Core ideas

- Phenotypic selection in PYT is challenged by limited seeds, resulting to few replications and environments.
- MTME-GP offers opportunity for enhancing prediction accuracy of multi-trait and diverse environments in PYT.
- MTME-GP enhances prediction by up to 2.5-fold, especially with numerous overlapping genotypes in various tested environments.
- RKHS MTME-GP models, excels in low-heritability, negatively correlated traits, like drought-affected conditions.

Multi-trait multi-environment genomic prediction of preliminary yield trials in pulse crops

Rica Amor Saludares[1], Sikiru Adeniyi Atanda[1], Lisa Piche[1], Hannah Worral[2], Francoise Dariva[1], Kevin McPhee[3], and Nonoy Bandillo[1*]

Affiliations:
[1]Department of Plant Sciences, North Dakota State University, Fargo, ND 58108-6050, USA

[2] North Central Research Extension Center, 5400 Hwy. 83 S, Minot, ND 58701, USA

[3] Department of Plant Science and Plant Pathology, Montana State University, 119 Plant BioScience Building, Bozeman, MT 59717-3150, US

*corresponding author: nonoy.bandillo@ndsu.edu

ORCID

Rica Amor Saludares: 0009-0007-4151-2837

Sikiru Adeniyi Atanda: 0000-0001-8758-2063

Lisa Piche: 0009-0006-2970-9648

Hannah Worral: 0000-0003-3832-5815

Kevin McPhee: 0000-0003-3103-7714

Nonoy Bandillo: 0000-0002-5941-9047

Abbreviations: BLUE, Best linear unbiased estimate; BRR, Bayesian GBLUP model; CAR22, Carrington 2022 dataset; G_RKHS, RKHS model excluding GxE interaction; GE_RKHS, RKHS model accounting for GxE interaction; GEBV, Genomic estimated breeding value; GP, Genomic prediction; GS- Genomic selection; GxE, genotpe-by-environment interaction; MAGIC, multi-parent advanced generation inter-cross; MOT20, Minot 2020 dataset; MOT21, Minot 2021 dataset; MT, Multivariate or Multi-trait; MTME, Multi-trait multi-environment; MTME-GP, Multi-trait multi-environment genomic prediction; NCREC, North Central Research Extension Center; NDSU, North Dakota State University; NIR, Near infrared spectroscopy; PA, Predictive ability; PYT, Preliminary yield trial; RKHS, Reproducing Kernel Hilbert Spaces model; UNI-Univariate

ABSTRACT

Phenotypic selection in preliminary yield trials (PYT) is challenged by limited seeds, resulting in trials with few replications and environments. The emergence of multi-trait multi-environment enabled genomic prediction (MTME-GP) offers opportunity for enhancing prediction accuracy and genetic gain across multiple traits and diverse environments. Using a set of 300 advanced breeding lines in the North Dakota State University (NDSU) pulse crop breeding program, we assessed the efficiency of a MTME-GP model for improving seed yield and protein content in field peas in stress and non-stress environments. MTME-GP significantly improved predictive ability, improving up to 2.5-fold, particularly when a significant number of genotypes overlapped across environments. Heritability of the training environments contributed

significantly to the overall prediction of the model. Average predictive ability ranged from 3 to 7-folds when environments with low heritability were excluded from the training set. Overall, the Reproducing Kernel Hilbert Spaces (RKHS) model consistently resulted in improved predictive ability across all breeding scenarios considered in our study. Our results lay the groundwork for further exploration, including integration of diverse traits, incorporation of deep learning techniques, and the utilization of multi-omics data in predictive modeling.

1.0 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. This places 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 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. 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., 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. Studies have also shown that the integration of genotype by environment (GxE) in the MT model further improves prediction accuracy (Gill et al., 2021; Sandhu et al., 2022).

In this study, 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 within- and across-environments using multiple years of data.

2.0 MATERIALS AND METHODS

2.1 Germplasm and phenotyping

The genetic materials consisted of 282 NDSU advanced elite breeding lines previously described in Bari et al. (2022)**.** 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 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. (2022), 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 $\geq$ 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).

2.3 Phenotyping

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

$$\mathbf{y} = \mathrm{f}(\mathbf{r}, \mathbf{c}) + \mathbf{Xb} + \mathbf{Z}_r\mathbf{u}_r + \mathbf{Z}_c\mathbf{u}_c + \boldsymbol{\varepsilon} \quad (1)$$

where $\mathbf{y}$ is the response variable for n-th phenotype, $\mathbf{b}$ 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, $\mathbf{I}$ 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, $\mathbf{I}$ is the same as above and σ_ε^2 is variance for the residual term or simply referred to as nugget. $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices for the fixed and random terms, respectively.

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):

$$\mathbf{y} = \mathbf{1_k}\mu + \mathbf{Zu} + \boldsymbol{\varepsilon} \quad (2)$$

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):

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{Z}_3\mathbf{u}_3 + \boldsymbol{\varepsilon} \quad (3)$$

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, $\sigma_k^2\mathbf{Z}_k\mathbf{K}\mathbf{Z}_\mathbf{k}'$) 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 $\mathbf{Z}_k\mathbf{K}\mathbf{Z}_\mathbf{k}'$ 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. $\mathbf{u}_\mathbf{2}$is the random effect of the pea lines and follows the multivariate normal distribution N(0, $\sigma_g^2\mathbf{Z}_\mathbf{g}\mathbf{G}\mathbf{Z}_\mathbf{g}'$), 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 ($\mathbf{G}$). $\mathbf{u}_3$is the random effect of the GxE effect and follows the multivariate normal distribution N(0, $\sigma_{gk}^2\mathbf{Z}_\mathbf{g}\mathbf{G}\mathbf{Z}_\mathbf{g}'\#\sigma_k^2\mathbf{Z}_k\mathbf{K}\mathbf{Z}_\mathbf{k}'$), where σ_{gk}^2 is the variance component of GE, # denotes the Hadamard product, and $\mathbf{Z}_\mathbf{g}\mathbf{G}\mathbf{Z}_\mathbf{g}'$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 $N(0, \sigma_\varepsilon^2\mathbf{I})$, where σ_ε^2 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 as follows:

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{1}_1\mu_1 \\ \vdots \\ \mathbf{1}_k\mu_n \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Z}_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad (4)$$

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, $\mathbf{u}_1 \ldots \mathbf{u}_n$ is the genomic effect of the lines for each n-th trait, and $\boldsymbol{\varepsilon}_1 \ldots \boldsymbol{\varepsilon}_n$ is the residual error for each n-th trait. The random term is assumed to follow the multivariate normal distribution $[\mathbf{u}_1 \ldots \mathbf{u}_n] \sim \text{MN}[0, (\mathbf{G} \otimes \mathbf{G}_o)]$, 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:

$$\mathbf{G}_o \otimes \mathbf{G} = \begin{bmatrix} \sigma^2_{g_1} & \sigma_{g_{12}} & \cdots & \sigma_{g_{1n}} \\ \sigma_{g_{21}} & \sigma^2_{g_2} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{g_{n1}} & \vdots & \ldots & \sigma^2_{g_n} \end{bmatrix} \otimes \mathbf{G} \qquad (5)$$

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 **I** is the same as above and **R** is a heterogeneous diagonal matrix of the residual variances for each n-th trait:

206 $$\mathbf{R}=\begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\varepsilon_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2_{\varepsilon_n} \end{bmatrix} \otimes \mathbf{I} \qquad (6)$$

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

209

210 For the multi-trait (MT) multi-environment GS model, Eq. 3 was expanded as described by
211 Montesinos et al. (2022):

212 $$\mathbf{y} = \mathbf{1}_{nK}\mu + \mathbf{Z}_{1.1}\mathbf{u}_{1.1} + \mathbf{Z}_{2.1}\mathbf{u}_{2.1} + \mathbf{Z}_{3.1}\mathbf{u}_{3.1} + \boldsymbol{\varepsilon} \quad (7)$$

213

214 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
215 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
216 the random effect of each environment of each trait with size k x n, $\mathbf{Z}_{2.1}$ is the incidence matrix
217 of genotypes of order i × j, $\mathbf{u}_{2.1}$ is the random effect of the genotypes i × n, and follows the
218 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
219 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
220 variance-covariance matrix of traits of order n × n. $\mathbf{Z}_{3.1}$ is the incidence matrix of GE of order i
221 × kj, $\mathbf{u}_{3.1}$ is the random effect of the genotypes by environment by trait of order kj × n and
222 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
223 $\mathbf{U}_{gk}$ is the unstructured variance-covariance matrix of order k by k. $\boldsymbol{\varepsilon}$ is the random term of the
224 residual and follows the multivariate normal distribution MN(0, $\mathbf{I}, \Sigma_t$). $\mathbf{I}$ is identity matrix of order
225 i ×n, and Σ_t is the unstructured variance-covariance matrix.

226

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 BLUE of each trait for the complete dataset. For within-environment predictions, datasets for each environment (MOT20, MOT21, and CAR22) were partitioned into different training set sizes (50%, 60%, 70%, and 80%) and the process was repeated 30 times. For across-environment predictions, models were trained on one environment and tested on a novel environment (e.g., MOT20 trained to predict MOT21). We also explored the effectiveness of training the model on multiple environments to predict a novel environment (e.g., MOT20 and MOT21 trained to predict CAR22).

3.0 RESULTS AND DISCUSSION

3.1 Predictive ability of different genomic prediction models

We evaluated the potential of GS to predict the genetic merit of two negatively-correlated complex traits across three environments with varying degrees of heritability (**Supp. Fig. 1**). Except in the case of UNI-GS model where G_BRR outperformed RKHS for both traits, the RKHS model consistently demonstrated superior predictive performance across models and in cross-validation strategies for both traits (**Supp. Fig.2:5**). The superiority of the RKHS model in all other scenarios evaluated in this study suggests 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 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. MT outperformed UNI by 11-fold across traits and environments (**Supp. Fig 6**:**7**). 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. These highlight the potential of MT models to enhance prediction accuracy, especially for challenging and resource-intensive traits.

The integration of GxE interaction by including environments with low heritability did not result in an increase in predictive ability as shown in Supplementary Figure 7F. Specifically, adding MOT21 to the training set, which has the lowest heritability estimate of 0.11, did not enhance predictive ability. This might suggest a nuanced interplay between environmental conditions and predictive models. This corroborates the findings of Rogers and Holland (2022), emphasizing that the addition of nuisance environments will not enhance overall predictive ability. This suggests that GxE interactions might be more relevant in environments with moderate to high heritability.

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). **Supp. Fig. 6:7** showed no clear trend in predictive ability for the traits with increasing training set size. The G_RKHS (MT) model consistently showed the highest

273 predictive ability for seed protein reaching 30% when 60% and 80% of MOT20 dataset were
274 trained to predict CAR22 (**Fig. 1**). However, in predicting for seed yield, the majority of the highest
275 predictive abilities were still observed under G_RKHS (MT), reaching 34% when 80% of MOT20
276 dataset was trained to test CAR22 (**Supp. Fig. 8**). Previous studies have emphasized a strong
277 relationship between prediction accuracy, training set size, and trait heritability (Luan et al., 2009;
278 Lorenz et al., 2011; Clark et al., 2012; Nyline et al., 2017; Kaler et al.,2022; Atanda et al., 2022).
279 Considering the varying heritability of our traits across environments, ranging from 1.57E-06 to
280 0.80 for grain yield and 0.12 to 0.53 for protein, and the negative correlation between traits, these
281 factors might contribute to the overall predictive ability across models in our study. Contrary to
282 our results, Bari et al. (2021) reported an increase in prediction accuracy with increased training
283 set size. Other studies (Budhlakoti et al. 2022; De Roos et al. 2020) have also reported the influence
284 of training set size and heritability on prediction accuracy. This underscores the importance of
285 careful consideration when selecting training set size for model training.

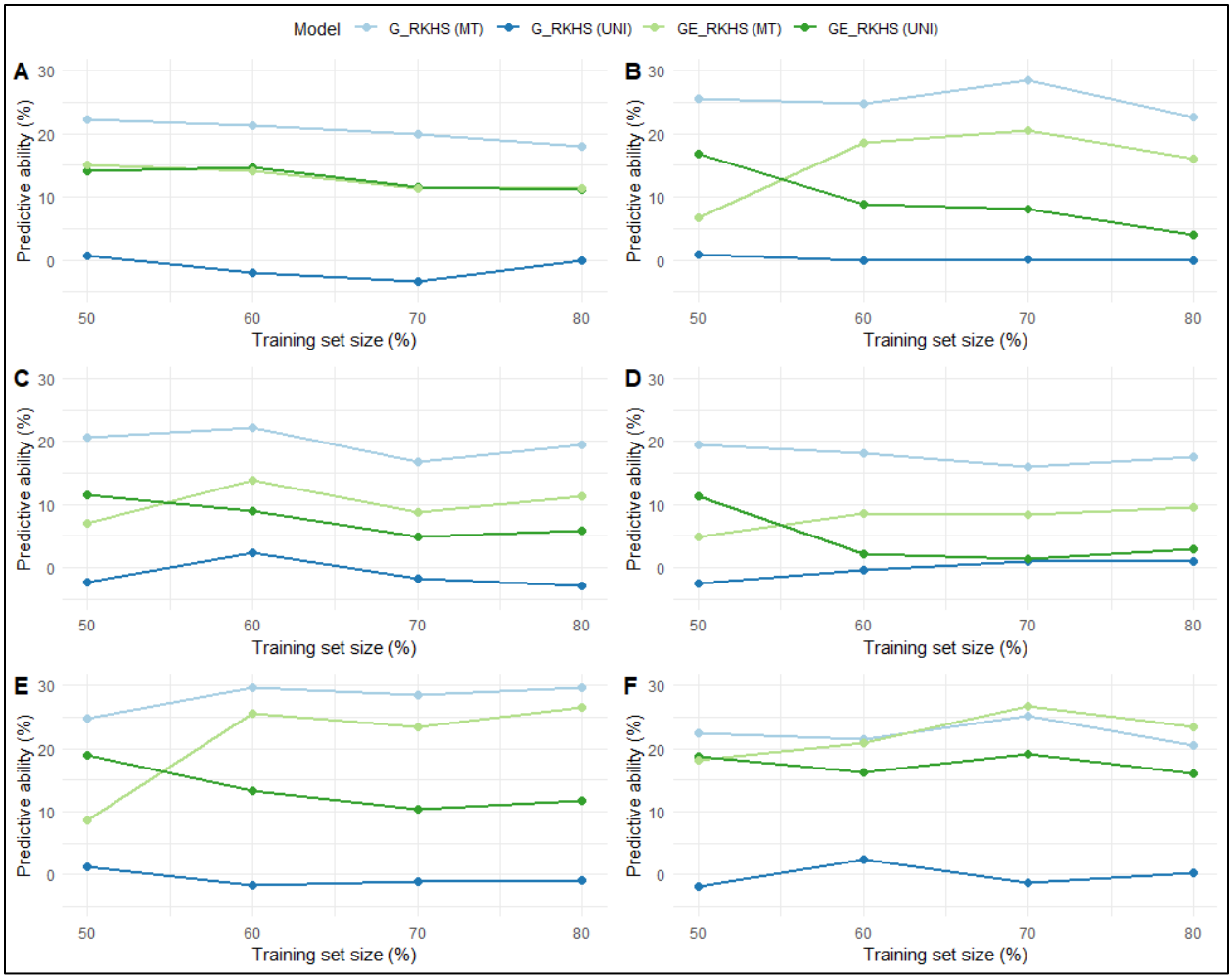
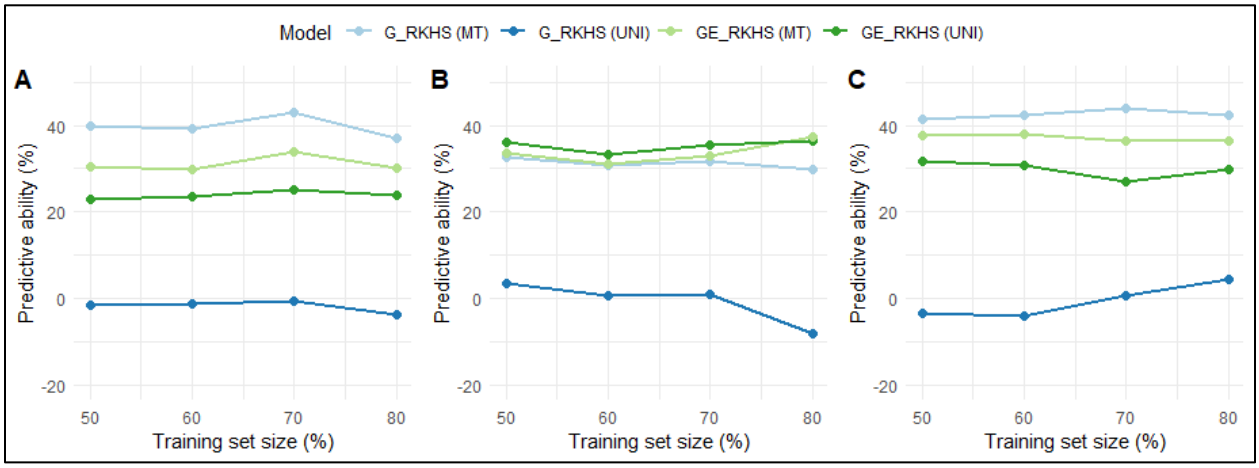


Figure 1: Average predictive ability with increasing training population size using RKHS models for seed protein content, RKHS is Reproducing Kernel Hilbert Spaces, MT is multivariate, UNI is univariate, G is prediction model considering genotype, GE is prediction model integrating GxE interaction. (A) 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.

3.3 Efficacy of MTME-GP for predictions within and across different environments

296 Generally, integration of GxE in the model improved the predictive ability (**Supp. Fig. 2B, 2E,**
297 **4B, 4E**) except for environment with low heritability as reported earlier. This further highlights
298 the importance of carefully managing testing environments to reduce the influence of
299 environmental nuisance on phenotyping. Ultimately, it underscores the significance of considering
300 heritability in the environment when developing training datasets for multi-environment GS
301 models, ensuring efficient capturing of the genetic relationship between environments and
302 borrowing information effectively across environments (Xu, 2016; van Eeuwijk et al. 2019;
303 Atanda et al., 2021). Similarly, Sapkota et al. (2020) reported varying prediction accuracy when
304 environments with different heritability were included in the training model to predict new
305 environments. Gill et al. (2021) emphasized the potential of MTME-GP in practical scenarios,
306 such as overcoming the challenges posed by the loss of complete or partial trials due to extreme
307 weather. As also shown in our study, also MTME-GP proved valuable in predicting the genetic
308 merit of the lines affected by drought condition for both traits (**Fig. 2:3**).
309



310

311

Figure 2: Average predictive abilities with increasing training population set size using RKHS model for seed yield, G is prediction model considering genotype, GE is prediction model integrating GxE interaction, MT is multivariate, UNI is univariate. (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.

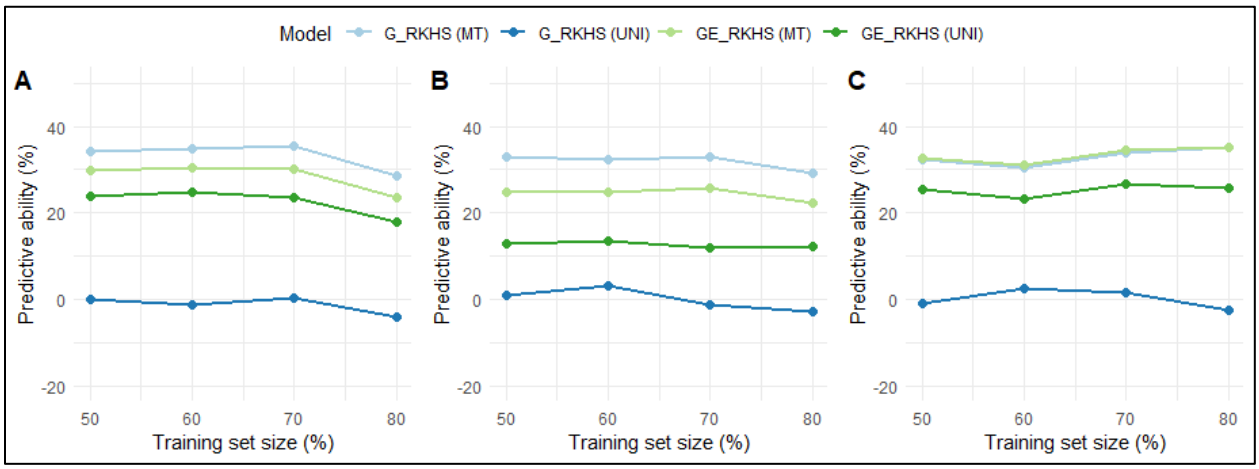


Figure 3: Average predictive abilities with increasing training set size using RKHS model for seed protein content, G is prediction model considering genotype, GE is prediction model integrating GxE interaction, MT is multivariate, UNI is univariate. (A) MOT21 and CAR22 datasets trained to predict MOT20, (B) MOT20 and CAR22 datasets trained to predict MOT21, (C) MOT20 and MOT21 datasets trained to predict CAR22.

CONCLUSION

Our research findings 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, 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 trait heritabilities such as grain 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.

ACKNOWLEDGMENTS

The authors express gratitude for the funding for this research from the North Dakota Department of Agriculture through the Specialty Crop Block Grant Program (19-429). Additionally, appreciation is extended to Bandillo Lab members for their assistance in conducting field research and collecting data.

CONFLICT OF INTEREST

The authors declare that the study was carried out without any existing commercial or financial associations that could be interpreted as posing a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The SNP dataset utilized in this study is accessible through: https://www.ncbi.nlm.nih.gov/sra/PRJNA730349. For access to the phenotype data, please contact the corresponding author.

REFERENCES

Achakzai, A.K.K. & Bangulzai, M.I. (2006). Effect of various levels of nitrogen fertilizer on the yield and yield attributes of pea (*Pisum sativum* L.) cultivars. *Pak. J. Bot*. 38: 331-340.

Ahmar, S., Gill, R.A., Jung, K., Faheem, A., Qasim, M.U., Mubeen, M., & Zhou, W. (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., Nazzicari, N., Pecetti, L., Romani, M., & Russi, L. (2019). Pea genomic selection for Italian environments. *BMC Genomics*, *20*(1), 603.

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

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

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

Atanda, S.A., Olsen, M., Crossa, J., Burgueño, J., Rincent, R., Dzidzienyo, D., Beyene, Y., Gowda, M., Dreher, K., Boddupalli, P.M., Tongoona, P., Danquah, E.Y., Olaoye, G., & Robbins, K.R.

(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., Cao, T-V., Bartholome, J., Ospina, Y., Quintero, C., Frouin, J., Bouvet J-M., & Grenier, C. (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., Jarquin, D., Posadas, L. G., Lorenz, A. J., & Graef, G. L. (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., Zheng, P., Viera, I., Worral, H., Szwiec, S., Ma, Y., Main, D., Coyne, C. J., McGee, R. J., & Bandillo, N. (2021). Harnessing genetic diversity in the USDA pea germplasm collection through genomic prediction. *Frontiers in Genetics*, *12*, 707754.

Bari, M. A., Fonseka, D., Stenger, J., Zitnick-Anderson, K., Atanda, S. A., Worral, H., Piche, L., Kim, J., Morales, M., Johnson, J., Saludares, R. A., Flores, P., Pasche, J., & Bandillo, N. (2022). A greenhouse-based high-throughput phenotyping platform for identification and genetic dissection of resistance to Aphanomyces root rot in field pea. In *bioRxiv* (p. 2022.08.01.502415). https://doi.org/10.1101/2022.08.01.502415.

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., & Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). *Plant Science: An International Journal of Experimental Plant Biology*, *242*, 23–36.

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

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

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

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

Budhlakoti, N., Kushwaha, A. K., Rai, A., Chaturvedi, K. K., Kumar, A., Pradhan, A. K., Kumar, U., Kumar, R. R., Juliana, P., Mishra, D. C., & Kumar, S. (2022). Genomic selection: a tool for accelerating the efficiency of molecular breeding for development of climate-resilient crops. *Frontiers in Genetics*, *13*, 832153.

Burrows, P. M. (1972). Expected selection differentials for directional selection. *Biometrics*, *28*(4), 1091–1100.

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

Clark, S.A., Hickey, J.M., Daetwyler, H.D., & Van der Werf, J.H. (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., Perez-Rodriguez, P., Cuevas, J., Montesinos-Lopez, O., Jarquin, D., de los Campos, G., Burgueño, J., Gonzales-Camacho, J.M., Perez-Elizalde, S., Beyene, Y., Dreisigacker, S.,

Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., & Varshney, R.K. Genomic selection in plant breeding: methods, models, and perspectives. (2017). *Trends in Plant Science*, *22*(11), 961–975.

Cuevas, J., Crossa, J., Soberanis, V., Perez-Elizalde, S., Perez-Rodriguez, P., de los Campos, G., Montesinos-Lopez, O.A., & Burgueño, J. (2016). Genomic prediction of genotype x environment interaction kernel regression models. *The Plant Genome*, 9: plantgenome2016.03.0024.

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

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

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. *PloS One*, *6*(5), e19379.

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

Gaire, R., Arruda, M.P., Mohammadi, M., Brown-Guedira, G., Kolb, F.L., & Rutkoski, J. (2022). Multi-trait genomic selection can increase selection accuracy for deoxynivalenol accumulation resulting from fusarium head blight in wheat. *The Plant Genome*, 15, e20188. https://doi.org/10.1002/tpg2.20188

441 Gill, H. S., Halder, J., Zhang, J., Brar, N. K., Rai, T. S., Hall, C., Bernardo, A., Amand, P. S., Bai,
442 G., Olson, E., Ali, S., Turnipseed, B., & Sehgal, S. K. (2021). Multi-trait multi-environment
443 genomic prediction of agronomic traits in advanced breeding lines of winter wheat. *Frontiers*
444 *in Plant Science*, *12*, 709545.
445 Gosal, S. S., & Wani, S. H. (2020). Accelerated plant breeding, Volume 3: Food Legumes.
446 Springer Nature.
447 Haile, T. A., Heidecker, T., Wright, D., Neupane, S., Ramsay, L., Vandenberg, A., & Bett, K. E.
448 (2020). Genomic selection for lentil breeding: Empirical evidence. *The Plant Genome*, *13*(1),
449 e20002.
450 Hayes, B. J., Panozzo, J., Walker, C. K., Choy, A. L., Kant, S., Wong, D., Tibbits, J., Daetwyler,
451 H. D., Rochfort, S., Hayden, M. J., & Spangenberg, G. C. (2017). Accelerating wheat breeding
452 for end-use quality with multi-trait genomic predictions incorporating near infrared and
453 nuclear magnetic resonance-derived phenotypes. *TAG. Theoretical and Applied Genetics.*
454 *Theoretische Und Angewandte Genetik*, *130*(12), 2505–2519.

455 Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., Piraux, F., Guerreiro,
456 L., Pérez, P., Calus, M., Burgueño, J., & de los Campos, G. (2013). A reaction norm model
457 for genomic selection using high-dimensional genomic and environmental data. *TAG.*
458 *Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *127*(3), 595–607.

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

463 Jia, Y., & Jannink, J.L. (2012). Multiple-trait genome selection methods increase genetic value
464 prediction accuracy. *Genetics*, 192, 1513-1522. https://doi.org/10.1534/genetics.112.144246.
465 Jiang, Y. & Reif, J.C. (2015). Modeling epistasis in genomic selection. *Genetics*. 201, 759-768.
466 https://doi.10.1534/genetics.115.177907.
467 Kaler, A.S., Purcell, L.C., Beissinger, T., & Gillman J.D. (2022). Genomic prediction models for
468 traits differing in heritability for soybean, rice, and maize. *BMC Plant Biol*. 2022 Feb 26;
469 22(1):87. doi. 10.1186/s12870-022-03479-y.
470 Keller, B., Ariza-Suarez, D., de la Hoz, J., Aparicio, J. S., Portilla-Benavides, A. E., Buendia, H.
471 F., Mayor, V. M., Studer, B., & Raatz, B. (2020). Genomic prediction of agronomic traits in
472 common bean (l.) under environmental stress. *Frontiers in Plant Science*, *11*, 1001.

473 Kreplak, J., Madoui, M.-A., Cápal, P., Novák, P., Labadie, K., Aubert, G., Bayer, P. E., Gali, K.
474 K., Syme, R. A., Main, D., Klein, A., Bérard, A., Vrbová, I., Fournier, C., d'Agata, L., Belser,
475 C., Berrabah, W., Toegelová, H., Milec, Z., … Burstin, J. (2019). A reference genome for pea
476 provides insight into legume genome evolution. *Nature Genetics*, *51*(9), 1411–1422.

477 Krishnappa, G., Savadi, S., Tyagi, B. S., Singh, S. K., Mamrutha, H. M., Kumar, S., Mishra, C.
478 N., Khan, H., Gangadhara, K., Uday, G., Singh, G., & Singh, G. P. (2021). Integrated genomic
479 selection for rapid improvement of crops. *Genomics*, *113*(3), 1070–1086.
480 Kumar, L., Chhogyel, N., Gopalakrishnan, T., Hasan, M.K., Jayasinghe, S.L., Kariyawasam, C.S.,
481 Kogo, B.K., & Ratnayake, S. (2022). Chapter 4- Climate change and future of agri-food
482 production. *Academic Press*, 49-79. https://doi.org/10.1016/B978-0-323-91001-9.00009-8.

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

Li, Y., Kaur, S., Pembleton, L. W., Valipour-Kahrood, H., Rosewarne, G. M., & Daetwyler, H. D. (2022). Strategies of preserving genetic diversity while maximizing genetic response from implementing genomic selection in pulse breeding programs. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *135*(6), 1813–1828.

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

Lush, J. L. (2013). *Animal Breeding Plans*. Read Books Ltd. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. (2019). *Plant Science: An International Journal of Experimental Plant Biology*, *282*, 23–39.

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

Mohammed, Y.A., Chen, C., Walia, M.K., Torrion, J.A., McVay, k., Lamb, P., Miller, P., Eckhoff, J., Miller, J., & Khan, Q. (2018). Dry pea (*Pisum sativum* L.) protein, starch, and ash concentrations as affected by cultivar and environment. Canadian Journal of Plant Science. 98(5): 1188-1198. https://doi.org/10.1139/cjps-2017-0338.

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

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

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

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

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

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

Pérez-Rodríguez, P., Crossa, J., Rutkoski, J., Poland, J., Singh, R., Legarra, A., Autrique, E., Campos, G. de L., Burgueño, J., & Dreisigacker, S. (2017). Single-step genomic and pedigree genotype × environment interaction models for predicting wheat lines in international environments. *The Plant Genome*, *10*(2). https://doi.org/10.3835/plantgenome2016.09.0089.

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

528 Poland, J. A., & Rife, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics.
529 *The Plant Genome*, *5*(3). https://doi.org/10.3835/plantgenome2012.05.0005.
530 Pratap, A., Kumar, S., Polowick, P. L., Blair, M. W., & Baum, M. (2022). Editorial: Accelerating
531 genetic gains in pulses. *Frontiers in Plant Science*, *13*, 879377.
532 Punia, S., & Kumar, M. (2022). Functionality and application of colored cereals: nutritional,
533 bioactive, and health aspects. Elsevier.
534 Rogers, A. R., & Holland, J. B. (2022). Environment-specific genomic prediction ability in maize
535 using environmental covariates depends on environmental similarity to training data. *G3* ,
536 *12*(2). https://doi.org/10.1093/g3journal/jkab440.
537 Samantara, K., Bohra, A., Mohapatra, S.R., Prihatini, R., Asibe, F., Singh, L., Reyes, V.P., Tiwari,
538 A., Maurya, A.K., Croser, J.S., Wani, S.H., Siddique, K.H., & Varshney, R. (2022). Breeding
539 more crops in less time: a perspective on speed breeding. *Biology* (*Bassel*). 11(2): 275.
540 Sandhu, K. S., Mihalyov, P. D., Lewien, M. J., Pumphrey, M. O., & Carter, A. H. (2021).
541 Combining genomic and phenomic information for predicting grain protein content and grain
542 yield in spring wheat. *Frontiers in Plant Science*, *12*, 613300.
543 Sandhu, K. S., Patil, S. S., Aoun, M., & Carter, A. H. (2022). Multi-trait multi-environment
544 genomic prediction for end-use quality traits in winter wheat. *Frontiers in Genetics*, *13*,
545 831020.
546 Sapkota, S., Lucas Boatwright, J., Jordan, K., Boyles, R., & Kresovich, S. (2020). Multi-trait
547 regressor stacking increased genomic prediction accuracy of sorghum grain composition.
548 *Agronomy*, 10, 1221. https://doi.org/10.3390/agronomy10091221.

549 Shanthakumar, P., Klepacka, J., Bains, A., Chawla, P., Dhull, S. B., & Najda, A. (2022). The
550 current situation of pea protein and its application in the food industry. *Molecules* , *27*(16).
551 https://doi.org/10.3390/molecules27165354.
552 Tilman, D., Balzer, C., Hill, J., & Befort, B.L. (2011). Global food demand and the sustainable
553 intensification of agriculture. *Proceedings of the National Academy of Sciences of the United*
554 *States of America*. 108(50) 20260-20264. https://doi.org/10.1073/pnas.1116437108.
555 Van Dijk, M., Morley, T., Rau, M.L.M & Saghai, Y. (2021). A meta-analysis of projected global
556 food demand and population at risk of hunger for the period 2010-2050. *Nat Food*, 494-501.
557 https://doi.org/10.1038/s43016-021-00322-9
558 van Eeuwijk, F.A., Bustos-Korts, D., Millet, E.J., Boer, M.P., Kruijer, W., Thompson, A.,
559 Malosetti, M., Itawa, H., Quiroz, R., Kuppe, C., Muller, O., Blazakis, K.N., Yu, K., Tardieu,
560 F., & Chapman, S.C. (2019). *Plant Science*, 282, 23-39.
561 Winn, Z. J., Larkin, D. L., Lozada, D. N., DeWitt, N., Brown-Guedira, G., & Mason, R. E. (2023).
562 Multivariate genomic selection models improve prediction accuracy of agronomic traits in
563 soft red winter wheat. *Crop Science*, *63*(4), 2115–2130.
564 Xu, Y. (2016). Envirotyping for deciphering environmental impacts on crop plants. *TAG.*
565 *Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *129*(4), 653–673.
566 Zhao, H., Lin, Z., Khansefid, M., Tibbits, J. F., & Hayden, M. J. (2023). Genomic prediction and
567 selection response for grain yield in safflower. *Frontiers in Genetics*, *14*, 1129433.