>9.82 x 10<sup>-7</sup></b>	4.24	4.03
S5A_PART1_18813	5A	18.81	<b>C/T</b>	<b>4.65 x 10<sup>-7</sup></b>	2.88	1.96

\*Alleles in bold reduce DON accumulation; PVE = Phenotypic Variance Explained.



**Figure 2.3.** Alleles reducing GC/MS-derived DON identified by GWAS.

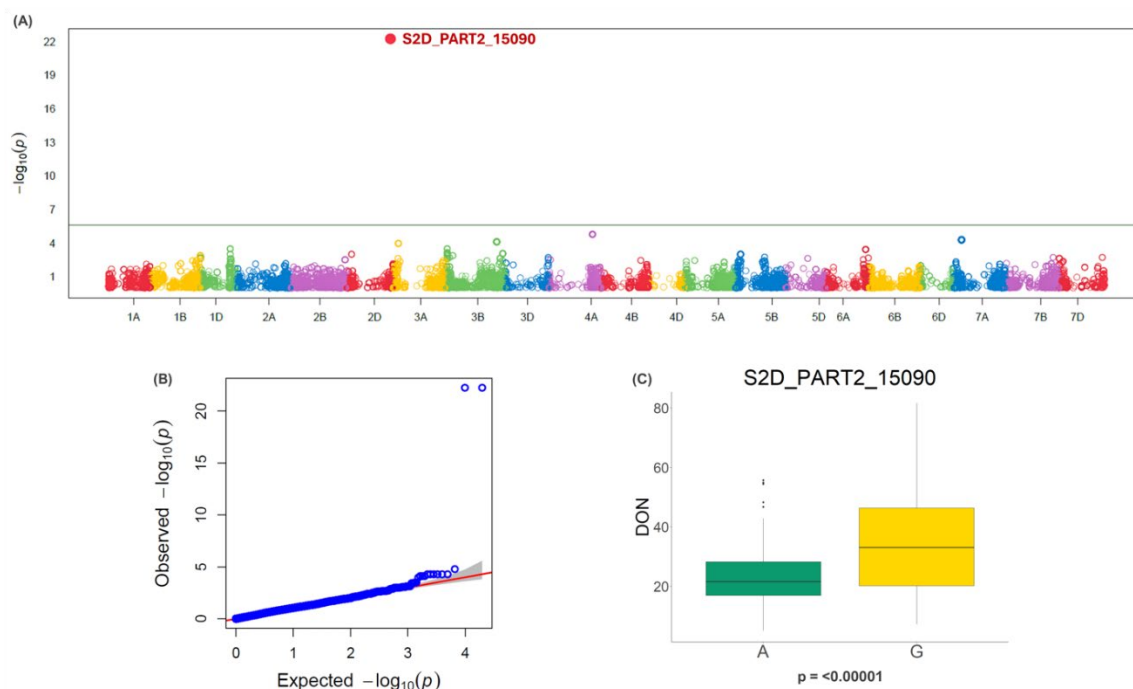
### 2.3.3. Genome wide association study for deoxynivalenol accumulation

GWAS was performed to identify marker trait associations (MTAs) associated with DON content (Supplementary Table 2.8). Five significant MTAs were identified in chromosomes 2A, 2B, 3A, and 5A explaining 1.96% to 4.03% of the variation in DON (Table 2.1, Supplementary

Table 2.8, Supplementary File 2.1). Favorable alleles at all five loci demonstrated a reduction in DON content (Figure 2.3).

GWAS using PC1 of all 204 wavebands identified a single locus (S2D\_PART2\_15090) at 613.12 Mb on chromosome 2D using PC1 that explained 26.4% of the phenotypic variation in the hyperspectral phenome of DON-infected wheat kernels (Figure 2.4a and 2.4b). Genotypes carrying the A allele demonstrate a 6.3 ppm reduction in DON compared with genotypes bearing the G allele (Figure 2.4c, Supplementary Table 2.8). The A allele at the 2D locus is the minor allele with a high frequency of 0.47.

GWAS using the PC1 value for all 38 waveband bins consistently identifies the 2D locus in 35 of the 38 waveband bins, explaining 6.4% to 28.3% of the phenotypic variation in PC1. The S2D\_PART2\_15090 single nucleotide polymorphism (SNP) demonstrates a negative allele effect



**Figure 2.4.** Genome wide association using PC1 of the 204 wavebands generated using hyperspectral imaging. (a) Manhattan plot identifying the MTA on chromosome 2DL and (b) its corresponding quantile-quantile plot, (c) variation in actual DON content of genotypes carrying the A and G allele in SNP S2D\_PART2\_15090.

from Windows 1 to 10 (397 nm to 584 nm) and a positive allele effect from Windows 11 to 25 (543 nm to 811 nm). The inversion of allele effect is consistent with the observed change in sign in the correlation between waveband bin PC1 values and DON content.

An additional 18 MTAs were identified across the hyperspectral phenome of DON infected wheat kernels on chromosomes 1A, 1B, 1D, 2B, 2D, 3A, 3B, 4A, 4B, 7A, 7B and 7D (Supplementary Table 2.8, Supplementary File 2.1) across different waveband ranges. A locus on 1B was identified from 396 nm to 467 nm explaining 18.8% and 17.6% of variation in PC1 of waveband reflectance values and reducing DON by 2.0 ppm and 2.2 ppm at waveband bins 1 and 2, respectively (Supplementary Table 2.8). Other MTAs explained 0.2% to 7.6% of variation in PC1 of reflectance values within individual waveband bins.

Several SNPs significantly associated with waveband bin PC1 values were found to be in linkage disequilibrium (LD) (Supplementary Table 2.9). On chromosome 2D, the SNPs S2D\_PART2\_13700 and S2D\_PART2\_15090, were found to be in high LD with  $r^2 = 0.88$  at substantial distance of 13.9 mb (Supplementary Table 2.9). Weaker LD was detected between SNPs S2B\_PART2\_28124 and S2B\_PART2\_27654 on chromosome 2B at a distance of 4.69 mb ( $r^2 = 0.72$ ), and between SNPs S1B\_PART2\_24837 and S1B\_PART2\_24652 on chromosome 1B (distance: 1.85 mb) ( $r^2=0.25$ ).

#### 2.3.4. Putative candidate gene identified on chromosome 2D

The SNP identified on 2D associated with PC1 of the entire hyperspectral phenome of DON infected wheat kernels and PC1 of nearly all sliding windows, S2D\_PART2\_15090, is located in the single exon of a 1,104 kb gene, *TraesCS2D02G524600*, coding for a protein with an F-box domain. *TraesCS2D02G524600* is upregulated in the spikelets and rachis in response to inoculation with *F. graminearum* in near-isogenic lines (NILs) bearing a 2DL introgression conferring FHB

resistance (Biselli et al. 2018) (Supplementary Figure 2.5) and has been implicated in *Fhb1* resistance with higher expression in NILs carrying *Fhb1* (Ma et al., 2021) (Supplementary Figure 2.6). The expression of *TraesCS2D02G524600* and 17 adjacent genes within a 2Mb region, 1Mb upstream and downstream, was investigated across tissues and developmental stages under *F. graminearum* infection (Wheat Expression Browser, Ramirez-Gonzalez et al., 2021) (Supplementary Figure 2.7). *TraesCS2D02G524600* is inducible by infection with *F. graminearum* and highly expressed in the spike at the reproductive stage, which is consistent across gene expression data sets. Only one gene in the 2Mb interval, *TraesCS2D02G524400*, located upstream of *TraesCS2D02G524600*, demonstrates the exact same expression profile. *TraesCS2D02G524600* is a likely candidate gene for the large effect locus on 2D that reduces DON accumulation in wheat kernels during infection by *F. graminearum*.

## **2.4. Discussion**

Breeding for resistance to FHB requires evaluation of the multiple components of resistance (Steiner et al., 2017). Each FHB resistance component has a different relationship to DON (Buerstmayr and Lemmens, 2015; Mesterhazy et al., 2015; Paul et al., 2005) and evaluating multiple traits can lead to better selection decisions in breeding. Visual observations of FHB severity and incidence are used to develop an overall visual FHB index (Steiner et al., 2017). The proportion of Fusarium damaged kernels can be estimated on samples of infected grain and a high correlation with DON has been demonstrated for this resistance component (Mesterhazy et al., 2015).

In this study, we generate multiple visual FHB resistance phenotypes using the hyperspectral phenome of DON infected wheat kernels. Reflectance values at individual wavebands can be considered unique phenotypes and high correlations were found between DON

and reflectance values at individual wavebands, especially at the visible light spectrum range. PC1 of the reflectance values from all wavebands compresses the entire hyperspectral phenome into a single phenotype that incorporates the information from all wavebands. The hyperspectral phenome was dissected further into 38 sliding windows and PC1 of each window was used as a separate phenotype that is correlated with DON.

In this study, we identified five MTAs on chromosomes 2A, 2B, 3A, and 5A that co-localize with previously reported genomic regions conferring resistance to DON (Supplementary Table 2.10). Individually, each of the MTAs identified for DON content explain only 1% to 4% of the variation in DON and reduce DON by 2.5 ppm to 4.2 ppm. FHB resistance traits can differ in their genetic architecture. Developing the hyperspectral phenome into a novel FHB resistance phenotype using PC1 of all wavebands, which is correlated to DON, led to identification of a comparably large effect locus on 2D explaining 26% of the variation in PC1 and reducing DON by 6.8 ppm. While several genomic regions were identified across waveband ranges, the 2D locus was identified using PC1, further establishing its association with the hyperspectral phenome of DON infected wheat kernels. By leveraging the hyperspectral phenome as a correlated trait, we were able to identify a locus influencing the target trait, DON mycotoxin content.

The large effect locus on 2D localizes to the exon of an F-box protein encoding gene and captures a large proportion of variation in the hyperspectral phenome of DON infected wheat kernels. Two SNPs were identified in the exon of this gene; however, the BLINK algorithm removes SNPs that are in LD. Multiple gene expression studies (Biselli et al. 2018; Schweiger et al. 2016; Ma et al. 2021) demonstrate the gene *TraesCS2D02G524600* is inducible upon infection with *F. graminearum* and is expressed exclusively in tissues of the spike and rachis as a component

of the defense response. It may be possible to select for the DON-reducing allele at 2D locus in a breeding context using a Kompetitive Allele Specific PCR (KASP) marker assay (He et al., 2014).

Evaluation of DON production during infection by *F. graminearum* is an integral part in developing wheat varieties with resistance to FHB, which has long been done using GC/MS (Tacke and Casper, 1996). Preparation and phenotyping of DON infected wheat kernel samples is time consuming, tedious, and labor intensive (Steiner et al., 2017). This study demonstrates that hyperspectral imaging can reduce the amount of time and physical resources necessary to make selection decisions in breeding for lower DON.

## REFERENCES

- Bai, G., Su, Z. & Cai, J. Wheat resistance to Fusarium head blight. *Canadian Journal of Plant Pathology* **40**, 336–346 (2018).
- Barnaby, J. Y. *et al.* Vis/NIR hyperspectral imaging distinguishes sub-population, production environment, and physicochemical grain properties in rice. *Sci Rep* **10**, 9284 (2020).
- Barreto, A., Paulus, S., Varrelmann, M. & Mahlein, A.-K. Hyperspectral imaging of symptoms induced by *Rhizoctonia solani* in sugar beet: comparison of input data and different machine learning algorithms. *J Plant Dis Prot* **127**, 441–451 (2020).
- Behmann, J. *et al.* Specim IQ: Evaluation of a New, Miniaturized Handheld Hyperspectral Camera and Its Application for Plant Phenotyping and Disease Detection. *Sensors* **18**, 441 (2018).
- Biselli, C. *et al.* Comparative Transcriptome Profiles of Near-Isogenic Hexaploid Wheat Lines Differing for Effective Alleles at the 2DL FHB Resistance QTL. *Front. Plant Sci.* **9**, 37 (2018).
- Buerstmayr, H. & Lemmens, M. Breeding healthy cereals: genetic improvement of Fusarium resistance and consequences for mycotoxins. *World Mycotoxin Journal* **8**, 591–602 (2015).
- Calamita, F., Imran, H. A., Vescovo, L., Mekhalfi, M. L. & La Porta, N. Early Identification of Root Rot Disease by Using Hyperspectral Reflectance: The Case of Pathosystem Grapevine/Armillaria. *Remote Sensing* **13**, 2436 (2021).
- Cambaza, E., Koseki, S. & Kawamura, S. Why RGB Imaging Should be Used to Analyze *Fusarium graminearum* Growth and Estimate Deoxynivalenol Contamination. *MPs* **2**, 25 (2019).
- da Silva, C. *et al.* QTL mapping of Fusarium head blight resistance and deoxynivalenol accumulation in the Kansas wheat variety ‘Everest’. *Mol Breeding* **39**, 35 (2019).
- Dhanapal, A. P. *et al.* Genome-wide association mapping of soybean chlorophyll traits based on canopy spectral reflectance and leaf extracts. *BMC Plant Biol* **16**, 174 (2016).
- Femenias, A., Llorens-Serentill, E., Ramos, A. J., Sanchis, V. & Marín, S. Near-infrared hyperspectral imaging evaluation of Fusarium damage and DON in single wheat kernels. *Food Control* **142**, 109239 (2022).
- Feng, H. *et al.* An integrated hyperspectral imaging and genome-wide association analysis platform provides spectral and genetic insights into the natural variation in rice. *Sci Rep* **7**, 4401 (2017).
- Foroud, N. A. *et al.* Trichothecenes in Cereal Grains – An Update. *Toxins* **11**, 634 (2019).

- Gage, J. L., De Leon, N. & Clayton, M. K. Comparing Genome-Wide Association Study Results from Different Measurements of an Underlying Phenotype. *G3 Genes|Genomes|Genetics* **8**, 3715–3722 (2018).
- Glaubitz, J. C. *et al.* TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLOS ONE* **9**, e90346 (2014).
- Haile, J. K. *et al.* Multi-locus genome-wide association studies reveal the genetic architecture of Fusarium head blight resistance in durum wheat. *Front. Plant Sci.* **14**, 1182548 (2023).
- He, C., Holme, J. & Anthony, J. SNP Genotyping: The KASP Assay. in Crop Breeding (eds. Fleury, D. & Whitford, R.) vol. 1145, p. 75–86 (Springer New York, New York, NY, 2014).
- He, X., Dreisigacker, S., Singh, R. P. & Singh, P. K. Genetics for low correlation between Fusarium head blight disease and deoxynivalenol (DON) content in a bread wheat mapping population. *Theor Appl Genet* **132**, 2401–2411 (2019).
- Herritt, M., Dhanapal, A. P. & Fritschi, F. B. Identification of Genomic Loci Associated with the Photochemical Reflectance Index by Genome-Wide Association Study in Soybean. *The Plant Genome* **9**, plantgenome2015.08.0072 (2016).
- Hijmans, R. J. *et al.* raster: Geographic Data Analysis and Modeling. (2023).
- Huang, M., Liu, X., Zhou, Y., Summers, R. M. & Zhang, Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8**, (2019).
- Jaillais, B., Roumet, P., Pinson-Gadais, L. & Bertrand, D. Detection of Fusarium head blight contamination in wheat kernels by multivariate imaging. *Food Control* **54**, 250–258 (2015).
- Jiang, L. *et al.* Functional mapping of N deficiency-induced response in wheat yield-component traits by implementing high-throughput phenotyping. *The Plant Journal* **97**, 1105–1119 (2019).
- Larkin, D. L. *et al.* Genome-wide analysis and prediction of Fusarium head blight resistance in soft red winter wheat. *Crop Science* **60**, 2882–2900 (2020).
- Lenth, R. V. *et al.* emmeans: Estimated Marginal Means, aka Least-Squares Means. (2024).
- Lipka, A. E. *et al.* GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399 (2012).
- Ma, S. *et al.* WheatOmics: A platform combining multiple omics data to accelerate functional genomics studies in wheat. *Mol Plant* **14**, 1965–1968 (2021).



- Mesterházy, A. *et al.* Breeding for FHB Resistance via Fusarium Damaged Kernels and Deoxynivalenol Accumulation as Well as Inoculation Methods in Winter Wheat. *Agricultural Sciences* **06**, 970 (2015).
- Mesterházy, Á. Breeding for Resistance to Fusarium Head Blight in Wheat. in *Mycotoxin Reduction in Grain Chains* p. 189–208 (John Wiley & Sons, Ltd, 2014). doi:10.1002/9781118832790.ch13 (2014).
- Mirocha, C. J. *et al.* Production of trichothecene mycotoxins by *Fusarium graminearum* and *Fusarium culmorum* on barley and wheat. *Mycopathologia* **128**, 19–23 (1994).
- Muraya, M. M. *et al.* Genetic variation of growth dynamics in maize (*Zea mays* L.) revealed through automated non-invasive phenotyping. *The Plant Journal* **89**, 366–380 (2017).
- Pane, C., Manganiello, G., Nicastro, N., Cardi, T. & Carotenuto, F. Powdery Mildew Caused by *Erysiphe cruciferarum* on Wild Rocket (*Diplotaxis tenuifolia*): Hyperspectral Imaging and Machine Learning Modeling for Non-Destructive Disease Detection. *Agriculture* **11**, 337 (2021).
- Paul, P. A., Lipps, P. E. & Madden, L. V. Relationship Between Visual Estimates of Fusarium Head Blight Intensity and Deoxynivalenol Accumulation in Harvested Wheat Grain: A Meta-Analysis. *Phytopathology* **95**, 1225–1236 (2005).
- Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J.-L. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE* **7**, e32253 (2012).
- R Development Core Team. R: A language and environment for statistical computing. (2021).
- Ramírez-González, R. H. *et al.* The transcriptional landscape of polyploid wheat. *Science* **361**, eaar6089 (2018).
- Rasheed, A. *et al.* Genome-wide association for grain morphology in synthetic hexaploid wheats using digital imaging analysis. *BMC Plant Biol* **14**, 128 (2014).
- Ropelewska, E. Post-harvest assessment of wheat and barley kernel infections with fungi of the genus *Fusarium* using thermal analysis. *Journal of Stored Products Research* **83**, 61–65 (2019).
- Schweiger, W. *et al.* Suppressed recombination and unique candidate genes in the divergent haplotype encoding Fhb1, a major Fusarium head blight resistance locus in wheat. *Theor Appl Genet* **129**, 1607–1623 (2016).
- Shen, G. *et al.* Rapid and nondestructive quantification of deoxynivalenol in individual wheat kernels using near-infrared hyperspectral imaging and chemometrics. *Food Control* **131**, 108420 (2022).

- Shi, Y., Liu, W., Zhao, P., Liu, C. & Zheng, L. Rapid and nondestructive determination of deoxynivalenol (DON) content in wheat using multispectral imaging (MSI) technology with chemometric methods. *Anal. Methods* **12**, 3390–3396 (2020).
- Steiner, B. *et al.* Breeding strategies and advances in line selection for Fusarium head blight resistance in wheat. *Trop. plant pathol.* **42**, 165–174 (2017).
- Su, W.-H. *et al.* Hyperspectral imaging and improved feature variable selection for automated determination of deoxynivalenol in various genetic lines of barley kernels for resistance screening. *Food Chemistry* **343**, 128507 (2021).
- Sun, D. *et al.* Using hyperspectral analysis as a potential high throughput phenotyping tool in GWAS for protein content of rice quality. *Plant Methods* **15**, 54 (2019).
- Tacke, B. K. & Casper, H. H. Determination of deoxynivalenol in wheat, barley, and malt by column cleanup and gas chromatography with electron capture detection. *J AOAC Int* **79**, 472–475 (1996).
- Teixido-Orries, I., Molino, F., Femenias, A., Ramos, A. J. & Marín, S. Quantification and classification of deoxynivalenol-contaminated oat samples by near-infrared hyperspectral imaging. *Food Chemistry* **417**, 135924 (2023).
- Tekle, S., Mage, I., Segtnan, V. H. & Bjornstad, Å. Near-Infrared Hyperspectral Imaging of Fusarium -Damaged Oats (*Avena sativa* L.). *Cereal Chem* **92**, 73–80 (2015).
- Wang, L. *et al.* Identification of the QTL-allele System Underlying Two High-Throughput Physiological Traits in the Chinese Soybean Germplasm Population. *Frontiers in Genetics* **12**, (2021).
- Wang, X. *et al.* Dynamic plant height QTL revealed in maize through remote sensing phenotyping using a high-throughput unmanned aerial vehicle (UAV). *Sci Rep* **9**, 3458 (2019).
- Wiersma, A. T. *et al.* Fine mapping of the stem rust resistance gene SrTA10187. *Theor Appl Genet* **129**, 2369–2378 (2016).
- Xavier, A., Hall, B., Hearst, A. A., Cherkauer, K. A. & Rainey, K. M. Genetic Architecture of Phenomic-Enabled Canopy Coverage in Glycine max. *Genetics* **206**, 1081–1089 (2017).
- Xiao, Q., Bai, X., Zhang, C. & He, Y. Advanced high-throughput plant phenotyping techniques for genome-wide association studies: A review. *Journal of Advanced Research* **35**, 215–230 (2022).
- Yates, S. *et al.* Precision Phenotyping Reveals Novel Loci for Quantitative Resistance to Septoria Tritici Blotch. *Plant Phenomics* **2019**, 2019/3285904 (2019).

Yoosefzadeh-Najafabadi, M., Torabi, S., Tulpan, D., Rajcan, I. & Eskandari, M. Genome-Wide Association Studies of Soybean Yield-Related Hyperspectral Reflectance Bands Using Machine Learning-Mediated Data Integration Methods. *Frontiers in Plant Science* **12**, (2021).

### **CHAPTER III: BLENDED GENOMIC AND HYPERSPECTRAL IMAGING-BASED PREDICTIONS ENABLE SELECTION FOR REDUCED DEOXYNIVALENOL CONTENT IN WHEAT GRAINS**

*[Manuscript Submitted: G3 (G3-2024-405605, Reviewed, revisions made)]*

#### **Abstract**

Breeding for low deoxynivalenol (DON) mycotoxin content in wheat is challenging due to the complexity of the trait and phenotyping limitations. Since phenomic prediction relies on non-additive effects and genomic prediction on additive effects, their complementarity can improve selection accuracy. In this study DON-infected wheat kernels were imaged using a hyperspectral camera to generate reflectance values across the spectrum of visible and near infrared light that were used in phenomic predictions. Five Bayesian generalized linear regression models, and two machine learning models were trained using phenomic and genomic predictions from advanced soft winter wheat breeding lines evaluated in 2021 and 2022. Across all training sets and models, phenomic predictions using wavebands in the visible light spectrum (400-700 nm) had higher predictive ability than genomic predictions or phenomic predictions using the full waveband range (400-1000 nm). Forward prediction using 2021 trial, 2022 trial, and combined trials as training set was performed using model blending on two sets of F<sub>4:5</sub> selection candidates evaluated independently in 2022 and 2023. The phenotypic and genetic correlations, as well as indirect selection accuracies, of the model averages of phenomic predictions and combined phenomic and genomic predictions were higher than genomic predictions alone. Accuracies depended on the combination of training set and selection candidates. Unsupervised K-Means clustering using the blended predicted values partitioned selection candidates into two groups with high and low mean observed DON content. This study demonstrates the potential of hyperspectral imaging-based

phenomic prediction to complement genomic prediction and highlights considerations for prediction-based selection of low DON in wheat.

### **3.1. Introduction**

Breeding for complex traits like reduced Deoxynivalenol (DON) mycotoxin accumulation, known as Type III resistance to Fusarium Head Blight (FHB), presents a significant challenge in wheat breeding (Bai et al., 2018). DON is commonly measured using gas chromatography/mass spectrometry (GC/MS) (Tacke and Casper, 1996), but this method is often time-consuming and labor-intensive due to the nature of sample preparation (Steiner et al., 2017). For breeding programs that outsource GC/MS analysis, the time required to generate DON data can extend the time frames for selection, potentially impacting decision making and the length of breeding cycles. To address these limitations, researchers have explored alternative approaches for DON measurement to reduce reliance on GC/MS (Buerstmayr et al., 2019, Alisaac and Mahlein, 2023).

Since its introduction, genomic prediction has been used extensively to improve complex traits (Haile et al., 2021, Shahi et al., 2022), including FHB resistance (Rutkoski et al., 2012), with moderate to high prediction accuracies. Genomic prediction enables the estimation of an individual's genetic breeding value based on genomic information, often using Single Nucleotide Polymorphisms (SNPs) (Heffner et al., 2009). One advantage of genomic prediction is enabling early and accurate predictions, thereby accelerating the breeding process and increasing selection efficiency, allowing breeders to evaluate a larger number of individuals, and increasing selection intensity. (Meuwissen et al., 2001, Heffner et al., 2009, Crossa et al., 2014) However, environmental effects and genotype-by-environment interactions still need to be considered for the target trait, which remains a weakness of genomic prediction, especially when multiple years and environments are considered (Jackson et al., 2023). While costs per sample have decreased

significantly, genotyping large numbers of individuals still requires substantial financial resources. For breeding programs conducting in-house sampling and DNA extraction, this process can also be time-consuming and labor-intensive.

In recent years, phenomic selection has emerged as a potential alternative to genomic prediction, addressing some of its practical challenges, particularly in terms of implementation (Rincent et al., 2018). The primary advantage of phenomic selection lies in its ability to achieve a comparable, and in some cases higher, level of accuracy than genomic prediction (Rincent et al., 2018; Ronald et al., 2022). Phenomic selection uses high-throughput phenotyping tools, such as near-infrared spectroscopy, to capture detailed phenomic data that reflect the physical and biochemical characteristics of the sample (Rincent et al., 2018). In the case of FHB resistance, several studies have demonstrated the potential application of image-based phenomic platforms, for FHB evaluation (Alisaac et al., 2018; Mahlein et al., 2019; Liang et al., 2024). Despite these advancements, phenomic prediction for DON, specifically in comparison with genomic prediction, has not been extensively explored, leaving room for further research to leverage its full potential in breeding programs targeting DON reduction.

Recently, efforts have been made to integrate information from varied phenotyping platforms to predict and select for traits in breeding programs (Martens et al., 2015; Rutkoski et al., 2016; Crain et al., 2018; Galan et al., 2020; Togninalli et al., 2023; Adak et al., 2023; Roth et al., 2023), with many using drone-based platforms to evaluate agronomic traits. Thapa et al. (2024) recently demonstrated the potential integration of hyperspectral imaging with genomic information to predict FHB-related traits, including DON accumulation, using deep learning approaches. Their

results demonstrated significant improvements in prediction accuracy compared to genomic prediction alone.

Model ensembling and blending approaches, such as Bayesian model averaging, weighted model averaging, and least squares model averaging, have long been used in fields outside of plant breeding (Raftery et al., 1997; Wasserman, 2000; Hansen, 2007; Magnus et al., 2010). An important contribution to the application of model ensembling in plant breeding was made by Kick and Washburn (2023), revealing that combining linear and non-linear models using genomic, environmental, and crop management data provided better accuracy than base (or individual) models for grain yield.

This study explores the use of hyperspectral imaging-derived spectral reflectance values in the phenomic prediction of DON mycotoxin content in soft winter wheat grains and compares phenomic prediction with genomic prediction. Simple model averaging was explored as a model blending approach to integrate both types of predictions. The utility of phenomic, genomic and blending of both phenomic and genomic predictions models was explored by forward prediction in a separate set of F<sub>4:5</sub> derived selection candidates. Model blending by unsupervised clustering of genotypes based on predicted DON values from various models was explored in a breeding selection scheme.

## **3.2. Materials and Methods**

### *3.2.1. Plant materials for training population and field establishment*

A total of 558 soft winter wheat genotypes (430 soft red and 128 soft white) comprised of advanced breeding lines and commercially released varieties were evaluated for DON mycotoxin content in response to Fusarium Head Blight (FHB) in 2021 (n=307) and 2022 (n=288). The 558 genotypes served as training populations in subsequent analysis. A set of 37 genotypes were

evaluated in both years. Limited overlap between years is due to the use of advanced breeding lines from an active breeding program where new entries are cycled through testing stages each year. The soft white winter wheat genotype, MI14W0190, was the FHB-resistant, low DON check. The soft white winter wheat variety, Ambassador, was the FHB-susceptible, high DON checks. In both years, genotypes were evaluated in a misted and inoculated *Fusarium* screening nursery in East Lansing, Michigan using a completely randomized design with two replicates.

### 3.2.2. *Fusarium* inoculum preparation and field inoculation

*Fusarium graminearum* cultures were collected in 2020 from Huron, Ingham, Monroe, Tuscola, and Sanilac counties and in 2021 from Huron, Ingham, Monroe, and Saginaw counties in Michigan, USA for the 2021 and 2022 trials, respectively. Isolates were grown by placing infected seeds in Nash-Synder Media for 5 to 7 days at room temperature. Approximately 1.5 kg of corn kernels were soaked in deionized water for 24 to 48 hours, placed in spawn bags with 0.2-micron filter patch (Unicorn Bags, TX, USA) and autoclaved three times for 90 minutes. A four-to six-day-old culture and 100 ml autoclaved deionized water were added to each autoclaved spawn bag. Isolates from different locations were cultured separately, and after approximately two weeks, infected grain spawns were dried in a biohazard hood at ambient temperature. The *F. graminearum* grain spawn cultures were pooled in equal proportions by weight prior to inoculation. Field inoculation began five weeks before flowering and grain spawn was applied four to five times. To promote proper condition for infection and disease development, a misting system was run throughout the nursery ten minutes every hour for 12 hours, 6:00 am to 6:00 pm.

### 3.2.3. *Deoxynivalenol* measurement

A sample of infected heads were collected from the middle 0.3 meter of each replicate of each genotype in the *Fusarium* screening nursery. A 10-gram subsample of infected wheat kernels



was ball-milled using Restch MM 400 miller (Retsch, PA, USA). The flour samples were then sent to the Department of Plant Pathology, University of Minnesota for DON concentration measurement using GC/MS.

#### *3.2.4. Hyperspectral imaging and processing*

A sample of 50 to 80 wheat kernels from each replicate of each genotype were placed against a dark background side-by-side a white reference panel and were imaged using a mobile, handheld hyperspectral camera, Specim IQ (Specim, Oulo, Finland). Imaging was done inside a 20 x 20 x 20 inches light box (Finnhomy, USA) using the attached LED light source with the camera mounted on a tripod and angled 45° facing downward over the kernels and the focus set at automatic. To capture the raw reflectance values from 204 wavebands (397 to 1004 nm), the camera was set at Default Recording Mode with an integration time of 30 to 40 seconds. Hyperspectral image files (.dat) were stored in Specim IQ studio and imported as raster layer to QGIS 3.10.2 (QGIS, 2020) for processing. Multiband color was used for rendering, with Band 088 (651.92 nm) as Red Band, Band 057 (560.30 nm) as Green Band, and Band 037 (501.72) as Blue Band. Color enhancements were set at Stretch to MinMax and normal blending mode. Raster calculated images at a 0.3 to 0.8 threshold were saved as GeoTIFF (.tif) file and converted to vector image (Polygonize) using default settings. Unnecessary features were removed using toggle edit to determine the region of interest (wheat kernels) and saved as ESRI shape file (vectorized image) (.shp). From the ESRI shape file, spectral reflectance values in each waveband were extracted using “raster” package (Hijmans and van Etten, 2012) in R v4.2.2 (R Core Team, 2023). Savitzky-Golay filter for smoothing for the generated spectral reflectance values across wavebands were done at window size of 11 and polynomial degree of 3 (3<sup>rd</sup> degree) using “signal” package in R (signal developers, 2023).

### 3.4.5. Statistical analysis

ANOVA F-Tests were carried out to assess the variation in DON content and spectral reflectance values among the genotypes tested in each year. Best Linear Unbiased Estimates (BLUES) of DON content and spectral reflectance values in all wavebands generated for all genotypes in each year were generated using “lsmeans” package (Lenth, 2016) following the model:

$$(1) Y_{ij} = \mu + G_i + \varepsilon_{ij}$$

Where  $Y_{ij}$  is the response of the  $i$ -th genotype in the  $j$ -th observation,  $\mu$  is the overall mean,  $G_i$  is the fixed effect of the  $i$ -th genotype and  $\varepsilon_{ij}$  is the residual error for the  $i$ -th genotype in the  $j$ -th observation.

BLUES of DON content and spectral reflectance values for all wavebands across years were modeled using “lme4” package in R (Bates et al., 2014) following the model:

$$(2) Y_{ijk} = \mu + G_i + T_j + (GT)_{ij} + \varepsilon_{ijk}$$

Where  $Y_{ijk}$  is the response of the  $i$ -th genotype in the  $j$ -th year for the  $k$ -th observation,  $\mu$  is the overall mean across genotypes and years,  $G_i$  is the fixed effect of the  $i$ -th genotype,  $T_j$  is the random effect of  $j$ -th trial,  $(GT)_{ij}$  is the random effect of the interaction between the  $i$ -th genotype and the  $j$ -th trial, and  $\varepsilon_{ijk}$  is the residual error. Mean DON content and spectral reflectance values across years were computed using “lsmeans” package (Lenth, 2016) in R after modeling. Pearson’s correlation coefficient was used to assess the relationship between BLUES of DON content and BLUES of reflectance values at all individual wavebands within and across years.

### 3.2.5. Heritability estimation

Heritability for DON content based on entry mean within years was calculated based on the following equation:

$$(3) H_{Geno}^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_e^2}{r}}$$

Where  $H_{Geno}^2$  is the estimated heritability on an entry mean basis,  $r$  is the number of replicates per genotype (entry),  $\sigma_e^2$  is the error variance (residual mean square), and  $\sigma_G^2$  is the genotype (entry) variance calculated following the equation:

Heritability based on entry mean across years was calculated based on the following equation:

$$(4) H_{GenoT}^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GT}^2}{r} + \frac{\sigma_e^2}{rN}}$$

Where  $H_{GenoT}^2$  is the estimated heritability on an entry mean basis,  $r$  is the number of replicates per genotype (entry),  $N$  is the number of trials,  $\sigma_e^2$  is the error variance (residual mean square), and  $\sigma_G^2$  is the genotype (entry) variance,  $\sigma_{GT}^2$  is the genotype by trial variance.

### 3.2.6. DNA isolation and single nucleotide polymorphisms genotyping

Tissue was collected from all genotypes evaluated and DNA was isolated according to Wiersma et al. (2016). Genotyping-by-sequencing libraries were prepared according to Poland et al. (2012) scaled to a 24uL volume in 384-well format. Libraries were sequenced at 384-plex on an Illumina HiSeq 4000 instrument. Single nucleotide polymorphisms were called using the TASSEL 5 GBS pipeline (Glaubitz et al., 2014). Reads were aligned to the RefSeq v2.0 wheat reference genome assembly (International Wheat Genome Sequencing Consortium) using default parameters. For the GBSSeqToTagDBPlugin and ProductionSNPCallerPluginV2 steps, the k-mer length was set to 64 base pairs and a minimum coverage of five reads was required for each k-mer. Default settings were used for all other steps. SNPs were initially called using all families and parents. SNPs were subsequently filtered for 0.70 call rate and 0.05 minor allele frequency. A total of 15,456 SNPs out of 386,651 SNPs remained after filtering.

### *3.2.7. Univariate genomic and phenomic prediction model assessment*

Two overlapping sets of hyperspectral bands were used in phenomic predictions of DON content, all 204 wavebands from 400 to 1000 nm in the Visible Light and Near Infrared Spectrum (VIS/NIR) and wavebands in the Visible light (VIS) region from 400 to 800 nm. Genomic predictions were made using 15,456 SNPs across the wheat genome. Both phenomic and genomic prediction for DON was assessed using BayesA, BayesB, BayesC, Bayesian-Ridge Regression (BRR), and Bayesian LASSO (BayesL) with 50,000 iterations and BurnIn set at 5000, implemented in the “BGLR” package (Perez and de los Campos, 2014). Ridge Regression Best Linear Unbiased Prediction (RRBLUP) was executed using “rrBLUP” package in R (Endelman, 2011). Extreme Gradient Boosting (XGBoost) was implemented using the “xgboost” package (Chen and Guestrin, 2016) with parameters eta (learning rate) = 0.01, gamma = 0.2, max\_depth (maximum tree depth) = 6, min\_child\_weight (minimum sum of instance weights) = 0.20, subsample = 1, and colsample\_bytree (fraction of features (columns) used to grow each tree) = 1. Random Forrest was executed using the “caret” package (Kuhn, 2008) with default parameters (ntree = 500, nodesize = 5, maxnodes = unlimited). A total of 100 individual five-fold cross validation was carried out with 80% of the genotypes designated in the training set for all the models. Predictive ability of phenomic and genomic prediction for all models were computed as the average Pearson’s correlation (predictive ability) between the actual DON content and predicted values from all 100 cycles in each model.

Genetic correlation as the Pearson correlation coefficient between the marker (or waveband for phenomic prediction)-based estimated breeding values (EBVs) derived from each models tested and the EBVs obtained from best linear unbiased prediction (BLUP) using the realized genomic relationship matrix (G). The realized G matrix was computed as the cross-product of

standardized marker data, and mixed model solutions were obtained using the “mixed.solve” function in the “rrBLUP” package (Endelman, 2011). From here, the genetic variance of the predicted and actual DON content, and its corresponding covariance was used to calculate the genetic correlation following the formula:

$$(5) r_g = \frac{Cov_{XI}}{\sqrt{(Var_X)(Var_I)}}$$

Where  $Cov_{XI}$  is the covariance between the observed DON content ( $X$ ) and the predicted DON content ( $I$ ),  $Var_X$  is the genetic variance of the observed DON content,  $Var_I$  is the genetic variance of the predicted DON content.

Prediction accuracy ( $Acc_{(A)}$ ) was also calculated from each cross-validation cycle following the formula adapted from Wang et al. (2025):

$$(6) Acc_{(A)} = r_g \sqrt{H_A^2}$$

Where  $r_g$  is the genetic correlation between predicted and actual DON content, and  $H_A^2$  is the estimated heritability of actual DON content in the testing set calculated by dividing the estimated genetic variance of the actual DON content in the testing set ( $\sigma^2_{G\_X}$ ) by its estimated phenotypic variance ( $\sigma^2_{P\_X}$ ).

All cross validations, model fitting, heritability estimates, phenotypic and genotypic correlations, and prediction accuracy assessment were carried out in R v4.2.2 (R Core Team, 2023).

### 3.2.7. Feature selection

The most informative wavebands were identified for each of the 2021 and 2022 trials and combined trials (across years) using recursive partitioning and regressive tree model with the “caret” package (Kuhn, 2008) carried out in R v4.2.2 (R Core Team, 2023). Feature importance

score (scaled relative importance percentage) of 100 or closer means higher association with DON content, whereas scores of 0 or closer means no or less association with DON.

### *3.2.8. Forward prediction*

To test the predictive ability of the trained models, a set of 239 F<sub>4:5</sub> generation breeding lines evaluated in 2022 and 217 F<sub>4:5</sub> breeding lines evaluated in 2023 were used as prediction sets of selection candidates. Field establishment, field inoculation, DON measurement, and hyperspectral imaging and processing of selection candidates followed the procedures described for the 2021 and 2022 training sets. To simulate actual breeding scenarios, models trained using 2021 and 2022 trials, and BLUEs generated across years were used to predict the DON content of selection candidates. Forward predictions were carried out separately for phenomic and genomic predictions using all the models tested and used for cross validation.

### *3.2.9. Evaluation of blended predictions using trained models*

Three approaches of model blending using all the models tested were investigated: 1) phenomic predictions of DON content, 2) genomic predictions of DON content, and 3) both phenomic and genomic predictions of DON content. Blended forward prediction by model averaging was carried out following the formula:

$$(7) \hat{y} = \frac{1}{M} \sum_{m=1}^M \hat{y}_m$$

Where  $\hat{y}$  is the blended (averaged) forward prediction,  $M$  is the total number of individual prediction models, and  $\hat{y}_m$  is the predicted value from the  $m$ -th model.

Phenotypic correlation between predicted and actual DON content in the selection candidates were carried out using Pearson's correlation. Genetic correlation ( $r_G$ ) and prediction accuracy ( $Acc_A$ ) was also determined following the approach and formula adapted from equations 5 and 6, respectively.

In addition, indirect selection accuracy,  $Acc(I)$ , of the blended predictions was determined based on the formula adapted from Lopez-Cruz et al. (2020):

$$(8) Acc(I) = H_I(cor_{XI})$$

Where  $cor_{XI}$  is the genetic correlation between the observed DON content of the selection candidates (X) and the predicted DON content of the selection candidates based on blended predictions (I), and  $H_I$  is the square root of the estimated broad sense heritability ( $H_I^2$ ) in the predicted DON content of the selection candidates based on blended predictions, where  $H_I^2$  is the estimated heritability of the blended DON content predictions of the selection candidates calculated by dividing the estimated genetic variance of the blended DON content predictions of the selection candidates ( $\sigma^2_{G\_I}$ ) by its estimated phenotypic variance ( $\sigma^2_{P\_I}$ ).

### 3.2.10 Blended clustering for selection DON content selection

Unsupervised K-Means Clustering was carried out using the “cluster” package in R (Maechler et al., 2023) in selection candidates using three different blending of eight phenomic, eight genomic and 16 phenomic and genomic prediction models. Clustering was carried out based on the objective function:

$$(9) J = \sum_{k=1}^K \sum_{i \in C_k} \|\hat{y}_i - \mu_k\|^2$$

Where  $J$  is the within-cluster sum of squares,  $K$  is the pre-defined number of clusters,  $C_k$  is the set of data points assigned to cluster  $k$ ,  $\hat{y}_i = [\hat{y}_{i1}, \hat{y}_{i2}, \dots, \hat{y}_{iM}]$  is the vector of predicted values from  $M$  different models for data point  $i$ ,  $\mu_k$  is the centroid of cluster  $k$ , defined as the mean vector of predicted values across all points in  $C_k$ . Two clusters were used for clustering (centers=2) as per result of Silhouette method employed to identify the optimum number of clusters. Two clusters were used for all K-means clustering runs for consistency.

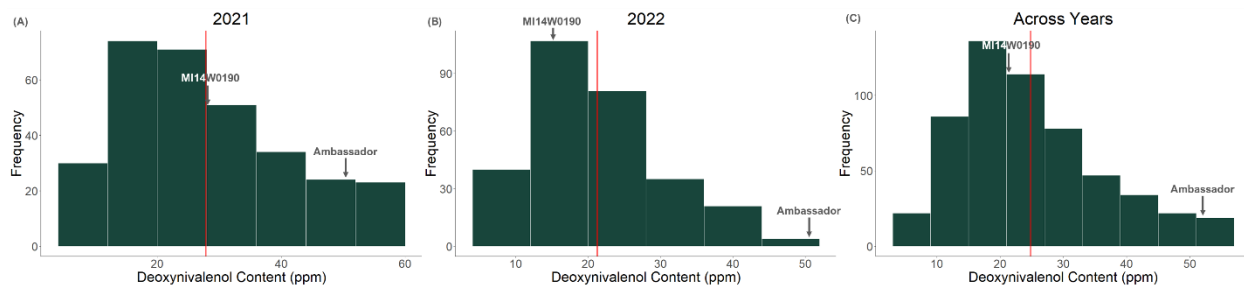
### 3.3. Results

#### 3.3.1. Deoxynivalenol evaluation

Significant variation in DON content was observed among the genotypes ( $p < 0.00001$ ) tested in 2021 and in 2022 (Figure 3.1). In 2021, 229 genotypes (74.6%) had significantly lower DON content than the susceptible check Ambassador, while 194 genotypes (63.5%) were not different from the resistant check MI14W0190, and 27 (8.8%) had significantly lower DON content than MI14W0190. In 2022, 279 genotypes (96.9%) had significantly lower DON content than Ambassador, and 194 (67.4%) demonstrated no difference from MI14W0190. No genotypes in 2022 had lower DON content than MI14W0190. Across both years, there was significant variation in DON due to genotype ( $p < 0.00001$ ) (Figure 3.1), year ( $p = 0.0000756$ ) and genotype-by-year interaction ( $p = 0.000174$ ). The entry mean heritability for DON across years at 0.55 was lower compared to single-year estimates for 2021 at 0.85 and 2022 at 0.81 (Supplementary Table 3.1.).

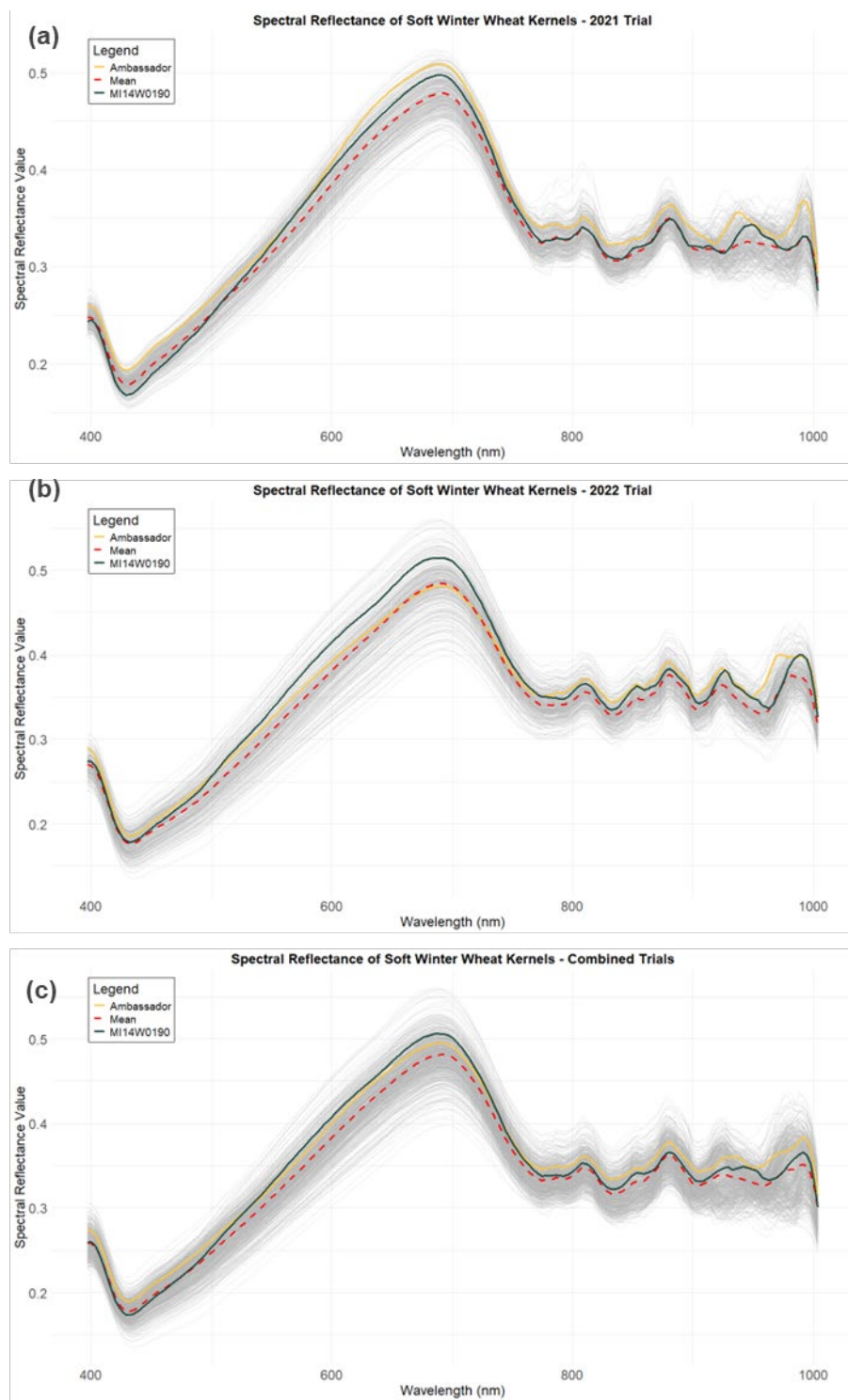
#### 3.3.2. Hyperspectral imaging of deoxynivalenol infected wheat kernels

Significant correlations were identified between the 204 wavebands in the visible and near-infrared regions and DON content for 2021, 2022, and across years (Figure 3.3). In 2021, all visible

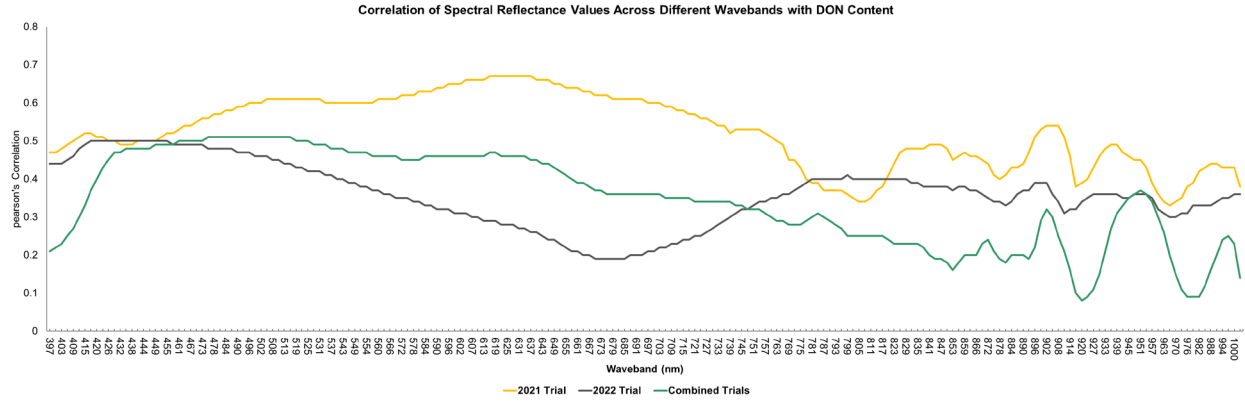


**Figure 3.1.** Frequency distributions of DON content in genotypes tested in (a) 2021 and (b) 2022. (c) 2021 and 2022 combined. Solid red bar represents average DON content among the genotypes tested for individual trials and across years. MI14W0190 and Ambassador, resistant and susceptible checks respectively, were also highlighted.





**Figure 3.2.** Spectral reflectance profile at the 400 to 1000 nm range of genotypes evaluated in (a) 2021 and (b) 2022, as well as in a (c) combined trials. Spectral profile of FHB-resistant check MI14W0190 and FHB-susceptible check Ambassador were highlighted. Broken line represents the average spectral profile for all the genotypes tested.



**Figure 3.3.** Pearson's correlation of spectral reflectance values with Deoxynivalenol for 2021, 2022 and 2021 and 2022 trials combined.

light wavebands from 409 nm to 763 nm, along with wavebands 896 n to 911 nm in the NIRs region, demonstrated correlations of at least 0.5 with DON content. Wavebands from 496 nm to 703 nm all demonstrated correlations of at least 0.6 (Figure 3.3). In 2022, correlations were lower, with only 14 wavebands from 418 nm to 455 nm reaching 0.5, and visible spectrum correlations ranged from 0.19 to 0.5. Interestingly, several near-infrared wavebands from 800 nm to 1000 nm had higher correlations with DON than those in the visible spectrum (Figure 3.3). Across years, correlations ranged from 0.21 to 0.51 in the visible spectrum from 400 nm to 800 nm and 0.09 to 0.37 in the near-infrared region. A moderate correlation of at least 0.5 was observed in wavebands from 461 nm to 525 nm (Figure 3.3).

### 3.3.3. Model comparison for deoxynivalenol prediction: predictive ability

To evaluate and compare DON prediction models trained using phenomic or genomic data, a five-fold cross-validation with an 80/20 training/testing split was performed 100 times for each model. Using the 2021 trial as the training set, wavebands in the visible light spectrum (Phenomic\_VIS) outperformed genomic prediction and phenomic prediction using all wavebands (Phenomic\_VIS/NIR) across models. Bayes B, Bayes C, and RRBLUP achieved the highest

**Table 3.1.** Cross validation predictive abilities (Pearson’s correlation) of phenomic and genomic prediction models in 2021, 2022 and combined trials.

Model	Predictor	2021	2022	2021 and 2022 Combined
Bayes A	Geno	0.69 ± 0.058	0.43 ± 0.096	0.61 ± 0.063
	Pheno_V	0.81 ± 0.047	0.65 ± 0.07	0.69 ± 0.049
	Pheno_V/N	0.77 ± 0.057	0.59 ± 0.078	0.68 ± 0.057
Bayes B	Geno	0.70 ± 0.065	0.45 ± 0.09	0.61 ± 0.060
	Pheno_V	0.82 ± 0.045	0.67 ± 0.082	0.68 ± 0.063
	Pheno_V/N	0.81 ± 0.046	0.63 ± 0.07	0.69 ± 0.056
Bayes C	Geno	0.70 ± 0.062	0.42 ± 0.11	0.62 ± 0.061
	Pheno_V	0.82 ± 0.042	0.63 ± 0.075	0.71 ± 0.052
	Pheno_V/N	0.81 ± 0.054	0.59 ± 0.092	0.68 ± 0.061
Bayesian LASSO	Geno	0.71 ± 0.059	0.43 ± 0.102	0.61 ± 0.065
	Pheno_V	0.81 ± 0.05	0.65 ± 0.075	0.69 ± 0.050
	Pheno_V/N	0.71 ± 0.068	0.58 ± 0.09	0.64 ± 0.060
BRR	Geno	0.70 ± 0.061	0.43 ± 0.088	0.61 ± 0.070
	Pheno_V	0.81 ± 0.051	0.61 ± 0.076	0.69 ± 0.055
	Pheno_V/N	0.67 ± 0.057	0.57 ± 0.096	0.60 ± 0.060
Random Forest	Geno	0.71 ± 0.061	0.39 ± 0.087	0.61 ± 0.056
	Pheno_V	0.68 ± 0.062	0.57 ± 0.07	0.66 ± 0.051
	Pheno_V/N	0.69 ± 0.043	0.55 ± 0.075	0.65 ± 0.049
RRBLUP	Geno	0.41 ± 0.14	0.68 ± 0.076	0.67 ± 0.062
	Pheno_V	0.82 ± 0.04	0.70 ± 0.069	0.70 ± 0.050
	Pheno_V/N	0.79 ± 0.044	0.63 ± 0.064	0.69 ± 0.049
XGBoost	Geno	0.61 ± 0.088	0.33 ± 0.096	0.60 ± 0.076
	Pheno_V	0.64 ± 0.089	0.53 ± 0.095	0.64 ± 0.047
	Pheno_V/N	0.62 ± 0.095	0.49 ± 0.08	0.61 ± 0.062

Bayesian LASSO – Bayesian Least Absolute Shrinkage and Selection Operator, BRR – Bayesian Ridge Regression, RRBLUP – Ridge Regression Best Linear Unbiased Prediction, XGBoost – Extreme Gradient Boosting, VIS – Visible Light Spectrum, VIS/NIR – Visible and Near Infrared Light Spectrum, Geno = Genomic prediction, Pheno\_V = Phenomic prediction using wavebands in the visible light spectrum, Pheno\_V/N = Phenomic Prediction using wavebands in the visible light and near infrared light spectrums

predictive ability at 0.82, followed by Bayes A and Bayesian LASSO (Table 3.1). In the 2022 trial, predictive ability was lower across all models and phenomic prediction using wavebands in the visible light spectrum again were better predictors of DON. RRBLUP had the highest had the highest predictive ability of 0.70 in 2022, followed by Bayes B and Bayes A (Table 3.1).

To develop a multi-year prediction, BLUEs were generated from genotypes tested in 2021 and 2022. Multi-year phenomic predictions consistently outperformed genomic predictions across all models (Table 3.1). As with individual trials as training set, wavebands in the visible light spectrum had significantly higher predictive ability, with Bayes C achieving the highest at 0.71.

**Table 3.2.** Genetic correlation from cross validation of phenomic and genomic prediction models in 2021, 2022 and combined trials.

Model	Predictor	2021	2022	2021 and 2022 Combined
Bayes A	Geno	$0.72 \pm 0.068$	$0.56 \pm 0.114$	$0.69 \pm 0.062$
	Pheno_V	$0.82 \pm 0.040$	$0.66 \pm 0.097$	$0.72 \pm 0.055$
	Pheno_V/N	$0.81 \pm 0.047$	$0.63 \pm 0.090$	$0.71 \pm 0.060$
Bayes B	Geno	$0.73 \pm 0.057$	$0.54 \pm 0.126$	$0.68 \pm 0.055$
	Pheno_V	$0.82 \pm 0.044$	$0.67 \pm 0.091$	$0.71 \pm 0.053$
	Pheno_V/N	$0.81 \pm 0.048$	$0.64 \pm 0.093$	$0.71 \pm 0.055$
Bayes C	Geno	$0.74 \pm 0.057$	$0.54 \pm 0.126$	$0.69 \pm 0.059$
	Pheno_V	$0.82 \pm 0.044$	$0.64 \pm 0.101$	$0.70 \pm 0.060$
	Pheno_V/N	$0.81 \pm 0.050$	$0.59 \pm 0.091$	$0.70 \pm 0.052$
Bayesian LASSO	Geno	$0.72 \pm 0.057$	$0.53 \pm 0.116$	$0.68 \pm 0.064$
	Pheno_V	$0.82 \pm 0.041$	$0.66 \pm 0.096$	$0.72 \pm 0.046$
	Pheno_V/N	$0.79 \pm 0.046$	$0.58 \pm 0.093$	$0.71 \pm 0.055$
BRR	Geno	$0.72 \pm 0.069$	$0.52 \pm 0.133$	$0.68 \pm 0.064$
	Pheno_V	$0.82 \pm 0.041$	$0.65 \pm 0.087$	$0.72 \pm 0.044$
	Pheno_V/N	$0.81 \pm 0.048$	$0.58 \pm 0.104$	$0.70 \pm 0.059$
Random Forest	Geno	$0.74 \pm 0.060$	$0.54 \pm 0.101$	$0.71 \pm 0.061$
	Pheno_V	$0.71 \pm 0.059$	$0.54 \pm 0.098$	$0.67 \pm 0.048$
	Pheno_V/N	$0.70 \pm 0.062$	$0.54 \pm 0.089$	$0.67 \pm 0.053$
RRBLUP	Geno	$0.45 \pm 0.158$	$0.73 \pm 0.068$	$0.71 \pm 0.062$
	Pheno_V	$0.83 \pm 0.041$	$0.67 \pm 0.080$	$0.72 \pm 0.049$
	Pheno_V/N	$0.80 \pm 0.045$	$0.61 \pm 0.076$	$0.71 \pm 0.048$
XGBoost	Geno	$0.72 \pm 0.053$	$0.47 \pm 0.118$	$0.69 \pm 0.065$
	Pheno_V	$0.66 \pm 0.067$	$0.52 \pm 0.099$	$0.64 \pm 0.056$
	Pheno_V/N	$0.66 \pm 0.070$	$0.49 \pm 0.092$	$0.62 \pm 0.054$

Bayesian LASSO – Bayesian Least Absolute Shrinkage and Selection Operator, BRR – Bayesian Ridge Regression, RRBLUP – Ridge Regression Best Linear Unbiased Prediction, XGBoost – Extreme Gradient Boosting, VIS – Visible Light Spectrum, VIS/NIR – Visible and Near Infrared Light Spectrum, Geno = Genomic prediction, Pheno\_V = Phenomic prediction using wavebands in the visible light spectrum, Pheno\_V/N = Phenomic Prediction using wavebands in the visible light and near infrared light spectrums

#### *3.3.4. Model comparison for deoxynivalenol prediction: genetic correlation*

Genetic correlation was also assessed to further compare phenomic and genomic prediction models from the cross validations. In the 2021 trial as training set, higher genetic correlations were observed in predictions using wavebands in the visible light spectrum (Pheno\_VIS) (Table 3.2). RRBLUP showed the highest genetic correlation (0.83), whereas Bayesian models showed slightly lower and similar genetic correlations (0.82) (Table 3.2). Using the 2022 trial, similar observations were obtained, where predictions using wavebands in the visible light spectrum (Pheno\_VIS) showed higher genetic correlations. BayesB and RRBLUP recorded the highest genetic correlations (0.67) followed by Bayes A and Bayesian LASSO (Table 3.2).

Multi-year phenomic predictions outperformed genomic predictions, except in Random Forest and XGBoost where genetic correlations from genomic predictions were higher. The highest genetic correlations were recorded using Bayes A, Bayesian LASSO, and RRBLUP (Table 3.2).

#### *3.3.5. Model comparison for deoxynivalenol prediction: prediction accuracy*

Prediction accuracy, the product between genetic correlations and square root of the estimated heritability of the actual DON content in the testing set, was also determined to fully assess the potential utility of the trained models for DON prediction. In the 2021 trial, phenomic predictions using wavebands in the visible light spectrum (Pheno-VIS) showed higher prediction accuracy over using the full waveband ranges or with genomic predictions (0.60 to 0.77) (Table 3.3). However, in Random Forest and XGBoost, higher prediction accuracy was observed in genomic predictions over phenomic predictions (Table 3.3). In the 2022 trial, phenomic predictions using the wavebands in the visible light spectrum were higher than genomic prediction, (0.26 to 0.33), with Bayes C where prediction accuracy being the highest using the full waveband ranges (0.49) (Table 3.3). In addition, genomic prediction using RRBLUP resulted in the highest

prediction accuracy at 0.50 (Table 3.3). In the multi-year predictions, phenomic and genomic prediction accuracy appears to be somewhat similar, with all of the prediction accuracies, regardless of whether it is phenomic or genomic prediction, ranged from 0.48 to 0.57 (Table 3.3)

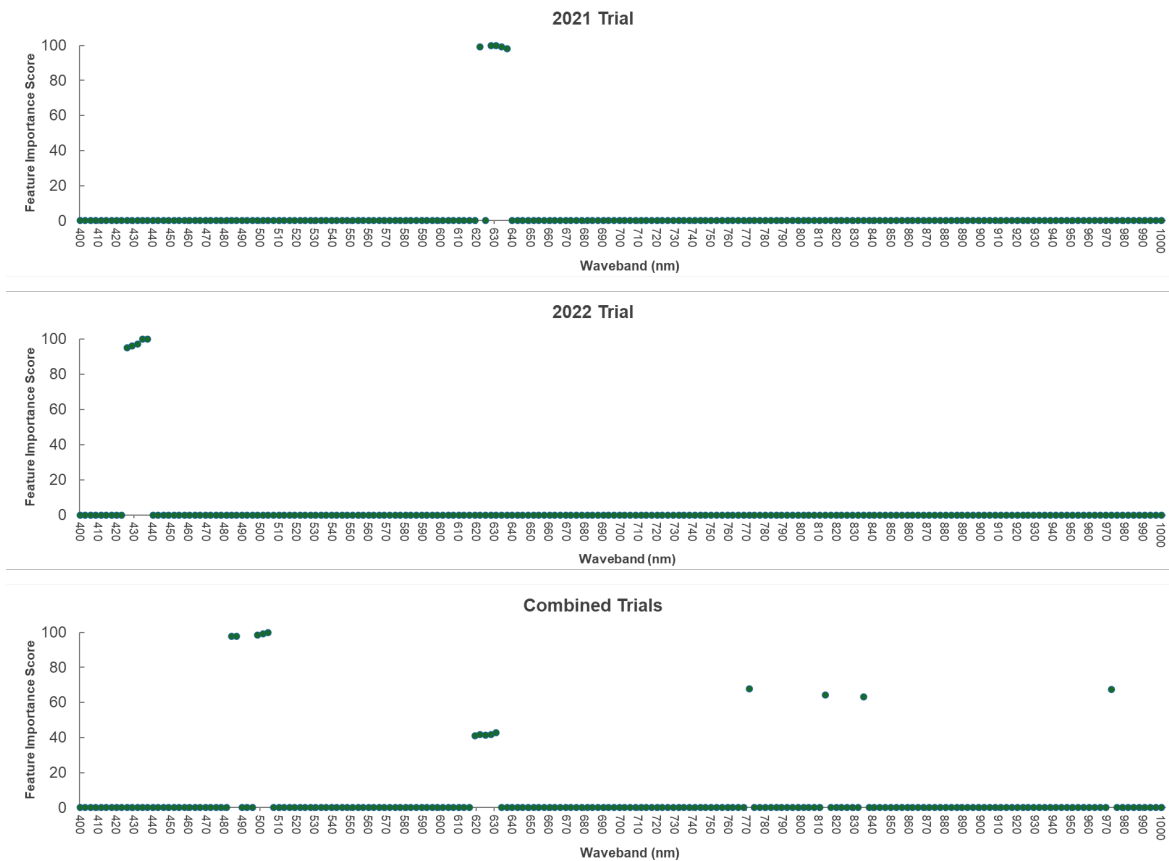
**Table 3.3.** Prediction accuracy from cross validation of phenomic and genomic prediction models in 2021, 2022 and combined trials.

Model	Predictor	2021	2022	2021 and 2022 Combined
Bayes A	Geno	$0.67 \pm 0.010$	$0.22 \pm 0.146$	$0.51 \pm 0.101$
	Pheno_V	$0.76 \pm 0.099$	$0.33 \pm 0.201$	$0.55 \pm 0.105$
	Pheno_V/N	$0.73 \pm 0.111$	$0.31 \pm 0.196$	$0.53 \pm 0.120$
Bayes B	Geno	$0.65 \pm 0.102$	$0.26 \pm 0.159$	$0.53 \pm 0.103$
	Pheno_V	$0.76 \pm 0.103$	$0.30 \pm 0.189$	$0.54 \pm 0.125$
	Pheno_V/N	$0.74 \pm 0.118$	$0.30 \pm 0.198$	$0.53 \pm 0.117$
Bayes C	Geno	$0.65 \pm 0.100$	$0.24 \pm 0.161$	$0.53 \pm 0.095$
	Pheno_V	$0.74 \pm 0.108$	$0.31 \pm 0.188$	$0.52 \pm 0.138$
	Pheno_V/N	$0.74 \pm 0.108$	$0.49 \pm 0.166$	$0.51 \pm 0.104$
Bayesian LASSO	Geno	$0.66 \pm 0.086$	$0.25 \pm 0.158$	$0.53 \pm 0.095$
	Pheno_V	$0.77 \pm 0.091$	$0.32 \pm 0.188$	$0.57 \pm 0.105$
	Pheno_V/N	$0.73 \pm 0.087$	$0.29 \pm 0.161$	$0.54 \pm 0.107$
BRR	Geno	$0.65 \pm 0.101$	$0.22 \pm 0.152$	$0.53 \pm 0.094$
	Pheno_V	$0.77 \pm 0.094$	$0.31 \pm 0.195$	$0.55 \pm 0.105$
	Pheno_V/N	$0.74 \pm 0.097$	$0.27 \pm 0.178$	$0.54 \pm 0.110$
Random Forest	Geno	$0.68 \pm 0.082$	$0.25 \pm 0.120$	$0.52 \pm 0.079$
	Pheno_V	$0.65 \pm 0.082$	$0.27 \pm 0.149$	$0.50 \pm 0.089$
	Pheno_V/N	$0.65 \pm 0.075$	$0.26 \pm 0.136$	$0.50 \pm 0.093$
RRBLUP	Geno	$0.32 \pm 0.176$	$0.50 \pm 0.176$	$0.51 \pm 0.139$
	Pheno_V	$0.76 \pm 0.083$	$0.31 \pm 0.167$	$0.54 \pm 0.095$
	Pheno_V/N	$0.74 \pm 0.082$	$0.28 \pm 0.149$	$0.53 \pm 0.093$
XGBoost	Geno	$0.66 \pm 0.079$	$0.22 \pm 0.107$	$0.51 \pm 0.070$
	Pheno_V	$0.60 \pm 0.088$	$0.26 \pm 0.146$	$0.49 \pm 0.087$
	Pheno_V/N	$0.60 \pm 0.088$	$0.24 \pm 0.129$	$0.48 \pm 0.087$

Bayesian LASSO – Bayesian Least Absolute Shrinkage and Selection Operator, BRR – Bayesian Ridge Regression, RRBLUP – Ridge Regression Best Linear Unbiased Prediction, XGBoost – Extreme Gradient Boosting, VIS – Visible Light Spectrum, VIS/NIR – Visible and Near Infrared Light Spectrum, Geno = Genomic prediction, Pheno\_V = Phenomic prediction using wavebands in the visible light spectrum, Pheno\_V/N = Phenomic Prediction using wavebands in the visible light and near infrared light spectrums

### 3.3.6. Feature selection

Wavebands in the visible light spectrum from 400 nm to 800 nm were identified to have the most important wavebands for both individual years and across years using recursive partitioning and regressive tree model (Figure 3.4). Specifically, in the 2021 trial, wavebands 622 nm, 628 nm, 631 nm, 634 nm, and 637 nm, all located in the red-light spectrum, were identified as the most informative wavebands having feature importance score of 98 to 100. Interestingly in the 2022 trial, wavebands in the blue light spectrum – 426 nm, 429 nm, 431 nm, 435 nm, and 438 nm – were identified to be the most informative wavebands with feature importance scores ranging from 95 to 100. In the combined trials, most informative wavebands having feature importance



**Figure 3.4.** Feature selection based on recursive partitioning and regressive tree model. Feature importance scores of each of the wavebands in the visible and near infrared regions from the (a) 2021 trial, (b) 2022 trial, and (c) 2021 and 2022 trials combination.

scores of 40 to 100 were identified across the entire spectrum, including four wavebands at the near infrared region – 814 nm, 835 nm, 972 nm, and 1004 nm. Majority of the informative wavebands still are located in the visible light spectrum – 484 nm, 487 nm, 499 nm, 502 nm, 505 nm, 619 nm, 622 nm, 625 nm, 628 nm, 631 nm, and 772 nm. Overall, wavebands in the red and green light spectrum appear to have the most important wavebands associated with DON concentration in infected wheat kernels.

### 3.3.7. Predictive ability of trained models in $F_{4:5}$ selection candidates

After evaluating phenomic and genomic model performance within training sets, we evaluated the model predictive ability outside of the training sets on  $F_{4:5}$  selection candidates evaluated in 2022 and 2023. All models were used for forward prediction and model blending.

When using the 2021 training set to predict DON content in the 2022 selection candidates, blended phenomic and genomic predictions demonstrated higher phenotypic correlations with

**Table 3.4.** Variation in correlations, prediction accuracy, and indirect selection accuracy of blended phenomic, genomic, and combined phenomic predictions in the selection candidates predicted using different training sets.

Training Set	Sel. Cand.	rP			rG			$H_{(I)}^2$			Acc <sub>(I)</sub>			Acc <sub>(A)</sub>		
		Ph	G	Ph+G	Ph	G	Ph+G	Ph	G	Ph+G	Ph	G	Ph+G	Ph	G	Ph+G
2021	2022 $F_{4:5}$	0.12	0.15	0.18	0.02	0.28	0.26	0.42	0.34	0.45	0.015	0.16	0.14	0.01	0.12	0.11
	2023 $F_{4:5}$	0.19	0.21	0.22	0.19	0.43	0.35	0.04	0.03	0.05	0.04	0.08	0.08	0.05	0.12	0.1
2022	2022 $F_{4:5}$	0.49	0.27	0.47	0.33	0.5	0.58	0.51	0.51	0.6	0.24	0.36	0.45	0.14	0.21	0.25
	2023 $F_{4:5}$	0.52	0.14	0.41	0.47	0.33	0.5	0.13	0.11	0.2	0.17	0.11	0.22	0.13	0.09	0.14
2021+ 2022	2022 $F_{4:5}$	0.3	0.2	0.31	0.14	0.36	0.38	0.44	0.37	0.5	0.09	0.22	0.27	0.06	0.16	0.16
	2023 $F_{4:5}$	0.38	0.18	0.31	0.31	0.38	0.41	0.07	0.04	0.08	0.08	0.08	0.12	0.07	0.11	0.12

Sel. Cand. = Selection Candidates, Ph = Blended Phenomic Predictions, G = Blended Genomic Predictions, Ph+G = Blended Phenomic and Genomic Predictions,  $r_P$  = phenotypic correlation of the blended predictions with observed DON content in the selection candidates,  $H_{(I)}^2$  = estimated broad-sense heritability calculated from the blended predicted DON values, Acc<sub>(I)</sub> = indirect selection accuracy, Acc<sub>(A)</sub> = prediction accuracy



observed DON content than the blended genomic predictions and the blended phenomic predictions, with the latter having slightly higher phenotypic correlation with DON. A similar observation was also obtained when the 2021 training set was used to predict the 2023 selection candidates. (Table 3.4).

Genotypes from the 2022 trial were used to predict DON content in 239 F<sub>4:5</sub> breeding lines also evaluated in 2022. The blended phenomic predictions had a higher phenotypic correlation with observed DON content than the blended genomic predictions and blended phenomic and genomic predictions, with the latter having relatively similar phenotypic correlation with blended phenomic predictions (Table 3.4). A similar pattern was observed in predicting the 2023 selection candidates with 2022 trial as the training set, however the difference between the blended phenomic predictions compared to blended phenomic and genomic predictions were more profound (Table 3.4).

When using BLUEs across years for training to predict DON content in the 2022 selection candidates, the blended phenomic and genomic predictions demonstrated a slightly higher phenotypic correlation with observed DON content, compared to blended genomic predictions and blended phenomic predictions (Table 3.4). Using BLUEs across years to predict DON content in 2023 selection candidates, the blended phenomic predictions had higher phenotypic correlation with observed DON content than blended phenomic and genomic predictions. Both outperformed the blended genomic predictions (Table 3.4).

#### *3.3.8. Genetic correlation of predicted and observed deoxynivalenol content in the selection candidates*

We evaluated the genetic correlation between predicted DON values and observed DON content in the selection candidates. Genomic estimated breeding values (GEBVs) generated from

genomic predictions are primarily based on additive genetic effects and assume no environmental influence (Meuwissen et al., 2001; Habier et al., 2010). In contrast, phenomic predictions incorporate non-additive effects and can include genotype-by-environment interactions and typically predict phenotype rather than genetic merit (Rincent et.al., 2018).

The blended phenomic and genomic predictions demonstrated higher genetic correlation with observed DON content (Table 3.4) in 2022 F<sub>4:5</sub> selection candidates predicted using the 2022 trial and combined trials as training set compared to blended phenomic or blended genomic predictions. A similar observation was also obtained in the 2023 selection candidates (Table 3.4). For both 2022 and 2023 selection candidates, when 2021 trial was used as training set, a higher genetic correlation was observed in blended genomic predictions compared to either blended phenomic or blended genomic predictions (Table 3.4). Overall, blended phenomic and genomic predictions appear to result in higher genetic correlations over either blended genomic or blended phenomic predictions.

### *3.3.9. Prediction accuracy and indirect selection accuracy estimation from forward prediction of DON content in the selection candidates*

For the 2022 selection candidates, a higher prediction accuracy was obtained in the blended phenomic and genomic predictions and in blended genomic predictions, with the latter having slightly higher prediction accuracy, when 2021 trial was used as training set (Table 3.4). Using the 2022 trial as training set, higher prediction accuracy was obtained in the blended phenomic and genomic predictions, whereas similar prediction accuracies were obtained in blended phenomic and genomic predictions and in blended genomic predictions using the combined trials as training set (Table 3.4).

For the 2023 selection candidates, a higher prediction accuracy was obtained in the genomic predictions, when the 2021 trial was used as training set (Table 3.4). When the 2022 trial and combined trials were used as training set, the blended phenomic and genomic predictions resulted in higher prediction accuracy (Table 3.4). Overall, blended phenomic and genomic predictions resulted in higher prediction accuracy, especially over blended phenomic predictions.

Indirect selection accuracy, derived by multiplying the square root of the estimated broad-sense heritability of predicted DON content with the genetic correlation between predicted and observed DON content in selection candidates. For the 2022 selection candidates, higher indirect selection accuracy was observed in blended genomic predictions with 2021 trial as training set, whereas higher indirect selection accuracy was observed in the blended phenomic and genomic predictions when the 2022 trial and the combined trials were used as training set (Table 3.4).

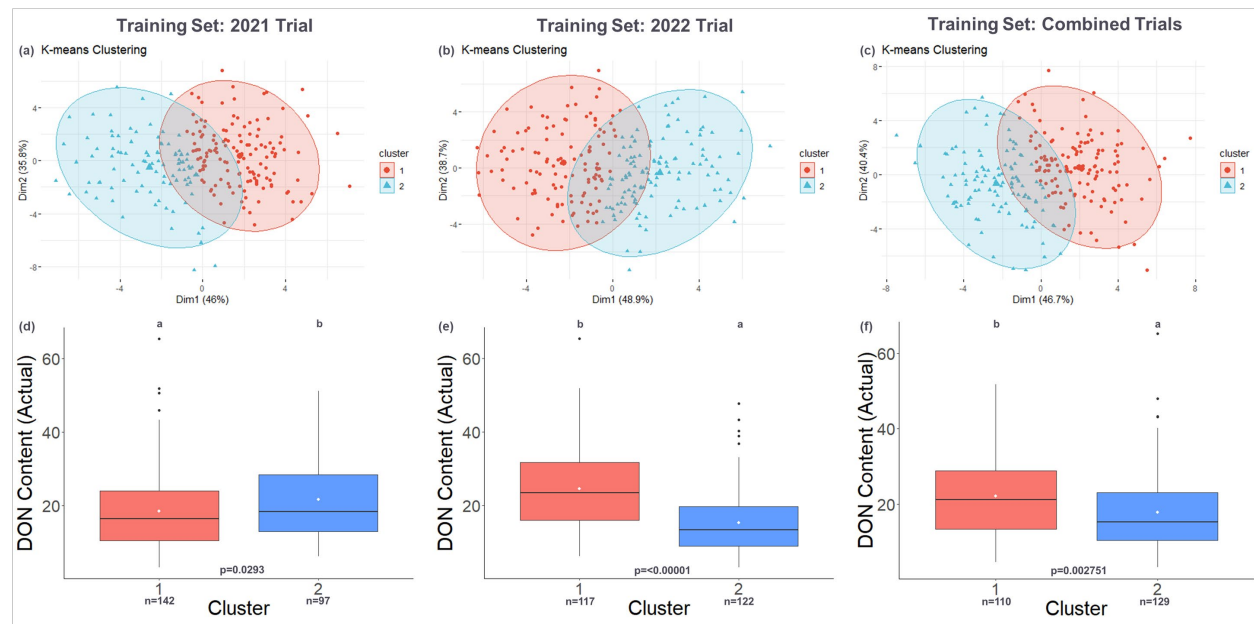
For the 2023 selection candidates, relatively similar indirect selection accuracy was obtained in the blended genomic predictions and in the blended phenomic and genomic predictions using 2021 trial as training set (Table 3.4). Higher indirect selection accuracy was obtained, however, in the blended phenomic and genomic predictions (Table 3.4). Overall, blended phenomic and genomic predictions resulted in higher indirect selection accuracy over blended phenomic and blended genomic predictions. This is likely due to the estimated heritability of predictions from blended phenomic and genomic predictions being relatively higher compared to blended phenomic or blended genomic predictions (Table 3.4).

#### *3.3.10. Unsupervised K-Means clustering of selection candidates using model blending*

Truncation selection can be applied to predicted DON content generated from a single phenomic or genomic prediction model. Blending of predictions from multiple models can enable classification of selection candidates for selection. We used unsupervised K-Means clustering to

separate the selection candidates into groups based on the blending of predictions generated across all models evaluated and used. Two groups were used for clustering based on the results of the Silhouette method.

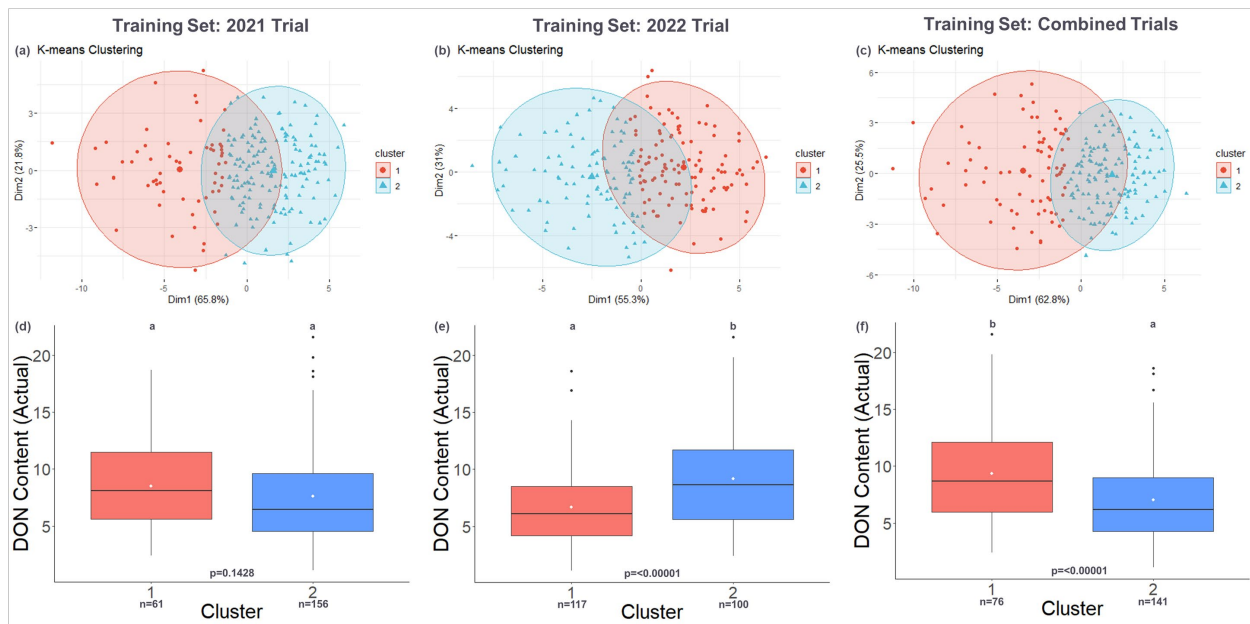
Clustering of the 2022 F<sub>4:5</sub> selection candidates with combined phenomic and genomic predictions generated two distinct groups, clusters 1 and 2, based on differences in the mean DON level of each group. These clusters demonstrated significantly different mean DON content using training sets from 2021 ( $p=0.02294$ ), 2022 ( $p=1.074 \times 10^{-11}$ ), and the combined 2021 and 2022 trials ( $p=0.002751$ ) (Figure 3.5, Supplementary Tables 3.5 to 3.7). Significant differences between Low and High DON groups were also observed from both blended phenomic and blended genomic



**Figure 3.5.** Unsupervised K-means clustering based on blended phenomic and genomic predictions of DON content in 2022 F<sub>4:5</sub> selection candidates using different training sets. (a, b, c) Clustering of 2022 F<sub>4:5</sub> selection candidates using genomically and predicted values. (d,e,f) Corresponding boxplots showing the actual DON content of the genotypes in Clusters 1 and 2 grouped based on phenomically and genomically predicted values. White dot represents the mean actual DON content of genotypes belonging to Clusters 1 and 2. Means with the same letter are not significantly different at alpha 0.05.

predictions using the 2022 trial and the combined 2021 and 2022 trials as training sets (Supplementary Figures 3.1 and 3.2, Supplementary Tables 3.5 and 3.8). However, when the 2021 trial was used as training set to predict 2022 selection candidates using only the blended genomic or blended phenomic predictions, no significant difference was observed between Low and High DON (Supplementary Figure 3.1 and 3.2, Supplementary Table 3.5). groups.

Clustering of the 2023 F<sub>4:5</sub> selection candidates with combined phenomic and genomic predictions also generated two distinct groups with significantly different mean DON content using training sets from 2022 ( $p=3.604 \times 10^{-6}$ ) and combined 2021 and 2022 trials ( $p=3.605 \times 10^{-5}$ ) (Figure 3.6, Supplementary Tables 3.9 and 3.10). Combined phenomic and genomic predictions



**Figure 3.6.** Unsupervised K-means clustering based on blended phenomic and genomic predictions of DON content in 2023 F<sub>4:5</sub> selection candidates using different training sets. (a, b, c) Clustering of 2023 F<sub>4:5</sub> selection candidates using genomically and predicted values. (d,e,f) Corresponding boxplots showing the actual DON content of the genotypes in Clusters 1 and 2 grouped based on phenomically and genomically predicted values. White dot represents the mean actual DON content of genotypes belonging to Clusters 1 and 2. Means with the same letter are not significantly different at alpha 0.05.

using the 2021 trial as the training set yielded no significant difference between groups of 2023 selection candidates (Figure 3.6, Supplementary Table 3.9). However, significant differences between groups of 2023 selection candidates were generated from the blended phenomic predictions using the 2021 trial ( $p=0.007936$ ), 2022 trial ( $p=4.018 \times 10^{-12}$ ) and combined 2021 and 2022 trials ( $p=1.536 \times 10^{-5}$ ) as training sets (Supplementary Figure 3.3, Supplementary Tables 3.10 and 3.11). Significant differences between groups were observed using the blended genomic predictions with the 2021 trial as the training set ( $p=0.02376$ ) (Supplementary Figure 3.4, Supplementary Table 3.9).

### **3.4. Discussion**

#### *3.4.1. Hyperspectral imaging of DON infected kernels and phenomic prediction*

Hyperspectral imaging offers a rapid, cost-effective, and non-destructive approach to evaluate DON content in wheat kernels in comparison to labor and resource intensive chemometric approaches (Femenias et al., 2020). Hyperspectral imaging methods produce multi-dimensional data from hundreds of wavebands and identifying the key wavebands provides crucial information on the basis of observed variation in DON content. In our study, the most important wavebands were found in the visible light spectrum, particularly in the red-light range around 600 nm. A noticeable noise was observed beyond 750 nm even after smoothing using Savitzky-Golay approach could be due to inaccurate spectra collection in this region. In this study, LED was used as the light source, potentially affecting the collection in the NIRs region, as most LEDs do not emit light beyond the 800 range. Likely, this may have also affected the correlation of the wavebands in the NIR regions to be relative lower compared to wavebands in the visible light region. Therefore, using light sources emitting or covering the full VIS/NIR range would definitely be more advantageous. Notable variation was detected between the 2021 and 2022 trials, reflected

in the wavebands identified as important in each year as well as the differing correlations between individual wavebands and DON content. This variation could be due to the environmental factors peculiar to each year or differences in infection levels across the disease nursery which could introduce error into observations of DON for individual genotypes.

Despite the noises in the NIR region and the difference between years, the importance of the visible light spectrum between 410 nm and 640 nm is in agreement with previous reports of wavebands associated with *Fusarium*-damaged kernels (Shahin and Symons, 2011; Shahin and Symons, 2012, Ropelewska and Zapotoczny, 2018). Cross validations done in this study highlight the observation that using only the wavebands in the visible light spectrum would be enough to predict DON content, with relatively higher accuracy over using the entire spectrum of visible and near infrared light.

#### *3.4.2. Comparison of phenomic and genomic predictions for deoxynivalenol*

Genomic prediction employs genetic information, usually SNPs, which capture additive effects primarily from small effect QTLs (Cerrudo et al. 2018). Since genetic information is stable across generations and is essentially unaffected by the environment, genomic prediction enables the estimation of a candidate's genetic breeding value or genetic merit (Meuwissen et al. 2001). This estimation is essential for ranking and identifying candidates with superior performance and for selecting promising parents for population development. Several studies have demonstrated the use of genomic prediction for low DON content (Rutkoski et al. 2012, Salam et al. 2015, Gaire et al. 2022). The main advantage of genomic prediction is the ability to estimate breeding values of selection candidates early in the breeding cycle. However, genomic prediction can be costly due to the required genotyping, and environmental effects on the trait of interest must still be considered.

Phenomic predictions capture phenotypic variation that reflects both genetic and environmental factors and account for non-additive effects (Robert et al., 2022, Rincent et al., 2018). Therefore, phenomic prediction focuses on predicting the phenotype rather than solely the genetic merit or breeding value making phenomic prediction an excellent approach for selection in complex traits (Zhu et al., 2021), as in the case of DON content. Several works have laid down the foundation and demonstrated the potential use of hyperspectral imaging for DON prediction (Femenias et al. 2020 and references therein). Among these, Alisaac et al. (2019) demonstrated the use of hyperspectral imaging in assessing mycotoxin, including DON, in wheat kernels. However, unlike in our study, the wavebands generated were not used as predictors for predicting DON content, rather, for direct comparison of spectral signatures in relation to DON content was done. Su et al. (2020) did use the spectral reflectance values for DON prediction using locally weighted partial least squares regression (LWPLSR) with relative success in barley kernels.

In this study, phenomic prediction demonstrated higher predictive ability consistent with other studies across different crop species (Parmley et al. 2015, Lane et al. 2020, Galan et al. 2020, Brault et al. 2022), including wheat (Cuevas et al. 2019, Krause et al. 2019, Liu et al. 2024). Bayesian generalized linear regression models (Bayes A, Bayes B, Bayes C, Bayesian LASSO, and Bayesian Ridge Regression) demonstrated relatively higher predictive ability in cross-validation studies compared to the machine learning models Random Forest and Extreme Gradient Boosting. Further, our results are in conjunction with the observations made by Zhu et al. (2021) where phenomic predictive ability using Ridge Regression Best Linear Unbiased Prediction (RR-BLUP), Bayes B and Bayes C was higher than random forest and gradient boosting approaches. This suggests a linear relationship between the phenomic signals captured by hyperspectral imaging and DON content in wheat kernels.



Recently, however, Wang et al. (2025) argued that the comparison between genomic and phenomic predictions simply by assessing the phenotypic correlation (predictive ability), especially from cross validation schemes, could be misleading. As genomic prediction estimates the breeding value or genetic merit of the target trait, in our case DON, Wang et al. (2025) proposed the comparison of between genomic and phenomic predictions should be assessed using prediction accuracy, which takes into account the genetic correlation between the predicted and actual value, and the heritability of the target trait, rather than solely relying on Pearson's correlation between observed and predicted phenotypic values. We do agree that using predictive ability, solely, in comparing GEBVs with predicted phenotype may be misleading, as predicted phenotypes would be more phenotypically correlated to the actual value than GEBVs, especially when the predictors used (e.g. phenomic data) have some level of correlation with the target trait, basically further inflating the Pearson's correlation observed. In our case, higher genetic correlations were still observed in phenomic predictions compared to genomic predictions. More importantly, overall, we have seen a higher prediction accuracy in phenomic prediction compared to genomic predictions, especially in the 2021 and 2022 trials when independently used as training set. With these addition metrics focusing on evaluating how well phenomic systems may have captured genetic signals, thereby representing the actual genetic variation of actual DON content, this further demonstrated the potential utility of hyperspectral imaging derived information in prediction of DON content in wheat kernels. To our knowledge, this is the first work in DON prediction using hyperspectral imaging comparing it with genomic predictions using prediction accuracy and genetic correlations as metrics.

### *3.4.3. Training set composition influences forward prediction in selection candidates*

The choice of the training set significantly impacted predictions of DON content in both phenomic and genomic predictions. While phenomic prediction generally demonstrated higher predictive ability, genetic correlation, and prediction accuracy than genomic prediction, the values observed varied depending on the training set used.

The models trained with phenomic data from the 2022 trial demonstrated relatively higher predictive ability in the F<sub>4:5</sub> selection candidates compared to models trained with 2021 data and the combined 2021 and 2022 BLUEs, especially in the 2022 F<sub>4:5</sub> selection candidates. This is likely influenced by the training set and selection candidate being in the same environment and receiving relatively similar disease pressure. A similar trend was observed for genomic prediction models. However, there was a larger difference in prediction accuracy between years in genomic prediction models compared to phenomic prediction models.

We also observed variation in the genetic correlations between blended predictions (model averages) and the actual DON content of the selection candidates, depending on the training set used. For the majority of cases, blended genomic predictions had higher genetic correlations over blended phenomic predictions, except when 2022 trial was used as training set to predict the 2022 F<sub>4:5</sub> selection candidates. This further establishes that solely relying on phenotypic correlation, both in cross validation and forward predictions, may not be enough or even misleading.

In addition, the predictive abilities and prediction accuracy observed in cross-validation studies within training sets did not always translate when models were used to predict a set of selection candidates. While cross-validation demonstrated higher predictive abilities and prediction accuracy for phenomic predictions using the 2021 trial, this was not reflected in the

prediction of the 2023 selection candidates, where the 2022 training set generated better phenotypic correlations, genetic correlation, and prediction accuracy with observed DON content.

These differences could be driven by the variation among the genotypes included in each training set and their genetic relatedness to the selection candidates (Zhu et al. 2021). Of the 307 genotypes evaluated in 2021, only 37 were evaluated in 2022, due to the dynamic nature of active breeding programs. The selection candidates in 2022 and 2023 were derived from cross combinations of different parental genotypes and therefore have genetic backgrounds differing from the training set. Differences in infection levels and the unique environmental parameters within each year may have contributed to differences in predictive ability. Indeed, Robert et al. (2022) highlight that the environments, years in our case, in which the lines were grown, play a crucial role in predictive ability of spectral reflectance data. In addition, relatively different levels of correlation were observed in 2021 and 2022 between the wavebands and with DON content, likely affecting the performance of phenomic prediction models in forward prediction.

These variations in predictive ability, genetic correlations, and prediction accuracy depending on the predictors used (phenomic or genomic information) and training set used (2021 trial, 2022 trial, and combined trials) prompted us to explore whether combining phenomic and genomic information would be more advantageous. Thapa et al. (2024) demonstrated in their work that there is a potential for increasing predictive ability in predicting DON content when phenomic and genomic information was combined. However, we wanted to see whether this observation could be extended beyond cross validations and observed in forward predictions. In addition, we employed a different approach in combining phenomic and genomic information, or rather, the phenomic and genomic predictions.

#### *3.4.4. Model averaging and complementation of genomic prediction with phenomic predictions*

In this study, all the model used—Bayes A, Bayes B, Bayes C, Bayesian LASSO, Bayesian Ridge Regression, Random Forest, RRBLUP, and XGBoost—were used to predict DON content in the selection candidates. Given the differing assumptions and fitting in these models, as well the varying observations between phenomic and genomic predictions, model averaging approach was explored to enhance prediction accuracy. Phenomic predictions estimate the phenotype, which also captures the non-additive effects, usually confounded environmental effects, whereas genomic predictions estimate the breeding value of candidates independent of environmental influences. In our study, we opted to average the predicted values from different models and compared the blended predictions from phenomic models to those from genomic models as well as blending both phenomic and genomic predictions.

In the majority of cases, the genetic correlation between the blended predictions from phenomic models and observed DON in the selection candidates was higher than that from genomic models, as well as its prediction accuracy and indirect selection accuracy. Most importantly, the estimated heritability in the predicted values were always higher in the blended phenomic and genomic predictions, over the blended phenomic or blended genomic predictions. This observation also held true for indirect selection accuracy, suggesting that blending phenomic predictions can effectively predict DON content in wheat. Here, we demonstrated the advantage and utility, in forward prediction, of combining phenomic and genomic predictions, simply by blending the models through simple averaging.

Considering these observations, both phenomic and genomic predictions can be used whenever feasible and when resources allow us to implement a balanced selection approach that leverages the strengths of both methods. However, as phenomic information potentially captures

non-additive and environmental effects, evaluating models taking into account non-additive effects (additive effects and epistatic effects), both phenomic and genomic predictions would give an even more compelling comparison between phenomic and genomic predictions as well as between models taking into account solely additive effects. In addition to this, we encourage further exploration of ways and approaches in integrating phenomic and genomic information such as multi-modal neural networks or multi-kernel approaches. We propose, however, to always include genetic correlations, prediction accuracy described by Wang et al. (2025) and indirect selection accuracy described by Lopez-Cruz et al. (2020) to avoid misleading interpretations, both in cross validation and forward predictions, rather than solely relying on predictive ability assessment by phenotypic (Pearson's) correlations.

#### *3.4.5. Clustering-model blending of selection candidates*

Unsupervised K-Means as a model blending approach was explored to group selection candidates based on their predicted DON contents. Rather than applying truncation selection on predicted values from multiple phenomic or genomic prediction models, the predictions from multiple models were used as variables for clustering. Clustering of selection candidates using blended predictions generated two groups which differed significantly for mean observed DON content. We assume that this approach shall enhance the reduction of the selection pool, especially when dealing with a large number of selection candidates.

### **3.5. Conclusion**

Phenomic prediction of DON content in wheat grains using reflectance values derived from hyperspectral imaging demonstrated higher predictive ability and prediction accuracy than genomic prediction. Phenomic prediction resulted in varying phenotypic correlation, genetic correlation, prediction accuracy and indirect selection accuracies depending on the predictor and

training set used for forward prediction. Blending phenomic and genomic prediction models through simple model averaging is a straightforward mechanism to potentially account for both genetic and environmental influences on the trait of interest. Additionally, clustering of selection candidates using blended predictions offers an approach to leverage both phenomic and genomic predictions in selection decisions. Careful consideration is required for training set design and implementation of phenomic predictions as the inflated predictive ability from cross validations does not always translate to forward predictions. Additional metrics – genetic correlation, prediction accuracy, and indirect selection accuracy – has to be assessed to acquire a proper comparison of model and approaches, especially when comparing phenomic and genomic predictions. This study highlights a mechanism to develop and leverage both phenomic and genomic predictions in selection decisions in an active plant breeding program.

## REFERENCES

- Adak, A. *et al.* Phenomic data-driven biological prediction of maize through field-based high-throughput phenotyping integration with genomic data. *Journal of Experimental Botany* **74**, 5307–5326 (2023).
- Alisaac, E., Behmann, J., Kuska, M. T., Dehne, H.-W. & Mahlein, A.-K. Hyperspectral quantification of wheat resistance to Fusarium head blight: comparison of two *Fusarium* species. *Eur J Plant Pathol* **152**, 869–884 (2018).
- Alisaac, E. *et al.* Assessment of Fusarium infection and mycotoxin contamination of wheat kernels and flour using hyperspectral imaging. *Toxins* **11**, 556 (2019).
- Alisaac, E. & Mahlein, A.-K. Fusarium Head Blight on Wheat: Biology, Modern Detection and Diagnosis and Integrated Disease Management. *Toxins* **15**, 192 (2023).
- Bai, G., Su, Z. & Cai, J. Wheat resistance to Fusarium head blight. *Canadian Journal of Plant Pathology* **40**, 336–346 (2018).
- Bates, D. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823* (2014).
- Biecek, P. DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research* **19**, 1–5 (2018).
- Brault, C. *et al.* Interest of phenomic prediction as an alternative to genomic prediction in grapevine. *Plant Methods* **18**, 108 (2022).
- Buerstmayr, M., Steiner, B. & Buerstmayr, H. Breeding for Fusarium head blight resistance in wheat—Progress and challenges. *Plant Breeding* **139**, 429–454 (2020).
- Cerrudo, D. *et al.* Genomic Selection Outperforms Marker Assisted Selection for Grain Yield and Physiological Traits in a Maize Doubled Haploid Population Across Water Treatments. *Front. Plant Sci.* **9**, (2018).
- Chen, T. & He, T. xgboost: eXtreme Gradient Boosting.
- Covarrubias-Pazaran, G. Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. *PLOS ONE* **11**, e0156744 (2016).
- Crain, J., Mondal, S., Rutkoski, J., Singh, R. P. & Poland, J. Combining High-Throughput Phenotyping and Genomic Information to Increase Prediction and Selection Accuracy in Wheat Breeding. *The Plant Genome* **11**, 170043 (2018).
- Crossa, J. *et al.* Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* **112**, 48–60 (2014).

- Cuevas, J. *et al.* Deep Kernel for Genomic and Near Infrared Predictions in Multi-environment Breeding Trials. *G3 Genes|Genomes|Genetics* **9**, 2913–2924 (2019).
- Dallinger, H. G. *et al.* Predictor bias in genomic and phenomic selection. *Theor Appl Genet* **136**, 235 (2023).
- Endelmann, J.B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, **4**, 250-255 (2011).
- Femenias, A., Gatiús, F., Ramos, A. J., Sanchis, V. & Marín, S. Use of hyperspectral imaging as a tool for *Fusarium* and deoxynivalenol risk management in cereals: A review. *Food Control* **108**, 106819 (2020).
- Gaire, R. *et al.* Multi-trait genomic selection can increase selection accuracy for deoxynivalenol accumulation resulting from fusarium head blight in wheat. *The Plant Genome* **15**, e20188 (2022).
- Galán, R. J. *et al.* Integration of genotypic, hyperspectral, and phenotypic data to improve biomass yield prediction in hybrid rye. *Theor Appl Genet* **133**, 3001–3015 (2020).
- Glaubitz, J. C. *et al.* TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLOS ONE* **9**, e90346 (2014).
- Gonçalves, M. T. V. *et al.* Near-infrared spectroscopy outperforms genomics for predicting sugarcane feedstock quality traits. *PLOS ONE* **16**, e0236853 (2021).
- Habier, D., Tetens, J., Seefried, F.-R., Lichtner, P. & Thaller, G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet Sel Evol* **42**, 5 (2010).
- Haile, T. A. *et al.* Genomic prediction of agronomic traits in wheat using different models and cross-validation designs. *Theor Appl Genet* **134**, 381–398 (2021).
- Hansen, B. E. Least Squares Model Averaging. *Econometrica* **75**, 1175–1189 (2007).
- Hijmans, R. J. The raster package.
- Jackson, R. *et al.* Phenomic and genomic prediction of yield on multiple locations in winter wheat. *Frontiers in Genetics* **14**, (2023).
- Kick, D. R. & Washburn, J. D. Ensemble of best linear unbiased predictor, machine learning and deep learning models predict maize yield better than each model alone. *in silico Plants* **5**, diad015 (2023).
- Krause, M. R. *et al.* Hyperspectral Reflectance-Derived Relationship Matrices for Genomic Prediction of Grain Yield in Wheat. *G3 Genes|Genomes|Genetics* **9**, 1231–1247 (2019).



- Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **28**, 1–26 (2008).
- Lane, H. M. *et al.* Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. *The Plant Phenome Journal* **3**, e20002 (2020).
- Lenth, R. V. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software* **69**, 1–33 (2016).
- Liang, K., Ren, Z., Song, J., Yuan, R. & Zhang, Q. Wheat FHB resistance assessment using hyperspectral feature band image fusion and deep learning. *International Journal of Agricultural and Biological Engineering* **17**, 240–249 (2024).
- Liu, Q. *et al.* Development of Machine Learning Methods for Accurate Prediction of Plant Disease Resistance. *Engineering* **40**, 100–110 (2024).
- Lopez-Cruz, M. *et al.* Regularized selection indices for breeding value prediction using hyperspectral image data. *Sci Rep* **10**, 8195 (2020).
- Maechler., M. cluster: Cluster Analysis Basics and Extensions. *R package version 2.0.7–1* (2018).
- Magnus, J. R., Powell, O. & Prüfer, P. A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics* **154**, 139–153 (2010).
- Mahlein, A.-K. *et al.* Comparison and Combination of Thermal, Fluorescence, and Hyperspectral Imaging for Monitoring Fusarium Head Blight of Wheat on Spikelet Scale. *Sensors* **19**, 2281 (2019).
- Märtens, K., Hallin, J., Warringer, J., Liti, G. & Parts, L. Predicting quantitative traits from genome and phenome with near perfect accuracy. *Nat Commun* **7**, 11512 (2016).
- Meher, P. K., Rustgi, S. & Kumar, A. Performance of Bayesian and BLUP alphabets for genomic prediction: analysis, comparison, and results. *Heredity* **128**, 519–530 (2022).
- Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**, 1819–1829 (2001).
- Parmley, K., Nagasubramanian, K., Sarkar, S., Ganapathysubramanian, B. & Singh, A. K. Development of Optimized Phenomic Predictors for Efficient Plant Breeding Decisions Using Phenomic-Assisted Selection in Soybean. *Plant Phenomics* **2019**, (2019).
- Pérez, P. & de los Campos, G. Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* **198**, 483–495 (2014).
- Pérez-Rodríguez, P. & de los Campos, G. Multitrait Bayesian shrinkage and variable selection models with the BGLR-R package. *Genetics* **222**, iyac112 (2022).

- Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J.-L. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE* **7**, e32253 (2012).
- R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2023).
- Raftery, A. E., Madigan, D. & Hoeting, J. A. Bayesian Model Averaging for Linear Regression Models. *Journal of the American Statistical Association* **92**, 179–191 (1997).
- Rincent, R. *et al.* 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**, 3961–3972 (2018).
- Robert, P. *et al.* Phenomic selection in wheat breeding: identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection. *Theor Appl Genet* **135**, 895–914 (2022).
- Robert, P. *et al.* Phenomic selection in wheat breeding: prediction of the genotype-by-environment interaction in multi-environment breeding trials. *Theor Appl Genet* **135**, 3337–3356 (2022).
- Ropelewska, E. & Zapotoczny, P. Classification of Fusarium-infected and healthy wheat kernels based on features from hyperspectral images and flatbed scanner images: a comparative analysis. *Eur Food Res Technol* **244**, 1453–1462 (2018).
- Roth, L., Fossati, D., Krähenbühl, P., Walter, A. & Hund, A. Image-based phenomic prediction can provide valuable decision support in wheat breeding. *Theor Appl Genet* **136**, 162 (2023).
- Rutkoski, J. *et al.* Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat. *The Plant Genome* **5**, (2012).
- Rutkoski, J. *et al.* 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**, 2799–2808 (2016).
- Sallam, A. H., Endelman, J. B., Jannink, J.-L. & Smith, K. P. Assessing Genomic Selection Prediction Accuracy in a Dynamic Barley Breeding Population. *The Plant Genome* **8**, plantgenome2014.05.0020 (2015).
- Shahi, D. *et al.* Multi-trait genomic prediction using in-season physiological parameters increases prediction accuracy of complex traits in US wheat. *BMC Genomics* **23**, 298 (2022).

- Shahin, M. A. & Symons, S. J. Detection of Fusarium damaged kernels in Canada Western Red Spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis. *Computers and Electronics in Agriculture* **75**, 107–112 (2011).
- Shahin, M. A. & Symons, S. J. Detection of fusarium damage in Canadian wheat using visible/near-infrared hyperspectral imaging. *Food Measure* **6**, 3–11 (2012).
- signal developers. Signal:signal processing. <https://r-forge.r-project.org/projects/signal/> , <https://cran.r-project.org/web/packages/signal/index.html> (2023).
- Steiner, B. *et al.* Breeding strategies and advances in line selection for Fusarium head blight resistance in wheat. *Trop. plant pathol.* **42**, 165–174 (2017).
- Tacke, B. K. & Casper, H. H. Determination of deoxynivalenol in wheat, barley, and malt by column cleanup and gas chromatography with electron capture detection. *J AOAC Int* **79**, 472–475 (1996).
- Thapa, S. *et al.* Integrating genomics, phenomics, and deep learning improves the predictive ability for Fusarium head blight–related traits in winter wheat. *The Plant Genome* **17**, e20470.
- Togninalli, M. *et al.* Multi-modal deep learning improves grain yield prediction in wheat breeding by fusing genomics and phenomics. *Bioinformatics* **39**, btad336 (2023).
- Wang, F., Feldmann, M.J. & Runcie, D.E. Why accuracy metrics fall short in comparing phenomic and genomic prediction models. *bioRxiv*, DOI:10.1101/2025.01.09.632209 (2025).
- Wasserman, L. Bayesian Model Selection and Model Averaging. *Journal of Mathematical Psychology* **44**, 92–107 (2000).
- Wiersma, A. T. *et al.* Fine mapping of the stem rust resistance gene SrTA10187. *Theor Appl Genet* **129**, 2369–2378 (2016).
- Zhu, X., Leiser, W. L., Hahn, V. & Würschum, T. Phenomic selection is competitive with genomic selection for breeding of complex traits. *The Plant Phenome Journal* **4**, e20027 (2021)

## **CHAPTER IV: UAV-DERIVED INFORMATION AND GENOMIC DATA INTEGRATION FOR ENHANCING GRAIN YIELD PREDICTIONS IN SOFT WINTER WHEAT**

### **Abstract**

The rapid advancement of high-throughput phenotyping platforms highlights the need to integrate phenomics with genomic information to enhance the prediction accuracy of complex, economically important traits such as grain yield. From UAV-based RGB and multispectral imaging of yield trials conducted in 2022 and 2023, 30 vegetation indices (VIs) were generated, with the majority showing significant phenotypic and genetic correlations with grain yield. Model training and predictive ability assessment were done each for 2022 and 2023 trials independently, and combined trials. Higher predictive ability was observed in phenomic prediction and upon integration of UAV-information with genomic data. Forward prediction in unique genotypes in untested environment to test the utility of trained models and prediction accuracies were executed in MSU-developed breeding lines evaluated in 2024 preliminary yield trial. Incorporating UAV-derived vegetation indices individually as fixed effects resulted in increased indirect selection accuracy. However, this improvement depends on the vegetation index and its correlation with grain yield when compared to genomic prediction models without UAV-derived information and phenomic prediction solely using UAV-derived information as predictors. Incorporating multiple vegetation indices as fixed effects resulted in varying indirect selection accuracy, with using the combined trial as training set shows stable forward prediction performance. While indirect selection accuracy depends on the training set used for forward prediction, training sets using best linear unbiased estimates (BLUEs) across years appeared to be more stable and produced higher prediction accuracy compared to most cases using individual trials as training sets. Overall, the study demonstrates the potential of UAV-derived information to improve accuracy for grain yield

in soft winter wheat. Nonetheless, careful selection and evaluation of UAV-derived information for integration into genomic prediction models remain essential.

#### **4.1. Introduction**

With the global population projected to reach 9.7 billion by 2050 (United Nations) and wheat production, along with total harvest area, declining both in the United States and worldwide (FAO, 2024), the development of high-yielding wheat varieties must be accelerated with greater precision. However, phenotyping remains a major bottleneck in breeding for quantitative traits such as grain yield, due to challenges in acquiring precise and accurate phenotypic data (Araus et al., 2018). Additionally, evaluating a large number of breeding lines or selection candidates to increase selection intensity remains resource-intensive, requiring substantial land area and labor.

Given the challenges of traditional phenotyping and the constraints in evaluating a larger number of breeding lines, marker-assisted selection (MAS) has been widely used for rapid and early-generation selection of economically important traits (Lubberstedt et al., 2023). MAS has been successfully employed in improving key traits such as disease resistance (Anderson, 2007; Yet et al., 2019; Liu et al., 2020) and grain quality (Gupta et al., 2011; Vishwakarma et al., 2015) in wheat. However, MAS is most effective when the target trait is highly heritable, controlled by large-effect quantitative trait loci (QTLs), or tightly linked to specific genes (Lubberstedt et al., 2023) – criteria that do not typically apply to grain yield.

To overcome the limitations of MAS for complex traits such as yield while enabling rapid and early-generation selection (Bonnett et al., 2022), genomic selection (or genome-wide prediction) was introduced (Meuwissen et al., 2001). Unlike MAS, genomic selection leverages single nucleotide polymorphism (SNP) markers distributed across the genome to estimate genomic estimated breeding values (GEBVs) (Meuwissen et al., 2001).

Several genomic prediction models have been developed to enhance prediction accuracy. Genomic Best Linear Unbiased Prediction (GBLUP) estimates GEBVs using a mixed model framework that incorporates genomic relationship matrix derived from SNP marker data (VanRaden, 2008). Similarly, ridge regression best linear unbiased prediction (RRBLUP) considers genomic relationships rather than pedigree-based relationships, but it differs by employing ridge regression to shrink estimated marker effects toward zero, with breeding values estimated as the sum of marker effects (Endelman, 2011). Bayesian approaches, including BayesA, BayesB, BayesC, Bayesian Ridge Regression (BRR), and Bayesian Least Absolute Shrinkage and Selection Operator (LASSO), provide additional flexibility by making different assumptions about marker effect distributions, variation of marker effects, and penalization strategies (Perez and de los Campos, 2012). More recently, machine learning models such as Random Forest, extreme gradient boosting (XGBoost), and support vector machines (SVM) have been explored for genomic prediction (Sirsat et al., 2022; Farooq et al., 2023; Laurencio et al., 2024).

Despite significant advancements in genomic prediction and reduced genotyping costs, several challenges remain in accurately predicting complex traits. These include training set size and composition, marker density and quality, genetic variability within the breeding population, relatedness between training and validation populations, trait heritability, genotype-by-environment (G×E) interactions, and the quality of phenotypic data used for model training (De Roos et al., 2009; Lorenzana and Bernardo, 2009; Luan et al., 2009; Daetwyler et al., 2010; Clark et al., 2011; Howard et al., 2014; Crossa et al., 2017; Yoosefzadeh-Najafabadi et al., 2022; Montesinos-Lopez et al., 2023).

To address these limitations, phenomic data from high-throughput sensing technologies such as RGB, multispectral, and hyperspectral imaging have been incorporated into breeding

programs. These imaging platforms offer precise estimates of complex traits while capturing potential G×E interactions (Jackson et al., 2023; Kaur et al., 2024). Furthermore, spectral reflectance data collected from these sensors serve as a link between genetic variation and trait expression (Osborne, 2006; Robert et al., 2022), providing valuable insights into selection candidate performance.

While phenomic selection has traditionally relied on near-infrared spectroscopy (NIRS) for trait prediction (Rincent et al., 2018), recent studies have demonstrated the utility of RGB and multispectral imaging for this purpose (Tanebe et al., 2023; Wei et al., 2023; Winn et al., 2023). Moreover, phenomic prediction has been shown to enhance trait predictive ability compared to genomic prediction alone (Montesinos-Lopez et al., 2023; Winn et al., 2023).

Although phenomic data improves predictive ability over genomic data alone, integrating both genomic and phenomic information provides a more comprehensive framework for trait prediction by potentially capturing additive and non-additive genetic effects while linking genetics to trait expression. Several studies have demonstrated that incorporating UAV-derived phenomic data into genomic prediction improves the predictive ability of quantitative trait prediction in wheat, including grain yield (Rutkoski et al., 2016; Montesinos-Lopez et al., 2023; Kaushal et al., 2024). However, as multi-omics-based breeding remains an emerging field, further exploration is needed to optimize strategies for combining genomic and phenomic information in complex trait prediction.

This study aims to further explore the integration of UAV-derived phenomic information with genomic data to enhance grain yield estimation. Specifically, the predictive ability and indirect selection accuracy of different integration strategies, including using individual and dimensionally reduced UAV-derived vegetation indices as fixed effects in a univariate genomic

prediction model were evaluated. Through this approach, we seek to develop an improved prediction approach that leverages both genomic and phenomic data for more accurate grain yield prediction in wheat breeding.

## **4.2. Materials and Methods**

### *4.2.1. Plant materials for yield trials: 2022 - 2023*

A set of 663 soft red and soft white winter wheat genotypes were evaluated for grain yield in 2022 (n=368) and 2023 (n=383) comprised of advanced breeding lines developed by the MSU Wheat Breeding and Genetics Program, advanced breeding lines developed by other public wheat breeding programs and commercially released varieties. A subset of only 85 genotypes were evaluated in both years since most genotypes were tested as part of an active variety development program and experimental breeding lines advanced from 2022 testing to 2023 testing were included in both years.

### *4.2.2. Field establishment, maintenance, and grain yield Evaluation*

Genotypes were established in an Alpha Lattice Design with three replicates. Yield plot lengths were 3.66 m long plot with 6 rows spaced at 19.05 cm. Planting was done using Almaco HD Grain Drill (Almaco, USA) with seeding rates standardized to equate Plots were planted at 1.8 million seeds per acre. Harvesting was done using Wintersteiger Quantum Plus (Wintersteiger, Austria) and yield data was acquired using Harvest Master II (Juniper Systems, USA) weigh system and standardized to 13.5% moisture. Details of management and maintenance are presented in Supplementary Table 4.1.



#### *4.2.3. Unmanned aerial vehicle (UAV) – flight, image processing, and vegetation index calculation*

Unmanned aerial vehicle (UAV) flights were carried out at maturity stage. Natural color (RGB) imaging was carried out using Autel Evo II Pro with attached XT705 RGB camera (Autel Robotics, Shenzhen, China) with 5472x3648 pixel resolution. Flight was carried out at a height of 200 ft (60.69 m), with a side overlap of 75% and forward overlap of 85%. Multispectral imaging was carried out using DJI M210V2 (DJI, Shenzhen, China) with attached Micasense RedEdge MX multispectral camera (AgEagle Aerial Systems Inc., Kansas, USA) with 1280 x 960 pixel resolution and covers five wavebands: red (668 nm, bandwidth: 10 nm), green (560 nm, bandwidth: 20 nm), blue (475 nm, bandwidth: 20 nm), RedEdge (717 nm, bandwidth: 10 nm), and near infrared (NIR) (840 nm, bandwidth: 40 nm). Flight was carried out at a height of 200 ft (60.69 m), with a side overlap of 75% and forward overlap of 80%. Singular flight was done for each trial at Feekes 11.02 (06/21/22 for the 2022 yield trial and 06/20/2023 for the 2023 trial) and used for downstream analyses. Image processing and orthorectification of collected images were carried out using Pix4Dmapper v 4.6.4 (Pix4D Inc, Denver, CO, USA). Orthorectified images were imported in Quantum Geographic Information System. Plot identification and generation of plot boundaries were carried out using the “grid” plugin in QGIS (Chen and Zhang, 2020) and saved as ESRI shapefile. A total of 30 vegetation indices (VIs) were calculated using the formulas presented in Supplementary Table 4.2. Zonal statistics plugin in QGIS was used to extract the mean calculated VI from all pixels at the plot level.

#### *4.2.4. Statistical analyses and principal component analysis*

Plot level grain yield and VIs were spatially adjusted using SpATs (Rodriguez-Alvarez, *et al.* 2018) package in R with genotype as fixed effect, and the range and row as random effect to

account for potential spatial variation. The adjusted plot level values were used in downstream analyses.

Analysis of variance (ANOVA) was carried out to assess the variation in grain yield and VIs among the genotypes within years following the model:

$$(1) Y_{ij} = \mu + G_i + e_{ij}$$

Where  $Y_{ij}$  represents the observed grain yield response for the  $i$ -th genotype in its  $j$ -th measurement,  $\mu$  is the overall mean response across all observations,  $G_i$  represents the fixed effect of the  $i$ -th genotype, and  $e_{ij}$  accounts for the residual error associated with the response of the  $i$ -th genotype in the  $j$ -th observation. Random spatial effects including range, row, complete blocks, and incomplete blocks were not incorporated in the model due to plot level spatial effects on grain yield being corrected by the SpATs algorithm.

To account and assess for variation due to year effects and compute multi-year means, best linear unbiased estimates (BLUEs) of grain yield and VIs were modeled using “lme4” package in R (Bates et al., 2015) following the model:

$$(2) Y_{ijk} = \mu + G_i + X_j + (GX)_{ij} + e_{ijk}$$

Where,  $Y_{ijk}$  represents the response of the  $i$ -th genotype in the  $j$ -th year for the  $k$ -th observation,  $\mu$  is the overall mean across all genotypes and years,  $G_i$  is the fixed effect of the  $i$ -th genotype,  $X_j$  is the random effect of the  $j$ -th year,  $(GX)_{ij}$  represents the random interaction between the  $i$ -th genotype and  $j$ -th year, and  $e_{ijk}$  is the residual error.

Means for grain yield and VIs as well as the corresponding variance components for each trait within and across years were computed using the “lsmeans” package (Lenth, 2016). Pearson’s correlation coefficient was computed between BLUEs for grain yield and VIs. All analyses were carried out in R v4.4.0 (R Core Team, 2023).

#### 4.2.5. Feature selection

Identification of most informative UAV-derived VIs generated from the 2022 and 2023 trials and combined trials from the generated BLUEs across years were carried out using Recursive Feature Elimination following a recursive partitioning and regressive tree model using “caret” package (Kuhn, 2008) in R v4.4.0 (R Core Team, 2023).

#### 4.2.6. Heritability estimation

Heritability for grain yield and VIs based on entry means for a single year was calculated based on the following equation:

$$(3) H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_e^2}{r}}$$

Where  $H^2$  is the estimated broad sense heritability on an entry mean basis,  $r$  is the number of replicates per genotype (entry),  $\sigma_e^2$  is the error variance (residual mean square), and  $\sigma_G^2$  is the genotype (entry) variance.

Heritability based on entry means across years was calculated based on the following equation:

$$(4) H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GT}^2}{r} + \frac{\sigma_e^2}{rT}}$$

Where  $H^2$  is the estimated broad sense heritability on an entry mean basis,  $r$  is the number of replicates per genotype (entry),  $T$  is the number of years,  $\sigma_e^2$  is the error variance (residual mean square), and  $\sigma_G^2$  is the genotype (entry) variance,  $\sigma_{GT}^2$  is the genotype by year variance.

#### 4.2.7. Genetic correlation estimation

Genetic correlations ( $r_g$ ) to further assess the relationships between grain yield and each of the VIs in “sommer” package in R (Covarrubia-Parazan, 2016) following the framework:

$$(5) y_1 = X_1\beta_1 + Z_1\mu_1 + e_1$$

$$(6) y_2 = X_2\beta_2 + Z_2\mu_2 + e_2$$

Where  $y_1$  and  $y_2$  are vectors of phenotypic values for trait 1 (grain yield) and trait 2 (VI), respectively,  $X_1$  and  $X_2$  are incidence matrices for fixed effects,  $\beta_1$  and  $\beta_2$  are vectors of fixed effects for each trait,  $Z_1$  and  $Z_2$  are incidence matrices for random effects of the genotypes,  $\mu_1$  and  $\mu_2$  are genetic effects for each trait which are assumed to follow a multivariate normal distribution with covariance structure based on genomic relationship matrix,  $e_1$  and  $e_2$  are residual error terms for each trait.

The genetic covariance structure, assuming that random genetic effects  $\mu_1$  and  $\mu_2$  have a covariance structure following the model:

$$(7) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \sim N \left( 0, \begin{pmatrix} G_{11} & A & G_{12} & A \\ G_{12} & A & G_{22} & A \end{pmatrix} \right)$$

Where  $G_{11}$  and  $G_{22}$  are the genetic variances for trait 1 (grain yield) and trait 2 (VI), respectively,  $G_{12}$  is the genetic covariance between traits 1 and 2,  $A$  is the genomic relationship matrix generated from SNP marker data.

The genetic correlation ( $r_G$ ) between trait 1 (grain yield) and trait 2 (VI) is then defined by the formula:

$$(8) r_G = \frac{G_{12}}{\sqrt{(G_{11})(G_{22})}}$$

Where  $G_{12}$  is the covariance grain yield and VI,  $G_{11}$  is the genetic variance of grain yield,  $G_{22}$  is the genetic variance of the VI.

#### 4.2.8. Principal component analysis (PCA) of VIs calculated

Principal Component Analysis was carried out for the 30 generated VIs to dimensionally reduce the phenomic data generated and identify the most important VI potentially contributing to

the variation observed among the genotypes. BLUEs from spatially adjusted VIs were used. Principal component analysis was carried out in base R v4.4.0 (R Core Team, 2023).

#### *4.2.9. DNA isolation and SNP genotyping*

DNA was extracted from all genotypes following the protocol of Wiersma et al. (2016). Genotyping-by-sequencing (GBS) libraries were prepared as described by Poland et al. (2012), scaled to a 24 µL volume in a 384-well format. The libraries were sequenced at 384-plex on an Illumina HiSeq 4000 system. Single nucleotide polymorphisms (SNPs) were identified using the TASSEL 5 GBS pipeline (Glaubitz et al., 2014). Sequencing reads were aligned to the wheat reference genome assembly (RefSeq v2.0) from the International Wheat Genome Sequencing Consortium, using default parameters. In the GBSSeqToTagDBPlugin and ProductionSNPCallerPluginV2 steps, a k-mer length of 64 base pairs was used, with a minimum coverage of five reads per k-mer. Default settings were applied for all other steps. SNPs were filtered based on a 0.70 call rate and a minor allele frequency (MAF) of 0.05, with 15,456 SNPs remaining after filtering.

#### *4.2.10. Genome wide association mapping*

Genome Wide Association Studies (GWAS) were carried out using the Bayesian-information and Linkage Disequilibrium Iteratively Nested Keyway (BLINK) (Huang et al., 2019) model in GAPIT v3 (Genomic Association and Prediction Integrated Tool) (Lipka et al., 2012; latest version: March 12, 2022). The VIs and grain yield were used as phenotypic input for GWAS. To address potential population structure, principal components were used, and the number of principal components used were adjusted based on review of QQ plots.

#### 4.2.11. Univariate Genomic Predictions

Genomic prediction for grain yield was assessed using BayesA, BayesB, BayesC, Bayesian Ridge Regression (BRR), and Bayesian LASSO (BayesL) with 50,000 iterations and a burn-in set at 5,000, using the “BGLR” package (Perez and de los Campos, 2014). Additionally, Ridge Regression Best Linear Unbiased Prediction (RRBLUP) was performed using the “rrblup” package (Endelman, 2011), Extreme Gradient Boosting (XGBoost) was conducted using the “xgboost” package (Chen and Guestrin, 2016) with parameters eta (learning rate) = 0.01, gamma = 0.2, max\_depth (maximum tree depth) = 6, min\_child\_weight (minimum sum of instance weights) = 0.20, subsample = 1, and colsample\_bytree (fraction of features (columns) used to grow each tree) = 1, and Random Forest was utilized with the “caret” package (Kuhn, 2008) with default parameters (ntree = 500, nodesize = 5, maxnodes = unlimited) for prediction. A total of 100 individual five-fold cross-validations were carried out, with 80% of the genotypes designated in the training set for all models. The predictive ability of genomic prediction for all models was computed as the average Pearson’s correlation between the actual grain yield and the predicted values from all 100 cycles in each model. All cross-validations were performed in R version 4.4.0 (R Core Team, 2023).

#### 4.2.12. Univariate genomic prediction with selected VIs as fixed effects

To assess how fixed effect covariates influence grain yield prediction accuracy, Ridge Regression Best Linear Unbiased Prediction (RRBLUP) was used as the base genomic prediction model following the model:

$$(9) y = X\beta + Zu + \varepsilon$$

Where  $y$  is the vector of phenotypic observations ( $n \times 1$ ), in this case grain yield,  $X$  is the matrix of fixed effects or covariates ( $n \times p$ ),  $\beta$  is the vector of fixed effect coefficients ( $p \times 1$ ),  $Z$  is

the matrix linking observations to genotypes ( $n \times g$ ),  $\mu$  is the vector of random genetic effects ( $g \times 1$ ) assuming  $\mu \sim N(0, \sigma_u^2 K)$ , where  $K$  is the genomic relationship matrix, and  $\epsilon$  vector of residuals.

Fourteen VIs were selected based on the results of feature importance scores from feature selection using recursive partitioning and regressive tree model. The VIs was used individually to evaluate how each VIs affects grain yield prediction, and with increments to assess how increasing the number of VIs used as fixed effect affects grain yield prediction.

#### 4.2.13. Forward prediction and indirect selection accuracy assessment

To test the utility of various approaches and models evaluated in an untested environment and unique genotypes, forward prediction of grain yield was executed in MSU-developed breeding lines ( $n = 216$ ) in the 2024 preliminary yield trial. The 2024 trial follows experimental design, field establishment, and grain yield evaluation conducted in the 2022 and 2023 trials. Additional details on field management and maintenance are presented in Supplementary Table 4.1. In addition, DNA isolation and SNP genotyping follows the same protocol as 2022 and 2023 trials.

The phenotypic correlation between the predicted grain yield in the 2024 breeding lines and the actual grain yield was done using Pearson's correlation in R. Calculation of genetic correlation and broad sense heritability follows the formula presented earlier. Genetic variance of predicted and actual grain yield and its covariance was determined by fitting in RRBLUP following the model:

$$(10) y = Zu + \epsilon$$

Where  $y$  is the vector of phenotypic values (predicted or actual grain yield),  $Z$  is the design matrix for random genetic effects of the genotypes,  $\mu$  is the vector of random genetic effects following the assumption that  $\mu \sim N(0, K\sigma_g^2)$ , where  $K$  is the genomic relationship matrix,  $\sigma_g^2$  is the genetic variance, and  $e$  is the vector of residual errors following the assumption that  $e \sim$

$N(0, I\sigma_e^2)$ , where  $I$  is the identity matrix, and  $\sigma_e^2$  is the residual variance. The variance-covariance matrix follows the model:

$$(11) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \sim N\left(0, \begin{pmatrix} G_{11} & G_{12} \\ G_{12} & G_{22} \end{pmatrix} \otimes A\right)$$

Where  $\mu_1$  is the genetic values of the actual grain yield of the breeding lines,  $\mu_2$  is the genetic values of the predicted grain yield,  $G$  is the genetic variance-covariance matrix,  $G_{11}$  is the genetic variance of actual grain yield,  $G_{22}$  is the genetic variance of the predicted grain yield, and  $G_{12}$  is the genetic covariance between the actual and predicted grain yield.

Indirect selection accuracy,  $Acc(I)$  representing prediction accuracy, was computed based on the formula adapted from Lopez-Cruz et al. (2020):

$$(12) Acc(I) = r_G(H_I)$$

Where  $r_G$  is the genetic covariance between the actual (X) and predicted (I) grain yield and  $H_I$  is the square root of the estimated heritability of the predicted grain yield.

### 4.3. Results

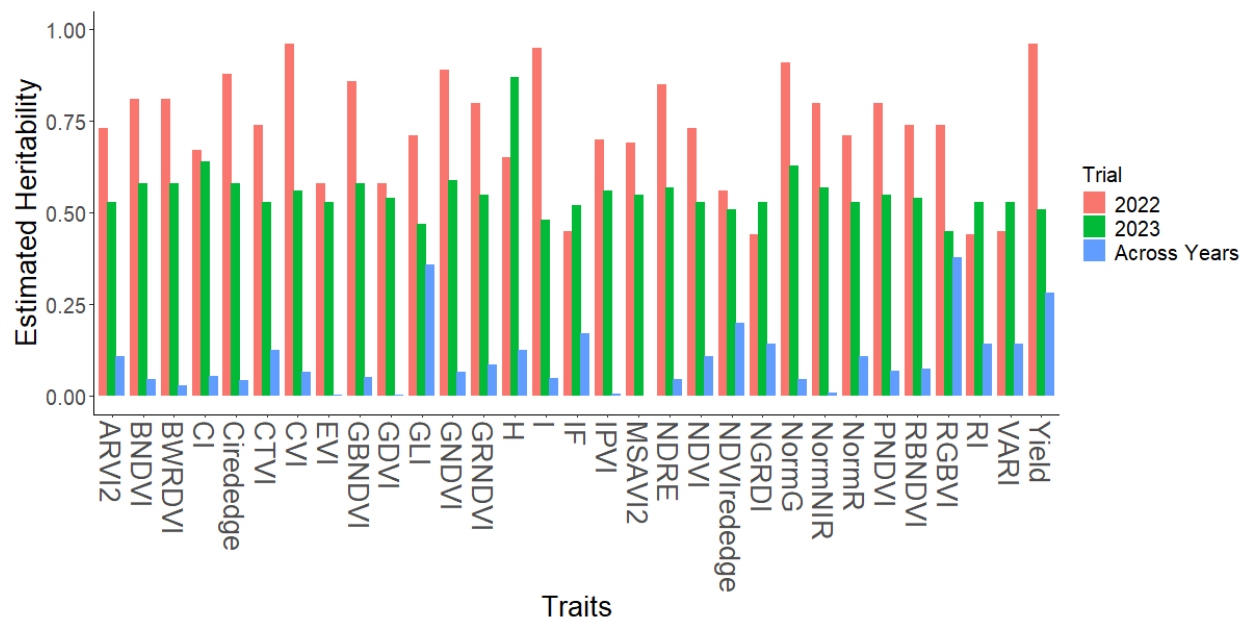
#### 4.3.1. Heritability and phenotypic variance component estimation

Relatively high broad sense heritability ranging from 0.44 to 0.96 was estimated for the majority of the VIs generated in 2022 while grain yield demonstrated the highest heritability at 0.96 (Figure 4.1). The VIs CVI and I demonstrated heritabilities similar to grain yield (Figure 4.1). The high estimated heritability of grain yield could be attributed to the high percentage of variance explained by genotype effects and accurate accounting for spatial variation across the trial, with similar observations in CVI and I (Supplementary Figure 4.1). Furthermore, the genotype effects accounted for 40% to 90% of the phenotypic variation observed among VIs (Supplementary Figure 4.1).



Heritability observed for grain yield in 2023 was 0.51, whereas the majority of the VIs demonstrated heritability higher than grain yield, ranging from 0.45 to 0.87, with H and CI having the highest estimated heritability at 0.87 and 0.64, respectively (Figure 4.1). The lower heritability for grain yield in 2023 could be attributed to relatively higher error (residual) variance, accounting for 50% to 60% of the phenotypic variation observed, with the exemption of the VI H where more than 80% of the phenotypic variation observed is due to genotype effect (Supplementary Figure 4.1). However, it is also worth noting that most of the genotypes evaluated in 2023 were not evaluated in the 2022 trial due to the trials being part of an active winter wheat breeding program. Further, 2022 and 2023 have relatively different weather conditions. Therefore, direct comparisons of heritability and variance components between 2022 and 2023 trials could be misleading.

However, the heritability of BLUEs generated across years for all trait was considerably lower for grain yield (0.283) and all VIs (0.001 to 0.378) (Figure 4.1). Only two VIs, RGBVI and



**Figure 4.1.** Estimated broad sense heritabilities of grain yield and vegetation indices in 2022 and 2023 trials separately, and across years.

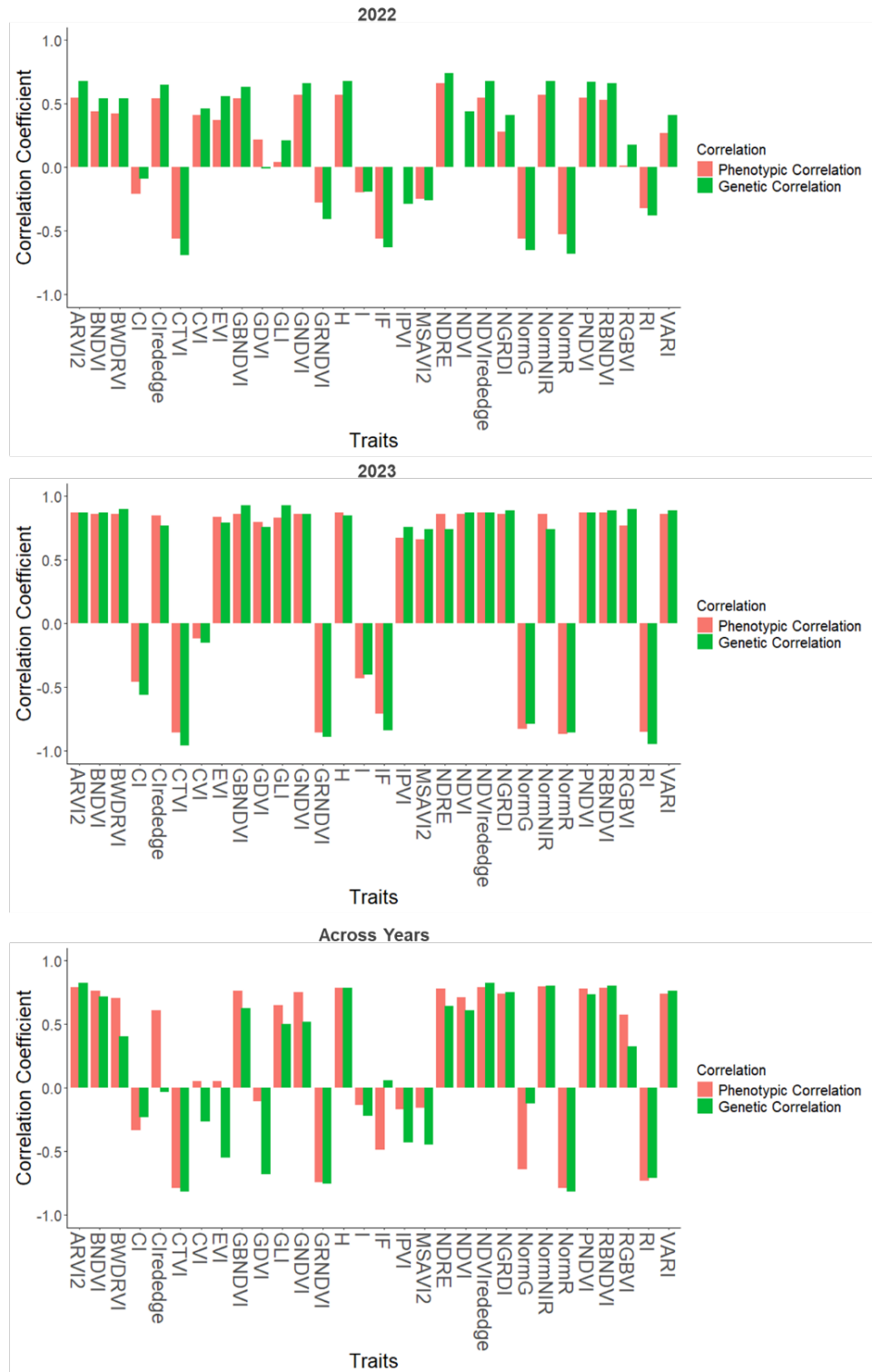
GLI, had higher estimated heritability than grain yield at 0.378 and 0.358, respectively (Figure 4.1). The substantial reduction in heritabilities across years can be attributed to year effects (Supplementary Figure 4.1) indicating that traits responded differently to the environmental conditions present in 2022 compared to 2023. For several of the VIs, almost all of the phenotypic variance was entirely due to year variance (Supplementary Figure 4.1). In VIs RGBVI and GLI, which demonstrated the highest heritability across years, the variance explained by year effect was very low, while variance explained by genotype, genotype by year interaction, and residual effect were relatively similar (Supplementary Figure 4.1).

#### *4.4.2 Phenotypic and genotypic correlation between VIs and grain yield*

In the 2022 trial, 25 of the VIs had significant phenotypic correlations with grain yield (Figure 4.2). Of these, 9 demonstrated negative phenotypic correlations with grain yield (-0.56 to -0.20). Higher genetic correlations (compared to phenotypic correlations) with grain yield were observed for 21 vegetation indices (Figure 4.2).

The majority of VIs generated in the 2023 trial demonstrated high phenotypic correlations with grain yield. All of the VIs had significant correlations with grain yield (Figure 4.2), of which 17 VIs recorded higher genetic correlations than phenotypic correlations with grain yield (Figure 4.2).

Trends of phenotypic and genetic correlations with grain yield varied among BLUEs from the combined 2022 and 2023 trials (Figure 4.2). Five VIs had higher positive genetic correlations with grain yield compared to individual years, whereas 13 VIs had higher positive phenotypic correlations with grain yield (Figure 4.2). Interestingly, the VIs Clrededge, CVI, and EVI had positive phenotypic correlation but negative genetic correlations with grain yield, which in contrast with that of IF (Figure 4.2).



**Figure 4.2.** Genetic and phenotypic correlations of the vegetation indices with grain yield in 2022 2023 and across year.

Altogether, 15 VIs demonstrated consistent positive phenotypic and genetic correlation with grain yield within and across years, while six VIs demonstrated consistent negative correlations (Figure 4.2). It is worth noting that CVI had positive phenotypic and genetic correlation with grain yield in 2022 and negative correlation in 2023 (Figure 4.2). In contrast, IPVI and MSAVI2 had negative phenotypic and genetic correlation in 2022 and positive correlation in 2023 (Figure 4.2).

#### *4.4.3. Principal component analysis of VIs*

Principal component analysis (PCA) was carried out for dimensionality reduction of VIs and to identify VIs contributing most to phenotypic variation observed. Grain yield was not included in PCA, therefore the variation observed and explained was solely from the VIs. The first three principal components (PC) explain 94.1% of the variation in VIs observed among the genotypes in the 2022 trial, with PC1 explaining 63.2% of the variance (Supplementary Figure 4.2). A total of 21 VIs contributed positively to the observed variation, with RBNDVI, ARVI2, and NDVI having the most substantial contributions (Supplementary Figure 4.3).

In the 2023 trial, the first three PCs explain 94.7% of the variation the VIs observed among the genotypes in the 2023 trial, with PC1 explaining 85.2% of the variance observed (Supplementary Figure 4.2). A total of 21 VIs contributed positively to the observed variation, with RBNDVI, PNDVI, and NDVI having the most substantial contributions (Supplementary Figure 4.3).

Similar results were observed when combining vegetative indices across years. The first three PCs explain 89.8% of the variation observed in multi-year indices, slightly lower than individual years, with PC1 explaining 67.7% of the variations observed (Supplementary Figure

4.2). A total of 20 VIs contributed positively to variation in VIs, with GRNDVI, RBNDVI, and PNDVI having the most substantial contributions (Supplementary Figure 4.3).

#### 4.4.4. Feature selection

In the 2022 trial, NDRE, NormG, GNDVI, GRNDVI, Clrededge, and NormNIR were identified as the most important VIs associated with grain yield using recursive partitioning and regressive tree model (Table 4.1). From the 2023 trial, NormNIR, Clrededge, NDVI, ARVI2, GRNDVI, NGRDI, RI, VARI, and CTVI were identified as the most important VIs associated with grain yield (Table 4.1) Across years, NormNIR, GRNDVI, NDRE, CTVI, NormR, PNDVI, NDVI, and ARVI2 were identified as the most important VIs associated with grain yield (Table 4.1).

Of all the VI selected, GRNDVI and NormNIR was selected consistently, whereas Clrededge was only selected in both 2022 and 2023 trials, but not in the combined trials. NDRE

**Table 4.1.** Feature importance scores of selected wavebands using recursive partitioning and regressive tree model from individual trial (2022, 2023) s and BLUEs from combined trials (2022-2023). Empty cells mean the VI was not selected for that year/trial.

Feature Selected VIs	Feature Importance Score			
	2022	2023	Combined Trials	Sum
GRNDVI	74.72	61.93	97.2	233.85
NormNIR	22.79	100	100	222.79
NDRE	100	-	59.73	159.73
Clrededge	50.28	62.11	-	112.39
NDVI	-	62.09	38.69	100.78
ARVI2	-	62.09	36.69	98.78
CTVI	-	37.91	58.4	96.31
NormG	75.84	-	-	75.84
GNDVI	74.72	-	-	74.72
NormR	-	-	57.53	57.53
PNDVI	-	-	38.78	38.78
NGRDI	-	38.39	-	38.39
RI	-	38.39	-	38.39
VARI	-	38.39	-	38.39

was selected in from the 2022 trial and using the combined trial, whereas ARVI2, CTVI, GRNDVI, and NDVI were all selected from the 2023 trial and using combined trials. NormG, GNDVI, NGRDI, RI, and VARI were all selected only once either from the 2022 trial or the 2023 trial. Interestingly, NormR and PNDVI were not selected from individual trials, but was selected when the trials were combined.

These fourteen VIs appear to be good predictors of grain yield based on feature selection using recursive partitioning and regressive model, arranged in descending order of importance and consistency of selection based on total importance score: GRNDVI, NormNIR, NDRE, Clrededge, NDVI, ARVI2, CTVI, NormG, GNDVI, PNDVI, NGRDI, RI and VARI.

#### *4.4.5. Grain Yield MTAs*

Genome wide association was carried out to identify significant MTAs potentially associated with grain yield. Seven MTAs were mapped in chromosomes 1B, 3A, 5B, 6B, and 7A, however, the identified MTAs were not consistent (Supplementary Table 4.7). The MTAs S3A\_1065055 in chromosome 3A, S5B\_607526225 in chromosome 5B, and S7A\_54345139 in chromosome 7A, explaining 6.35%, 11.05%, and 31.05% of the phenotypic variance explained respectively, were only mapped from the 2022 trial (Supplementary Table 4.7). The MTAs S1B\_686820474 mapped in chromosome 1B and S3A\_532458794 in chromosome 3A, explaining 17.36% and 7.13% of the phenotypic variance observed respectively, were only mapped from the 2023 trial (Supplementary Table 4.7). Interestingly, the MTAs S3A\_551647066 mapped in chromosome 3A and S6B\_163179890 in chromosome 6B, explaining 8.66% and 6.18% of the phenotypic variance observed respectively, were mapped using the BLUEs across years, but were not identified from either of the 2022 or 2023 trials alone (Supplementary Table 4.7).

#### *4.4.6. MTAs associated with VIs*

Since the VIs are essentially phenotypes, genome wide association was conducted for all the 30 VIs for 2022 trial, 2023 trial, and combined 2022-2023 trials. A total of 88 MTAs were mapped in different VIs and trials (Supplementary Table 4.7). A total of 17 MTAs were found to be putatively associated with the VI Hue (H), distributed in various chromosomes (Supplementary Table 4.7). Interestingly, the MTAs associated with this VI was not associated or mapped with any other vegetation indices, whereas several MTAs identified is associated to multiple vegetation indices (Supplementary Table 4.7). Most notably, the 1A SNP at 295.5Mb and the 1D SNP at 0.5Mb are associated with most VIs, both of which were mapped from 12 different VIs, respectively (Supplementary Table 4.7). In addition, several mapped MTAs were mapped using BLUEs from combined trials and in the 2022 trial: S1A\_312552669 associated with EVI, S1D\_38087525 associated with VIs GDVI, IPVI, and MSAVI, and S1D\_464487 associated with VIs BWDRVI, CIrededge, GBNDVI, and NormG (Supplementary Table 4.7). However, none of the MTAs mapped from 2023 trial were mapped in the 2022 trial or in the combined trials.

#### *4.4.7. Coincident MTA in chromosome 7A*

A coincidental MTA, S7A\_54345139 in chromosome 7A, was found to be associated with grain yield and with VIs NDRE and IF (Supplementary Table 4.7). This MTA could be considered a large effect MTA for grain yield as it explains 31.05% of the variation observed in grain yield in the 2022 trial (Supplementary Table 4.7). The MTA also explains 15.69% and 7.03% of the phenotypic variation observed in NDRE and IF in the 2022 trial (Supplementary Table 4.7). Interestingly, when looking at the grain yield of the genotypes bearing this allele, we found a significant difference between genotypes bearing homozygous G and homozygous T alleles in the

2022 trial, 2023 trial, and combined trials. However, the spread or variation in grain yields are still wide and not clearcut (Supplementary Figure 4.5).

#### 4.4.8. Predictive Ability of Univariate Genomic and Phenomic Predictions

Predictive ability assessment was carried out in a cross-validation scheme within training sets using a 80/20 training and testing set split. Each of the 2022 and 2023 trials were considered as separate training sets, and the combined trials was considered as another separate training set. Predictive ability was determined by calculating the average phenotypic Pearson's correlation between the predicted and actual grain yield in the testing set in 100 cycles of cross validation. Eight models, including five Bayesian models, ridge regression best linear unbiased prediction (RRBLUP) and two machine learning models, were tested and compared.

Significant variation for predictive ability was identified among genomic prediction models ( $p = 4.99 \times 10^{-11}$ ) in the 2022 trial, with BayesC having significantly higher predictive ability

**Table 4.2.** Comparison of predictive abilities of genomic prediction and phenomic prediction derived from cross-validation using 2022 trial, 2023 trial, and combined 2022 and 2023 trials.

Model	2022		2023		Combined Trials	
	Genomic <sup>#</sup>	Phenomic <sup>#</sup>	Genomic <sup>#</sup>	Phenomic <sup>#</sup>	Genomic <sup>#</sup>	Phenomic <sup>#</sup>
BayesA	0.402 ± 0.082 <sup>b</sup>	0.709 ± 0.065 <sup>b</sup>	0.277 ± 0.068 <sup>ab</sup>	0.864 ± 0.040 <sup>ab</sup>	0.298 ± 0.124 <sup>ab</sup>	0.862 ± 0.041 <sup>cd</sup>
BayesB	0.429 ± 0.089 <sup>cde</sup>	0.719 ± 0.058 <sup>b</sup>	0.305 ± 0.086 <sup>bc</sup>	0.868 ± 0.036 <sup>ab</sup>	0.313 ± 0.112 <sup>bc</sup>	0.862 ± 0.039 <sup>cd</sup>
BayesC	0.444 ± 0.086 <sup>e</sup>	0.719 ± 0.060 <sup>b</sup>	0.303 ± 0.088 <sup>bc</sup>	0.865 ± 0.035 <sup>ab</sup>	0.309 ± 0.122 <sup>bc</sup>	0.866 ± 0.036 <sup>cd</sup>
BL	0.422 ± 0.095 <sup>bcd</sup>	0.717 ± 0.067 <sup>b</sup>	0.307 ± 0.086 <sup>bc</sup>	0.867 ± 0.031 <sup>ab</sup>	0.301 ± 0.126 <sup>ab</sup>	0.870 ± 0.029 <sup>d</sup>
BRR	0.437 ± 0.095 <sup>de</sup>	0.709 ± 0.070 <sup>b</sup>	0.289 ± 0.087 <sup>abc</sup>	0.866 ± 0.033 <sup>ab</sup>	0.304 ± 0.110 <sup>abc</sup>	0.861 ± 0.043 <sup>c</sup>
RF	0.408 ± 0.098 <sup>bc</sup>	0.641 ± 0.057 <sup>a</sup>	0.316 ± 0.073 <sup>c</sup>	0.873 ± 0.030 <sup>b</sup>	0.352 ± 0.014 <sup>d</sup>	0.803 ± 0.027 <sup>a</sup>
RRBLUP	0.356 ± 0.089 <sup>a</sup>	0.712 ± 0.060 <sup>b</sup>	0.284 ± 0.078 <sup>ab</sup>	0.872 ± 0.026 <sup>ab</sup>	0.283 ± 0.109 <sup>a</sup>	0.814 ± 0.028 <sup>b</sup>
XGBoost	0.418 ± 0.096 <sup>bcd</sup>	0.649 ± 0.063 <sup>a</sup>	0.268 ± 0.081 <sup>a</sup>	0.863 ± 0.027 <sup>a</sup>	0.325 ± 0.101 <sup>c</sup>	0.805 ± 0.028 <sup>ab</sup>

BL – Bayesian LASSO, BRR – Bayesian Ridge Regression, RF – Random Forest, RRBLUP – Ridge Regression Best Linear Unbiased Prediction, XGBoost – Extreme Gradient Boosting

#Means with the same letters are not significantly difference at alpha 0.05. Comparisons are within columns.



(0.444), and RRBLUP having the lowest predictive ability (0.356) (Table 4.2). In the 2023 trial, significant variation in predictive ability was also observed among the models (p.value: 0.03566), with RF having the highest predictive ability (0.316) and xgBoost having the lowest predictive ability (0.268) (Table 4.2). Significant variation among the models was also observed in the combined 2022 and 2023 trials as a training set (p.value:  $1.53 \times 10^{-07}$ ) with RF having significantly highest predictive ability (0.352), whereas RRBLUP showed significantly lower predictive ability (0.283) (Table 4.2). Across all the training sets, RRBLUP showed consistently low predictive ability. In addition, using the 2022 trial as training set resulted in higher predictive ability compared to using 2023 trial or using combined trials (Table 4.2).

The same models for genomic prediction were used for phenomic prediction, where all the 30 VIs generated were used as predictors in place of genomic information (SNPs). Significant variation in predictive ability was observed with the 2022 trial (p.value:  $8.75 \times 10^{-33}$ ). The Bayesian models and RRBLUP showed no significant differences while BayesB and BayesC demonstrated the highest predictive ability (0.719) and the machine learning models showing significantly lower predictive ability (Table 4.2). No significant variation in predictive ability was observed among models in the 2023 trial (p.value: 0.32173), however, RF demonstrated the highest predictive ability (0.873) (Table 4.2). In the combined trials, BL demonstrated the highest predictive ability (0.87), while machine learning models RF and xgBoost demonstrated the lowest predictive ability (Table 4.2). Overall, 2022 trial had lower predictive abilities compared to 2023 trial and the combined trials (Table 4.2). Surprisingly, machine learning models RF and xgBoost showed relatively lower predictive ability compared to Bayesian models, which is most evident in the 2022 and in the combined trials (Table 4.2).

Phenomic predictions showed higher predictive ability compared to genomic predictions regardless of the training set and model used (Table 4.2). The higher predictive ability of phenomic predictions could be accounted by the correlation of the VIs with grain yield. Comparing the training sets, 2022 trial recorded higher predictive ability using genomic information, which is in contrast with phenomic prediction where the 2022 trial recorded lower overall predictive ability (Table 4.2). Notably, RRBLUP showed poor predictive ability using genomic information, whereas it showed relatively high predictive ability using phenomic information. Bayesian models showed consistently good predictive abilities in both genomic and phenomic predictions (Table 4.2).

#### *4.4.9. Predictive ability of univariate genomic prediction with feature selected VIs as covariates*

To assess the impact of incorporating VIs as fixed effects in genomic predictions, the 13 feature selected VIs were all tested individually as fixed effects. RRBLUP was chosen as the base genomic prediction model due to consistently poor genomic predictive ability compared with the other models.

Significant variation in genomic predictive ability was identified among VIs used as fixed effect in the 2022 trial (p.value:  $<0.00001$ ). Incorporating NDRE as fixed effect in the RRBLUP model demonstrated the highest predictive ability (0.66) and improved the accuracy relative to the base model. Incorporating RI, VARI and NGRDI decreased the predictive ability relative to the base model (Table 4.3).

Similarly, significant variation was identified using VIs as a fixed effect (p.value:  $3.67 \times 10^{-35}$ ) in the 2023 trial, however the differences among models were minor (0.828 to 0.875) (Table 4.3). The highest predictive ability was observed using GRNDVI, PNDVI, NDVI, NormR, and

**Table 4.3.** Comparison of predictive abilities of genomic prediction using RRBLUP with selected VIs used individually as fixed effect derived from cross-validation using 2022 trial, 2023 trial, and combined 2022 and 2023 trials.

VI	2022 <sup>#</sup>	2023 <sup>#</sup>	Combined Trials <sup>#</sup>
ARVI2	0.543 ± 0.069 <sup>cd</sup>	0.875 ± 0.026 <sup>d</sup>	0.791 ± 0.024 <sup>fg</sup>
CIrededge	0.543 ± 0.063 <sup>cd</sup>	0.857 ± 0.034 <sup>b</sup>	0.613 ± 0.041 <sup>a</sup>
CTVI	0.552 ± 0.068 <sup>cde</sup>	0.869 ± 0.032 <sup>cd</sup>	0.790 ± 0.026 <sup>fg</sup>
GNDVI	0.573 ± 0.062 <sup>e</sup>	0.862 ± 0.028 <sup>bc</sup>	0.751 ± 0.029 <sup>d</sup>
GRNDVI	0.566 ± 0.064 <sup>de</sup>	0.875 ± 0.031 <sup>d</sup>	0.786 ± 0.027 <sup>ef</sup>
NDRE	0.661 ± 0.058 <sup>f</sup>	0.869 ± 0.032 <sup>cd</sup>	0.784 ± 0.028 <sup>ef</sup>
NDVI	0.543 ± 0.068 <sup>cd</sup>	0.873 ± 0.031 <sup>d</sup>	0.792 ± 0.026 <sup>fg</sup>
NGRDI	0.268 ± 0.087 <sup>b</sup>	0.864 ± 0.026 <sup>bc</sup>	0.741 ± 0.030 <sup>c</sup>
NormG	0.561 ± 0.062 <sup>de</sup>	0.828 ± 0.034 <sup>a</sup>	0.641 ± 0.037 <sup>b</sup>
NormNIR	0.566 ± 0.064 <sup>de</sup>	0.869 ± 0.027 <sup>cd</sup>	0.795 ± 0.025 <sup>g</sup>
NormR	0.530 ± 0.070 <sup>c</sup>	0.873 ± 0.031 <sup>d</sup>	0.789 ± 0.026 <sup>fg</sup>
PNDVI	0.549 ± 0.066 <sup>cde</sup>	0.874 ± 0.031 <sup>d</sup>	0.779 ± 0.028 <sup>e</sup>
RI	0.158 ± 0.087 <sup>a</sup>	0.864 ± 0.026 <sup>bc</sup>	0.747 ± 0.030 <sup>c</sup>
VARI	0.163 ± 0.099 <sup>a</sup>	0.859 ± 0.026 <sup>b</sup>	0.747 ± 0.037 <sup>c</sup>

<sup>#</sup>Means with the same letters are not significantly different at  $\alpha=0.05$ . Comparisons are within columns.

ARVI with no significant differences observed in the predictive abilities of these VIs ranging from 0.87 to 0.88 (Table 4.3).

Using the combined 2022 and 2023 trials, significant variation was also observed in the genomic predictive abilities from incorporating VIs as a fixed effect (p.value: <0.00001). Incorporating NormNIR resulted in the highest predictive ability (0.795), with the lowest when CIrededge was incorporated (0.613) (Table 4.2).

Incorporating VIs as fixed effect resulted in the highest predictive abilities in the 2023 trial compared to the 2022 trial and in the combined trials (Table 4.3). This is likely due to the higher correlation of the VIs with grain yield in 2023. Incorporation of GRNDVI, NDRE, and NormNIR

in RRBLUP as fixed effect resulted in high predictive ability, whereas incorporating VARI and Clrededge consistently resulted in poor predictive ability regardless of the training set used (Table 4.3). Overall, predictive ability increases when feature selected VIs are incorporated in RRBLUP as fixed effects compared to the base model that includes only genomic information (Table 4.3).

#### 4.4.10. Predictive ability of univariate genomic prediction incorporating multiple vegetative indices as covariates

To determine if including additional VIs to the base RRBLUP model improve predictive ability, GRNDVI was included as a common fixed effect and an increasing number of VIs were added as fixed effects. Predictive ability of the base model decreased with the addition of two to

**Table 4.4.** Comparison of predictive abilities of genomic prediction using RRBLUP with multiple and increasing number of VIs as fixed effect derived from cross-validation using 2022 trial, 2023 trial, and combined 2022 and 2023 trials.

Number of VIs	2022 <sup>#</sup>	2023 <sup>#</sup>	Combined Trials <sup>#</sup>	VIs Incorporated as Covariate
1	0.53 ± 0.198 <sup>d</sup>	0.87 ± 0.026 <sup>ef</sup>	0.79 ± 0.027 <sup>g</sup>	GRNDVI
2	0.61 ± 0.162 <sup>c</sup>	0.88 ± 0.036 <sup>f</sup>	0.77 ± 0.035 <sup>fg</sup>	GRNDVI, NDRE
3	0.56 ± 0.171 <sup>d</sup>	0.87 ± 0.043 <sup>ef</sup>	0.77 ± 0.052 <sup>fg</sup>	GRNDVI, NDRE, NDVI
4	0.55 ± 0.162 <sup>d</sup>	0.85 ± 0.047 <sup>de</sup>	0.75 ± 0.061 <sup>ef</sup>	GRNDVI, NDRE, NDVI, CTVI
5	0.51 ± 0.179 <sup>cd</sup>	0.84 ± 0.055 <sup>d</sup>	0.75 ± 0.060 <sup>de</sup>	GRNDVI, NDRE, NDVI, CTVI, NormG
6	0.48 ± 0.188 <sup>bc</sup>	0.80 ± 0.105 <sup>c</sup>	0.73 ± 0.067 <sup>cd</sup>	GRNDVI, NDRE, NDVI, CTVI, NormG, Clrededge
7	0.46 ± 0.190 <sup>abc</sup>	0.78 ± 0.109 <sup>bc</sup>	0.72 ± 0.071 <sup>bc</sup>	GRNDVI, NDRE, NDVI, CTVI, NormG, Clrededge, PNDVI
8	0.44 ± 0.192 <sup>ab</sup>	0.77 ± 0.112 <sup>ab</sup>	0.71 ± 0.075 <sup>ab</sup>	GRNDVI, NDRE, NDVI, CTVI, NormG, Clrededge, PNDVI, NGRDI
9	0.41 ± 0.198 <sup>a</sup>	0.75 ± 0.120 <sup>a</sup>	0.70 ± 0.080 <sup>a</sup>	GRNDVI, NDRE, NDVI, CTVI, NormG, Clrededge, PNDVI, NGRDI, VARI

<sup>#</sup>Means with the same letters are not significantly difference at alpha 0.05. Comparisons are within columns.

**Table 4.5.** Comparison of predictive abilities of genomic prediction using RRBLUP with increasing number of principal components (PCs) as fixed effect derived from cross-validation using 2022 trial, 2023 trial, and combined 2022 and 2023 trials.

Number of PCs	2022 <sup>#</sup>	2023 <sup>#</sup>	Combined Trials <sup>#</sup>
1	0.516 ± 0.069 <sup>b</sup>	0.870 ± 0.026 <sup>b</sup>	0.780 ± 0.027 <sup>b</sup>
2	0.465 ± 0.143 <sup>a</sup>	0.841 ± 0.055 <sup>a</sup>	0.760 ± 0.045 <sup>a</sup>
3	0.472 ± 0.168 <sup>a</sup>	0.842 ± 0.047 <sup>a</sup>	0.756 ± 0.76 <sup>a</sup>

<sup>#</sup>Means with the same letters are not significantly difference at alpha 0.05. Comparisons are within columns.

nine VIs. The trend of decreasing predictive ability as VIs are added was observed for the 2022 ( $p=2.90 \times 10^{-17}$ ), 2023 ( $p=6.54 \times 10^{-51}$ ) and combined years ( $p=1.45 \times 10^{-30}$ ).

In order to take advantage of the information in all of the VIs simultaneously rather than relying solely on one VI or few selected VIs, we added one, two, or three PCs as fixed effects in RRBLUP. Consistent with the observation of decreased predictive ability by increasing the number of VIs as fixed effect in RRBLUP (Table 4.4), the predictive ability decreased as the number of PCs increased in the 2022 ( $p=0.0158$ ), 2023 ( $p=1.814 \times 10^{-6}$ ) and combined years (0.00037) (Table 4.4). This could potentially be due to the high percentage of variation contributed by the first PC.

#### 4.4.11. Forward prediction and indirect selection accuracy

To assess indirect selection accuracy and evaluate the utility of the models and approaches, the 2022 and 2023 trials, independently, and combined trials were used as training sets for forward prediction of grain yield in a set of genotypes evaluated in 2024, hereby termed testing set. Phenotypic and genetic correlations between the predicted and actual grain yield in the testing set was evaluated. Indirect selection accuracy was calculated by multiplying the genetic correlation between predicted and actual grain yield with the square root of heritability of the predicted values in the testing set.

#### 4.4.12 Forward prediction: genetic correlation

Genomically predicted values using 2022 trial as the training set showed a negative genetic correlation with actual grain yield in all models, except XGBoost (Table 4.6). Using the 2023 trial, the highest genetic correlation was observed from BL (0.346), while the lowest from RRBLUP (Table 4.6). When using combined trials, low genetic correlation was observed, and predicted values using RRBLUP showed a negative genetic correlation (Table 4.6).

Predicted grain yield using VIs (phenomic prediction), with the 2022 trial as the training set, showed varied genetic correlations, with the highest observed in XGBoost (0.256), as well as when using the 2023 trial as training set (0.392) (Table 4.6). In contrast, negative correlation was observed with XGBoost (-0.152) when using combined trials as the training set, whereas highest genetic correlation was obtained using RRBLUP (0.330) (Table 4.6).

**Table 4.6.** Phenotypic and genetic correlations between predicted grain yield from forward prediction in genotypes evaluated in 2024 with different training sets and models used.

Model	Genetic Correlation						Phenotypic Correlation					
	Training Set:						Training Set:					
	2022		2023		Combined Trials		2022		2023		Combined Trials	
	Geno.	Pheno.	Geno.	Pheno.	Geno.	Pheno.	Geno.	Pheno.	Geno.	Pheno.	Geno.	Pheno.
BayesA	-0.225	0.164	0.265	0.117	0.020	0.250	-0.129	0.472	0.045	0.138	-0.100	0.645
BayesB	-0.197	0.142	0.291	0.073	0.016	0.234	-0.121	0.427	0.047	0.038	-0.042	0.639
BayesC	-0.221	0.185	0.277	0.103	0.022	0.239	-0.126	0.495	0.046	-0.241	-0.040	0.639
BL	-0.221	0.113	0.346	0.232	0.019	0.224	-0.126	0.369	0.054	0.367	-0.043	0.643
BRR	-0.221	0.069	0.267	-0.024	0.017	0.238	-0.127	0.281	0.044	-0.067	-0.040	0.644
RF	-0.018	0.170	0.037	0.364	0.022	0.304	-0.020	0.538	-0.065	0.455	-0.047	0.587
RRBLUP	-0.302	0.224	0.025	0.322	-0.001	0.330	-0.164	0.613	0.016	0.510	-0.050	0.557
XGBoost	0.093	0.256	0.043	0.392	0.00001	-0.152	-0.023	0.571	0.002	0.416	0.005	-0.384

Geno = genomic prediction, pheno = phenomic prediction, BL = Bayesian Lasso, BRR = Bayesian Ridge Regression, RF = Random Forest, RRBLUP = Ridge Regression Best Linear Unbiased Prediction, XGBoost = Extreme Gradient Boosting

Comparing genomic and phenomic predictions, using the 2023 trial as the training set resulted in higher genetic correlations in genomic prediction, whereas using combined trials in phenomic prediction resulted in higher genetic correlations (Table 4.6). Interestingly, predicted values obtained from genomic prediction using RRBLUP resulted in low genetic correlation, whereas predicted values obtained from phenomic prediction using RRBLUP resulted in higher genetic correlation (Table 4.7). Overall, phenomically predicted values showed higher genetic correlation compared to genomically predicted values (Table 4.7).

The genetic correlations in the testing set using different training sets and approaches for integrating UAV-derived information (VIs) with genomic data for grain yield predictions were also evaluated. RRBLUP was used as the base model for genomic prediction of grain yield, with VIs as fixed effects.

Genetic correlations varied depending on the VI used as a fixed effect with the 2022 trial as the training set. The highest genetic correlation observed using GNDVI (0.193) and PNDVI (0.192), while similar genetic correlations were also observed in several VIs ranging from 0.165 to 0.189 (Table 4.7). Similarly, the highest genetic correlation was observed using PNDVI with the 2023 trial as the training set (0.380) and with combined trials as the training set (0.334) (Table 4.7). Cirededge resulted in negative genetic correlation using the 2022 trial (-0.307), the 2023 trial (-0.213), and combined trials (-0.226) as the training set (Table 4.8). Overall, using the 2023 trial resulted in higher genetic correlations compared to using either the 2022 trial or combined trials (Table 4.7).

Genetic correlation of predicted grain yield in the testing set using RRBLUP with multiple and increasing number of VIs was also evaluated. The trend observed in the predictive ability assessment from cross validation was observed when 2022 and 2023 trial was used as training set

independently. Highest genetic correlation was observed when only one VI (GRNDVI) was used as fixed effect using the 2022 trial as training set (Table 4.7). Genetic correlation decreased as more VIs were incorporated, with negative correlations observed when six to eight VIs were used (Table 4.7). Using the 2023 trial as the training set, highest genetic correlation was observed when incorporating two VIs (0.39) (Table 4.7). Relatively similar genetic correlation was observed when

**Table 4.7.** Phenotypic and genetic correlations between predicted grain yield from forward prediction in genotypes evaluated in 2024 with different training sets and different UAV-data information integration.

Model and Approach	Genetic Correlation			Phenotypic Correlation		
	Training Set:			Training Set:		
	2022	2023	Combined Trial	2022	2023	Combined Trial
<i>Selected VIs used Individually as Fixed Effect</i>						
ARVI2	0.171	0.354	0.307	0.387	0.572	0.518
CIrededge	-0.307	-0.213	-0.226	-0.458	-0.558	0.505
CTVI	0.169	0.358	0.305	0.389	0.573	0.514
GNDVI	0.193	0.363	0.315	0.468	0.593	0.551
GRNDVI	0.189	0.364	0.319	0.442	0.596	0.550
NDRE	0.167	0.279	0.276	0.509	0.113	0.564
NDVI	0.171	0.354	0.307	0.387	0.572	0.518
NGRDI	0.046	0.289	0.235	0.097	0.368	0.293
NormG	0.188	0.344	0.305	0.459	0.573	0.530
NormNIR	0.189	0.361	0.307	0.442	0.590	0.536
NormR	0.165	0.350	0.304	0.369	0.563	0.508
PNDVI	0.192	0.380	0.334	0.425	0.588	0.540
RI	0.046	0.289	0.235	0.097	0.368	0.293
VARI	0.045	0.302	0.243	0.084	0.353	0.278
<i>Multiple and Increasing Number of VIs as Fixed Effect</i>						
1	0.19	0.36	0.32	0.44	0.6	0.55
2	0.12	0.37	0.29	0.46	0.57	0.57
3	0.12	0.35	0.26	0.47	0.54	0.56
4	0.10	0.35	0.26	0.46	0.54	0.57
5	0.10	0.31	0.19	0.45	0.46	0.43
6	-0.28	0.24	0.29	-0.24	0.21	0.61
7	-0.31	-0.3	0.3	-0.31	-0.59	0.61
8	-0.08	-0.28	0.29	0.08	-0.61	0.61
9	0.06	-0.32	0.34	0.39	-0.56	0.61
<i>Multiple and Increasing Number of Dimensionally Reduced VIs (PCs) as Fixed Effect</i>						
1	0.160	-0.205	0.312	0.354	-0.524	0.495
2	0.180	-0.209	0.313	0.459	-0.503	0.527
3	0.302	-0.180	0.339	0.425	-0.503	0.511



using one to six VIs (Table 4.7). However, genetic correlations began to decrease upon integrating seven to nine VIs as fixed effects (Table 4.7). Interestingly, when combined trials were used as the training set, genetic correlation was also observed to be similar, with the exemption of when five VIs were used as fixed effect (Table 4.7).

The 30 VIs generated were dimensionally reduced into PCs, accounting for collinearity among some VIs while maximizing the information retained in each VI, and were then used as fixed effects (Table 4.7). Highest genetic correlation was observed when the first three PCs were used as fixed effect with the 2022 trial as the training set (0.302), increasing from 0.160 using only the first PC (Table 4.7). Using the 2023 trial as the training set, negative correlation was observed when PCs were used as fixed effect (Table 4.7). With combined trials as the training set, highest genetic correlation was observed when the first three PCs were used (0.339) (Table 4.7). Overall, genetic correlation increased as more PCs were used as fixed effects (Table 8). Furthermore, utilizing PCs from combined trials resulted in higher genetic correlations compared to using either of the individual trials (Table 4.7).

While there were variations on how the predicted values were genetically correlated with actual grain yield across in the testing set using different UAV-based information and genomic data integration approaches, models, and training sets, overall, we observed higher genetic correlations when UAV and genomic data were integrated compared to using genomic data alone.

#### *4.4.13. Forward prediction: phenotypic correlation*

Phenotypic correlation between predicted values in the testing set with actual grain yield was assessed. Negative phenotypic correlation was observed in genomically predicted values using 2022 trial as training set (Table 4.6). Low phenotypic correlation was also observed in genomically predicted values using 2023 trial as training set, with the highest only having a

correlation of 0.054 (BL) (Table 4.6). Negative phenotypic correlations was also observed when combined trials were used as training set (Table 4.6).

Using the 2022 trial as training set, phenotypic correlation across models ranged from 0.281 to 0.613, with the highest observed using RRBLUP (Table 4.6). Similarly, RRBLUP had the highest phenotypic correlation when 2023 trial was used as training set (0.510), whereas negative phenotypic correlation was observed in BayesC and BRR models (Table 4.6). Using the combined trials, bayesian models showed higher phenotypic correlation compared when other models were used, with the highest observed using BayesA (0.645) (Table 4.6). Comparing how training set affected phenotypic correlation, using combined trials as training set resulted in higher phenotypic correlations (Table 4.6).

Overall, phenomically predicted grain yield showed higher phenotypic correlations compared to genomically predicted grain yield.

Using VIs individually as fixed effect with 2022 trial as training set, a wider range of phenotypic correlation was observed (-0.458 to 0.509) across the different VIs, with the highest observed using NDRE (Table 4.7). Using the 2023 trial as training set, relatively similar phenotypic correlation was observed in ten VIs, with the highest observed in GRNDVI (0.596) (Table 4.7). In both 2022 trial and 2023 trial as training set, incorporating CIrededge as fixed effect resulted in negative correlation (Table 4.7). Using the combined trials as training set, eleven VIs had relatively similar phenotypic correlations, with the highest observed when NDRE was used (Table 4.7). Comparing the training sets, relatively higher phenotypic correlations was observed across different VIs when 2023 trial was used as training set, with the exemption of CIrededge (Table 4.8). NDRE, interestingly, had the highest phenotypic correlation using 2022 trial and the

combined trials as training set, and had one of the lowest phenotypic correlation using 2023 trial as training set (Table 4.7).

Using multiple VIs incorporated in RRBLUP as fixed effect with 2022 trial as training set, phenotypic correlation started to decreased from 0.44 using only one VI to  $-0.31$  when seven VIs were used (Table 4.7). Interestingly, the phenotypic correlation when nine VIs were used was only slightly lower with the phenotypic correlation when one to five VIs were used (Table 4.7). Higher phenotypic correlation was observed when one VI (0.6) using the 2023 trial as training set (Table 4.7). From thereon, phenotypic correlation started to decrease as more number of VIs incorporated (Table 4.7). Using the combined trials as training set, phenotypic correlation increased as the number of VIs incorporated increased, eventually reaching the highest phenotypic correlation using six to nine VIs (0.61) (Table 4.7). While the different training sets used resulted in different patterns of how increasing the number of VIs affected phenotypic correlations, higher correlations was observed when the combined trial was used compared to using the individual trials as training set (Table 4.7).

Phenotypic correlation increased when the first two (0.459) or the first three (0.425) PCs were used as fixed effect in RRBLUP with 2022 trial as training set (Table 4.7). Negative phenotypic correlation, however, was observed when the 2023 trial was used as training set (Table 4.7). Compared to both individual trials, using the combined trial as training set resulted in higher phenotypic correlation (Table 4.7). In addition, highest phenotypic correlation was observed when two (0.527) or three (0.511) PCs were used as fixed effect (Table 4.7).

Similar with the obervations in genetic correlation assessment, phenotypic correlations varied depending on the training set, model, and UAV-data integration approach used. Comparing the different approaches, RRBLUP with multiple VIs as fixed effect using combined trials as

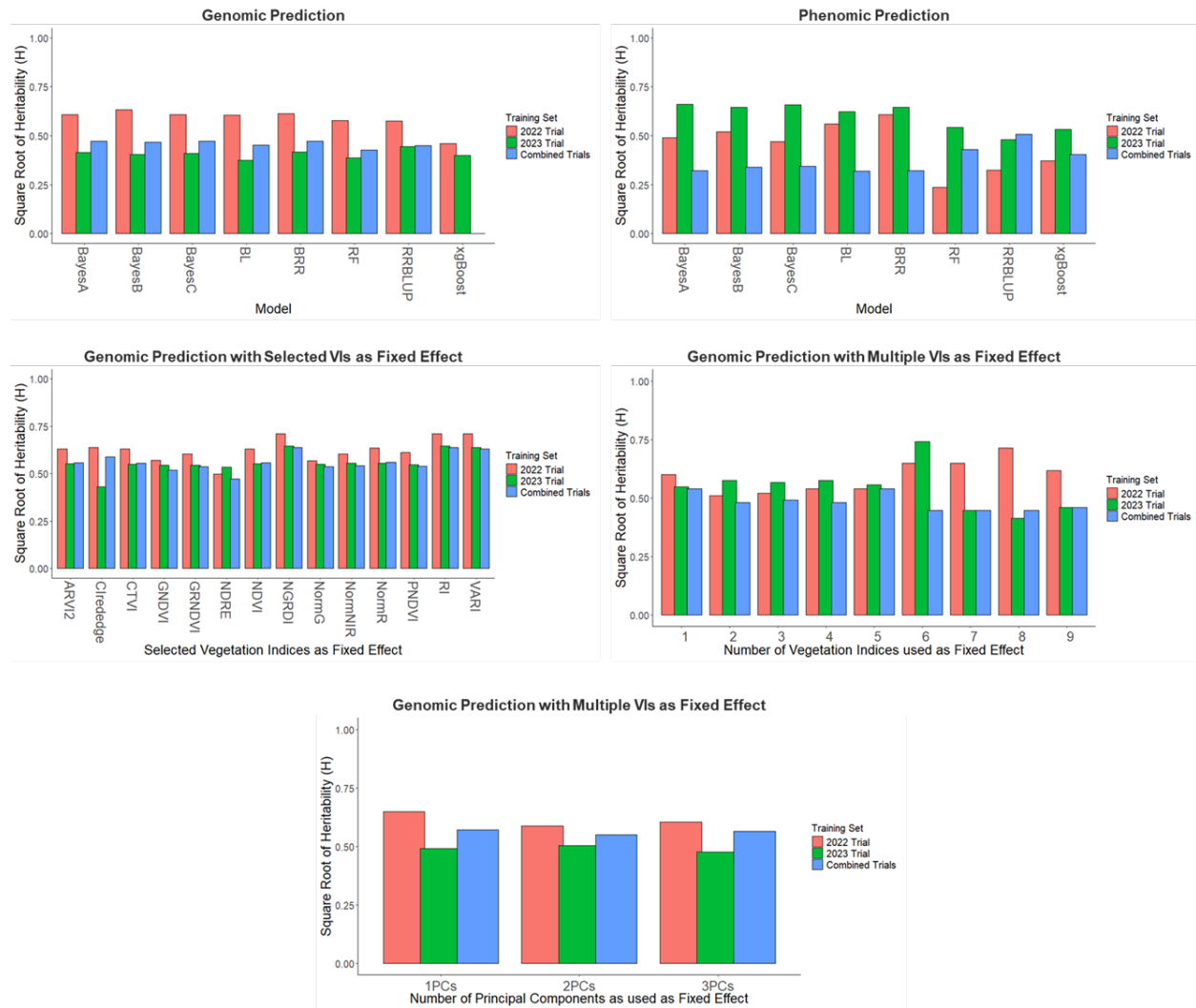
training set resulted in higher phenotypic correlations. Overall, incorporating UAV-information with genomic data, regardless of the approach, resulted in higher phenotypic correlations compared to using genomic data alone.

#### *4.4.14. Forward prediction: variation in heritability in predicted grain yield*

Heritability, ratio of estimated genetic variance and estimated phenotypic variance, was assessed in the predicted grain yield of the testing set using different training sets, models, and approaches used. The square root of heritability, simply termed H from this point forward, was computed in predicted values to assess the proportion of variation controlled by genetics in forward prediction. In a broader sense, this serves as an indirect measure of the accuracy of the models and approaches in capturing genetic variance.

Bayesian models showed higher H in genomically predicted values using the 2022 trial, with the highest observed in BayesB (0.63) (Figure 4.3). Using the 2023 trial as training set, H ranged from 0.374 to 0.444. Contrary to when the 2022 trial was used, RRBLUP had the highest H when 2023 trial was used as training set (0.444) (Figure 4.3). Combined trials as training set resulted in H ranging from 0.00003 to 0.473, with the highest observed in BayesC (Figure 4.3). Comparing H from different training sets used, the 2022 trial as training set resulted in higher H (Figure 4.3). In addition, regardless of the training set used, BayesA, BayesB, BayesC and BRR consistently had higher H, whereas xgBoost performed poorly (Figure 4.3).

Phenomically predicted grain yield with 2022 trial as training set showed higher H when Bayesian models were used, with the highest in BRR (0.607), compared to other phenomic prediction models (Figure 4.3) Using the 2023 trial as training set, H ranged from 0.481 to 0.660, with the highest observed in BayesB (Figure 4.3). With combined trials as training set, H ranged from 0.319 to 0.507, with the highest observed in RRBLUP (Figure 4.3). Utilization of the 2023



**Figure 4.3.** Square root of estimated heritability (H) predicted grain yield in the breeding lines evaluated in 2024 (testing set) from forward prediction using various models and data integration approaches with different training sets.

trial resulted in better H compared to using either the 2022 trial or combined trials as training set (Figure 4.3). Interestingly, when individual trials were used as training set, Bayesian models performed better (Figure 4.3). However, when combined trials were used as training set, RF, RRBLUP, and XGBoost performed better (Figure 4.3).

Interestingly enough, H was relatively higher in genomically predicted value using 2022 trial, while H was higher in phenomically predicted values using the 2023 trial as training set

(Figure 4.3). In addition, for both phenomically and genomically predicted values, Bayesian models showed relatively better H compared to RF, RRBLUP, and xgBoost (Figure 4.3).

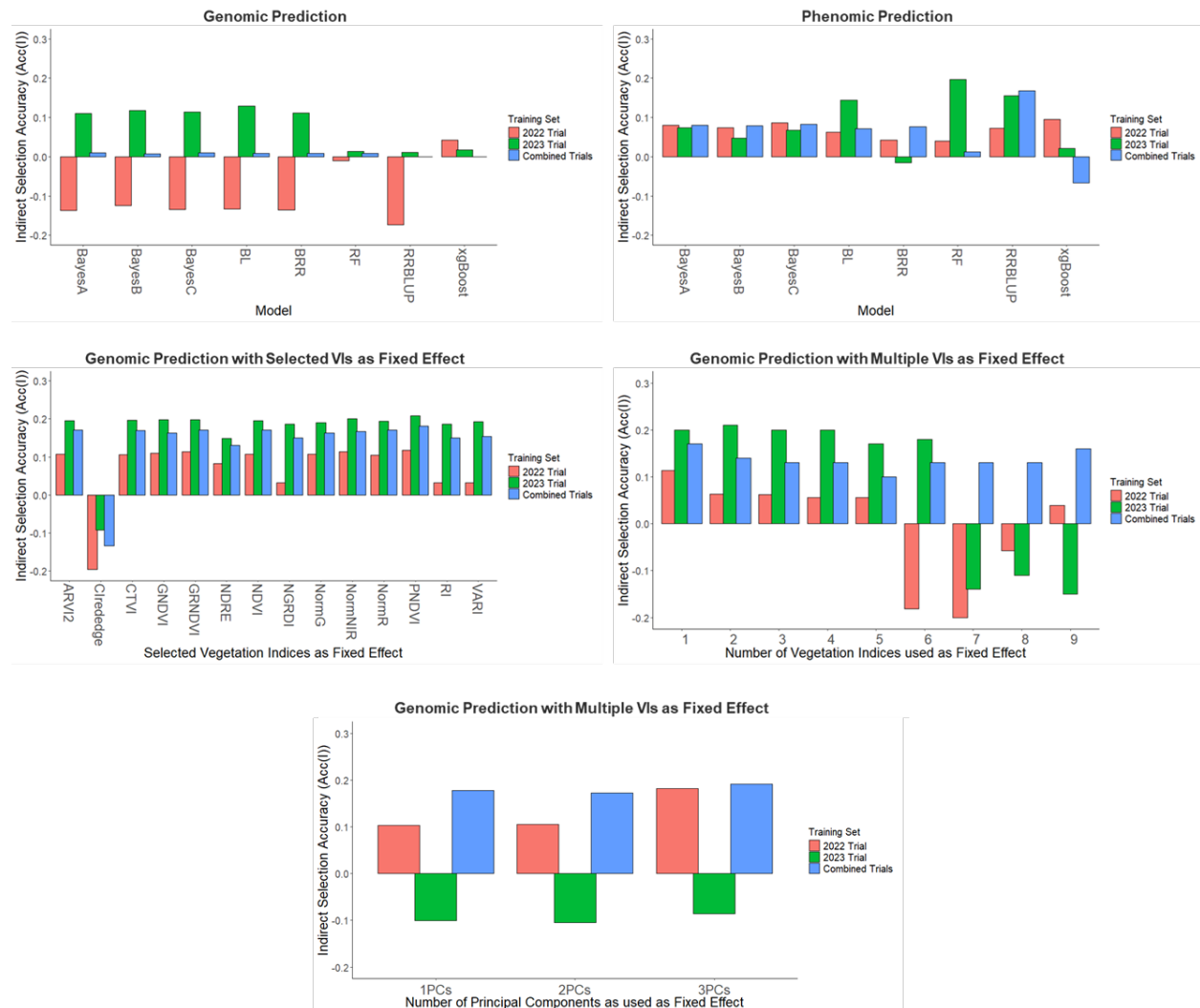
The H of predicted values from RRBLUP using individual VIs as fixed effect ranged from 0.498 to 0.711 when 2022 trial was used as fixed effect, 0.431 to 0.646 when 2023 trial was used, and 0.479 to 0.638 when combined trials were used (Figure 4.3). Across all training sets, incorporation of NGRDI, RI, and VARI resulted in the highest H, while lowest was obtained when NDRE was used as fixed effect (Figure 4.3).

The H of predicted values from RRBLUP with multiple VIs as fixed effect using the 2022 trial as training set ranged from 0.510 to 0.714, with an observable increase in H as the number of VIs incorporated increased (Figure 4.3). Highest H was obtained when eight VIs were incorporated, and lowest using only two VIs (Figure 4.3). Using the 2023 trial, H ranged from 0.412 to 0.742, with the highest observed when seven VIs were used as fixed effect (Figure 4.3). Using combined trials, a different trend was observed. Using the combined trials as training set, similar H was observed when one or five VIs were used, whereas the rest had similar H (Figure 4.3). Taking the observations altogether, using individual trials resulted in higher H, with an increasing trend as the number of VI as fixed effect increased, as compared to when using combined trial as training set (Figure 4.3).

PCs used as fixed effect showed no discernable pattern in H of predicted values in all of the training sets used (Figure 4.3). However, highest H was observed when only the first principal component was used as fixed effect when 2022 trial and combined trials were used as training set, whereas higher H was observed when two PCs were used in 2023 trial as training set (Figure 4.3). In addition, higher H was observed when 2022 trial was used as training set.

#### 4.4.15. Forward prediction: indirect selection accuracy

While other metrics were explored and evaluated in forward prediction, assessing the indirect selection accuracy serves as an excellent metric in comparing models and approaches, and ultimately, the effectivity and utility of trained models and approaches. Indirect selection accuracy was calculated by multiplying the genetic correlations and the square root of heritability of predicted grain yield.



**Figure 4.4.** Indirect selection accuracy (Acc(I)) of predicted grain yield in the breeding lines evaluated in 2024 (testing set) from forward prediction using various models and data integration approaches with different training sets.

Low indirect selection accuracy was observed in genomic prediction models using the 2022 trial as training set (Figure 4.4). However, it is worth noting that the negative indirect selection accuracy was a result of the predicted and actual grain yield having negative correlations. Low indirect selection accuracy was also observed across models using 2023 trial and combined trials as training set (Figure 4.4). Across all models, genomic prediction using the 2023 trial resulted in slightly higher indirect selection accuracy (0.011 to 0.130) over genomic prediction using combined trials as training set ( $-0.0001$  to 0.01) (Figure 4.4).

Overall, phenomic indirect selection accuracy was higher compared to genomic indirect selection accuracy. With 2022 trial as training set, phenomic indirect selection accuracy ranged from 0.042 to 0.095, whereas using the 2023 trial as training set ranged from  $-0.015$  to 0.197, and  $-0.067$  to 0.168 when combined trials were used as training set (Figure 4.4). Highest indirect selection accuracy was observed in BL with 2023 trial as training set using phenomic information, which was higher than any other model both in genomic and phenomic prediction (Figure 4.4).

Indirect selection accuracy varied based on the individual VI used as fixed effect. With the 2022 trial as training set, indirect selection accuracy ranged from  $-0.196$  to 0.118,  $-0.092$  to 0.208 using the 2023 trial, and  $-0.133$  to 0.181 using the combined trials (Figure 4.4). Across training sets, highest indirect selection accuracy was observed when PNDVI was incorporated (Figure 4.4). The negative indirect selection accuracy observed in CIrededge was driven by the negative correlation between predicted and actual grain yield values (Figure 4.4).

Indirect selection accuracy decreased as the number of VIs incorporated increased with the 2022 trial as training set (Figure 4.4). Using the 2023 trial as training set, similar indirect selection accuracy was observed upon integration of one to four VIs as fixed effect, and started to decrease upon integration of five to nine VIs (Figure 4.4). Using combined trials as training set, the indirect



selection accuracies were relatively similar, ranging from 0.13 to 0.17, regardless of the number of VIs used as fixed effect, with the exemption however of when five VIs were used as fixed effect (Figure 4.4). Overall, using the 2023 trial as training set resulted in higher indirect selection accuracy over using either 2022 trial or combined trials as training set. However, similar with the trend observed in predictive ability assessment, indirect selection accuracy somehow decreased as more VIs were used as fixed effect (Figure 4.4).

PCs as fixed effect in genomic prediction resulted in varied prediction accuracies. With the 2022 trial as training set, highest indirect selection accuracy was obtained when the first three PCs were used (Figure 4.4). Negative prediction accuracies were obtained when 2023 trial was used as training set. Overall, higher prediction accuracies were obtained when combined trials were used as training set over using individual trials, with the highest indirect selection accuracy was observed when the first three PCs were used as fixed effect (Figure 4.4).

While the values generated for indirect selection accuracy seem low, it is worth noting that the point of reference was the indirect selection accuracy of genomic prediction models, especially RRBLUP. Considering this, higher indirect selection accuracies were obtained from phenomic prediction using UAV-information. More importantly, indirect selection accuracy was also observed in models and approaches where VIs were either incorporated as fixed effect or used as secondary trait. However, we would like to point out the obvious differences between the obtained indirect selection accuracies, regardless of the model and approach used, between the 2022 and 2023 trial when used independently as training set. Therefore, using combined trials would be a better option to take into account the difference between years.

## 4.4. Discussion

### 4.4.5. UAV-based high throughput phenotyping for grain yield in soft winter wheat

The heritability of grain yield and vegetation indices (VIs) varied across trials, primarily due to differences in the genotypes evaluated in 2022 and 2023. More importantly, environmental conditions varied between years. In 2023, drought conditions led to lower soil moisture content, particularly toward the harvest stage, along with higher temperatures during plant development (Supplementary Table 4.16). A significant portion of the phenotypic variation was attributed to the year effect and genotype-by-year interactions. These findings align with previous studies emphasizing the impact of genotype-by-environment interactions on wheat grain yield (Nehe et al., 2019; Lozada & Carter, 2020; Saeidinia et al., 2023).

Interestingly, VIs exhibited stronger phenotypic and genotypic correlations with grain yield in the 2023 trial, despite greater error variance relative to genotype variance. This suggests that VIs may capture more than just yield variation, potentially reflecting other environmental and physiological factors.

Based on phenotypic and genetic correlations, as well as feature selection analyses, several VIs demonstrated potential associations with grain yield. NDRE emerged as a particularly promising index for indirect yield evaluation and selection. Zhang et al. (2019) demonstrated NDRE's ability to differentiate rice genotypes based on final grain yield under varying nitrogen treatments. Similarly, Torino et al. (2014) reported that NDRE provided more accurate yield predictions in maize compared to NDVI. In this study, NDRE exhibited higher heritability and stronger correlations with grain yield than several other VIs, possibly due to the timing of UAV flights at crop maturity. Boiarskii & Hasegawa (2019) noted that different VIs are useful at different growth stages, particularly highlighting NDRE's sensitivity to chlorophyll content at later

stages. Similarly, Kanke et al. (2010) suggested that the red-edge wavebands, from which NDRE is derived, are more responsive to plant physiological changes during advanced growth stages.

NormNIR was also consistently selected through feature selection as a potential VI associated with grain yield. While no prior studies have directly linked NormNIR to grain yield, Mendes dos Santos et al. (2022) found that it effectively distinguished between healthy and leaf miner-infected coffee leaves. To our knowledge, this is the first report associating NormNIR with grain yield in wheat.

GRNDVI was another VI consistently identified to be associated with grain yield. It has been widely used as an indicator of crop growth (Zhao et al., 2011) and is frequently employed for monitoring vegetation development (Fan et al., 2020). In sugarcane, Akbarian et al. (2022) found that combining NDRE and GRNDVI yielded the most accurate predictions of biomass-related yield. Although sugarcane yield differs from wheat grain yield, NDRE and GRNDVI may capture biomass variation in wheat, potentially influencing yield.

Although NDVI remains one of the most commonly used VIs for grain yield evaluation and selection (Wall et al., 2007; Duan et al., 2017; Hassan et al., 2019; Aranguren et al., 2020), its phenotypic and genetic correlations with grain yield in this study, along with its heritability, were relatively lower than those of NDRE and GRNDVI. This contrasts with several studies where NDVI exhibited the strongest correlations with yield (Duan et al., 2017). These results highlight the importance of carefully selecting VIs for yield evaluation and indirect selection, rather than relying solely on NDVI.

#### *4.4.2. Genome wide association mapping of VIs and grain yield*

Significant marker-trait associations (MTAs) were identified for grain yield: three from the 2022 trial, two from the 2023 trial, and two from BLUEs calculated across years. However, none

of these MTAs were consistently detected across multiple trials, suggesting that they may be genotype-dependent. This is likely due to differences in the genotypes evaluated each year, as well as environmental variations between 2022 and 2023. Nonetheless, two MTAs from the 2022 trial exhibited large effects: S5B\_607526225, explaining 11.05% of the phenotypic variance, and S7A\_54345139, explaining 31.05%.

Interestingly, MTA S7A\_54345139, identified in the 2022 trial, was associated with grain yield, NDRE, and Shape Index (IF), explaining 31.05%, 15.69%, and 7.03% of the observed phenotypic variance, respectively. Notably, in 2022, NDRE exhibited the highest positive phenotypic and genotypic correlation with grain yield, whereas IF showed a significant negative correlation with yield. Furthermore, MTA S7A\_54345139 had negative marker effects for both grain yield and NDRE, while exhibiting a positive marker effect for IF—contradicting its correlation with grain yield. This suggests that NDRE and IF may share a similar genetic control despite being negatively correlated, and both may also be genetically linked to grain yield. However, since grain yield is a complex quantitative trait influenced by multiple contributing factors, it is essential to determine which specific yield-related traits NDRE and IF are capturing, ultimately contributing to yield variation.

When examining MTAs associated with VIs, we observed that several VIs may share common genetic control, even when negatively correlated. For instance, MTA S1A\_295503711 was mapped to 12 VIs in the 2022 trial, with marker effects varying by VI. This MTA had positive marker effects for CTVI, NormG, and NormR, but negative marker effects for the remaining eight VIs. Notably, CTVI, NormG, and NormR were negatively correlated with both the other VIs and grain yield, while the remaining VIs, except for GRNDVI, were positively correlated with grain yield. A similar pattern was observed for MTA S6B\_18908474. In contrast, MTA S5D\_515694610

followed a different pattern, exhibiting positive marker effects for BNDVI, GLI, NDVIrededge, NGRDI, and VARI, but negative marker effects for CTVI, CVI, EVI, IF, and RI. Interestingly, the marker effects of this MTA mirrored the phenotypic correlations of these VIs with grain yield, which contrasts with the patterns observed for S1A\_295503711 and S6B\_18908474.

Although strong phenotypic and genetic correlations were observed between most VIs and grain yield, these results emphasize the complex genetic architecture underlying both grain yield and vegetation indices. While the MTAs identified in this study may be associated with grain yield in soft winter wheat, many of them did not show a significant yield difference among genotypes carrying these MTAs. These findings highlight the challenges of dissecting the genetic basis of grain yield and understanding the genetic links between VIs and yield. Further investigation is needed to elucidate the shared genetic control between these VIs and grain yield.

#### *4.4.3. Predictive ability using UAV-derived information and in combination with genomic information outperforms predictive ability of genomic prediction*

Grain yield prediction using genomic information varied across trials, with higher genomic predictive ability observed in 2022 compared to 2023 and the combined trials. However, since the genotypes evaluated in 2022 and 2023 were largely different, this may have influenced predictive ability. Additionally, environmental differences between the two years likely contributed to the observed variation in predictive ability (Supplementary Table 4.16). In the 2023 trial and combined trials used in cross-validation, Random Forest had the highest predictive ability, although only marginally better than other genomic prediction models. This suggests that both linear and non-linear relationships may exist between genomic information and grain yield, as evidenced by BayesC yielding the highest accuracy in the 2022 trial.

Using 30 vegetation indices (VIs) as predictors for grain yield resulted in higher predictive ability than genomic prediction alone. This finding aligns with Montesinos-Lopez et al. (2023), who demonstrated that incorporating VIs derived from multispectral sensors, along with raw wavebands, improved predictive ability in wheat. Similarly, Wan et al. (2020) reported enhanced grain yield predictive ability in rice when multiple UAV-based multispectral imaging datasets were included as input variables in Random Forest. Several other studies have confirmed that UAV-derived RGB and multispectral imaging data improve grain yield predictive ability in maize (Sunoj et al., 2021; Kumar et al., 2023), soybean (Herrero-Huerta et al., 2020; da Silva et al., 2020; Maimaitijiang et al., 2020), and sorghum (Galli et al., 2020; Varela et al., 2021).

While UAV-derived phenomic data, such as VIs, demonstrated higher grain yield predictive ability compared to genomic data alone, integrating both data types is essential. Several studies have explored integrating UAV-based RGB and multispectral imaging data with genomic information to predict grain yield in wheat using various models and approaches (Rutkoski et al., 2016; Sun et al., 2019; Sandhu et al., 2021; Montesinos-Lopez et al., 2023; Kaushal et al., 2024). Genomic selection enables prediction based on genomic estimated breeding values (GEBVs), primarily capturing additive effects and genomic relationships among selection candidates (Meuwissen et al., 2001; Habier et al., 2010). In contrast, phenomic selection accounts for non-additive effects, which genomic prediction may miss (Rincent et al., 2018) and incorporates environmental and genotype-by-environment interactions, as phenomic data essentially serve as a proxy for phenotype (Fernandez et al., 2017). Since both approaches provide complementary information, combining genomic and phenomic selection is crucial (Robert et al., 2022). Additionally, linking genetic data (genomic) with phenotypic expression (phenomic) ensures that no important trait insights are overlooked (Robert et al., 2022).

Predictive ability nearly doubled when selected VIs were included as fixed effects in RRBLUP, with the highest accuracy achieved when NDRE was used. However, NGRDI, RI, and VARI resulted in the lowest predictive abilities and even reduced predictive ability in the 2022 trial compared to genomic prediction using RRBLUP without fixed effects. This underscores the importance of selecting the appropriate VIs as fixed effects to improve grain yield predictive ability.

To leverage the full range of information from all VIs while addressing multicollinearity, dimensionality reduction was applied. The first three principal components (PCs) were used as fixed effects, resulting in predictive abilities higher than genomic prediction alone and comparable to phenomic prediction or the use of individual and multiple VIs as fixed effects. An advantage of this approach is that dimensionally reduced phenomic data can capture more environmental and physiological effects relevant to grain yield than relying on a small set of individual indices. Given this, increasing the number of VIs to capture additional environmental and physiological effects could further enhance grain yield prediction when processed through dimensionality reduction techniques. However, this approach requires additional data processing, such as principal component analysis, and extra computational time to generate and analyze a larger set of VIs.

#### *4.4.4. Predictive ability of using integrated genomic and UAV-data depends on the correlation of vegetation indices with grain yield*

Several metrics and considerations have been established the potential of phenomic information—specifically UAV-derived VIs—to evaluate agronomic traits such as grain yield. These include the heritability of phenomic traits and their phenotypic and genetic correlations with the target trait. However, our findings suggest that phenotypic correlation is a stronger indicator of predictive ability.

When NDRE was used as a fixed effect in RRBLUP, the predictive ability was 0.65 in the 2022 trial, 0.87 in the 2023 trial, and 0.78 in the combined trials. These values closely mirrored the phenotypic correlations of NDRE with grain yield, which were 0.66 in the 2022 trial, 0.86 in the 2023 trial, and 0.78 in the combined trials. Interestingly, these results contrast with the estimated heritability of NDRE, which was 0.85 for the 2022 trial, 0.57 for the 2023 trial, and only 0.045 across years. Despite this discrepancy, heritability remains a relevant metric for assessing the potential of a trait as a predictor.

Following this, RGBVI should be considered a viable index for evaluating grain yield, particularly when accounting for environmental variation (e.g., year effects). Among all VIs and grain yield, RGBVI had the highest estimated heritability across years. However, its predictive ability when used as a fixed effect in RRBLUP was relatively low: 0.05 in the 2022 trial, 0.77 in the 2023 trial, and 0.57 in the combined trials. Despite its lower predictive ability compared to other VIs, the trend of predictive ability aligning with phenotypic correlation persisted. The phenotypic correlations between RGBVI and grain yield were 0.01 for the 2022 trial, 0.77 for the 2023 trial, and 0.58 for the combined trials.

Additionally, when dimensionally reduced vegetation indices were used as secondary traits in a multi-trait genomic prediction model with a factor analytic structure, the trend of phenotypic correlation influencing predictive ability remained evident. The phenotypic correlation between the first principal component (PC1) and grain yield was 0.52 for the 2022 trial, 0.86 for the 2023 trial, and 0.77 for the combined trials. While the second and third principal components (PC2 and PC3) had much lower and non-significant correlations with grain yield, the predictive ability followed a similar trend. When the first PCs were used as secondary traits, the predictive abilities were 0.65, 0.87, and 0.77 for the 2022 trial, 2023 trial, and combined trials, respectively. When



only the first two PCs were used, the predictive abilities were slightly lower at 0.59, 0.87, and 0.77, respectively.

Thus, our results suggest that the phenotypic correlation between the VIs and the target trait is a strong predictor of predictive ability, especially when integrated as fixed effect in genomic prediction models.

#### *4.4.5. Increasing the number of covariates in genomic prediction models may reduce predictive ability*

Including multiple VIs as fixed effects in RRBLUP initially improved predictive ability compared to genomic prediction without fixed effects. However, as more VIs were added, a decreasing trend in predictive ability was observed. This suggests that incorporating too many VIs may introduce noise into the model due to multicollinearity among some indices, leading to redundant information. Therefore, careful selection of fixed effects is crucial when integrating multiple VIs into genomic prediction models to maximize predictive ability while minimizing potential overfitting or loss of model efficiency.

#### *4.4.6. Training set using combined trials is more advantageous in forward prediction of unique genotypes in untested environment*

Indirect selection accuracy, defined as the genetic correlation between predicted values from forward prediction and actual grain yield, multiplied by the square root of the estimated heritability of the predicted values, varied depending on the training set used. Forward prediction using the 2023 trial as the training set resulted in higher phenotypic correlation, genetic correlation, and indirect selection accuracy, whereas using the 2022 trial generally led to lower values for these metrics. The discrepancy between using either of the independent trials as training sets was substantial across all metrics. These differences can largely be attributed to the environmental

variations between 2022 and 2023 (Supplementary Table 4.16). Additionally, relying on a single trial or year may fail to adequately represent the environmental conditions of the target environment where forward prediction is conducted.

Furthermore, the genotypes present in the individual trials were largely different. This effect was particularly evident in the indirect selection accuracy and genetic correlation assessments when genomic prediction without covariates was used for forward predictions. These findings highlight the importance of considering genotype-by-environment interactions when training a model for forward prediction.

Given this, combining trials and training models using the BLUEs calculated across years offers clear advantages over relying on a single-year or single-environment trial. In most approaches, including phenomic prediction and genomic prediction incorporating multiple VIs or principal components as fixed effects, models trained on the combined 2022 and 2023 trials exhibited relatively higher phenotypic and genetic correlations, heritability of predicted values, and indirect selection accuracy compared to those trained on individual trials, particularly the 2022 trial. Interestingly, in some cases, the sign of the phenotypic or genetic correlation, as well as indirect selection accuracy, flipped between positive and negative depending on which individual trial was used as the training set. This was especially evident when principal components were used as fixed effects or when multiple VIs were used as fixed effects. This phenomenon likely results from differences in how VIs correlated with grain yield in 2022 and 2023, and potentially in 2024 as well.

Thus, combining trials and using BLUEs across years or environments is more advantageous, as it accounts for environmental effects not only on the target trait but also on predictor variables such as UAV-derived information. Additionally, combining trials increases the

number of genotypes used to train the model, which is crucial for capturing genetic variation. Since the genotypes used in the trials have different pedigrees and genetic backgrounds, incorporating a larger and more diverse training set enhances the model's ability to account for genotype variation. In this study, forward prediction was intentionally conducted on unique genotypes, and it is likely that combining trials captured more genotype variance, further improving the model's ability to represent the variation present in the testing set.

#### *4.4.7. Integrating genomic and UAV-information results in higher indirect selection accuracy*

Incorporating UAV-derived information into genomic prediction improved indirect selection accuracy compared to using either genomic or phenomic prediction alone in forward prediction. While differences were observed in the performance of models and approaches integrating UAV data with genomic prediction, this overall trend remained consistent. However, several considerations must be taken into account depending on how UAV-derived information is incorporated into genomic prediction.

When using an individual vegetation index (VI) as a fixed effect in genomic prediction, careful selection is crucial. Feature selection performed in trials prior to cross-validation indicated that NDRE was a strong candidate as a fixed effect. However, this did not translate effectively in forward prediction, where lower phenotypic and genetic correlations between predicted and observed yield were observed.

As previously noted, the correlation of VIs with grain yield likely influences both predictive ability and indirect selection accuracy in forward prediction. For example, when RBNDVI was used as a fixed effect, it resulted in higher indirect selection accuracy than NDRE, regardless of the training set used. Comparing NDRE and RBNDVI, NDRE exhibited higher phenotypic and genetic correlations with grain yield in 2022, whereas RBNDVI showed stronger

correlations in 2023 and across years. This pattern likely influenced forward prediction performance in the 2024 preliminary yield trial, where NDRE had a higher phenotypic correlation with yield ( $r = 0.62$ ) compared to RBNDVI ( $r = 0.58$ ). However, this was not always the case.

For instance, Cirededge exhibited positive phenotypic and genetic correlations with yield in the 2022 and 2023 trials and across years. However, in the 2024 trial, Cirededge showed a negative phenotypic correlation with yield ( $r = -0.58$ ). This shift was reflected in the indirect selection accuracy when Cirededge was used as a fixed effect in genomic prediction. These observations suggest that relying on individual VIs as fixed effects may be problematic in the long run due to their varying correlations with the target trait across different years.

Using multiple VIs as fixed effects could be a potential solution to mitigate the limitations of individual VIs. However, indirect selection accuracy remained inconsistent, and in some cases, negative correlations between predicted and actual yield were observed. This presents a challenge in forward prediction, as it could lead to the unintended selection of low-yielding genotypes for advancement instead of high-yielding ones.

To address these concerns, we propose two key strategies: (1) using training sets from combined trials, which have already been shown to be more advantageous than single-year trials, and (2) employing dimensionally reduced UAV-derived information for multiple VIs in model training. Given the presence of multicollinearity among vegetation indices, dimensionality reduction—such as principal component analysis (PCA)—effectively addresses this issue while maximizing the variation captured by different VIs. More importantly, using dimensionally reduced UAV-derived information from combined trials accounts for differences in how these predictors correlate with the target trait. This is particularly crucial, as UAV-derived information is highly susceptible to environmental and year-to-year variations.

#### **4.5. Conclusion**

Our findings highlight that the phenotypic correlation between vegetation indices (VIs) and grain yield is a reliable indicator of predictive ability, regardless of whether VIs are used individually or dimensionally reduced into principal components. More importantly, integrating UAV-derived information as a fixed effect or secondary trait significantly improves indirect selection accuracy compared to using genomic or phenomic predictions alone. The use of dimensionally reduced UAV-derived data further enhances indirect selection accuracy by capturing comprehensive information within VIs while mitigating multicollinearity. Based on these results, we recommend incorporating dimensionally reduced UAV-derived information as a fixed effect for genomic prediction models such as RRBLUP. Additionally, training models on BLUEs from multiple trials provides a more robust approach than relying on single-trial or single-year data. Taken altogether, these strategies offer a more comprehensive and accurate framework for grain yield prediction, reinforcing the importance of integrating UAV-derived phenomic data with genomic selection.

## REFERENCES

- Akbarian, S. *et al.* Sugarcane yields prediction at the row level using a novel cross-validation approach to multi-year multispectral images. *Computers and Electronics in Agriculture* **198**, 107024. <https://doi.org/10.1016/j.compag.2022.107024> (2022).
- Anderson, J.A. Marker-assisted selection for Fusarium head blight resistance in wheat. *International Journal of Food Microbiology* **119**, 51–53. <https://doi.org/10.1016/j.ijfoodmicro.2007.07.025> (2007).
- Aranguren, M., Castellón, A. & Aizpurua, A. Wheat Yield Estimation with NDVI Values Using a Proximal Sensing Tool. *Remote Sensing* **12**, 2749. <https://doi.org/10.3390/rs12172749> (2020).
- Araus, J.L. *et al.* Translating High-Throughput Phenotyping into Genetic Gain. *Trends in Plant Science* **23**, 451–466. <https://doi.org/10.1016/j.tplants.2018.02.001> (2018).
- Aspilcueta-Borquis, R.R. *et al.* Genetic parameters of total milk yield and factors describing the shape of lactation curve in dairy buffaloes. *Journal of Dairy Research* **79**, 60–65. <https://doi.org/10.1017/S0022029911000823> (2012).
- Bates, D. *et al.* Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48 (2015).
- Biecek, P. DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research* **19**, 1–5 (2018).
- Boiarskii, B. & Hasegawa, H. Comparison of NDVI and NDRE Indices to Detect Differences in Vegetation and Chlorophyll Content. *JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES*, spl1. <https://doi.org/10.26782/jmcms.spl.4/2019.11.00003> (2019).
- Bonnett, D. *et al.* (2022). Response to Early Generation Genomic Selection for Yield in Wheat. *Frontiers in Plant Science* **12**. <https://doi.org/10.3389/fpls.2021.718611> (2022).
- Chen, C.J. & Zhang, Z. GRID: a python package for field plot phenotyping using aerial images. *Remote Sensing* **12**, 1697 (2020).
- Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *In* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). New York, NY, USA: ACM. (2016).
- Clark, S.A., Hickey, J.M. & van der Werf, J.H. Different models of genetic variation and their effect on genomic evaluation. *Genetics Selection Evolution* **43**, 18. <https://doi.org/10.1186/1297-9686-43-18> (2011).

- Covarubias-Pazarán, G. Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS ONE* **11**, e0156744 (2016).
- Crossa, J. *et al.* Genomic Selection and Prediction in Plant Breeding. *Journal of Crop Improvement* **25**, 239–261. <https://doi.org/10.1080/15427528.2011.558767> (2011).
- Crossa, J. *et al.* Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science* **22**, 961–975. <https://doi.org/10.1016/j.tplants.2017.08.011> (2017).
- Daetwyler, H.D. *et al.* Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Animal Production Science* **50**, 1004–1010. <https://doi.org/10.1071/AN10096> (2010).
- da Silva, E.E. *et al.* UAV-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation. *Remote Sensing Applications: Society and Environment* **18**, 100318. <https://doi.org/10.1016/j.rsase.2020.100318> (2020).
- de Roos, A.P.W., Hayes, B.J., & Goddard, M.E. Reliability of Genomic Predictions Across Multiple Populations. *Genetics* **183**, 1545–1553. <https://doi.org/10.1534/genetics.109.104935> (2009).
- Duan, T., Chapman, S.C., Guo, Y., & Zheng, B. Dynamic monitoring of NDVI in wheat agronomy and breeding trials using an unmanned aerial vehicle. *Field Crops Research* **210**, 71–80. <https://doi.org/10.1016/j.fcr.2017.05.025> (2017).
- Endelman, J.B. Ridge regression and other kernels for genomic selection with R package RRBLUP. *The Plant Genome* **4**, 250–255 (2011).
- FAOSTAT. <https://www.fao.org/faostat/en/#data/FBS> (accessed January 30, 2025).
- Fan, J., Xu, Y., Ge, H. & Yang, W. Vegetation growth variation in relation to topography in Horqin Sandy Land. *Ecological Indicators* **113**, 106215. <https://doi.org/10.1016/j.ecolind.2020.106215> (2020).
- Farooq, M. *et al.* Genomic prediction in plants: opportunities for ensemble machine learning based approaches. *F1000Research* **11**, 802. <https://doi.org/10.12688/f1000research.122437.2> (2023).
- Galli, G. *et al.* Optimization of UAS-based high-throughput phenotyping to estimate plant health and grain yield in sorghum. *The Plant Phenome Journal* **3**, e20010. <https://doi.org/10.1002/ppj2.20010> (2020).
- Glaubitz, J. C. *et al.* TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* **9**, e90346 (2014).

- Gupta, P. *et al.* QTL analysis, association mapping and marker-assisted selection for some quality traits in bread wheat - An overview of the work done at CCS University, Meerut. *Journal of wheat research* **3**, 1 (2011).
- Habier, D. *et al.* The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution* **42**, 5. <https://doi.org/10.1186/1297-9686-42-5> (2010).
- Hassan, M.A. *et al.* A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant Science* **282**, 95–103. <https://doi.org/10.1016/j.plantsci.2018.10.022> (2019).
- Herrero-Huerta, M., Rodriguez-Gonzalvez, P. & Rainey, K.M. Yield prediction by machine learning from UAS-based multi-sensor data fusion in soybean. *Plant Methods* **16**, 78. <https://doi.org/10.1186/s13007-020-00620-6> (2020).
- Howard, R., Carriquiry, A.L. & Beavis, W.D. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes, Genomes, Genetics* **4**, 1027–1046 (2014).
- Huang, M. *et al.* BLINK: a package for next level genome-wide association studies with individuals and markers in the millions. *GigaScience* **8**, giy154 (2019).
- Jackson, R. *et al.* Phenomic and genomic prediction of yield on multiple locations in winter wheat. *Frontiers in Genetics* **14**, 1164935. <https://doi.org/10.3389/fgene.2023.1164935> (2023).
- Kanke, Y. *et al.* Red Edge as a Potential Index for Detecting Differences in Plant Nitrogen Status in Winter Wheat. *Journal of Plant Nutrition*, **35**, 1526–1541 (2012).
- Kaur, S. *et al.* Hyperspectral imaging combined with machine learning for high-throughput phenotyping in winter wheat. *The Plant Phenome Journal* **7**, e20111. <https://doi.org/10.1002/ppj2.20111> (2024).
- Kaushal, S. *et al.* Enhancing the potential of phenomic and genomic prediction in winter wheat breeding using high-throughput phenotyping and deep learning. *Frontiers in Plant Science* **15**. <https://doi.org/10.3389/fpls.2024.1410249> (2024).
- Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **28**, 1–26 (2008).
- Kumar, C. *et al.* Multi-Stage Corn Yield Prediction Using High-Resolution UAV Multispectral Data and Machine Learning Models. *Agronomy* **13**, 1277. <https://doi.org/10.3390/agronomy13051277> (2023).
- Lipka, A.E. *et al.* GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–9 (2012).



- Liu, R. *et al.* Developing stripe rust resistant wheat (*Triticum aestivum* L.) lines with gene pyramiding strategy and marker-assisted selection. *Genetic Resources and Crop Evolution* **67**, 381–391. <https://doi.org/10.1007/s10722-019-00868-5> (2020).
- Lopez-Cruz M. *et al.* Regularized selection indices for breeding value prediction using hyperspectral imaging data. *Scientific Reports* **10**, 8195 (2020).
- Lorenzana, R.E. & Bernardo, R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and Applied Genetics* **120**, 151–161. <https://doi.org/10.1007/s00122-009-1166-3> (2009).
- Lourenço, V.M. *et al.* Genomic prediction using machine learning: a comparison of the performance of regularized regression, ensemble, instance-based and deep learning methods on synthetic and empirical data. *BMC Genomics* **25**, 152. <https://doi.org/10.1186/s12864-023-09933-x> (2024).
- Lozada, D.N. *et al.* Genomic Prediction and Indirect Selection for Grain Yield in US Pacific Northwest Winter Wheat Using Spectral Reflectance Indices from High-Throughput Phenotyping. *International Journal of Molecular Sciences* **21**, 165. <https://doi.org/10.3390/ijms21010165> (2020).
- Luan, T. *et al.* The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. *Genetics* **183**, 1119–1126 (2009).
- Lübberstedt, T., Beavis, W., & Suza, W. Chapter 7: Marker Assisted Selection and Genomic Selection (2023).
- Maimaitijiang, M. *et al.* Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sensing of Environment* **237**, 111599. <https://doi.org/10.1016/j.rse.2019.111599> (2020).
- Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**, 1819–1829. <https://doi.org/10.1093/genetics/157.4.1819> (2001).
- Montesinos-López, O.A. *et al.* Genomics combined with UAS data enhances prediction of grain yield in winter wheat. *Frontiers in Genetics* **14**. <https://doi.org/10.3389/fgene.2023.1124218> (2023).
- Nehe, A. *et al.* Genotype x environment interaction and genetic gain for grain yield and grain quality traits in Turkish spring wheat released between 1964 and 2010. *PLoS ONE* **14**, e0219432. <https://doi.org/10.1371/journal.pone.0219432> (2019).
- Paixão, P.T.M. *et al.* Factor analysis applied in genomic selection studies in the breeding of *Coffea canephora*. *Euphytica* **218**, 42. <https://doi.org/10.1007/s10681-022-02998-x> (2022).

- Poland, J. A. *et al.* Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLoS ONE* **7**, e32253 (2012).
- Pérez, P. & de los Campos, G. Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* **198**, 483–495 (2014).
- R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2023).
- Rincent, R. *et al.* 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**, 3961–3972. <https://doi.org/10.1534/g3.118.200760> (2018).
- Robert, P. *et al.* Phenomic selection in wheat breeding: identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection. *Theoretical and Applied Genetics* **135**, 895–914. <https://doi.org/10.1007/s00122-021-04005-8> (2022).
- Rodriguez-Alvarez, M.X. *et al.* Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics* **23**, 52–71 (2018).
- Rutkoski, J. *et al.* 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**, 2799–2808. <https://doi.org/10.1534/g3.116.032888> (2016).
- Saeidnia, F., Taherian, M. & Nazeri, S.M. Graphical analysis of multi-environmental trials for wheat grain yield based on GGE-biplot analysis under diverse sowing dates. *BMC Plant Biology* **23**, 198. <https://doi.org/10.1186/s12870-023-04197-9> (2023).
- Salas Fernandez, M.G. *et al.* A High-Throughput, Field-Based Phenotyping Technology for Tall Biomass Crops. *Plant Physiology* **174**, 2008–2022. <https://doi.org/10.1104/pp.17.00707> (2017).
- Sandhu, K. *et al.* Multitrait machine- and deep-learning models for genomic selection using spectral information in a wheat breeding program. *The Plant Genome* **14**, e20119. <https://doi.org/10.1002/tpg2.20119> (2021).
- Santos, L.M.D. *et al.* Vegetation indices applied to suborbital multispectral images of healthy coffee and coffee infested with coffee leaf miner. *AgriEngineering* **4**, 311–319. <https://doi.org/10.3390/agriengineering4010021> (2022).
- Sirsat, M.S., Oblessuc, P.R. & Ramiro, R.S. Genomic Prediction of Wheat Grain Yield Using Machine Learning. *Agriculture* **12**, 1406. <https://doi.org/10.3390/agriculture12091406> (2022).

- Sun, J. *et al.* High-throughput phenotyping platforms enhance genomic selection for wheat grain yield across populations and cycles in early stage. *Theoretical and Applied Genetics* **132**, 1705–1720. <https://doi.org/10.1007/s00122-019-03309-0> (2019).
- Tanabe, R., Matsui, T. & Tanaka, T.S.T. Winter wheat yield prediction using convolutional neural networks and UAV-based multispectral imagery. *Field Crops Research* **291**, 108786. <https://doi.org/10.1016/j.fcr.2022.108786> (2023).
- Tolhurst, D.J. *et al.* Genomic selection using random regressions on known and latent environmental covariates. *Theoretical and Applied Genetics* **135**, 3393–3415. <https://doi.org/10.1007/s00122-022-04186-w> (2022).
- Torino, M.S. *et al.* Evaluation of Vegetation Indices for Early Assessment of Corn Status and Yield Potential in the Southeastern United States. *Agronomy Journal* **106**, 1389–1401. <https://doi.org/10.2134/agronj13.0578> (2014).
- VanRaden, P.M. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science* **91**, 4414–4423. <https://doi.org/10.3168/jds.2007-0980> (2008).
- Varela, S. *et al.* Understanding Growth Dynamics and Yield Prediction of Sorghum Using High Temporal Resolution UAV Imagery Time Series and Machine Learning. *Remote Sensing* **13**, 1763. <https://doi.org/10.3390/rs13091763> (2021).
- Vishwakarma, M.K. *et al.* Marker-assisted improvement of grain protein content and grain weight in Indian bread wheat. *Euphytica* **208** 313–321. <https://doi.org/10.1007/s10681-015-1598-6> (2016).
- Wall, L., Larocque, D. & Léger, P. The early explanatory power of NDVI in crop yield modelling. *International Journal of Remote Sensing* **29**, 2211–2225. <https://doi.org/10.1080/01431160701395252> (2008).
- Wan, L. *et al.* Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer – a case study of small farmlands in the South of China. *Agricultural and Forest Meteorology* **291**, 108096. <https://doi.org/10.1016/j.agrformet.2020.108096> (2020).
- Wei, L. *et al.* Wheat biomass, yield, and straw-grain ratio estimation from multi-temporal UAV-based RGB and multispectral images. *Biosystems Engineering* **234**, 187–205. <https://doi.org/10.1016/j.biosystemseng.2023.08.002> (2023).
- Wiersma, A. T. *et al.* Fine mapping of the stem rust resistance gene SrTA10187. *Theor Appl Genet.* **129**, 2369–2378 (2016).
- Winn, Z.J. *et al.* Phenomic versus genomic prediction—A comparison of prediction accuracies for grain yield in hard winter wheat lines. *The Plant Phenome Journal* **6**, e20084. <https://doi.org/10.1002/ppj2.20084> (2023).

- Ye, X. *et al.* Improvement of three commercial spring wheat varieties for powdery mildew resistance by marker-assisted selection. *Crop Protection* **125**, 104889. <https://doi.org/10.1016/j.cropro.2019.104889> (2019).
- Yoosefzadeh-Najafabadi, M., Rajcan, I. & Eskandari, M. Optimizing genomic selection in soybean: An important improvement in agricultural genomics. *Heliyon* **8**. <https://doi.org/10.1016/j.heliyon.2022.e11873> (2022).
- Zhao, H. *et al.* Improvement and comparative analysis of indices of crop growth condition monitoring by remote sensing. *Transactions of the Chinese Society of Agricultural Engineering* **27**, 243–249 (2011).

## **CHAPTER V: EVALUATING CLOSE-RANGE INFRARED THERMAL IMAGING AND UNSUPERVISED K-MEANS CLUSTERING FOR FUSARIUM HEAD BLIGHT RESISTANCE SELECTION**

### **Abstract**

Development of Fusarium head blight (FHB)-resistant winter wheat varieties relies heavily on large-scale field evaluations, primarily depending on visual scoring of disease severity and incidence. However, this manual assessment remains one of the major bottlenecks in FHB resistance breeding due to its labor-intensive nature and inherent subjectivity. Close-range infrared thermal imaging was explored as a potential alternative for large-scale field evaluations, given its demonstrated ability to distinguish resistance and susceptibility at the single-plant and single-spike levels. Plot-level infrared thermal imaging was conducted using a handheld infrared thermal camera, Flir E60 (Teledyne FLIR, USA), in the Michigan State University Wheat Breeding and Genetics FHB nursery from 2022 to 2024, with each trial treated independently. The scalability of close-range infrared thermal imaging for field applications remains challenging, as no direct relationship was established between plot-level radiometric temperature and FHB-related traits under field conditions, despite the clear association observed at the spike level under greenhouse conditions. Despite this, plot-level radiometric temperature has shown potential as an additional trait in unsupervised K-means clustering, facilitating better reduction of the selection pool compared to simple means comparison. The integration of plot-level radiometric temperature into clustering analyses resulted in improved selection intensity compared to clustering based solely on FHB-related traits. However, further considerations and caution are necessary when attempting to scale close-range infrared thermal imaging for large-scale field evaluations, as its applicability remains uncertain.

## 5.1. Introduction

Fusarium head blight (FHB) remains one of the most economically devastating diseases in wheat, primarily caused by *Fusarium* species (Xu et al., 2008), with *F. graminearum* being the most prevalent in the world, including the United States (McMullen et al., 2012). Each year, FHB affects vast wheat-growing regions, leading to significant reductions in yield and grain quality. FHB-inflicted losses amounted to approximately \$1.2 billion between 2015-2016 (Wilson et al., 2018), highlighting the persistent threat this disease poses to wheat production.

Efforts to mitigate FHB rely on an integrated approach combining agronomic practices, chemical control, and genetic resistance. However, the most effective and sustainable strategy remains to be the development of FHB-resistant wheat varieties (Steiner et al., 2017). Several morphological and physiological traits have been targeted to enhance resistance, including increased plant height to reduce splash dispersal of *Fusarium* spores (Mesterhazy, 1995), awn reduction to limit spore retention (Mesterhazy, 1995), and shorter flowering duration to minimize floret exposure to inoculum (Steiner, et al., 2017). More critically, breeding efforts focus on improving five key resistance mechanisms: Type I resistance, which prevents initial infection (Schroeder and Christensen, 1963); Type II resistance, which limits fungal spread within the spike (Schroeder and Christensen, 1963); Type III resistance, which reduces kernel infection (Mesterhazy et al., 1999); Type IV resistance, which enhances tolerance to FHB infection (Mesterhazy et al., 1999); and Type V resistance, which mitigates mycotoxin (deoxynivalenol, DON) accumulation (Foroud et al., 2019). Among these, Type I and Type II resistance are often prioritized in breeding programs.

Developing FHB-resistant wheat requires extensive phenotypic screening and selection. Traditional methods rely on induction of the disease followed by visual assessment of response to

the infection (Steiner et al., 2017). Large-scale evaluations are typically conducted in disease nurseries under field conditions to further assess resistance and facilitate selection. However, field-based phenotyping poses additional challenges, including labor demands, time constraints, genotype-by-environment interactions, and the subjectivity of visual scoring, which remains a major bottleneck in FHB resistance breeding (Miedaner et al., 2008).

To address these limitations, researchers have increasingly explored the use of imaging technologies such as RGB, multispectral, thermal, and hyperspectral imaging for high-throughput phenotyping of plant diseases (Mahlein et al., 2016). UAV-based imaging has been particularly explored for field-level FHB assessments, providing large-scale, non-destructive disease evaluation (Francesconi et al., 2021). Imaging approaches have also been adopted for single-plant and single-spike level assessments, offering high-resolution data and improved efficiency (Zhang et al., 2019; Mahlein et al., 2019; Mustafa et al., 2023). Among these technologies, close-range infrared thermal imaging has shown promise in detecting FHB-related temperature variations associated with disease progression at the single plant or spike-level (Masri et al., 2017; Mahlein et al., 2019). However, the scalability of close-range infrared thermal imaging for field applications remains largely unexplored, as most thermal imaging studies have been conducted using UAVs at broader spatial scales.

This study aims to explore the potential scaling up of close-range infrared thermal imaging to plot-level phenotyping under field conditions. By integrating thermal imaging into phenotyping pipelines, this research could contribute to more efficient and objective selection strategies for developing FHB-resistant wheat varieties.

## 5.2. Materials and Methods

### 5.2.1. Greenhouse experiment – FHB inoculation, experimental design and establishment

Prior to infrared thermal imaging under field conditions, an initial experiment was done in 2022 under greenhouse conditions. Known susceptible genotype Ambassador and resistant genotype MI14W0190 were subjected to FHB-evaluation under greenhouse conditions to build on the hypothesis that radiometric temperature or infrared thermal readings (simply termed “thermal readings” from this point forward) could detect response to FHB and delineate resistant and susceptible lines. Individual spikes were sprayed with *Fusarium graminearum* macroconidia at  $1 \times 10^5$  concentration at flowering stage using Graco TC Pro Cordless Sprayer (Graco Inc., USA) sprayer. Sprayed individual spikes were bagged for 3 days to create a suitable humid environment. The experiment was laid out in a completely randomized design (CRD) with six individual heads per genotype treated as replicates, with two conditions: non-inoculated (control) and inoculated (FHB-infected). Spike-Level FHB-severity at a single plant scale were assessed at 4, 8, 12, 16, and 20 days after inoculation (DAI) following the formula:

$$(1) \% \text{ Severity} = \frac{\text{No.of Infected Spikelets}}{\text{Total No.of Spikelets}} \times 100$$

### 5.2.2. Greenhouse experiment – single spike-level infrared thermal imaging

Individual spikes from both genotypes and both conditions were imaged at all timepoints, where the spikes were detached from the plant and placed against a white background placed inside a 20 x 20 x 20 inches photobox (Finnhomy). Imaging was carried out using Flir E60 Infrared Camera (320 x 240 IR resolution, 18 mm focal length) (Teledyne FLIR, USA), with default of 0.95 emissivity, 1 meter distance, 20°C external optic temperature. Images were stored and processed in Flir Tools to extract thermal readings with reflectance temperature parameter in Flir Tools adjusted based on measured temperature of the white background. Adjustment for all thermal



readings was carried out using “emmeans” package in R 4.2.2 using ambient temperature as covariate due to sudden change in the temperature in the greenhouse at 16 days after inoculation.

### 5.2.3. Greenhouse experiment – statistical analysis

Analysis of Variance (ANOVA) was carried out to assess the difference in spike-level FHB-severity among the genotypes and timepoints following the model:

$$(2) y_{ijk} = \mu + G_i + T_j + (GT)_{ij} + e_{ijk}$$

Where  $y_{ijk}$  represents the response (FHB-severity) of the  $i$ -th genotype in the  $j$ -th timepoint for the  $k$ -th observation,  $\mu$  is the overall mean across genotype and timepoints,  $G_i$  is the fixed effect of the  $i$ -th genotype,  $T_j$  is the random effect of  $j$ -th timepoint,  $(GT)_{ij}$  represents the random effect of the interaction between genotype and timepoints, and  $e_{ijk}$  is the residual error.

To assess the variation in spike-level thermal reading, the data was fitted in a linear mixed model:

$$(3) y_{ijkl} = \mu + G_i + T_j + C_k + A + (GT)_{ij} + (GC)_{ik} + (TC)_{jk} + (GTC)_{ijk} + e_{ijkl}$$

Where  $y_{ijkl}$  represents the thermal reading (°C) of the  $i$ -th genotype in the  $j$ -th timepoint at  $k$ -th condition (inoculated, non-inoculated),  $\mu$  is the overall mean across genotypes, timepoints, and conditions,  $G_i$  is the fixed effect of genotype,  $T_j$  is the random effect of timepoints,  $C_k$  is the fixed effect of conditions (inoculated, non-inoculated),  $GT_{ij}$  is the random effect of the interaction between genotype (G) and timepoint (T),  $GC_{ik}$  is the fixed effect of the interaction between genotype (G) and conditions (C),  $TC_{jk}$  is the random effect of the interaction between timepoint (t) and conditions (C),  $GTC_{ijk}$  is the random effect of the interactions between genotype, timepoints, and conditions, and  $e_{ijkl}$  is the residual effect.

Analysis was not executed in a repeated measures manner since different observational units (single spike) were used at each timepoint, and no repeated measurements for the same observational unit was done.

#### *5.2.4. Plot level infrared thermal imaging FHB evaluation – plant material*

Three separate trials were used to test whether infrared thermal imaging could be leveraged under field conditions and its potential application in FHB-resistance selection. The trials were composed of soft red and soft white winter wheat genotypes (305 genotypes in 2022, 398 in 2023, and 309 in 2024) developed by the Michigan State University Wheat Breeding and Genetics team, as well as commercially released varieties and breeding lines developed by various breeding programs. In addition, genotypes MI14W0190 and Ambassador were used as resistant and susceptible checks, respectively, for all the trials.

#### *5.2.5. Plot level infrared thermal imaging FHB evaluation – field establishment, inoculation, and evaluation*

*Fusarium graminearum* cultures were collected from multiple locations: Huron, Ingham, Monroe, Tuscola, Sanilac, and Saginaw counties in Michigan, and were used the trials. FHB nursery was established in a completely randomized design (CRD) with two replications per genotype. Preparation of isolates, field inoculation, and maintenance of FHB nursery was carried out following the procedure described by Concepcion et al., (2024).

FHB related traits were determined by visual and subjective assessment of FHB-severity (average % infection in each spike in the plot) and FHB-Incidence (average % of spikes with FHB infection in the plot). FHB-Index was calculated following the formula:

$$(4) \text{ FHB Index} = \frac{(\text{FHB Severity} * \text{FHB Incidence})}{100}$$

#### *5.2.6. Plot level infrared thermal imaging FHB evaluation – image acquisition and processing*

Imaging was carried out using the same Flir E60 Infrared Camera (320 x 240 IR resolution, 18 mm focal length) (Teledyne FLIR, USA) used in the greenhouse experiment, with default of 0.95 emissivity, 1 meter distance, 20°C external optic temperature. Images were taken above the plots (approximately 175 cm from the bases of the plant) at an angle of approximately 45°C. Images were stored and processed in Flir Tools, where the heads were segmented from the rest of the plot using the interval option to adjust the upper and lower limit of the temperature range. The average between upper and lower limit was generated to determine the estimated plot-level thermal readings.

#### *5.2.7. Plot level infrared thermal imaging FHB evaluation – data analysis and FHB-resistance selection*

Prior to downstream analyses, plot level observations of FHB-severity, FHB-incidence, FHB-index, and thermal readings were spatially adjusted, treating range and row as random effect, and genotype as fixed effect using the R package “SpATs” (Rodriguez-Alvarez et al., 2018).

After spatial adjustment, analysis of variance (ANOVA) was carried out to assess the variation in FHB-severity, FHB-incidence, FHB-index, and thermal readings among the genotypes following a linear model:

$$(5) y_{ij} = \mu + G_i + e_{ij}$$

Where  $y_{ij}$  is the response of the  $i$ -th genotype (G) in the  $j$ -th observation,  $G_i$  is the fixed effect of genotypes, and  $e_{ij}$  is the residual effect.

Performance of the genotypes were compared against the checks Ambassador and MI14W0190 using Dunnett’s Test at alpha 0.05 for all FHB related traits and thermal readings. Pearson’s and Spearman Rank Correlation were determined to assess the relationship between

FHB-related traits and thermal readings. All analyses were conducted in R 4.4.0 (R Core Team, 2021).

Clustering-based selection among the genotypes was carried out using Unsupervised K-Means clustering using “nbclust” package in R (Charrad et al., 2014) following the formula:

$$(6) J = \sum_{i=1}^n \sum_{k=1}^K w_{ik} \|x_i - \mu_k\|^2$$

Where  $J$  is the K-means objective function (within-cluster sum of squared errors),  $n$  is the total number of data points,  $K$  is the number of clusters,  $x_i$  represents data points or observations,  $\mu_k$  is the centroid (mean vector) of cluster  $k$ ,  $\|x_i - \mu_k\|^2$  is the squared Euclidean distance between  $x_i$  and its assigned cluster centroid,  $w_{ik}$  is the binary indicator variable ensuring that each point belongs to only one cluster (1, if  $x_i$  belongs to cluster  $k$ , 0 if otherwise). For consistency, three clusters (centers = 3) for all the clustering approaches (trait combinations) conducted.

Following the clustering, selection differential ( $S$ ) was determined for each cluster following the formula:  $S = \bar{X}_s - \bar{X}$ , where  $\bar{X}_s$  is the mean of phenotype of genotypes in each cluster (cluster mean) and  $\bar{X}$  is the mean phenotype of all genotypes. Selection intensity was carried out using the formula:

$$(7) i = \frac{S}{\sigma_P}$$

Where  $S$  is the selection differential and  $\sigma_P$  is the phenotype standard deviation. Selection intensity was calculated separately for severity, incidence, and index for all clusters from all trait combinations.

### 5.3. Results

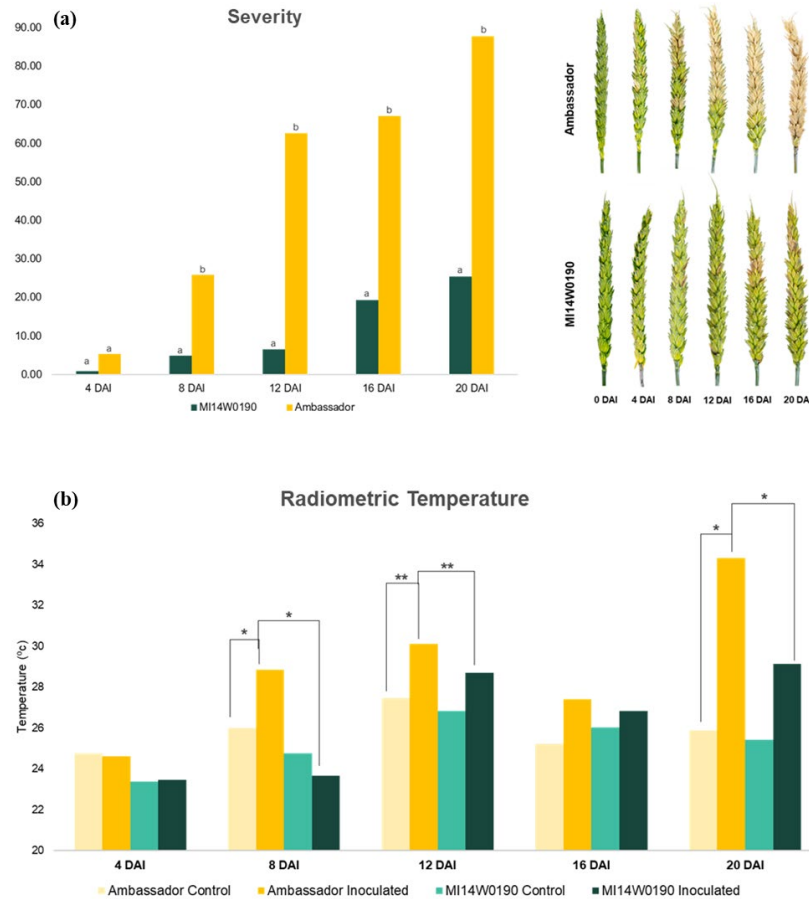
#### 5.3.1. Spike-level FHB evaluation and infrared thermal imaging

To evaluate the potential utility of infrared thermal imaging in differentiating the responses of resistant and susceptible genotypes to FHB infection, spike-level FHB-severity and thermal readings were compared between the resistant genotype MI14W0190 and the susceptible genotype Ambassador.

A rapid progression of FHB infection was observed in Ambassador, with genotype, timepoint, and their interaction showing a significant effect on FHB severity ( $p < 0.0001$ ) (Figure 5.1a). Compared to MI14W0190, Ambassador exhibited significantly higher FHB severity starting at 8 days after inoculation, with differences ranging from 21.02% to 62.36% between the two genotypes (Figure 5.1a).

Consistent with the observations in FHB severity, genotype, timepoint, condition (inoculated vs. non-inoculated), and all interactions had a significant effect on variations in thermal readings ( $p < 0.0001$ ) (Figure 5.1b). In addition, the progression of FHB infection was recorded and observed using infrared thermal imaging (Figure 5.1b). A significant temperature difference of 5.20°C was recorded between the inoculated heads of MI14W0190 and Ambassador starting at eight days after inoculation (Figure 5.1b).

Notably, at around 16 days after inoculation, greenhouse conditions were less ideal compared to other timepoints, as increased temperatures inside the greenhouse affected the thermal readings. Despite this, differences between MI14W0190 and Ambassador, as well as between inoculated and non-inoculated conditions, remained observable (Figure 5.1b).



**Figure 5.1.** Development and variation of FHB-infection between resistant and susceptible genotypes. (a,b) Progression and variation of FHB-infection in resistant genotype MI14W0190 and susceptible check Ambassador at 4 to 20 days after inoculation. (b) Variation in spike-level thermal readings between inoculated (FHB-infected) and non-inoculated spikes of Ambassador and MI14W0190 at 4 to 20 days after inoculation. Means of the same letter are not significantly different (p.value: <0.05). \*Significantly different at alpha 0.1. \*\*Significantly different at alpha 0.05. Comparisons are within a timepoint.

### 5.3.2 Field evaluation: FHB-severity

Significant variation in FHB-severity was observed among the genotypes in the 2022 trial (p.value: <0.0001), 2023 trial (p.value: <0.0001), and 2024 trial (p.value:<0.0001). Across the trials, most genotypes exhibited significantly lower FHB-severity than the susceptible check Ambassador (p < 0.05). In the 2022 trial, 273 (83%) genotypes had FHB-severity lower than Ambassador by 11.13% to 32.99%, while MI21R0256 and MI21R0275 had significantly higher

severity. Compared to the resistant check MI14W0190, 188 (57%) showed no significant difference. In the 2023 trial, 349 (88%) genotypes had 30% to 49% lower FHB-severity than Ambassador, and 328 (83%) showed no difference from MI14W0190. In the 2024 trial, 285 (93%) genotypes had significantly lower FHB-severity than Ambassador, while 199 (65%) showed no difference from MI14W0190. Overall, most genotypes demonstrated significantly lower FHB-severity than Ambassador while remaining comparable to MI14W0190.

#### *5.3.3. Field evaluation: FHB-incidence*

Significant variation in FHB-incidence was observed among the genotypes in the 2022 trial (p.value: <0.0001), 2023 trial (p.value: <0.0001), and 2024 trial (p.value: <0.0001). In the 2022 trial, 92 (28%) genotypes had significantly lower FHB-incidence than the susceptible check Ambassador by 17% to 51% ( $p < 0.05$ ), while MI210256 and MI21R0038 exhibited significantly higher FHB-incidence. Additionally, 42 (13%) genotypes showed no significant difference from the resistant check MI14W0190. In the 2023 trial, although most genotypes had lower FHB-incidence, only 13 (3%) showed significantly lower FHB-incidence compared to Ambassador by approximately 50% ( $p < 0.05$ ), while 382 (96%) genotypes had no significant difference from MI14W0190. In the 2024 trial, 152 (49%) genotypes exhibited significantly lower FHB-incidence than Ambassador (14% to 43% lower,  $p < 0.05$ ). Compared to MI14W0190, 220 (71%) showed no significant difference, while 54 (17.5%) had significantly lower FHB-incidence.

#### *5.3.4. Field evaluation: FHB-index*

Significant variation in FHB-index was observed in the 2022 trial (p.value:<0.0001), 2023 trial (p.value: <0.0001), and 2024 trial (p.value: <0.0001). In the 2022 trial, 252 (77%) genotypes had significantly lower FHB-index than the susceptible check Ambassador by 10 to 28 units, while MI21R0256 and MI210275 showed significantly higher values ( $p < 0.05$ ). Additionally, 130 (40%)

genotypes showed no significant difference from the resistant check MI14W0190. In the 2023 trial, 349 (88%) genotypes exhibited significantly lower FHB-index than Ambassador by 20 to 31 units ( $p < 0.05$ ), while 377 (95%) showed no significant difference from MI14W0190. In the 2024 trial, 288 (94%) genotypes had significantly lower FHB-index than Ambassador by 7 to 26 units ( $p < 0.05$ ), whereas 217 (70%) showed no significant difference from MI14W0190. Overall, most genotypes displayed a significantly lower FHB-index than Ambassador while remaining comparable to MI14W0190.

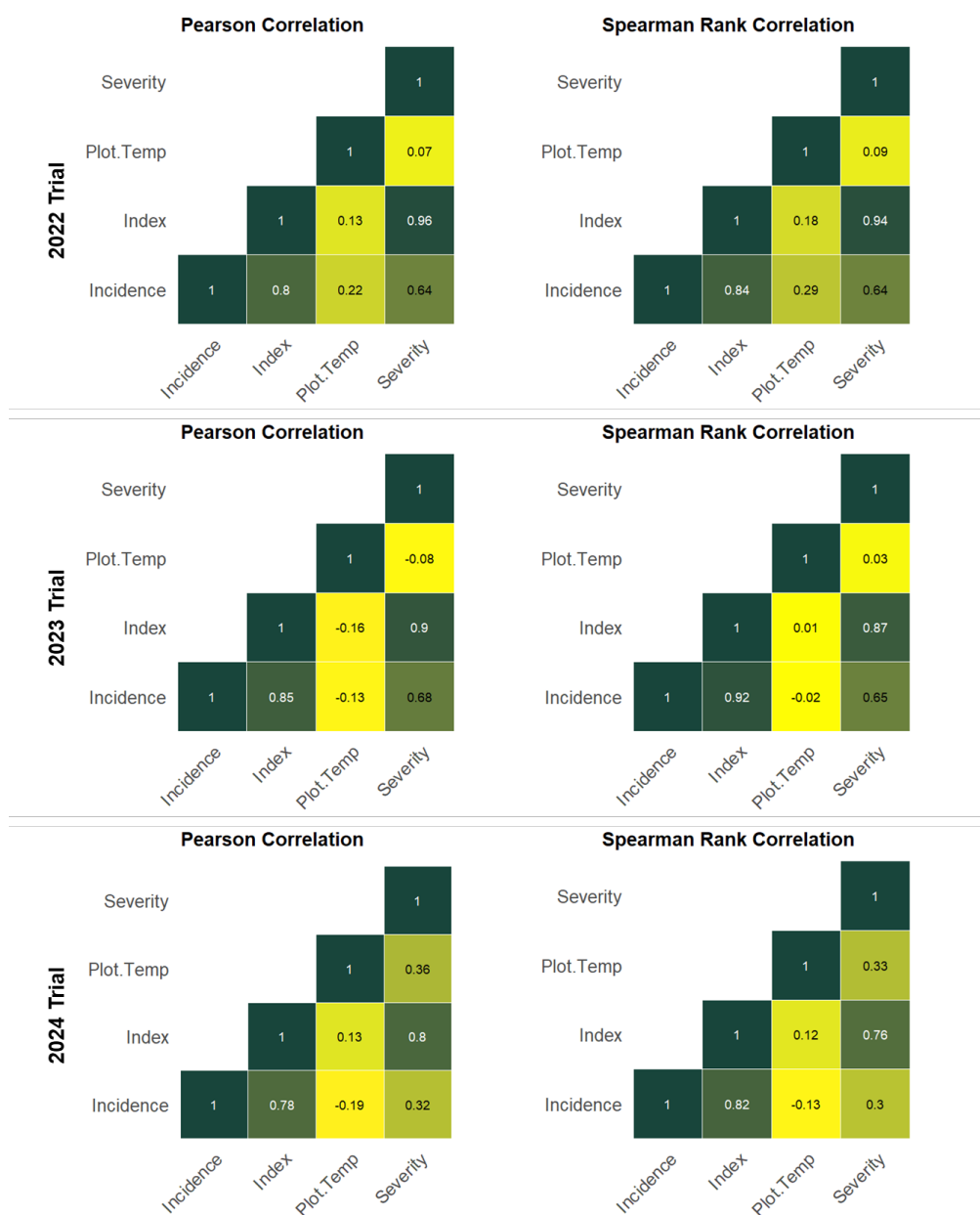
#### *5.3.5. Field Evaluation: plot-level thermal readings*

Significant variation in plot-level thermal readings was observed among the genotypes (p.value:  $< 0.0001$ ) in the 2022 trial. However, only MI21R0207 exhibited a significantly lower thermal reading than the susceptible check Ambassador by 7°C ( $p < 0.0001$ ), while 33 (10%) genotypes had significantly higher thermal reading than the resistant check MI14W0190 ( $p < 0.05$ ). From the 2023 trial no significant variation was observed among the genotypes (p.value: 0.9967) and no genotypes showed a significant difference from either of the checks ( $p < 0.05$ ). In the 2024 trial, significant variation in plot-level thermal readings was observed (p.value:  $< 0.0001$ ), with 129 (42%) genotypes recording lower thermal readings than Ambassador by 2°C to 7°C ( $p < 0.05$ ), while 179 (58%) showed no significant difference from MI14W0190.

#### *5.3.6. Phenotypic relationship among FHB-related traits and with plot-level thermal readings*

Correlation analysis was conducted to examine the relationship between plot-level thermal readings and FHB-related traits—severity, incidence, and index. Both Pearson's and Spearman's rank correlations were assessed to determine potential linear or non-linear relationships, respectively.





**Figure 5.2.** Pearson and spearman rank correlation showing potential relationship between FHB-related traits and plot-level thermal readings (Plot.Temp) at different trials.

Moderate to high Pearson correlations were observed between FHB-index and FHB-incidence (0.78 to 0.85) and between FHB-index and FHB-severity (0.80 to 0.96) (Figure 5.2). Similarly, Spearman's rank correlation showed moderate to high correlation for FHB-index with FHB-incidence (0.82 to 0.92) and FHB-severity (0.76 to 0.94) (Figure 5.2). These results were expected, as FHB-index is a function of FHB-incidence and FHB-severity. Interestingly, higher

Spearman's rank correlations compared to Pearson correlations for FHB-index and FHB-incidence suggest a possible non-linear relationship, whereas lower Spearman's rank correlations for FHB-index and FHB-severity indicate a predominantly linear association.

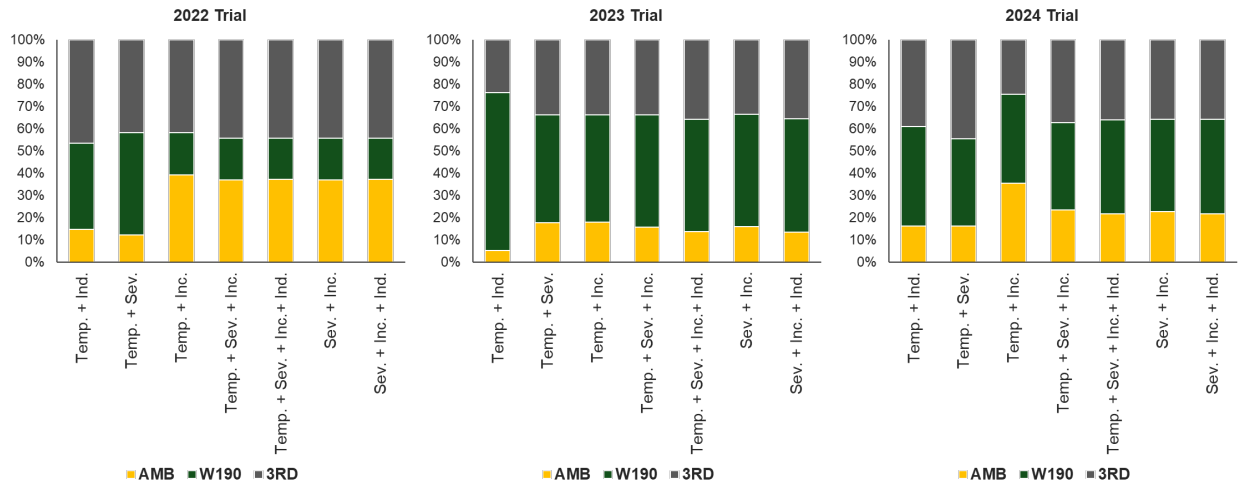
Correlations between plot-level thermal readings and FHB related traits were weak and inconsistent, with negative correlations observed in the 2023 trial, suggesting no direct relationship under field conditions (Figure 5.2). However, in 2024, a moderately positive Pearson correlation (0.36) was observed between FHB-severity and thermal readings, indicating a potential linear relationship (Figure 5.2). In 2022, a weak Spearman's rank correlation (0.29) between FHB-incidence and thermal readings suggested a possible non-linear association (Figure 5.2).

Overall, these findings suggest that under field conditions, plot-level thermal readings may not be as reliable for indirect selection alone for FHB resistance evaluation and selection compared to spike-level and single-plant scale evaluation under greenhouse conditions.

#### *5.3.7. Unsupervised K-means clustering*

Since spike-level radiometric temperature demonstrated strong potential for distinguishing between resistance and susceptibility to FHB while potentially lacking a direct relationship with FHB-related traits under field conditions, we investigated whether plot-level thermal readings could still complement FHB-related traits—severity, incidence, and index—for FHB resistance selection. To streamline the selection process, we applied unsupervised K-means clustering using various trait combinations. Additionally, we assessed selection intensity each cluster to evaluate the effectiveness of this clustering approach in FHB resistance selection.

Across all trials, the number of genotypes per cluster varied depending on the trait combinations used for clustering. A consistent pattern emerged: more genotypes clustered with the resistant



**Figure 5.3.** Differences in the percentage (number) of genotypes belonging to each cluster using unsupervised k-means clustering with various trait combinations. AMB – cluster with susceptible check Ambassador, W190 – cluster with resistant check MI14W0190, 3RD – third cluster not containing either of the checks.

check MI14W0190 in Temp. + Ind. cluster, whereas fewer genotypes clustered with MI14W0190 in other clusters (Figure 5.3).

Selection intensity was calculated by dividing the selection differential (presented in Supplementary Tables 5.1 to 5.3) by the phenotypic standard deviation. Lower selection intensity values were desirable, as FHB resistance selection was based on minimizing FHB severity, incidence, and index.

In the 2022 trial, lower selection intensity values were observed for FHB-related traits in clusters Temp+Ind and Temp+Sev. Notably, cluster Temp+Sev+Inc+Ind and Sev+Inc+Ind appeared to be the most effective in selecting for lower FHB-severity. Similar trends were observed for FHB-incidence and FHB-index (Table 5.1).

In the 2023 trial, Temp+Sev resulted in lower selection intensity value for FHB severity. For FHB incidence, Temp+Inc had the lowest selection intensity value, primarily grouping genotypes with reduced FHB incidence. These results align with the expectations, as clustering were done using

**Table 5.1.** Selection intensity in severity, incidence, and index among genotypes evaluated in 2022 belonging to various K-means derived clusters using different trait combinations. Lower selection intensity value is more desirable.

Trait combinations for clustering	2022 Trial								
	Severity			Incidence			Index		
	AMB	W190	3RD	AMB	W190	3RD	AMB	W190	3RD
Temp. + Ind.	1.59	-0.85	0.21	1.10	-0.93	0.42	5.61	-0.91	0.22
Temp. + Sev.	1.83	-0.81	0.36	0.87	-0.84	0.38	1.76	-0.77	0.33
Temp. + Inc.	0.56	-1.04	-0.06	0.95	-1.59	-0.18	0.69	-1.19	-0.18
Temp. + Sev. + Inc.	0.65	-1.09	-0.01	0.99	-1.59	-0.14	0.85	-1.22	-0.19
Temp. + Sev. + Inc. Ind.	0.72	-1.10	-0.14	0.96	-1.60	-0.14	0.90	-1.23	-0.24
Sev. + Inc.	0.65	-1.09	-0.07	0.99	-1.57	-0.14	0.85	-1.14	-0.19
Sev. + Inc. + Ind.	0.72	-1.10	-0.14	0.96	-1.60	-0.14	0.90	-1.23	-0.24

AMB – cluster with susceptible check Ambassador, W190 – cluster with resistant check MI14W0190, 3RD – third cluster not containing either of the checks. Temp. – Plot-level thermal readings, Sev. – FHB-Severity, Inc. – FHB-Incidence, Ind. – FHB-Index

**Table 5.2.** Selection intensity in severity, incidence, and index among genotypes evaluated in 2023 belonging to various K-means derived clusters using different trait combinations. Lower selection intensity value is more desirable.

Trait combinations for clustering	2023 Trial								
	Severity			Incidence			Index		
	AMB	W190	3RD	AMB	W190	3RD	AMB	W190	3RD
Temp. + Ind.	2.47	-0.48	0.88	2.15	-0.48	0.96	3.11	-0.50	0.81
Temp. + Sev.	1.72	-0.78	0.21	1.15	-0.54	0.17	1.63	-0.60	0.00
Temp. + Inc.	1.20	-0.54	0.12	1.65	-0.85	0.32	1.46	-0.64	0.07
Temp. + Sev. + Inc.	1.52	-0.57	-0.86	1.69	-0.78	0.42	1.83	-0.64	0.09
Temp. + Sev. + Inc. Ind.	1.68	-0.57	0.16	1.73	-0.81	0.48	1.50	-0.64	0.13
Sev. + Inc.	1.61	-0.55	0.13	1.68	-0.78	0.41	1.81	-0.64	0.09
Sev. + Inc. + Ind.	1.71	-0.57	0.17	1.73	-0.80	0.50	2.03	-0.63	0.14

AMB – cluster with susceptible check Ambassador, W190 – cluster with resistant check MI14W0190, 3RD – third cluster not containing either of the checks. Temp. – Plot-level thermal readings, Sev. – FHB-Severity, Inc. – FHB-Incidence, Ind. – FHB-Index

**Table 5.3.** Selection intensity in severity, incidence, and index among genotypes evaluated in 2024 belonging to various K-means derived clusters using different trait combinations. Lower selection intensity value is more desirable.

Trait combinations for clustering	2024 Trial								
	Severity			Incidence			Index		
	AMB	W190	3RD	AMB	W190	3RD	AMB	W190	3RD
Temp. + Ind.	1.48	-0.64	0.12	1.06	-0.78	0.45	1.75	-0.84	0.23
Temp. + Sev.	1.72	-0.95	0.22	0.46	0.45	0.12	1.40	-0.74	0.14
Temp. + Inc.	0.35	-0.08	-0.38	1.05	-0.10	-1.35	0.87	-0.17	-0.97
Temp. + Sev. + Inc.	0.97	-0.45	-0.33	0.88	0.49	-1.07	1.35	-0.09	-0.80
Temp. + Sev. + Inc. Ind.	1.28	-0.35	-0.36	1.00	0.42	-1.09	1.52	-0.07	-0.84
Sev. + Inc.	1.26	-0.38	-0.36	0.96	0.42	-1.10	1.48	-0.09	-0.84
Sev. + Inc. + Ind.	1.28	-0.36	-0.36	1.00	0.41	-1.10	1.52	-0.08	-0.84

AMB – cluster with susceptible check Ambassador, W190 – cluster with resistant check MI14W0190, 3RD – third cluster not containing either of the checks. Temp. – Plot-level thermal readings, Sev. – FHB-Severity, Inc. – FHB-Incidence, Ind. – FHB-Index

plot-level thermal readings were done with FHB- severity or -incidence, respectively. For FHB-index, clusters Temp+Ind, Temp+Sev, and Sve+Inc+Ind showed higher selection intensity values. Overall, Temp+Inc consistently yielded low selection intensity value across all traits, whereas Temp+Ind performed the worst, with higher selection intensity values (Table 5.2).

Similar to the 2023 trial, the 2024 trial showed that Temp+Sev had the lowest selection intensity value for FHB severity, while Temp+Inc had the lowest selection intensity for FHB incidence and FHB index. Interestingly, in 2024, the least favorable clusters (with the highest selection intensity values) were Temp+Sev+Inc+Ind, Sev+Inc, and Sev+Inc+Ind (Table 5.3).

Comparing all trials, Temp+Inc—plot-level thermal readings and FHB-incidence—emerged as the most effective approach for identifying genotypes with lower FHB severity, incidence, or index. In contrast, clusters relying solely on FHB-related traits, performed worse than

clusters that incorporated plot-level thermal readings, suggesting that including plot-level thermal reading may enhance the identification of resistant genotypes.

## **5.4. Discussion**

### *5.4.1. Infrared thermal imaging discerns susceptible and resistant genotypes at single-spike level*

Single-plant evaluations have been established for FHB-resistance evaluation and selection to increase the number of genotypes evaluated and reduce the spatial resources (e.g., experimental field) required. However, this approach requires manual counting of spikelets – both healthy and infected – to assess the FHB-resistance, which in turn increases labor intensity and evaluation time. Over the years, the use of infrared thermal imaging for FHB-resistance evaluation and selection has been explored and demonstrated (Masri et al., 2017; Mahlein et al., 2019). The results of these studies were in conjunction with our results where the use of handheld infrared thermal camera Flir E60 allowed the delineation between resistant and susceptible genotypes, as early as eight days post inoculation which is earlier compared to previous reports (Masri et al., 2017). Thus, this solidifies the use of radiometric temperature to select between resistance and susceptible genotypes at the single plant and single spike level.

However, there was a significant drop in radiometric temperature at 16 DAI. This is due to the sudden increase in the temperature in the greenhouse during the time of imaging, requiring adjustment of the observations based on ambient temperature during the course of the greenhouse experiment. When the environmental conditions cannot be fully controlled, as in the case of our study, recording and applying ambient temperatures to adjust the radiometric temperatures should be done. Provided that all conditions are controlled and consistent throughout the evaluation period, we believe that adjustment using ambient temperature should not be necessary.

#### *5.4.2. Enhancing selection for FHB resistance using K-means clustering and infrared thermal imaging*

Evaluation and selection for Fusarium Head Blight (FHB) resistance in field trials primarily rely on two key traits: severity (percentage of infection in spikes) and incidence (percentage of spikes infected). While both traits are critical and interconnected, selecting based on a single trait is often insufficient. For instance, a genotype may exhibit high incidence but low severity, or vice versa, making selection decisions challenging. To address this, the FHB index is commonly used to integrate severity and incidence. However, extreme values in one trait can disproportionately influence the index, making selection potentially misleading.

To refine selection, we explored the use of unsupervised k-means clustering, which allows simultaneous consideration of multiple FHB traits independently rather than relying solely on the FHB index. If selection prioritizes lower severity, for example, a genotype may be chosen if it has significantly lower severity than the susceptible check and is comparable to or better than the resistant check. However, when selection is based on individual trait means, a large proportion of genotypes may still qualify, limiting selection intensity. By contrast, k-means clustering across multiple FHB-related traits improves selection efficiency by reducing the number of selected genotypes while still prioritizing trait reduction.

Given the subjective nature of traditional FHB trait evaluation, we sought a more objective metric and explored the use of plot-level infrared thermal imaging. Our preliminary findings demonstrated that infrared thermal imaging effectively distinguished resistant from susceptible genotypes under controlled conditions. To extend this approach, we implemented plot-level infrared thermal imaging in the field. While prior studies have used UAV-based imaging for FHB evaluation with varying success, handheld infrared thermal imaging offers a more practical and

logistically feasible alternative. However, we anticipated a reduced resolution and increased variability due to confounding field conditions. This expectation was confirmed, as plot-level radiometric temperature showed no direct correlation with FHB-related traits.

Despite this, we assessed whether plot-level infrared thermal imaging could contribute to selection in a novel way—not as a proxy or indirect selection trait, as UAV-derived indices have been used (Zhang et al., 2019; Francesconi et al., 2021), but as an independent, uncorrelated trait. Surprisingly, despite its lack of direct correlation with FHB traits, incorporating plot-level radiometric temperature into unsupervised k-means clustering improved selection intensity for FHB resistance. Notably, clustering based on plot-level radiometric temperature and FHB-incidence enhanced selection not only for reduced FHB incidence but also for lower severity.

These findings do not suggest prioritizing incidence over severity but may be explained by canopy compensation effects during thermal imaging. Higher incidence indicates a higher percentage of infected spikes, which could elevate overall plot temperatures, even if severity is low. Conversely, in plots with low incidence but high severity, uninfected spikes emitting lower temperatures may dilute the thermal signal.

Despite observations, implementing close-range infrared thermal imaging in field conditions requires further refinement. Improved sensitivity and resolution may enhance detection of subtle temperature differences among genotypes, while advancements in image processing could enhance scalability and implementation. Future research should focus on optimizing these aspects to fully leverage infrared thermal imaging as a selection tool for FHB resistance.

#### *5.4.3. Lack of direct relationship between plot-level thermal readings with FHB-related traits*

Under greenhouse conditions, at a single-spike resolution, we observed the expected response: higher radiometric temperature was associated with higher disease severity and



increased Fusarium Head Blight (FHB) susceptibility. In field conditions, despite some level of correlation between radiometric temperature and FHB severity or incidence in certain trials, no direct relationship between radiometric temperature and FHB-related traits could be established. Several confounding factors likely contributed to this discrepancy.

First, imaging resolution played a significant role. At the single-spike level in the greenhouse, images were captured at a much closer range, allowing for higher spatial resolution and more detailed information in each pixel. In contrast, plot-level imaging in the field required an overhead perspective to capture as many spikes as possible. This broader view inherently reduced spatial resolution, limiting the ability to extract precise temperature information from individual spikes. Additionally, at the plot level, the proportion of pixels representing spike tissue decreased, further affecting temperature accuracy.

Second, features captured in the image varied between conditions. In the greenhouse, individual spikes were detached from the plant and imaged against a solid white background, minimizing extraneous features that could influence radiometric temperature readings. Since the white background maintained a consistent temperature, it had little to no impact on the thermal readings of the spikes. However, at the plot level in the field, the imaging frame contained multiple additional elements, including foliage, stems, soil, and weeds. These features, which emit and absorb heat differently, introduced variability in radiometric temperature readings, ultimately influencing the measured temperature of the spikes.

Third, image segmentation presented a challenge, particularly in infrared thermal imaging. Unlike RGB image segmentation, where color differences help distinguish plant parts, thermal segmentation is more complex because all pixels represent temperature values. Under field conditions, spike pixels may share similar radiometric temperatures with other features, such as

senescing foliage. As a result, the final plot-level radiometric temperature reading incorporates temperature information from multiple elements. In contrast, at the single-spike level, segmentation is more straightforward because the temperature contrast between the wheat spike and the white background allows for clear differentiation at the pixel level.

Fourth, environmental effects and level of infection. While the majority of the reasons laid out here have been on how the imaging side: resolution, confounding factors, and segmentation, we cannot oversee the potential environmental effects affecting the trial, more importantly the thermal readings. Microclimatic conditions varied across developmental stages and around evaluation periods (complete weather data is presented in Supplementary Table 5.4). Variation in precipitation, soil and ambient temperature, and relative humidity varied at different developmental stages, especially during flowering. As environmental factors affect how thermal readings were recorded by the sensor, these microclimatic changes most likely have affected the thermal readings as well. Further, microclimatic conditions varied during the field inoculation dates, which may have directly affected the level of infection in the FHB nursery and therefore may have affected the radiometric temperature emitted by the wheat spikes - regardless of whether it is infected or not.

## **5.5. Conclusion**

Infrared thermal imaging successfully delineated the response and the progression of FHB-infection at a single-spike level, showing potential use in single plant evaluations under greenhouse conditions. However, the scalability of close-range infrared thermal imaging for plot-level FHB-evaluation was met with certain challenges, particularly the lack of direct relationship between plot-level radiometric temperature with FHB-related traits. Despite this, there is a potential use of plot-level radiometric temperature as an uncorrelated and independent trait in unsupervised k-

means clustering to aid in selection for reduced FHB-related traits resulting in better selection intensity.

## REFERENCES

- Al Masri, A., Hau, B., Dehne, H.-W., Mahlein, A.-K. & Oerke, E.-C. Impact of primary infection site of *Fusarium* species on head blight development in wheat ears evaluated by IR-thermography. *Eur J Plant Pathol* **147**, 855–868 (2017).
- Concepcion, J. S., Noble, A. D., Thompson, A. M., Dong, Y. & Olson, E. L. Genomic regions influencing the hyperspectral phenome of deoxynivalenol infected wheat. *Scientific reports* **14**, 19340 (2024).
- Foroud, N. A. *et al.* Trichothecenes in Cereal Grains – An Update. *Toxins* **11**, 634 (2019).
- Francesconi, S., Harfouche, A., Maesano, M. & Balestra, G. M. UAV-Based Thermal, RGB Imaging and Gene Expression Analysis Allowed Detection of Fusarium Head Blight and Gave New Insights Into the Physiological Responses to the Disease in Durum Wheat. *Front. Plant Sci.* **12**, (2021).
- Lenth R. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.10.7-100001. <https://rvlenth.github.io/emmeans/>, <https://rvlenth.github.io/emmeans/> (2025).
- Mahlein, A.-K. Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Disease* **100**, 241–251 (2016).
- Mahlein, A.-K. *et al.* Comparison and Combination of Thermal, Fluorescence, and Hyperspectral Imaging for Monitoring Fusarium Head Blight of Wheat on Spikelet Scale. *Sensors* **19**, 2281 (2019).
- McMullen, M. *et al.* A Unified Effort to Fight an Enemy of Wheat and Barley: Fusarium Head Blight. *Plant Disease* **96**, 1712–1728 (2012).
- Mesterházy, A. Types and components of resistance to Fusarium head blight of wheat. *Plant Breeding* **114**, 377–386 (1995).
- Mesterházy, Á., Bartók, T., Mirocha, C. G. & Komoróczy, R. Nature of wheat resistance to Fusarium head blight and the role of deoxynivalenol for breeding. *Plant Breeding* **118**, 97–110 (1999).
- Miedaner, T. *et al.* Effects of genotype and genotype-environment interaction on deoxynivalenol accumulation and resistance to Fusarium head blight in rye, triticale, and wheat. *Plant Breeding* **120**, 97-105 (2008).
- Mustafa, G. *et al.* Fusarium head blight monitoring in wheat ears using machine learning and multimodal data from asymptomatic to symptomatic periods. *Front. Plant Sci.* **13**, (2023).

- R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2023).
- Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A. & Eilers, P. H. C. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics* **23**, 52–71 (2018).
- Steiner, B. *et al.* Breeding strategies and advances in line selection for Fusarium head blight resistance in wheat. *Trop. plant pathol.* **42**, 165–174 (2017).
- Wilson, W., Dahl, B. & Nganje, W. Economic costs of Fusarium Head Blight, scab and deoxynivalenol. *World Mycotoxin Journal* **11**, 291–302 (2018).
- Zhang, N. *et al.* Development of Fusarium head blight classification index using hyperspectral microscopy images of winter wheat spikelets. *Biosystems Engineering* **186**, 83–99 (2019).
- Xu, X.-M. *et al.* Relationship Between the Fungal Complex Causing Fusarium Head Blight of Wheat and Environmental Conditions. *Phytopathology*® **98**, 69–78 (2008).

## **CHAPTER VI: INTEGRATED PHENOMIC AND GENOMICS: PROSPECTS, CAVEATS, IMPLICATION IN GENETIC GAIN AND APPLICATION IN WHEAT BREEDING**

This dissertation explored the integration of phenomic information with genomic data to enhance prediction accuracy and understand the genetic basis of phenomic traits in soft winter wheat. Specifically, this work aimed to (1) increase the prediction accuracy for economically important traits (grain yield, Fusarium head blight resistance) and (2) provide insights into the potential genetic control of phenomic traits and their associations with these target traits. By examining the relationship between high-throughput phenomic data and genomic information, this research sought to establish a framework for leveraging multi-omics information in plant breeding.

An overarching goal of this dissertation was to demonstrate how integrating phenomic and genomic data improves genetic gain. Understanding the interplay between phenomic traits and genomic information allows for a more precise assessment of genetic potential, ultimately enhancing breeding efficiency.

The following sections present the author's perspective on multi-omic breeding, emphasizing its implications predictive breeding and future advancements in plant breeding.

### *6.1. Phenotype vs. Breeding Value: What are we actually predicting?*

In modern plant breeding, distinguishing between genomic and phenotypic prediction is essential for understanding how different types of data contribute to assessing a genotype's potential. When using genomic information, we generate genomically estimated breeding values (GEBVs) to assess the genetic potential of individuals for a given trait, independent of environmental effects or genotype-by-environment (G×E) interactions. While true breeding values remain impossible to predict due to the complexity of inheritance and environmental influences, GEBVs provide a powerful approximation of genetic merit. The primary applications of genomic

prediction include generation advancement, where superior genotypes are identified for the next breeding cycle, and parental selection, which involves choosing individuals with the highest breeding potential for population development. Since genomic prediction operates under the assumption that genetic effects are stable across environments, it allows breeders to make informed long-term genetic improvement decisions.

In contrast, when using non-genomic information, such as phenomic data from high-throughput imaging sensors, the focus shifts from predicting breeding values to predicting phenotypic performance. The phenotype of an individual is influenced by genetic effects (G), environmental effects (E), and their interaction (G×E). Our findings establish that phenomic traits exhibit genetic control, as evidenced by the identification of genomic regions associated with specific hyperspectral waveband ranges linked to deoxynivalenol accumulation in wheat kernels and genomic loci associated with RGB and multispectral vegetation indices (VIs). Furthermore, we have demonstrated that year-to-year environmental variability significantly impacts phenomic traits, reinforcing their role as a bridge between genetic potential and real-world performance. Since phenomic traits potentially captured both genetic and environmental influences, they serve as secondary or correlated traits in predictive models for complex traits like grain yield and disease resistance. This concept underpins phenomic selection, where phenomic data is used to infer the potential performance of genotypes under specific environmental conditions.

When combining genomic and phenomic information, the objective remains to be phenotypic prediction, but with an improved understanding of how genetic potential interacts with environmental conditions. Unlike genomic prediction alone, which provides an intrinsic estimate of genetic merit, integrating phenomic data allows us to contextualize performance in real-world environments. For example, consider a genotype with a GEBV for yield of 80 bu/ac, a

phenomically predicted phenotype of 90 bu/ac, and a multi-omic prediction of 85 bu/ac, which accounts for both genetic potential and environmental interactions. In this case, the 85 bu/ac prediction represents the genotype's expected performance, integrating its genetic foundation (GEBV) and phenotypic expression under the given environmental conditions. If this genotype is evaluated under drought stress, its yield is expected to fall below 80–85 bu/ac, whereas if it is grown under optimal management conditions, its yield should approach or exceed 80–90 bu/ac.

This framework highlights the power of multi-omic prediction, which provides a holistic assessment of a genotype's potential by simultaneously considering its breeding value and real-world performance. By leveraging both genomic and phenomic information, breeders can identify stable performers, ensuring consistent yield across environments; optimize selection strategies, balancing genetic potential with environmental adaptability; and enhance prediction accuracy, reducing uncertainty in breeding decisions. Ultimately, the integration of genomics, phenomics, and environmental data enables a more precise and dynamic approach to plant breeding, accelerating genetic gain while ensuring adaptability to changing agricultural conditions.

## *6.2. Implications on Genetic Gain*

Enhancing genetic gain relies on a few important aspects: increasing prediction accuracy, enhancing selection intensity, broadening genetic diversity, and reducing generation interval.

Enhancing selection intensity follows a simple principle: "more entries, more chances of winning." Phenomics enables rapid data collection, reduces labor, and allows for evaluating more genotypes, increasing the likelihood of identifying superior candidates. For example, traditional deoxynivalenol (DON) content estimation requires grinding samples, sending them to an external lab, and waiting for results—often delaying selections until the next trial is already established. In



contrast, hyperspectral imaging eliminates the need for sample processing, significantly reducing labor and time, enabling the evaluation of more entries, and enhancing selection intensity.

Integrating phenomic and genomic data enhances prediction accuracy. As demonstrated in this dissertation, incorporating UAV-derived information as a fixed effect in genomic prediction, or simply averaging phenomic and genomic predictions for deoxynivalenol (DON) content, led to improved accuracy. These multi-omic approaches thereby appear to perform better over relying solely on genomic or phenomic prediction. Additionally, higher prediction accuracy directly translates to increased selection intensity—the more precise the evaluations, the more effective the selection process becomes within a larger candidate pool.

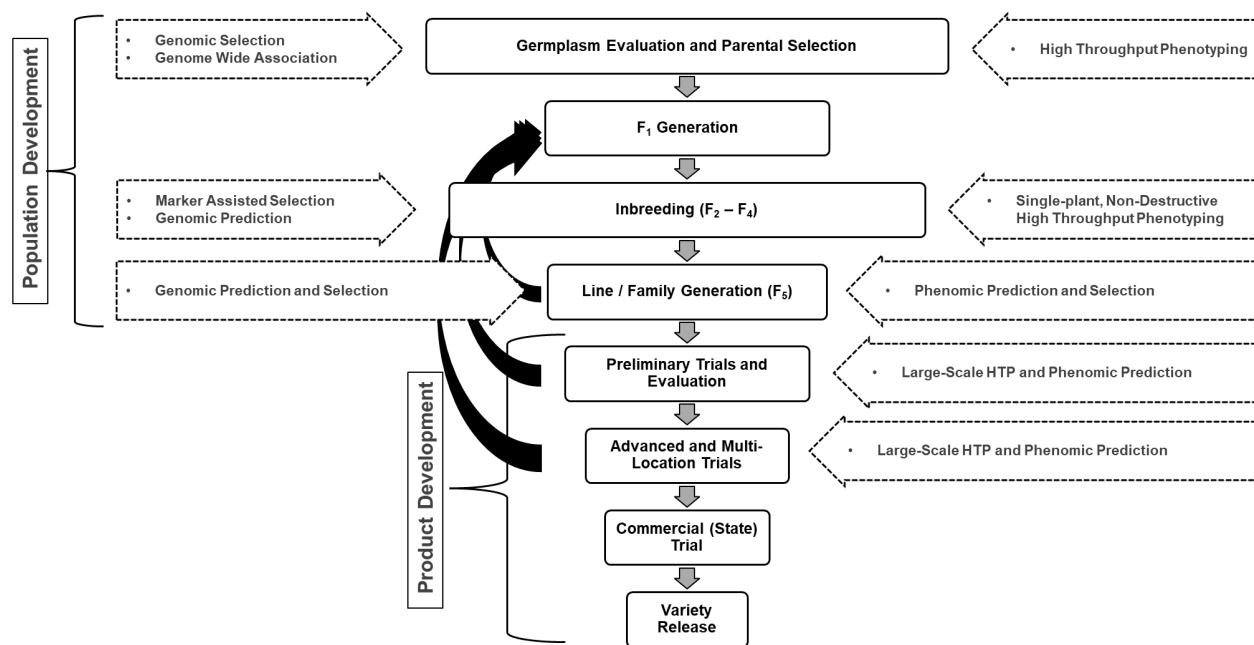
Evaluating a larger number of candidates not only increases selection intensity but also enables the assessment of more diverse populations, potentially enhancing genetic diversity within breeding programs. Beyond evaluation and selection, identifying key genomic regions and significant SNPs means additional QTLs or variants that could be introgressed into breeding materials, thereby broadening the genetic background of breeding pools.

Taken altogether, an integrated phenomic and genomic breeding approach could have positive implications in both short- and long-term genetic gains.

### *6.3. Prospects and Application of Integrated Phenomics and Genomics in a Breeding Pipeline*

The introduction of new breeding approaches, such as phenomics-assisted or multi-omic-based breeding, is often met with skepticism. While further validation and metrics are needed for full implementation, we can explore how phenomic information integrates into a genomics-assisted breeding pipeline.

Assuming an inbred breeding pipeline where establishment of families are done at F<sub>4</sub> to F<sub>5</sub> stages with accompanying genomic selection, or potentially at the preliminary trial, or advanced



**Figure 6.1.** Potential application of genomics and phenomics in a breeding pipeline.

trial. In addition, genotypes selected following genomic prediction are funneled back into population development as parents, basically similar to a recurrent selection-based breeding. Product development, including field trials, multi-location trials, and regional trials follow after family establishment. With this type of breeding pipeline, where does phenomic information fit?

1. *Germplasm Evaluation and Parental Selection* – Genomic prediction has significantly improved parental selection by assessing genetic merit. However, in breeding programs without the capacity to do genomic prediction, phenotypic performance remains a key factor in parental selection. Phenomics can augment genomics-based selection by providing rapid, high-throughput trait evaluation, enabling breeders to assess a larger number of genotypes efficiently. This, in turn, increases the likelihood of identifying superior parental lines and enhances genetic diversity within breeding pools.

2. *Reduction of Selection Pool* – At early breeding stages, rapid and accurate phenotype estimation is essential for efficiently narrowing the selection pool. Phenomics facilitates this by offering non-destructive, high-throughput evaluations that improve selection precision. For instance, in this study, deoxynivalenol (DON) content in F<sub>4:5</sub> breeding lines was predicted using phenomics-assisted, genomics-assisted, and multi-omic-based approaches. These findings suggest that integrating phenomic and genomic predictions at early generations can accelerate selection decisions and improve breeding efficiency.
3. *Enhancing and Complementing Field Performance Evaluation* – Field trials are critical for assessing the performance of breeding lines, particularly for complex traits like grain yield. Field performance trials, including multi-year and multi environment trials, are essential for selecting elite genotypes, but phenomic tools can streamline this process. In this work, UAV-based evaluations at the preliminary trial stage, combined with genomic data, improved grain yield predictions compared to genomic prediction alone. This approach enables the identification of elite genotypes with greater accuracy and reduces evaluation time.

#### 6.4. Caveats and Considerations

The application of phenomics in evaluating important traits – ranging from disease resistant, abiotic stress tolerance, agronomic traits, and quality traits – there are several considerations that must be looked into at when considering using phenomic platforms:

1. *Resources* – currently, there is an abundance of genomic resources available to breeders. More importantly, the cost of genotyping has significantly gone down, allowing breeders to evaluate more genotypes. Therefore, the cost of phenomic platforms is still the biggest consideration in adopting phenomic platforms. Currently, some specialized sensors (e.g.,

hyperspectral camera, 3D sensors) could be expensive and potentially a huge initial advancement. A potential trade-off would either be considering a bigger investment up front and saving later (in phenomics) or continuously investing lesser resources in every trial, saving some resources every trial but would accumulate long term (in genomics).

2. *Plant architecture and the target phenotype* – Selecting the appropriate phenomic platform depends on how the target trait manifests, particularly for traits like disease resistance. For example, infrared thermal imaging is well-suited for Fusarium Head Blight (FHB) detection at the spike level, as temperature variations can indicate infection. However, applying the same approach for deoxynivalenol (DON) content in post-harvest kernels is ineffective, as matured kernels may emit similar radiometric temperatures. Similarly, if a disease manifests as discoloration, RGB and hyperspectral imaging are ideal, whereas if it leads to tissue necrosis, thermal imaging may be more appropriate. Plant architecture plays a crucial role in selecting phenomic platforms. In wheat, handheld, rover-based, and UAV systems are effective as they can capture whole plants or entire plots under field conditions. However, for taller crops like maize, handheld and rover-based systems pose challenges due to plant height, making UAV-based imaging the preferred option. Additionally, the spatial expression of traits should guide platform selection. For example, traits expressed in kernels, such as DON accumulation, are best captured with close-range imaging systems, whereas agronomic traits like yield or plant height are more effectively measured using UAV-based systems. Ultimately, selecting the right phenomic platform requires a comprehensive understanding of plant architecture, the target trait, and its spatial manifestation. This integrated approach ensures optimal sensor selection, enhancing data accuracy and efficiency in phenomics-assisted breeding.

3. *Computational and processing demand* – Image processing for complex phenomic data presents significant computational and automation challenges. Before adopting high-dimensional sensors like hyperspectral cameras, it is crucial to assess whether simpler alternatives, such as RGB imaging using a mobile phone, can provide sufficient information. Simpler approaches should always be prioritized when feasible, as they reduce computational overhead, data storage needs, and processing time. As phenomic platforms capture higher-dimensional data, computational demands increase significantly. For instance, hyperspectral imaging, while highly informative, requires extensive data processing and storage, which may not always be practical. Therefore, balancing data complexity, ease of processing, and resource availability is essential when integrating phenomic tools into breeding programs. Selecting the most efficient yet effective platform ensures both scalability and practicality in phenomics-assisted breeding.
4. *Potential trade-offs of multi-omics-based breeding* – employing phenomic, genomic, or multi-omic approaches involves trade-offs, particularly in terms of time, expertise, and cost. While UAV-based imaging for agronomic traits is rapid, processing and analyzing the images can be time-intensive. Similarly, while genomic data analysis is well-established, it requires specialized expertise, as does phenomic data processing. Thus, breeding programs must decide whether to rely on existing expertise, invest in skill development, or collaborate with specialists to optimize efficiency. Cost is another critical factor. If genotyping becomes more affordable, it may be more practical to genotype more entries rather than invest in phenomic approaches. Conversely, if phenomic evaluations help refine selection before harvest, leaving unselected entries in the field could be viewed as a waste of resources. Balancing these trade-offs requires strategic decision-making, ensuring that

resource allocation aligns with breeding program goals while maximizing efficiency and genetic gain.

#### *6.5. Future Directions in Phenomics and Multi-Omic Breeding*

1. *Rapid and More Automated Image Processing* – Phenomics is rapidly evolving, with significant advancements in image and data processing, particularly for RGB and multispectral imaging. However, hyperspectral imaging for instance still requires further automation and optimization. The development of efficient, high-throughput processing pipelines presents a major opportunity to enhance phenomics' impact on plant breeding by improving data accessibility and reducing processing time.
2. *Cost-Effective and Affordable Phenomics Platforms* – Despite its potential, phenomic platforms remain costly, especially for advanced sensors. As the field advances, increased competition is likely to drive costs down, similar to how genotyping became more affordable over time. Additionally, scientists and engineers have opportunities to develop lower-cost alternatives that provide the same quality of data, making phenomics more accessible for breeding programs.
3. *Integration of Machine Learning and AI* – The rapid advancements in machine learning and AI provide unprecedented opportunities for improving prediction accuracy and complex trait analysis. Neural networks are already enhancing phenomic prediction, leading to more precise selection. Furthermore, AI-driven models enable the integration of diverse datasets, including environomics, transcriptomics, metabolomics, and ionomics. The application of multi-modal neural networks in multi-omics breeding could revolutionize the way complex traits are predicted and utilized, paving the way for more sophisticated data-driven breeding strategies.

While the integration of phenomics into genomics-based breeding is promising, plant breeders must critically assess its implementation within existing breeding pipelines. Rather than hastily adopting new technologies, it is essential to carefully evaluate the advantages, limitations, and feasibility of phenomic approaches. Considerations such as cost, computational demand, processing time, and practical application in selection decisions should be weighed to ensure that phenomics enhances, rather than complicates, breeding efficiency. A strategic and well-integrated approach will maximize the benefits of phenomics while maintaining the effectiveness of established breeding programs.