A MULTI-OMICS MULTI-ENVIRONMENT PREDICTION IN PULSE CROP

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Rica Amor Gregorio Saludares

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Plant Sciences

April 2024

Fargo, North Dakota

North Dakota State University
Graduate School

Title

A MULTI-OMICS MULTI-ENVIRONMENT PREDICTION IN PULSE CROP

By

Rica Amor Gregorio Saludares

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Nonoy Bandillo
Chair

Dr. Qi Zhang

Dr. Megan Orr

Dr. Julie Pasche

Approved:

April 8, 2024
Date

Dr. Richard Horsley
Department Chair

ABSTRACT

Understanding the genetic bases underlying seed yield and protein, and eventually recombining them in desired genetic backgrounds, continues to be a challenge to pulse crop breeders. Phenotypic selection for seed yield and protein in preliminary yield trials is hindered by the need to phenotype a large number of early-generation lines (>10,000) with limited seeds, resulting to trials with few replications and limited environments. In this study, we evaluated and applied a multi-trait multi-environment (MTME) and a multi-omics prediction framework to address phenotyping bottleneck and the complexities underlying negatively correlated traits, and maximize connectivity among genotypes for predicting performance of untested genotypes in diverse set of environments. Using over 200 NDSU modern advanced breeding lines and 300 USDA diverse accessions, our findings demonstrated that MTME prediction significantly enhanced predictive ability by 1.3 and 1.8-fold for yield and protein, respectively. For the environments with low heritability of tested trait, however, using the MTME prediction led to small increases in prediction accuracy. To further maximize connectivity among genotypes and environments, a subset of individuals was included from the testing population that led to 1.6 and 1.2-fold improvement for yield and protein, respectively. Incorporating additional orthogonal information such as gene expression (RNA) into the prediction framework showed potential for further increasing prediction accuracy. Using ~300 USDA diverse accessions assessed in two environments, integrating genotypic and expression data (DNA+RNA) resulted to higher predictive ability (0.48-0.55) over using DNA only (0.42) or RNA only (0.43-0.53). Overall, we found that maximizing the relationship among genotypes and environments, along with integration of additional orthogonal information (e.g. RNA) into genomic prediction framework can further enhance predicting performance of untested genotypes in diverse environments.

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to all those who have supported me throughout this journey. Firstly, I am deeply thankful to my major advisor, Dr. Nonoy Bandillo, for believing in me and providing me with the opportunity to pursue my PhD at North Dakota State University. Your mentorship, guidance, and unwavering support have been invaluable to me. I am also grateful to the rest of my graduate committee members, Dr. Qi Zhang, Dr. Megan Orr, and Dr. Julie Pasche, for their time, expertise, and valuable insights that have enriched my research.

I am especially indebted to Dr. Sikiru Atanda for his expertise, patience, and mentorship. His guidance has been instrumental in completing my analyses and growing as a researcher. I am truly grateful for his support through every stage of this journey, and most especially for believing in me.

I would like to extend my appreciation to the members of the Bandillo Lab, Princy Johnson, Shailesh Acharya, Mario Morales, and Harry Navasca, for their camaraderie, expertise, and assistance during fieldwork. Special thanks to Lisa Piche for her expertise, unwavering support and words of wisdom and encouragement, and to Hannah Worral for her assistance, especially in the field operations in Minot, and to the hourly staff of the Pulse Crops Breeding and Genetics Lab for their hard work and dedication.

I express my gratitude to the Department of Plant Sciences staff, for their invaluable assistance throughout my degree, especially to Shannon Ueker, Eileen Buringrud, and Karen Jevning.

I am also thankful to all my friends at NDSU, my classmates and officemates, for their companionship and support. Your friendship has made my time here truly memorable.

To Dr. Tonette Laude, my MS advisor, former supervisor and an esteemed alumna of NDSU, I extend my heartfelt gratitude for your invaluable advice, unwavering encouragement and support throughout my academic journey.

To my uncle, Dr. Glenn Gregorio, who encouraged me to pursue plant breeding and genetics, you have been a constant inspiration.

Lastly, I would like to thank my husband, parents, siblings, and relatives for their love, encouragement, and understanding throughout this journey. Your unwavering support has been my strength.

DEDICATION

I dedicate this dissertation to my husband, Rovel Austria, whose unwavering support, kindness, and understanding have been my pillars of strength. Your patience and encouragement have fueled my determination to succeed, and I am forever grateful for your love.

To my parents, Dr. Guia Saludares and Manuel-Popoveck Saludares, your endless support and guidance have been invaluable throughout my academic journey. Your words of encouragement, constant motivation, and heartfelt prayers have been a source of inspiration.

To my siblings, Reina Angeli, Ruth Althea, and Royale Albert, thank you for being my source of strength and for always believing in me. Your love and support have been my rock during challenging times, and I am blessed to have you in my life.

Above all, I dedicate this work to the Almighty God, who continues to make the impossible possible, for the gift of life, wisdom, knowledge, and for giving me the strength to persevere despite the challenges.

Lastly, to all who have been with me through the ups and downs of this journey, your presence and support have meant the world to me. Thank you for believing in me and for being a part of this incredible journey.

Maraming Salamat po sa inyong lahat!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDIX TABLES

LIST OF APPENDIX FIGURES

CHAPTER 1: LITERATURE REVIEW

1.1. Pea (*Pisum sativum*)

1.1.1. A historical overview

Field pea, also known as dry pea, is a member of the third largest flowering plant family, Fabaceae or Leguminosae with approximately 18,000 species and 800 genera (Foyer et al., 2016; Lewis, 2005). It is considered as one of the world's first domesticated crops and is cultivated in most temperate regions (Warkentin et al., 2015). Vavilov and Dorofeev (1992) identified the Mediterranean region as the center of origin for most cultivated vegetables, including peas, along with Central Asia, the Near East, and Ethiopia (Kalloo and Bergh, 2012). Peas were introduced to the Americas by European immigrants around 1500 AD, following early explorers. They quickly became one of the first seed crops cultivated in the United States. According to (Shoemaker and Delwiche, 1934), the initial pea-seed growing operations were established near Lake Ontario, situated in northeastern New York and Canada. Over time, the cultivation of field peas expanded, with significant production occurring in the Washington and Idaho Palouse region. Subsequently, North Dakota, South Dakota, Montana, and Minnesota joined the ranks of pea-producing states, beginning in the 1990s (Endres and Kandel, 2021). The period from 1949 to 1960s saw the formation of several growers' associations aimed at enhancing trade within the industry. Ultimately, in 1965, the USA Dry Pea & Lentil Council (USADPLC) was established to further these objectives which plays a crucial role in facilitating and funding research for improved varieties, developing new products, and analyzing the nutritional profiles of these commodities (USA Pulses, 2024).

1.1.2. Characteristics and growth patterns

Peas are diploid plants with a chromosome number of x= 7 and a genome size of 4.45 GB (Kreplak et al., 2019). Field pea is one of the two main classes of pea cultivars, the other being garden pea. Garden peas are typically harvested at a young, tender stage by shelling them from the pods. In contrast, field peas are harvested at the dry mature stage from pods that are not edible. The seeds of field peas are typically round, containing either yellow or green cotyledons, and their seed coats may be clear or pigmented (Endres and Kandel, 2021). Field pea varieties can be broadly categorized into two main types based on their leaf structure: the conventional leaf type, characterized by normal leaves and vine lengths that can reach up to 9 feet; and the semi-leafless type, which features leaves that have been modified into well-developed tendrils and usually has shorter vines, measuring about 2 ft (Miller et al., 2005; Uzun et al., 2005; Mikic et al., 2011, Tafesse et al., 2019; Endres and Kandel, 2021). Field peas can display two growth habits: indeterminate, where the terminal bud remains vegetative and continues growing as long as conditions permit; or semi-determinate, where vegetative growth continues after plant transitions to reproductive mode and begins flowering, but may stop before moisture becomes scarce, depending on the variety (Krall et al., 2006; Clark, 2019). They are classified as hypogeal plants, with cotyledons remaining below the ground and inside the seed coat during germination, providing protection against frost, wind erosion, and insect damage, as new stems can sprout from buds at or below ground level (Wulf and Reid, 2020; GRDC, 2018; Lamb and Podder, 2008).

1.1.3. Economic importance

Field pea production is primarily for human consumption or as livestock feed (Miller et al., 2005). Often referred to as the “poor man’s meat”, field pea is valued for its high protein

content, rich vitamins and mineral profile, and affordability, making it a popular choice in vegetarian diets, particularly among lower-income consumers (Amarakoon et al., 2012). Compared to wheat and other cereals, peas boast higher levels of protein, total dietary fiber, and total sugar content (Tulbek, 2014). This inherent nutritional richness highlights the importance of peas as a source of protein and dietary fiber, particularly for developing and underdeveloped countries. Dry peas exhibit a wide range of market classifications, with green and yellow cotyledon types being the primary ones. However, only selected varieties meet the standards for being sold in the green or yellow human edible market (Endres and Kandel, 2021).

1.2. Challenges in pea production

Peas are cultivated primarily in Canada, Russia, USA, France, and Australia (Tulbek et al., 2017). World production of dry peas has shown significant fluctuations in recent years culminating at 14.2 million metric tons in 2022, according to (FAOSTAT, 2024). Apart from their nitrogen-fixing ability, sustaining gains in grain yield is crucial for peas to remain an attractive option in crop rotations. However, expanding world pea production poses several challenges that require concerted efforts from pea breeders globally. Peas are highly vulnerable to various biotic stresses, including powdery mildew (León et al., 2020; Sulima & Zhukov, 2022; Rana et al., 2023), ascochyta blight (Bretag, 2006; Joshi et al., 2022; Tivoli & Banniza, 2007), rust (Osuna-Caballero et al., 2022; Chand et al., 2004), wilt (Kraft, 1994; Haglund and Kraft, 2001) and root rots (Chatterton et al., 2018; Wu et al., 2022). Additionally, abiotic factors such as heat (Devi et al., 2022; Mohapatra et al., 2020), frost (Shafiq et al., 2012), drought (Stagnari et al., 2016; Sorensen et al., 2003), salinity (Ehtaiwwesh and Emsahel, 2020; Popova et al., 2023; Duzdemir et al., 2009) and soil pH (Rice et al., 2000; Chaudhari et al., 2020) present major constraints on yield. Addressing these biotic and abiotic stresses is crucial for breeders aiming to

stabilize and increase grain yield (Warkentin et al., 2015). Breeders prioritize improving crop yield as the main trait for ensuring the long-term success of cultivars, along with disease resistance and other agronomic factors. Consequently, they primarily select for traits related to yield such as plant height, vegetative growth form, maturity, number of pods, and others. On the other hand, nutrition and quality are seen as secondary attributes, such as protein content, cooking quality, color, and taste, which impact the acceptability of varieties in the global marketplace (Tulbek et al., 2017).

1.3. Addressing pea production problems through breeding

Continued improvement for pea grain yield is challenged by limited available genetic resources. To overcome this, international germplasm exchange and the utilization of diverse Pisum accessions are crucial for achieving new yield gains (Warkentin et al., 2015). These genetic resources often possess unique traits that are essential sources of variation for enhancing germplasm. They provide a valuable reservoir of genetic diversity that can sustain long-term genetic gains in breeding programs (Bari et al., 2021).

Over the past four decades, molecular markers have played a significant role in revealing DNA-level polymorphisms, thereby enhancing breeding efforts, thus has revolutionized plant breeding. Marker-assisted selection (MAS) allows breeders to identify and select plants carrying specific genes or genomic regions associated with target traits. This technology has become increasingly important in recent years and has expedited the breeding process by reducing the time and resources required for traditional breeding phenotypic selection, especially in enhancing tolerance to both abiotic and biotic stressors (Barman and Kundu, 2019; Javid et al., 2015). The USDA Pea Single Plant Plus Collection (PSPPC) has been instrumental in pea breeding. Holdsworth et al. (2017) assembled this collection, which contains 431 *P. sativum* accessions

with a wide range of morphological, geographic and taxonomic diversity. Through genotyping-by-sequencing, 66,591 high-quality SNPs were generated, aiding in the identification of novel sources of favorable alleles. Notably, accessions from Central Asia exhibit diversity comparable to the sister species *P. fulvum* and subspecies, *P. sativum* subsp. *elatius*, indicating their potential for breeding programs.

The emergence of next-generation sequencing technologies and various genotyping platforms has significantly reduced the cost of genotyping, making it more affordable than phenotyping. This has opened up new avenues for plant breeding, particularly through the implementation of genomic selection (GS). GS allows breeders to predict an individual's breeding value using genome-wide genotypic data, enabling the selection of superior genotypes at an early stage (Belamkar et al., 2018; Desta and Ortiz, 2014; Jarquín et al., 2017; Pérez-Rodríguez et al., 2017; Poland, 2015; Poland and Rife, 2012). In GS, a training set or population with both phenotypic and genome-wide marker data is used to develop a prediction model through cross-validation (Meuwissen et al., 2001). This model is then utilized to calculate the genome-estimated breeding value (GEBV) of a breeding or testing population, which only requires genotypic data. Selection of the best-performing lines is based on these predicted GEBVs (Bhat et al., 2016). The adoption of GS in pea breeding has allowed breeders to accelerate the development of elite lines with enhanced traits that can adapt to changing environmental conditions. This approach is crucial for meeting the increasing demand for improved pea varieties (Annicchiarico et al., 2019; Budhlakoti et al., 2022; Cazzola et al., 2021; Gosal and Wani, 2020; Haile et al., 2020; Li et al., 2022; Pratap et al., 2022).

1.4. Genomic selection models

Genomic selection has emerged as a powerful tool in modern plant breeding, offering significant potential for accelerating crop improvement (Meuwissen et al., 2001). By reducing the cost per breeding cycle and shortening the generation interval, it enables more rapid genetic gain, leading to increased efficiency in the crop breeding process (Bhat et al., 2016; Kaler et al., 2022). In GS, a training set or population with both phenotypic and genome-wide marker data is used to develop a prediction model through cross-validation. This model is then utilized to calculate the genome-estimated breeding value (GEBV) of a breeding or testing population, which only requires genotypic data (Newell & Jannink, 2014). Selection of the best-performing lines is based on these predicted GEBVs (Bhat et al., 2016).

Univariate or single-trait (UNI) models have been widely employed in GS, focusing on predicting individual traits independently while assuming no correlation between traits (Atanda et al., 2022; Sandhu et al., 2022; Montesinos-López et al., 2022). Multi-trait GS (MT-GS) models integrate information from correlated traits and shared genetic information between lines to improve the accuracy (Jia and Jannink, 2012; Gill et al. 2021; Atanda et al., 2022; Montesinos-López et al., 2022;). As traits are genetically correlated, these MT-GS models have demonstrated their ability to enhance prediction accuracy, particularly for traits with inherently low heritability. Hayes et al. (2017) reported increased genomic prediction accuracy by ~40% for wheat end-use quality traits using a MT-GS model compared to a UNI-GS model. In barley, Bhatta et al. (2020) reported an increase of 57 to 61% prediction accuracy for agronomic and malting quality traits. In a recent study, Atanda et al. (2022) proposed a sparse-phenotyping-aided MT-GS model and demonstrated a notable improvement of over 12% in prediction accuracy across nutritional traits in field pea. Generally, prediction accuracy in MT-GS improves

as correlation between traits increases. However, in practice, the correlation between traits ranges from positive to negative, along with varying degrees of heritability. Addressing this challenge, Atanda et al. (2022) emphasized composition of traits in the training and prediction sets based on the heritability and genetic correlation between traits to enhance the prediction accuracy.

Economically important crop traits are often considered complex due to their polygenic nature and strong influence by environmental factors (Campbell et al., 2019; Mondal et al., 2023; Riedelsheimer et al., 2012; Shi et al., 2009). To effectively evaluate these traits, it is necessary to assess lines across multiple environments to account for genotype-by-environment (GxE) interactions. Multi-trait, multi-environment models in GS integrate the analysis of several traits evaluated across different environments. This approach, known as multi-trait multi-environment genomic prediction (MTME-GP), allows for the selection of pea lines that demonstrate consistent and robust performance across various traits and environments. MTME-GP has shown promising results in improving prediction accuracy and genetic gain for economically important traits. Studies have demonstrated that integrating GxE interactions in the MTME model further enhances prediction accuracy. For example, Sandhu et al. (2022) found that MT-based GS models outperformed UNI models for within-environment and across-location predictions of end-use quality traits in winter wheat. Gill et al. (2021) concluded that MT and MTME models offered significant advantages when considering environments and correlated traits. In their study on advanced breeding lines of winter wheat, the MTME model proved superior in predicting all agronomic traits. Overall, MTME-GP has emerged as a valuable tool for plant breeders, enabling them to effectively handle large breeding populations each season and address the challenges associated with complex traits and diverse environments.

1.5. Multi-omics prediction of crop traits

Technological advancements have significantly enhanced plant breeding practices, thereby playing a crucial role in addressing the challenges of feeding the rapidly expanding global population. Omics technologies, encompassing genomics, transcriptomics, and metabolomics, have revolutionized crop improvement strategies by providing detailed insights into molecular mechanisms underlying complex traits (Dai & Shen, 2022; Yang et al., 2021)

1.5.1. Genomics

The field of genomics has experienced significant advancements in recent years, driven by the emergence of next-generation sequencing and various genotyping platforms, which have made genotyping more accessible and cost-effective. These advancements have not only facilitated genome advancement but have also enhanced functional research (Shendure et al., 2017). Genomics has provided researchers with unprecedented insights into the genetic makeup and diversity of various plant species, revolutionizing plant breeding and crop improvement strategies (Roychowdhury et al., 2023).

One key approach in genomics is Quantitative trait locus (QTL) mapping, which aims to identify regions of the genome that contribute to variation in traits of interest (Ahmad et al., 2022). Numerous QTLs related to seed protein content and yield traits in pea have been identified through various QTL linkage analysis studies (Burstin et al., 2007; Gali et al., 2018; Klein et al., 2020; Ma et al., 2017; Timmerman-Vaughan et al., 1996; Ubayasena et al., 2011). However, traditional QTL mapping has limitations in mapping resolution due to the low recombination in mapping populations (Yan et al., 2017).

Genome-wide association studies (GWAS) have furthered our understanding of complex traits by associating genetic variants with phenotype variation across diverse germplasm (Korte

& Farlow, 2013). This approach, which relies on historic recombination and higher marker density (Gali et al., 2018), has been particularly effective in crops with diverse germplasm, such as maize and rice, leading to the discovery of novel genes underlying important agronomic traits. For example, (Anilkumar et al., 2022) identified three Meta-QTL for grain weight in rice while (Hu et al., 2022) identified five QTLs associated with relative spikelet fertility. Similarly, (Zhao et al., 2022) discovered three QTLs controlling rice grain length, and (Fei et al., 2022a; Fei et al., 2022b) identified numerous QTLs for maize yield traits.

GS has emerged as a transformative tool in plant breeding, utilizing genetic markers across the genome to predict the genetic merit of individuals (Meuwissen et al., 2001). While effective for many economically important traits (Beyene et al., 2015; Das et al., 2020; Huang et al., 2019; Rio et al., 2019; Rutkoski et al., 2012; Sarinelli et al., 2019; Yabe et al., 2018), it may not capture all the genetic variation underlying complex traits, due to complex biological processes that can influence phenotypes (Guo et al., 2016; Li et al., 2019). To address this limitation, researchers have turned to multi-omics prediction, which integrates data from multiple omics layers to provide a more comprehensive view of the genetic architecture of traits.

1.5.2. Transcriptomics

Transcriptomes, which capture the complete set of RNA transcripts in a cell and tissue, offer valuable insights into gene expression patterns that directly influence phenotypic traits (Zhou et al., 2022). Studies have shown the utility of gene expression data in predicting complex traits, such as hybrid maize yield performance, where a set of genes associated with hybrid performance led to higher prediction accuracy compared to using genetic markers alone (Fu et al., 2012). Li et al. (2019) integrated gene expression data with the whole-genome SNP data to predict various traits in *Drosophila melanogaster*, employing five different kernel-based

methods. Similarly, Azodi et al. (2020) demonstrated that transcriptome-based models outperformed baseline predictions derived from genetic marker data, suggesting that transcriptome data contributes valuable information to genomic prediction.

Despite these advancements, integrating omics data has presented challenges. Some studies have reported lower predictive abilities of models using transcriptome data compared to the benchmark model, GBLUP (Li et al., 2019; Xu et al., 2017). In contrast, others found that combining genetic markers and transcript data did not substantially improve performance over genetic marker data alone (Azodi et al., 2020). However, despite these challenges, the integration of omics data holds promise for advancing genomic prediction models and enhancing our understanding of complex trait inheritance.

In parallel, multi-environment trials in plant breeding provide critical insights into the adaptability and stability of breeding lines across diverse environmental conditions (Burgueño et al., 2012; Fehr, 1991; Mathew et al., 2018). While much of the focus in integrating multi-omics prediction has been on single-environment trials and single traits, a study by Hu et al. (2021) stands out for its evaluation of multi-omics multi-trait prediction models in multi-environment trials in oats. Their findings, demonstrating the superiority of these models over traditional approaches, highlight the potential of multi-omics integration in enhancing prediction accuracy and robustness across different environments.

These advancements underscore the importance of incorporating multi-omics approaches in plant breeding to develop cultivars that exhibit consistent and superior performance across a wide range of conditions. Such integration not only contributes to sustainable agriculture but also enhances food security by ensuring the resilience and productivity of crop varieties.

1.6. Literature cited

Ahmad, N., S. Ibrahim, Z. Tian, L. Kuang, X. Wang, H. Wang & X. Dun. (2022). Quantitative trait loci mapping reveals important genomic regions controlling root architecture and shoot biomass under nitrogen, phosphorus, and potassium stress in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* 13: 994666.

Amarakoon, D., D. Thavarajah, K. McPhee & P. Thavahrajah. (2012). Iron-, zinc-, and magnesium-rich field peas (*Pisum sativum* L.) with naturally low phytic acid: A potential food-based solution to global micronutrient malnutrition. *J. Food Compos. Anal.* 27(1):8–13.

Anilkumar, C., R. P. Sah, T.P. Muhammed Azharudheen, S. Behera, N. Singh, N.R. Prakash, N.C. Sunitha, B.N. Devanna, B.C. Marndi, B.C. Patra & S.K. Nair. (2022). Understanding complex genetic architecture of rice grain weight through QTL-meta analysis and candidate gene identification. *Sci. Rep.* 12(1):13832.

Annicchiarico, P., N. Nazzicari, L. Pecetti, M. Romani & L. Russi. (2019). Pea genomic selection for Italian environments. *BMC Genomics*. 20:603.

Atanda, S.A., J. Steffes, Y. Lan, M.A.A Bari, J. Kim, M. Morales, J.P. Johnson, R. Saludares, H. Worral, L. Piche, A. Ross, M. Grusak, C. Coyne, R. McGee, J. Rao & N. Bandillo. (2022). Multi-trait genomic prediction improves selection accuracy for enhancing seed mineral concentrations in pea. *The Plant Genome.* 15(4):e20260.

Azodi, C.B., J. Pardo, R. VanBuren, G. de Los Campos & S.H. Shiu. (2020). Transcriptome-based prediction of complex traits in maize. *Plant Cell.* 32(1):139–151.

Bari, M.A.A, P. Zheng, I. Viera, H. Worral, S. Szwiec, Y. Ma, D. Main, C.J. Coyne, R.J. McGee & N. Bandillo. (2021). Harnessing genetic diversity in the USDA pea germplasm collection through genomic prediction. *Front. Genet.* 12:707754.

Barman, M. & S. Kundu. (2019). Molecular markers and a new vista in plant breeding: a review: *Int. J. curr. microbial. Appl. sci.* 8(12):1921-1929.

Belamkar, V., M. J. Guttieri, W. Hussain, D. Jarquín, I. El-Basyoni, J. Poland, A.J. Lorenz & P.S. Baenziger (2018). Genomic selection in preliminary yield trials in a winter wheat breeding program. *G3*. 8(8): 2735–2747.

Beyene, Y., K. Semagn, S. Mugo, A. Tarekegne, R. Babu, B. Meisel, P. Sehabiague, D. Makumbi, C. Magorokosho, S. Oikeh, J. Gakunga, M. Vargas, M. Olsen, B. Prasanna, M. Banziger & J. Crossa (2015). Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Sci.* 55(1):154–163.

Bhat, J.A., S. Ali, R.K. Salgotra, Z.A. Mir, S. Dutta, V. Jadon, A. Tyagi, M. Mushtaq, N. Jain, P.K. Singh, G.P. Singh & K.V. Prabhu. (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front. Genet.* 7:221.

Bhatta, M., L. Gutierrez, L. Cammarota, F. Cardozo, S. Germán, B. Gómez-Guerrero, M.F. Pardo, V. Lanaro, M. Sayas & A.J. Castro. (2020). Multi-trait genomic prediction model increased the predictive ability for agronomic and malting quality traits in barley (*Hordeum vulgare* L.). *G3.* 10(3):1113–1124.

Bretag, T. W. (2006). The epidemiology and control of ascochyta blight in field peas: a review. *Aust. J. Res.* 57, 883-902. 10.1071/AR05222.

Budhlakoti, N., A.K. Kushwaha, A. Rai, K.K. Chaturvedi, A. Kumar, A.K. Pradhan, U. Kumar, R.R. Kumar, P. Juliana, D.C. Mishra & S. Kumar. (2022). Genomic selection: A tool for

accelerating the efficiency of molecular breeding for development of climate-resilient crops. *Front. Genet.* 13:832153.

Burgueño, J., G. de los Campos, K. Weigel & J. Crossa. (2012). Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52(2):707–719.

Burstin, J., P. Marget, M. Huart, A. Moessner, B. Mangin, C. Duchene, B. Desprez, N. Munier-Jolain & G. Duc. (2007). Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. *J. Plant Physiol.* 144(2):768–781.

Campbell, B.J., A.F. Berrada, C, Hudalla, S. Amaducci & J.K. McKay. (2019). Genotype × environment interactions of industrial hemp cultivars highlight diverse responses to environmental factors. *Agrosystems Geosci. Environ.* 2(1):1–11.

Cazzola, F., C.J. Bermejo, I. Gatti & E. Cointry. (2021). Speed breeding in pulses: an opportunity to improve the efficiency of breeding programs. *Crop Pasture Sci.* 72(3): 165–172.

Chand, R., C.P. Srivastava & C. Kushwaha. (2004). Screening technique for pea (*Pisum sativum* L.) genotypes against rust diseases (*Uromyces fabae*). *Indian J Agric Sci*. 74:166-167.

Chatterton, S., M. Harding, R. Bowness, D.L. Mclaren, S. Banniza & B.D. Gossen. (2018). Importance and causal agents of root rot on field pea and lentil on the Canadian prairies, 2014-2017. *Can. J. Plant Pathol.* https://doi.org/10.1080/07060661.2018.1547792.

Chaudhari, D., K. Rangappa, A. Das, J. Layek, S. Basavaraj, B.K. Kandpal, Y. Shouche, Y. & P. Rahi, P. (2020). Pea (*Pisum sativum* L.) plant shapes its rhizosphere microbiome for

nutrient uptake and stress amelioration in acidic soils of the north-east region of India. *Front. microbiol.* 11:538448.

Clark, S. (2019). Pea (*Pisum sativum* L.) characteristics for use and successful planting. United States Department of Agriculture. Natural Resources Conservation Service. Big Flats Plant Materials Center. Plant Materials Technical Note No. 19-01.

Dai, X. & L. Shen. (2022). Advances and trends in omics technology development. *Front. Med.* 9:911861.

Das, R.R., M.T. Vinayan, M.B. Patel, R.K. Phagna, S.B. Singh, J.P. Shahi, A. Sarma, N.S. Barua, R. Babu, K. Seetharam, J.A. Burgueño & P.H. Zaidi. (2020). Genetic gains with rapid-cycle genomic selection for combined drought and waterlogging tolerance in tropical maize (*Zea mays* L.). *The Plant Genome.* 13(3):e20035.

Desta, Z.A. & R. Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* 19(9):592–601.

Devi, J., V. Sagar, G.P. Mishra, P.K. Jha, N. Gupta, R.K. Dubey, P.M. Singh, T. Behera & P.V.V. Prasad. (2022). Heat stress tolerance in peas (*Pisum sativum* L.): Current status and way forward. *Front. Plant Sci.* 13:1108276.

Duzdemir, O., A. Kurunc & A. Unlukara. (2009). Response of pea (*Pisum sativum*) to salinity and irrigation water regime. *Bulg. J. Agric. Sci.* 15(5):400-409.

Ehtaiwwesh, A.F. & M.J. Emsahel. (2020). Impact of salinity stress on germination and growth of pea (*Pisum sativum* L) plants. *Al-Mukhtar Journal of Sciences*. 35(2):146–159.

Endres, G. & H. Kandel. (2021). Field pea production NDSU extension service. North Dakota State University, Fargo, ND. 1-11.

FAOSTAT (2024). Food and agricultural commodities production. Retrieved February 14, 2024 from: http://faostat.fao.org.

Fehr, W. R. (1991). Principles of cultivar development. Vol.1 Theory and Technique. Macmillan, New York.

Fei, J., J. Lu, Q. Jiang, Z. Liu, D. Yao, J. Qu, S. Liu, S. Guan & Y. Ma. (2022a). Maize plant architecture trait QTL mapping and candidate gene identification based on multiple environments and double populations. *BMC Plant Biol*. 22(1):110.

Fei, X., Y. Wang, Y. Zheng, X. Shen, X., E. Lizhu, J. Ding, J. Lai, W. Song & H. Zhao. (2022b). Identification of two new QTLs of maize (*Zea mays* L.) underlying kernel row number using the HNAU-NAM1 population. *BMC Genet*. 23(1): 593.

Foyer, C.H., H-M. Lam, H.T. Nguyen, K.H.M Siddique, R.K. Varshney, T.D. Colmer, W. Cowling, H. Bramley, T.A. Mori, J.M. Hodgson, J.W. Cooper, A.J. Miller, K. Kunert, J. Vorster, C. Cullis, J.A. Ozga, M.L. Wahlqvist, Y. Liang, H. Shou, K. Shi, J. Yu, N. Fodod, B.N. Kaiser, F-L Wong, B. Valliyodan & M.J. Considine. (2016). Neglecting legumes has compromised human health and sustainable food production. *Nat. Plants*. 2: 16112.

Fu, J., K.C. Falke, A. Thiemann, T.A. Schrag, A.E. Melchinger, S. Scholten, & M. Frisch. (2012). Partial least squares regression, support vector machine regression, and transcriptome-based distances for prediction of maize hybrid performance with gene expression data. *Theor. Appl. Genet*. 124(5): 825–833.

Gali, K.K., Y. Liu, A. Sindhu, M. Diapari, A.S.K. Shunmugam, G. Arganosa, K. Daba, C. Caron, R.V.B. Lachagari, B. Tar'an & T.D. Warkentin. (2018). Construction of high-

density linkage maps for mapping quantitative trait loci for multiple traits in field pea (*Pisum sativum* L.). *BMC Plant Biol*. 18(1): 172.

Gill, H.S., J. Halder, J. Zhang, N.K. Brar, T.S. Rai, C. Hall, A. Bernardo, P.S. Amand, G. Bai, E. Olson, S. Ali, B. Turnipseed & S.K. Sehgal. (2021). Multi-trait multi-environment genomic prediction of agronomic traits in advanced breeding lines of winter wheat. *Front. Plant Sci.* 12:709545.

Gosal, S.S & S.H. Wani. (2020). Accelerated Plant Breeding, Volume 3: Food Legumes. *Springer Nat*.

Grains Research & Development Corporation. (2018). Section 5 Plant growth and physiology. In: Field Pea. GRDC Grownotes.

Guo, Z., M.M. Magwire, C.J. Basten, Z. Xu & D. Wang. (2016). Evaluation of the utility of gene expression and metabolic information for genomic prediction in maize. *Theor. Appl. Genet*. 129(12): 2413-2427.

Haglund, W.A. & J.M. Kraft. (2001). Fusarium wilt. In Compendium of Pea Diseases and Pests, 2nd edn. St. Paul, MN: APS press.

Haile, T.A., T. Heidecker, D. Wright, S. Neupane, L. Ramsay, A. Vandenberg & K.E. Bett. (2020). Genomic selection for lentil breeding: Empirical evidence. *The Plant Genome.* 13(1): e20002.

Hayes, B. J., Panozzo, J., Walker, C. K., Choy, A. L., Kant, S., Wong, D., Tibbits, J., Daetwyler, H. D., Rochfort, S., Hayden, M. J., & Spangenberg, G. C. (2017). Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. *Theor. Appl. Genet*. 130(12): 2505-2519.

Holdsworth, W.L., E. Gazave, P. Cheng, J. Myers, M.A. Gore, C.J. Coyne, R.J. McGee & M. Mazourek. (2017). A community resource for exploring and utilizing genetic diversity in the USDA pea single plant plus collection. *Hort. Res*. 4: 17017. doi:10.1038/hortres.2017.17.

Hu, C., J. Jiang, Y. Li, S. Song, Y. Zou, C. Jing, Y. Zhang, D. Wang, Q. He & X. Dang. (2022). QTL mapping and identification of candidate genes using a genome-wide association study for heat tolerance at anthesis in rice (*Oryza sativa* L.). *Front. Genet.* 13: 983525.

Hu, H., M.T. Campbell, T.H. Yeats, X. Zheng, D.E. Runcie, G. Covarrubias-Pazaran, C. Broeckling, L. Yao, M. Caffe-Treml, L.A. Gutiérrez, K.P. Smith, J. Tanaka, O.A. Hoekenga, M.E. Sorrells, M.A. Gore & J-L. Jannink. (2021). Multi-omics prediction of oat agronomic and seed nutritional traits across environments and in distantly related populations. *Theor. Appl. Genet*. 134(12): 4043-4054.

Huang, M., E.G. Balimponya, E.M. Mgonja, L.K. McHale, A. Luzi-Kihupi, G-L.Wan & C.H. Sneller. (2019). Use of genomic selection in breeding rice (*Oryza sativa* L.) for resistance to rice blast (*Magnaporthe oryzae*). Molecular Breeding: New Strategies in Plant Improvement, 39(8). https://doi.org/10.1007/s11032-019-1023-2.

Jarquín, D., C. Lemes da Silva, R.C. Gaynor, J. Poland, A. Fritz, R. Howard, S. Battenfield & J. Crossa. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in Kansas wheat. *The Plant Genome.* 10(2). https://doi.org/10.3835/plantgenome2016.12.0130.

Javid, M., G.M. Rosewarne, S. Sudheesh, P. Kant, A. Leonforte, N. Lombardi, P.R. Kennedy, N.O.I. Cogan, A.T. Slater & S. Kaur. (2015). Validation of molecular markers associated

with boron tolerance, powdery mildew resistance and salinity tolerance in field peas. *Front. Plant. Sci*. 6: 917. doi: 10.3389/fpls.2015.00917.

Jia, Y. & J.L. Jannink. (2012). Multiple-trait genome selection methods increase genetic value prediction accuracy. *Genetics*. 192(4): 1513-1522. doi.org/10.1534/genetics.112.144246.

Joshi, S., B.R. Pandey & G. Rosewarne. (2022). Characterization of field pea (*Pisum sativum*) resistance against *Peyronellaea pinodes* and *Didymella pinodella* that cause Ascochyta blight. *Front. Plant Sci.*13: 976375.

Kaler, A.S., L.C. Purcell, T. Beissinger & J.D. Gillman. (2022). Genomic prediction models for traits differing in heritability for soybean, rice, and maize. *BMC Plant Biol*. 22(1), 87.

Kalloo, G. & O.B. Bergh. (2012). Genetic improvement of vegetable crops. *Elsevier Science*. 409-425.

Klein, A., Houtin, H., Rond-Coissieux, C., Naudet-Huart, M., Touratier, M., Marget, P., & Burstin, J. (2020). Meta-analysis of QTL reveals the genetic control of yield-related traits and seed protein content in pea. *Sci. Rep.* 10(1): 15925.

Korte, A. & A. Farlow. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*. 9: 29.

Kraft, J.M. (1994). Fusarium wilt of peas (a review). *Agron.* 14: 561-567.

Krall, J.M., S.D. Miller, J.T. Cecil, C. Bastian, T. Foulke, D.D. Baltensperger, B.M. Harveson, P.A. Burgener, G.W. Hergert, G.K. Hein, D.J. Lyon, T. Nleya, J. Rickersten & S. Blodgett. (2006). Pea production in the high plains. South Dakota State University Extension, FS932; University of Wyoming, B-1175; and University of Nebraska-Lincoln, EC187.

Kreplak, J., M.A. Madoui, P. Cápal, P. Novák, K. Labadie, G. Aubert, P.E. Bayer, K.K. Gali, R.A. Syme, D. Main, A. Klein, A. Bérard, I. Vrbová, C. Fournier, L. d'Agata, C. Belser, W. Berrabah, H. Toegelová, Z. Milec, J. Vrána, H. Lee, A. Kougbeadjo, M. Térézol, C. Huneau, C.J. Turo, N. Mohellibi, P. Neumann, M. Falque, K. Gallardo, R. McGee, B. Tar'an, A. Bendahmane, J.M. Aury, J. Batley, M.C. Le Paslier, N. Ellis, T. Warkentin, C.J. Coyne, J. Salse, D. Edwards, J. Lichtenzveig, J. Macas, J. Doležel, P. Wincker & J. Burstin. (2019). A reference genome for pea provides insight into legume genome evolution. *Nat. Genet*. 51(9), 1411–1422.

Lamb, J. & A. Podder. (2008). Grain legume handbook for the pulse industry. Grain Legume Hand Book Committee.

León, D.P., Ó.E. Checa & Obando, P.A. (2020). Inheritance of resistance of two pea lines to powdery mildew. *J. Agron.* 112(4)L 2466-2471.

Lewis, G.P. (2005). Legumes of the World. Royal Botanic Gardens Kew.

Li, Y., S. Kaur, L.W. Pembleton, H. Valipour-Kahrood, G.M. Rosewarne & H.D. Daetwyler, (2022). Strategies of preserving genetic diversity while maximizing genetic response from implementing genomic selection in pulse breeding programs. *Theor. Appl. Genet*. 135(6), 1813–1828.

Li, Z., N. Gao, J.W.R. Martini & H. Simianer. (2019). Integrating gene expression data into genomic prediction. *Front. Genet.* 10: 126.

Ma, Y., C.J. Coyne, M.A. Grusak, M. Mazourek, P. Cheng, D. Main & R.J. McGee. (2017). Genome-wide SNP identification, linkage map construction and QTL mapping for seed mineral concentrations and contents in pea (*Pisum sativum* L.). *BMC Plant Biol*. 17(1): 43.

Mathew, B., J. Léon & M.J. Sillanpää. (2018). Impact of residual covariance structures on genomic prediction ability in multi-environment trials. *PLoS One*. 13(7): e0201181.

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

Mikić A., V. Mihailović, B. Ćupina, V. Kosev, T. Warkentin, K. McPhee, M. Ambrose, J. Hofer and N. Ellis. (2011). Leaf types in pea genetic background and agronomic value of leaf types in pea (*Pisum sativum*). *Field Veg. Crop Res.* 48: 275-284.

Miller, P., K. McKay, C. Jones, S. Blodgett, F. Menalled, J. Riesselman, C. Chen, and D. Wichman. (2005). Growing dry pea in Montana. Montana State University Extension Service.

Mohapatra, C., R. Chand, J.K. Tiwari & A.K. Singh. (2020). Effect of heat stress during flowering and pod formation in pea (*Pisum sativum* L.). *Physiol. Mol. Biol. Plants.* 26(6): 1119-1125.

Mondal, R., A. Kumar & B.N. Gnanesh. (2023). Crop germplasm: Current challenges, physiological-molecular perspective, and advance strategies towards development of climate-resilient crops. *Heliyon.* 9(1): e12973.

Montesinos-López, O. A., A. Montesinos-López, D.A. Bernal Sandoval, B.A. Mosqueda-Gonzalez, M.A. Valenzo-Jiménez & J. Crossa. (2022). Multi-trait genome prediction of new environments with partial least squares. *Front. Genet.* 13: 966775.

Newell, M.A. & J-L. Jannink. (2014). Genomic selection in plant breeding. *Methods Mol Biol.* 1145: 117-130. doi:10.1007/978-1-4939-0446-4_10.

Osuna-Caballero, S., N. Rispail, E. Barilli & D. Rubiales. (2022). Identification and characterization of novel sources of resistance to rust caused by *Uromyces pisi* in *Pisum* spp. Plants, 11(17): 2268.

Pérez-Rodríguez, P., J. Crossa, J. Rutkoski, J. Poland, R. Singh, A. Legarra, E. Autrique, G. de los Campos, J. Burgueño & S. Dreisigacker. (2017). Single-step genomic and pedigree genotype × environment interaction models for predicting wheat lines in international environments. *The Plant Genome*. 10(2).

Poland, J. (2015). Breeding-assisted genomics. Current Opinion in Plant Biology, 24, 119–124.

Poland, J.A. & T.W. Rife. (2012). Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome*. 5(3). https://doi.org/10.3835/plantgenome2012.05.00005.

Popova, A.V., P. Borisova & D. Vasilev. (2023). Response of pea plants (*Pisum* sativum cv. Ran 1) to NaCl treatment in regard to membrane stability and photosynthetic activity. *Plants*. 12(2): 324.

Pratap, A., S. Kumar, P.L. Polowick, M.W. Blair & M. Baum. (2022). Editorial: Accelerating genetic gains in pulses. *Front. Plant Sci.* 13: 879377.

Rana, C., A. Sharma, R. Rathour, D. Bansuli, D.K. Banyal, R.S. Rana & P. Sharma. (2023). In vivo and in vitro validation of powdery mildew resistance in garden pea genotypes. *Sci. Rep.* 13(1):1-11.

Rice, W.A., W. Clayton, P.E. Olsen & N.Z. Lupwayi. (2000). Rhizobial inoculant formulations and soil pH influence field pea nodulation and nitrogen fixation. *Can. J. Soil Sci*. 80(3): 395-400. https://doi.org/10.4141/S99-059.

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, R. Sulpice, T. Altmann, M. Stitt, L. Willmitzer & A.E. Melchinger. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* 44(2): 217-220.

Rio, S., T. Mary-Huard, L. Moreau & A. Charcosset. (2019). Genomic selection efficiency and a priori estimation of accuracy in a structured dent maize panel. *Theor. Appl. Genet*. 132(1): 81-96.

Roychowdhury, R., S.P. Das, A. Gupta, P. Parihar, K. Chandrasekhar, U. Sarker, A. Kumar, D.P. Ramrao & C. Sudhakar. (2023). Multi-omics pipeline and omics-integration approach to decipher plant's abiotic stress tolerance responses. *Genes*, 14(6). doi: 10.3390/genes14061281.

Rutkoski, J., J. Benson, Y. Jia, G. Brown-Guedira, J-L. Jannink & M. Sorrells. (2012). Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *The Plant Genome*. 5(2): 51-61.

Sandhu, K.S., P.D. Mihalyov, M.J. Lewien, M.O. Pumphrey & A.H. Carter. (2021). Combining genomic and phenomic information for predicting grain protein content and grain yield in spring wheat. *Front. Plant Sci.* 12: 613300.

Sandhu, K.S., S.S. Patil, M. Aoun & A.H. Carter. (2022). Multi-trait multi-environment genomic prediction for end-use quality traits in winter wheat. *Front. Genet.* 13: 831020.

Sarinelli, J.M., J.P. Murphy, P. Tyagi, J.B. Holland, J. W. Johnson, M. Mergoum, R.E. Mason, A. Babar, S. Harrison, R. Sutton, C.A. Griffey & G. Brown-Guedira. (2019). Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theor. Appl. Genet.* 132(4): 1247-1261.

Shafiq, S., D.E. Mather, M. Ahmad & J.G. Paull. (2012). Variation in tolerance to radiant frost at reproductive stages in field pea germplasm. *Euphytica.* 186(3): 831-845.

Shendure, J., S. Balasubramanian, G.M. Church, W. Gilbert, J. Rogers, J.A. Schloss & R.H. Waterston. (2017). DNA sequencing at 40: past, present and future. *Nature*. 550: 345-353.

Shoemaker, D.N. & E.J. Delwiche. (1934). Descriptions of types of principal American varieties of garden peas.

Shi, J., R. Li, D. Qiu, C. Jiang, Y. Long, C. Morgan, I. Bancroft, J. Zhao & J. Meng. (2009). Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. *Genetics*. 182(3): 851-861.

Sorensen, J.N., M. Edelenbos, and L. Weinberg. (2003). Drought effects on green pea texture and related physical-chemical properties at comparable maturity. *J. Amer. Soc. Hort. Sci.* 128(1): 128-135.

Stagnari, F., A. Galieni & M. Pisante. (2016). Drought stress effects on crop quality. In Water Stress and Crop Plants (pp. 375–392). John Wiley & Sons, Ltd.

Sulima, A.S. & V.A. Zhukov. (2022). War and peas: Molecular bases of resistance to powdery mildew in pea (*Pisum sativum* L.) and Other Legumes. *Plants.* 11(3): 339.

Tafesse, E., T. Warkentin, & R. Bueckert. (2019). Canopy architecture and leaf type as traits of heat resistance in pea. *Field Crops Res.* 241: 107561.

Timmerman-Vaughan, G.M., J.A. McCallum, T.J. Frew, N.F. Weeden & A.C. Russell. (1996). Linkage mapping of quantitative trait loci controlling seed weight in pea (*Pisum sativum* L.). *Theor. Appl. Genet.* 93(3): 431-439.

Tivoli, B. & S. Banniza (2007). Comparison of the epidemiology of Ascochyta blights on grain legumes. *Eur. J. Plant Pathol.* 119(1): 59-76.

Tulbek, M.C., R.S.H. Lam, Y.C. Wang, P. Asavajaru & A. Lam. (2017). Chapter 9- Pea: A sustainable vegetable protein crop. In: Nadathur, S.R., J.P.D. Wanasundra, and L. Scanlin. (eds). Sustainable Protein Sources. *Academic Press*. 145-164. https://doi.org/10.1016/b978-0-12-802778-3.00009-3.

Tulbek, M.C. (2014). Pulse flours as functional ingredients. In: IUFOST annual conference, Montreal QC, Canada.

Ubayasena, L., K. Bett, B. Tar'an, & T. Warkentin. (2011). Genetic control and identification of QTLs associated with visual quality traits of field pea (*Pisum sativum* L.). *Genome.* 54(4), 261–272.

USA Pulses. (2024). USA Dry Pea & Lentil Council. Retrieved February 20, 2004 from: http://www.usapulses.org/membership/usadplc.

Uzun, A., U. Bilgili, M. Sincik, I. Filya & E. Acikgoz. (2005). Yield and quality of forage type pea lines of contrasting leaf types. *Europ. J. Agron.* 22: 85-94.

Vavilov, N.I & V.F. Dorofeev. (1992). Origin and geography of cultivated plants. Cambridge Univeristy Press.

Warkentin, T.D., P. Smykal, C.J. Coyne, N. Weeden, C. Domoney, D. Bing, A. Leonforte, Z. Xuxiao, P. Dixit, L. Boros, K.E. McPhee, R.J. McGee, & T.H.N. Ellis. (2015). Pea. In: De Ron, A.M. (ed). Handbook in Plant Breeding: Grain Legumes. *Springer*. 37-33.

Wu, L., R. Fredua-Agyeman, S.E. Strelkov, K-F. Chang & S-F. Hwang. (2022). Identification of quantitative trait loci associated with partial resistance to fusarium root rot and wilt caused by *Fusarium graminearum* in field pea. *Front. Plant Sci.* 12: 784593.

Wulf, K. & J. Reid. (2020). What drives interspecies graft union success? Exploring the role of phylogenetic relatedness and stem anatomy. *Physiol. Plant*. doi: 10.1111/ppl.13118.

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

Yabe, S., H. Yoshida, H. Kajiya-Kanegae, M. Yamasaki, H. Iwata, K. Ebana, T. Hayashi & H. Nakagawa. (2018). Description of grain weight distribution leading to genomic selection for grain-filling characteristics in rice. *PloS One.* 13(11): e0207627.

Yan, L., N. Hofmann, S. Li, M.E. Ferreira, B. Song, G. Jiang, S. Ren, C. Quigley, E. Fickus, P. Cregan & Q. Song. (2017). Identification of QTL with large effect on seed weight in a selective population of soybean with genome-wide association and fixation index analyses. *BMC Genomics*. 18(1): 529.

Yang, Y., M.A. Saand, L. Huang, W.B. Abdelaal, J. Zhang, Y. Wu, J. Li, M.H. Sirohi & F. Wang. (2021). Applications of multi-omics technologies for crop improvement. *Front. Plant Sci.* 12: 563953.

Zhao, M., Y. Wang, N. He, X. Pang, L. Wang, Z. Ma, Z. Tang, H. Gao, L. Zhang, L. Fu, C. Wang, J. Liu & W. Zheng. (2022). QTL Detection for Rice Grain Length and Fine Mapping of a Novel Locus qGL6.1. *Rice*. 15(1): 60.

Zhou, R., F. Jiang, L. Niu, X. Song, L. Yu, Y. Yang & Z. Wu. (2022). Increase crop resilience to heat stress using omic strategies. *Front. Plant Sci.* 13: 891861.

CHAPTER 2: MULTI-TRAIT MULTI-ENVIRONMENT GENOMIC PREDICTION OF PRELIMINARY YIELD TRIALS IN PULSE CROP[1]

2.1. Introduction

The challenges posed by a rapidly expanding global population and climate change underscore the imperative for sustainable food production (Tilman et al. 2011; van Dijk et al. 2021; Kumar et al. 2022). Field pea (*Pisum sativum*) emerges as a desirable crop, not only meeting the criteria for sustainability but also standing out as an affordable and nutritious plant-based protein source, placing field pea at the forefront of leguminous crops in the food industry (Punia & Kumar, 2022; Shanthakumar et al., 2022). However, the conventional process of developing a promising line for release to farmers involves rigorous phenotypic assessments across multiple seasons and environments, especially for polygenic traits with complex genetic architecture (Samantara et al., 2022). Accelerating the development of crop varieties to meet the needs of a growing population stands out as a viable strategy to help feed the world (Ahmar et al. 2020).

Genomic selection for complex traits in early breeding cycles has the potential to significantly reduce the selection cycle time and expedite genetic gain (Ertiro et al., 2015; Crossa et al., 2017; Bernardo, 2020). The advent of next-generation sequencing and various genotyping platforms has rendered genotyping more accessible and cost-effective than traditional phenotyping methods (Atanda et al., 2021). This transformative shift provides a unique

[1] This chapter has been published in a preprint server as *Saludares, R.A., S.A. Atanda, L. Piche, H. Worral, F. Dariva, K. McPhee & N. Bandillo. 2024. Multi-trait multi-environment genomic prediction of preliminary yield trials in pulse crops. doi: https://doi.org/10.1101/2024.02.18.580909*. It was also previously submitted for publication to an open-access journal and is currently under review. Rica Amor Saludares generated and analyzed data, conducted the research and investigation process, visualized data, and prepared the manuscript.

opportunity to seamlessly integrate genomic selection (GS), leveraging DNA information to predict the genetic merit of new genotypes (Meuwissen et al., 2001; Atanda et al., 2021). Studies have shown the potential of GS in pulse breeding programs for genetic improvement of seed yield, seed protein content, and wider adaptability to ever-changing environmental conditions (Annicchiarico et al., 2019; Budhlakoti et al., 2022; Cazzola et al., 2021; Gosal & Wani, 2020; Haile et al., 2020; Li et al., 2022; Pratap et al., 2022). The North Dakota State University (NDSU) pulse breeding program is undergoing a fundamental shift from phenotypically-driven approaches to a more modern GS-based approach at the preliminary yield trial (PYT) stage. Specifically, GS has great potential in early generation selection or culling in PYT using information from advanced trials. Improving accuracy in the early yield testing stage for selection of top-performing lines is essential for efficient resource allocation, shortening the breeding cycle, and, ultimately, increasing genetic gain (Bassi et al., 2016; Atanda et al., 2021; Bandillo et al., 2022).

Univariate or single-trait (UNI) models have been widely employed in GS, focusing on predicting individual traits independently while assuming no correlation between traits (Atanda et al., 2022; Sandhu et al., 2022; Montesinos-López et al., 2022a). Multi-trait GS (MT-GS) models integrate information from correlated traits and shared genetic information between lines to improve the accuracy (Jia and Jannink, 2012; Gill et al. 2021; Atanda et al., 2022; Montesinos-López et al., 2022b). As traits are genetically correlated, these MT-GS models have demonstrated their ability to enhance prediction accuracy, particularly for traits with inherently low heritability. Studies have also shown that the integration of genotype by environment (GxE) in the MT model further improves prediction accuracy.

In this chapter, we explored the merit of a multi-trait multi-environment enabled genomic prediction model (MTME-GP) in enhancing the prediction accuracy of two highly important, yet negatively correlated, traits: seed protein content and seed yield in field pea. Additionally, we further assessed the potential of MTME-GP models for predicting performance in single and cross-environment predictions using multiple years of data.

2.2. Materials and methods

2.2.1. Germplasm and phenotyping

The genetic materials consisted of 282 NDSU advanced elite breeding lines previously described in Bari et al. (2021). The lines were planted in 1.5- x 7.6-m plots at 0.30-m spacing between plots with 840 pure live seeds per plot, arranged in an augmented incomplete block design with five diagonal repeated checks for preliminary yield trials. Seed yield and agronomic data were collected in 3-year experiments from 2020 to 2022, including two environments at the NDSU North Central Research Extension Center (NCREC) near Minot, ND (MOT20 and MOT21) and one environment at the Carrington Research Extension Center near Carrington, ND (CAR22). Standard cultural practices were followed. Plots were harvested at physiological maturity (90-120 days after planting) and dried to 13% moisture content. A total of 0.11 kg clean and dried harvested seeds per line was used for protein analysis at the NCREC using near infrared (NIR) spectroscopy.

2.2.2. Genotyping

Young leaves were harvested from seedlings of each pea line planted in a greenhouse environment. DNA extraction was carried out using the DNeasy® Plant Mini Kit (Qiagen, Germantown, MD, USA) following the manufacturer's instructions, and elution was performed with 100µl. Subsequently, the DNA samples obtained were quantified using the Qubit dsDNA

BR Assay kit and Qubit 4.0 fluorometer (Life Technologies Corporation, Eugene, OR). As described by Bari et al. (2021), DNA samples were standardized to a final concentration of 25 ng/µl for subsequent genotyping-by-sequencing (GBS) at a genomic center. The prepared dual-indexed GBS libraries using the restriction enzyme ApeKI (Elshire et al. 2011) were combined into a single pool and sequenced across 1.5 lanes of NovaSeq S1x100-pb run, producing approximately 1,000 million pass filter reads with mean quality scores of > 30. The resulting quality reads were aligned to the established pea reference genome (Kreplak et al. 2019) yielding a total of 28,832 SNP markers. After removal of SNPs with minor allele frequency less than 1%, heterozygosity exceeding 20%, and those having over 90% missing values, the remaining 11,858 SNPs were used for downstream analysis. SNPs with missing values were imputed using Beagle v.5.1 (Browning et al., 2018).

2.2.3. Phenotypic data analysis

A mixed linear model was used to extract best linear unbiased estimates (BLUEs) for all traits evaluated using the following model:

$$\mathbf{y} = \mathrm{f}(\mathbf{r}, \mathbf{c}) + \mathbf{Xb} + \mathbf{Z}_\mathrm{r}\mathbf{u}_\mathbf{r} + \mathbf{Z}_\mathrm{c}\mathbf{u}_\mathrm{c} + \boldsymbol{\varepsilon} \qquad (2.1)$$

where **y** is the response variable for n-th phenotype, **b** is the fixed effect of the genotype, $\mathbf{u_r}$ and $\mathbf{u_c}$ are row and column random effects accounting for discontinuous field variation with multivariate normal distribution: $\mathbf{u_r} \sim \mathrm{N}(0, \mathbf{I}\sigma_\mathrm{r}^2)$ and $\mathbf{u_c} \sim \mathrm{N}(0, \mathbf{I}\sigma_\mathrm{c}^2)$ respectively, wherein, **I** is an identity matrix and σ_r^2 and σ_c^2 are variances due to row and column effect. f(**r, c**) is a smooth bivariate function defined over the row and column positions, **ε** is the measurement error from each plot with distribution of $\boldsymbol{\varepsilon} \sim \mathrm{N}(0, \mathbf{I}\sigma_\varepsilon^2)$, wherein, **I** is the same as above and σ_ε^2 is variance for the residual term or simply referred to as nugget. **X** and **Z** are incidence matrices for the fixed

and random terms, respectively. A total of 188 genotypes were found to overlap across three environments (Table A1).

2.2.4. Genomic selection models

The univariate (UNI) single environment GS model was fitted using the Bayesian approach and implemented in the BGLR R package (Pérez & de los Campos, 2014):

$$\mathbf{y} = \mathbf{1_k}\mu + \mathbf{Zu} + \boldsymbol{\varepsilon} \quad (2.2)$$

where $\mathbf{y}$ is the vector (n x 1) of adjusted means (BLUEs) for j-th pea lines for a targeted trait; μ is the overall mean; $\mathbf{1}_k$ $(k \times 1)$ is a vector of ones; $\mathbf{u}$ is the genomic effect of the j-th pea line and assumed to follow the multivariate normal distribution expressed as $\mathbf{u} \sim N(0, \mathbf{G}\sigma_g^2)$, where $\mathbf{G}$ is the genomic relationship matrix and σ_g^2 is the additive genetic variance; and $\mathbf{Z}$ is the incidence matrix for genomic effect of the lines.

The UNI multi-environment GS model was fitted using a reaction norm model which accounts for genotype by environment interaction (GxE) described in Jarquin et al. (2013):

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{Z}_3\mathbf{u}_3 + \boldsymbol{\varepsilon} \quad (2.3)$$

where $\mathbf{y}$ (n×1) is the vector of phenotypes of the pea lines measured in the environments (1...k), μ is the overall mean and $\mathbf{1}_n$ (nx1) is a of vector ones. $\mathbf{u}_1$is the random effect of the k-th environment and follows the multivariate normal distribution $N(0, \sigma_k^2\mathbf{Z}_k\mathbf{K}\mathbf{Z}_k')$ where σ_k^2 is the variance of the main effect of the environment, $\mathbf{K}$ is a relationship matrix between the environments which is an identity matrix, $\mathbf{Z}_k$is an incidence matrix that relates the phenotypes to the mean of the environments, and $\mathbf{Z}_k\mathbf{K}\mathbf{Z}_k'$ is a block diagonal matrix that uses a 1 for all pairs of observations in the same environment and a 0 for off-diagonal elements. $\mathbf{u_2}$is the random effect of the pea lines and follows the multivariate normal distribution $N(0, \sigma_g^2\mathbf{Z_g}\mathbf{G}\mathbf{Z_g}')$, where σ_g^2 is the variance of the main effect of the pea lines, $\mathbf{Z}_g$is an incidence matrix that relates the phenotypes

with the genomic relationship between the pea lines (**G**). $\mathbf{u}_3$is the random effect of the GxE effect and follows the multivariate normal distribution N(0, $\sigma^2_{gk}\mathbf{Z_g}\mathbf{G}\mathbf{Z}'_\mathbf{g}\#\sigma^2_k\mathbf{Z}_k\mathbf{K}\mathbf{Z}'_\mathbf{k}$), where σ^2_{gk} is the variance component of GE, # denotes the Hadamard product, and $\mathbf{Z_g}\mathbf{G}\mathbf{Z}'_\mathbf{g}$and $\mathbf{Z}_k\mathbf{K}\mathbf{Z}'_\mathbf{k}$ are the same as previously described. $\boldsymbol{\varepsilon}$ is the random term of the residual and follows the multivariate normal distribution N(0, $\sigma^2_\varepsilon\mathbf{I}$), where σ^2_ε is the homogenous residual variance. For the Bayesian Reproducing Kernel Hilbert Spaces Regressions (RHKS), the **G** matrix was replaced by kernel matrix (see Pérez & de los Campos, 2014 for details).

The multi-trait (MT) single environment GS model was fitted by extending Eq. 2.2 as follows:

$$\begin{bmatrix}\mathbf{y}_1\\ \vdots\\ \mathbf{y}_n\end{bmatrix} = \begin{bmatrix}\mathbf{1}_1\mu_1\\ \vdots\\ \mathbf{1}_k\mu_n\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \mathbf{Z}_n\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\\ \vdots\\ \mathbf{u}_n\end{bmatrix} + \begin{bmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\end{bmatrix} \quad (2.4)$$

where $\mathbf{y}_1 \ldots \mathbf{y}_n$ are the vector of phenotypes, $\mu_1 \ldots \mu_n$ are the overall mean for each n-th trait, $\mathbf{Z}_1 \ldots \mathbf{Z}_n$ is the incidence matrix for genomic effect of the lines for each n-th trait, $\mathbf{u}_1$ .... $\mathbf{u}_n$ is the genomic effect of the lines for each n-th trait, and $\boldsymbol{\varepsilon}_1$ ... $\boldsymbol{\varepsilon}_n$ is the residual error for each n-th trait. The random term is assumed to follow the multivariate normal distribution [$\mathbf{u}_1$ .... $\mathbf{u}_n$] $\sim$ MN[0, $(\mathbf{G} \otimes \mathbf{G}_o)$], where **G** is the same as above and $\mathbf{G}_o$ is an n x n unstructured variance-covariance matrix of the genetic effect of the traits, this is represented as follows:

$$\mathbf{G}_o \otimes \mathbf{G} = \begin{bmatrix}\sigma^2_{g_1} & \sigma_{g_{12}} & \cdots & \sigma_{g_{1n}}\\ \sigma_{g_{21}} & \sigma^2_{g_2} & \cdots & \cdots\\ \sigma_{g_{n1}} & \vdots & \ddots & \vdots\\ & \vdots & \ldots & \sigma^2_{g_n}\end{bmatrix} \otimes \mathbf{G} \quad (2.5)$$

The diagonal elements represent variance for each trait and covariances between traits are the off-diagonal elements.

Further, the residual term for each n-th trait is assumed to follow the multivariate normal distribution $[\boldsymbol{\varepsilon}_1 \ldots \boldsymbol{\varepsilon}_n] \sim \text{MN}[0, (\mathbf{I} \otimes \mathbf{R})]$, where $\mathbf{I}$ is the same as above and $\mathbf{R}$ is a heterogeneous diagonal matrix of the residual variances for each n-th trait:

$$\mathbf{R}=\begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\varepsilon_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2_{\varepsilon_n} \end{bmatrix} \otimes \mathbf{I} \tag{2.6}$$

The diagonal elements represent the residual variance for each n-th trait and off-diagonal elements of the $\mathbf{R}$ matrix equal zero.

For the multi-trait (MT) multi-environment GS model, Eq. 2.3 was expanded as described by Montesinos et al. (2022):

$$\mathbf{y} = \mathbf{1}_{\text{nK}}\mu + \mathbf{Z}_{1.1}\mathbf{u}_{1.1} + \mathbf{Z}_{2.1}\mathbf{u}_{2.1} + \mathbf{Z}_{3.1}\mathbf{u}_{3.1} + \boldsymbol{\varepsilon} \tag{2.7}$$

where $\mathbf{y}$ is of size i x n and i =j x k, n is the number of traits, j is the number of genotypes and k is the number of environments. $\mathbf{Z}_{1.1}$ is the incidence matrix of environment of size i x k, $\mathbf{u}_{1.1}$ is the random effect of each environment of each trait with size k x n, $\mathbf{Z}_{2.1}$ is the incidence matrix of genotypes of order i × j, $\mathbf{u}_{2.1}$ is the random effect of the genotypes i × n, and follows the multivariate normal distribution $\text{MN}(0, \sigma^2_g\mathbf{Z_g}\mathbf{G}\mathbf{Z}'_\mathbf{g}, \mathbf{U}_g)$, where $\mathbf{Z_g}$is an incidence matrix of the genotypes of order i x j. $\mathbf{G}$, $\mathbf{Z_g}\mathbf{G}\mathbf{Z}'_\mathbf{g}$and $\mathbf{Z}_k\mathbf{K}\mathbf{Z}'_\mathbf{k}$ are the same as above and $\mathbf{U}_g$is the unstructured variance-covariance matrix of traits of order n × n. $\mathbf{Z}_{3.1}$ is the incidence matrix of GE of order i × kj, $\mathbf{u}_{3.1}$ is the random effect of the genotypes by environment by trait of order kj × n and follows the matrix multivariate normal distribution $\text{MN}(0, \sigma^2_{gk}\mathbf{Z_g}\mathbf{G}\mathbf{Z}'_\mathbf{g} \# \sigma^2_k\mathbf{Z}_k\mathbf{K}\mathbf{Z}'_\mathbf{k}, \mathbf{U}_{gk})$, where $\mathbf{U}_{gk}$ is the unstructured variance-covariance matrix of order k by k. $\boldsymbol{\varepsilon}$ is the random term of the residual and follows the multivariate normal distribution $\text{MN}(0, \mathbf{I}, \Sigma_t)$. $\mathbf{I}$ is identity matrix of order i ×n, and Σ_t is the unstructured variance-covariance matrix.

2.2.5. Cross validation scheme

Evaluation of the predictive performance was assessed using various validation scenarios mean to mimic possible utilization scenarios of genomic selection in the NDSU field pea breeding program. Models were trained to predict seed yield and total seed protein content within and across different environments. Predictive ability (PA) was estimated as the Pearson correlation coefficient between predicted genomic estimated breeding value (GEBV) and BLUEs of each trait of the entire dataset.

In whole-environment predictions, we trained models using the complete dataset for each environment (MOT20, MOT21, and CAR22) to predict another entire environment (whole single-environment prediction). Alternatively, we trained models using the entire datasets of two environments to predict the entire third environment (whole cross-environment prediction).

For split-environment predictions, datasets for each environment were partitioned into different training set sizes (50%, 60%, 70%, and 80%). These subsets were used as training set, including the entire dataset of another environment, to predict the remaining datasets of the testing set (split single-environment prediction). This process was repeated 30 times. For split cross-environment predictions, datasets for each environment were divided into different training set sizes and used as the training set, including the whole datasets of the remaining environments, to predict the remaining dataset of the testing set. For example, 60% of the MOT20 datasets were used as a training set, including the entire MOT21 and CAR22 datasets to predict the remaining 40% of MOT20.

2.3. Results and discussion

2.3.1. Predictive ability of different genomic prediction models

We assessed the potential of genomic selection (GS) to predict the genetic merit of two negatively-correlated complex traits across three environments with varying heritability (Fig. 2.1). Notably, MOT21 exhibited extremely low heritability estimates for yield (1.56E-06) and

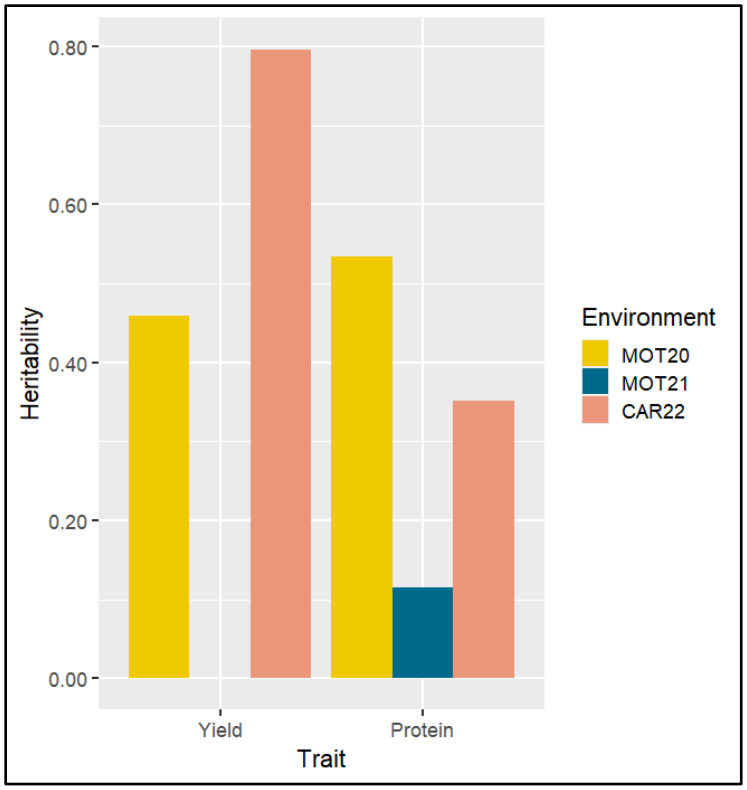

Figure 2.1. Heritability estimates for yield and protein under three environments, MOT20 is Minot 2020, MOT21 is Minot 2021, CAR22 is Carrington 2022.

protein (0.11), while CAR22 displayed the highest heritability for yield (0.80), and MOT20 had the highest for protein (0.53). The substantially lower heritability estimates observed under MOT21, a year characterized by drought, can be primarily attributed to significant environmental variation. This variation masked genetic effects, complicating the accurate estimation of the genetic component for these traits. Additionally, the presence of genotype-by-environment (GxE) interactions further challenged the precise estimation of genetic merit in this environment.

Here, we conducted whole single-environment prediction, considering either one trait at a time (univariate or UNI) or multiple traits simultaneously (multivariate or MT), and incorporating only the genetic factors (G) or also the interaction between the genotype and environment (GE). With the exception of UNI-GS, where G_BRR performed better than RKHS for both traits (Fig. 2.2 and 2.3), the RKHS model consistently outperformed other models across

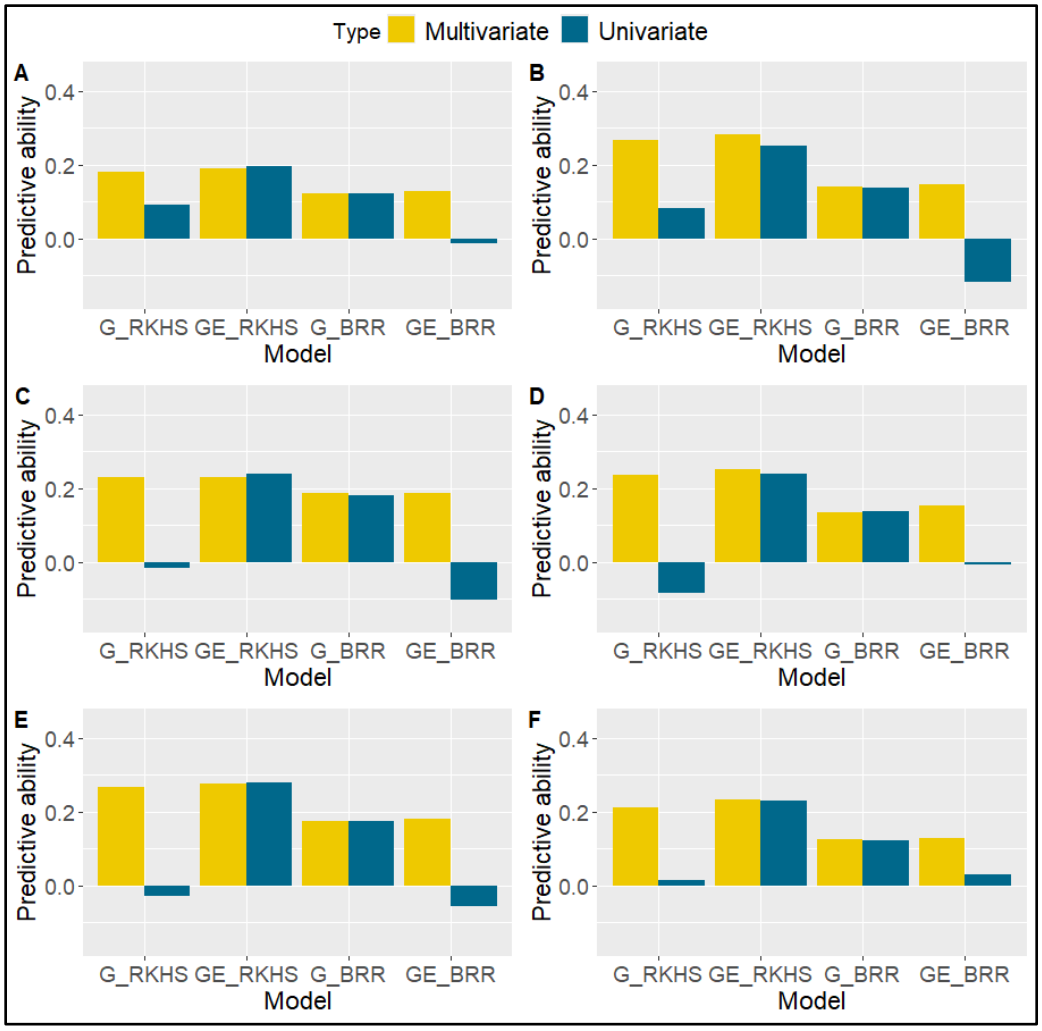

Figure 2.2. Predictive ability for seed yield using different genomic prediction models under single-environment prediction, BRR is Bayesian Ridge Regression model, RKHS is Reproducing Kernel Hilbert Spaces model, MT is multivariate, UNI is univariate, G is prediction model considering genotype, GE is prediction model integrating GxE interaction. (A) MOT21 dataset trained to predict MOT20, (B) CAR22 dataset trained to predict MOT20, (C) MOT20 dataset trained to predict MOT21, (D) CAR22 dataset trained to predict MOT21, (E) MOT20 dataset trained to predict CAR22, (F) MOT21 dataset trained to predict CAR22.

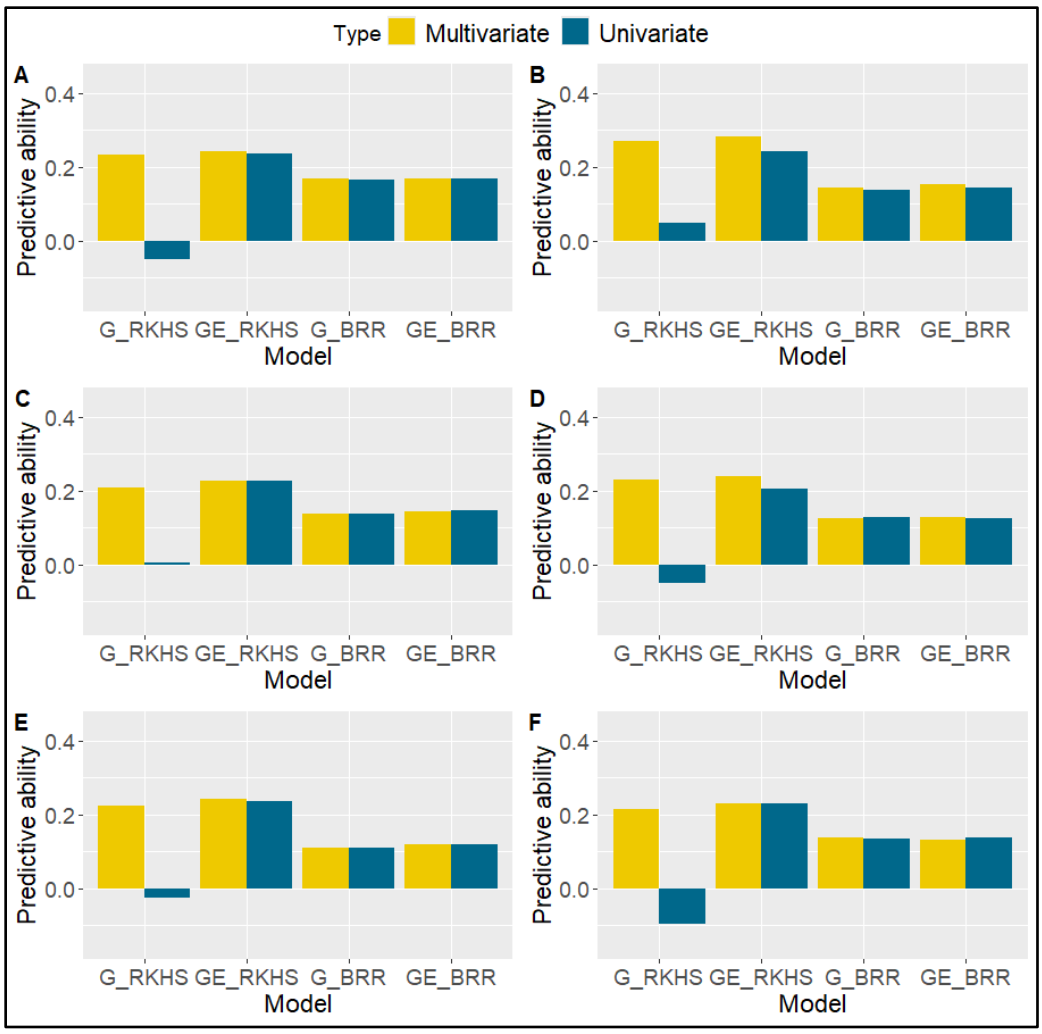

Figure 2.3. Predictive ability for seed protein content using different genomic prediction models under single-environment prediction, BRR is Bayesian Ridge Regression model, RKHS is Reproducing Kernel Hilbert Spaces model, MT is multivariate, UNI is univariate, G is prediction model considering genotype, GE is prediction model integrating GxE interaction. (A) MOT21 dataset trained to predict MOT20, (B) CAR22 dataset trained to predict MOT20, (C) MOT20 dataset trained to predict MOT21, (D) CAR22 dataset trained to predict MOT21, (E) MOT20 dataset trained to predict CAR22, (F) MOT21 dataset trained to predict CAR22.

scenarios. In the case of whole cross-environment prediction, similar trends were observed, with the RKHS model showing superior performance for both traits (Fig. 2.4 and 2.5). The superiority of the RKHS model in all other scenarios evaluated in this study suggests the robustness and reliability of the model in capturing not only additive effects but also non-linear effects and complex GxE interactions (Baertschi et al. 2021; Jiang and Reif 2015). These findings align with

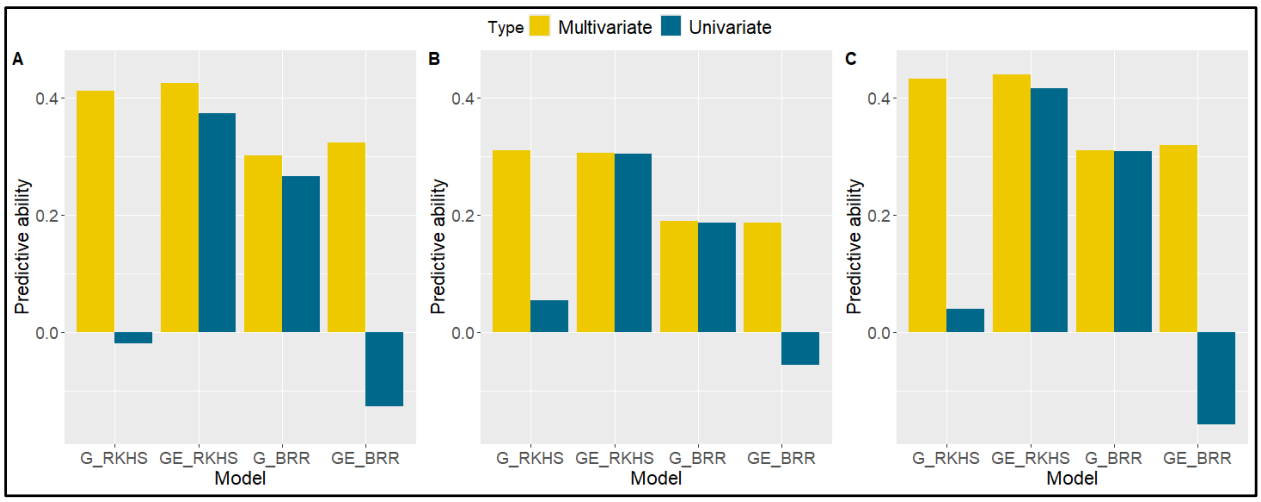

Figure 2.4. Predictive ability for seed yield using different genomic prediction models under cross-environment prediction, BRR is Bayesian Ridge Regression, RKHS is Reproducing Kernel Hilbert Spaces model, MT is multivariate, UNI is univariates, G is prediction model considering genotype, GE is prediction model integrating GxE interaction. (A) MOT21 and CAR22 datasets trained to predict MOT20, (B) MOT20 and CAR22 datasets trained to predict MOT21, (C) MOT20 and MOT21 datasets trained to predict CAR22.

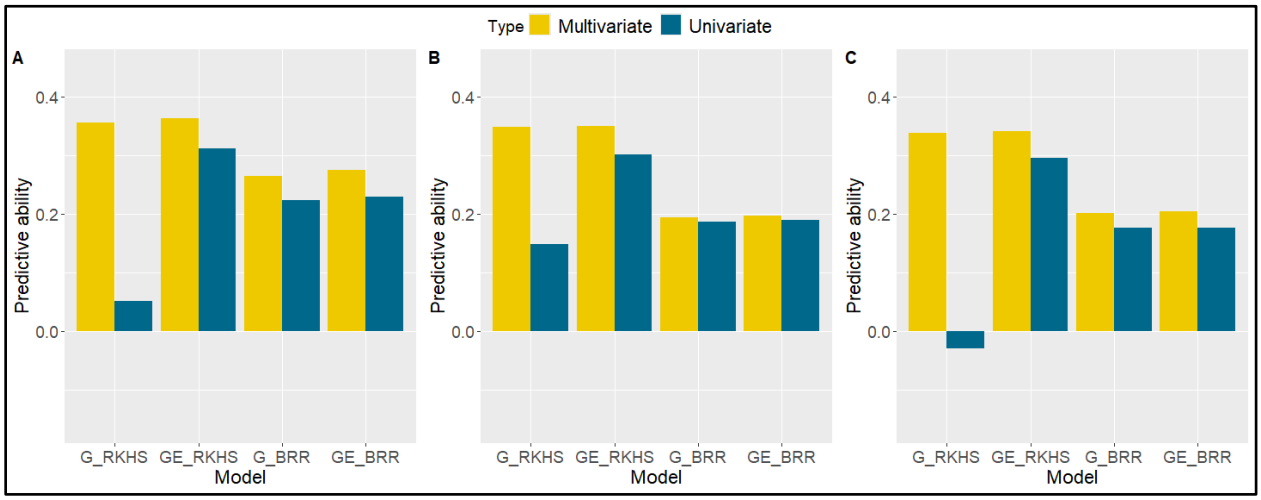

Figure 2.5. Predictive ability for seed protein content using different genomic prediction models under cross-environment prediction, BRR is Bayesian Ridge Regression, RKHS is Reproducing Kernel Hilbert Spaces model, MT is multivariate, UNI is univariates, G is prediction model considering genotype, GE is prediction model integrating GxE interaction. (A) MOT21 and CAR22 datasets trained to predict MOT20, (B) MOT20 and CAR22 datasets trained to predict MOT21, (C) MOT20 and MOT21 datasets trained to predict CAR22.

those of Bari et al. (2021), which observed subtle but favorable advantages of the RKHS model for predicting seed yield in field peas. To compare UNI with MT, we focused on the RKHS model due to its superiority over the BRR model across all validation scenarios.

Across all scenarios, including whole single and cross-environment predictions, MT consistently outperformed UNI under both G and GE_RKHS, with average predictive abilities improved by 1.9 to 2.4-fold for yield and 2-fold for protein. Okeke et al. (2017) also reported an improvement in predictive ability (average of 40%) with MT compared to UN for various traits in African cassava. Similarly, Arojju et al. (2020) reported improvements in prediction accuracy ranging from 24% to 59% for dry matter yield and 67% to 105% for nutritive quality traits in perennial ryegrass. Most recently, Winn et al. (2023) demonstrated substantial enhancement in prediction accuracy for various combinations of soft red winter wheat traits. The aforementioned highlight the potential of MT models to enhance prediction accuracy, especially for challenging and resource-intensive traits.

While integrating GxE interactions in the model using the MT approach did not lead to a significant improvement, showing only a 3 to 9% increase in results, this outcome could be due to the inclusion of environments with low heritability. These findings align with Rogers and Holland (2022), indicating that GxE interactions might be more relevant in environments with moderate to high heritability.

2.3.2. Optimal training set size for improved predictive performance of RKHS model

The training set size is one of the major factors influencing the prediction accuracy of un-tested lines (Norman et al. 2018). In split-environment prediction, we employed various training set sizes for training the RKHS model. Figure 2.6 illustrates how different training set sizes

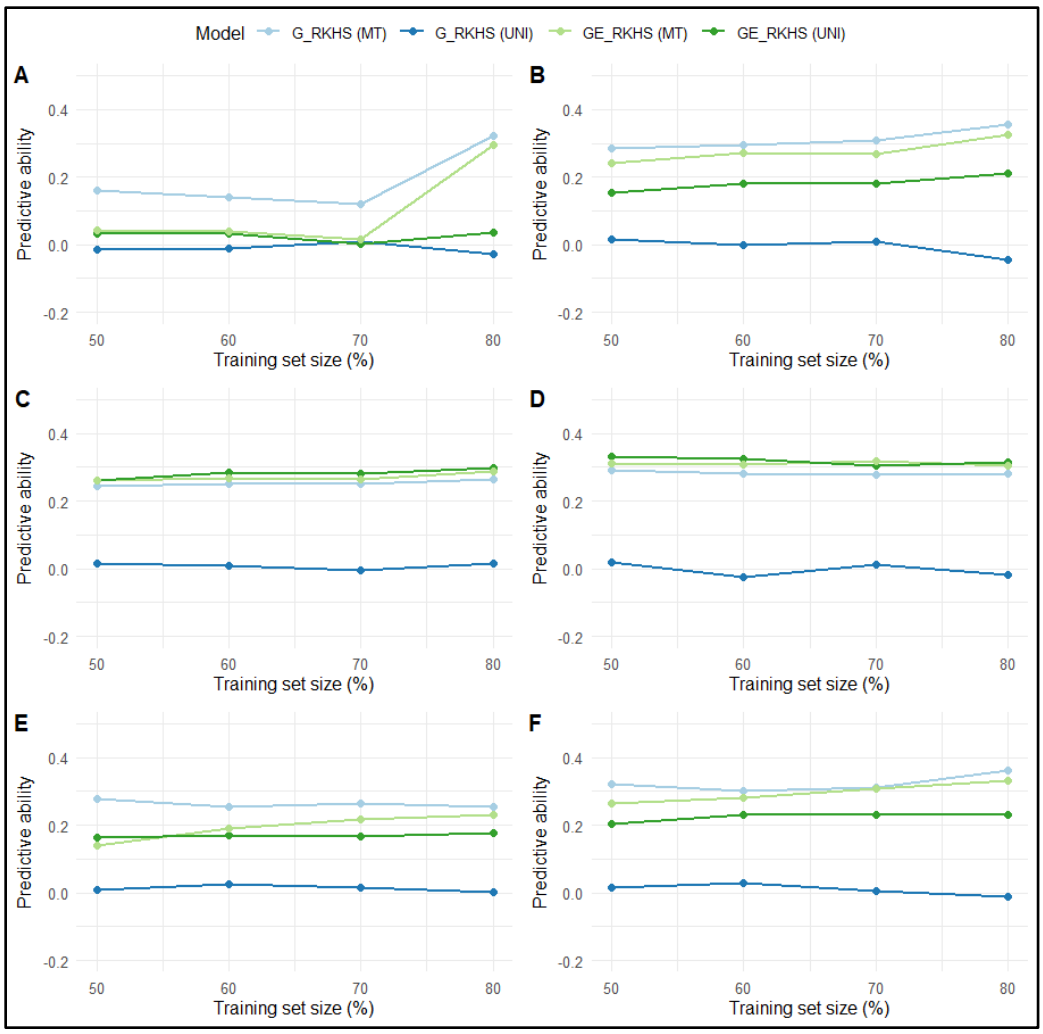

Figure 2.6. Average predictive ability with increasing training population size using RKHS models for seed yield, RKHS is Reproducing Kernel Hilbert Spaces, MT is multivariate, UNI is univariate, G is prediction model considering genotype, GE is prediction model integrating GxE interaction. (A) MOT21 dataset trained to predict MOT20, (B) CAR22 dataset trained to predict MOT20, (C) MOT20 dataset trained to predict MOT21, (D) CAR22 dataset trained to predict MOT21, (E) MOT20 dataset trained to predict CAR22, (F) MOT21 dataset trained to predict CAR22.

affect the model's predictive performance in split single-environment prediction for seed yield.

In predicting seed yield, the majority of the highest predictive abilities were observed under G_RKHS (MT), reaching 34% when 80% of the CAR22 dataset was trained to test MOT20 (Fig. 2.6B). On the other hand, the G_RKHS (MT) model consistently showed the highest predictive ability for seed protein (Fig. 2.7) reaching 30% when 60% and 80% of the MOT20 dataset were

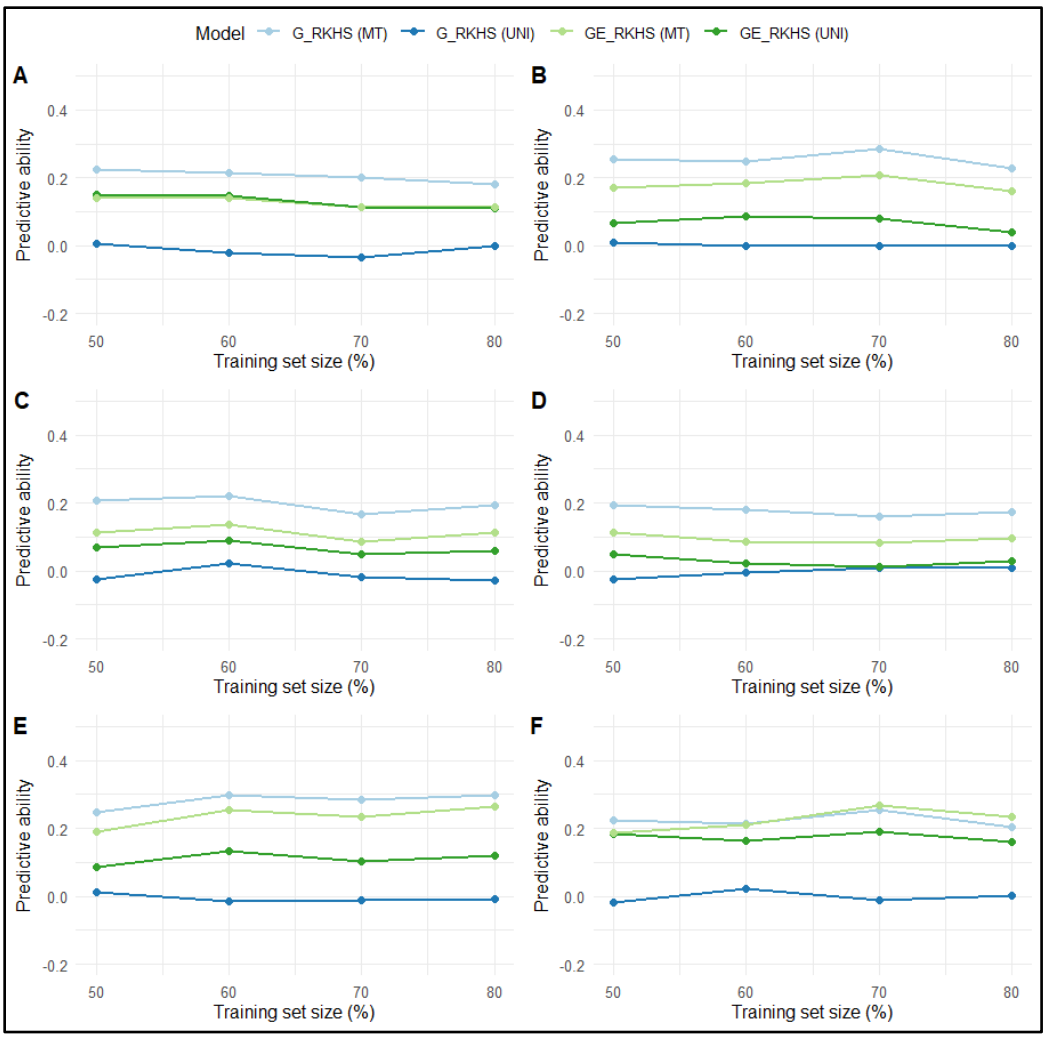

Figure 2.7. Average predictive ability with increasing training population size using RKHS models for seed protein content, RKHS is Reproducing Kernel Hilbert Spaces, MT is multivariate, UNI is univariate, G is prediction model considering genotype, GE is prediction model integrating GxE interaction. (A) MOT21 dataset trained to predict MOT20, (B) CAR22 dataset trained to predict MOT20, (C) MOT20 dataset trained to predict MOT21, (D) CAR22 dataset trained to predict MOT21, (E) MOT20 dataset trained to predict CAR22, (F) MOT21 dataset trained to predict CAR22.

trained to predict CAR22 (Fig. 2.7E). Previous studies have emphasized a strong relationship between prediction accuracy, training set size, and trait heritability (Luan et al., 2009; Lorenz et al., 2011; Clark et al., 2012; Nyline et al., 2017; Kaler et al.,2022; Atanda et al., 2022). Considering the varying heritability of the traits across environments, ranging from 1.57E-06 to 0.80 for grain yield and 0.12 to 0.53 for protein, and the negative correlation between traits, these

factors might contribute to the overall predictive ability across models in our study. Contrary to our results, Bari et al. (2021) reported an increase in prediction accuracy with increased training set size. Other studies (Budhlakoti et al. 2022; De Roos et al. 2009) have also reported the influence of training set size and heritability on prediction accuracy. This underscores the importance of careful consideration when selecting training set size for model training.

2.3.3. Efficacy of MTME-GP for predictions across different environments

Generally, we found that the mean predictive abilities of the RKHS model were higher under the training set size of 80%, particularly under the integration of GxE interaction in the model (GE_RKHS). Thus, this aspect was a focal point of our discussion in this section, aimed at comparing the efficacy of MTME-GP for both split single and cross-environment predictions. Our analysis revealed a clear trend showing improved predictive ability under cross-environment prediction, as depicted in Figure 2.8. However, an exception was noted in predicting the MOT20 yield (Fig. 2.8A). This discrepancy could be attributed to the significant influence of the very low heritability traits in other environments that were used to train the model for predicting MOT20 yield.

This observation underscores the importance of carefully managing testing environments to reduce the influence of environmental nuisance on phenotyping. Ultimately, it underscores the significance of considering heritability in the environment when developing training datasets for multi-environment GS models, ensuring efficient capturing of the genetic relationship between environments and borrowing information effectively across environments (Xu, 2016; van Eeuwijk et al. 2019; Atanda et al., 2021). Similarly, Sapkota et al. (2020) reported varying prediction accuracy when environments with different heritability were included in the training model to predict new environments. Additionally, Gill et al. (2021) emphasized the potential of

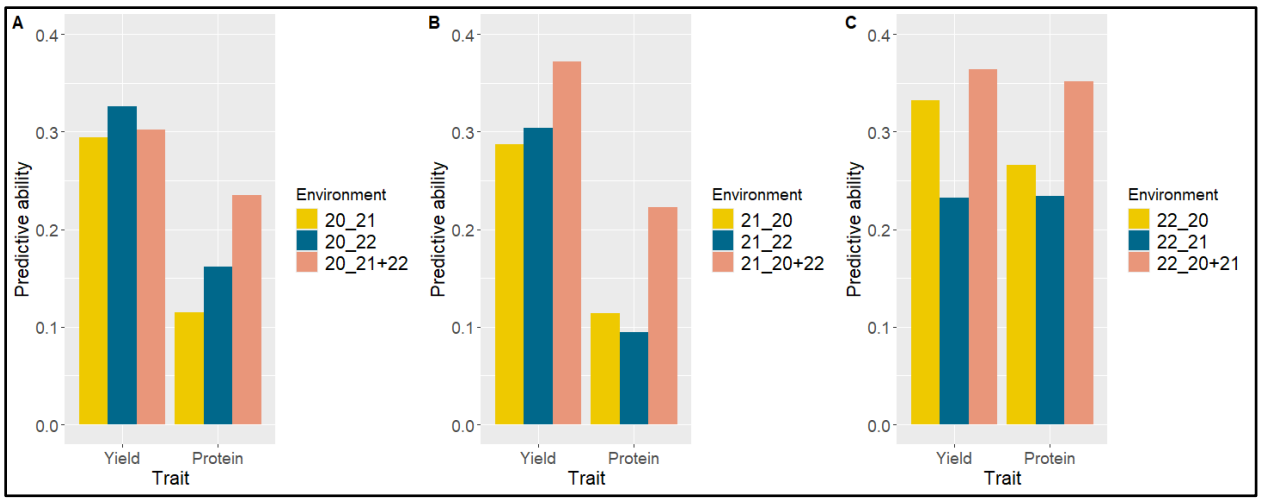

Figure 2.8. Average predictive abilities under 80% training set size using GE_RKHS model for seed yield and protein content, (A) MOT20 prediction utilizing MOT21 (20_21), CAR22 (20_22) and combination of MOT21 and CAR22 (20_21+22) as training sets; (B) MOT21 prediction utilizing MOT20 (21_20), CAR22 (21_22) and combination of MOT20 and CAR22 (21_20+22) as training sets; (C) CAR22 prediction utilizing MOT20 (22_20), MOT21 (22_21) and combination of MOT20 and MOT21 (22_20+21) as training sets.

MTME-GP in practical scenarios, such as overcoming the challenges posed by the loss of complete or partial trials due to extreme weather. MTME-GP proved valuable in predicting the genetic merit of the lines under MOT21, which were affected by drought conditions for both traits.

2.4. Conclusion

Our research findings in this chapter highlight the intricate dynamics of genomic prediction for seed yield and seed protein content in the face of diverse environmental conditions. The consistent superiority of the RKHS model, particularly in capturing GxE interactions, highlights its robustness and as a choice model in GS. Furthermore, the adoption of MTME-GP has proven instrumental in addressing the complexities associated in predicting inherently low heritability estimates of traits such as seed yield and total protein content. To fully harness the potential of genomic prediction in plant breeding, composition of the training set in terms of the individuals as well as the heritability of the environments for MTME-enabled GS

should be carefully considered. More so, including a wider array of correlated traits in prediction models, integrating deep learning for a more profound understanding of genetic architecture, and incorporating multi-omics data for a comprehensive view of trait genetics and molecular foundations all hold promise. This research marks a significant stride towards unlocking the potential of genomics in public plant breeding programs and offers valuable insights into the challenges and opportunities entailed by complex traits and diverse environments.

2.5. Literature cited

Ahmar, S., R.A. Gill, K. Jung, A. Faheem, M.U. Qasim, M. Mubeen & W. Zhou. (2020). Conventional and molecular techniques from simple breeding to speed breeding in crop plants: recent advances and future outlook. *Int J Mol Sci*. 21(7): 2590. doi: 10.3390/ijms21072590.

Annicchiarico, P., N. Nazzicari, L. Pecetti, M. Romani & L. Russi. (2019). Pea genomic selection for Italian environments. *BMC Genomics*. 20(1): 603.

Arojju, S.K., M. Cao, M. Trolove, B.A. Barrett, C. Inch, C. Eady, A. Stewart & M.J. Faville (2020). Multi-trait genomic prediction improves predictive ability for dry matter yield and water-soluble carbohydrates in perennial ryegrass. *Front. Plant Sci.* 11: 1197.

Atanda, S.A., M. Olsen, J. Burgueño, J. Crossa, D. Dzidzienyo, Y. Beyene, M. Gowda, K. Dreher, X. Zhang, B.M. Prasanna, P. Tangoona, E.Y. Danquah, G. Olaoye & K.R. Robbins. (2021). Maximizing efficiency of genomic selection in CIMMYT's tropical maize breeding program. *Theor. Appl. Genet.* 134: 279-294.

Atanda, S.A., J. Steffes, Y. Lan, M.A. Bari, J. Kim, M. Morales, J.P. Johnson, R. Saludares, H. Worral, L. Piche, A. Ross, M. Grusak, C. Coyne, R. McGee, J. Rao & N. Bandillo.

(2022). Multi-trait genomic prediction improves selection accuracy for enhancing seed mineral concentrations in pea. *The Plant Genome*. 15(4): e20260.

Atanda, S.A., M. Olsen, J. Crossa, J. Burgueño, R. Rincent, D. Dzidzienyo, Y. Beyene, M. Gowda, K. Dreher, P.M. Boddupalli, P. Tongoona, E.Y. Danquah, G. Olaoye & K.R. Robbins (2021). Scalable sparse testing genomic selection strategy for early yield testing stage. *Front. Plant Sci.* 12: 658978. doi: 10.3389/fpls.2021.658978.

Baertschi, C., T-V. Cao, J. Bartholome, Y. Ospina, C. Quintero, J. Frouin, J-M. Bouvet & C. Grenier. (2021). Impact of early genomic prediction for recurrent selection in an upland rice synthetic population. *G3.* 11(12). http://doi.10.1093/g3journal/jkab320.

Bandillo, N.B., D. Jarquin, L.G. Posadas, A.J. Lorenz & G.L. Graef. (2022). Genomic selection performs as effectively as phenotypic selection for increasing seed yield in soybean. *The Plant Genome*. 16(1): e20285.

Bari, M.A.A., P. Zheng, I. Viera, H. Worral, S. Szwiec, Y. Ma, D. Main, C.J. Coyne, R.J. McGee & N. Bandillo. (2021). Harnessing genetic diversity in the USDA pea germplasm collection through genomic prediction. *Front. Genet.* 12: 707754.

Bassi, F. M., A.R. Bentley, G. Charmet, R. Ortiz & J. Crossa. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Sci.* 242: 23-36.

Bernardo, R. (2020). Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. *Hered.* 125(6): 375-385.

Browning, B.L., Y. Zhou & S.R. Browning. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103(3): 338-348.

Budhlakoti, N., A.K. Kushwaha, A. Rai, K.K. Chaturvedi, A. Kumar, A.K. Pradhan, U. Kumar, R.R. Kumar, P. Juliana, D.C. Mishra & S. Kumar. (2022). Genomic selection: a tool for

accelerating the efficiency of molecular breeding for development of climate-resilient crops. *Front. Genet.* 13: 832153.

Cazzola, F., C.J. Bermejo, I. Gatti & E. Cointry (2021). Speed breeding in pulses: an opportunity to improve the efficiency of breeding programs. *Crop Pasture Sci.* 72(3): 165-172.

Clark, S.A., J.M. Hickey, H.D. Daetwyler & J.H. Van der Werf. (2012). The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestocks breeding schemes. *Genet Sel Evol.* 44(1):1-9.

Crossa, J., P. Perez-Rodriguez, J. Cuevas, O. Montesinos-Lopez, D. Jarquin, G. de los Campos, J. Burgueño, J.M. Gonzales-Camacho, S. Perez-Elizalde, Y. Beyene, S. Dreisigacker, R. Singh, X. Zhang, M. Gowda, M. Roorkiwal, J. Rutkoski & R.K. Varshney. Genomic selection in plant breeding: methods, models, and perspectives. (2017). *Trends Plant Sci.* 22(11): 961-975.

De Roos, A.P.W., B.J. Hayes & M.E. Goddard. (2009). Reliability of genomic predictions across multiple populations. *Genetics.* 183: 1545-1553. doi:10.1534/GENETICS.109.104935.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J. Poland, K. Kawamoto, E.S. Buckler & S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One. 5: e19349-10.* doi: 10.1371/journal.pone.0019379.

Ertiro, B.T., V. Ogugo, M. Worku, B. Das, M. Olsen, M. Labuschagne & K. Semagn, K. (2015). Comparison of Kompetitive Allele Specific PCR (KASP) and genotyping by sequencing (GBS) for quality control analysis in maize. *BMC Genomic.* 16(1): 1-12.

Gill, H.S., J. Halder, J. Zhang, N.K. Brar, T.S. Rai, C. Hall, A. Bernardo, P.S. Amand, G. Bai, E. Olson, S. Ali, B. Turnipseed & S.K. Sehgal. (2021). Multi-trait multi-environment

genomic prediction of agronomic traits in advanced breeding lines of winter wheat. *Front Plant Sci.* 12: 709545.

Gosal, S.S. & S.H. Wani. (2020). Accelerated plant breeding, Volume 3: Food Legumes. Springer Nature.

Haile, T. A., T. Heidecker, D. Wright, S. Neupane, L. Ramsay, A. Vandenberg & K.E. Bett. (2020). Genomic selection for lentil breeding: Empirical evidence. *The Plant Genome*. 13(1): e20002.

Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Pérez, M. Calus, J. Burgueño & G. de los Campos. (2013). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3), 595-607.

Jia, Y. & J-L. Jannink. (2012). Multiple-trait genome selection methods increase genetic value prediction accuracy. *Genetics.* 192: 1513-1522. https://doi.org/10.1534/genetics.112.144246.

Jiang, Y. & J.C. Reif. (2015). Modeling epistasis in genomic selection. *Genetics.* 201, 759-768. https://doi.10.1534/genetics.115.177907.

Kaler, A.S., L.C. Purcell, T. Beissinger & J.D. Gillman. (2022). Genomic prediction models for traits differing in heritability for soybean, rice, and maize. *BMC Plant Biol.* 22(1):87. doi. 10.1186/s12870-022-03479-y.

Kreplak, J., M.A. Madoui, P. Cápal, P. Novák, K. Labadie, G. Aubert, P.E. Bayer, K.K. Gali, R.A. Syme, D. Main, A. Klein, A. Bérard, I. Vrbová, C. Fournier, L. d'Agata, C. Belser, W. Berrabah, H. Toegelová, Z. Milec, J. Vrána, H. Lee, A. Kougbeadjo, M. Térézol, C. Huneau, C.J. Turo, N. Mohellibi, P. Neumann, M. Falque, K. Gallardo, R. McGee, B.

Tar'an, A. Bendahmane, J.M. Aury, J. Batley, M.C. Le Paslier, N. Ellis, T. Warkentin, C.J. Coyne, J. Salse, D. Edwards, J. Lichtenzveig, J. Macas, J. Doležel, P. Wincker & J. Burstin. (2019). A reference genome for pea provides insight into legume genome evolution. *Nat. Genet*. 51(9), 1411–1422.

Kumar, L., N. Chhogyel, T. Gopalakrishnan, M.K. Hasan, S.L. Jayasinghe, C.S. Kariyawasam, B.K. Kogo & S. Ratnayake. (2022). Chapter 4- Climate change and future of agri-food production. *Academic Press,* 49-79. https://doi.org/10.1016/B978-0-323-91001-9.00009-8.

Li, Y., S. Kaur, L.W. Pembleton, H. Valipour-Kahrood, G.M. Rosewarne & H.D. Daetwyler. (2022). Strategies of preserving genetic diversity while maximizing genetic response from implementing genomic selection in pulse breeding programs. *Theor. Apply. Genet.* 135(6): 1813-1828.

Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells & J-L. Jannink. (2011). Genomic selection in plant breeding: knowledge and prospects. *Adv Agron.* 110: 77-123.

Luan, T., J.A. Woolliams, S. Lien, M. Kent, M. Svendsen & T.H. Meuwissen. (2009). The accuracy of genomic selection in Norweigan red cattle assessed by cross-validation. *Genetics.* 183(3):1119-26.

Meuwissen, T.H., B.J. Hayes & M.E. Goddard. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 157(4): 1819-29. doi: 10.1093/genetics/157.4.1819.

Montesinos-López, O.A., A. Montesinos-López, D.A. Bernal Sandoval, B.A. Mosqueda-Gonzalez, M.A. Valenzo-Jiménez & J. Crossa. (2022a). Multi-trait genome prediction of new environments with partial least squares. *Front. Genet.* 13: 966775.

Montesinos-López, O. A., J.C. Montesinos-López, A. Montesinos-López, J.M. Ramírez-Alcaraz, J. Poland, R. Singh, S. Dreisigacker, L. Crespo, S. Mondal, V. Govidan, P. Juliana, J.H. Espino, J. H. S. Shrestha, R.K. Varshney & J. Crossa. (2022b). Bayesian multitrait kernel methods improve multienvironment genome-based prediction. *G3*. 12(2): jkab406.

Norman, A., J. Taylor, J. Edwards & H. Kuchel (2018). Optimizing genomic selection in wheat: effect of marker density, population size and population structure on prediction accuracy. *G3*. 8(9): 2889-2899. https://doi.org/10.1534/g3.118.200311.

Nyline, M., B. Uwimana, R. Swennen, M. Batte, A. Brown, P. Chistelova, E. Hribova, J. Lorenzen & J. Dolezel. (2017). Trait variation and genetic diversity in a banana genomic selection training population. *PLoS One.* 12(6): e0178734.

Okeke, U. G., D. Akdemir, I. Rabbi, P. Kulakow & J-L. Jannink. (2017). Accuracies of univariate and multivariate genomic prediction models in African cassava. *Genet. Sel.* 49(1): 1-10.

Pérez, P., & G. de los Campos. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics.* 198(2): 483-495.

Pratap, A., S. Kumar, P.L. Polowick, M.W. Blair & M. Baum. (2022). Editorial: Accelerating genetic gains in pulses. *Front. Plant Sci.*, 13: 879377.

Punia, S., & M. Kumar. (2022). Functionality and application of colored cereals: nutritional, bioactive, and health aspects. Elsevier.

Rogers, A. R., & Holland, J. B. (2022). Environment-specific genomic prediction ability in maize using environmental covariates depends on environmental similarity to training data. *G3.* 12(2). https://doi.org/10.1093/g3journal/jkab440.

Samantara, K., A. Bohra, S.R. Mohapatra, R. Prihatini, F. Asibe, L. Singh, V.P. Reyes, A. Tiwari, A.K. Maurya, J.S. Croser, S.H. Wani, K.H. Siddique & R. Varshney. (2022). Breeding more crops in less time: a perspective on speed breeding. *Biology (Basel).* 11(2): 275.

Sandhu, K.S., S.S. Patil, M. Aoun & A.H. Carter. (2022). Multi-trait multi-environment genomic prediction for end-use quality traits in winter wheat. *Front. Genet.* 13: 831020.

Sapkota, S., J. Lucas Boatwright, K. Jordan, R. Boyles, & S. Kresovich, S. (2020). Multi-trait regressor stacking increased genomic prediction accuracy of sorghum grain composition. *Agron.* 10: 1221. https://doi.org/10.3390/agronomy10091221.

Shanthakumar, P., J. Klepacka, A. Bains, P. Chawla, S.B. Dhull & A. Najda. (2022). The current situation of pea protein and its application in the food industry. *Mol.* 27(16). https://doi.org/10.3390/molecules27165354.

Tilman, D., C. Balzer, J. Hill & B.L. Befort. (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences of the United States of America. 108(50) 20260-20264. https://doi.org/10.1073/pnas.1116437108.

van Dijk, M., T. Morley, M.L.M. Rau & Y. Saghai. (2021). A meta-analysis of projected global food demand and population at risk of hunger for the period 2010-2050. *Nat Food.* 494-501. https://doi.org/10.1038/s43016-021-00322-9

van Eeuwijk, F.A., D. Bustos-Korts, E.J. Millet, M.P. Boer, W. Kruijer, A. Thompson, M. Malosetti, H. Itawa, R. Quiroz, C. Kuppe, O. Muller, K.N. Blazakis, K. Yu, F. Tardieu & S.C. Chapman. (2019). Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Sci.* 282: 23-39.

Winn, Z.J., D.L. Larkin, D. N. Lozada, N. DeWitt, G. Brown-Guedira & R.E. Mason (2023). Multivariate genomic selection models improve prediction accuracy of agronomic traits in soft red winter wheat. *Crop Sci.*, 63(4), 2115–2130.

Xu, Y. (2016). Envirotyping for deciphering environmental impacts on crop plants. *Theor. Appl. Genet.* 129(4), 653–-73.

CHAPTER 3: INTEGRATING MULTI-OMICS DATA INTO GENOMIC PREDICTION FRAMEWORK IN FIELD PEA

3.1. Introduction

The integration of gene expression data into the genomic prediction framework represents a promising avenue for enhancing accuracy of predicting complex traits (Azodi et al., 2020; Li et al., 2019; Mahmood et al., 2022). In field pea (*Pisum sativum* L.) breeding, genomic prediction has emerged as a valuable tool for improving breeding efficiency and accelerating genetic gain (Annicchiarico et al., 2019; Atanda et al., 2022; Bari et al., 2021; Burstin et al., 2015; Castro-Urrea et al., 2023; Zhao et al., 2022). Field pea, a significant pulse crop, esteemed for its high protein content and adaptability across diverse environments (Kindie et al., 2019; Powers & Thavarajah, 2019), faces the challenges in breeding for complex traits such as seed yield and protein content, owing to their polygenic nature and susceptibility to environmental influences (Campbell et al., 2019; Mondal et al., 2023; Riedelsheimer et al., 2012; Shi et al., 2009).

Traditional genomic prediction methods rely on genetic markers distributed throughout the genome to predict the genetic merit of individuals (Meuwissen et al., 2001). While effective for many crops, these methods may not fully capture all the genetic variation associated with complex traits (Li et al., 2019 Azodi et al., 2020). However, relying solely on mono-omics approach, may not provide sufficient insight into the complexity of plant responses under stress conditions or environment, as highlighted by (Roychowdhury et al., 2023).

In response to these limitations, advancements in high-throughput technologies have facilitated the exploration of omics data, adding a multi-omics layer to achieve a more comprehensive understanding of the genetic architecture of complex traits (Chen et al., 2023;

Yang et al., 2021). Gene expression data, offering insights into gene activity under varying conditions, holds promise for unraveling the genetic underpinnings of complex traits (Mahmood et al., 2022). Recent years have witnessed the widespread adoption of RNA-sequencing (RNA-Seq) as a means to study gene expression patterns, enabling the quantification of gene expression levels and identification of differentially expressed genes (Zhang et al., 2017; Roychowdhury et al., 2023).

Integrating gene expression data into genomic prediction models offers the potential to enhance prediction accuracy and deepen insights into the genetic mechanisms governing complex traits. While numerous studies have highlighted the advantages of multi-omics prediction over traditional genomic prediction, limited research has been done in the context of field pea.

In this chapter, we aimed to investigate the utility of integrating gene expression data into genomic prediction of complex traits in field pea. We hypothesize that integrating gene expression will enhance the accuracy of predicting complex traits and provide valuable insights into their genetic basis. To address this, we evaluated the performance of the integrated models in predicting complex traits such as seed yield and protein content, in comparison to traditional genomic prediction models.

3.2. Materials and methods

3.2.1. Germplasm and phenotyping

The genetic materials consisted of 300 USDA germplasm accessions previously described by Bari et al. (2021). The lines were planted following an augmented incomplete block design with four diagonal repeated checks and two replications. Seed yield and agronomic data were collected from the 2022 experiment, across two environments: North Dakota State

University (FAR22), and Washington State University (WSU22). Standard cultural practices were implemented, and plots were harvested at physiological maturity (90-120 days after planting) and dried to 13% moisture content. For protein analysis, 0.11 kg of clean and dried harvested seeds per line was used, employing near infrared (NIR) spectroscopy.

3.2.2. Genotyping

Young leaves were harvested from seedlings of each pea line planted in a greenhouse under controlled conditions of 65 to 70°F temperature and a 16-hour light cycle. DNA extraction was conducted using the DNeasy® Plant Mini Kit (Qiagen, Germantown, MD, USA) according to the manufacturer's protocol, and elution was performed with 100µl. Subsequently, the concentration of DNA samples was quantified using the Qubit dsDNA BR Assay kit and Qubit 4.0 fluorometer (Life Technologies Corporation, Eugene, OR), standardized to a final concentration of 25 ng/µl. Whole genome resequencing (WGR) at a depth of 10x was carried out at HudsonAlpha Genome Sequencing Center (Huntsville, AL, USA) using Illumina sequencing technology. The sequencing generated 5.9 terabytes of raw data, comprising 103 billion paired-end reads. Quality assessment of the reads was conducted using Fast QC, followed by trimming using Trimmomatic. The trimmed reads were aligned to the *Pisum sativum* Chinese reference genome (Yang et al., 2022) using bwa-mem2. PCR duplicates were identified and removed using Picard's 'MarkDuplicates' function and Samtools-1.10, respectively. Variant calling was performed using BCFtools, and filtering was applied using VCFtools, with parameters set to a minimum depth of 5, maximum missingness of 5%, and a minor allele frequency of 5%, resulting in the retrieval of 6,720,968 SNPs. Further filtering was conducted using Plink v.19 to remove SNPs with less than 10% missing values, resulting in a final set of 870,224 SNPs for downstream analysis.

3.2.3. Sample collection

An initial field experiment with 12 accessions was conducted to optimize the timing of pod sampling. Timepoints were determined starting from the reproductive growth stage of pea, specifically when the first flat pod was observed at one of more nodes (R3 stage, T_0). Subsequently, timepoints were added at 6-day intervals from the previous sampling until reaching 18 days (T_1, T_2, and T_3). Pod samples were collected from each timepoint per line for RNA extraction (Fig. 3.1). Upon conducting the initial expression analysis, T_1 (6 days after the first flat pod observed) was found to exhibit the highest number of expressed genes across all lines with varying maturity periods, thus identified as the optimal timepoint for pod sampling. During the actual 2022 field experiment, pods from three plants per plot of 300 accessions were tagged at T_0. Three pods per plot, one from each tagged plant, were harvested at T_1, which was 6 days after tagging (Fig. 3.2A). The harvested pods were placed in a 50 ml tube and immediately stored in dry ice until they could be transferred to -80^{o}C for subsequent RNA extraction (Fig. 3.2B).

3.2.4. RNA extraction and sequencing

RNA extraction was conducted using the *Quick*-RNA^{TM} Plant Miniprep (ZYMO Research, Orange, CA, USA) according to the manufacturer's protocol including proteinase K treatment, with elution performed using 100µl. RNA sample concentrations were quantified and quality assessed using the Qubit RNA BR and RNA IQ Assay kits with Qubit 4.0 fluorometer (Life Technologies Corporation, Eugene, OR, USA), and standardized to a final concentration of 100 ng/µl. 3' RNA-Seq Library creation, multiplexing, and Illumina NextSeq500 single-end sequencing were carried out by the Cornell Institute of Biotechnology.

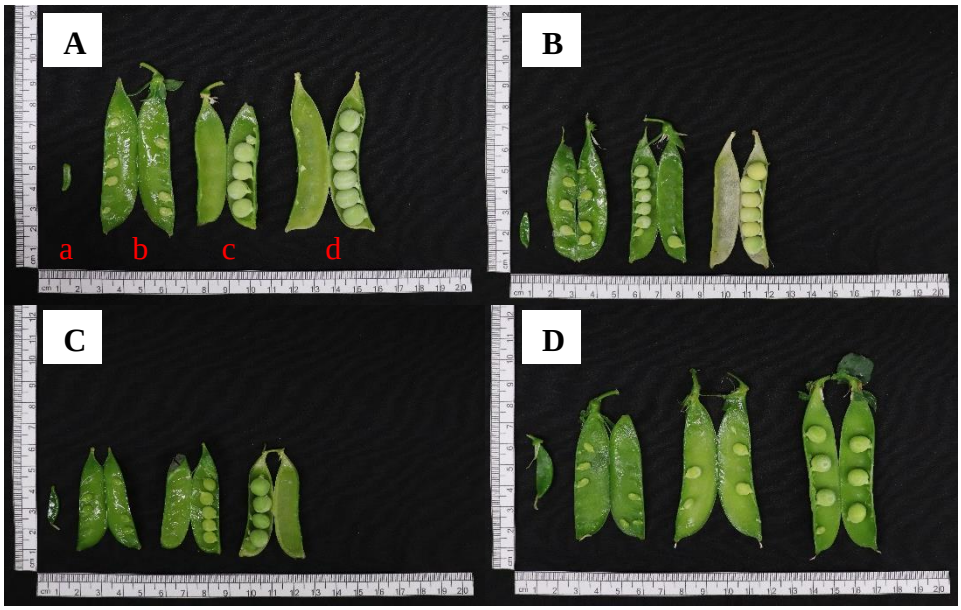

Figure 3.1. Longitudinal section of pods from four accessions with different maturity periods at different collection timepoints. (A) Agassiz, check variety, (B) PI 280617 PSP, early maturing, (C) PI 249645 PSP, mid maturing, (D) PI 340126, late maturing, (a) T_0, (b) T_1, (c) T_2, (d) T_3.

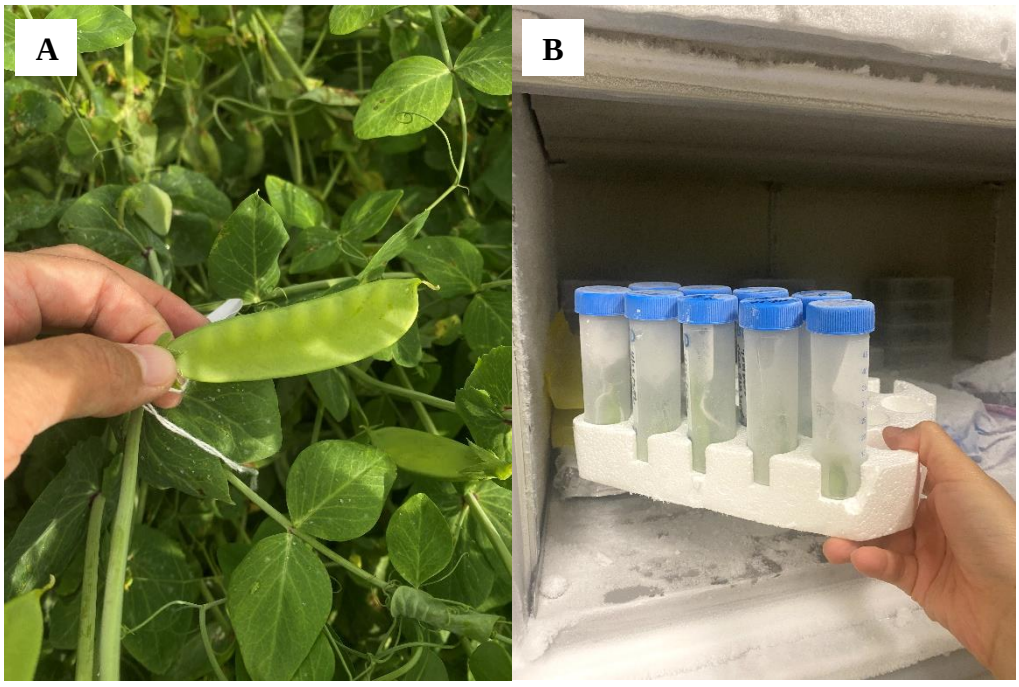

Figure 3.2. Pod collection and storage.

For samples with at least 200,000 reads, the first 12 bases corresponding to the random priming sites and any Illumina adapters were removed using Trimmomatic v.0.36 (Bolger et al., 2014). Subsequently, poly-A tails and poly-G stretches of at least 10 bases in length were removed using BBDuk program from the BBMap package (Bushnell, 2014), with reads kept at a minimum length of 18 bases after trimming. The trimmed reads were aligned to the Pea ZW6 genome assembly using STAR aligner v.2.7.10b (Dobin et al., 2013). For the STAR indexing step, the gff3 annotation file was converted to gtf formal using the gffread program from Cufflinks (Trapnell et al., 2010). The resulting SAM files were converted to BAM format using SAMtools v.1.15.1 (Danecek et al., 2021), and the number of reads overlapping each gene in the gff3 file on the forward strand were counted using HTSeq-count v.0.6.1 (Anders et al., 2015). The R package DeSeq2 v.1.36.0 (Love et al., 2014) was employed to obtain normalized and variance-stabilized counts.

3.2.5. Phenotypic data analysis

A mixed linear model was used to extract best linear unbiased estimates (BLUEs) for all traits evaluated using the following model:

$$\mathbf{y} = \mathrm{f}(\mathbf{r}, \mathbf{c}) + \mathbf{Xb} + \mathbf{Z}_r\mathbf{u}_r + \mathbf{Z}_c\mathbf{u}_c + \boldsymbol{\varepsilon} \quad (3.1)$$

where **y** is the response variable for n-th phenotype, **b** is the fixed effect of the genotype, $\mathbf{u_r}$ and $\mathbf{u_c}$ are row and column random effects accounting for discontinuous field variation with multivariate normal distribution: $\mathbf{u_r} \sim N(0, \mathbf{I}\sigma_r^2)$ and $\mathbf{u_c} \sim N(0, \mathbf{I}\sigma_c^2)$ respectively, wherein, **I** is an identity matrix and σ_r^2 and σ_c^2 are variances due to row and column effect. f(**r, c**) is a smooth bivariate function defined over the row and column positions, $\boldsymbol{\varepsilon}$ is the measurement error from each plot with distribution of $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma_\varepsilon^2)$, wherein, **I** is the same as above and σ_ε^2 is variance for the residual term or simply referred to as nugget. **X** and **Z** are incidence matrices for the fixed

and random terms, respectively. A total of 221 and 223 genotypes were included in FAR22 (Table A2) and WSU22 (Table A3), respectively.

3.2.6. Statistical models

We employed a Bayesian approach for predicting complex traits, specifically yield and protein content, in field pea using different sets of predictors. The BayesB model assumes that the phenotype y is a linear combination of the predictors, with a normally distributed error term. This model can be expressed as follows:

$$y_{ji} = \beta_0 + \sum_{j=1}^{p} X_{gj}\beta_j + \sum_{i=1}^{n} X_{ei}\alpha_i + \epsilon \quad (3.2)$$

where: y_{ij} is the phenotype of i-th sample, β_0 is the intercept treated as fixed effect, X_{gj} is the genotype indicator variable for SNP, β_j is the effect size of SNP j, X_{ei} is the expression level of gene i, α_i is the effect size of gene expression of gene i, n is the number of genotypes, p is the number of SNPs, and ϵ is the residual error term.

For genotypic data only (DNA):

$$y_j = \beta_0 + \sum_{j=1}^{p} X_{gj}\beta_j + \epsilon \quad (3.3)$$

For expression data only (RNA):

$$y_i = \beta_0 + \sum_{i=1}^{n} X_{ei}\alpha_i + \epsilon \quad (3.4)$$

When combining genotypic and expression data as predictors (DNA+RNA), the full model (Eq. 3.2) is used. The combination of these predictors aims to capture both the genetic and transcriptomic information that contribute to the phenotype.

The residual error term ϵ is assumed to be independent and identically distributed (*iid*) with normal distribution centered at zero with variance σ_ϵ^2. The conditional distribution of the data given effects and variance parameters is

$$P(y|\theta) = \prod_{i=1}^{n} \quad N(\mu_i, \sigma_\epsilon^2), \quad (3.5)$$

where $y = \{y_i\}$, θ represents the collection of model parameters $\theta = \{\beta_0, \beta, \sigma_\epsilon^2\}$, $N(\mu_i, \sigma_\epsilon^2)$ is a normal distribution centered at $\mu_i = \beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j$ and with variance σ_ϵ^2, and $\beta = \{\beta_j\}$ is a vector containing the effects of the individual spectra-derived wavelengths (Perez and de los Campos, 2014). The prior density assigned to the effects in Bayes B, $p(\beta_j|\Omega)$ is a mixture of a point of mass at zero and a scaled-*t* density, that is, $(\beta_j|\Omega)^{iid} \sim \pi \times t(\beta_j|df_\beta, S_\beta) + (1-\pi) \times 1(\beta_j = 0)$; therefore, a priori, with probability π, β_j is drawn from the *t*-density and with probability $(1-\pi)$ $\beta_j = 0$ (Ferragina et al., 2015).

We fitted the BayesB model using the BGLR R package (Pérez & de los Campos, 2014). The model was trained using 50% of the data in each environment and tested on the remaining 50%. To ensure robustness, we repeated the cross-validation procedure 30 times. For each trait (yield and protein content) and environment (FAR22 and WSU22), we evaluated the predictive ability of the model using Pearson correlation coefficient between the observed (BLUE) and predicted values (GEBV).

3.3. Results and discussion

In this chapter, we transition from a multi-trait multi-environment genomic prediction approach to a univariate, single-trait, single environment prediction using BayesB model for yield and protein content due to the limited availability of expression data from only two environments. The heritability estimates for yield and protein content exhibited substantial variability across environments, with WSU22 showing the lowest estimate (yield: 0.07, protein: 0.49) [Fig. 3.3]. This indicates a significant influence of environmental factors on yield performance in WSU22.

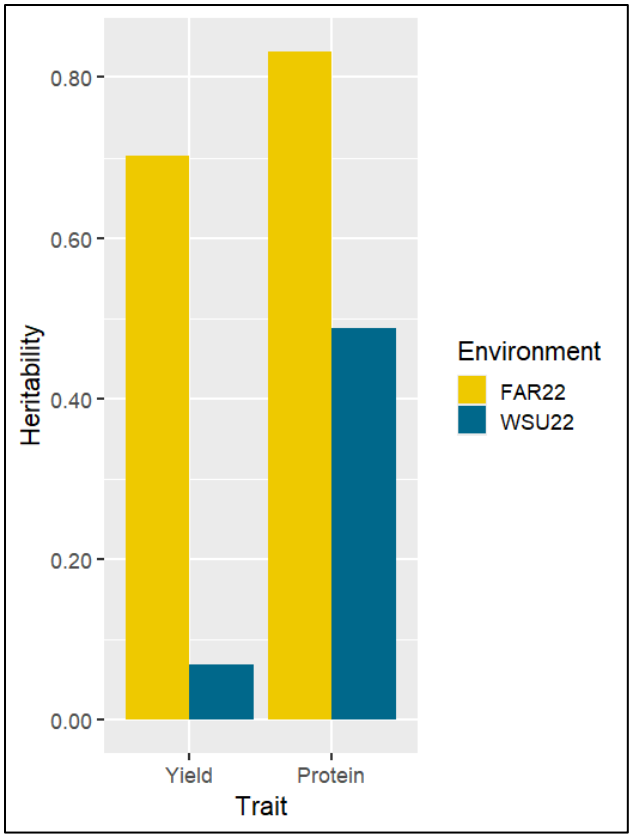

Figure 3.3. Heritability estimates for yield and protein under two environments, FAR22 is Fargo 2022, WSU22 is Washington State University 2022.

In terms of predictive ability, FAR22 demonstrated the highest mean predictive ability for yield across different data integration scenarios (Fig. 3.4). Interestingly, while the integration of multi-omics data (DNA+RNA) yielded the highest predictive ability for yield, the integration of expression data (RNA) alone in WSU22 resulted in the lowest mean predictive ability. These contrasting outcomes suggest a complex interplay between genetic architecture, environment, tissue specificity, influencing the predictive performance of the models.

In contrast, a clear trend emerged with multi-omics prediction for protein under FAR22. The model achieved a predictive ability of 0.42 with genomic data (DNA) alone, 0.53 with RNA

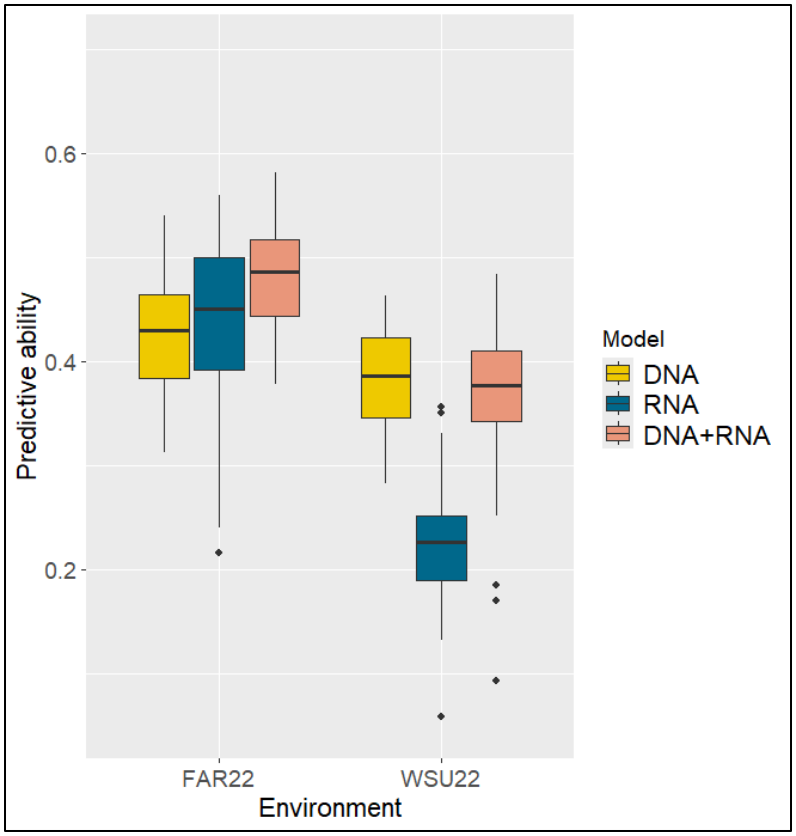

Figure 3.4. Mean distribution of predictive ability for yield across environments, DNA is genomic data, RNA is expression data, DNA+RNA is both DNA and RNA data integrated in prediction model.

alone, and 0.55 when both DNA and RNA were integrated (Fig. 3.5). The integration of expression data improved predictive ability by 26% compared to using genomic data alone.

However, only a 3% improvement was observed when multi-omics were integrated compared to using RNA alone. Conversely, in WSU22, the integration of DNA alone had the lowest mean predictive ability of 0.23, but this improved by 74% when RNA was used instead. Black (2000) highlighted that predictive ability improvement when using transcriptomic data can be attributed to alternative splicing, a mechanism wherein a single gene can produce multiple distinct transcripts, leading to increased protein diversity and potentially more phenotypic variation. Surprisingly, no improvement was observed with multi-omics integration, and instead, predictive ability declined by 17%. This unexpected outcome suggests that the integration of

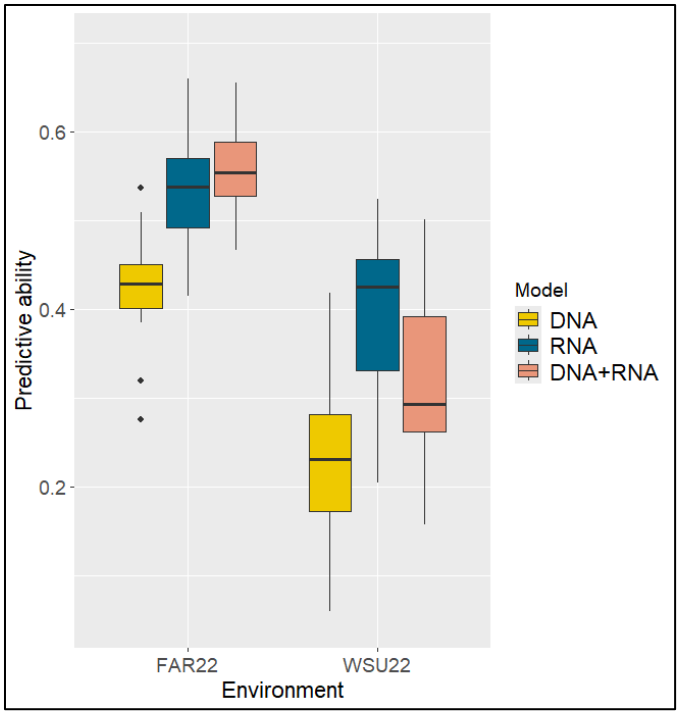

Figure 3.5. Mean distribution of predictive ability for protein content across environments, DNA is genomic data, RNA is expression data, DNA+RNA is both DNA and RNA data integrated in prediction model.

RNA data in this particular environment did not contribute positively to the predictive model, highlighting the complex and content-dependent nature of multi-omics data integration in prediction models.

The contrasting results observed between WSU22 and FAR22 can be partially explained by the heritability estimates for protein and yield in these environments. The moderate heritability estimated for protein in WSU22 likely contributed to the significant improvement in predictive ability when using RNA data alone. However, the very low heritability estimates for yield in WSU22 had a notable impact on the performance of multi-omics prediction, with no clear improvement over using genomic data alone. This observation aligns with findings by Wu et al. (2022), noting that regardless of the predictor, traits with higher heritabilities generally exhibit higher predictive abilities. Despite the limited improvement in protein prediction with

multi-omics data in FAR22, the modest 15% improvement suggests that the additional RNA data did contribute to the prediction models, albeit not as significant as expected. The complex interplay between genetic architecture, environment, and tissue specificity likely influenced these outcomes.

Hu et al. (2021) also found limited value in using transcripts alone to improve prediction accuracy, either by themselves or in combination with SNPs, particularly in single-environment prediction scenarios. This finding is consistent with other studies reporting either lower or comparable predictive abilities of transcripts compared to baseline GBLUP, with predictions being influenced by various factors (Guo et al., 2016; Westhues et al., 2017; Xu et al., 2017).

Several factors identified by Guo et al. (2016) may contribute to the limited utility of transcripts in prediction models. For instance, the tissue sampled may be limited to a single developmental time point, failing to capture dynamic changes occurring later, unsampled developmental stages. Additionally, Wu et al. (2022) observed that predictive ability was notably higher for datasets from seedlings compared to leaf datasets on average across traits. This discrepancy might be explained by the fact that more diverse genes are expressed in certain tissue types than in others. They also emphasized the importance of using predictors that are biologically closer to the phenotype of interest, suggesting that this approach may enhance the predictive ability in genomic predictions. Furthermore, it is possible that transcripts and SNPs capture similar genetic signals for the predicted traits, resulting in redundancy in information captured by these markers (Guo et al., 2016). Further research is needed to explore these factors and to develop more effective strategies for integrating multi-omics data to enhance prediction accuracy in plant breeding.

3.4. Conclusion

This chapter highlights the complexity and context-dependency of integrating gene expression data into genomic prediction models for complex traits in field pea. The contrasting results between WSU22 and FAR22 underscore the importance of considering the interplay between genetic architecture, environment, and tissue specificity when integrating multi-omics data. While the integration of gene expression data led to improvements in predictive ability for protein content in some environments, it did not always enhance predictive performance for yield. These findings suggest that the utility of transcriptomic data in prediction models may vary depending on the trait and environment under consideration.

Further research directions should focus on elucidating the factors influencing the effectiveness of integrating multi-omics data, such as the timing and specificity of tissue sampling, and the genetic architecture of the traits. Additionally, developing more sophisticated models that can better capture the complex interactions between genetic and environmental factors may further improve prediction accuracy. Overall, integrating gene expression data holds great promise for enhancing genomic prediction in field pea breeding, but further research is needed to fully realize its potential.

3.5. Literature cited

Anders, S., P.T. Pyl & W. Huber. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinform.* 31(2), 166-169.

Annicchiarico, P., N. Nazzicari, L. Pecetti, M. Romani & L. Russi, L. (2019). Pea genomic selection for Italian environments. *BMC Genomics.* 20(1): 603.

Atanda, S.A., J. Steffes, Y. Lan, M.A. Bari, J. Kim, M. Morales, J.P. Johnson, R. Saludares, H. Worral, L. Piche, A. Ross, M. Grusak, C. Coyne, R. McGee, J. Rao & N. Bandillo.

(2022). Multi-trait genomic prediction improves selection accuracy for enhancing seed mineral concentrations in pea. *The Plant Genome*. 15(4): e20260.

Azodi, C. B., Pardo, J., VanBuren, R., de Los Campos, G., & Shiu, S.-H. (2020). Transcriptome-Based Prediction of Complex Traits in Maize. The Plant Cell, 32(1), 139–151.

Bari, M.A.A., P. Zheng, I. Viera, H. Worral, S. Szwiec, Y. Ma, D. Main, C.J. Coyne, R.J. McGee & N. Bandillo. (2021). Harnessing genetic diversity in the USDA pea germplasm collection through genomic prediction. *Front. Genet.* 12: 707754.

Black, D. L. (2000). Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell.* 103(3): 367-370.

Bolger, A.M., M. Lohse & B. Usadel. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinform.* 30(15): 2114-2120.

Burstin, J., P. Salloignon, M. Chabert-Martinello, J-B. Magnin-Robert, M. Siol, F. Jacquin, A. Chauveau, C. Pont, G. Aubert, C. Delaitre, C. Truntzer & G. Duc. (2015). Genetic diversity and trait genomic prediction in a pea diversity panel. *BMC Genomics*. 16(1): 105.

Bushnell, B. (2014). BBmap: A fast, accurate, splice-aware aligner. United States. https://www.osti.gov/servlets/purl/1241166.

Campbell, B. J., Berrada, A. F., Hudalla, C., Amaducci, S., & McKay, J. K. (2019). Genotype × environment interactions of industrial hemp cultivars highlight diverse responses to environmental factors. *Agronsystems Geosci. Environ.* 2(1): 1-11.

Castro-Urrea, F.A., M.P. Urricariet, K.T. Stefanova, L. Li, W.M. Moss, A.L. Guzzomi, O. Sass, K.H.M. Siddique & W.A. Cowling. (2023). Accuracy of selection in early generations of

field pea breeding increases by exploiting the information contained in correlated traits. Plants: 12(5). https://doi.org/10.3390/plants12051141.

Chen, C., J. Wang, D. Pan, X. Wang, Y. Xu, J. Yan, L. Wang, X. Yang, M. Yang & G-P. Liu. (2023). Applications of multi-omics analysis in human diseases. *MedComm.* 4(4): e315.

Danecek, P., J.K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M.O. Pollard, A. Whitwham, T. Keane, S.A. McCarthy, R.M. Davies & H. Li. (2021). Twelve years of SAMtools and BCFtools. *Gigascience.* 10(2). https://doi.org/10.1093/gigascience/giab008.

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson & T.R. Gingeras. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinform.* 29(1), 15-21.

Ferragina, A., G. de los Campos, A.I. Vasquez & A. Cecchinato. 2015. Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. *J. Dairy Sci.* 98(11): 8133-8151.

Guo, Z., M.M. Magwire, C.J. Basten, Z. Xu & D. Wang. (2016). Evaluation of the utility of gene expression and metabolic information for genomic prediction in maize. *Theor. Appl. Genet.* 129(12): 2413-2427.

Hu, H., M.T. Campbell, T.H. Yeats, X. Zheng, D.E. Runcie, G. Covarrubias-Pazaran, C. Broeckling, L. Yao, M. Caffe-Treml, L.A. Gutiérrez, K.P. Smith, J. Tanaka, O.A. Hoekenga, M.E. Sorrells, M.A. Gore & J-L. Jannink. (2021). Multi-omics prediction of oat agronomic and seed nutritional traits across environments and in distantly related populations. *Theor. Appl. Genet.* 134(12): 4043-4054.

Kindie, Y., A. Bezabih, W. Beshir, Z. Nigusie, Z. Asemamaw, A. Adem, B. Tebabele, G. Kebede, T. Alemayehu & F. Assres. (2019). Field pea (*Pisum sativum* L.) variety

development for moisture deficit areas of Eastern Amhara, Ethiopia. *Adv Agric.* 2019: 1-6.

Li, Z., N. Gao, J.W.R. Martini & H. Simianer, H. (2019). Integrating gene expression data into genomic prediction. *Front. Genet.* 10: 126.

Love, M.I., W. Huber & S. Anders. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15(12): 550.

Mahmood, U., X. Li, Y. Fan, W. Chang, Y. Niu, J. Li, C. Qu & K. Lu. (2022). Multi-omics revolution to promote plant breeding efficiency. *Front. Plant Sci.* 13: 1062952.

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

Mondal, R., A. Kumar & B.N. Gnanesh. (2023). Crop germplasm: Current challenges, physiological-molecular perspective, and advance strategies towards development of climate-resilient crops. *Heliyon*. 9(1): e12973.

Pérez, P. & G. de los Campos. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genet.* 198(2): 483-495.

Powers, S.E. & D. Thavarajah. (2019). Checking agriculture's pulse: Field pea (*Pisum sativum* L.), sustainability, and phosphorus use efficiency. *Front. Plant Sci.* 10: 1489.

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, R. Sulpice, T. Altmann, M. Stitt, L. Willmitzer & A.E. Melchinger. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat Genet.* 44(2): 217-220.

Roychowdhury, R., S.P. Das, A. Gupta, P. Parihar, K. Chandrasekhar, U. Sarker, A. Kumar, D.P. Ramrao & C. Sudhakar. (2023). Multi-omics pipeline and omics-integration approach to

decipher plant's abiotic stress tolerance responses. *Genes*. 14(6). https://doi.org/10.3390/genes14061281.

Shi, J., Li, R., Qiu, D., Jiang, C., Long, Y., Morgan, C., Bancroft, I., Zhao, J., & Meng, J. (2009). Unraveling the complex trait of crop yield with quantitative trait loci mapping in Brassica napus. Genetics, 182(3), 851–861.

Trapnell, C., B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold & L. Pachter. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28(5): 511-515.

Westhues, M., T.A. Schrag, C. Heuer, G. Thaller, H.F. Utz, W. Schipprack, A. Thiemann, F. Seifert, A. Ehret, A. Schlereth, M. Stitt, Z. Nikoloski, L. Willmitzer, C.C. Schön, S. Scholten & A.E. Melchinger. (2017). Omics-based hybrid prediction in maize. *Theor. Appl. Genet.* 130(9): 1927-1939.

Wu, P-Y., B. Stich, M. Weisweiler, A. Shrestha, A. Erban, P. Westhoff & D. Van Inghelandt. (2022). Improvement of prediction ability by integrating multi-omic datasets in barley. *BMC Genomics*. 23(1): 200.

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

Yang, T., R. Liu, Y. Luo, S. Hu, D. Wang, C. Wang, M.K. Pandey, S. Ge, Q. Xu, N. Li, G. Li, Y. Huang, R.K. Saxena, Y. Ji, M. Li, X. Yan, Y. He, Y. Liu, X. Wang, C. Xiang, R.K. Varshney, H. Ding, S. Gao & X. Zong. (2022). Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. *Nat Genet.* 54(10): 1553–1563.

Yang, Y., M.A. Saand, L. Huang, W.B. Abdelaal, J. Zhang, Y. Wu, J. Li, M.H. Sirohi & F. Wang, F. (2021). Applications of multi-omics technologies for crop improvement. *Front. Plant Sci.* 12: 563953.

Zhang, X., X. Liu, D. Zhang, H. Tang, B. Sun, C. Li, L. Hao, C. Liu, Y. Li, Y. Shi, X. Xie, Y. Song, T. Wang & Y. Li. (2017). Genome-wide identification of gene expression in contrasting maize inbred lines under field drought conditions reveals the significance of transcription factors in drought tolerance. *PLoS one*. 12(7): e0179477.

Zhao, H., B.R. Pandey, M. Khansefid, H.V. Khahrood, S. Sudheesh, S. Joshi, S. Kant, S. Kaur, & G.M. Rosewarne. (2022). Combining NDVI and bacterial blight score to predict grain yield in field pea. *Front. Plant Sci.* 13: 923381.

CHAPTER 4: MULTI-TRAIT MULTI-ENVIRONMENT GENOMIC PREDICTION ACROSS DIVERSE PEA ACCESSIONS

4.1. Introduction

The field of genomics has undergone rapid expansion in recent years, driven by advancements in sequencing technologies and bioinformatics tools (Goodwin et al., 2016; Levy & Myers, 2016; Satam et al., 2023). This growth has led to a wealth of genomic data for numerous species, fundamentally altering our understanding of genetic diversity and makeup across the biological spectrum. Pea, as a model organism, has been extensively studied for its genetic diversity, often utilizing Simple Sequence Repeats (SSR) markers (Baranger et al., 2004; Kwon et al., 2012; Smýkal et al., 2008; Tar'an et al., 2005; Zong et al., 2009) or polymorphisms of insertion sites of PDR1 Ty1-copia group retrotransposons (RBIP) (Jing et al., 2010, 2012; Smýkal et al., 2008). However, the advent of next-generation sequencing, led to a rapid expansion of SNP discovery and genotyping array development (Deulvot et al., 2010; Duarte et al., 2014; Leonforte et al., 2013; Sindhu et al., 2014). SNP markers, with their abundance, wide distribution in genomes, and bi-allelic nature, have become the marker of choice for population genetics approaches, enabling genetic mapping and diversity assessment in various living organisms (Burstin et al., 2015).

The increasing availability of high-throughput genetic markers has not only transformed our understanding of genetic diversity but has also revolutionized crop genetic improvement methodologies. Traditional approaches, reliant on the phenotypic evaluation of related individuals and the calculation of their breeding value, have been supplemented by genomic prediction (GP). GP is a transformative tool in plant breeding, allowing for the selection of superior individuals based on their genetic profiles (Meuwissen et al., 2001). This shift toward

genomic-based selection has the potential to greatly enhance the efficiency and effectiveness of crop breeding programs, ultimately leading to the development of improved and more resilient cultivars to meet the challenges of global food insecurity.

The Reproducing Kernel Hilbert Spaces (RKHS) model has emerged as a prominent technique in GP, utilizing genetic marker information to predict the phenotypic performance and aiding in the selection of elite genotypes (Gianola et al., 2006; Gianola & van Kaam, 2008). While RKHS models exhibit significant potential in predicting complex traits in both animal (Long et al., 2010) and plant breeding (Crossa et al., 2010; Cuevas et al., 2016, 2018), their performance, like other GP models, can be significantly impacted by various factors. These factors include genetic relationships within the studied population, the heritability of the traits, and genotype by environment interactions (GxE).

In breeding programs where individuals exhibit low genetic relatedness, predicting phenotypic outcomes becomes more challenging due to the complex traits, as accuracy of GEBV can result in a large part from genetic relationships captured by markers (Habier et al., 2007; Werner et al., 2020). This complexity necessitates robust prediction models capable of capturing intricate genetic relationships within the population. Additionally, the heritability of the traits plays a crucial role in prediction accuracy (Kaler et al., 2022). Several studies showed a strong relationship between prediction accuracy and trait heritability (Clark et al., 2012; Lorenz et al., 2011; Luan et al., 2009; Nyine et al., 2017). Traits exhibiting high heritability are more predictable using genetic markers, while those with low heritability are more influenced by environmental factors, posing challenges for accurate prediction.

Furthermore, the genotype and environment (GxE) interaction is a critical consideration in GP due to its significant impact on quantitative or complex traits. Accounting for GxE

interaction has been shown to enhance prediction accuracy, especially in environments where these interactions are significant (Mageto et al., 2020; Rogers & Holland, 2022). This phenomenon has been observed in studies across different species, highlighting its importance in GP (Burgueño et al., 2012; Jarquín et al., 2014; Monteverde et al., 2018; Pérez-Rodríguez et al., 2015; Saint Pierre et al., 2016).

This chapter explores the prediction accuracy of the RKHS model in diverse pea accessions, investigating the influence of genetic diversity, trait heritability, and GxE interactions. By examining these factors, we seek to improve the efficacy of RKHS models in GP and contribute to the advancement of breeding strategies in plant breeding programs.

4.2. Materials and methods

4.2.1. Germplasm and phenotyping

The study utilized 300 germplasm accessions sourced from the USDA Pea Core Collection, originating from diverse geographical regions worldwide, as detailed by Bari et al. (2021). The lines were planted following an augmented incomplete block design with four diagonal repeated checks. Seed yield and agronomic data were collected in a 2-year experiment, from 2021 to 2022, across six environments: two at North Dakota State University (FAR21 and FAR22), two at Washington State University (WSU21 and WSU22), and two at Montana State University (MON21 and MON22). Standard cultural practices were implemented, and plots were harvested at physiological maturity (90-120 days after planting) and dried to 13% moisture content. For protein analysis, 0.11 kg of clean and dried harvested seeds per line was used, employing near infrared (NIR) spectroscopy.

4.2.2. Genotyping

Young leaves were harvested from seedlings of each pea line planted in a greenhouse under controlled conditions of 65 to 70°F temperature and a 16-hour light cycle. DNA extraction was conducted using the DNeasy® Plant Mini Kit (Qiagen, Germantown, MD, USA) according to the manufacturer's protocol, and elution was performed with 100µl. Subsequently, the concentration of DNA samples was quantified using the Qubit dsDNA BR Assay kit and Qubit 4.0 fluorometer (Life Technologies Corporation, Eugene, OR) and standardized to a final concentration of 25 ng/µl. Whole genome resequencing (WGR) at a depth of 10x was carried out at HudsonAlpha Genome Sequencing Center (Huntsville, AL, USA) using Illumina sequencing technology. The sequencing generated 5.9 terabytes of raw data, comprising 103 billion paired-end reads. Quality assessment of the reads was conducted using Fast QC, followed by trimming using Trimmomatic. The trimmed reads were aligned to the *Pisum sativum* Chinese reference genome (Yang et al., 2022) using bwa-mem2. PCR duplicates were identified and removed using Picard's 'MarkDuplicates' function and Samtools-1.10, respectively. Variant calling was performed using BCFtools, and filtering was applied using VCFtools, with parameters set to a minimum depth of 5, maximum missingness of 5%, and a minor allele frequency of 5%, resulting in the retrieval of 6,720,968 SNPs. Further filtering was conducted using Plink v.19 to remove SNPs with less than 10% missing values, resulting in a final set of 870,224 SNPs for downstream analysis.

4.2.3. Phenotypic data analysis

A mixed linear model was used to extract best linear unbiased estimates (BLUEs) for all traits evaluated using the following model:

$$\mathbf{y} = \mathrm{f}(\mathbf{r}, \mathbf{c}) + \mathbf{Xb} + \mathbf{Z}_\mathrm{r}\mathbf{u}_r + \mathbf{Z}_\mathrm{c}\mathbf{u}_\mathrm{c} + \boldsymbol{\varepsilon} \quad (4.1)$$

where **y** is the response variable for n-th phenotype, **b** is the fixed effect of the genotype, $\mathbf{u_r}$ and $\mathbf{u_c}$ are row and column random effects accounting for discontinuous field variation with multivariate normal distribution: $\mathbf{u_r} \sim N(0, \mathbf{I}\sigma_r^2)$ and $\mathbf{u_c} \sim N(0, \mathbf{I}\sigma_c^2)$ respectively, wherein, **I** is an identity matrix and σ_r^2 and σ_c^2 are variances due to row and column effect. f(**r, c**) is a smooth bivariate function defined over the row and column positions, $\boldsymbol{\varepsilon}$ is the measurement error from each plot with distribution of $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma_\varepsilon^2)$, wherein, **I** is the same as above and σ_ε^2 is variance for the residual term or simply referred to as nugget. **X** and **Z** are incidence matrices for the fixed and random terms, respectively. A total of 302 genotypes were found to overlap across six environments (Table A4).

4.2.4. Statistical models

The multi-trait (MT) single environment GS model was fitted by extending univariate single environment GS model (see Chapter 2 Eq. 2.2) as follows:

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{1}_1\mu_1 \\ \vdots \\ \mathbf{1}_k\mu_n \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Z}_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{4.2}$$

where $\mathbf{y}_1 \ldots \mathbf{y}_n$ are the vector of phenotypes, $\mu_1 \ldots \mu_n$ are the overall mean for each n-th trait, $\mathbf{Z}_1 \ldots \mathbf{Z}_n$ is the incidence matrix for genomic effect of the lines for each n-th trait, $\mathbf{u}_1$ .... $\mathbf{u}_n$ is the genomic effect of the lines for each n-th trait, and $\boldsymbol{\varepsilon}_1$ ... $\boldsymbol{\varepsilon}_n$ is the residual error for each n-th trait. The random term is assumed to follow the multivariate normal distribution [$\mathbf{u}_1$ .... $\mathbf{u}_n$] $\sim MN[0, (\mathbf{G} \otimes \mathbf{G}_o)]$, where **G** is the same as above and $\mathbf{G}_o$ is an n x n unstructured variance-covariance matrix of the genetic effect of the traits, this is represented as follows:

$$\mathbf{G}_o \otimes \mathbf{G} = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_{12}} & \cdots & \sigma_{g_{1n}} \\ \sigma_{g_{21}} & \sigma_{g_2}^2 & \cdots & \cdots \\ \sigma_{g_{n1}} & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ldots & \sigma_{g_n}^2 \end{bmatrix} \otimes \mathbf{G} \tag{4.3}$$

The diagonal elements represent variance for each trait and covariances between traits are the off-diagonal elements. Further, the residual term for each n-th trait is assumed to follow the multivariate normal distribution $[\boldsymbol{\varepsilon}_1 \ldots \boldsymbol{\varepsilon}_n] \sim MN[0, (\mathbf{I} \otimes \mathbf{R})]$, where **I** is the same as above and **R** is a heterogeneous diagonal matrix of the residual variances for each n-th trait:

$$\mathbf{R}=\begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\varepsilon_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2_{\varepsilon_n} \end{bmatrix} \otimes \mathbf{I} \tag{4.4}$$

The diagonal elements represent the residual variance for each n-th trait and off-diagonal elements of the **R** matrix equal zero.

For the multi-trait (MT) multi-environment GS model, univariate multi-environment HS model (See Chapter 2 Eq. 2.3) was expanded as described by Montesinos et al. (2022):

$$\mathbf{y} = \mathbf{1}_{nK}\mu + \mathbf{Z}_{1.1}\mathbf{u}_{1.1} + \mathbf{Z}_{2.1}\mathbf{u}_{2.1} + \mathbf{Z}_{3.1}\mathbf{u}_{3.1} + \boldsymbol{\varepsilon} \tag{4.5}$$

where **y** is of size i x n and i =j x k, n is the number of traits, j is the number of genotypes and k is the number of environments. $\mathbf{Z}_{1.1}$ is the incidence matrix of environment of size i x k, $\mathbf{u}_{1.1}$ is the random effect of each environment of each trait with size k x n, $\mathbf{Z}_{2.1}$ is the incidence matrix of genotypes of order i × j, $\mathbf{u}_{2.1}$ is the random effect of the genotypes i × n, and follows the multivariate normal distribution $MN(0, \sigma^2_g\mathbf{Z_g}\mathbf{G}\mathbf{Z}'_\mathbf{g}, \mathbf{U}_g)$, where $\mathbf{Z_g}$is an incidence matrix of the genotypes of order i x j. $\mathbf{G}$, $\mathbf{Z_g}\mathbf{G}\mathbf{Z}'_\mathbf{g}$and $\mathbf{Z}_k\mathbf{K}\mathbf{Z}'_\mathbf{k}$ are the same as above and $\mathbf{U}_g$is the unstructured variance-covariance matrix of traits of order n × n. $\mathbf{Z}_{3.1}$ is the incidence matrix of GE of order i × kj, $\mathbf{u}_{3.1}$ is the random effect of the genotypes by environment by trait of order kj × n and follows the matrix multivariate normal distribution $MN(0, \sigma^2_{gk}\mathbf{Z_g}\mathbf{G}\mathbf{Z}'_\mathbf{g}\#\sigma^2_k\mathbf{Z}_k\mathbf{K}\mathbf{Z}'_\mathbf{k}, \mathbf{U}_{gk})$, where $\mathbf{U}_{gk}$ is the unstructured variance-covariance matrix of order k by k. $\boldsymbol{\varepsilon}$ is the random term of the

residual and follows the multivariate normal distribution $MN(0, \mathbf{I}, \Sigma_t)$. **I** is identity matrix of order i ×n, and Σ_t is the unstructured variance-covariance matrix.

4.2.5. Cross validation scheme

Predictive ability (PA) was estimated as the Pearson correlation coefficient between predicted genomic estimated breeding values (GEBVs) and best linear unbiased estimates (BLUEs) of each trait for the complete dataset. To evaluate the whole-environment prediction, we tested each environment using data from all the other environments combined. For example, to predict FAR21, we trained the model on all data from FAR22, MON21, MON22, WSU21, and WSU22. In split-environment prediction, we divided each environment into two halves. One half was used for testing, while the other half, along with the data from the remaining environments, was used for training. For instance, to predict the performance in FAR21, we combined 50% of the FAR21 data with all data from FAR22, MON21, MON22, WSU21, and WSU22 for training. Both cross-validation schemes were repeated 30 times to ensure the robustness and reliability of the results.

4.3. Results and discussion

4.3.1. Predictive performance of RKHS model for whole-environment prediction

In Chapter 2, we assessed the capability of GS models to predict the genetic potential of two negatively correlated traits in elite pea lines across three distinct environments. Building on the superior performance of the RKHS model observed in the previous chapter, our focus shifted to evaluating this model's predictive capacity for the same traits in a highly diverse set of pea accessions (Cheng et al. 2015).

Heritability estimates showed substantial variability across different traits and environments (Fig. 4.1). Protein content exhibited a wide range of heritability estimates, ranging

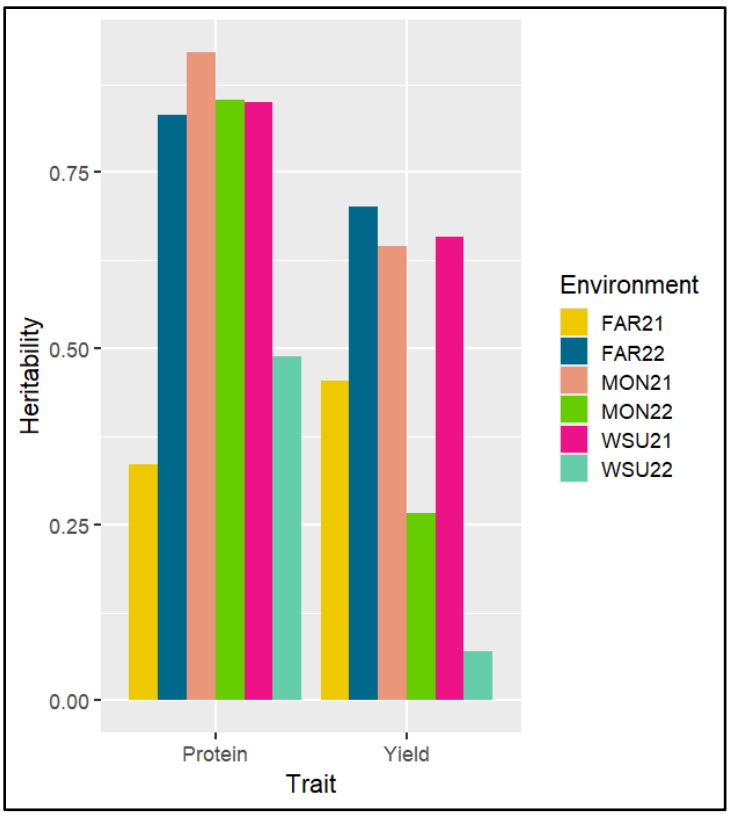

Figure 4.1. Heritability estimates for yield and protein under six environments, FAR21 is Fargo 2021, FAR22 is Fargo 2022, MON21 is Montana 2021, MON22 is Montana 2022, WSU21 is Washington State University 2021, WSU22 is Washington State University 2022.

from 0.34 to 0.92. Among the accessions, FAR21 displayed the lowest heritability, while MON21 exhibited the highest. On the other hand, yield heritability ranged from 0.07 to 0.70, with WSU22 demonstrating the lowest and FAR22 the highest heritability. The use of diverse accessions evaluated across a range of environments for complex traits appeared to enhance heritability estimates, likely due to the improved estimation of genetic effects. This suggests that the inclusion of diverse genetic backgrounds and environmental conditions can provide a more comprehensive assessment of the genetic basis of traits, leading to more accurate heritability estimates (Becker and Andreotti 2010).

Figure 4.2 illustrates the predictive performance of the RKHS model for yield under whole-environment prediction, where each environment was tested using data from

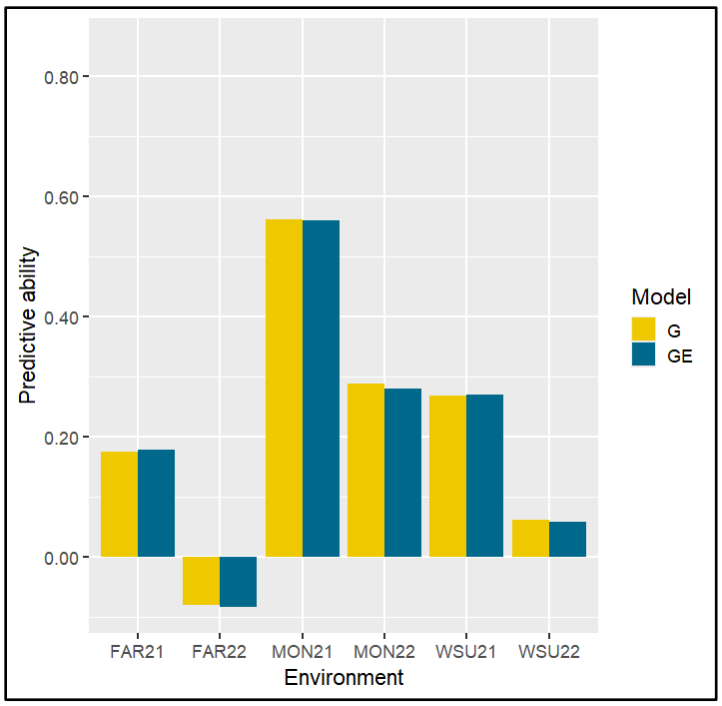

Figure 4.2. Predictive ability for yield using the RKHS model across multi-environments with whole-environment prediction, RKHS is Reproducing Kernel Hilbert Spaces, G is prediction model considering genotypic factor, GE is prediction model integrating GxE interaction.

the remaining environments. A comparison with the same RKHS model in Chapter 2, which focused on advanced breeding lines across three environments, revealed a 25% decrease in the mean predictive ability for yield (from 0.28 to 0.21). This decline in predictive ability, despite the trait's higher heritability, may be attributed to the strong influence of environmental factors on yield (Stewart-Brown et al. 2019). Interestingly, similar observations were made by de Oliveira et al. (2018) in their study of complex traits like plant height and days to flower, which exhibited low predictive abilities despite high heritability estimates. Their findings suggest that the genetic variability of the training population can significantly impact predictive ability, indicating a complex interplay of factors influencing trait heritability and prediction accuracy (Asoro et al., 2011; Daetwyler et al., 2008; de Los Campos et al., 2013; de Oliveira et al., 2018; Grattapaglia & Resende, 2011; Nakaya & Isobe, 2012).

Despite including an environment with the lowest heritability (WSU22), only FAR22, when tested, negatively affected the predictive ability (-0.08). This observation highlights the potential challenge of integrating environments with distinct genetic and environmental factors into a unified prediction model. Generally, multi-trait models can enhance the prediction accuracy of low heritability traits if they exhibit at least a moderate correlation with the highly predictive ability (Jia and Jannink 2012; Montesinos-López et al. 2016). However, the high negative correlation between seed yield and protein content in this study may have limited the improvement in predictive abilities even for traits with moderate to high heritability (Bhatta et al. 2020; Uhlarik et al. 2022).

The persistent issue of poor predictive ability poses a significant challenge in the widespread adoption of genomic selection (Crossa et al. 2014). Despite this, numerous studies have highlighted the potential benefits of genomic selection or complex traits (Crossa et al. 2014; Burgueño et al. 2012), such as grain yield in wheat (Belamkar et al. 2018; Juliana et al. 2018; Lado et al. 2018; Michel et al. 2018).

In contrast, the model's performance in predicting protein content (Fig. 4.3) showed a significant improvement when using diverse accessions across six environments. The mean predictive ability for protein content increased 129% (from 0.27 to 0.62) compared to Chapter 2 which utilized modern advanced breeding lines. This enhancement can be attributed to the higher heritability of protein content. It is well-documented that traits with higher heritability tend to exhibit higher prediction accuracies (Kaler et al. 2022; Lorenz et al. 2011; Luan et al. 2009; Nyine et al. 2017).

The superior predictive performance of the RKHS model for protein could be attributed to the higher heritability estimates of the trait across environments, indicating that the genetic

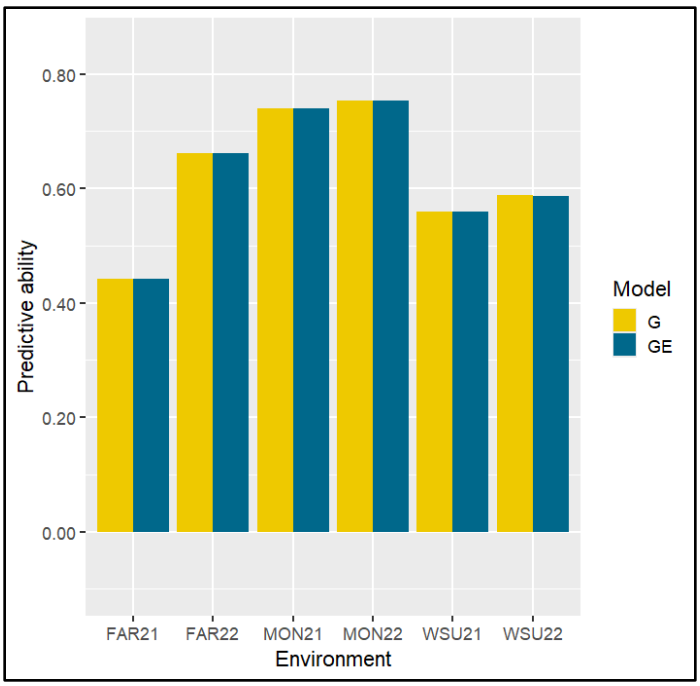

Figure 4.3. Predictive ability for seed protein content using the RKHS model across multi-environments with whole-environment prediction, RKHS is Reproducing Kernel Hilbert Spaces, G is prediction model considering genotypic factor, GE is prediction model integrating GxE interaction.

effects influencing protein content are consistent and stable regardless of the varying conditions (Sallam et al. 2015). Moreover, the inclusion of diverse accessions allowed the model to capture a wide spectrum of genetic effects influencing the trait. Furthermore, Persa et al. (2023) noted that higher genetic diversity in the soybean population training set resulted in less pronounced decay in predictive ability when using complex models that accounts for GxE interactions. This indicates that diverse alleles associated with stress resilience may contribute to higher stability in trait performance across environments.

The comparable performance of the RKHS model when considering only the genotype factor (G) compared to integrating the genotype by environment (GxE) interaction could be related to a strong population structure. Population structure can lead to inflamed genomic prediction accuracies obtained from random cross-validation (Werner et al. 2020). Breeding

populations often exhibit strong population structure due to their diverse genetic backgrounds. Several other studies have also indicated that predictive ability is primarily influenced by differences in the mean performance of breeding populations, a factor that can be highly affected in structured populations (Windhausen et al. 2012; Hickey et al. 2014; Massman et al. 2013; Lehermeier et al. 2014; Würschum et al. 2017).

4.3.2. Predictive performance of RKHS model for split-environment prediction

For split-environment prediction, we divided each environment into two halves. One half was used for testing, while the other half, along with the data from the remaining environments, was used for training. In predicting yield, we observed improved predictive abilities of the model per environment, with an average improvement of 57% (Fig. 4.4) compared to whole-environment prediction (Section 4.3.1). Notably, testing 50% of FAR21 showed the lowest improvement (-0.081 to -0.085), while testing 50% of WSU22 showed the highest improvement (0.06 to 0.31). The mean distribution of predictive ability for yield across environments is illustrated in Figure A1.

Regarding protein prediction, a significant improvement in the mean predictive ability of the RKHS model was also observed, with an average improvement of 19% (Fig. 4.5). Here, MON21 exhibited the lowest mean predictive ability improvement (0.74 to 0.75), while FAR21 showed the highest mean predictive ability improvement (0.44 to 0.75). The mean distribution of predictive ability for seed protein content across environments is illustrated in Figure A2. This improvement in predictive ability can be attributed to the variability in environments. By including a subset of the population in the training set, the model can better adapt to the specific environmental conditions present in the environment being tested. This aligns with the findings of Atanda et al. (2022), who observed that when 50% of the genotypes overlapped across the

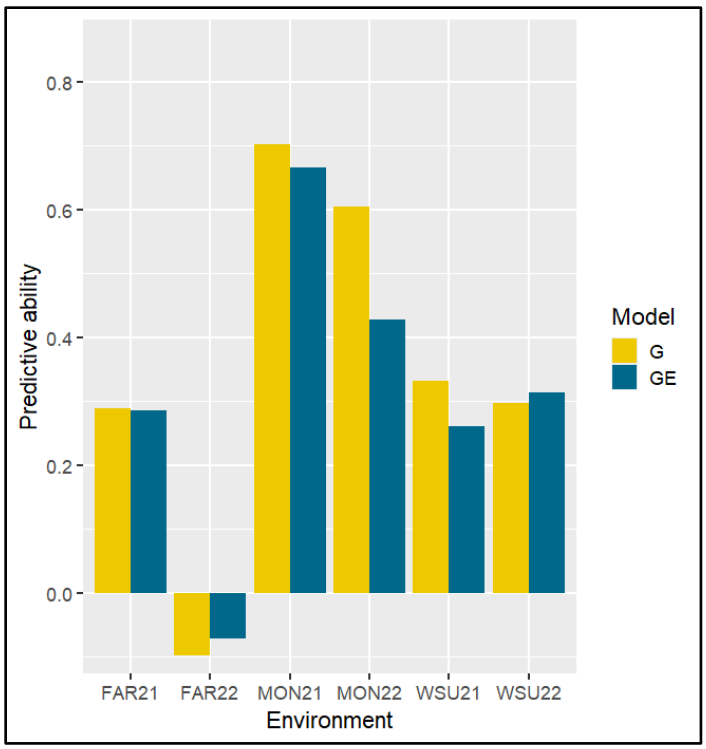

Figure 4.4. Mean predictive ability for yield using the RKHS model across multi-environments with split-environment prediction, RKHS is Reproducing Kernel Hilbert Spaces, G is prediction model considering genotypic factor, GE is prediction model integrating GxE interaction.

environments used for training, predictive ability improved compared to non-overlapping genotypes across environments. This adaptability leads to more accurate predictions for environments similar to those in the training set, compared to using a model trained on different sets of environments.

Utilizing a subset of lines that overlap across environments allows for borrowing of information across environments and serves as connectivity across environments. This approach enables the model to leverage information from closely related individuals within and across environments using multi-environment models, as noted by Atanda et al. (2021), Burgueño et al. (2012), Jarquín et al. (2020) and Atanda et al. (2022). These studies highlight that genetic correlation between environments influences predictive ability in multi-environment genomic prediction, further supporting the benefits of using a subset of the population in the training set

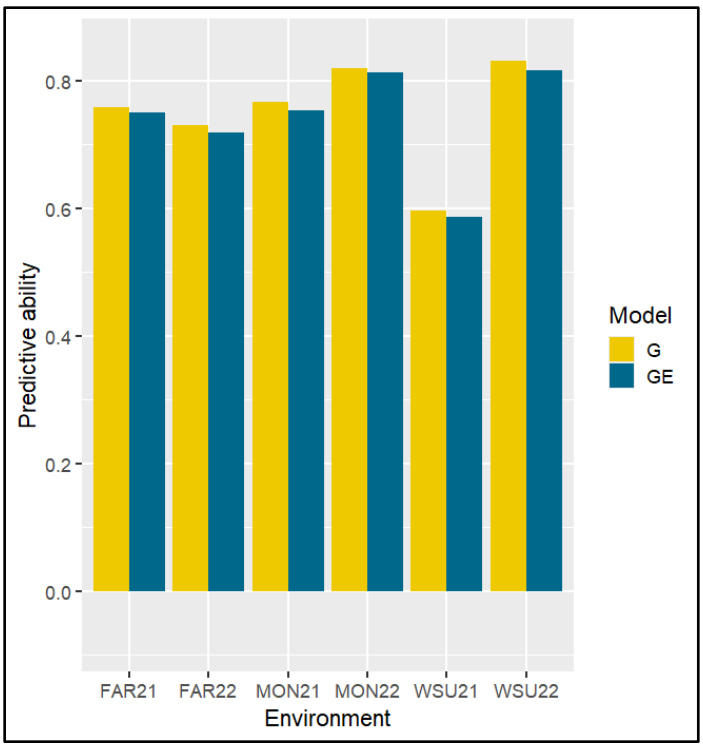

Figure 4.5. Mean predictive ability for seed protein content using the RKHS model across multi-environments with split-environment prediction, RKHS is Reproducing Kernel Hilbert Spaces, G is prediction model considering genotypic factor, GE is prediction model integrating GxE interaction.

for the improved predictive performance across environments.

4.4. Conclusion

This chapter highlights the intricate interplay between genetic diversity, trait heritability, and genotype-by-environment interactions in the predictive performance of the Multi-trait Multi-Environment RKHS model for complex traits in pea accessions. The study reveals a nuanced dynamic wherein while predictive abilities for yield may decline under strong environmental influences, the model's performance for protein content significantly improves, particularly when diverse accessions are included. This improvement underscores the importance of integrating considerations of genetic diversity and environmental variability in genomic prediction models.

Nonetheless, our findings also highlight the challenges in seamlessly integrating environments with distinct genetic backgrounds into unified prediction models, emphasizing the need for robust multi-trait models capable of capturing complex genetic relationships. Future research should focus on refining models to adapt to specific environmental conditions, thereby enhancing the development of resilient crop cultivars to address global food security challenges.

It is imperative to explore strategies to mitigate the impact of population structure and genotype-by-environment interactions in the training set. Strategies such as incorporating structured populations into the training sets, utilizing advanced statistical methods to account for population structure, and integrating environmental covariates in the model are all promising avenues for improving the accuracy and robustness of genomic prediction models. Furthermore, investigating the genetic basis of GxE interactions and developing models that can effectively capture and predict these interactions will be crucial for advancing the predictive performance of genomic selection in plant breeding programs.

4.5. Literature cited

Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott & J-L. Jannink. (2011). Accuracy and training population design for genomic selection on quantitative traits in elite north American oats. *The Plant Genome*. 4(2): 132-144.

Atanda, S.A., V. Govindan, R. Singh, K.R. Robbins, J. Crossa & A.R. Bentley. (2022). Sparse testing using genomic prediction improves selection for breeding targets in elite spring wheat. *Theor. Appl. Genet.* 135(6): 1939-1950.

Atanda, S. A., M. Olsen, J. Crossa, J. Burgueño, R. Rincent, D. Dzidzienyo, Y. Beyene, M. Gowda, K. Dreher, P.M. Boddupalli, P. Tongoona, E.Y. Danquah, G. Olaoye & K.R. Robbins. (2021). Scalable sparse testing genomic selection strategy for early yield testing stage. *Front. Plant Sci.* 12: 658978.

Baranger, A., G. Aubert, G. Arnau, A.L. Lainé, G. Deniot, J. Potier, C. Weinachter, I. Lejeune-Hénaut, J. Lallemand & J. Burstin. (2004). Genetic diversity within *Pisum sativum* using protein- and PCR-based markers. *Theor. Appl. Genet.* 108(7): 1309-1321.

Bari, M.A.A., P. Zheng, I. Viera, H. Worral, S. Szwiec, Y. Ma, D. Main, C.J. Coyne, R.J. McGee & N. Bandillo. (2021). Harnessing genetic diversity in the USDA pea germplasm collection through genomic prediction. *Front. Genet.* 12: 707754.

Becker, R.C. & F. Andreotti. (2010). Genetics and genomics in the management of hemostasis and thrombosis. In Essentials of Genomic and Personalized Medicine (374–389). *Elsevier*.

Belamkar, V., M.J. Guttieri, W. Hussain, D. Jarquín, I. El-Basyoni, J. Poland, A.J. Lorenz & P.S. Baenziger. (2018). Genomic selection in preliminary yield trials in a winter wheat breeding program. *G3*. 8(8): 2735-2747.

Bhatta, M., L. Gutierrez, L. Cammarota, F. Cardozo, S. Germán, B. Gómez-Guerrero, M.F. Pardo, S.V. Lanaro, M. Sayas & A.J. Castro. (2020). Multi-trait genomic prediction model increased the predictive ability for agronomic and malting quality traits in barley (*Hordeum vulgare* L.) *G3*. 10(3): 1113-1124.

Burgueño, J., G. de los Campos, K. Weigel & J. Crossa. (2012). Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52(2): 707-719.

Burstin, J., P. Salloignon, M. Chabert-Martinello, J-B. Magnin-Robert, M. Siol, F. Jacquin, A. Chauveau, C. Pont, G. Aubert, C. Delaitre, C. Truntzer & G. Duc. (2015). Genetic diversity and trait genomic prediction in a pea diversity panel. *BMC Genomics*. 16(1): 105.

Cheng, P., W. Holdsworth, Y. Ma, C.J. Coyne, M. Mazourek, M.A. Grusak, S. Fuchs & R.J. McGee. (2015). Association mapping of agronomic and quality traits in USDA pea single-plant collection. *Mol Breeding*, 35(2): 75. https://doi.org/10.1007/s11032-015-0277-6.

Clark, S. A., Hickey, J. M., Daetwyler, H. D., & van der Werf, J. H. J. (2012). The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet. Sel. Evol.* 44(1): 4.

Crossa, J., G. de los Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus, D. Makumbi, R.P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Banziger & H-J. Braun. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genet.* 186(2), 713-724.

Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, J. Cerón-Rojas, X. Zhang, S. Dreisigacker, R. Babu, Y. Li, D. Bonnett & K. Mathews. (2014). Genomic prediction in CIMMYT maize and wheat breeding programs. *Hered.* 112(1): 48-60.

Cuevas, J., J. Crossa, V. Soberanis, S. Pérez-Elizalde, P. Pérez-Rodríguez, G. de los Campos, G. O.A. Montesinos-López & J. Burgueño. (2016). Genomic prediction of genotype × environment interaction kernel regression models. *The Plant Genome*. 9(3). https://doi.org/10.3835/plantgenome2016.03.0024.

Cuevas, J., I. Granato, R. Fritsche-Neto, O.A. Montesinos-Lopez, J. Burgueño, E. Bandeira, M. Sousa & J. Crossa. (2018). Genomic-enabled prediction kernel models with random intercepts for multi-environment trials. *G3*. 8(4): 1347-1365.

Daetwyler, H.D., B. Villanueva & J.A. Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PloS one*. 3(10): e3395.

de Los Campos, G., J.M. Hickey, R. Pong-Wong, H.D. Daetwyler & M.P.L. Calus. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics, 193(2), 327-345.

de Oliveira, A.A., M.M. Pastina, V.F. de Souza, R.A. da Costa Parrella, R.W. Noda, M.L.F. Simeone, R.E. Schaffert, J.V. de Magalhães, C.M.B. Damasceno & G.R.A. Margarido. (2018). Genomic prediction applied to high-biomass sorghum for bioenergy production. *Mol Breed.* 38(4): 49.

Deulvot, C., H. Charrel, A. Marty, F. Jacquin, C. Donnadieu, I. Lejeune-Hénaut, J. Burstin & G. Aubert. (2010). Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. *BMC Genomics*. 11: 468.

Duarte, J., N. Rivière, A. Baranger, G. Aubert, J. Burstin, L., Cornet, C. Lavaud, I. Lejeune-Hénaut, J-P. Martinant, J-P. Pichon, M-L. Pilet-Nayel & G. Boutet. (2014). Transcriptome sequencing for high throughput SNP development and genetic mapping in Pea. *BMC Genomics*. 15: 126.

Gianola, D., R.L. Fernando & A. Stella. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genet.* 173(3): 1761-1776.

Gianola, D. & J.B. van Kaam (2008). Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genet.*178(4): 2289-2303.

Goodwin, S., J.D. McPherson & W.R. McCombie. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17(6): 333-351.

Grattapaglia, D. & M.D.V. Resende. (2011). Genomic selection in forest tree breeding. *Tree Genet. Genomes.* 7(2): 241–255.

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

Hickey, J.M., S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B.M. Prasanna, M. Grondona, A. Zambelli, V.S. Windhausen, K. Mathews & G. Gorjanc (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Sci.* 54(4): 1476-1488.

Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Pérez, M. Calus, J. Burgueño & G. de los Campos. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3): 595–607.

Jarquín, D., R. Howard, J. Crossa, Y. Beyene, M. Gowda, J.W.R. Martini, G. Covarrubias Pazaran, J. Burgueño, A. Pacheco, M. Grondona, V. Wimmer & B.M. Prasanna. (2020). Genomic prediction enhanced sparse testing for multi-environment trials. *G3*. 10(8): 2725-2739.

Jia, Y. & J-L. Jannink. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genet.* 192(4): 15131522.

Jing, R., M.A. Ambrose, M.R. Knox, P. Smykal, M. Hybl, Á. Ramos, C. Caminero, J. Burstin, G. Duc, L.J.M. van Soest, W.K. Święcicki, M.G. Pereira, M. Vishnyakova, G.F. Davenport, A.J. Flavell & T.H.N. Ellis. (2012). Genetic diversity in European Pisum germplasm collections. *Theor. Appl. Genet.* 125(2): 367-380.

Jing, R., A. Vershinin, J. Grzebyta, P. Shaw, P. Smýkal, D. Marshall, M.J. Ambrose, T.H.N. Ellis & A.J. Flavell. (2010). The genetic diversity and evolution of field pea (*Pisum sativum* L.) studied by high throughput retrotransposon-based insertion polymorphism (RBIP) marker analysis. *BMC Evol. Biol.* 10: 44.

Juliana, P., R.P. Singh, J. Poland, S. Mondal, J. Crossa, O.A. Montesinos-López, S. Dreisigacker, P. Pérez-Rodríguez, J. Huerta-Espino, L. Crespo-Herrera & V. Govindan. (2018). Prospects and challenges of applied genomic selection-a new paradigm in breeding for grain yield in bread wheat. *The Plant Genome*. 11(3). https://doi.org/10.3835/plantgenome2018.03.0017.

Kaler, A.S., L.C. Purcell, T. Beissinger & J.D. Gillman. (2022). Genomic prediction models for traits differing in heritability for soybean, rice, and maize. *BMC Plant Biol.* 22(1), 87.

Kwon, S-J., A.F. Brown, J. Hu, R. McGee, C. Watt, T. Kisha, G. Timmerman-Vaughan, M. Grusak, K.E. McPhee & C.J. Coyne. (2012). Genetic diversity, population structure and

genome-wide marker-trait association analysis emphasizing seed nutrients of the USDA pea (*Pisum sativum* L.) core collection. *Genes Genom.* 34(3): 305-320.

Lado, B., D. Vázquez, M. Quincke, P. Silva, I. Aguilar & L. Gutiérrez, L. (2018). Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theor. Appl. Genet.* 131(12): 2719-2731.

Lehermeier, C., N. Krämer, E. Bauer, C. Bauland, C. Camisan, L. Campo, P. Flament, A.E. Melchinger, M. Menz, N. Meyer, L. Moreau, J. Moreno-González, M. Ouzunova, H. Pausch, N. Ranc, W. Schipprack, M. Schönleben, H. Walter, A. Charcosset & C-C. Schön. (2014). Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. *Genet.* 198(1): 3-16.

Leonforte, A., S. Sudheesh, N.O.I. Cogan, P.A. Salisbury, M.E. Nicolas, M. Materne, J.W. Forster & S. Kaur. (2013). SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.). *BMC Plant Biol.* 13: 161.

Levy, S.E. & R.M. Myers. (2016). Advancements in next-generation sequencing. *Annu Rev Genomics Hum Genet.* 17: 95-115.

Long, N., D. Gianola, G.J.M. Rosa, K.A. Weigel, A. Kranis & O. González-Recio. (2010). Radial basis function regression methods for predicting quantitative traits using SNP markers. *Genet. Res.* 92(3): 209-225.

Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells & J-L. Jannink. (2011). Genomic Selection in Plant Breeding. In Advance Agronomy (pp. 77–123). *Elsevier*.

Luan, T., J.A. Woolliams, S. Lien, M. Kent, M. Svendsen & T.H.E. Meuwissen. (2009). The accuracy of genomic delection in Norwegian red cattle assessed by cross-validation. *Genet.* 183(3): 1119-1126.

Mageto, E.K., J. Crossa, P. Pérez-Rodríguez, T. Dhliwayo, N. Palacios-Rojas, M. Lee, R. Guo, F. San Vicente, X. Zhang & V. Hindu. (2020). Genomic prediction with genotype by environment interaction analysis for kernel zinc concentration in tropical maize germplasm. *G3.* 10(8): 2629-2639.

Massman, J.M., A. Gordillo, R.E. Lorenzana & R. Bernardo. (2013). Genome-wide predictions from maize single-cross data. *Theor. Appl. Genet.* 126(1): 13-22.

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

Michel, S., C. Kummer, M. Gallee, J. Hellinger, C. Ametz, B. Akgöl, D. Epure, H. Güngör, F. Löschenberger & H. Buerstmayr. (2018). Improving the baking quality of bread wheat by genomic selection in early generations. *Theor. Appl. Genet.* 131(2): 477-493.

Montesinos-López, O.A., A. Montesinos-López, J. Crossa, F.H. Toledo, O. Pérez-Hernández, K.M. Eskridge & J. Rutkoski. (2016). A genomic bayesian multi-trait and multi-environment model. *G3.* 6(9): 2725-2744.

Monteverde, E., J.E. Rosas, P. Blanco, F. Pérez de Vida, V. Bonnecarrère, G. Quero, L. Gutierrez & S. McCouch (2018). Multienvironment models increase prediction accuracy of complex traits in advanced breeding lines of rice. *Crop Sci.* 58(4): 1519-1530.

Nakaya, A. & S.N. Isobe. (2012). Will genomic selection be a practical method for plant breeding? *Ann. Bot.* 110(6): 1303-1316.

Nyine, M., B. Uwimana, R. Swennen, M. Batte, A. Brown, P. Christelová, E. Hřibová, J. Lorenzen & J. Doležel. (2017). Trait variation and genetic diversity in a banana genomic selection training population. *PloS one*. 12(6): e0178734.

Pérez-Rodríguez, P., J. Crossa, K. Bondalapati, G. De Meyer, F. Pita & G. de los Campos. (2015). A pedigree-based reaction norm model for prediction of cotton yield in multienvironment trials. *Crop Sci.* 55(3), 1143-1151.

Persa, R., C. Canella Vieira, E. Rios, V. Hoyos-Villegas, C.D. Messina, D. Runcie & D. Jarquin. (2023). Improving predictive ability in sparse testing designs in soybean populations. *Front. Genet.* 14: 1269255.

Rogers, A.R. & J.B. Holland/ (2022). Environment-specific genomic prediction ability in maize using environmental covariates depends on environmental similarity to training data. *G3.* 12(2). https://doi.org/10.1093/g3journal/jkab440.

Saint Pierre, C., J. Burgueño, J. Crossa, G. Fuentes Dávila, P. Figueroa López, E. Solís Moya, J. Ireta Moreno, V.M. Hernández Muela, V.M. Zamora Villa, P. Vikram, K. Mathews, C. Sansaloni, D. Sehgal, D. Jarquin, P. Wenzl & S. Singh, S. (2016). Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. S*ci. Rep.* 6: 27312.

Sallam, A.H., J.B. Endelman, J-L. Jannink & K.P. Smith. (2015). Assessing genomic selection prediction accuracy in a dynamic barley breeding population. *The Plant Genome*. 8(1): eplantgenome2014.05.0020.

Satam, H., K. Joshi, U. Mangrolia, S. Waghoo, G. Zaidi, S. Rawool, R.P. Thakare, S. Banday, A.K. Mishra, G. Das & S.K. Malonia. (2023). Next-generation sequencing technology: current trends and advancements. *Biol.* 12(7). https://doi.org/10.3390/biology12070997.

Sindhu, A., L. Ramsay, L-A. Sanderson, R. Stonehouse, R. Li, J. Condie, A.S.K. Shunmugam, Y. Liu, A.B. Jha, M. Diapari, J. Burstin, G. Aubert, B. Tar'an, K.E. Bett, T.D. Warkentin & A.G. Sharpe. (2014). Gene-based SNP discovery and genetic mapping in pea. *Theor. Appl. Genet.* 127(10): 2225-2241.

Smýkal, P., M. Hýbl, J. Corander, J. Jarkovský, A.J. Flavell & M. Griga. (2008). Genetic diversity and population structure of pea (*Pisum sativum* L.) varieties derived from combined retrotransposon, microsatellite and morphological marker analysis. *Theor. Appl. Genet.* 117(3): 413-424.

Stewart-Brown, B.B., Q. Song, J.N. Vaughn & Z. Li. (2019). Genomic selection for yield and seed composition traits within an applied soybean breeding program. *G3*. 9(7), 2253-2265.

Tar'an, B., C. Zhang, T. Warkentin, A. Tullu & A. Vandenberg (2005). Genetic diversity among varieties and wild species accessions of pea (*Pisum sativum* L.) based on molecular markers, and morphological and physiological characters. *Genome.* 48(2): 257–272.

Uhlarik, A., M. Ćeran, D. Živanov, R. Grumeza, L. Skøt, E. Sizer-Coverdale & D. Lloyd. (2022). Phenotypic and genotypic characterization and correlation analysis of pea (*Pisum sativum* L.) diversity panel. *Plants.* 11(10). https://doi.org/10.3390/plants11101321.

Werner, C.R., R.C. Gaynor, G. Gorjanc, J.M. Hickey, T. Kox, A. Abbadi, G. Leckband, R.J. Snowdon & A. Stahl. (2020). How population structure impacts genomic selection accuracy in cross-validation: implications for practical breeding. *Front. Plant Sci.* 11: 592977.

Windhausen, V.S., G.N. Atlin, J.M. Hickey, J. Crossa, J-L. Jannink, M.E. Sorrells, B. Raman, J.E. Cairns, A. Tarekegne, K. Semagn, Y. Beyene, P. Grudloyma, F. Technow, C.

Riedelsheimer & A.E. Melchinger (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3*. 2(11): 1427–1436.

Würschum, T., H.P. Maurer, S. Weissmann, V. Hahn & W.L. Leiser. (2017). Accuracy of within- and among-family genomic prediction in triticale. *Plant Breed.* 136(2): 230-236.

Yang, T., R. Liu, Y. Luo, S. Hu, D. Wang, C. Wang, M.K. Pandey, S. Ge, Q. Xu, N. Li, G. Li, Y. Huang, R.K. Saxena, Y. Ji, M. Li, X. Yan, Y. He, Y. Liu, X. Wang, C. Xiang, R.K. Varshney, H. Ding, S. Gao & X. Zong. (2022). Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. *Nat Genet.* 54(10): 1553–1563.

Zong, X., R.J. Redden, Q. Liu, S. Wang, J. Guan, J. Liu, Y. Xu, X. Liu, J. Gu, L. Yan, P. Ades & R. Ford. (2009). Analysis of a diverse global *Pisum* sp. collection and comparison to a Chinese local *P. sativum* collection with microsatellite markers. *Theor. Appl. Genet.* 118(2): 193–204.

APPENDIX

Table A1. List of overlapping NDSU pea accessions across three environments.

GENOTYPE	PEDIGREE
NDP080169	BIG-DADDY//STO_4031_AM2_160_8321/PS310150
NDP080175	PS01101184/SUPRA
NDP080176	PS01101184/SUPRA
NDP101185	SUPRA/PS01102929
NDP101186	SUPRA/PS01102929
NDP120018	CDC MEADOW/PS05ND310
NDP120080	THUNDERBIRD/PS05ND325
NDP120150	PS05ND327/CDC MEADOW
NDP120180	PS05ND330/THUNDERBIRD
NDP121556	PS02100151/STIRLING
NDP121608	DS ADMIRAL/PS03100278
NDP121638	PS03100278/DS ADMIRAL
NDP121688	UNIVERSAL/PS01102958
NDP130001	DS ADMIRAL/PS05ND218
NDP130002	DS ADMIRAL/PS05ND218
NDP130010	DS ADMIRAL/PS05ND310
NDP130013	DS ADMIRAL/PS05ND310
NDP130046	MEDORA/PS05ND327
NDP130059	LIFTER/PS05ND310
NDP130079	STIRLING/PS05ND330
NDP130085	CDC MOZART/PS05ND218
NDP130110	CDC GOLDEN/PS05ND310
NDP130134	CDC MEADOW/PS05ND310
NDP130152	COOPER/PS05ND227
NDP130158	COOPER/PS05ND310
NDP130167	COOPER/PS05ND430
NDP130302	STIRLING/PS03100546
NDP130337	DS ADMIRAL/PS01102958
NDP130340	DS ADMIRAL/PS01102958
NDP140005	DS ADMIRAL/PS05ND218
NDP140006	DS ADMIRAL/PS05ND218
NDP140366	PS05ND327/CDC GOLDEN
NDP140390	PS05ND327/THUNDERBIRD

Table A1. List of overlapping NDSU pea accessions across three environments (continued).

GENOTYPE	PEDIGREE
NDP150001	GSP-Ae-D9904-17/MEDORA
NDP150013	DS ADMIRAL/PS05ND218
NDP150037	DS ADMIRAL/PS05ND310
NDP150038	DS ADMIRAL/PS05ND310
NDP150042	DS ADMIRAL/PS05ND310
NDP150045	MEDORA/PS05ND218
NDP150046	MEDORA/PS05ND218
NDP150047	MEDORA/PS05ND218
NDP150051	MEDORA/PS05ND218
NDP150052	MEDORA/PS05ND218
NDP150053	MEDORA/PS05ND218
NDP150054	MEDORA/PS05ND218
NDP150055	MEDORA/PS05ND218
NDP150058	MEDORA/PS05ND218
NDP150059	MEDORA/PS05ND218
NDP150060	MEDORA/PS05ND218
NDP150062	MEDORA/PS05ND218
NDP150063	MEDORA/PS05ND218
NDP150066	MEDORA/PS05ND227
NDP150068	MEDORA/PS05ND227
NDP150069	MEDORA/PS05ND227
NDP150070G	MEDORA/PS05ND227
NDP150073	MEDORA/PS05ND227
NDP150075	MEDORA/PS05ND227
NDP150076	MEDORA/PS05ND227
NDP150077	MEDORA/PS05ND227
NDP150079	MEDORA/PS05ND227
NDP150080	MEDORA/PS05ND227
NDP150081	MEDORA/PS05ND227
NDP150082	MEDORA/PS05ND227
NDP150084	MEDORA/PS05ND227
NDP150085	MEDORA/PS05ND310
NDP150087	MEDORA/PS05ND310
NDP150089	MEDORA/PS05ND310

Table A1. List of overlapping NDSU pea accessions across three environments (continued).

GENOTYPE	PEDIGREE
NDP150091	MEDORA/PS05ND310
NDP150094	MEDORA/PS05ND310
NDP150099	MEDORA/PS05ND310
NDP150100	MEDORA/PS05ND310
NDP150105	STIRLING/PS05ND430
NDP150106	STIRLING/PS05ND430
NDP150108	STIRLING/PS05ND430
NDP150109	STIRLING/PS05ND430
NDP150110	STIRLING/PS05ND430
NDP150112	STIRLING/PS05ND430
NDP150113	STIRLING/PS05ND430
NDP150114	STIRLING/PS05ND430
NDP150117	STIRLING/PS05ND430
NDP150119	STIRLING/PS05ND430
NDP150121	CDC GOLDEN/PS05ND227
NDP150125	CDC GOLDEN/PS05ND227
NDP150127	CDC GOLDEN/PS05ND227
NDP150128	CDC GOLDEN/PS05ND227
NDP150129	CDC GOLDEN/PS05ND227
NDP150130	CDC GOLDEN/PS05ND227
NDP150131	CDC GOLDEN/PS05ND227
NDP150140	CDC GOLDEN/PS05ND310
NDP150142	CDC GOLDEN/PS05ND310
NDP150151	CDC GOLDEN/PS05ND310
NDP150162	CDC MEADOW/PS05ND227
NDP150168	CDC MEADOW/PS05ND227
NDP150169	CDC MEADOW/PS05ND227
NDP150176	CDC MEADOW/PS05ND227
NDP150178	CDC MEADOW/PS05ND310
NDP150179	CDC MEADOW/PS05ND310
NDP150184	CDC MEADOW/PS05ND310
NDP150191	CDC MEADOW/PS05ND310
NDP150192	CDC MEADOW/PS05ND310
NDP150197	THUNDERBIRD/PS05ND310

Table A1. List of overlapping NDSU pea accessions across three environments (continued).

GENOTYPE	PEDIGREE
NDP150198G	THUNDERBIRD/PS05ND310
NDP150199	THUNDERBIRD/PS05ND310
NDP150200	THUNDERBIRD/PS05ND310
NDP150201	THUNDERBIRD/PS05ND310
NDP150203	THUNDERBIRD/PS05ND310
NDP150206	THUNDERBIRD/PS05ND310
NDP150210	THUNDERBIRD/PS05ND310
NDP150213	THUNDERBIRD/PS05ND430
NDP150214	THUNDERBIRD/PS05ND430
NDP150215	THUNDERBIRD/PS05ND430
NDP150216	THUNDERBIRD/PS05ND430
NDP150217	THUNDERBIRD/PS05ND430
NDP150218	THUNDERBIRD/PS05ND430
NDP150220	THUNDERBIRD/PS05ND430
NDP150222	THUNDERBIRD/PS05ND430
NDP150224	THUNDERBIRD/PS05ND430
NDP150225	THUNDERBIRD/PS05ND430
NDP150227	THUNDERBIRD/PS05ND430
NDP150228	THUNDERBIRD/PS05ND430
NDP150230	THUNDERBIRD/PS05ND430
NDP150232	PS05ND218/STIRLING
NDP150235	PS05ND218/STIRLING
NDP150237	PS05ND218/STIRLING
NDP150258	PS05ND227/DS ADMIRAL
NDP150269	PS05ND325/DS ADMIRAL
NDP150288	MEDORA/PS05ND327
NDP150289	MEDORA/PS05ND327
NDP150317	CDC MOZART/PS05ND430
NDP150318	CDC MOZART/PS05ND430
NDP150321	CDC MEADOW/PS05ND327
NDP150326	CDC MEADOW/PS05ND327
NDP150338	PS05ND218/THUNDERBIRD
NDP150378	PS05ND430/DS ADMIRAL
NDP150380	PS05ND430/DS ADMIRAL

Table A1. List of overlapping NDSU pea accessions across three environments (continued).

GENOTYPE	PEDIGREE
NDP150382	PS05ND430/DS ADMIRAL
NDP150386	PS05ND430/DS ADMIRAL
NDP150387	PS05ND430/DS ADMIRAL
NDP150392	PS05ND430/MEDORA
NDP150401	PS05ND430/CDC GOLDEN
NDP150407	PS05ND430/CDC MEADOW
NDP150416	PS05ND430/CDC MEADOW
NDP150419	PS05ND430/CDC MEADOW
NDP150456	NDP080174/NDP080169
NDP150459	NDP080174/NDP080169
NDP150476	CDC GOLDEN/PS05ND310
NDP150495	CDC MEADOW/PS05ND227
NDP150501	CDC MEADOW/PS05ND227
NDP150513	CDC MEADOW/PS05ND310
NDP150528	PS05ND325/DS ADMIRAL
NDP160010	CDC GOLDEN/PS05ND227
NDP160022	CDC MEADOW/PS05ND227
NDP160034	CDC MEADOW/PS05ND310
NDP160049	THUNDERBIRD/PS05ND430
NDP160051	THUNDERBIRD/PS05ND430
NDP160055	THUNDERBIRD/PS05ND430
NDP160062	PS05ND218/STIRLING
NDP160066	PS05ND218/STIRLING
NDP160071	PS05ND227/DS ADMIRAL
NDP160075	PS05ND227/DS ADMIRAL
NDP160169	PS05ND325/DS ADMIRAL
NDP160176	PS05ND325/MEDORA
NDP160180	PS05ND325/MEDORA
NDP160183	PS05ND325/MEDORA
NDP160188	PS05ND330/CDC MEADOW
NDP160193	PS05ND330/CDC MEADOW
NDP160195	PS05ND430/DS ADMIRAL
NDP160196	PS05ND430/DS ADMIRAL
NDP160197	PS05ND430/DS ADMIRAL

Table A1. List of overlapping NDSU pea accessions across three environments (continued).

GENOTYPE	PEDIGREE
NDP160201	PS05ND430/DS ADMIRAL
NDP160204	PS05ND430/DS ADMIRAL
NDP160216	PS05ND430/MEDORA
NDP160218	PS05ND430/MEDORA
NDP160226	PS05ND430/CDC MEADOW
NDP160231	PS05ND430/CDC MEADOW
NDP160274	PS05ND0232/STIRLING
NDP160278	PS07ND0102/STIRLING
NDP160279	PS07ND0102/STIRLING
NDP160281	NDP080138/STIRLING
NDP160305	NDP080142/LIFTER
NDP170004G	N16P097/PS07ND0190
NDP170028G	N16P097/PS07ND0190
NDP170039G	PS07ND0190/N16P097
NDP170089G	N16P106/NDP121166
NDP170111G	N16P106/NDP121166
PS07100972	BIG-DADDY//MARO/PS310148
PS07100995	PS01101184/SUPRA
PS07101014	MARROWFAT/WV135C*6af/2/PS210713/3/CEB_1221/4/MARO/PS310148

Table A2. List of USDA pea accessions under Fargo 2022 (FAR22) experiment.

GENOTYPE	DESCRIPTION
PI 102887	Inbred line
PI 102888 PSP	Inbred line
PI 116056 PSP	Inbred line
PI 116944 PSP	Inbred line
PI 117998 PSP	Inbred line
PI 121352 PSP	Inbred line
PI 123246	Inbred line
PI 125839 PSP	Inbred line
PI 125840 PSP	Inbred line
PI 137118	Inbred line
PI 137119 PSP	Inbred line
PI 137120	Inbred line
PI 138945	Inbred line
PI 140295	Inbred line
PI 140296	Inbred line
PI 140298 PSP	Inbred line
PI 142774	Inbred line
PI 142777	Inbred line
PI 143485 PSP	Inbred line
PI 155109 PSP	Inbred line
PI 156720 PSP	Inbred line
PI 162909 PSP	Inbred line
PI 163125	Inbred line
PI 163126 PSP	Inbred line
PI 164285	Inbred line
PI 164346	Inbred line
PI 164417	Inbred line
PI 164612 PSP	Inbred line
PI 164614	Inbred line
PI 164669	Inbred line
PI 164779 PSP	Inbred line
PI 164836	Inbred line
PI 164838	Inbred line
PI 164971 PSP	Inbred line

Table A2. List of USDA pea accessions under Fargo 2022 (FAR22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 164972 PSP	Inbred line
PI 165949 PSP	Inbred line
PI 166084 PSP	Inbred line
PI 166142	Inbred line
PI 166159 PSP	Inbred line
PI 166187	Inbred line
PI 167250	Inbred line
PI 167253	Inbred line
PI 171810 PSP	Inbred line
PI 171814	Inbred line
PI 174321	Inbred line
PI 174921 PSP	Inbred line
PI 174922	Inbred line
PI 174925	Inbred line
PI 175228	Inbred line
PI 175231 PSP	Inbred line
PI 175232	Inbred line
PI 179019	Inbred line
PI 179449	Inbred line
PI 179451 PSP	Inbred line
PI 179722 PSP	Inbred line
PI 180693 PSP	Inbred line
PI 180702 PSP	Inbred line
PI 181800	Inbred line
PI 183467 PSP	Inbred line
PI 184130 PSP	Inbred line
PI 184784 PSP	Inbred line
PI 189171	Inbred line
PI 193578 PSP	Inbred line
PI 193586	Inbred line
PI 193588	Inbred line
PI 193590 PSP	Inbred line
PI 193836	Inbred line
PI 193837	Inbred line

Table A2. List of USDA pea accessions under Fargo 2022 (FAR22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 193838	Inbred line
PI 194340	Inbred line
PI 195020 PSP	Inbred line
PI 195404 PSP	Inbred line
PI 195631 PSP	Inbred line
PI 196017	Inbred line
PI 196026	Inbred line
PI 196027	Inbred line
PI 197990 PSP	Inbred line
PI 198072 PSP	Inbred line
PI 198735 PSP	Inbred line
PI 204306 PSP	Inbred line
PI 206006 PSP	Inbred line
PI 207508 PSP	Inbred line
PI 210571 PSP	Inbred line
PI 212031 PSP	Inbred line
PI 212112	Inbred line
PI 220174 PSP	Inbred line
PI 220189 PSP	Inbred line
PI 222117 PSP	Inbred line
PI 223527 PSP	Inbred line
PI 226561	Inbred line
PI 226562	Inbred line
PI 227258 PSP	Inbred line
PI 227457	Inbred line
PI 236492 PSP	Inbred line
PI 240516 PSP	Inbred line
PI 241593 PSP	Inbred line
PI 242027 PSP	Inbred line
PI 244093 PSP	Inbred line
PI 244121 PSP	Inbred line
PI 244129	Inbred line
PI 244191 PSP	Inbred line
PI 244262	Inbred line

Table A2. List of USDA pea accessions under Fargo 2022 (FAR22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 249645 PSP	Inbred line
PI 249646	Inbred line
PI 250438 PSP	Inbred line
PI 250440 PSP	Inbred line
PI 250448 PSP	Inbred line
PI 253968 PSP	Inbred line
PI 257244 PSP	Inbred line
PI 257592 PSP	Inbred line
PI 261623 PSP	Inbred line
PI 261677 PSP	Inbred line
PI 263011	Inbred line
PI 263030 PSP	Inbred line
PI 266070 PSP	Inbred line
PI 269761 PSP	Inbred line
PI 269762 PSP	Inbred line
PI 269763	Inbred line
PI 269771	Inbred line
PI 269774	Inbred line
PI 269776	Inbred line
PI 269777 PSP	Inbred line
PI 269802 PSP	Inbred line
PI 269804 PSP	Inbred line
PI 269818 PSP	Inbred line
PI 270536 PSP	Inbred line
PI 271035 PSP	Inbred line
PI 271116 PSP	Inbred line
PI 272148 PSP	Inbred line
PI 272161	Inbred line
PI 272171 PSP	Inbred line
PI 272184 PSP	Inbred line
PI 272194 PSP	Inbred line
PI 272204 PSP	Inbred line
PI 272215 PSP	Inbred line
PI 272216 PSP	Inbred line

Table A2. List of USDA pea accessions under Fargo 2022 (FAR22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 272218 PSP	Inbred line
PI 273605 PSP	Inbred line
PI 273676	Inbred line
PI 274307 PSP	Inbred line
PI 274308 PSP	Inbred line
PI 274584 PSP	Inbred line
PI 280252 PSP	Inbred line
PI 280607	Inbred line
PI 280613 PSP	Inbred line
PI 280617 PSP	Inbred line
PI 280619 PSP	Inbred line
PI 280621	Inbred line
PI 285710 PSP	Inbred line
PI 285727 PSP	Inbred line
PI 286430 PSP	Inbred line
PI 286607 PSP	Inbred line
PI 299023	Inbred line
PI 306590	Inbred line
PI 306591 PSP	Inbred line
PI 307666 PSP	Inbred line
PI 308796 PSP	Inbred line
PI 314794 PSP	Inbred line
PI 314800	Inbred line
PI 314803	Inbred line
PI 324695 PSP	Inbred line
PI 324699	Inbred line
PI 324702 PSP	Inbred line
PI 324703 PSP	Inbred line
PI 324706 PSP	Inbred line
PI 331413 PSP	Inbred line
PI 331414 PSP	Inbred line
PI 340126	Inbred line
PI 343263	Inbred line
PI 343267	Inbred line

Table A2. List of USDA pea accessions under Fargo 2022 (FAR22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 343268	Inbred line
PI 343278	Inbred line
PI 343284	Inbred line
PI 343286	Inbred line
PI 343292 PSP	Inbred line
PI 343296	Inbred line
PI 343298	Inbred line
PI 343312	Inbred line
PI 343321 PSP	Inbred line
PI 343331 PSP	Inbred line
PI 343936	Inbred line
PI 344003 PSP	Inbred line
PI 347295 PSP	Inbred line
PI 347457 PSP	Inbred line
PI 347477 PSP	Inbred line
PI 347496 PSP	Inbred line
PI 356980 PSP	Inbred line
PI 356984 PSP	Inbred line
PI 356986 PSP	Inbred line
PI 356991 PSP	Inbred line
PI 356992 PSP	Inbred line
PI 358300 PSP	Inbred line
PI 358620 PSP	Inbred line
PI 358633 PSP	Inbred line
PI 378157 PSP	Inbred line
PI 381334 PSP	Inbred line
PI 393490 PSP	Inbred line
PI 404225 PSP	Inbred line
PI 409031 PSP	Inbred line
PI 413678 PSP	Inbred line
PI 429839 PSP	Inbred line
PI 429845 PSP	Inbred line
PI 476413 PSP	Inbred line
PI 477371 PSP	Inbred line

Table A2. List of USDA pea accessions under Fargo 2022 (FAR22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 486131 PSP	Inbred line
PI 499982 PSP	Inbred line
PI 505108 PSP	Inbred line
PI 505127 PSP	Inbred line
PI 594358 PSP	Inbred line
PI 638516 PSP	Inbred line
PI 639976 PSP	Inbred line
PI 639977 PSP	Inbred line
PI 639978 PSP	Inbred line
PI 639979 PSP	Inbred line
PI 639980 PSP	Inbred line
PI 639981 PSP	Inbred line
W6 12723 PSP	Advanced breeding line
W6 12739 PSP	Advanced breeding line
W6 17293 PSP	Advanced breeding line
W6 26157 PSP	Advanced breeding line
W6 26160 PSP	Advanced breeding line

Table A3. List of USDA pea accessions under Washington State University 2022 (WSU22) experiment.

GENOTYPE	DESCRIPTION
PI 102888 PSP	Inbred line
PI 116056 PSP	Inbred line
PI 116844	Inbred line
PI 116944 PSP	Inbred line
PI 117998 PSP	Inbred line
PI 118501 PSP	Inbred line
PI 121352 PSP	Inbred line
PI 124478 PSP	Inbred line
PI 125839 PSP	Inbred line
PI 125840 PSP	Inbred line
PI 134271 PSP	Inbred line
PI 137118	Inbred line
PI 137119 PSP	Inbred line
PI 137120	Inbred line
PI 138945	Inbred line
PI 140295	Inbred line
PI 140296	Inbred line
PI 140298 PSP	Inbred line
PI 142774	Inbred line
PI 142777	Inbred line
PI 143485 PSP	Inbred line
PI 155109 PSP	Inbred line
PI 156720 PSP	Inbred line
PI 162909 PSP	Inbred line
PI 163125	Inbred line
PI 163126 PSP	Inbred line
PI 163127	Inbred line
PI 163129 PSP	Inbred line
PI 164346	Inbred line
PI 164417	Inbred line
PI 164548 PSP	Inbred line
PI 164612 PSP	Inbred line
PI 164614	Inbred line
PI 164669	Inbred line

Table A3. List of USDA pea accessions under Washington State University 2022 (WSU22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 164779 PSP	Inbred line
PI 164836	Inbred line
PI 164838	Inbred line
PI 164971 PSP	Inbred line
PI 166142	Inbred line
PI 166159 PSP	Inbred line
PI 167250	Inbred line
PI 167253	Inbred line
PI 169603 PSP	Inbred line
PI 171810 PSP	Inbred line
PI 171814	Inbred line
PI 173930	Inbred line
PI 174321	Inbred line
PI 174921 PSP	Inbred line
PI 175231 PSP	Inbred line
PI 179019	Inbred line
PI 179449	Inbred line
PI 179450 PSP	Inbred line
PI 179451 PSP	Inbred line
PI 179459 PSP	Inbred line
PI 179722 PSP	Inbred line
PI 180329 PSP	Inbred line
PI 180702 PSP	Inbred line
PI 181800	Inbred line
PI 183467 PSP	Inbred line
PI 184130 PSP	Inbred line
PI 184784 PSP	Inbred line
PI 189171	Inbred line
PI 193590 PSP	Inbred line
PI 193836	Inbred line
PI 193837	Inbred line
PI 193838	Inbred line
PI 194339	Inbred line
PI 194340	Inbred line

Table A3. List of USDA pea accessions under Washington State University 2022 (WSU22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 194349	Inbred line
PI 195631 PSP	Inbred line
PI 196017	Inbred line
PI 196026	Inbred line
PI 196027	Inbred line
PI 196031	Inbred line
PI 197044 PSP	Inbred line
PI 198074 PSP	Inbred line
PI 198735 PSP	Inbred line
PI 201390 PSP	Inbred line
PI 203067 PSP	Inbred line
PI 207508 PSP	Inbred line
PI 209507 PSP	Inbred line
PI 210558 PSP	Inbred line
PI 210569 PSP	Inbred line
PI 210571 PSP	Inbred line
PI 212031 PSP	Inbred line
PI 212112	Inbred line
PI 220174 PSP	Inbred line
PI 220175	Inbred line
PI 220189 PSP	Inbred line
PI 221697 PSP	Inbred line
PI 222071 PSP	Inbred line
PI 222117 PSP	Inbred line
PI 226561	Inbred line
PI 226562	Inbred line
PI 227258 PSP	Inbred line
PI 236492 PSP	Inbred line
PI 240516 PSP	Inbred line
PI 241593 PSP	Inbred line
PI 242027 PSP	Inbred line
PI 244093 PSP	Inbred line
PI 244129	Inbred line
PI 244150 PSP	Inbred line
PI 194349	Inbred line

Table A3. List of USDA pea accessions under Washington State University 2022 (WSU22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 244191 PSP	Inbred line
PI 244262	Inbred line
PI 249644	Inbred line
PI 249645 PSP	Inbred line
PI 250438 PSP	Inbred line
PI 250440 PSP	Inbred line
PI 250447 PSP	Inbred line
PI 250448 PSP	Inbred line
PI 253968 PSP	Inbred line
PI 257244 PSP	Inbred line
PI 257592 PSP	Inbred line
PI 261623 PSP	Inbred line
PI 261666	Inbred line
PI 263011	Inbred line
PI 263030 PSP	Inbred line
PI 263032 PSP	Inbred line
PI 266070 PSP	Inbred line
PI 269543 PSP	Inbred line
PI 269761 PSP	Inbred line
PI 269762 PSP	Inbred line
PI 269774	Inbred line
PI 269775	Inbred line
PI 269777 PSP	Inbred line
PI 269802 PSP	Inbred line
PI 269804 PSP	Inbred line
PI 269818 PSP	Inbred line
PI 271035 PSP	Inbred line
PI 271511 PSP	Inbred line
PI 272148 PSP	Inbred line
PI 272161	Inbred line
PI 272184 PSP	Inbred line
PI 272194 PSP	Inbred line
PI 272204 PSP	Inbred line
PI 272215 PSP	Inbred line

Table A3. List of USDA pea accessions under Washington State University 2022 (WSU22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 272216 PSP	Inbred line
PI 273676	Inbred line
PI 274307 PSP	Inbred line
PI 274308 PSP	Inbred line
PI 274584 PSP	Inbred line
PI 275821 PSP	Inbred line
PI 277851	Inbred line
PI 277852 PSP	Inbred line
PI 279823 PSP	Inbred line
PI 280607	Inbred line
PI 280611 PSP	Inbred line
PI 280617 PSP	Inbred line
PI 280621	Inbred line
PI 285710 PSP	Inbred line
PI 285727 PSP	Inbred line
PI 285739	Inbred line
PI 286607 PSP	Inbred line
PI 299023	Inbred line
PI 306590	Inbred line
PI 307666 PSP	Inbred line
PI 308796 PSP	Inbred line
PI 314794 PSP	Inbred line
PI 314803	Inbred line
PI 320972 PSP	Inbred line
PI 324697 PSP	Inbred line
PI 324702 PSP	Inbred line
PI 324703 PSP	Inbred line
PI 331413 PSP	Inbred line
PI 331414 PSP	Inbred line
PI 340126	Inbred line
PI 340128 PSP	Inbred line
PI 343263	Inbred line
PI 343267	Inbred line
PI 343268	Inbred line

Table A3. List of USDA pea accessions under Washington State University 2022 (WSU22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 343277	Inbred line
PI 343278	Inbred line
PI 343286	Inbred line
PI 343292 PSP	Inbred line
PI 343295	Inbred line
PI 343298	Inbred line
PI 343312	Inbred line
PI 343331 PSP	Inbred line
PI 343824 PSP	Inbred line
PI 343936	Inbred line
PI 343987 PSP	Inbred line
PI 347281 PSP	Inbred line
PI 347295 PSP	Inbred line
PI 347329 PSP	Inbred line
PI 347337	Inbred line
PI 347477 PSP	Inbred line
PI 347496 PSP	Inbred line
PI 356980 PSP	Inbred line
PI 356984 PSP	Inbred line
PI 356986 PSP	Inbred line
PI 356991 PSP	Inbred line
PI 356992 PSP	Inbred line
PI 358300 PSP	Inbred line
PI 358620 PSP	Inbred line
PI 358633 PSP	Inbred line
PI 371796 PSP	Inbred line
PI 378157 PSP	Inbred line
PI 381334 PSP	Inbred line
PI 393488 PSP	Inbred line
PI 393489 PSP	Inbred line
PI 393490 PSP	Inbred line
PI 404225 PSP	Inbred line
PI 409031 PSP	Inbred line
PI 429839 PSP	Inbred line

Table A3. List of USDA pea accessions under Washington State University 2022 (WSU22) experiment (continued).

GENOTYPE	DESCRIPTION
PI 429843 PSP	Inbred line
PI 429849 PSP	Inbred line
PI 476409 PSP	Inbred line
PI 476413 PSP	Inbred line
PI 477371 PSP	Inbred line
PI 486131 PSP	Inbred line
PI 499982 PSP	Inbred line
PI 505062 PSP	Inbred line
PI 505080 PSP	Inbred line
PI 505108 PSP	Inbred line
PI 505127 PSP	Inbred line
PI 619079 PSP	Inbred line
PI 639978 PSP	Inbred line
PI 639981 PSP	Inbred line
W6 12723 PSP	Advanced breeding line
W6 17293 PSP	Advanced breeding line
W6 26157 PSP	Advanced breeding line
W6 26160 PSP	Advanced breeding line
W6 39762 PSP	Advanced breeding line

Table A4. List of overlapping USDA pea accessions across six environments.

GENOTYPE	DESCRIPTION
DS Admiral	Check variety
Hampton	Check variety
PI 102887	Inbred line
PI 102888 PSP	Inbred line
PI 116056 PSP	Inbred line
PI 116844	Inbred line
PI 116944 PSP	Inbred line
PI 117264 PSP	Inbred line
PI 117998 PSP	Inbred line
PI 118501 PSP	Inbred line
PI 121352 PSP	Inbred line
PI 123246	Inbred line
PI 124478 PSP	Inbred line
PI 125839 PSP	Inbred line
PI 125840 PSP	Inbred line
PI 134271 PSP	Inbred line
PI 137118	Inbred line
PI 137119 PSP	Inbred line
PI 137120	Inbred line
PI 138945	Inbred line
PI 140295	Inbred line
PI 140296	Inbred line
PI 140298 PSP	Inbred line
PI 142774	Inbred line
PI 142775 PSP	Inbred line
PI 142777	Inbred line
PI 143485 PSP	Inbred line
PI 155109 PSP	Inbred line
PI 156720 PSP	Inbred line
PI 162909 PSP	Inbred line
PI 163125	Inbred line
PI 163126 PSP	Inbred line
PI 163127	Inbred line

Table A4. Overlapping genotypes across six environments (continued).

GENOTYPE	DESCRIPTION
PI 163129 PSP	Inbred line
PI 164285	Inbred line
PI 164346	Inbred line
PI 164396	Inbred line
PI 164417	Inbred line
PI 164548 PSP	Inbred line
PI 164612 PSP	Inbred line
PI 164614	Inbred line
PI 164669	Inbred line
PI 164779 PSP	Inbred line
PI 164836	Inbred line
PI 164838	Inbred line
PI 164971 PSP	Inbred line
PI 164972 PSP	Inbred line
PI 165949 PSP	Inbred line
PI 166084 PSP	Inbred line
PI 166142	Inbred line
PI 166159 PSP	Inbred line
PI 166187	Inbred line
PI 167250	Inbred line
PI 167253	Inbred line
PI 169603 PSP	Inbred line
PI 171810 PSP	Inbred line
PI 171814	Inbred line
PI 173930	Inbred line
PI 174321	Inbred line
PI 174921 PSP	Inbred lines
PI 174922	Inbred line
PI 174925	Inbred line
PI 175228	Inbred line
PI 175231 PSP	Inbred line
PI 175232	Inbred line
PI 179019	Inbred line
PI 179449	Inbred line

Table A4. Overlapping genotypes across six environments (continued).

GENOTYPE	DESCRIPTION
PI 179450 PSP	Inbred line
PI 179451 PSP	Inbred line
PI 179459 PSP	Inbred line
PI 179722 PSP	Inbred line
PI 180329 PSP	Inbred line
PI 180693 PSP	Inbred line
PI 180702 PSP	Inbred line
PI 181800	Inbred line
PI 183467 PSP	Inbred line
PI 184130 PSP	Inbred line
PI 184784 PSP	Inbred line
PI 189171	Inbred line
PI 193578 PSP	Inbred line
PI 193586	Inbred line
PI 193588	Inbred line
PI 193590 PSP	Inbred line
PI 193836	Inbred line
PI 193837	Inbred line
PI 193838	Inbred line
PI 194339	Inbred line
PI 194340	Inbred line
PI 194349	Inbred line
PI 195020 PSP	Inbred line
PI 195404 PSP	Inbred line
PI 195631 PSP	Inbred line
PI 196017	Inbred line
PI 196026	Inbred line
PI 196027	Inbred line
PI 196031	Inbred line
PI 197044 PSP	Inbred line
PI 197990 PSP	Inbred line
PI 198072 PSP	Inbred line
PI 198074 PSP	Inbred line
PI 198735 PSP	Inbred line

Table A4. Overlapping genotypes across six environments (continued).

GENOTYPE	DESCRIPTION
PI 201390 PSP	Inbred line
PI 203067 PSP	Inbred line
PI 203069 PSP	Inbred line
PI 204306 PSP	Inbred line
PI 206006 PSP	Inbred line
PI 207508 PSP	Inbred line
PI 209507 PSP	Inbred line
PI 210558 PSP	Inbred line
PI 210569 PSP	Inbred line
PI 210571 PSP	Inbred line
PI 212031 PSP	Inbred line
PI 212112	Inbred line
PI 220174 PSP	Inbred line
PI 220175	Inbred line
PI 220189 PSP	Inbred line
PI 221697 PSP	Inbred line
PI 222071 PSP	Inbred line
PI 222117 PSP	Inbred line
PI 223527 PSP	Inbred line
PI 226561	Inbred line
PI 226562	Inbred line
PI 227258 PSP	Inbred line
PI 227457	Inbred line
PI 236492 PSP	Inbred line
PI 240516 PSP	Inbred line
PI 241593 PSP	Inbred line
PI 242027 PSP	Inbred line
PI 244093 PSP	Inbred line
PI 244121 PSP	Inbred line
PI 244129	Inbred line
PI 244150 PSP	Inbred line
PI 244191 PSP	Inbred line
PI 244262	Inbred line
PI 248181 PSP	Inbred line

Table A4. Overlapping genotypes across six environments (continued).

GENOTYPE	DESCRIPTION
PI 249644	Inbred line
PI 249645 PSP	Inbred line
PI 249646	Inbred line
PI 250438 PSP	Inbred line
PI 250440 PSP	Inbred line
PI 250447 PSP	Inbred line
PI 250448 PSP	Inbred line
PI 253968 PSP	Inbred line
PI 257244 PSP	Inbred line
PI 257592 PSP	Inbred line
PI 261623 PSP	Inbred line
PI 261666	Inbred line
PI 261677 PSP	Inbred line
PI 263011	Inbred line
PI 263014 PSP	Inbred line
PI 263027 PSP	Inbred line
PI 263030 PSP	Inbred line
PI 263031 PSP	Inbred line
PI 263032 PSP	Inbred line
PI 266070 PSP	Inbred line
PI 269543 PSP	Inbred line
PI 269761 PSP	Inbred line
PI 269762 PSP	Inbred line
PI 269763	Inbred line
PI 269771	Inbred line
PI 269774	Inbred line
PI 269775	Inbred line
PI 269776	Inbred line
PI 269777 PSP	Inbred line
PI 269802 PSP	Inbred line
PI 269804 PSP	Inbred line
PI 269818 PSP	Inbred line
PI 269825 PSP	Inbred line
PI 270536 PSP	Inbred line

Table A4. Overlapping genotypes across six environments (continued).

GENOTYPE	DESCRIPTION
PI 271035 PSP	Inbred line
PI 271116 PSP	Inbred line
PI 271511 PSP	Inbred line
PI 272148 PSP	Inbred line
PI 272161	Inbred line
PI 272171 PSP	Inbred line
PI 272184 PSP	Inbred line
PI 272194 PSP	Inbred line
PI 272204 PSP	Inbred line
PI 272215 PSP	Inbred line
PI 272216 PSP	Inbred line
PI 272218 PSP	Inbred line
PI 273605 PSP	Inbred line
PI 273676	Inbred line
PI 274307 PSP	Inbred line
PI 274308 PSP	Inbred line
PI 274584 PSP	Inbred line
PI 275821 PSP	Inbred line
PI 277851	Inbred line
PI 277852 PSP	Inbred line
PI 279823 PSP	Inbred line
PI 280252 PSP	Inbred line
PI 280607	Inbred line
PI 280609 PSP	Inbred line
PI 280611 PSP	Inbred line
PI 280613 PSP	Inbred line
PI 280617 PSP	Inbred line
PI 280619 PSP	Inbred line
PI 280621	Inbred line
PI 285708	Inbred line
PI 285710 PSP	Inbred line
PI 285718 PSP	Inbred line
PI 285727 PSP	Inbred line
PI 285739	Inbred line

Table A4. Overlapping genotypes across six environments (continued).

GENOTYPE	DESCRIPTION
PI 286430 PSP	Inbred line
PI 286607 PSP	Inbred line
PI 299023	Inbred line
PI 306590	Inbred line
PI 306591 PSP	Inbred line
PI 307666 PSP	Inbred line
PI 308796 PSP	Inbred line
PI 311112	Inbred line
PI 314794 PSP	Inbred line
PI 314800	Inbred line
PI 314803	Inbred line
PI 319374 PSP	Inbred line
PI 320972 PSP	Inbred line
PI 324695 PSP	Inbred line
PI 324697 PSP	Inbred line
PI 324699	Inbred line
PI 324702 PSP	Inbred line
PI 324703 PSP	Inbred line
PI 324706 PSP	Inbred line
PI 331413 PSP	Inbred line
PI 331414 PSP	Inbred line
PI 340126	Inbred line
PI 340128 PSP	Inbred line
PI 340130 PSP	Inbred line
PI 343263	Inbred line
PI 343267	Inbred line
PI 343268	Inbred line
PI 343277	Inbred line
PI 343278	Inbred line
PI 343284	Inbred line
PI 343286	Inbred line
PI 343292 PSP	Inbred line
PI 343295	Inbred line
PI 343296	Inbred line

Table A4. Overlapping genotypes across six environments (continued).

GENOTYPE	DESCRIPTION
PI 343298	Inbred line
PI 343312	Inbred line
PI 343321 PSP	Inbred line
PI 343331 PSP	Inbred line
PI 343824 PSP	Inbred line
PI 343936	Inbred line
PI 343958 PSP	Inbred line
PI 343987 PSP	Inbred line
PI 344003 PSP	Inbred line
PI 347281 PSP	Inbred line
PI 347290	Inbred line
PI 347295 PSP	Inbred line
PI 347329 PSP	Inbred line
PI 347337	Inbred line
PI 347457 PSP	Inbred line
PI 347477 PSP	Inbred line
PI 347496 PSP	Inbred line
PI 356980 PSP	Inbred line
PI 356984 PSP	Inbred line
PI 356986 PSP	Inbred line
PI 356991 PSP	Inbred line
PI 356992 PSP	Inbred line
PI 358300 PSP	Inbred line
PI 358620 PSP	Inbred line
PI 358633 PSP	Inbred line
PI 371796 PSP	Inbred line
PI 378157 PSP	Inbred line
PI 381334 PSP	Inbred line
PI 393488 PSP	Inbred line
PI 393489 PSP	Inbred line
PI 393490 PSP	Inbred line
PI 404225 PSP	Inbred line
PI 409031 PSP	Inbred line
PI 413678 PSP	Inbred line

Table A4. Overlapping genotypes across six environments (continued).

GENOTYPE	DESCRIPTION
PI 413688 PSP	Inbred line
PI 429839 PSP	Inbred line
PI 429843 PSP	Inbred line
PI 429845 PSP	Inbred line
PI 429849 PSP	Inbred line
PI 476409 PSP	Inbred line
PI 476413 PSP	Inbred line
PI 477371 PSP	Inbred line
PI 486131 PSP	Inbred line
PI 494077 PSP	Inbred line
PI 499982 PSP	Inbred line
PI 505062 PSP	Inbred line
PI 505080 PSP	Inbred line
PI 505108 PSP	Inbred line
PI 505122 PSP	Inbred line
PI 505127 PSP	Inbred line
PI 594358 PSP	Inbred line
PI 619079 PSP	Inbred line
PI 638516 PSP	Inbred line
PI 639976 PSP	Inbred line
PI 639977 PSP	Inbred line
PI 639978 PSP	Inbred line
PI 639979 PSP	Inbred line
PI 639980 PSP	Inbred line
PI 639981 PSP	Inbred line
W6 12723 PSP	Advanced breeding line
W6 12739 PSP	Advanced breeding line
W6 17293 PSP	Advanced breeding line
W6 26157 PSP	Advanced breeding line
W6 26160 PSP	Advanced breeding line
W6 39762 PSP	Advanced breeding line

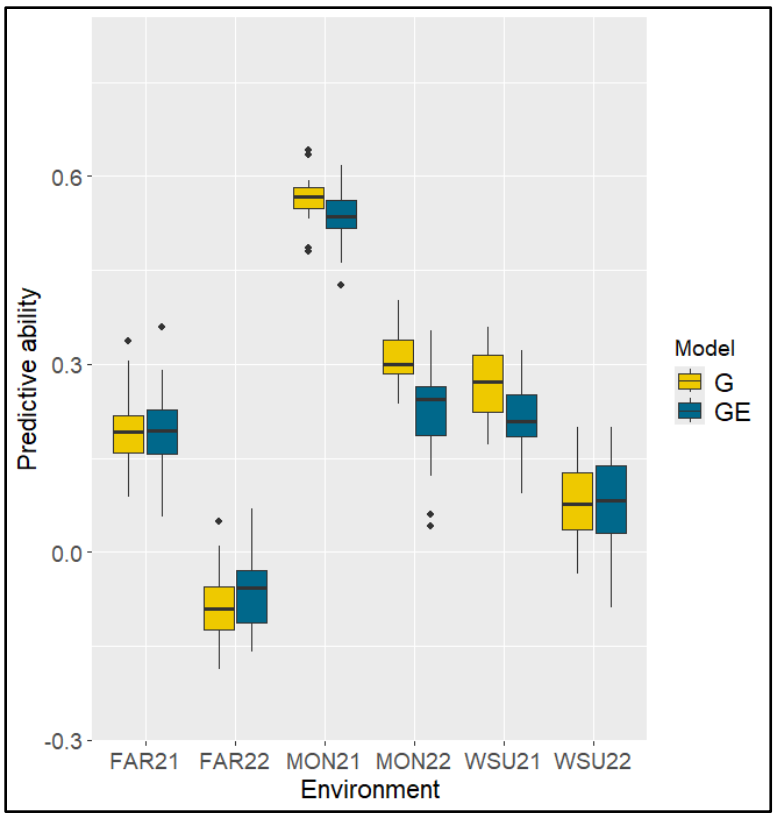

Figure A1. Mean distribution of predictive ability for yield across environments with split-environment prediction, RKHS is Reproducing Kernel Hilbert Spaces, G is prediction model considering genotypic factor, GE is prediction model integrating GxE interaction.

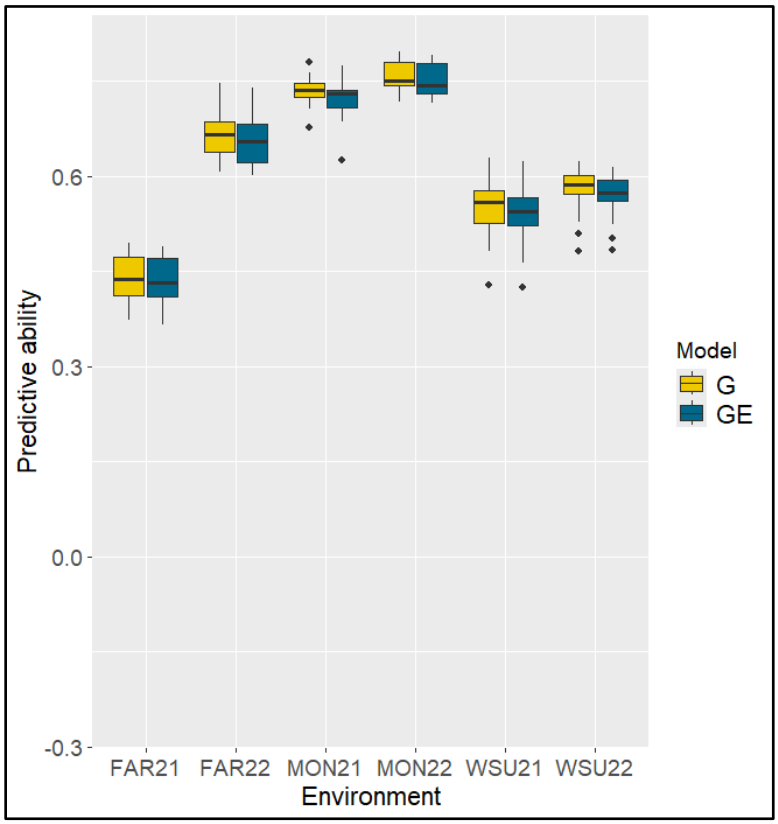

Figure A2. Mean distribution of predictive ability for seed protein content across environments with split-environment prediction, RKHS is Reproducing Kernel Hilbert Spaces, G is prediction model considering genotypic factor, GE is prediction model integrating GxE interaction.