

Article

Near-Infrared Spectroscopy-Based Phenomics Data Can Improve Genomic Prediction of Agronomic and Grain Quality Traits Across Multi-Environment Sorghum Hybrid Trials

Pradip Sapkota ^{1,*} , Jales Fonseca ², Ramasamy Perumal ³ , José Crossa ⁴ and William L. Rooney ¹ ¹ Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843, USA² Research and Development Breeding, Bayer Crop Science-US, Stanton, MN 55018, USA³ Agriculture Research Center, Kansas State University, Hays, KS 67601, USA⁴ Maize and Wheat Improvement Center (CIMMYT), Texcoco de Mora 56237, Mexico

* Correspondence: prdp@tamu.edu

Abstract

In recent years, phenotyping approaches in plant breeding have expanded in both methodology and data collection capacity. One such tool, Near-Infrared Spectroscopy (NIRS) generates a wealth of reflectance values for biological samples. To test the potential of NIRS-based predictions, a hundred grain sorghum hybrids generated from a 10 × 10 factorial mating design were evaluated across eight environments. Hybrids were phenotyped for grain yield, days to anthesis, plant height, kernel hardness index, kernel diameter, and kernel weight. Hybrid grain samples were scanned with NIRS to generate phenomic data while parental lines were genotyped using genotyping by sequencing. Three different predictive models: genomic prediction (GP), phenomic prediction (PP), and GP + PP were fitted. Three different cross-validation schemes of untested hybrids in characterized environments (CV1), tested hybrids in uncharacterized environments (CV2), and untested hybrids in uncharacterized environments (CV3) were completed. GP + PP significantly improved over GP for days to anthesis, kernel hardness index, kernel diameter, and kernel weight for CV1. Prediction accuracy of GP + PP was also significantly improved for the kernel hardness index and kernel weight for CV2 and CV3. Depending on logistics, phenomic prediction has the potential to complement or supplement genomic data for predictive strategies in sorghum.

Keywords: genomic selection; grain quality; near-infrared spectroscopy; phenomic selection; sorghum



Academic Editor: Yin Li

Received: 31 July 2025

Revised: 4 September 2025

Accepted: 9 September 2025

Published: 15 September 2025

Citation: Sapkota, P.; Fonseca, J.; Perumal, R.; Crossa, J.; Rooney, W.L. Near-Infrared Spectroscopy-Based Phenomics Data Can Improve Genomic Prediction of Agronomic and Grain Quality Traits Across Multi-Environment Sorghum Hybrid Trials. *Plants* **2025**, *14*, 2871. <https://doi.org/10.3390/plants14182871>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Classical crop breeding strategies have integrated new prediction methodologies to meet increased productivity demands for food, feed, fiber, and forage [1]. Of these, genomic selection (GS) and novel approaches to phenomics are among the most important new technologies [2]. While these technologies have the greatest impact on the most widely grown and marketed crops, other crops such as sorghum [*Sorghum bicolor* (L.) Moench] could benefit from their use as well. Grain sorghum is an internationally grown grain crop that is used as a food, feed, and industrial purpose [3]. While it has always been a food crop in Africa and India, food-grade sorghum has increased in other cultures due to its low glycemic index and gluten-free properties [4,5].

While improvements in sorghum genetics and agronomic management have increased grain yield, the rate of genetic improvement has slowed in recent years [6,7]. The adoption of newer breeding technologies could improve the rate of genetic gain. To date, a primary limitation to their use in sorghum is the absence of the logistical infrastructure [8].

One of these approaches is GS, which trains a statistical model by integrating phenotypic data and genome-wide molecular markers to predict the performance of untested cultivars [9]. GS has accelerated rates of genetic gain and fits well within most breeding programs [10]. Another approach, high-throughput phenotyping, can measure traits faster and more efficiently than the traditional approaches [11]. One such method, near-infrared spectroscopy (NIRS), is commonly used to predict grain compositional traits without extensive sample preparation and destruction. NIRS-based phenomic data can be collected at a lower cost and easily as compared to genomic data. Predictions are based on reflectance values from different wavelengths that are associated with targeted organic compounds [12,13].

In addition to direct prediction of a specific trait, raw NIRS band reflectance values could also be used for more general predictive modeling purposes [14]. Performing selection for targeted traits based on large numbers of reflectance as variables through predictive modeling is known as phenomic selection (PS) [14,15]. Although the cost of genotyping has dropped, implementation of GS can be challenging for an exponentially larger number of genotypes [16]. In addition, PS is logistically less demanding in comparison to GS; therefore, smaller breeding programs as well as crops with limited genomic resources can successfully adopt it [14]. Combined or independently, GS and PS have the potential to improve selection intensity, accuracy, and ultimately genetic gain improvement in a breeding program, and the successful implementation of these tools has been demonstrated in corn, wheat, and soybean [17,18].

Earlier work in sorghum demonstrated that genomic prediction could estimate breeding values for parental lines and the performance of specific hybrid performances for traits such as grain yield (GY), plant height (PH), and days to anthesis (DA) [19–22]. Given the increased use of sorghum as a food grain, there is increased interest in grain quality improvement in sorghum [23]. Of those, kernel physical traits such as the kernel hardness index (KHI), kernel diameter (KD), and kernel weight (KW) are typically characterized at the level of a single kernel with a single-kernel characterization system (SKCS), manufactured by Perten Instruments AB. KHI represents end-use quality traits such as mold resistance, grain storage ability, and insect resistance [24,25]. KD and KW, which represent grain size and kernel weight, are also directly relevant to end uses dictating human consumption [26]. However, these SKCS instruments are no longer available for commercial purposes, creating a void for phenotyping. To accelerate grain quality improvement in sorghum, alternative methods of phenotyping for these traits must be developed. In this context, widespread adoption of NIRS within breeding programs provides an opportunity to optimize NIR-based phenomic selection approaches to accelerate grain quality improvement in sorghum.

To test the potential of integrating GS and PS in sorghum, a set of 100 sorghum hybrids generated from a 10×10 factorial mating design were evaluated across eight environments in Texas and Kansas dryland locations. The hypothesis of this study is that NIR reflectance data obtained on grain samples can serve as low-cost phenomic data, alone or in combination with genomic data, and could accurately predict agronomic performance and grain quality traits of hybrid sorghums. The objectives of the study were to (1) estimate variance components and heritability for NIRS values by band reflection, (2) build prediction models for agronomic and kernel traits in a single environment, and (3) build multi-environment prediction models by integrating genomic and phenomic data across environments for different prediction scenarios.

2. Results

2.1. Combined Analysis and Variance Components

Significant variation was observed for GY, DA, PH, KHI, KD, and KW among hybrids within all environments. The average best linear unbiased estimate (BLUE) for grain yield combined across eight environments was 6.44 t ha^{-1} with a range from 2.2 to 10.98 t ha^{-1} . The average BLUE across eight environments for DA varied from 54 to 87.5 days, with a mean of 70.8 days, and PH varied from 101.6 to 177.8 cm with an average of 135.4 cm. For kernel physical characteristics, BLUE of KHI ranged from 26.1 to 98.3 with a mean of 75.2, KD ranged from 1.96 to 3.002 mm with a mean of 2.49 mm, and KW ranged from 17.82 to 33.19 mg with a mean of 25.87 mg (Figure 1).

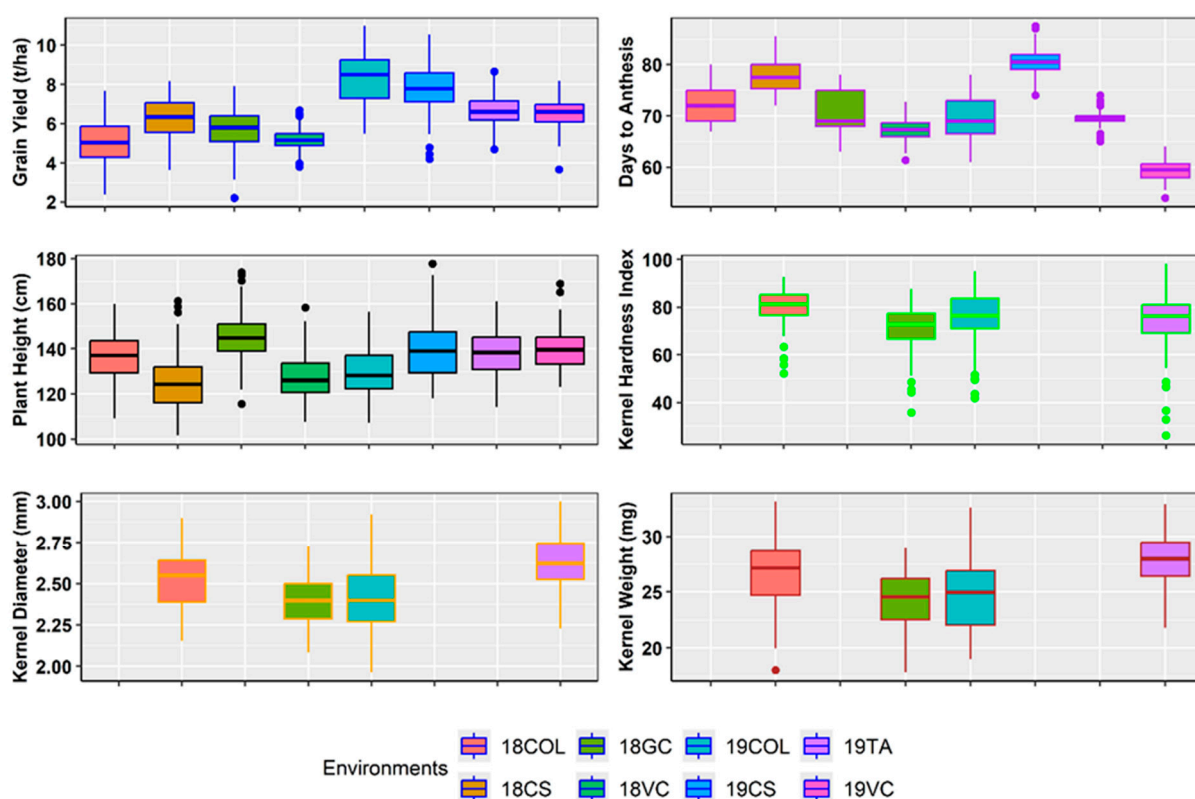


Figure 1. Distribution of best linear unbiased estimates of grain yield (t ha^{-1}), days to anthesis, and plant height (cm) for hybrids evaluated across eight environments: 18COL (2018 Colby, KS, USA), 18CS (2018 College Station, TX, USA), 18GC (2018 Garden City, KS, USA), 18VC (2018 Victoria, TX, USA), 19COL (2019 Colby, KS, USA), 19CS (2019 College Station, TX, USA), 19TA (2019 Taft, TX, USA), and 19VC (2019 Victoria, TX, USA); best linear unbiased estimates of kernel physical factors: kernel hardness index, kernel diameter, and kernel weight for four representative environments: 18CS (2018 College Station, TX, USA), 18VC (2018 Victoria, TX, USA), 19COL (2019 Colby, KS, USA), and 19VC (2019 Victoria, TX, USA).

The likelihood ratio test (LRT) for combined analysis across environments (using Equation (3)) revealed significant effects of hybrids, environment, and interaction of GCA and SCA effects with environments for GY, DA, PH, KHI, KD, and KW (Table 1). $G \times E$ interactions for GY, DA, and PH ranged from 4.9 to 12.1; the same for KHI, KD, and KW were slightly higher (10.8–19.4) (Table 1). Environmental effects accounted for 49%, 78%, and 24% of the variation for GY, DA, and PH, respectively. Likewise, hybrid effects accounted for 9%, 12%, and 47% of the variation for GY, DA, and PH, respectively. In general, kernel traits had larger hybrid effects than environmental effects; hybrid effects accounted for 57%, 54%, and 42% of KHI, KD, and KW, respectively. In terms of specific

genetic effects, the additive effect was larger in males than in females for all three kernel traits (Table 1). The $G \times E$ interactions were significant for both agronomic and kernel traits but lower than the main effects for both hybrids and environments. Broad-sense heritability (H^2) was high for all traits, and narrow-sense heritability (h^2) was only slightly lower, indicating that additive gene action predominates for these traits (Table 1). The h^2 estimates for male and female parents were similar for GY and DA, but h^2 estimates for the male parent were larger than those of female parents for PH, KHI, KD, and KW (Table 1).

Table 1. Variance components and heritability estimates for agronomic and grain traits in 100 grain sorghum hybrids from a 10×10 complete factorial that were evaluated in eight Texas and Kansas multi-locations in summer 2018 and 2019.

Variance Components ^a	Agronomic Traits						Kernel Traits					
	Grain Yield (GY)		Days to Anthesis (DA)		Plant Height (PH)		Kernel Hardness INDEX (KHI)		Kernel Diameter (KD)		Kernel Weight (KW)	
	Estimate	%	Estimate	%	Estimate	%	Estimate	%	Estimate	%	Estimate	%
(Hybrid)	0.39	14.5	6.23	11.8	94.6	48.1	76.6	56.7	0.023	54.1	4.76	41.5
GCA _f	0.18 ***	6.8	3.03 ***	5.7	35.1 ***	17.8	18.6 ***	13.8	0.009 ***	20.8	1.51 ***	13.2
GCA _m	0.18 ***	6.6	2.93 ***	5.5	51.0 ***	25.9	42.8 ***	31.7	0.013 ***	29.3	2.34 ***	20.5
SCA	0.03 **	1.2	0.28 ***	0.5	8.5 ***	4.3	15.2 ***	11.2	0.001 ***	4.05	0.90 ***	7.85
Env	1.30 ***	48.6	41.20 ***	78.0	51.0 ***	25.9	11.1 ^{NS}	8.2	0.01 **	23.1	2.45 **	21.4
(Hybrid \times Env)	0.32	12.1	2.58	4.9	15.2	7.7	58.2	19.4	0.003	10.8	1.76	15.4
GCA _f \times Env	0.16 ***	6.0	1.41 ***	2.7	4.1 ***	2.1	9.3 ***	6.9	0.001 ***	3.65	0.57 ***	4.94
GCA _m \times Env	0.09 ***	3.5	1.03 ***	1.9	8.5 ***	4.3	10.8 ***	8.0	0.001 ***	3.21	0.50 ***	4.39
SCA \times Env	0.078 *	2.6	0.14 ^{NS}	0.3	2.7 *	1.4	6.1 ***	4.5	0.001 ***	3.93	0.69 ***	6.05
Rep (Env)	0.02 ^{NS}	0.9	0.14 ^{NS}	0.3	1.3 ^{NS}	0.6	2.3 ^{NS}	1.6	0 ^{NS}	1.54	0.170 ^{NS}	1.48
Residual	0.634	23.8	2.65	5.0	34.6	17.6	18.7	13.8	0.004	10.5	2.31	20.2
H^2	0.83		0.92		0.96		0.90		0.93		0.87	
h^2	0.76		0.88		0.87		0.72		0.86		0.70	
h^2_f	0.38		0.44		0.35		0.22		0.36		0.27	
h^2_m	0.38		0.45		0.52		0.50		0.50		0.43	
CVe	12.50		2.31		4.36		5.76		2.75		5.88	

^a GCA_f, female general combining ability; GCA_m, male general combining ability; SCA, specific combining ability. ^{NS} Non-significant. * significant at the 0.05 probability level. ** significant at the 0.01 probability level. *** significant at the 0.001 probability level. H^2 : Broad sense heritability, h^2 : narrow sense heritability; h^2_f : female narrow sense heritability; h^2_m : male narrow sense heritability; CVe: coefficient of variation.

2.2. Variance Components and Heritability of NIR Spectra

In the combined analysis, using Equation (3) for all 4200 reflectance bands, female effects explained an overall of 17.6% of the variation, ranging from 1.42 to 30.6%. Male effects explained 26.3% of the variation with a range of 8.91 to 64.3% (Figure 2a). Female \times male effects ranged from 1.1 to 12.7% of the variation with an average of 5.1%. Combined, the total hybrid variation (female + male + female \times male) accounted for 49%, which indicates that spectra can delineate genotypic differences across environments. Environment variation was 37.1%, ranging from 3.75 to 58.3% (Figure 3). $G \times E$ effects explained 2.1% of total variation at spectrum of 487.5 nm, whereas about 20.8% of the variation was explained at spectrum of 1127.5 nm with a mean of about 9.99% (Figure 2a). Majority of variations were attributed to genetic effects with prominent male effects when analyzed within environments (Supplementary Figure S1). H^2 for reflectance values of spectra ranged from 0.91 (1098.5 nm) to 0.99 (487.5 nm). Most of these variations were driven by additive effects, as the h^2 range was from 0.85 (1857 nm) to 0.90 (630 nm). Further partitioning revealed that female h^2 ranged from 0.21 (487.5 nm) to 0.57 (805.5 nm); however, male h^2 ranged from 0.32 (807 nm) to 0.69 (487.5 nm) (Figure 2b).

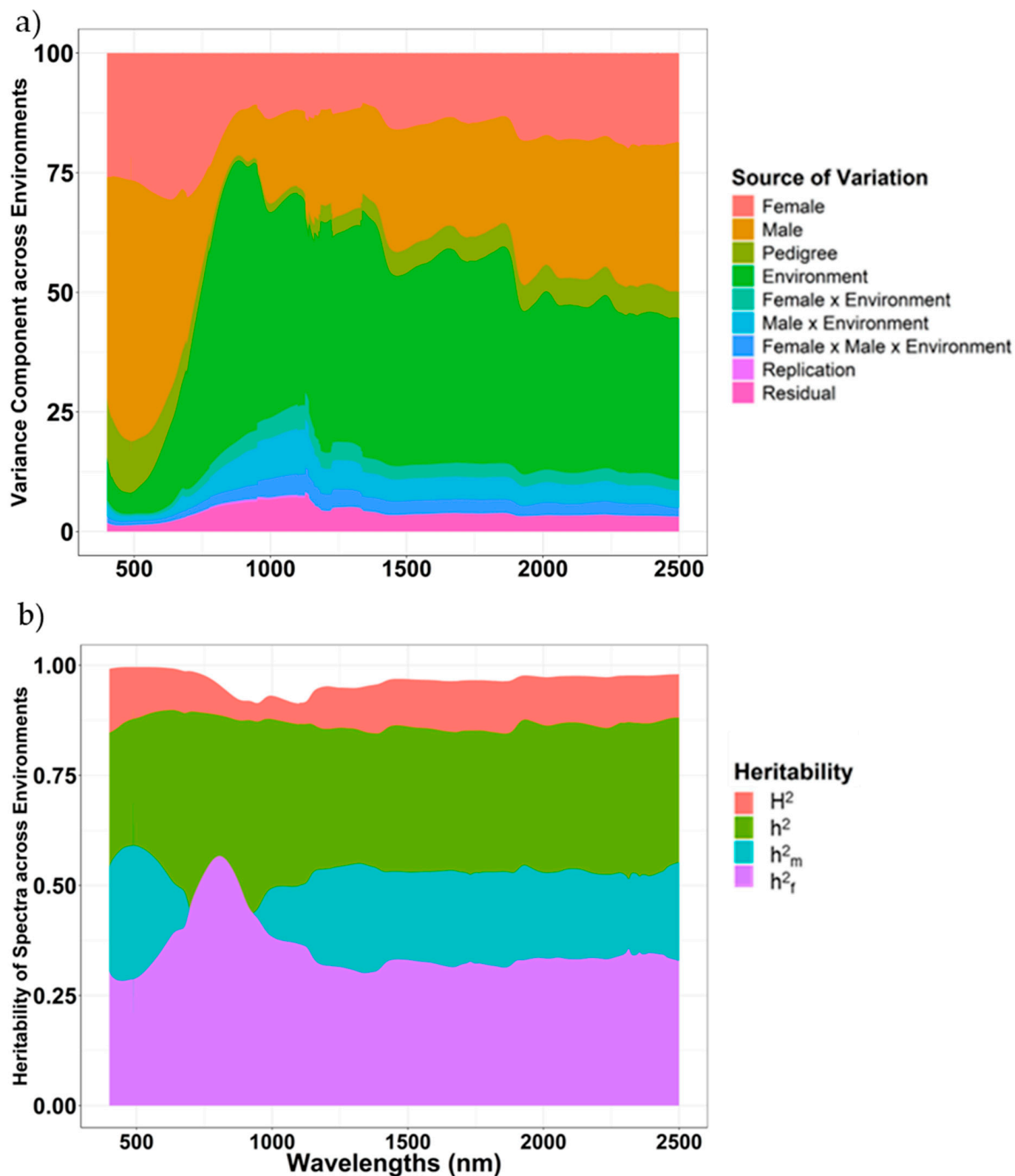


Figure 2. (a) Stacked graph presenting percent variation explained by female effects, male effects, female \times male effects, environment effects, female \times environment effects, male \times environment effects, and replication effects across spectral bands of 400 nm to 2500 nm across eight environments by Equation (3), where $y = 1 + Z_1f + Z_2m + Z_3h + Z_4E + Z_5fE + Z_6mE + Z_7fmE + Z_8r(E) + e$. (b) Overlaying graphs presenting broad sense heritability (H^2), total narrow sense heritability (h^2), female NSH (h^2_f), and male NSH (h^2_m) for spectral bands of 400 nm to 2500 nm across environments.

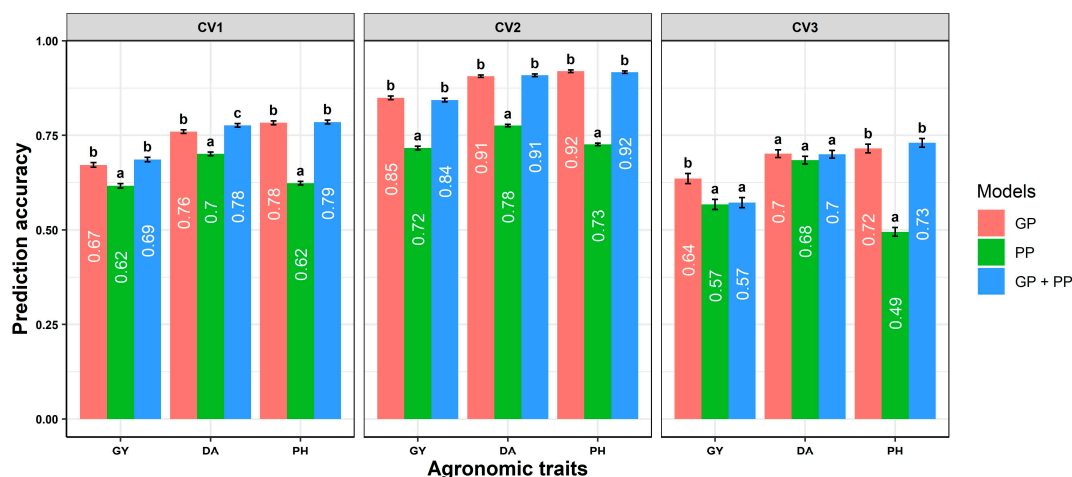


Figure 3. Prediction ability of genomic prediction (GP), phenomic prediction (PP), and GP + PP models for agronomic traits grain yield (GY), days to anthesis (DA), and plant height (PH) in three prediction scenarios: CV1 (untested hybrids in a characterized environment), CV2 (tested hybrids in an uncharacterized environment), and CV3 (untested hybrids in an uncharacterized environment) across environments. Error bars and significant letters a, b, and c were assigned for statistical differences determined by Tukey's honest test between models.

2.3. Predictive Abilities of Genomic or Phenomic Prediction Models

The prediction accuracy significantly varied across traits and models. In the combined analysis, GP consistently had a higher prediction accuracy in CV1 than PP for agronomic traits in hybrids (Figure 3). While GP + PP numerically outperformed GP in the scenario of CV1, it significantly outperformed GP for DA (Figure 3). Likewise, PP statistically performed similarly to GP for DA in CV3 (Figure 3).

For kernel traits, PP numerically outperformed GP in the CV1 scenario, and prediction was significantly improved for the GP + PP combined model (Figure 4). In the CV2 and CV3 prediction scenarios, GP was less effective than PP or GP + PP for KHI and KW; however, GP was more effective than PP for KD; and there was significant improvement in the GP + PP model for KHI and KW.

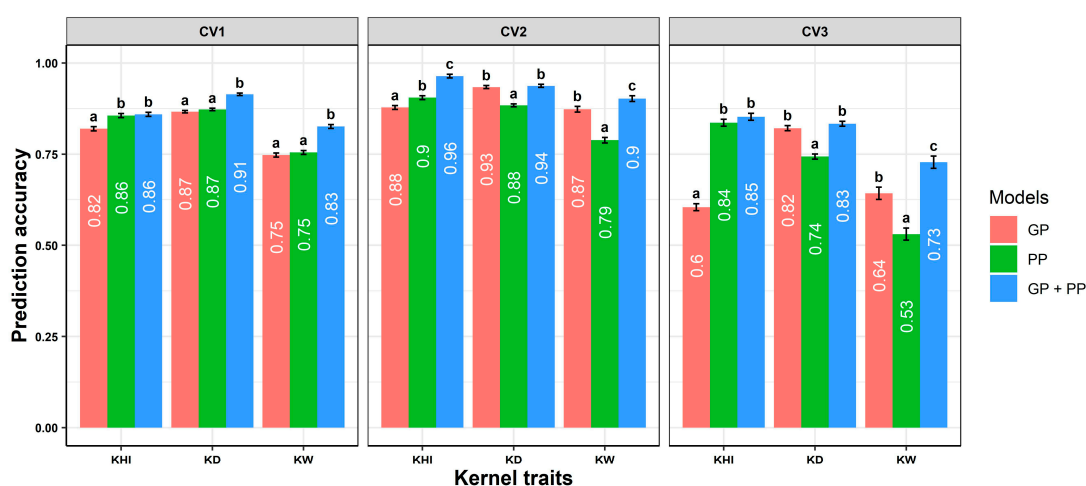


Figure 4. Prediction ability of genomic prediction (GP), phenomic prediction (PP), and GP + PP models for kernel physical characteristics: kernel hardness index (KHI), kernel diameter (KD), and kernel weight (KW) in three different prediction scenarios: CV1 (untested hybrids in a characterized environment), CV2 (tested hybrids in an uncharacterized environment), and CV3 (untested hybrids in an uncharacterized environment) across environments. Error bars and significant letters a, b, and c were assigned for statistical differences determined by Tukey's honest test between models.

2.4. Relationship of NIRS Bands with Phenotypic Traits

Correlations between all six traits and the specific bandwidth varied (Figure 5). For example, correlations between bandwidths and GY were highest ($r = 0.56$) at 1370 nm, while KHI correlations peaked ($r = 0.46$) at 1297 nm. Other peak correlations were $r = 0.51$ and 0.46 for KD (1331 nm) and KW (1923 nm), respectively. All the traits with higher prediction accuracy had their highest correlation with bands from 1200 nm to 1400 nm, which is the short-wavelength infrared region. Both DA and PH had no functional relationship with bandwidth at any wavelength (Figure 5). Interestingly, these traits had comparable prediction accuracies with phenomic and genomic prediction models.

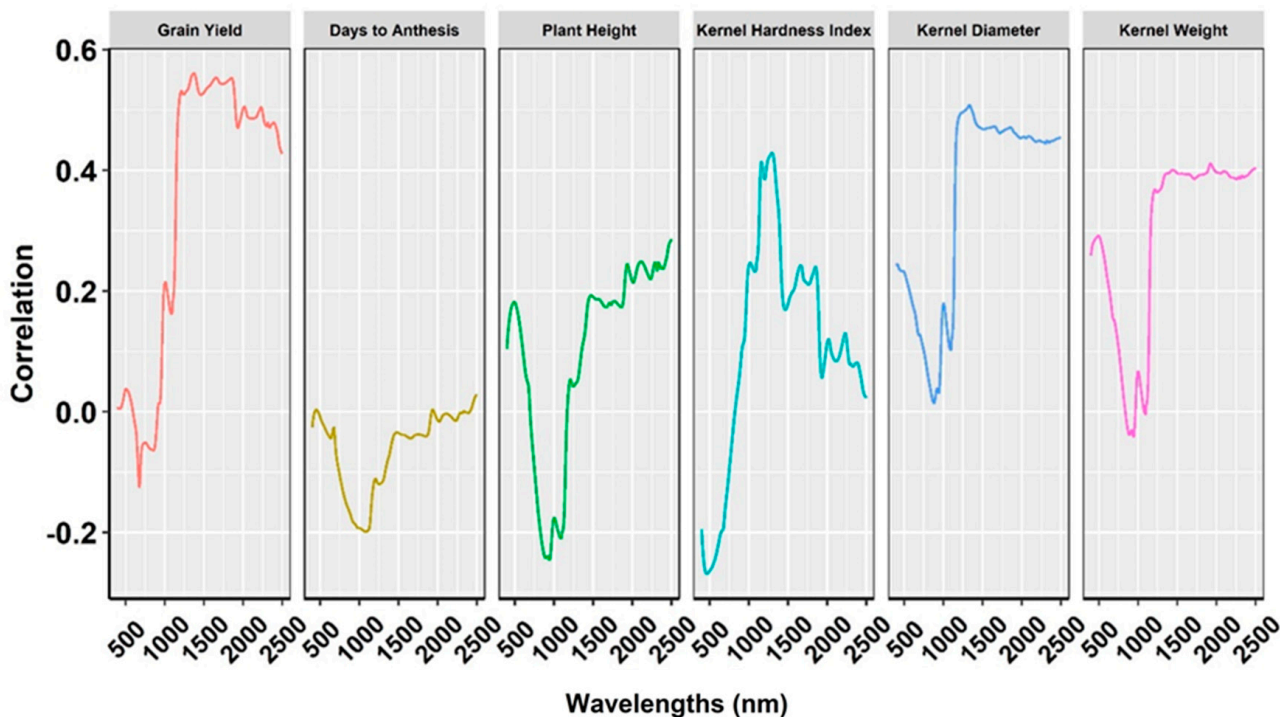


Figure 5. Pearson's correlation of NIR bands with grain yield, days to anthesis, plant height, kernel hardness index, kernel diameter, and kernel weight across spectral wavelengths of 400 nm–2500 nm across environments.

3. Discussion

Previous studies have reported the inheritance and the genetic variances of GY, PH, and DA [19,21], but the inheritance and the genetic variances of kernel traits are unique and represent some insight into the quantitative nature of these traits. Significant variation was detected among the hybrids, and the traits had high heritability for all three kernel traits (Table 1). Variation was detected in both the male and female parents, but the male parents accounted for a higher proportion of additive variation. Interestingly, the SCA effects for kernel traits were high relative to the agronomic traits, and this implies the influence of dominant gene effects for these traits. The largest variation for each of the three agronomic traits was due to the environmental effect, whereas genetic and $G \times E$ effects were the primary sources of variation for kernel traits. Given that selection for these traits is, at best, indirect, implies that it could shift means quickly towards any directional selection, although the significant $G \times E$ interactions for these traits must be considered in selection [27].

Earlier studies highlighted remote sensing as a useful phenomic selection tool for assessing plant health and grain yield in sorghum [28]. Motivated by studies in other crops such as corn, wheat, and coffee, this study represents the first report of NIR-based phenomic

prediction methods for agronomic performance and grain quality traits in sorghum [29–31]. As breeding programs have started implementing the GS approach, optimization of GS models is necessary to maintain consistent prediction accuracy, which could be achieved with the inclusion of other omics data [30,32,33]. NIR reflectance information, as low-cost omics data, has proven effective in complementing or supplementing genomic data for predictive modeling approaches [29,34]. This study explored the potential of implementing NIR bands alone or combining them with genomic data to train predictive models in sorghum hybrids.

NIR-based phenomic data appear useful in predicting economically important traits in sorghum. Using variance estimations from NIR bands, it was possible to consistently detect genetic, environmental, and $G \times E$ interaction effects across different bands (Figure 2a). The majority of variation in NIR spectra was attributable to additive genetic effects (female and male), highlighting the potential for genetic improvement through phenomic selection strategies [31]. Variations in NIR bands have been dissected in a similar manner in other crops as well [14,35,36]. The highly heritable nature of NIR spectra, along with major variations attributable to additive effects, demonstrates replicability for their regular use in a breeding program (Figure 2b). Overall, genetic variation in spectra was primarily attributed to male effects (Supplementary Figure S1), revealing that NIR spectra could be used to distinguish between hybrids in a manner akin to molecular markers. It should be noted that grain quality traits had larger genetic effects and more accurate predictions than agronomic traits. However, they were evaluated in fewer environments, which may have subjected these traits to lower $G \times E$ interaction effects. Consequently, higher prediction accuracies of phenomic models for three grain quality traits could be due to efficient discernment of NIR bands among grain samples. As expected, the relative effect of any bandwidth varied, and some were more correlated with genetics, while specific bandwidths were affected by the environment.

For untested hybrids in characterized environments (CV1), GP resulted in higher prediction accuracy than PP for agronomic traits, and prediction accuracy slightly increased (approximately 1.5%) for GP + PP, except for plant height (Figure 4). Similar results were reported in wheat [35,37]. The higher prediction accuracy observed in the GP + PP model may be due to the ability of NIR to consistently differentiate the hybrids. While the relationship between NIR reflectance and maturity is not obvious, the results indicated that NIR captured variation unexplained by GP, given the improvement in prediction accuracy for GP + PP. NIR can capture non-additive effects more consistently [35]. However, this is not always true; sorghum plant height is strongly influenced by non-additive gene action, but prediction accuracy did not improve with NIR band reflectance data. Thus, the typical improvements seen with NIR-based prediction models likely depend on the relationship between the trait and the physiochemical properties of the grain.

GP outperformed PP for GY, DA, and PH for untested hybrids in characterized environments (CV2), and GP + PP did not improve prediction accuracy. For untested hybrids in uncharacterized environments (CV3), GP outperformed PP for GY and PH, while PP was statistically similar to GP for DA. Similarly, within-environment predictability of GP was consistently better than that of PP, with additional gains observed when combining GP + PP for agronomic traits (Supplementary Figure S2).

In the CV1 scenario, PP statistically outperformed GP for KHI, and further significant improvements were observed in the GP + PP for KD and KW. Notably, PP significantly outperformed GP for KHI by approximately 24% in CV3 (Figure 4). Collecting grain samples from new environments can accurately predict KHI via phenomic models as compared to genomic models. The combined model GP + PP also improved prediction accuracy for KW by about 4% in CV2 and 9% in CV3. The high prediction accuracy

of PP relative to GP in CV2 and CV3 indicated its potential for predicting hybrids in uncharacterized environments. Similar patterns of improvement were recorded for GP + PP over GP for most of the single-environment models (Supplementary Figure S2). The superior performance of PP in predicting grain quality traits stems from their association with the physical structure of grain (Figure 4) [38]. Higher correlation was observed for most of the traits with respective NIR bands, except for plant height and days to anthesis (Figure 5). Grain yield and kernel weight were highly correlated with spectral bands in corn [30]. Inclusion of LASSO-nominated significant bands in corn recorded no improvement in prediction accuracy; therefore, fitting all the available bands would be effective for predictive modeling [30].

The whole-kernel NIR is effective at predicting hardness in wheat [39] and results herein indicate the same is true of sorghum. Thus, with further testing, NIR could offer a non-destructive alternative to the SKCS method. This could provide alternative methods for time-consuming and destructive SKCS methods that are rapidly disappearing from grain quality labs. Phenomic selection models can utilize the same single spectral scan to predict complex traits; therefore, it is rapid and less time-consuming. While these methodologies can be extended to other crops and traits, the efficacy of phenomic models is largely subject to the relationship between grain structure and traits of interest. For instance, phenomic models can be implemented to improve traits such as amylose, amylopectin, lysine and 3-DOA (3-deoxyanthocyanidins), etc., in suitable germplasm, specialty sorghum such as black sorghum and waxy sorghum, to accelerate trait-based breeding efforts.

The effectiveness of phenomic prediction in sorghum varies depending on the trait, with the greatest success observed for traits that are closely associated with the chemical composition of grain. This is a logical extension as NIRS detects chemical bonds that determine tissue composition, which are in turn influenced by both genetic and environmental factors [35]. NIRS effectively captures and tags specific regions of the reflectance spectrum, analogous to how QTL are identified in genomics [40]. As such, NIR reflectance data may also have the potential to consistently reveal genetic architecture underlying certain traits.

The positive correlation between phenomic and genomic predictions across traits indicated that spectral data exhibits a level of consistency comparable to that of genetic markers and phenotypic data [41]. Unlike genomic prediction, phenomic prediction is less dependent on the size of training sets and genetic relationships between individual hybrids and more influenced by $G \times E$ interactions [42]. Additionally, phenomic prediction is less prone to overestimation due to relatedness between training and testing sets, which can affect the accuracy of genomic prediction. The integration of high-dimensional phenomic datasets that capture underlying sources of variations has shown potential for enhancing prediction accuracy in this study and in others [43].

Although phenomic-assisted breeding has demonstrated efficiency, it remains less commonly used than genomic-assisted breeding in cultivar development pipelines [44]. Limited application of PS might have been associated with the challenges associated with cost, labor, time, and processing [16]. One of the major caveats of grain-based omics data, as compared to genomic data, is the requirement for multi-location trials to capture spectral data across diverse environments. In addition, phenomic data are influenced by environmental factors and $G \times E$ interactions, which can obscure reliable genetic differences among lines [42]. However, the high repeatability of NIR spectra observed in this study indicates that fewer environments may be sufficient for reliable data collection.

This study acknowledges the disadvantage of PP over GP, as there is still a need for hybrid development, grain sampling, and NIR scanning, while GP can be implemented by using molecular markers obtained from parents. GP models could be effective for breeding programs that already have logistics developed, while PP can be useful to the

breeding programs that are at a disadvantage in genotyping. Smaller breeding programs, having limited genomic resources, for example, in developing countries, e.g., Asian and African countries, can take advantage of this method. While GP demands integration of genotyping platforms, maintaining relatedness, and appropriate training size; PP is less influenced in such scenarios. Implementation of PP into a breeding program does not require extensive redesigning in terms of logistics. While deciding on the efficacy of PP, it is usually benchmarked against GP. Prediction accuracy was typically considered as a metric to show that a model is better than others. However, more accurate models do not necessarily have better prediction accuracy, and relying solely on prediction accuracy is not sufficient. The application of prediction models impacts other parameters of the breeder's equation and the rate of genetic gain must be estimated by considering the full effects of their application [45].

In a sorghum improvement program employing GP, a subset of lines, such as doubled haploids and lines from the F_4 generation, are testcrossed to generate hybrids for advanced testing and to serve as training sets. The success of GP in such a system depends on maintaining consistent prediction accuracy across cycles. Integrating the phenomics data, specifically the NIR spectra obtained from grain samples collected during hybrid trials, could enhance the prediction accuracy of GP models. Notably, since mechanical harvesting systems often include built-in NIR prediction capabilities, these spectral data are already available and can be utilized without incurring additional resource demands.

Multi-environment breeding trials are often subject to biotic and abiotic stresses, which can limit the collection of reliable phenotypic data. In such cases, collecting grain samples from selected environments may reduce the need for extensive phenotyping, as NIRS has the potential to capture genotype, environment, and $G \times E$ interaction effects [46]. PP models can be trained to predict hybrid performance across environments by leveraging spectral similarities between both hybrids and environments [47]. This could improve the phenotypic data availability to train GP models to improve breeding outcomes [45]. These methodologies could be further enhanced by using spectral data collected at early growth stages such as from vegetative tissues using a hand-held spectrophotometer rather than waiting until grain maturity. Training PP models with early growth stage spectra may improve practical applications in the breeding program. Improvement in prediction accuracy in multi-trait genomic selection models for grain quality traits, suggests a direction for similar advancements in phenomic selection [48]. Consequently, multi-trait phenomic prediction models should be explored further in sorghum.

4. Conclusions

This study investigated the potential of NIR-based phenomic prediction models to increase the selection efficacy of hybrids from applied sorghum breeding programs. This first-time study indicates that NIR-based phenomic prediction can be effective for certain traits in sorghum. Specifically, phenomic predictions were comparable to or exceeded genomic predictions for grain yield, days to anthesis, kernel hardness index, kernel diameter, and kernel weight. However, phenomic prediction was notably less effective for plant height. A consistent improvement in prediction accuracy was observed when phenomic data were integrated into genomic prediction models. These results suggest that genomic prediction remains preferable for traits not closely associated with grain characteristics, whereas phenomic data can add value when relevant samples are available. Nonetheless, the utility of phenomic prediction is limited by the prerequisite of obtaining phenotypic samples, which may not align with early-stage genomic selection strategies that precede phenotyping. Therefore, phenomic prediction could serve as a complementary tool to genomic prediction, particularly for materials lacking genotypic data or in later stages of

selection. Future research studies should explore the integration of phenomic data derived from remote sensing technologies and include more genetically diverse materials across a broader range of environments. In this study, smoothing phenomic data using first derivatives produced results like genomic prediction models. Advancing this work will require more robust machine learning models capable of integrating high-dimensional data to improve the prediction of economically important traits in sorghum.

5. Materials and Methods

5.1. Experimental Design

A total of 100 hybrids were developed using a 10×10 complete factorial mating design. These hybrids were grown in eight different environments in summer 2018 and 2019. Details on hybrid development and field evaluation were described in a previous publication [19]. In brief, the experimental design in each environment was a randomized complete block design (RCBD) with three replications in 2018 and two replications in 2019. The experiments in 2018 were conducted in four environments: Victoria, Texas: 18VC (28°47'24.4" N, 96°50'22.6" W), College Station, Texas: 18CS (30°32'56.6" N, 96°26'11.5" W), Garden City, Kansas: 18GC (37°59'21.4" N, 100°48'52.5" W), Colby, Kansas: 18COL (39°22'56.6" N, 101°04'45.0" W). In 2019, the same experiments were conducted in Victoria, College Station and Taft (19TA, 28°00'05.4" N, 97°15'12.4" W), Texas, and Colby, Kansas.

5.2. Agronomic Evaluation and Traits

Data on three agronomic traits: days to anthesis (DA), plant height (PH), and grain yield (GY) were recorded on a plot basis uniformly in all the locations studied. The DA was the number of days from planting to when half of the plot flowered halfway down the panicle. For PH, a representative plant from a plot was measured from the ground to the tip of the panicle. At physiological maturity, the crop was harvested using a plot combine; grain was weighed, and moisture content was measured. Plot weights were later adjusted for 14% of moisture to calculate GY on a ton/hectare (t/ha).

5.3. Grain Characterization

A grain sample, collected on a plot basis, and composed of five panicles typical of the plot was hand-harvested prior to combine harvest. The panicles were threshed in Almaco BT14E Belt Thresher (ALMACO), and grain samples were dried to a shelf-stable moisture content averaging 12%. A 100 g grain sample from each hybrid in each environment was scanned using a FOSS 2500 spectrophotometer (FOSS North America, Eden Prairie, MN, USA, 2005), which measures reflectance between 400 and 2500 nm at an interval of 0.5 nm. By using the calibrated curves available in the Texas A&M sorghum breeding and genetics laboratory, these spectra were estimated for starch, protein, fiber, fat, ash, and moisture content in the grain (on a % basis). In addition, 300 grains of each plot from four representative environments (18CS, 18VC, 19COL, and 19VC) were analyzed for the three physical traits such as the kernel hardness index (KHI), kernel diameter (KD), and kernel weight (KW), using the single-kernel characterization system (SKCS 4100, Perten Instruments North America Inc., Springfield, IL, USA). The value for KHI was presented in numbers ranging from 0 to 100, where a higher number indicates a harder kernel.

5.4. Phenomic Data

The raw NIR spectra generated to estimate grain composition were also used to calculate the best linear unbiased estimates (BLUEs) for each hybrid within environments for all wavelengths by solving mixed models presented in Equation (1). 'y' represents the vector of the response variable, h represents the hybrid fixed effect, r represents replications,

and e represents residuals in Equation (1). Likewise, Z_1 and Z_2 are incidence matrices relating phenotypes with hybrids and replications, respectively. The BLUEs of phenomic data have dimensions of 100×4200 and 800×4200 within and across all environments, respectively. Data smoothing and pretreatment with the first derivative were performed using the ‘SavitzkyGolay’ function with a window of 11 to remove low-quality spectra. Filtered data were scaled around zero to generate even distribution. The dimensions of phenomic data were reduced to 100×4190 , and 800×4190 , respectively, for single and all environments. All pretreatments were performed using the package ‘prospectr’ in R 4.3.1 [49]. These clean data were used to build the relationship among hybrids. Phenomic relationship matrices were computed as $\frac{NIR \times NIR'}{4190}$, where NIR was the scaled first derivative of phenomic data, and NIR' was the transpose of NIR.

$$y = 1 + Z_1h + Z_2r + e \quad (1)$$

5.5. Genotypic Data

DNA extraction, quality control, and library preparation were performed to perform genotyping by sequencing (GBS) for parents used in this study [50]. All the steps for mapping and variant calling were detailed in a previous publication [19]. Single Nucleotide Polymorphic (SNP) markers with more than 5% missing values were removed and missing values were imputed. Finally, 35,546 SNP markers were used to build a genomic relationship matrix, where alleles were coded as 0, 1, 2; 0 = homozygous recessive, 1 = heterozygote, and 2 = homozygous dominant. A genomic relationship matrix was calculated using the ‘vanraden’ method in R [51]. In the genomic relationship matrix among parents, G would be $G = (XX')/n$, where X is the SNP marker of parents centered around zero, X' is the transpose of X , and n is the number of 35,546 SNP markers used. These relationships were used to build kernels to model genetic effects via general combining abilities (GCAs) and specific combining abilities (SCAs) to fit into predictive models.

5.6. Variance Component Estimations

Phenotypic data and reflectance values were fitted in a linear mixed model, and the effects of female and male parents, hybrid (female \times male) effects, and replications within environment were estimated using Equation (2).

$$y = 1\mu + Z_1f + Z_2m + Z_3h + Z_4r + e \quad (2)$$

where y is a vector of response variables; μ is an intercept; f is a random effect of female, $f \sim N(0, \sigma_f^2 I)$; m is a random effect of male, $m \sim N(0, \sigma_m^2 I)$; h is a random effect of hybrid $h \sim (0, \sigma_h^2 I)$; r is a vector of replications $r \sim N(0, \sigma_r^2 I)$; e is a vector of residuals, $e \sim N(0, \sigma_e^2 I)$; 1 is a vector of ones; Z_1 , Z_2 , Z_3 , and Z_4 are incidence matrices; σ_f^2 , σ_m^2 , σ_r^2 , and σ_e^2 are variance components of females, males, hybrids, replicates and residuals, respectively. For combined-environment analysis, Equation (2) was extended to include environment and genotype \times environment ($G \times E$) and performed the following combined analysis:

$$y = 1\mu + Z_1f + Z_2m + Z_3h + Z_4s + Z_5fs + Z_6ms + Z_7fms + Z_8r(s) + e \quad (3)$$

For combined analysis, s is a vector of environmental effects, $s \sim N(0, \sigma_s^2 I)$; fs is a vector of the interaction effect between the GCA of female and environmental effects, $fs \sim N(0, \sigma_{fs}^2 I)$; ms is a vector of the interaction effect between the GCA of male and environmental effects, $ms \sim N(0, \sigma_{ms}^2 I)$; fms is a vector of the interaction effect between the SCA of hybrid combinations and environmental effects, $fms \sim N(0, \sigma_{fms}^2 I)$; $r(s)$ is the

vector of replication nested within environments, $r(s) \sim N(0, \sigma_{r(s)}^2 I)$. Variance components for response variables were estimated via restricted maximum likelihood (REML), and significance was assessed by the likelihood ratio test (LRT) at a 5% level. All analyses were performed on the lme4 package in R. Based on variance component estimations, broad sense heritability and narrow sense heritability were computed for a single environment: $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{r}}$ and $h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \frac{\sigma_e^2}{r}}$. where σ_g^2 is the variance component due to hybrids, σ_a^2 is additive genetic variance, σ_e^2 is error variance, and r is number of replications in each experiment. Narrow sense heritability for females (h_f^2) and males (h_m^2) within an environment was computed: $h_f^2 = \frac{\sigma_f^2}{\sigma_g^2 + \frac{\sigma_e^2}{r}}$ and $h_m^2 = \frac{\sigma_m^2}{\sigma_g^2 + \frac{\sigma_e^2}{r}}$, respectively. These equations were extended for combined analysis: $H^2 = \frac{\sigma_g^{2*}}{\sigma_g^{2*} + \frac{\sigma_{gs}^{2*}}{t} + \frac{\sigma_e^2}{tr}}$, $h^2 = \frac{\sigma_a^{2*}}{\sigma_g^{2*} + \frac{\sigma_{gs}^{2*}}{t} + \frac{\sigma_e^2}{tr}}$, $h_f^2 = \frac{\sigma_f^2}{\sigma_g^{2*} + \frac{\sigma_{gs}^{2*}}{t} + \frac{\sigma_e^2}{tr}}$, and $h_m^2 = \frac{\sigma_m^2}{\sigma_g^{2*} + \frac{\sigma_{gs}^{2*}}{t} + \frac{\sigma_e^2}{tr}}$ where σ_g^{2*} is the variance component due to hybrids that were computed as the summation of the variance component due to female (σ_f^2), male (σ_m^2), and female and male interaction (σ_{fm}^2) effects. σ_{gs}^{2*} is the variance component due to hybrids \times environment interaction, which was computed as a summation of female \times environment (σ_{fs}^2), male \times environment (σ_{ms}^2), and female \times male \times environment (σ_{fms}^2). Likewise, σ_a^{2*} is additive genetic variance and is calculated as a summation of σ_f^2 and σ_m^2 . 't' and 'r' are the number of environments and replications within each environment, respectively. Coefficient of variation (CV_e) was calculated using $CV_e = \frac{\sqrt{\sigma_e^2}}{\bar{x}}$, where σ_e^2 is residual variance and \bar{x} is overall mean of the traits.

BLUEs were calculated for all phenotypic data by fitting fixed hybrid effects at each environment using Equation (1). Extracted BLUEs for hybrids were used to train prediction models. Kernel-based genomic best linear unbiased prediction (GBLUP) models were trained using SNP markers in silico for the genomic prediction (GP) model. The phenomic prediction (PP) model used cleaned NIR spectral data after filtering and pretreatment of the first derivative. Phenomic data were integrated with genomic data for the GP + PP model. All the statistical models were fitted for single and multi-environments and defined as single environment and combined environments.

5.7. Genomic and Phenomic Prediction Models

5.7.1. Single-Environment Prediction Models

Three different predictive models were fitted to predict the phenotypic performance of the hybrids. These include genomic prediction (GP), phenomic prediction (PP), and genomic + phenomic prediction (GP + PP). GP was trained by modeling the GCA effects of females and males and SCA effects of female \times male interactions using SNP markers. The first derivative of NIR was considered as a PP. Finally, phenomic data was integrated with genomic data for GP + PP models.

GP Model:

$$y = 1\mu + Z_1f + Z_2m + Z_3h + e \quad (4)$$

PP Model:

$$y = 1\mu + NIR_1 + e \quad (5)$$

GP + PP Model:

$$y = 1\mu + NIR_1 + Z_1f + Z_2m + Z_3h + e \quad (6)$$

where y is a vector of response variables; μ is intercept; f is a vector of GCA effects of females, $f \sim N(0, \sigma_{fj}^2 G_f)$, where j is environments; $m \sim N(0, \sigma_{mj}^2 G_m)$, m is a vector of GCA effects of males, $h \sim N(0, \sigma_{hj}^2 H)$, where h is a vector of SCA effects of hybrid combinations; $e \sim N(0, \sigma_{ej}^2 I)$, e is a vector of residuals. Notably, 1 is a vector of ones; Z_1, Z_2 , and Z_3 are incidence matrices for females, males, and hybrids, respectively. H is calculated by the Kronecker product in silico using G_f and G_m . For phenomic kernels, cleaned NIR bands, i.e., the first derivative of NIR, were modeled like genomic markers; therefore, it follows the distribution of $NIR_1 \sim^{i.i.d.} N(0, NIR\sigma_p^2)$, where σ_p^2 and NIR represent a phenomic variance component and phenomic effects estimated using the first derivative of NIR.

5.7.2. Combined Environment Prediction Models

The single-environment models were extended to incorporate interactions with multi-environments. Environmental and $G \times E$ effects were modeled by including incidence matrices that relate phenotypic observation with the environment [52]. NIR \times E interaction was computed using the Hadamard product since NIR was collected from grain samples of each environment. The Hadamard product (where \odot denotes the element-wise product) of the covariance structure of the interaction was modeled as follows: $f \times E \sim^{i.i.d.} N(0, \sigma_{f \times E}^2 V_f)$, $m \times E \sim^{i.i.d.} N(0, \sigma_{m \times E}^2 V_m)$ and $f \times m \times E \sim^{i.i.d.} N(0, \sigma_{H \times E}^2 V_H)$; where $\sigma_{f \times E}^2$, $\sigma_{m \times E}^2$, and $\sigma_{H \times E}^2$ are variance components associated with female \times environment, male \times environment, and female \times male \times environment, respectively. V_F, V_M , and V_H represent variance–covariance matrices for female \times environment, male \times environment, and female \times male \times environments, respectively. Variance–covariance matrices were calculated as $V_F = Z_1 G_f Z_1' \odot Z_E Z_E'$, $V_M = Z_2 G_m Z_2' \odot Z_E Z_E'$, and $V_H = Z_3 G_H Z_3' \odot Z_E Z_E'$. For the interaction of phenomic effects, $NIR_1 \times E \sim^{i.i.d.} N(0, \sigma_{P \times E}^2 V_P)$ where $\sigma_{P \times E}^2$ are variance components associated with phenomic effects \times environment and $V_P = Z_P P_H Z_P' \odot Z_E Z_E'$.

GP Model:

$$y = + Z_1 f + Z_2 m + Z_3 h + Z_4 E + Z_5 fE + Z_6 mE + Z_7 fmE + e \quad (7)$$

PP Model:

$$y = \mu + NIR_1 + E + NIR_1 E + e \quad (8)$$

GP + PP Model:

$$y = + Z_1 f + Z_2 m + Z_3 h + Z_4 E + Z_5 fE + Z_6 mE + Z_7 fmE + NIR_1 + NIR_1 E + e \quad (9)$$

5.7.3. Cross-Validation Performance Evaluation

For the combined environment analysis, three different cross-validation (CV) scenarios were tested and repeated 50 times. These three scenarios represent the selection situations encountered in a breeding program. The CV1 method predicts hybrids that have not been evaluated in any environments; however, sets of genetically related hybrids have been evaluated in the same environment. The CV2 method (leave one environment out) predicts previously tested hybrids in uncharacterized environments. Likewise, the CV3 method predicts untested hybrids in uncharacterized environments (new lines in new environments). For CV1, 100 hybrids were partitioned into 70:30 (70 hybrids for training and 30 for testing purposes for all environments) (Supplementary Figure S3a). Training datasets of $(70 \times 8 = 560)$ were used to predict $(30 \times 8 = 240)$ records for agronomic traits: grain yield, days to anthesis, and plant height across eight environments. After fitting the models, Pearson's correlation coefficients were calculated for testing sets of each environment for CV1. In a similar way, training datasets of $(70 \times 4 = 280)$ were used to predict testing sets of $(30 \times 4 = 120)$ for kernel characteristic traits from four environments.

For CV2 and CV3, phenotypic records for one of the environments were removed in addition to 70:30 partitions for CV1. In this case, the training set consists of $(70 \times 7 = 490)$ and testing sets of $[(30 \times 7) + 100 = 310]$ for agronomic traits (Supplementary Figure S3b). Similarly, training sets consist of $(70 \times 3 = 210)$ to predict testing sets comprising $[100 + (30 \times 3) = 190]$ for kernel characteristic traits. Pearson's coefficient of correlation of test environments was computed for CV2 and CV3. In this analysis, 18CS and 19COL were chosen as uncharacterized environments for CV2 and CV3 schemes. Likewise, the models were fitted in a 70:30 ratio for a single environment without $G \times E$ kernels.

Pearson's correlation coefficient on CV, model, and trait combinations was finally used to calculate mean and standard error for multi-environment models. Finally, post hoc analyses with Tukey's honest test at the significance level of 0.05 were performed for prediction accuracies to detect differences between model performances for all traits.

5.8. Software

All models were fitted in R where SNP markers and NIR spectra were used for GP and PP models, respectively [53]. The first derivative of the matrix of NIR was calculated using the 'SavitzkyGolay' function from the library 'prospectr' to build phenomic relationship matrices [49]. All the prediction models were fitted with 5000 iterations, and burnIn of 1000 with thin of 10 using Gibbs sampler using the 'BGLR' package in R [54]. Genomic and phenomic priors were fitted as Reproducing Kernel Hilbert Spaces ("RKHS"); however, environment kernels were fitted using "BRR". Three models, as explained above, were fitted for a single environment and multi-environment. $G \times E$ and $NIR \times E$ were calculated using Hadamard products.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants14182871/s1>, Figure S1: Percent variation explained by male effects, female effects, female \times male effects, and replication across all 4200 near-infrared spectroscopy (NIRS) bands across eight environments by equation $y = 1 + Z_1f + Z_2m + Z_3h + e$. Abbreviations for locations are COL: Colby, KS, CS: College Station, TX, GC: Garden City, KS, VC: Victoria, TX, TA: Taft, TX. Figure S2: Prediction abilities of Genomic Prediction (GP), Phenomic Prediction (PP), GP + PP models in predicting GY: Grain Yield, PH: Plant Height, DA: Days to Anthesis, KHI: Kernel Hardness Index, KD: Kernel Diameter, and KW: Kernel Weight for eight environments (18COL: 2018 Colby, KS; 18CS: 2018 College Station, TX; 18GC: 2018 Garden City, KS; 18VC: 2018 Victoria, TX; 19COL: 2018 Colby, Kansas; 19CS: 2019 College Station, TX; 19TA: 2019 Taft, TX; 19VC: 2019 Victoria, TX). Bars represent standard deviation for each model within environments for respective traits. Figure S3: Cross validation scheme across environments for genomic and phenomic models in predicting sorghum hybrid performance. (a) CV1 cross validation scheme across environments with 30% testing hybrids i.e., $30 \times 8 = 240$, which were predicted by training 70% of hybrids across environments i.e., $70 \times 8 = 560$. (b) CV2 and CV3 cross validation schemes were performed by leaving one environment out on top of 30% of hybrids from other seven environments, prediction accuracy was calculated for 70% tested hybrids on uncharacterized environments for CV2 and 30% untested hybrids in uncharacterized environments for CV3 as shown above. Prediction accuracy was calculated by calculating Pearson's correlation coefficients within testing sets for all CV schemes.

Author Contributions: Conceptualization, P.S., J.F., R.P., and W.L.R.; formal analysis, P.S.; funding acquisition, W.L.R.; investigation, P.S., J.F., R.P., J.C., and W.L.R.; methodology, P.S., J.F., R.P., J.C., and W.L.R.; writing—original draft, P.S.; writing—review and editing, P.S., J.F., R.P., J.C., and W.L.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Sapkota, P. (2025). Supplementary data—phenomic and genomic models for predicting sorghum agronomic and grain characteristics related traits across environments. https://github.com/sapkotapradip/Phenomic_NIR-Genomic_Pub (accessed on 27 July 2025).

Acknowledgments: Authors would like to thank Robert Klein, Patricia Klein, and Julie McCollum for assisting on the effort of genotyping of parents used in this study. The authors would like to thank Cereal Quality Laboratory at Texas A&M for facilitating Single Kernel Characterization Systems to facilitating phenotyping of kernel characterizations. Also, we thank the Dudley Smith family for the research enhancement fund.

Conflicts of Interest: Author Jales Fonseca was employed by the Bayer Crop Science-US. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GS	Genomic selection
NIRS	Near-infrared spectroscopy
PS	Phenomic selection
KHI	Kernel hardness index
GY	Grain yield
DA	Days to anthesis
PH	Plant height
KD	Kernel diameter
KW	Kernel weight
SKCS	Single-kernel characteristics system
BLUE	Best linear unbiased estimates
LRT	Likelihood ratio test
GCA	General combining ability
SCA	Specific combining ability
G × E	Genotype × environment
H ²	Broad sense heritability
h ²	Narrow sense heritability
GP	Genomic prediction
PP	Phenomic prediction

References

1. Crossa, J.; Fritsche-neto, R.; Montesinos-lopez, O.A.; Costa-neto, G.; Dreisigacker, S.; Montesinos-lopez, A. The modern plant breeding triangle: Optimizing the use of genomics, phenomics, and enviromics data. *Front. Plant Sci.* **2021**, *12*, 651480. [CrossRef]
2. Herr, A.W.; Adak, A.; Carroll, M.E.; Elango, D.; Kar, S.; Li, C.; Jones, S.E.; Carter, A.H.; Murray, S.C.; Paterson, A.; et al. Unoccupied aerial systems imagery for phenotyping in cotton, maize, soybean, and wheat breeding. *Crop Sci.* **2023**, *63*, 1722–1749. [CrossRef]
3. Hulse, J.H.; Laing, E.M.; Pearson, O.E. *Sorghum and the Millets: Their Composition and Nutritive Value*; Academic Press: Cambridge, MA, USA, 1980; p. 997.
4. Rooney, L.W.; Awika, J.M. Overview of products and health benefits of specialty sorghums1. *Cereals Foods World* **2005**, *50*, 109.
5. Taylor, J.R.; Schober, T.J.; Bean, S.R. Novel food and non-food uses for sorghum and millets. *J. Cereal Sci.* **2006**, *44*, 252–271. [CrossRef]
6. Pfeiffer, B.K.; Pietsch, D.; Schnell, R.W.; Rooney, W.L. Long-term selection in hybrid sorghum breeding programs. *Crop Sci.* **2019**, *59*, 150–164. [CrossRef]
7. Leon, F.; Harvey, L.; Rooney, W. Historical multi-environmental trials reveal trends in hybrid grain sorghum performance from 1970 to 2021 in Texas. *Crop Sci.* **2024**, *64*, 3014–3027. [CrossRef]
8. Xin, Z.; Wang, M.; Cuevas, H.E.; Chen, J.; Harrison, M.; Pugh, N.A.; Morris, G. Sorghum genetic, genomic, and breeding resources. *Planta* **2021**, *254*, 114. [CrossRef]
9. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [CrossRef]

10. Crossa, J.; Pérez-Rodríguez, P.; Cuevas, J.; Montesinos-López, O.; Jarquín, D.; de los Campos, G.; Burgueño, J.; González-Camacho, J.M.; Pérez-Elizalde, S.; Beyene, Y.; et al. Genomic selection in plant breeding: Methods, models, and perspectives. *Trends Plant Sci.* **2017**, *22*, 961–975. [[CrossRef](#)]
11. Yang, W.; Feng, H.; Zhang, X.; Zhang, J.; Doonan, J.H.; Batchelor, W.D.; Xiong, L.; Yan, J. Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol. Plant* **2020**, *13*, 187–214. [[CrossRef](#)]
12. Pasquini, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.* **2003**, *14*, 198–219. [[CrossRef](#)]
13. Engel, J.; Gerretzen, J.; Szymańska, E.; Jansen, J.J.; Downey, G.; Blanchet, L.; Buydens, L.M.C. Breaking with trends in pre-processing? *TrAC-Trends Anal. Chem.* **2013**, *50*, 96–106. [[CrossRef](#)]
14. Rincet, R.; Charpentier, J.P.; Faivre-Rampant, P.; Paux, E.; Le Gouis, J.; Bastien, C.; Segura, V. Phenomic selection is a low-cost and high-throughput method based on indirect predictions: Proof of concept on wheat and poplar. *G3 Genes Genomes Genet.* **2018**, *8*, 3961–3972. [[CrossRef](#)] [[PubMed](#)]
15. Barclay, V.J.; Bonner, R.F.; Hamilton, I.P. Application of wavelet transforms to experimental spectra: Smoothing, denoising, and data set compression. *Anal. Chem.* **1997**, *69*, 78–90. [[CrossRef](#)]
16. Winn, Z.J.; Amsberry, A.L.; Haley, S.D.; DeWitt, N.D.; Mason, R.E. Phenomic versus genomic prediction—A comparison of prediction accuracies for grain yield in hard winter wheat lines. *Plant Phenome J.* **2023**, *6*, e20084. [[CrossRef](#)]
17. Rutkoski, J.; Poland, J.; Mondal, S.; Autrique, E.; Pérez, L.G.; Crossa, J.; Reynolds, M.; Singh, R. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes Genomes Genet.* **2016**, *6*, 2799–2808. [[CrossRef](#)]
18. Sandhu, K.S.; Merrick, L.F.; Sankaran, S.; Zhang, Z.; Carter, A.H. Prospectus of genomic selection and phenomics in cereal, legume and oilseed breeding programs. *Front. Genet.* **2022**, *6*, 2799–2808. [[CrossRef](#)]
19. Fonseca, J.M.O.; Klein, P.E.; Crossa, J.; Pacheco, A.; Perez-Rodriguez, P.; Ramasamy, P.; Klein, R.; Rooney, W.L. Assessing combining abilities, genomic data, and genotype × environment interactions to predict hybrid grain sorghum performance. *Plant Genome* **2021**, *14*, e20127. [[CrossRef](#)]
20. Crozier, D.; Leon, F.; Fonseca, J.M.O.; Klein, P.E.; Klein, R.R.; Rooney, W.L. Inbred phenotypic data and non-additive effects can enhance genomic prediction models for hybrid grain sorghum. *Crop Sci.* **2023**, *63*, 1183–1196. [[CrossRef](#)]
21. Kent, M.A.; Fonseca, J.M.O.; Klein, P.E.; Klein, R.R.; Hayes, C.M.; Rooney, W.L. Use of genomic prediction to screen sorghum B-lines in hybrid testcrosses. *Plant Genome* **2023**, *16*, e20369. [[CrossRef](#)]
22. Winans, N.D.; Klein, R.R.; Fonseca, J.M.O.; Klein, P.E.; Rooney, W.L. Evaluating introgression sorghum germplasm selected at the population level while exploring genomic resources as a screening method. *Plants* **2023**, *12*, 444. [[CrossRef](#)]
23. Are, A.K.; Srivastava, R.K.; Mahalingam, G.; Gorthy, S.; Gaddameedi, A.; Kunapareddy, A.; Kotla, A.; Jaganathan, J. Application of plant breeding and genomics for improved sorghum and pearl millet grain nutritional quality. In *Sorghum and Millets*; AACC International Press: Saint Paul, MN, USA, 2019; pp. 51–68.
24. Jambunathan, R.; Kherdekar, M.S.; Stenhouse, J.W. Sorghum grain hardness and its relationship to mold susceptibility and mold resistance. *J. Agric. Food Chem.* **1992**, *40*, 1403–1408. [[CrossRef](#)]
25. Menkir, A.; Ejeta, G.; Butler, L.; Melakeberhan, A. Physical and chemical kernel properties associated with resistance to grain mold in sorghum. *Cereal Chem.* **1996**, *73*, 613–617.
26. Lee, W.J.; Pedersen, J.F.; Shelton, D.R. Relationship of sorghum kernel size to physiochemical, milling, pasting, and cooking properties. *Food Res. Int.* **2002**, *35*, 643–649. [[CrossRef](#)]
27. Kent, M.A.; Crozier, D.S.; Rooney, W.L. Assessment of kernel characteristics to predict popping performance in grain sorghum. *Crop Sci.* **2022**, *62*, 1051–1059. [[CrossRef](#)]
28. Galli, G.; Horne, D.W.; Collins, S.D.; Jung, J.; Chang, A.; Fritsche-Neto, R.; Rooney, W.L. Optimization of UAS-based high-throughput phenotyping to estimate plant health and grain yield in sorghum. *Plant Phenome J.* **2020**, *3*, e20010. [[CrossRef](#)]
29. Cuevas, J.; Montesinos-López, O.; Juliana, P.; Guzmán, C.; Pérez-Rodríguez, P.; González-Bucio, J.; Burgueño, J.; Montesinos-López, A.; Crossa, J. Deep kernel for genomic and near infrared predictions in multi-environment breeding trials. *G3 Genes Genomes Genet.* **2019**, *9*, 2913–2924. [[CrossRef](#)]
30. DeSalvio, A.J.; Adak, A.; Murray, S.C.; Jarquín, D.; Winans, N.D.; Crozier, D.; Rooney, W.L. Near-infrared reflectance spectroscopy phenomic prediction can perform similarly to genomic prediction of maize agronomic traits across environments. *Plant Genome* **2024**, *17*, e20454. [[CrossRef](#)]
31. Adunola, P.; Tavares Flores, E.; Riva-Souza, E.M.; Ferrão, M.A.G.; Senra, J.F.B.; Comério, M.; Espindula, M.C.; Verdin Filho, A.C.; Volpi, P.S.; Fonseca, A.F.A.d.; et al. A comparison of genomic and phenomic selection methods for yield prediction in *Coffea canephora*. *Plant Phenome J.* **2024**, *7*, e20109. [[CrossRef](#)]
32. Alemu, A.; Åstrand, J.; Montesinos-López, O.A.; Isidro y Sánchez, J.; Fernández-González, J.; Tadesse, W.; Vetukuri, R.R.; Carlsson, A.S.; Ceplitis, A.; Crossa, J.; et al. Genomic selection in plant breeding: Key factors shaping two decades of progress. *Mol. Plant* **2024**, *17*, 552–578. [[CrossRef](#)]

33. Ye, S.; Li, J.; Zhang, Z. Multi-omics-data-assisted genomic feature markers preselection improves the accuracy of genomic prediction. *J. Anim. Sci. Biotechnol.* **2020**, *11*, 109. [CrossRef] [PubMed]
34. Gullifa, G.; Barone, L.; Papa, E.; Giuffrida, A.; Materazzi, S.; Risoluti, R. Portable NIR spectroscopy: The route to green analytical chemistry. *Front. Chem.* **2023**, *11*, 1214825. [CrossRef]
35. Robert, P.; Goudemand, E.; Auzanneau, J.; Oury, F.X.; Rolland, B.; Heumez, E.; Bouchet, S.; Caillebotte, A.; Mary-Huard, T.; Le Gouis, J.; et al. Phenomic selection in wheat breeding: Prediction of the genotype-by-environment interaction in multi-environment breeding trials. *Theor. Appl. Genet.* **2022**, *135*, 3337–3356. [CrossRef]
36. Meyenberg, C.; Braun, V.; Friedrich, C.; Longin, H.; Thorwarth, P. Feature engineering and parameter tuning: Improving phenomic prediction ability in multi-environmental durum wheat breeding trials. *Theor. Appl. Genet.* **2024**, *137*, 188. [CrossRef]
37. Krause, M.R.; González-Pérez, L.; Crossa, J.; Pérez-Rodríguez, P.; Montesinos-López, O.; Singh, R.P.; Dreisigacker, S.; Poland, J.; Rutkoski, J.; Sorrells, M.; et al. Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3 Genes Genomes Genet.* **2019**, *9*, 1231–1247. [CrossRef]
38. Osborne, B.G.; Anderssen, R.S. Single-kernel characterization principles and applications. *Cereal Chem.* **2003**, *80*, 613–622. [CrossRef]
39. Maghirang, E.B.; Dowell, F.E. Hardness measurement of bulk wheat by single-kernel visible and near-infrared reflectance spectroscopy. *Cereal Chem.* **2003**, *80*, 316–322. [CrossRef]
40. Cruickshank, R.H.; Munck, L. It's barcoding Jim, but not as we know it. *Zootaxa* **2011**, *2933*, 55–56. [CrossRef]
41. Brault, C.; Lazerges, J.; Doligez, A.; Thomas, M.; Ecartot, M.; Roumet, P.; Bertrand, Y.; Berger, G.; Pons, T.; François, P.; et al. Interest of phenomic prediction as an alternative to genomic prediction in grapevine. *Plant Methods* **2022**, *18*, 108. [CrossRef]
42. Zhu, X.; Leiser, W.L.; Hahn, V.; Würschum, T. Phenomic selection is competitive with genomic selection for breeding of complex traits. *Plant Phenome J.* **2021**, *4*, e20027. [CrossRef]
43. Montesinos-López, O.A.; Herr, A.W.; Crossa, J.; Montesinos-López, A.; Carter, A.H. Enhancing winter wheat prediction with genomics, phenomics and environmental data. *BMC Genom.* **2024**, *25*, 544. [CrossRef]
44. Parmley, K.; Nagasubramanian, K.; Sarkar, S.; Ganapathysubramanian, B.; Singh, A.K. Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. *Plant Phenomics* **2019**, *2019*, 5809404. [CrossRef] [PubMed]
45. Wang, F.; Feldmann, M.J.; Runcie, D.E. Don't benchmark phenomic prediction against genomic prediction accuracy. *Plant Phenome J.* **2025**, *8*, e70029. [CrossRef]
46. Delwiche, S.R.; Graybosch, R.A.; Nelson, L.A.; Hruschka, W.R. Environmental effects on developing wheat as sensed by near-infrared reflectance of mature grains. *Cereal Chem.* **2002**, *79*, 885–891. [CrossRef]
47. Robert, P.; Brault, C.; Rincet, R.; Vincent, S. Phenomic selection: A new and efficient alternative to genomic selection. In *Genomic Prediction of Complex Traits: Methods and Protocols*; Springer: Cham, Switzerland, 2022; pp. 397–420. [CrossRef]
48. Sapkota, P.; Perumal, R.; Fonseca, J.; Klein, P.; Klein, R.R.; Awika, J.; Rooney, W.L. Heritability, combining ability and genomic selection for grain quality traits in sorghum hybrids across multi-environment trials. *Crop Sci.* **2025**, submitted.
49. Stevens, A.; Ramirez-Lopez, L. An Introduction to the Prospectr Package. 2014. Available online: <https://CRAN.R-project.org/package=prospectr> (accessed on 24 July 2025).
50. Morishige, D.T.; Klein, P.E.; Hilley, J.L.; Sahraeian, S.M.E.; Sharma, A.; Mullet, J.E. Digital genotyping of sorghum—A diverse plant species with a large repeat-rich genome. *BMC Genom.* **2013**, *14*, 448. [CrossRef]
51. Vanraden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef]
52. Acosta-Pech, R.; Crossa, J.; de Los Campos, G.; Teyssèdre, S.; Claustres, B.; Pérez-Elizalde, S.; Pérez-Rodríguez, P. Genomic models with genotype × environment interaction for predicting hybrid performance: An application in maize hybrids. *Theor. Appl. Genet.* **2017**, *130*, 1431–1440. [CrossRef]
53. R Core Team. A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. 2023. Available online: <https://www.r-project.org/> (accessed on 27 July 2025).
54. Pérez, P.; de Los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **2014**, *198*, 483–495. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.