Received: 28 February 2025 | Accepted: 21 October 2025

DOI: 10.1002/ppj2.70051

The Plant Phenome Journal OPEN ACCESS

ORIGINAL ARTICLE

Factors influencing phenomic prediction: A case study on a large sorghum back cross nested association mapping population

Clément Bienvenu[1,2] | **Vincent Garin**[1,2] | **Nicolas Salas**[1,2] | **Korotimi Théra**[3] | **Mohamed Lamine Tekete**[3] | **Madathiparambil Chandran Sarathjith**[4,5] | **Chiaka Diallo**[6,7] | **Angélique Berger**[1,2] | **Caroline Calatayud**[1,2] | **Fabien De Bellis**[1,2] | **Jean-François Rami**[1,2] | **Michel Vaksmann**[1,2] | **Vincent Segura**[2,8] | **David Pot**[1,2] | **Hugues de Verdal**[1,2]

[1]CIRAD, UMR AGAP Institut, Montpellier, France

[2]UMR AGAP Institut, University of Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

[3]Institut d'Economie Rurale, Bamako, Mali

[4]Centre for Water Resources Development and Management (CWRDM), Kozhikode, Kerala, India

[5]KCAEFT Tavanur, Kerala Agricultural University, Malappuram, Kerala, India

[6]Sorghum Program, International Crops Research Institute for the Semi-Arid Tropics, Bamako, Mali

[7]Département d'Enseignement et de Recherche des Sciences et Techniques Agricoles, Institut polytechnique rural de formation et de recherche appliquée de Katibougou, Koulikoro, Mali

[8]Geno-Vigne, IFV-INRAE-Institut Agro, Montpellier, France

Correspondence
Hugues de Verdal, CIRAD, UMR AGAP Institut, Montpellier, F-34398, France.
Email: hugues.de_verdal@cirad.fr

Assigned to Associate Editor Jessica Rutkoski.

Hugues de Verdal, David Pot, and Vincent Segura should be considered joint senior authors.

Funding information
Generation Challenge Programme, Grant/Award Numbers: G7010.05.01, G7010.05.02

Abstract

Plant breeding is crucial to develop varieties able to cope with climate change and support food and feed value chains. Genomic prediction (GP) has been a major step in increasing their efficiency and recently, phenomic prediction (PP) has gained attention as a promising complementary approach to GP, potentially further increasing this efficiency. Factors impacting PP are not fully clarified. Thus, we studied the impacts of spectra preprocessing, prediction methods, population structure, training set size, near infrared reflectance spectroscopy (NIRS) acquisition environment, and wavelength selection on a large multi-parental sorghum population including 2498 BC1F3:5 families from 29 crosses with a strong population structure. Using 51,545 single nucleotide polymorphisms and 1154 NIRS features, we show that PP can reach predictive abilities (PAs) similar to GP, that it is less affected by population structure, and can reach its maximal PA with smaller training sets than GP, but its performances are trait dependent. We also show that NIRS can be acquired in a

Abbreviations: BCNAM, back cross nested association mapping; BLUEs, best linear unbiased estimators; BLUP, best linear unbiased prediction; CV, cross validation; DER2, second derivative; DT, detrend; FLAG, flag leaf appearance; GP, genomic prediction; NIN, number of internodes; NIRS, near infrared reflectance spectroscopy; PA, predictive ability; PAN, panicle length; PCA, principal component analysis; PED, peduncle length; PH, plant height; PP, phenomic prediction; QTL, quantitative trait loci; SNP, single nucleotide polymorphism; STEM, stem length.

reference environment to perform prediction in other environments and that it is possible to randomly select wavelengths to perform predictions. Finally, we show that spectra preprocessing and statistical methods have an inconsistent impact on PA. Our study confirms that PP is a relevant trait prediction method that deserves attention to optimize breeding schemes. The main challenges for the future will be to better understand the information contained in the spectra and disentangle their genetic and proxy components to optimize the use of PP in breeding programs.

Plain Language Summary

Plants grown by farmers are the result of a breeding process. Efficient plant breeding requires the estimation of breeding values retrieved by statistical models and genetic data. Phenomic prediction refers to the use of these models with a new type of data: infrared spectrum, which is easier to collect and less expensive than genetic data. We found that using infrared spectrum allows breeding value estimations sometimes as good as with genetic data, and that it may lead to more robust statistical models. However, phenomic prediction performance is variable and depends on many factors that are not fully understood. Our work is the first of this kind on sorghum, which is a staple crop in West Africa. Sorghum could benefit a lot from phenomic prediction as breeding programs in the south often have limited financial means.

1 | INTRODUCTION

Plant breeding plays a key role in food and feed production, by allowing the development of varieties adapted to the needs and stakes of the value chain stakeholders. Thus, improving breeding programs' efficiency is crucial to guarantee food security especially in the context of climate change (Tester & Langridge, 2010). By referring to the breeder's equation ($\Delta g = \frac{i \times r \times \sigma_A}{L}$; Lush, 1937), a breeding program can be improved by increasing selection intensity (i), better estimating breeding values (r), screening more diversity (σ_A), or reducing the time and cost of field experiments (L).

Genomic prediction (GP) was conceptualized in the late 1990s and is a method to estimate breeding values using only molecular markers, thus reducing the need for expensive field experiments once a predictive model is set (Meuwissen et al., 2001). It has been routinely implemented in a wide range of species to better estimate breeding values and increase selection intensities. Through its two decades of existence, GP has been tested in a variety of contexts, and some key factors impacting its reliability have been studied: genotyping density, genetic architecture of the predicted trait, training set size, relatedness between training and validation sets, statistical methods and modeling, and the importance of genotype-by-environment (G × E) interactions (Lado et al., 2016; Lorenz et al., 2011).

Recently, Rincent et al. (2018) suggested an innovative approach called phenomic prediction (PP) to further reduce the costs of breeding values estimations. PP globally mimics GP but uses near infrared reflectance spectroscopy (NIRS) wavelengths of plant tissues or canopies as predictors instead of molecular markers. PP is low cost, high throughput, non-destructive and often ready to be implemented in breeding programs that already gather infrared spectroscopy data for other uses. Considering a broad definition of PP, the technique has been implemented in mainly three forms: (i) with tissue or plant organ based spectra using a wide range of wavelengths, (ii) with canopy based spectral measurements often done with unmanned aerial vehicles carrying multispectral or hyperspectral cameras covering a narrower range of wavelengths in the visible and near infrared, and (iii) with vegetation indices (VI) also based on canopy measurement but with spectra preprocessing to calculate VIs that play the role of the predictor. Factors impacting these three approaches can differ and this study focuses on definition (i), which corresponds to PP as defined by Rincent et al. (2018).

PPs have been reported on a wide range of plant species, including annual crops such as wheat (Dallinger et al., 2023), maize (Adak et al., 2022), soybean (Zhu et al., 2021), rye (Galán et al., 2020), triticale (Zhu et al., 2022), mungbean (Fumia, Nair, et al., 2023), rapeseed (Roscher-Ehrig et al., 2024), potato (Maggiorelli et al., 2024), pepper (Fumia, Kantar, et al., 2023), and rice (de Verdal et al., 2024), as

well as perennials such as alfalfa (Feng et al., 2020), sugar cane (Gonçalves et al., 2020), coffee (Adunola et al., 2024), grapevine (Brault et al., 2022), and even forest trees such as poplar (Rincent et al., 2018), slash pine (Li et al., 2023), and eucalyptus (Mora-Poblete et al., 2024). PPs have also recently been reported for dairy sheep (Machefert et al., 2024).

Some key factors impacting GP reliability can also impact PP and some of those factors have also been studied for PP but less extensively (training set size [Dallinger et al., 2023; Zhu et al., 2021, 2022], population structure [Laurençon et al., 2024; Roscher-Ehrig et al., 2024; Weiß et al., 2022; Zhu et al., 2021, 2022], statistical method [Brault et al., 2022; Cuevas et al., 2019; Meyenberg et al., 2024; Mora-Poblete et al., 2024; Roscher-Ehrig et al., 2024; Zhu et al., 2021], and genetic architecture of the trait [Robert, Auzanneau, et al., 2022; Roscher-Ehrig et al., 2024; Zhu et al., 2022]). Moreover, some factors specific to PP exist and have already been studied, namely, spectra preprocessing (Brault et al., 2022; Cuevas et al., 2019; Meyenberg et al., 2024) and number of wavelengths used (DeSalvio et al., 2024; Zhu et al., 2021). In addition, contrary to genotyping data, which are fixed across environments for a given genotype, NIRS data capture environmental information. Thus, environment of NIRS acquisition (including phenological stage, tissue, and environmental context among others) and environmental relatedness between training and test sets may also impact the reliability of PP and have also been studied (Mora-Poblete et al., 2024; Rincent et al., 2018; Robert, Goudemand, et al., 2022; Roscher-Ehrig et al., 2024; Zhu et al., 2021, 2022). As spectra contain genetic and environmental information, it should be advantageous to replicate measurements and better isolate the genetic part while also taking advantage of G × E information.

Overall, it has been found that PP can perform similarly to GP in many different species (Adunola et al., 2024; Brault et al., 2022; Dallinger et al., 2023; Rincent et al., 2018; Robert, Auzanneau, et al., 2022; Robert, Goudemand, et al., 2022; Roscher-Ehrig et al., 2024; Thapa et al., 2024; Weiß et al., 2022; Zhu et al., 2021, 2022) and that statistical methods do not seem to impact PP predictive abilities (PAs) (Roscher-Ehrig et al., 2024; Zhu et al., 2021), but nonlinear methods and deep learning algorithms may provide some improvement over more classical linear models (Cuevas et al., 2019; Mora-Poblete et al., 2024). PP is also less sensitive than GP to population structure (Laurençon et al., 2024; Roscher-Ehrig et al., 2024; Weiß et al., 2022; Zhu et al., 2021, 2022), and it requires fewer genotypes in the training population to reach its maximal PA (Dallinger et al., 2023; Zhu et al., 2021, 2022). The quantitative nature of the trait (high/low number of QTL [quantitative trait loci]) seems to affect PP PA, with better performance for complex traits such as yield compared to GP (Roscher-Ehrig et al., 2024; Zhu et al., 2022). It also seems that selecting wavelength based on least absolute shrinkage and selection operator (LASSO) weights or partial least square regression (PLSR) loadings can greatly reduce the number of wavelengths needed with little impact on PA (DeSalvio et al., 2024; Zhu et al., 2021). One hypothesis to explain these performances of PP is that each wavelength would capture the effects of many genes at the same time (Zhu et al., 2021). Thus, one wavelength would be more informative than one single nucleotide polymorphism (SNP), which is likely due to their quantitative information as opposed to the binary information of SNPs. From this statement, it seems possible that PP can perform as high as GP with only a few dozen of wavelengths as they already cover a large part of the genome (Zhu et al., 2021). It is also coherent with the fact that PP can outperform GP on complex traits but not on Mendelian traits. Indeed, by capturing the effects of many genes at the same time, wavelengths can also capture non additive information such as epistasis or G × E interaction. As mono/oligogenic traits are driven by a few QTL with strong additive effects explaining a large part of the phenotypic variance, PP could differentiate lines differing at few loci only if they directly affect a wavelength of the spectrum. On the other hand, traits less dependent on additive effects could be better predicted by PP (Zhu et al., 2022). Finally, this hypothesis can also explain the need for smaller training sets, as more information is carried by each spectrum, less spectra are required to train a model (Zhu et al., 2022).

More specifically to PP, NIRS environment acquisition has an impact on PA, with a drop of performance when spectra of the training and validation sets are not from the same environment (Zhu et al., 2021), likely due to the environmental, G × E, and endophenotype-like information in the spectra. Combining spectra from different environments, or tissue may improve prediction abilities (Rincent et al., 2018; Robert, Auzanneau, et al., 2022; Robert, Goudemand, et al., 2022)

Core Ideas

- Phenomic prediction works on sorghum but is highly trait dependent.
- Phenomic prediction is less affected than genomic prediction by training set size and population structure.
- Near infrared reflectance spectroscopy (NIRS) data can be acquired in a reference environment to predict genotypes phenotyped in another environment.
- A dozen of randomly selected wavelengths can be sufficient to reach the same predictive abilities as the full spectra in certain traits.
- It is likely that spectral information is poorly used by geneticists' models.

maybe because different tissues/environment of acquisition have different gene expression profiles and can complement each other's missing genetic information. It has also been shown that having spectra in a reference environment to predict genotypes' performances in other environments is feasible (Brault et al., 2022; Rincent et al., 2018; Roscher-Ehrig et al., 2024; Zhu et al., 2021, 2022) which supports the hypothesis that NIRS can capture genetic information and relatedness between genotypes. Lastly, PP and GP have been tested in many scenarios relevant for breeding programs. PP yielded its best results for sparse testing, poorest results for predicting unknown genotypes in unknown environments and intermediate and contrasted results across studies for unknown genotypes in known environments and known genotypes in unknown environments (Adak et al., 2022; Adunola et al., 2024; DeSalvio et al., 2024; Lane et al., 2020; Robert, Goudemand, et al., 2022).

Sorghum (*Sorghum bicolor*) is the fifth most important cereal in the world after corn, wheat, rice, and barley with a total production of 57 million metric tons produced in 2022. Africa is the biggest producer accounting for 51% of the world production (FAOSTAT, 2024). It is a very versatile crop as its use include food (raw and processed), cattle feed (grain and fodder), and energy production (as a bioethanol source, mostly in the United States) and even phytoremediation of cadmium-contaminated soils (Liu et al., 2020). Moreover, its tolerance to drought, low input levels, or salinity makes it a very interesting crop to cope with climate change (Hossain et al., 2022). Despite being an important crop, PP has not been investigated much on this species, which could benefit a lot from low-cost breeding tools as sorghum is mainly produced in developing countries. Pioneer work on this species have been initiated by Galli et al. (2020) who tested PP on sorghum using drone imagery and VI as predictors of yield and anthracnose sensitivity, studying the impacts of combining different VI, different flight dates, and comparing results of neural network and PLSR models.

The present study was conducted on a large (multi parental) sorghum back cross nested association mapping (BCNAM) population (Garin et al., 2024) from which a subset of two recurrent parents and 22 donor parents grown during 2 years in four environments was considered. The strong genetic structure of the population, and the several environments in which it was phenotyped make it particularly suited to study the impact of population structure and NIRS environment acquisition on PP PA. Moreover, multi parental populations are relevant for breeding programs but have not been frequently used in PP studies. In this context, we investigated the potential of PP for sorghum breeding by studying the factors impacting PP: training set size, population structure, NIRS acquisition environment, wavelength selection, spectra preprocessing, and statistical method.

2 | MATERIALS AND METHODS

The plant material and phenotypic and genomic data acquisition were described in Garin et al. (2024). We briefly recall those points in the following sections. Thus, "Section 2.1," "Section 2.2," and "Section 2.3" are taken from Garin et al. (2024) with adaptations to fit the analysis made in this study. The general framework of the study is schematized in Figure 1.

2.1 | Plant material

The West and Central Africa Back Cross Nested Association Mapping population is composed of three populations obtained by crossing three elite recurrent parents to 24 donor parents, backcrossing their F_1 hybrids to obtain BC_1 families which were then selfed for three generations to obtain BC_1F_4 genotypes. All data were acquired on $BC_1F_{3:4}$ generations. NIRS data were acquired in only two of these populations, which were retained in this study: Grinkan (GR) and Kenin-Keni (KK). The recurrent parents are elite lines selected in Mali through farmer variety testing. GR was developed through pedigree breeding methods. KK was derived from a recurrent selection population involving local parents of different botanical types (Leroy et al., 2014). These two recurrent parents were chosen for their productivity, their adaptation to soil and climate, and their resistance to major biotic and abiotic stresses, but they have poor grain quality and mold susceptibility (GR), and low yield and yield stability (KK). The 24 donor parents cover diverse racial (Guinea, Caudatum, and Durra [Harlan & de Wet, 1972]) and geographical origins. They are characterized by key adaptive traits like height, maturity, and photoperiod sensitivity. Those parents were also selected for traits like tolerance to *Striga hermonthica*, soil phosphorus deficiency and/or drought, and good grain quality that could increase farmer acceptance (see Garin et al., 2024, tab. 1).

2.2 | Phenotypic data

Genotypes were phenotyped during 2 years (2012 and 2013) in Mali at two sites: Sotuba (12.65 lat.; −7.93 long.) and Cinzana (13.25 lat.; −5.96 long.) with two sowing dates (Sowing 1: end of June, Sowing 2: 3–4 weeks later) for a total of four environments per year. In each environment, the genotypes were laid out as an augmented block design using the recurrent parents as checks. Each unit plot consisted in a line of 10 plants spaced by 0.4 m, and lines (unit plots) were separated by 0.75 m. Because of logistic constraints, all crosses could not be grown in one season. Therefore, some of them were grown in 2012 and others in 2013, each year containing both

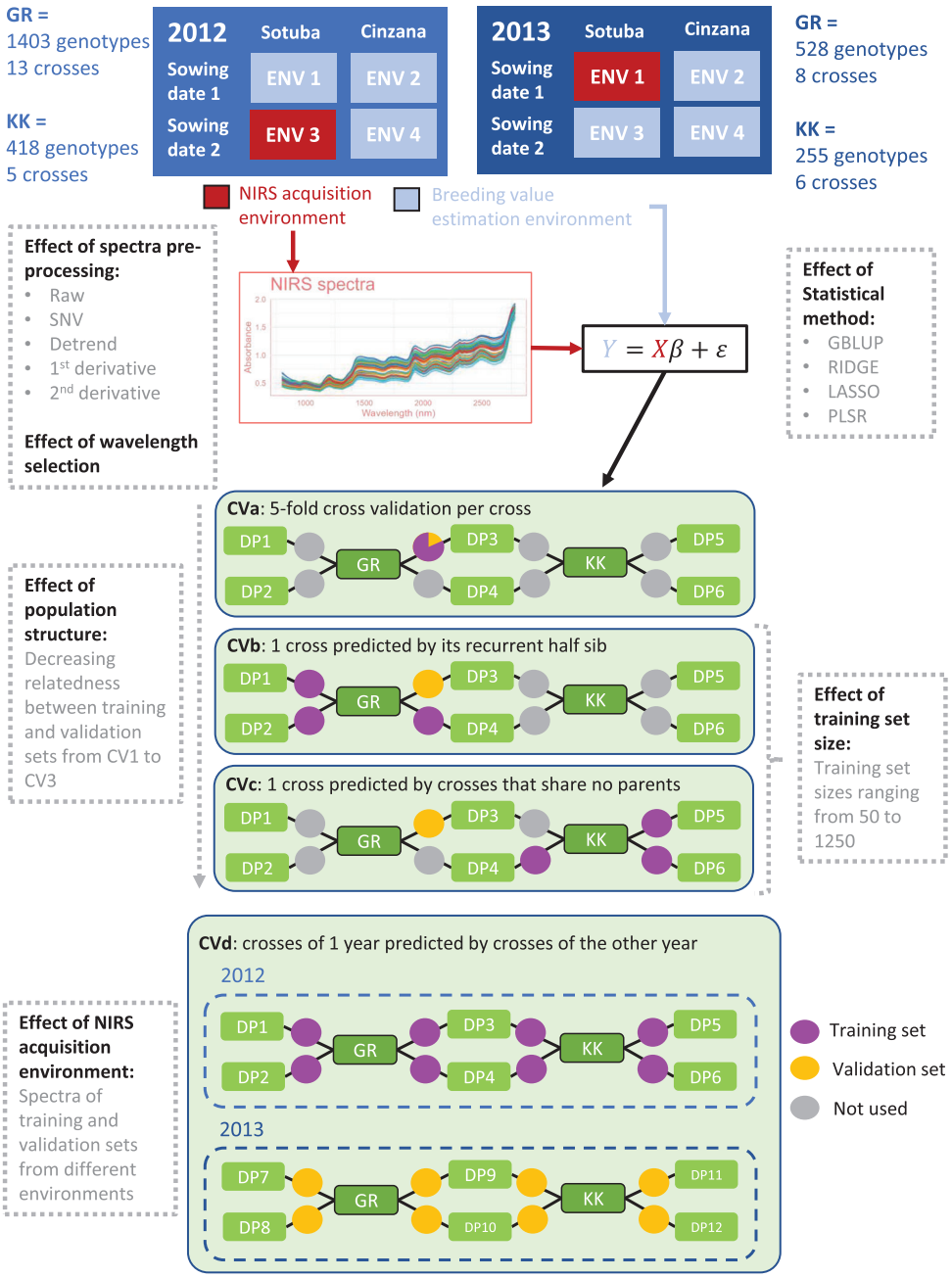

FIGURE 1 Schematic representation of the data, the cross-validation scenarios, and the factors studied. DP, donor parent; GR, Grinkan; KK, Kenin Keni.

KK and GR subpopulations (only donor parents were different across years). For each year, only genotypes present in all four environments were kept. In 2012, 1821 $BC_1F_{3:5}$ families derived from 18 crosses were kept (GR: 13 crosses and 1403 families, KK: five crosses and 418 families), while in 2013, 783 $BC_1F_{3:5}$ families derived from 14 crosses were kept (GR: eight crosses and 528 families, KK: six crosses and 255 families). In total, 2498 families from 29 crosses were considered across years (GR: 1848 families, 19 crosses; KK: 650 families, 10 crosses) and only 106 families belonging to three crosses were common to the 2 years.

Seven traits were manually measured: flag leaf appearance (FLAG; CO_324:0000631) as the number of days after sowing when half of the plot had their ligulated flag leaves visible, peduncle length (PED; CO_324:0000622) as the distance in cm between the final node and the panicle bottom, panicle length (PAN; CO_324:0000620) as the distance in cm from the end of the peduncle to the panicle top, stem length (STEM; CO_324:0000627) as the distance in cm from the soil to the top of the stem, plant height (PH; CO_324:0000623) as the distance in cm between the soil and the panicle top, number of internodes (NIN; CO_324:0000605), and grain yield

(YIELD; CO_324:0000403) in kg.ha^{-1} at the plot level by harvesting all panicles from a plot (except the border) drying, hand threshing them, and weighting the grains. All traits except FLAG were measured at harvest.

2.3 | Genotypic data

The 2498 families and their parents were genotyped using genotyping by sequencing (GBS; Elshire et al., 2011) with 384-plex libraries on an Illumina HiSeq 2000 sequencer. The offspring were genotyped at generation BC_1F_3. The sequence data were analyzed running the reference genome-based TASSEL GBS pipeline (Glaubitz et al., 2014). Unique tags (3,844,911) were aligned on the sorghum reference genome v2.1 (Paterson et al., 2009). After the filtering of raw genotype data for minor allele frequencies (<0.05) and single marker missing data (<0.9), 51,545 segregating SNPs were identified among the parents with between 11,856 and 26,128 SNPs segregating in the individual crosses. Missing values in the parents were imputed using Beagle (Browning et al., 2018). Missing values in the offspring genotypes were imputed using FSFHap (Bradbury et al., 2007). We determined a unique genetic consensus map by projecting the physical distance of the 51,545 markers on a high-quality genetic consensus map (Guindo et al., 2019) using the R package ziplinR (https://github.com/jframi/ziplinR).

2.4 | NIRS data

For each genotype in each year, spectra were acquired in one environment (one site and one sowing date) on sun-dried grain samples after harvest (Figure 1). Spectra were obtained with a MPA Bruker FT-NIR spectrometer (Bruker Optics Inc.) for 1154 wavelengths ranging from 800 to 2800 nm. For each sample, the average of three technical repetitions was used and outlier spectra were removed from the data set afterward.

To assess the effect of spectra preprocessing, smoothing using Savitsky-Golay procedure was applied (filter length of 11, polynomial order of 1, and derivation order of 0) and four classical chemometric pretreatments were applied to the smoothed spectra: First derivative (DER1), second derivative (DER2), standard normal variate (SNV), and detrend (DT).

2.5 | Estimation of phenotypic breeding values

Due to the low number of common genotypes between years, 2012 and 2013 data were analyzed separately. For each year, breeding values were estimated with the following model computed with the asreml R package (Butler, 2021):

$$\begin{aligned} \underline{Y_{ijklm}} = \mathrm{CROSS}_i + \mathrm{GENO}_{(i)j} + \underline{\mathrm{SOW}_k} + \underline{\mathrm{SITE}_l} \\ + \underline{\mathrm{CROSS:SOW}_{ik}} + \underline{\mathrm{CROSS:SITE}_{il}} \\ + \underline{\mathrm{GENO:SOW}_{(i)jk}} + \underline{\mathrm{GENO:SITE}_{(i)jl}} \\ + \underline{\mathrm{BLOCK}_{(kl)m}} + \underline{\varepsilon_{ijklm}} \end{aligned}$$

where Y_{ijklm} are the phenotypes; CROSS_i is the fixed effect of the cross i; $GENO_{(i)j}$ is the fixed effect of the genotype j nested into cross j; $\mathrm{SOW} \sim N(0, I\sigma^2_{\mathrm{sow}})$ is the random effect of sowing date; $\mathrm{SITE} \sim N(0, I\sigma^2_{\mathrm{SITE}})$ is the random effect of site; $\mathrm{CROSS} \times \mathrm{SOW} \sim N(0, I\sigma^2_{\mathrm{crso}})$ is the random effect of the interaction between cross and sowing date; $\mathrm{CROSS} \times \mathrm{SITE} \sim N(0, I\sigma^2_{\mathrm{crsi}})$ is the random effect of the interaction between cross and site; $\mathrm{GENO} \times \mathrm{SOW} \sim N(0, I\sigma^2_{\mathrm{geso}})$ is the random effect of the interaction between genotype and sowing date; $\mathrm{GENO} \times \mathrm{SITE} \sim N(0, I\sigma^2_{\mathrm{gesi}})$ is the random effect of the interaction between genotype and site; $\mathrm{BLOCK} \sim N(0, V_b)$ is the random effect of the blocks nested into site–sow combinations, V_b being a diagonal matrix containing the variance components $\sigma^2_{(kl)b}$ of $\mathrm{BLOCK}_{(kl)m}$ at each level of site–sow combination, that is, one block variance per site–sow combination; and $\varepsilon \sim N(0, R)$ is the residual, R being a diagonal matrix containing the variance components $\sigma^2_{(kl)r}$ of ε_{ijklm} at each level of site–sow combination, that is, one residual variance per site–sow combination. The breeding values were estimated as the sum of the cross and genotype BLUEs (best linear unbiased estimators). To minimize the effect of direct relationships between spectra and phenotypes and thus getting closer to information brought by genetic signal, phenotypic data from the NIRS acquisition environment (i.e., one site–sow combination) were not included in breeding values estimations (Figure 1). As there were only two sowing dates and two sites, the removal of one environment hindered the specification of a SITE:SOW interaction term in the model. This may impact the precision of estimated breeding values, but is relevant to avoid a proxy-like prediction mechanisms and biases in PA estimation due to the covariance between breeding values and environmental information in the spectra as pointed out by Wang et al. (2025).

2.6 | Variance components and heritability

Heritabilities were calculated according to the method proposed by Schmidt et al. (2019) to account for the unbalanced and incomplete design of the experiment. Calculations were done using codes provided in the article (https://github.com/PaulSchmidtGit/Heritability) with the asreml R package (Butler, 2021). This method is a pairwise heritability calculation, meaning that it is based on entry differences rather than entry means and provides one heritability estimation for each possible pair of genotypes. It is the squared correlation between the

true genotypic values differences and the estimated/predicted genetic values differences for each possible pair of genotypes. For a pair of genotypes i and j:

$$H_{ij}^2 = \text{cor}(g_i - g_j, \hat{g}_i - \hat{g}_j)^2 = \frac{2\sigma_g^2 - v_{ij}^{\text{BLUP}}}{2\sigma_g^2}$$

where σ_g^2 is the genetic variance, and v_{ij}^{BLUP} is the prediction error variance of a difference between BLUPs (best linear unbiased prediction) of genotypes i and j. The parameters needed to calculate the heritability were estimated with the breeding values estimation model but considering the genotypic effect as the only genetic effect (i.e., no cross effect) considered random with GENO $\sim N(0, I\sigma_g^2)$. The heritabilities reported for each trait are the mean of heritabilities across pairs of genotypes (i.e., BCF3:5 families in our case) (Schmidt et al., 2019):

$$\overline{H_{\cdot\cdot}^2} = \frac{2}{n(n-1)} \sum_i \sum_{j<i} H_{ij}^2$$

2.7 | Prediction models

Estimation of prediction accuracies were performed in R (R Core Team, 2023) using four different models (Figure 1): GBLUP as a reference with the package sommer (Covarrubias-Pazaran, 2016), PLSR with the package rchemo (Brandolini-Bunlon et al., 2023), ridge regression (RIDGE) and LASSO with the package glmnet (Friedman et al., 2010). GBLUP (genomic best linear unbiased prediction) was used with both genotypic and NIRS data as a reference as there is generally little difference between models and that it is one of the simplest, of the most computationally efficient, with no parameter tuning, and that it has competitive predictive performances (Lourenço et al., 2024), while the other models were only used with NIRS data.

2.7.1 | GBLUP model

$$Y_i = \underline{G_i} + \varepsilon_i$$

where Y_i is the breeding values as calculated in Model 1; $G \sim N(0, \sigma_g^2 M)$ is the random effect of genotype, σ_g^2 being the genetic variance and M being the kinship matrix designated as K when it was based on genotypic information or the pseudo-kinship matrix designated as H when it was based on NIRS information; and $\varepsilon_i \sim N(0, I\sigma^2)$ iid is the residual, σ^2 being the residual variance.

The genotypic kinship matrix K was calculated with the Van Raden method (VanRaden, 2008), and the hyperspectral pseudo-kinship matrix H was calculated as $H = SS^t/n$ with S being the scaled and centred spectral matrix, S^t being the transpose matrix of S, and n being the number of wavelengths (Robert, Auzanneau, et al., 2022).

2.7.2 | RIDGE and LASSO models

$$Y_i = \sum_{k=1}^{p} x_{ik}\beta_k + \varepsilon_i$$

where marker effects β are estimated by minimizing the following loss function:

$$\hat{\beta} = \text{Argmin}_{\beta \in \mathbb{R}^p} \left\{ \|Y - X\beta\|^2 + \lambda . \left(\frac{1-\alpha}{2} \|\beta\|_2^2 + \alpha . \|\beta\|_1 \right) \right\}$$

where constraint $\|\beta\|_1 = \sum_{k=1}^{p} |\beta_k|$, $\|\beta\|_2^2 = \sum_{k=1}^{p} \beta^2$, and $\lambda > 0$, with Y_i being the breeding values of the genotypes i based on Model (1), x_{ik} is the absorbance at wavelength k of genotype i, β_k is the effect of the wavelength k, and $\varepsilon_i \sim N(0, \sigma^2)$ iid is the residual. LASSO is obtained with $\alpha = 1$ and RIDGE is obtained with $\alpha = 0$. The value of λ was optimized by internal 10-fold cross validation (CV) for every model that was computed with built-in functions of the glmnet package.

2.7.3 | PLSR model

PLSR is based on two principal component analysis (PCA)-like decompositions of the data into latent variables:

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E}$$

$$Y = \boldsymbol{U}\boldsymbol{Q}^T + F$$

where X is the $n \times m$ spectral matrix (n being observations and m being predictors i.e., wavelengths). Y is the $n \times 1$ vector of the response variable (breeding values). T and U are $n \times l$ matrices (for l latent variables) and are respectively the scores (i.e., projected values of individuals on the latent variables) for X and Y matrices. They are estimated so that the covariances between column i of U and column i of T are maximized while covariances between column i of U and column j of T (with $i \neq j$) are zero (i.e., latent variables of T and U are built to be orthogonal). P and Q are $m \times l$ matrices and are, respectively, the loading matrices of X and Y. E and F are, respectively, the error terms of X and Y assumed to be independent and identically distributed normal random variables.

Predictions of Y are obtained by calculating the scores (T matrix) of the validation X matrix on the latent variables

estimated, then estimating the scores of the unknown Y (U matrix) matrix through covariances between U and T matrices and then calculating the estimated values of Y through the estimated loadings of Y (Q matrix).

Predictions were tested with several latent variables ranging from one to 20. Final results were obtained with 10 latent variables as this number maximized PA and minimized root mean square error of prediction overall.

2.8 | CV scenarios

To assess the effects of population structure on prediction accuracy, three CV scenarios were used (Figure 1). Those scenarios are comparable to the ones in Lehermeier et al. (2014):

- CVa: intra cross scenario, where a fivefold CV is performed using training and validation set data coming from a single cross.
- CVb: intra recurrent parent leave one cross out scenario, where one cross in the validation set is predicted by all the crosses that share the same recurrent parent in the training set.
- CVc: inter recurrent parent scenario, where a cross of one population (e.g., GR) is predicted using the crosses of the other population (KK), excluding the crosses with a shared parent.

Training set sizes of CVa were dependent on the specific cross size and ranged from 40 to 108 genotypes. CVb and CVc scenarios were also used to assess the effect of training set size on prediction accuracies. In these scenarios, different training set sizes were set by increasing the number of genotypes from 50 to 1250 with a step of 50. For each training set size, sampling was made so that the structure of the population was respected (i.e., the proportion of genotypes belonging to each cross was the same as in the total population in all training sets). To study the other impacting factors, the CVb scenario was used with training sets sizes of 1000 genotypes in 2012 to have a large training set and keep a reasonable computing time, and 450 in 2013 to have the largest possible training set given the lower number of genotypes in 2013. For each scenario and different population size within scenario (CVb and CVc), 10 replications were carried out.

An additional scenario (CVd) was developed to study the impact of NIRS acquisition environment (Figure 1): inter-year scenario with crosses from 1 year predicted by a model trained on crosses from the other year so that spectra from the training and validation set were acquired in different environments and different genotypes. CVd was trained with data collected in 2012 and validated with data from 2013 (CVd_2013) and vice versa (CVd_2012). CVd_2012 has no practical use in breeding but mimics the same situation as CVd_2013 where new genotypes are predicted in a new environment, which is relevant. Even though training and validation sets share no genotypes, crosses from both recurrent parents are present in both sets. For these scenarios, 30 repetitions were made, and training set size was set to 600, respecting the structure of the population, to have the biggest training set size possible given the number of genotypes in 2013. Table S1 is a summary of scenarios and training set sizes used to study each factor.

2.9 | CDmean and Fst calculation

To quantify relatedness between training and validation sets in CVa, CVb, and CVc, CDmean and Fst indicators were calculated. Using polymorphism data, Fst estimates the genetic differentiation between two populations by comparing the average number of different base pairs between two individuals coming from the same population or coming from the two populations to be compared. If the differences between two individuals coming from the same population and two individuals coming from different populations are similar, the two populations are not differentiated, and on the contrary, if these differences are different, the populations are differentiated. CDmean is an indicator based on the GBLUP model. It uses the generalized coefficient of determination (CD), which is the squared correlation between the true and the predicted contrast of genetic values for a given contrast. For CDmean calculation, the contrast used is the difference between the genetic value of an individual and the mean of the whole population (training + and validation sets). Given this contrast, CDmean is the mean of all the CDs of the individuals from the validation set given a training set. To summarize, CDmean estimates the reliability of the predictions of contrasts between genetic values of the validation set given a training set using a GBLUP model, thus accounting for population structure through the kinship matrix used in GBLUP.

Fst values were calculated with the R package BEDASSLE (Bradburd, 2024), which estimates Fst according to Weir and Hill (2002). CDmean values were calculated according to Rincent et al. (2012) with R codes given in the TrainSel package (Akdemir et al., 2021) documentation. Then, linear regressions between PA and Fst or CDmean were carried out, and the effects of these indicators on PA were tested with a Student test on the slopes of the regressions.

2.10 | Wavelength selection

To study the effect of wavelength selection, three steps were implemented. First, LASSO and PLSR models were used to estimate the importance of each wavelength in prediction with the loadings of the PLSR and the weights of the LASSO. Then, wavelengths were selected based on these loadings

and weights. Finally, the selected wavelengths were used in GBLUP models and their results were compared with classical PP using all wavelengths. Training and validation sets were not the same in the LASSO/PLSR and in the GBLUP model to avoid overfitting and inflated PAs that typically occur when selecting variables (Krstajic et al., 2014; Utz et al., 2000). PLSR and LASSO models were trained with CVb training sets with 2012 and 2013 data with 450 genotypes, which correspond to the largest training set that could be composed using 2013 data. According to these analyses, the loadings of the first latent variable of PLSR and the weights attributed to each wavelength in LASSO were retrieved. The mean loadings and weights were calculated for each year. The selection of wavelength was based on the local maxima and minima of the mean loadings and maxima of the weights. The selected wavelengths were then used in a GBLUP model in CVb scenario with training set sizes of 1000 for 2012 data and 450 for 2013 data. See Figure S1 for a schematic representation.

Another method of wavelength selection was random selection. The goal of this approach was to determine whether a statistical method is really needed and useful to identify the most informative wavelengths, or if the NIRS information is redundant enough to be reduced to a few wavelengths without having to select wavelengths precisely. In this method, 10 samples from 1 to 10, 20, 30, 40, 50, and 60 wavelengths were randomly chosen (with a computer random number generator) to calculate the kinship matrix. Each sample was then used in predictions using the CVb scenario with training set sizes of 1000 for 2012 data and 450 with 2013 data.

All analysis related to wavelength selection were performed on raw spectra to mimic at best what could be achieved with cheaper sensors measuring few wavelengths as preprocessing the same wavelength with 1154 or less than 60 wavelengths would not yield the same spectral values.

2.11 | Models' accuracy metrics

The PA of all models were calculated with the Pearson's correlation coefficient between observed and predicted breeding values. As CV scenarios were based on a leave one cross out (or alike) methods, PAs were always calculated within cross for each scenario. Thus, the results correspond to the within cross PA metrics average over the different crosses. Significant differences between PA were tested using pairwise Wilcoxon tests with a Bonferroni correction if not mentioned otherwise.

2.12 | Graphics

All graphics were produced with the ggplot2 package (Wickham, 2016) of R (R Core Team, 2023), or with PowerPoint.

3 | RESULTS

3.1 | Description of phenotypic data

Coefficients of variation for each trait in each environment are available in Table S2. The trait with the highest variability was YIELD; followed by PED, STEM, and PH; followed by NIN and PAN; followed by FLAG which was the least variable trait. For PAN, PED, PH, FLAG, and NIN, coefficients of variation were stable across environments and years. STEM was more variable for 2012 than 2013 overall. YIELD was more variable in 2013 than 2012 overall except for SB1 in 2012 where the highest coefficient of variation (0.7) occurred. A complete description of means and standard deviations for each trait in each environment is available in Table S3.

Correlations between environments for each trait are presented in Figure S2. PED was the trait with the highest correlation between environments ranging from 0.62 to 0.83. YIELD was the trait with the lowest correlations between environments, ranging from 0.06 to 0.39. FLAG, PH, and STEM had medium to high correlations between environments, ranging from 0.52 to 0.83. NIN had low to medium correlation between environments, ranging from 0.2 to 0.47. Finally, PAN had medium-high correlation between environments in 2013 ranging from 0.65 to 0.67, and medium-low correlations in 2012, ranging from 0.39 to 0.52.

3.2 | Variance components of phenotypic traits and spectra

Partitions of variance for each trait are presented in Figure 2. In 2012, PH-related traits (PED, STEM, and PH) and FLAG were the most heritable with heritability ranging from 0.73 to 0.8. NIN and PAN were less heritable with heritabilities of 0.52 and 0.47, respectively. YIELD was the least heritable trait at 0.3. In 2013 FLAG, PH, and STEM were the most heritable traits with heritabilities from 0.80 to 0.83, respectively. NIN, PAN, and PED were less heritable with values of 0.57, 0.61, and 0.68, respectively. Finally, YIELD had the lowest heritability of 0.4. For NIRS wavelengths, heritability estimates were stable across the spectra with a severe drop around 2700 nm likely due to these wavelengths being close to the upper limit of the wavelength range of the spectrometer. Heritability values were close between years with mean heritabilities of 0.52 and 0.46 for 2012 and 2013, respectively. There were no noticeable effects of the interactions between the genetic variables (GENO and CROSS), and experimental design variables (SITE and SOW) except for YIELD in 2012. Sowing date was responsible for a large part of the variances of FLAG, NIN, PH, STEM, and YIELD in both years. The site did not impact FLAG, PAN, PED, and YIELD in 2012, but had a large impact on NIN this year. The impact of site

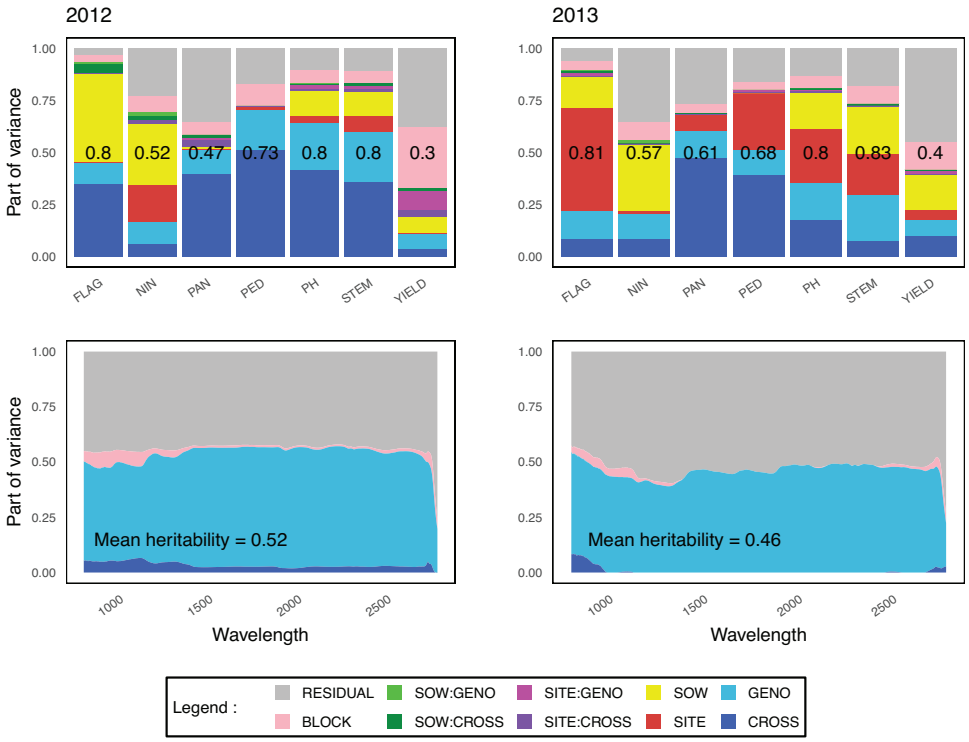

FIGURE 2 Partitions of variance for the studied traits and wavelengths along spectra in 2012 and 2013. RESIDUAL, ENV, CROSS, and GENO stand for the factors contributing to the phenotypic variance. Numbers in black are heritability of traits.

was more pronounced in 2013, being responsible for a large part of the variances of FLAG, PED, PH, and STEM.

3.3 | Overall comparison of GP and PP

Figure 3 presents PAs of genomic and PPs for raw and preprocessed spectra in the CVb scenario with 1000 genotypes in the training set in 2012 and 450 in 2013. In 2012, for FLAG, NIN, and PAN, PP and GP performed similarly except for PP with spectra preprocessed with DER2 for NIN and PAN, and spectra preprocessed with DT for PAN. For PH-related traits (PED, PH, and STEM) and YIELD, PP systematically had lower accuracy than GP, with DER2 preprocessing and raw spectra having the worst results for PH-related traits.

Slightly contrasting results were observed in 2013. For this year, PP and GP performed similarly for PH and STEM. PP consistently outperformed GP for FLAG, NIN, and YIELD, but GP consistently outperformed PP for PED. In addition, PP was not able to predict PAN for this year.

3.4 | Effect of spectra preprocessing

In 2012 (Figure 3), DER2 lead to significantly lower PAs than other preprocessing methods and results for all other pretreatments were similar overall with few significant differences except for PED where results were more contrasted. In 2013, significant differences between preprocessing were found only for PED where raw spectra outperformed preprocessed spectra. In 2012, the maximum difference between the highest and lowest PA was of 0.2, and occurred for PED. In 2013, the maximum difference between the highest and lowest PA was of 0.14, and also occurred for PED. As raw spectra had no difference with preprocessed spectra most of the time and was sometimes better than preprocessed spectra, all following analyses were carried out using raw spectra.

3.5 | Effect of statistical method

Figure 4 presents PAs for PP obtained from four statistical models: GBLUP, RIDGE, LASSO, and PLSR. PAs were obtained in the CVb scenario with a training set size of 1000 genotypes in 2012 and 450 in 2013. In 2012, no effect of statistical method was observed for NIN. RIDGE gave significantly lower PAs for all other traits, and LASSO gave significantly lower PAs for PED, PH, and STEM. The highest difference was found between PLSR and RIDGE for PH, being 0.18. In 2013, no significant differences were found between models for FLAG and NIN. RIDGE gave significantly lower PAs for all other traits, and LASSO gave significantly lower PAs for PH and YIELD. The highest difference was found between PLSR and RIDGE for YIELD and correspond to

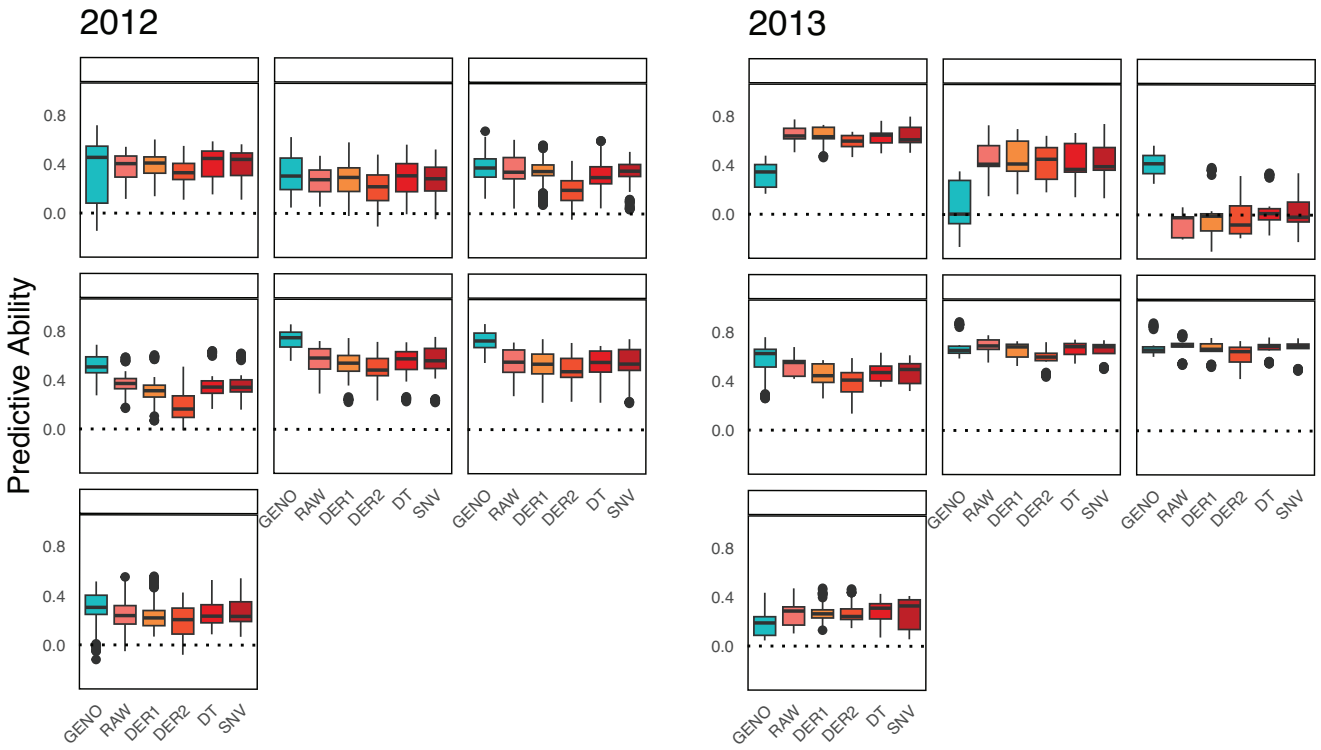

FIGURE 3 Genomic prediction and effects of spectra preprocessing on phenomic prediction predictive abilities. Predictive abilities were calculated using GBLUP (Genomic Best Linear Unbiased Prediction) in the CVb scenario with a training set size of 1000 in 2012 and 450 in 2013. Letters represent the results of pairwise Wilcoxon tests corrected by Bonferroni method for a threshold of 5%. GENO denotes genomic prediction, RAW denotes phenomic prediction with raw spectra, DER1 and DER2 denotes phenomic prediction with first and second derivative of the spectra, DT denotes phenomic prediction with spectra processed with detrend, and SNV denotes phenomic prediction with spectra processed with the standard normal variate method. Horizontal dotted line is at the 0 value. FLAG, flag leaf appearance; NIN, number of internodes; PAN, panicle length; PED, peduncle length; PH, plant height; STEM, stem length; YIELD, yield.

0.17. GBLUP and PLSR consistently gave similar results in both years. Overall, RIDGE and LASSO seem to give lower PAs in more situations, but the effect of statistical model are inconsistent across traits. Further analyses were carried out using the GBLUP reference model.

3.6 | Effect of training set size

The evolution of PAs according to the training set size for the CVb scenario in 2012 using a GBLUP model is presented in Figure 5. For NIN, PH, and STEM, PP reached a plateau of PAs around 400 genotypes in the training set while PAs of GP continued increasing. For PED, FLAG, and PAN, the plateau was reached for PP around 700 genotypes in the training set, while PAs of GP also continued to increase. For YIELD, GP reached a plateau of PAs around 700 genotypes in the training set while for PP it happened at 1100 genotypes. The effect of training set size was studied with 2013 data (not shown) but the lower number of genotypes available (maximum training set size of 450) was not enough to draw conclusions. The effect of training set size was also studied with CVc scenario with 2012 data (Figure S3). In this scenario, PAs of GP and PP had similar evolution for PH, STEM, and YIELD, PAs of PP being superior for PH and STEM, and equivalent for YIELD. For PED, NIN, and PAN, PP reached a plateau of PA around 400 genotypes in the training set while PAs of GP continued to increase, PAs of PP being superior for PED and NIN. For FLAG, PAs of GP increased with the training set size, while PAs of PP were stable and even slightly decreased for large training sets. Standard error around the mean is very large in this scenario making the analysis less relevant.

3.7 | Effect of population structure

Figure 6 presents the PAs for CVa, CVb, and CVc scenarios for a training set size of 100 (which is the maximum training set size available in CVa) using a GBLUP model on 2012 data. The BCNAM population is strongly structured as can be seen by a PCA on SNP or spectral data (Figure S4). Thus, it is relevant to build prediction scenarios with varying

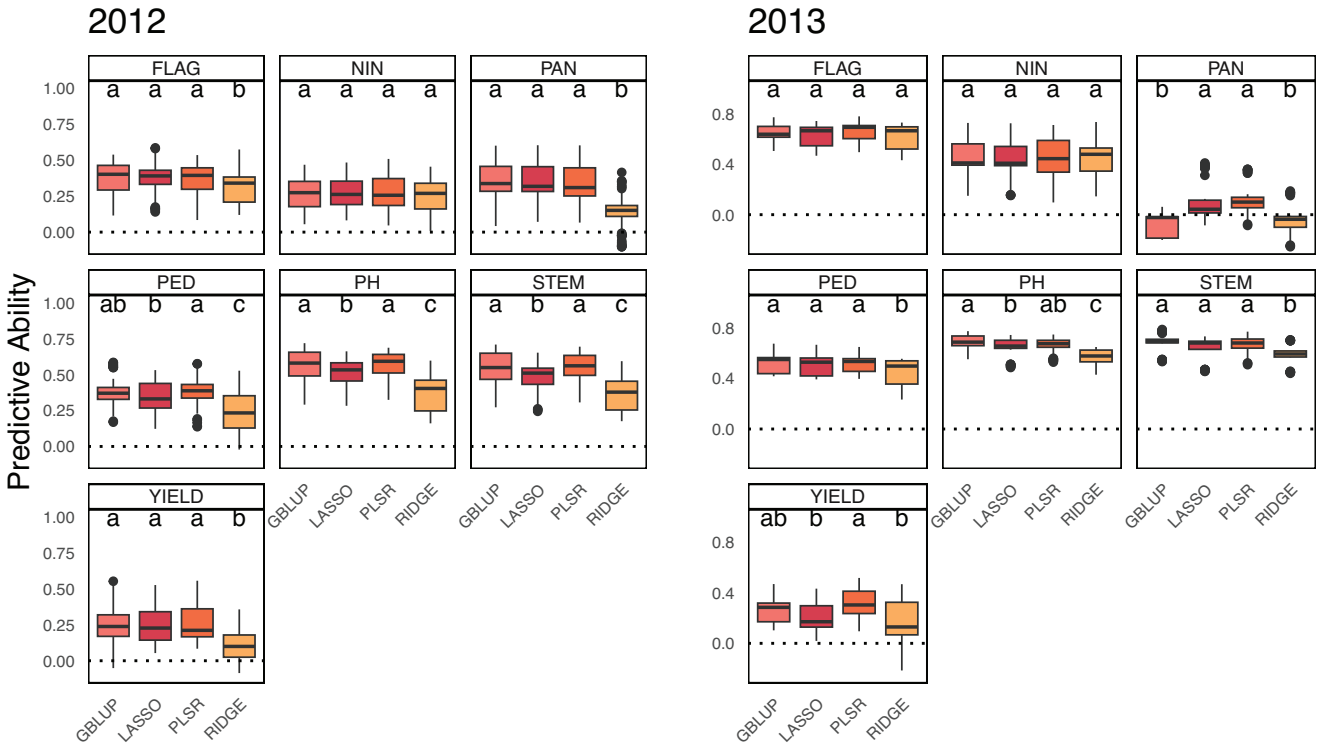

FIGURE 4 Effects of statistical method on predictive abilities. Predictive abilities were calculated in the CVb scenario with a training set size of 1000 for 2012 and 450 for 2013, and spectra preprocessed with DER1. Letters represent the results of pairwise Wilcoxon tests corrected by Bonferroni method for a threshold of 5%. Horizontal dotted line is at the 0 value. FLAG, flag leaf appearance; GBLUP (Genomic Best Linear Unbiased Prediction), reference gblup model; LASSO, least absolute selection and shrinkage operator; NIN, number of internodes; PAN, panicle length; PED, peduncle length; PH, plant height; PLSR, partial least square regression; RIDGE, ridge regression; STEM, stem length; YIELD, yield.

relatedness between training and validation sets. This relatedness decreased from CVa to CVc (Figure S5), which caused PAs to decrease both in PP and GP (Figure 6). Results from the different CV scenarios were different for GP in each trait, whereas, for PP, results from CVb and CVc were never significantly different except for STEM. Thus, for GP, more significant differences between CV scenarios were observed than for PP and it was especially the case for PH-related traits. For instance, the median PAs of PP for PH dropped from 0.52 to 0.21 between CVa and CVc, while for GP it dropped from 0.74 to 0.06. For FLAG, NIN, PED, PH, and STEM, PP had significantly higher PAs than GP in CVc scenario while GP had significantly superior PAs in CVa. For YIELD, PAs of PP were not significantly different from PAs of GP in CVc while they were different in CVa.

CDmean and Fst indicators were calculated for every CV partition used (Figure S5) to study population structure, and linear regressions were carried out to test the effect of CDmean and Fst between training and validation sets on PA (Figure S6). Slopes for PP were systematically closer to 0 than slopes for GP. For STEM, the slope for PP was not even significantly different from 0 with the CDMEAN indicator. There were no CDmean values between 0.32 and 0.56, and no Fst values between 0.11 and 0.51 in our partitions. Optimizing training and validation set composition to target specific CDmean or Fst values can be extremely time consuming. Given the variability of PA for each CDmean or Fst value, we did not seek to optimize partitions to access intermediate values because it would have taken too much time to have a representative number of partitions.

The effect of population structure was not studied in 2013 because the maximum training set size for CVa scenario was too low to achieve relevant PA.

3.8 | Effect of NIRS acquisition environment

Results for inter-year prediction are presented in Figure 7. CVb_2012 is the reference scenario with data of the training and validation set coming from 2012, and predicted with the CVb scenario. In CVd_2012, spectra and breeding values used to train the model were acquired in 2013 while spectra and breeding values used to validate the model were acquired in 2012. In CVd_2013, training data were acquired in 2012 and validation data were acquired in 2013. Training and validation sets of CVd scenarios share no common environment nor common genotypes. No differences were found between GP and PP for the CVd scenario except for YIELD where PP

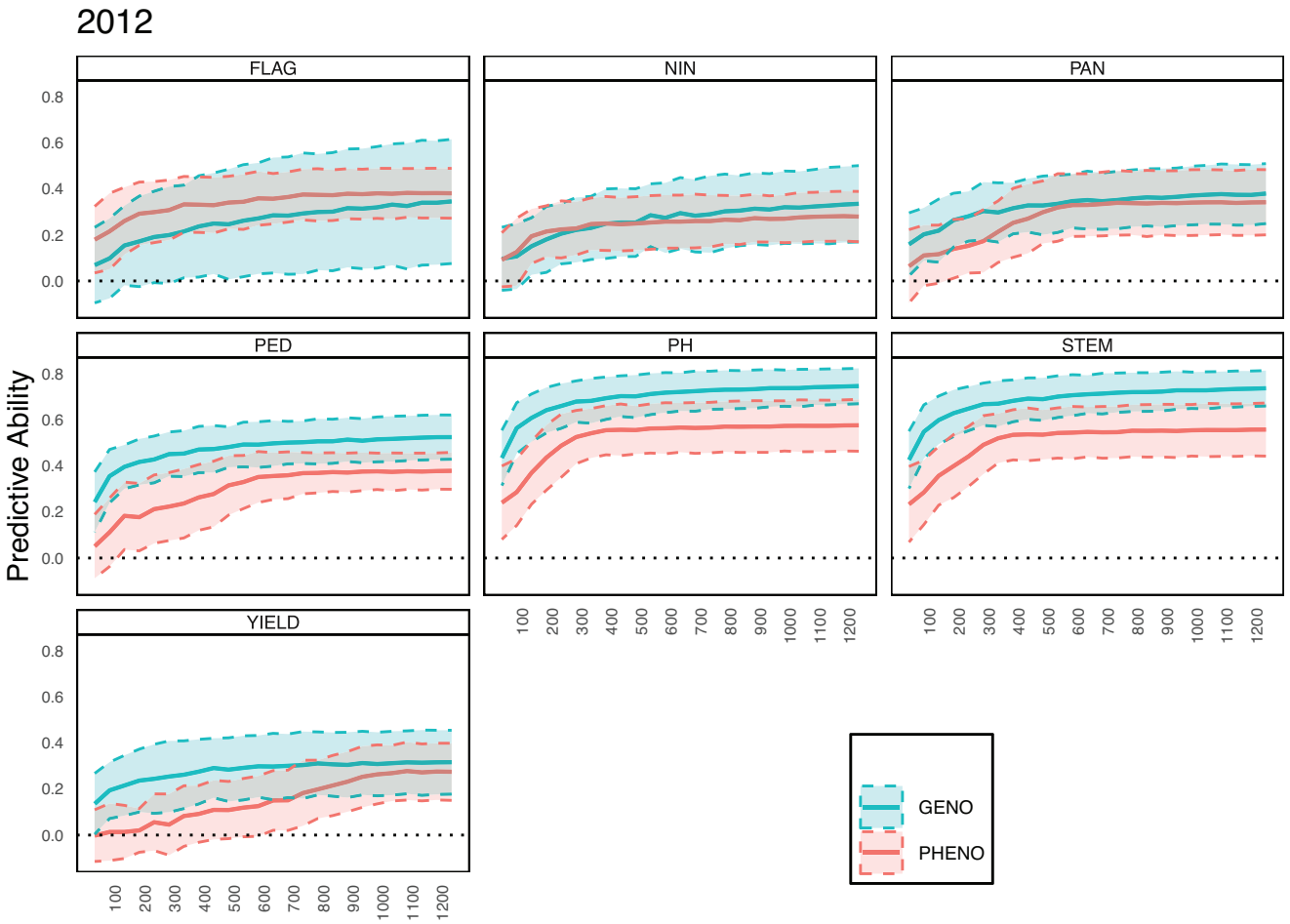

FIGURE 5 Effect of training set size on predictive abilities. Predictive abilities were calculated with 2012 data, using GBLUP (Genomic Best Linear Unbiased Prediction) in the CVb scenario and spectra preprocessed with DER1. Horizontal dotted line is at the 0 value. Full lines represent the mean predictive abilities (PAs) and dotted lines represent the standard error around the mean PA. FLAG, flag leaf appearance; NIN, number of internodes; PAN, panicle length; PED, peduncle length; PH, plant height; STEM, stem length; YIELD, yield.

had lower PAs than GP in CVd_Y12. Inter-year scenarios had lower PAs than CVb for all traits except FLAG and NIN. For all traits except YIELD, training models on 2012 data gave significantly better results than training models on 2013 data.

3.9 | Effect of wavelength selection

The results of PP using wavelengths selected by partial least square (PLS), LASSO, or randomly are presented in Figure 8. These results are from GBLUP models, but using only wavelengths selected by LASSO, PLS, or randomly. Models with wavelength selection were compared to PP models with all wavelengths and GP. The wavelengths that were selected by LASSO or PLSR are presented in Figures S7–S10. The sets of wavelengths selected by LASSO gave the same PAs as using all wavelengths for FLAG, NIN, PAN, and YIELD in 2012 and for all traits in 2013 except PH. Wavelengths selected by PLS gave the same PAs as using all wavelengths for NIN and FLAG in 2012 and NIN, FLAG, and YIELD in 2013. The number of wavelengths selected was dependent on the trait, but LASSO selected less than 10 wavelengths for FLAG, PAN, PED, PH, and STEM in 2012 and FLAG, NIN, PED, PH, and STEM in 2013. LASSO selected around 20 wavelengths otherwise. In 2012, PLS selected around 30 wavelengths for each trait and year except for NIN with 38 wavelengths selected in 2012 and 40 in 2013. In 2013, PLS selected around 35 wavelengths except for PED with 28 wavelengths selected.

Considering randomly selected wavelengths, the number of wavelengths required to reach the same PAs as using all wavelengths differed across traits. In 2012, 30 wavelengths were enough to predict STEM, PED, and PH, 20 wavelengths were enough to predict YIELD, 15 were enough to predict PAN, 10 to predict FLAG, and six to predict NIN with the same accuracy as with all wavelengths. In 2013, 30 wavelengths were needed to predict PED, 20 for YIELD and PH, 15 for STEM, nine for FLAG, and three for NIN.

For equivalent number of wavelengths selected, LASSO selection generally outperformed random selection, but

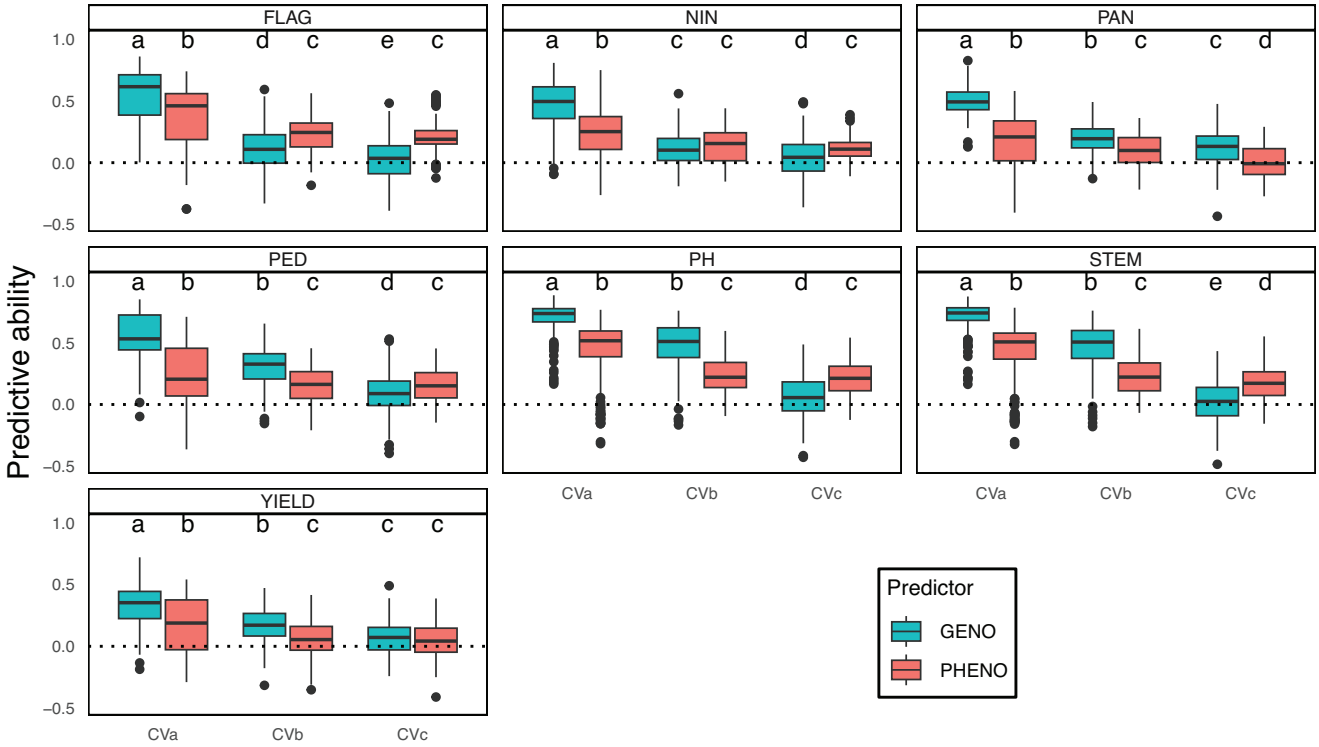

FIGURE 6 Effects of population structure on predictive abilities. Predictive abilities were calculated with 2012 data, using GBLUP (Genomic Best Linear Unbiased Prediction) scenario with a training set size of 100 and spectra preprocessed with DER1. Horizontal dotted line is at the 0 value. Letters represent the results of pairwise Wilcoxon tests corrected by Bonferroni method for a threshold of 5%. CVa, CVb, and CVc refer to cross-validation scenarios with decreasing genetic relatedness between training and validation sets from CVa to CVc. CVa denotes intra-cross fivefold cross validation, CVb denotes intra recurrent parent leave one cross out cross-validation, and CVc denotes inter recurrent parent cross-validation. FLAG, flag leaf appearance; NIN, number of internodes; PAN, panicle length; PED, peduncle length; PH, plant height; STEM, stem length; YIELD, yield.

random selection often outperformed PLS selection. Lastly, randomly selecting a very low number of wavelengths gave significantly lower PAs than using all the wavelengths in most traits, yet, the predictive power reached at only a dozen of randomly selected wavelengths was surprisingly high.

4 | DISCUSSION

In this study, we implemented PP on sorghum using a large BCNAM population. Our design comprised 2 years with four environments each, and a strongly structured population allowed us to study factors impacting PP, namely, spectra preprocessing, statistical method, training set size, population structure, NIRS acquisition environment, and wavelength selection.

4.1 | Phenomic selection for sorghum breeding

Despite being the fifth most produced cereal worldwide, a major crop for food security in arid and semiarid areas, and a promising crop to face climate change issues, sorghum genetics and breeding programs have not received much attention and means for development and optimization (Khoury et al., 2014; Pingali & Traxler, 2002). GP has already been tested on this species with promising results (Hunt et al., 2018; Maulana et al., 2023; Yu et al., 2016) in temperate programs. Moreover, more and more cereal breeding programs from temperate and semiarid regions are mobilizing NIRS technologies on a routine basis to assess grain and forage quality. (For cereals, see Osborne [2006]; for rice, see Sun et al. [2014], for sorghum, see personal communications from LIDEA-Seeds [France, 2024], RAGT2N [France, 2024], and the IAVAO Network [Jean-François Rami, CIRAD, France, 2024]). Thus, the additional costs to implement PP in these programs is low, and PP fits well in the current context of sorghum breeding.

More generally, PP as a low-cost, or even low-tech method is especially relevant for breeding programs with low financial support and reduced access to molecular technologies. Our study demonstrated the possibility of using PP even in a rather unfavorable very low cost set up, that is, with only one reference NIRS acquisition environment and one spectrum per genotype. Having only one spectrum per genotype, that is, no biological repetitions, makes it likely that genetic information is hard to retrieve, but still, models were able to grasp

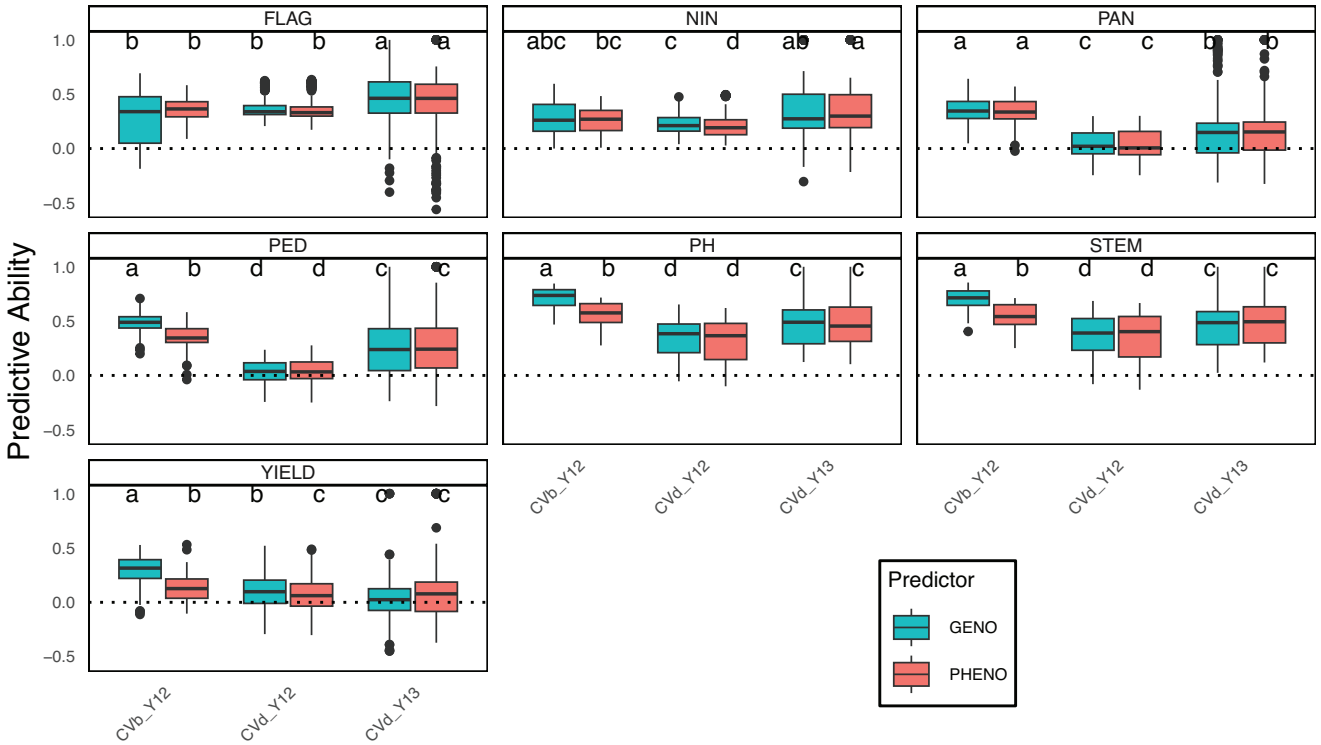

FIGURE 7 Effects of near infrared reflectance spectroscopy (NIRS) acquisition environment on predictive abilities. Predictive abilities were calculated using GBLUP (Genomic Best Linear Unbiased Prediction) with training set size of 600 and spectra preprocessed with DER1. CVb_2012 is the reference model, in CVb scenario for 2012 data with 600 individuals in the training set and spectra preprocessed with DER1. CVd_2012 is the model trained on 2013 data and validated on 2012 data. CVd_2013 is the model trained on 2012 data and validated on 2013 data. Horizontal dotted line is at the 0 value. Letters represent the results of pairwise Wilcoxon tests corrected by Bonferroni method for a threshold of 5%. FLAG, flag leaf appearance; NIN, number of internodes; PAN, panicle length; PED, peduncle length; PH, plant height; STEM, stem length; YIELD, yield.

information out of them. Moreover, we showed that using a single reference NIRS acquisition environment is possible as shown by previous studies (Brault et al., 2022; Rincent et al., 2018; Zhu et al., 2021, 2022). We also showed that predicting untested genotypes in untested environment is possible even though PP often produces lower PAs than GP in this case. One interesting approach to further evaluate the possibilities of using PP in such unfavorable design would be to better characterize environments or quantify environmental relatedness between training and validation sets to see how it affects PA. Such approach currently applied for GP yielded significant improvements (Montesinos-López et al., 2024).

4.2 | Comparisons between PP and GP

A recent study demonstrated that comparing PA of PP and GP is not relevant because both can be biased in different directions(Wang et al., 2025). Results of Wang et al. (2025) show that differences between PAs of PP and GP tend to be overestimated when PP outperforms GP, whereas they tend to be underestimated when GP outperforms PP. The authors argue that this is due to correlations between phenomic predicted breeding values, and the environmental, nonadditive genetics, and measurement error information that is present in the spectra. In our case, breeding values used for the training and validation of models were calculated without taking into account the environment of spectra acquisition. Thus, there should be no, or little correlation between predicted breeding values and environmental information of the spectra. Moreover, we can argue that environmental information may be the most impacting factor in biasing PAs before nonadditive genetic effects (which cannot be addressed) and measurement errors (which can be addressed by using different spectrometers for different genotypes). Finally, even when comparisons between GP and PP are not relevant, it is important to note that PP on its own had encouraging results and was able to predict genetic values to a certain extent.

4.3 | Factors impacting PP

Our results show that PP can impact the choice of the training set. Our results are in accordance with previous studies showing that PP allows smaller training sets than GP to reach a plateau of PAs (Dallinger et al., 2023; Zhu et al., 2021, 2022).

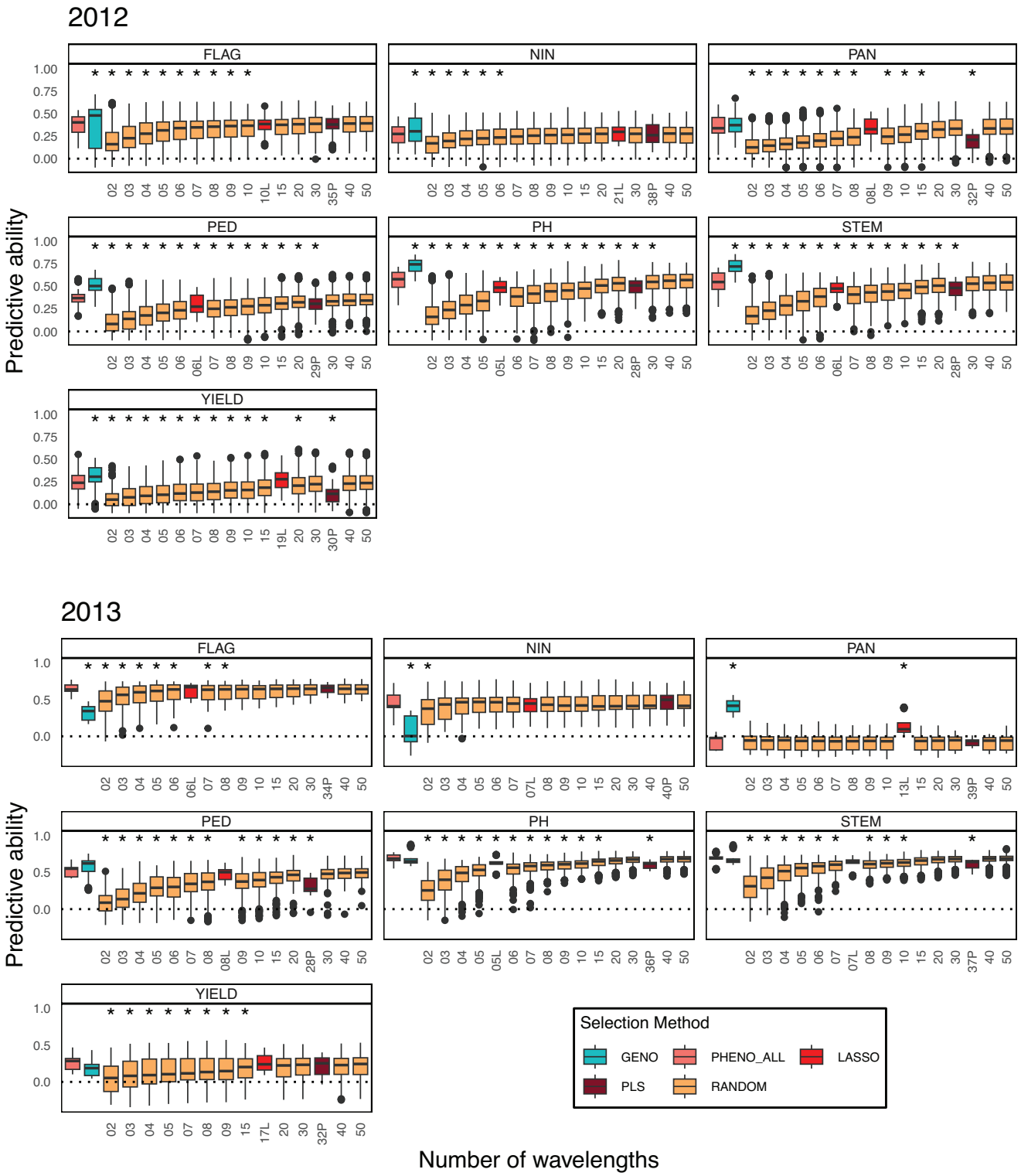

FIGURE 8 Effects of wavelength selection on predictive abilities. Predictive abilities were using GBLUP (Genomic Best Linear Unbiased Prediction) in the CVb scenario with a training set size of 1000 in 2012 and 450 in 2013, on raw spectra. Wavelengths were randomly selected (RAND), or selected using least absolute shrinkage and selection operator (LASSO) (numbers with L) and partial least square (PLS) (numbers with P) models. PHENO_ALL is the phenomic prediction with all wavelengths, and GENO is the genomic prediction. Horizontal dotted line is at the 0 value. Asterisks indicate predictive abilities that are significantly different from the PHENO_ALL modality, tested with Dunnett test with a threshold of 5%. The *x*-axes are not linear and differ across traits and years as different number of wavelengths were selected by PLS and LASSO in these different contexts. FLAG, flag leaf appearance; NIN, number of internodes; PAN, panicle length; PED, peduncle length; PH, plant height; STEM, stem length; YIELD, yield.

Results on PH-related traits and FLAG are in accordance with previous studies showing that PP is less affected by population structure (Laurençon et al., 2024; Roscher-Ehrig et al., 2024; Weiß et al., 2022; Zhu et al., 2021, 2022). Most importantly, PP had higher PA than GP for almost all traits in the scenario with the least relatedness between training and validation sets, highlighting a real advantage of PP over GP in less favorable prediction scenarios. Nevertheless, as training set sizes available for comparison between all CV scenarios were small, it would be interesting to test similar scenarios with larger training set sizes to generalize results. This study also highlighted the trait dependency of these features of PP mitigating the advantages PP can have over GP. This trait dependency could not be explained by the genetic architecture (number of QTL discovered for the trait). Indeed, no pattern emerged when trying to establish a link between our prediction results and the genetic architecture of traits previously studied in the same population (Garin et al., 2024). A few authors have studied the impact of genetic architecture on PP (Roscher-Ehrig et al., 2024; Zhu et al., 2022). They found that PP performs better than GP for complex traits determined by many QTL. The absence of such pattern in our results may arise from the fact that the number of QTL detected for each trait in our population (Garin et al., 2024) is always lower than six and that the explained variance varies a lot for each trait across environments and subpopulations.

Considering factors related to data processing, we found small or no effects of spectra preprocessing with poor performances when spectra were preprocessed with DER2 but few significant differences in PA in most cases with other preprocessing. This result is hardly generalizable as no consensus exists in the literature regarding the best preprocessing method. Even though some studies found no effect of spectra preprocessing with classical chemometric methods (Brault et al., 2022; Dallinger et al., 2023), it was proposed by Meyenberg et al. (2024) that fine tuning parameters of preprocessing methods could have an impact on prediction. Some improvement of PP performance was found by using genetic values of reflectance or absorbances at each wavelength (Brault et al., 2022). Unfortunately, spectra of our data set were only acquired in one environment per genotype hindering the calculation of genetic values at each wavelength without using a marker-based kinship matrix. While this approach could have been interesting to test, it would have been more a particular case of integrating genomic and phenomic data together than a kind of preprocessing that would involve only spectral information.

Considering factors related to data analysis, we found some significant effect of the statistical method on the performance of PP, RIDGE being generally less performant, GBLUP and PLSR being the most performant. One limit of our study is that we have only tested frequentist linear methods. Regarding this factor, it is also hard to generalize results as some studies including Bayesian and nonlinear frameworks did not find major differences between statistical methods (Roscher-Ehrig et al., 2024; Zhu et al., 2021) while others found that using nonlinear methods (Cuevas et al., 2019) or deep learning (Mora-Poblete et al., 2024) could enhance PA of PP. Nonetheless, using Bayesian or deep learning methods requires more computational resources than classic frequentist methods and in addition implementing deep learning strategies requires expertise and energy demanding strategies to fine tune model parameters (Lee et al., 2022; Zhang et al., 2019). In this context, applications of such methodologies will be dependent on the effective gains they will provide to compensate their higher investments requirements.

Finally, one important feature of this work is the fact that spectra were acquired in a set of environmental conditions that were not used to estimate breeding values, showing that it is possible to perform predictions using a reference environment for NIRS acquisition in a given year. Moreover, our results with inter-year CV scenarios also show that PP can be as good as GP in predicting untested genotypes in untested environments.

4.4 | Information contained in NIRS is poorly valued by prediction models

Selecting wavelengths with PLSR or LASSO models worked well to reduce the number of wavelengths without substantial loss of PA but did not always outperform random selection for equivalent numbers of wavelengths selected. The best selection method was always LASSO with which in most cases, less than 15 wavelengths enabled predictions as good as using all 1154 wavelengths. As spectra measurements were done with only one acquisition time and no repetition, this result is in line with the "minimal setup" proposed by Mróz et al. (2024) in which only three bands in the visible range acquired on a single date for canopy measurement could predict wheat YIELD with an accuracy ranging from 0.51 to 0.55. One must be aware that this study used a diversity panel and canopy measurement, which is a totally different context than ours making comparisons irrelevant, but it still shows that it is possible to reach decent PA with limited information.

In this study, PP was used as defined by Rincent et al. (2018), that is, using near infrared spectroscopy on grains to estimate a kinship matrix and perform predictions. Since 2018, many studies have widened the concept and have also used canopy measurement to capture fewer visible and NIRS bands with multispectral cameras (Fumia, Nair, et al., 2023; Galán et al., 2020; Krause et al., 2019), sometimes capturing less than 10 bands (Li et al., 2023; Maggiorelli et al., 2024; Mróz et al., 2024). Other studies used VI also calculated with very few bands but often densely measured throughout the growing cycle and coupled with PH measurements (Adak

et al., 2022, 2024; Adak, Anderson, et al., 2023; Adak, Kang, et al., 2023; Biswas et al., 2021; Kaushal et al., 2024; Sandhu et al., 2021; Shafiee et al., 2023; Togninalli et al., 2023; Washburn et al., 2024; Winn et al., 2023), or even temperature and canopy area (Parmley et al., 2019). It would be interesting to study the costs and benefits of using only a few bands in a minimal setup or with dense acquisition along the plant cycle, compared to PP as defined in Rincent et al. (2018).

We hypothesized that using less than 15 selected wavelengths or even a few randomly selected wavelengths is possible either because of low amount of useful information in spectra or because of poor use of complete spectral information. The low information amount hypothesis is that only a few wavelengths are informative in the spectra and other wavelengths contain either no information or an information due to their correlation with informative wavelengths (a linkage-disequilibrium-like relation between wavelengths). This hypothesis is supported by the high multicollinearity that typically occur between adjacent wavelengths and the PCA of raw spectra (Figure S4) showing that the two first components of the PCA account for more than 97% of the total variance of spectra, meaning that there are only two independent features in the spectra. The poor use of complete spectral information hypothesis is that models do not really extract all the information possible when all wavelengths are used and in fact extract a quantity of information equivalent to what is contained in only a few wavelengths and can also extract noise that disturbs predictions. This could be because wavelengths are considered individually and independently in models (e.g., when calculating a spectral kinship matrix) while they actually are very correlated to each other. More classical uses of NIRS made in the field of chemometrics rather calculate "features" of spectra (e.g., principal components) to further analyze them (J.-M. Roger [personal communication, January 7, 2025]). The possibility to use few wavelengths questions the need for "omic" information in prediction, which are typically quite costly (genomic, transcriptomic among others). As stated in Zhu et al. (2021, 2022), each wavelength, behaving like endophenotypes may contain additional information including additive, nonadditive, environmental, and genotype–environment–interaction effects. As such, using a reduced number of bands could in fact contain an omic type of information despite the low number of variables measured compared to classic omics.

That being said, all the nonadditive information supposedly contained in the spectra were not explicitly taken into account in our models, and it is impossible at this stage to confirm that this information is responsible for PP performances. The growing body of results concerning PP that could be explained by the endophenotype behavior hypothesis calls for a better understanding of the information contained in the spectra and how it can be linked to genetic relationship matrices. It would be interesting to decompose the phenomic signal into its additive and nonadditive parts to compare models that contain both or neither. Further analysis could also focus on an efficient way to make this decomposition, which may borrow from the chemometrics framework.

5 | CONCLUSION AND PERSPECTIVES

The interest of PP for sorghum breeding has been demonstrated. We studied factors impacting PP, confirmed some of its advantages over GP concerning training population definition and size, but we also highlighted a strong trait dependency of these advantages. We also stressed the robustness of PP in unfavorable application conditions, and when using only a few selected wavelengths, opening prospects for the adoption of PP in breeding programs from the global south. Further studies are still needed to better understand and decompose phenomic information to use each component at best in predictive models.

AUTHOR CONTRIBUTIONS

Clément Bienvenu: Conceptualization; data curation; formal analysis; investigation; methodology; writing—original draft. **Vincent Garin**: Data curation; methodology; writing—review and editing. **Nicolas Salas**: Formal analysis; writing—review and editing. **Korotimi Théra**: Formal analysis; resources; writing—review and editing. **Mohamed Lamine Tekete**: Formal analysis; resources; writing—review and editing. **Madathiparambil Chandran Sarathjith**: Resources; writing—review and editing. **Chiaka Diallo**: Formal analysis; resources; writing—review and editing. **Angélique Berger**: Resources; writing—review and editing. **Caroline Calatayud**: Resources; writing—review and editing. **Fabien De Bellis**: Resources; writing—review and editing. **Jean-François Rami**: Formal analysis; funding acquisition; resources; writing—review and editing. **Michel Vaksmann**: Formal analysis; resources; writing—review and editing. **Vincent Segura**: Conceptualization; investigation; supervision; validation; writing—review and editing. **David Pot**: Conceptualization; funding acquisition; investigation; supervision; validation; writing—review and editing. **Hugues de Verdal**: Conceptualization; investigation; supervision; validation; writing—review and editing.

ACKNOWLEDGMENTS

We acknowledge the MGX-Genotyping platform (UMR AGAP—CIRAD, Montpellier, France) for its support regarding the genotyping of the BCNAM population. We would like to thank the anonymous reviewers of the previous versions of this article for their highly relevant comments and suggestions regarding the genetic value estimations used to train the prediction models and Section 4 that greatly contributed to the current version. This work was supported by grant

from the Generation Challenge Programme (Project Numbers G7010.05.01 and G7010.05.02). The work of Clément Bienvenu was supported by a doctoral allowance from the French Ministry of Higher Education and Research.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data used in this study are available at https://doi.org/10.18167/DVN1/TZVGLS.

ORCID

Clément Bienvenu https://orcid.org/0009-0004-9234-6216
Vincent Garin https://orcid.org/0000-0002-5571-1841
Nicolas Salas https://orcid.org/0009-0009-3462-7673
Korotimi Théra https://orcid.org/0000-0002-9776-8408
Mohamed Lamine Tekete https://orcid.org/0000-0002-9812-0187
Madathiparambil Chandran Sarathjith https://orcid.org/0000-0001-7555-7733
Chiaka Diallo https://orcid.org/0009-0001-7984-9028
Angélique Berger https://orcid.org/0000-0003-1249-3706
Fabien De Bellis https://orcid.org/0000-0001-7070-7691
Jean-François Rami https://orcid.org/0000-0002-5679-3877
Michel Vaksmann https://orcid.org/0000-0002-5258-1279
Vincent Segura https://orcid.org/0000-0003-1860-2256
David Pot https://orcid.org/0000-0001-6144-8448
Hugues de Verdal https://orcid.org/0000-0002-1923-8575

REFERENCES

Adak, A., Anderson, S. L., & Murray, S. C. (2023). Pedigree-management-flight interaction for temporal phenotype analysis and temporal phenomic prediction. *The Plant Phenome Journal*, *6*(1), e20057. https://doi.org/10.1002/ppj2.20057

Adak, A., DeSalvio, A. J., Arik, M. A., & Murray, S. C. (2024). Field-based high-throughput phenotyping enhances phenomic and genomic predictions for grain yield and plant height across years in maize. *G3 Genes|Genomes|Genetics*, *14*(7), jkae092. https://doi.org/10.1093/g3journal/jkae092

Adak, A., Kang, M., Anderson, S. L., Murray, S. C., Jarquin, D., Wong, R. K. W., & Katzfuß, M. (2023). Phenomic data-driven biological prediction of maize through field-based high-throughput phenotyping integration with genomic data. *Journal of Experimental Botany*, *74*(17), 5307–5326. https://doi.org/10.1093/jxb/erad216

Adak, A., Murray, S. C., & Anderson, S. L. (2022). Temporal phenomic predictions from unoccupied aerial systems can outperform genomic predictions. *G3 Genes|Genomes|Genetics*, *13*(1), jkac294. https://doi.org/10.1093/g3journal/jkac294

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., Ferrão, R. G., Munoz, P. R., & Ferrão, L. F. V. (2024). A comparison of genomic and phenomic selection methods for yield prediction in *Coffea canephora*. *The Plant Phenome Journal*, *7*(1), e20109. https://doi.org/10.1002/ppj2.20109

Akdemir, D., Rio, S., & Isidro y Sánchez, J. (2021). TrainSel : An R package for selection of training populations. *Frontiers in Genetics*, *12*, 655287. https://doi.org/10.3389/fgene.2021.655287

Biswas, A., Andrade, M. H. M. L., Acharya, J. P., de Souza, C. L., Lopez, Y., de Assis, G., Shirbhate, S., Singh, A., Munoz, P., & Rios, E. F. (2021). Phenomics-assisted selection for herbage accumulation in alfalfa (*Medicago sativa* L.). *Frontiers in Plant Science*, *12*, 756768. https://doi.org/10.3389/fpls.2021.756768

Bradburd, G. (2024). *BEDASSLE: Quantifies effects of geo/eco distance on genetic differentiation*. https://CRAN.R-project.org/package=BEDASSLE

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brandolini-Bunlon, M., Jallais, M., Roger, J. M., & Lesnoff, M. (2023). *R package rchemo : Dimension reduction, regression and discrimination for chemometrics*. https://github.com/ChemHouse-group/rchemo

Brault, C., Lazerges, J., Doligez, A., Thomas, M., Ecarnot, M., Roumet, P., Bertrand, Y., Berger, G., Pons, T., François, P., Le Cunff, L., This, P., & Segura, V. (2022). Interest of phenomic prediction as an alternative to genomic prediction in grapevine. *Plant Methods*, *18*(1), 108. https://doi.org/10.1186/s13007-022-00940-9

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, *103*(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Butler, D. (2021). *asreml: Fits the linear mixed model*. https://asreml.kb.vsni.co.uk/knowledge-base/asreml/

Covarrubias-Pazaran, G. (2016). Genome assisted prediction of quantitative traits using the R package sommer. *PLoS ONE*, *11*, 1–15. https://doi.org/10.1371/journal.pone.0156744

Cuevas, J., Montesinos-López, O., Juliana, P., Guzmán, C., Pérez-Rodríguez, P., González-Bucio, J., Burgueño, J., Montesinos-López, A., & Crossa, J. (2019). Deep kernel for genomic and near infrared predictions in multi-environment breeding trials. *G3 Genes|Genomes|Genetics*, *9*(9), 2913–2924. https://doi.org/10.1534/g3.119.400493

Dallinger, H. G., Löschenberger, F., Bistrich, H., Ametz, C., Hetzendorfer, H., Morales, L., Michel, S., & Buerstmayr, H. (2023). Predictor bias in genomic and phenomic selection. *Theoretical and Applied Genetics*, *136*(11), 235. https://doi.org/10.1007/s00122-023-04479-8

DeSalvio, A. J., Adak, A., Murray, S. C., Jarquín, D., Winans, N. D., Crozier, D., & Rooney, W. L. (2024). Near-infrared reflectance spectroscopy phenomic prediction can perform similarly to genomic prediction of maize agronomic traits across environments. *The Plant Genome*, *17*(2), e20454. https://doi.org/10.1002/tpg2.20454

de Verdal, H., Segura, V., Pot, D., Salas, N., Garin, V., Rakotoson, T., Raboin, L.-M., VomBrocke, K., Dusserre, J., Pacheco, S. A. C., & Grenier, C. (2024). Performance of phenomic selection in rice : Effects of population size and genotype-environment interactions on predictive ability. *PLoS ONE*, *19*(12), e0309502. https://doi.org/10.1371/journal.pone.0309502

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. https://doi.org/10.1371/journal.pone.0019379

FAOSTAT. (2024). *Crops and livestock products*. https://www.fao.org/faostat/en/#data/QCL

Feng, L., Zhang, Z., Ma, Y., Du, Q., Williams, P., Drewry, J., & Luck, B. (2020). Alfalfa yield prediction using UAV-based hyperspectral imagery and ensemble learning. *Remote Sensing*, *12*(12), Article 12. https://doi.org/10.3390/rs12122028

Friedman, J., Tibshirani, R., & Hastie, T. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Fumia, N., Kantar, M., Lin, Y., Schafleitner, R., Lefebvre, V., Paran, I., Börner, A., Diez, M. J., Prohens, J., Bovy, A., Boyaci, F., Pasev, G., Tripodi, P., Barchi, L., Giuliano, G., & Barchenger, D. W. (2023). Exploration of high-throughput data for heat tolerance selection in *Capsicum annuum*. *The Plant Phenome Journal*, *6*(1), e20071. https://doi.org/10.1002/ppj2.20071

Fumia, N., Nair, R., Lin, Y.-P., Lee, C.-R., Chen, H.-W., von Wettberg, E. B., Kantar, M., & Schafleitner, R. (2023). Leveraging genomics and phenomics to accelerate improvement in mungbean : A case study in how to go from GWAS to selection. *The Plant Phenome Journal*, *6*(1), e20088. https://doi.org/10.1002/ppj2.20088

Galán, R. J., Bernal-Vasquez, A.-M., Jebsen, C., Piepho, H.-P., Thorwarth, P., Steffan, P., Gordillo, A., & Miedaner, T. (2020). Integration of genotypic, hyperspectral, and phenotypic data to improve biomass yield prediction in hybrid rye. *Theoretical and Applied Genetics*, *133*(11), 3001–3015. https://doi.org/10.1007/s00122-020-03651-8

Galli, G., Horne, D. W., Collins, S. D., Jung, J., Chang, A., Fritsche-Neto, R., & Rooney, W. L. (2020). Optimization of UAS-based high-throughput phenotyping to estimate plant health and grain yield in sorghum. *The Plant Phenome Journal*, *3*(1), e20010. https://doi.org/10.1002/ppj2.20010

Garin, V., Diallo, C., Tékété, M. L., Théra, K., Guitton, B., Dagno, K., Diallo, A. G., Kouressy, M., Leiser, W., Rattunde, F., Sissoko, I., Touré, A., Nébié, B., Samaké, M., Kholovà, J., Frouin, J., Pot, D., Vaksmann, M., Weltzien, E., … Rami, J.-F. (2024). Characterization of adaptation mechanisms in sorghum using a multireference back-cross nested association mapping design and envirotyping. *Genetics*, *226*(4), iyae003. https://doi.org/10.1093/genetics/iyae003

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE*, *9*(2), e90346. https://doi.org/10.1371/journal.pone.0090346

Gonçalves, M. T. V., Morota, G., Costa, P. M. D. A., Vidigal, P. M. P., Barbosa, M. H. P., & Peternelli, L. A. (2020). Near-infrared spectroscopy outperforms genomics for predicting sugarcane feedstock quality traits. *bioRxiv*. https://doi.org/10.1101/2020.07.16.206110

Guindo, D., Teme, N., Vaksmann, M., Doumbia, M., Vilmus, I., Guitton, B., Sissoko, A., Mestres, C., Davrieux, F., Fliedel, G., Kouressy, M., Courtois, B., & Rami, J.-F. (2019). Quantitative trait loci for sorghum grain morphology and quality traits : Toward breeding for a traditional food preparation of West-Africa. *Journal of Cereal Science*, *85*, 256–272. https://doi.org/10.1016/j.jcs.2018.11.012

Harlan, J. R., & de Wet, J. M. J. (1972). A simplified classification of cultivated sorghum. *Crop Science*, *12*(2), 172–176. https://doi.org/10.2135/cropsci1972.0011183X001200020005x

Hossain, M. S., Islam, M. N., Rahman, M. M., Mostofa, M. G., & Khan, M. A. R. (2022). Sorghum : A prospective crop for climatic vulnerability, food and nutritional security. *Journal of Agriculture and Food Research*, *8*, 100300. https://doi.org/10.1016/j.jafr.2022.100300

Hunt, C. H., van Eeuwijk, F. A., Mace, E. S., Hayes, B. J., & Jordan, D. R. (2018). Development of genomic prediction in sorghum. *Crop Science*, *58*(2), 690–700. https://doi.org/10.2135/cropsci2017.08.0469

Kaushal, S., Gill, H. S., Billah, M. M., Khan, S. N., Halder, J., Bernardo, A., Amand, P. S., Bai, G., Glover, K., Maimaitijiang, M., & Sehgal, S. K. (2024). Enhancing the potential of phenomic and genomic prediction in winter wheat breeding using high-throughput phenotyping and deep learning. *Frontiers in Plant Science*, *15*, 1410249. https://doi.org/10.3389/fpls.2024.1410249

Khoury, C. K., Bjorkman, A. D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., Rieseberg, L. H., & Struik, P. C. (2014). Increasing homogeneity in global food supplies and the implications for food security. *Proceedings of the National Academy of Sciences*, *111*(11), 4001–4006. https://doi.org/10.1073/pnas.1313490111

Krause, M. R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R. P., Dreisigacker, S., Poland, J., Rutkoski, J., Sorrells, M., Gore, M. A., & Mondal, S. (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3: Genes, Genomes, Genetics*, *9*(4), 1231–1247. https://doi.org/10.1534/g3.118.200856

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics*, *6*(1), 10. https://doi.org/10.1186/1758-2946-6-10

Lado, B., Barrios, P. G., Quincke, M., Silva, P., & Gutiérrez, L. (2016). Modeling genotype × environment interaction for genomic selection with unbalanced data from a wheat breeding program. *Crop Science*, *56*(5), 2165–2179. https://doi.org/10.2135/cropsci2015.04.0207

Lane, H. M., Murray, S. C., Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Rooney, D. K., Barrero-Farfan, I. D., Fuente, G. N. D. L., & Morgan, C. L. S. (2020). Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. *The Plant Phenome Journal*, *3*(1), e20002. https://doi.org/10.1002/ppj2.20002

Laurençon, M., Legrix, J., Wagner, M.-H., Demilly, D., Baron, C., Rolland, S., Ducournau, S., Laperche, A., & Nesi, N. (2024). Genomic and phenomic predictions help capture low-effect alleles promoting seed germination in oilseed rape in addition to QTL analyses. *Theoretical and Applied Genetics*, *137*(7), 156. https://doi.org/10.1007/s00122-024-04659-0

Lee, B. D., Gitter, A., Greene, C. S., Raschka, S., Maguire, F., Titus, A. J., Kessler, M. D., Lee, A. J., Chevrette, M. G., Stewart, P. A., Britto-Borges, T., Cofer, E. M., Yu, K.-H., Carmona, J. J., Fertig, E. J., Kalinin, A. A., Signal, B., Lengerich, B. J., Jr, T. J. T., & Boca, S. M. (2022). Ten quick tips for deep learning in biology. *PLOS Computational Biology*, *18*(3), e1009803. https://doi.org/10.1371/journal.pcbi.1009803

Lehermeier, C., Krämer, N., Bauer, E., Bauland, C., Camisan, C., Campo, L., Flament, P., Melchinger, A. E., Menz, M., Meyer, N., Moreau, L., Moreno-González, J., Ouzunova, M., Pausch, H., Ranc, N., Schipprack, W., Schönleben, M., Walter, H., Charcosset, A., & Schön, C.-C. (2014). Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. *Genetics*, *198*(1), 3–16. https://doi.org/10.1534/genetics.114.161943

Leroy, T., Coumaré, O., Kouressy, M., Trouche, G., Sidibé, A., Sissoko, S., Touré, A. O., Guindo, T., Sogoba, B., & Dembélé, F. (2014). Inscription d'une variété de sorgho obtenue par sélection participative au Mali dans des projets multiacteurs. *Agronomie, Environnement et Sociétés*, *4*, 143–152.

Li, Y., Yang, X., Tong, L., Wang, L., Xue, L., Luan, Q., & Jiang, J. (2023). Phenomic selection in slash pine multi-temporally using UAV-multispectral imagery. *Frontiers in Plant Science*, *14*, 1156430. https://doi.org/10.3389/fpls.2023.1156430

Liu, Z.-Q., Li, H.-L., Zeng, X.-J., Lu, C., Fu, J.-Y., Guo, L.-J., Kimani, W. M., Yan, H.-L., He, Z.-Y., Hao, H.-Q., & Jing, H.-C. (2020). Coupling phytoremediation of cadmium-contaminated soil with safe crop production based on a sorghum farming system. *Journal of Cleaner Production*, *275*, 123002. https://doi.org/10.1016/j.jclepro.2020.123002

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. In D. L. Sparks (Ed.), *Advances in agronomy* (Vol. 110, pp. 77–123). Academic Press. https://doi.org/10.1016/B978-0-12-385531-2.00002-5

Lourenço, V. M., Ogutu, J. O., Rodrigues, R. A. P., Posekany, A., & Piepho, H.-P. (2024). Genomic prediction using machine learning : A comparison of the performance of regularized regression, ensemble, instance-based and deep learning methods on synthetic and empirical data. *BMC Genomics*, *25*(1), 152. https://doi.org/10.1186/s12864-023-09933-x

Lush, J. L. (1937). *Animal breeding plans*. Collegiate Press, Inc.

Machefert, C., Robert-Granié, C., Astruc, J. M., & Larroque, H. (2024). Genetic parameters of milk mid-infrared spectra and their genetic relationships with milk production and feed efficiency traits in French Lacaune dairy sheep. *Journal of Dairy Science*, *107*(12), 11239–11253. https://doi.org/10.3168/jds.2024-25127

Maggiorelli, A., Baig, N., Prigge, V., Bruckmüller, J., & Stich, B. (2024). Using drone-retrieved multispectral data for phenomic selection in potato breeding. *Theoretical and Applied Genetics*, *137*(3), 70. https://doi.org/10.1007/s00122-024-04567-3

Maulana, F., Perumal, R., Serba, D. D., & Tesso, T. (2023). Genomic prediction of hybrid performance in grain sorghum (*Sorghum bicolor* L.). *Frontiers in Plant Science*, *14*, 1139896. https://doi.org/10.3389/fpls.2023.1139896

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Meyenberg, C., Braun, V., Longin, C. F. H., & Thorwarth, P. (2024). Feature engineering and parameter tuning : Improving phenomic prediction ability in multi-environmental durum wheat breeding trials. *Theoretical and Applied Genetics*, *137*(8), 188. https://doi.org/10.1007/s00122-024-04695-w

Montesinos-López, O. A., Crespo-Herrera, L., Pierre, C. S., Cano-Paez, B., Huerta-Prado, G. I., Mosqueda-González, B. A., Ramos-Pulido, S., Gerard, G., Alnowibet, K., Fritsche-Neto, R., Montesinos-López, A., & Crossa, J. (2024). Feature engineering of environmental covariates improves plant genomic-enabled prediction. *Frontiers in Plant Science*, *15*, 1349569. https://doi.org/10.3389/fpls.2024.1349569

Mora-Poblete, F., Mieres-Castro, D., do Amaral, A. T., Jr., Balach, M., & Maldonado, C. (2024). Integrating deep learning for phenomic and genomic predictive modeling of *Eucalyptus* trees. *Industrial Crops and Products*, *220*, 119151. https://doi.org/10.1016/j.indcrop.2024.119151

Mróz, T., Shafiee, S., Crossa, J., Montesinos-Lopez, O. A., & Lillemo, M. (2024). Multispectral-derived genotypic similarities from budget cameras allow grain yield prediction and genomic selection augmentation in single and multi-environment scenarios in spring wheat. *Molecular Breeding*, *44*(1), 5. https://doi.org/10.1007/s11032-024-01449-w

Osborne, B. G. (2006). Applications of near infrared spectroscopy in quality screening of early-generation material in cereal breeding programmes. *Journal of Near Infrared Spectroscopy*, *14*(2), 93–101. https://doi.org10.1255/jnirs.595

Parmley, K., Nagasubramanian, K., Sarkar, S., Ganapathysubramanian, B., & Singh, A. K. (2019). Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. *Plant Phenomics*, *2019*, 5809404. https://doi.org/10.34133/2019/5809404

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A. K., Chapman, J., Feltus, F. A., Gowik, U., … Rokhsar, D. S. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, *457*(7229), 551–556. https://doi.org/10.1038/nature07723

Pingali, P. L., & Traxler, G. (2002). Changing locus of agricultural research : Will the poor benefit from biotechnology and privatization trends?. *Food Policy*, *27*(3), 223–238. https://doi.org/10.1016/S0306-9192(02)00012-X

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rincent, R., Charpentier, J.-P., Faivre-Rampant, P., Paux, E., Le Gouis, J., Bastien, C., & Segura, V. (2018). Phenomic selection is a low-cost and high-throughput method based on indirect predictions : Proof of concept on wheat and poplar. *G3 Genes|Genomes|Genetics*, *8*(12), 3961–3972. https://doi.org/10.1534/g3.118.200760

Rincent, R., Laloë, D., Nicolas, S., Altmann, T., Brunel, D., Revilla, P., Rodríguez, V. M., Moreno-Gonzalez, J., Melchinger, A., Bauer, E., Schoen, C.-C., Meyer, N., Giauffret, C., Bauland, C., Jamin, P., Laborde, J., Monod, H., Flament, P., Charcosset, A., & Moreau, L. (2012). Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals : Comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics*, *192*(2), 715–728. https://doi.org/10.1534/genetics.112.141473

Robert, P., Auzanneau, J., Goudemand, E., Oury, F.-X., Rolland, B., Heumez, E., Bouchet, S., Le Gouis, J., & Rincent, R. (2022). Phenomic selection in wheat breeding : Identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection. *Theoretical and Applied Genetics*, *135*(3), 895–914. https://doi.org/10.1007/s00122-021-04005-8

Robert, P., Goudemand, E., Auzanneau, J., Oury, F.-X., Rolland, B., Heumez, E., Bouchet, S., Caillebotte, A., Mary-Huard, T., Le Gouis, J., & Rincent, R. (2022). Phenomic selection in wheat breeding : Prediction of the genotype-by-environment interaction in multi-environment breeding trials. *Theoretical and Applied Genetics*, *135*(10), 3337–3356. https://doi.org/10.1007/s00122-022-04170-4

Roscher-Ehrig, L., Weber, S. E., Abbadi, A., Malenica, M., Abel, S., Hemker, R., Snowdon, R. J., Wittkop, B., & Stahl, A. (2024). Phenomic selection for hybrid rapeseed breeding. *Plant Phenomics*, *6*, 0215. https://doi.org/10.34133/plantphenomics.0215

Sandhu, K. S., Mihalyov, P. D., Lewien, M. J., Pumphrey, M. O., & Carter, A. H. (2021). Combining genomic and phenomic information for predicting grain protein content and grain yield in spring wheat. *Frontiers in Plant Science*, *12*, 613300. https://doi.org/10.3389/fpls.2021.613300

Schmidt, P., Hartung, J., Bennewitz, J., & Piepho, H.-P. (2019). Heritability in plant breeding on a genotype-difference basis. *Genetics*, *212*(4), 991–1008. https://doi.org/10.1534/genetics.119.302134

Shafiee, S., Mroz, T., Burud, I., & Lillemo, M. (2023). Evaluation of UAV multispectral cameras for yield and biomass prediction in wheat under different sun elevation angles and phenological stages. *Computers and Electronics in Agriculture*, *210*, 107874. https://doi.org/10.1016/j.compag.2023.107874

Sun, C., Yu, Y., Duan, B., & Zhu, Z. (2014). Rapid prediction of rice quality characteristics by near-infrared reflectance spectroscopy for breeding programs. *Cereal Chemistry*, *91*(3), 270–275. https://doi.org/10.1094/CCHEM-08-13-0164-R

Tester, M., & Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science*, *327*(5967), 818–822. https://doi.org/10.1126/science.1183700

Thapa, S., Gill, H. S., Halder, J., Rana, A., Ali, S., Maimaitijiang, M., Gill, U., Bernardo, A., St. Amand, P., Bai, G., & Sehgal, S. K. (2024). Integrating genomics, phenomics, and deep learning improves the predictive ability for Fusarium head blight–related traits in winter wheat. *The Plant Genome*, *17*(3), e20470. https://doi.org/10.1002/tpg2.20470

Togninalli, M., Wang, X., Kucera, T., Shrestha, S., Juliana, P., Mondal, S., Pinto, F., Govindan, V., Crespo-Herrera, L., Huerta-Espino, J., Singh, R. P., Borgwardt, K., & Poland, J. (2023). Multi-modal deep learning improves grain yield prediction in wheat breeding by fusing genomics and phenomics. *Bioinformatics*, *39*(6), btad336. https://doi.org/10.1093/bioinformatics/btad336

Utz, H. F., Melchinger, A. E., & Schön, C. C. (2000). Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. *Genetics*, *154*(4), 1839–1849. https://doi.org/10.1093/genetics/154.4.1839

VanRaden, P. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Wang, F., Feldmann, M. J., & Runcie, D. E. (2025). Do not benchmark phenomic prediction against genomic prediction accuracy. *The Plant Phenome Journal*, *8*(1), e70029. https://doi.org/10.1002/ppj2.70029

Washburn, J. D., Adak, A., DeSalvio, A. J., Arik, M. A., & Murray, S. C. (2024). High temporal resolution unoccupied aerial systems phenotyping provides unique information between flight dates. *The Plant Phenome Journal*, *7*(1), e20113. https://doi.org/10.1002/ppj2.20113

Weir, B. S., & Hill, W. G. (2002). Estimating F-statistics. *Annual Review of Genetics*, *36*, 721–750. https://doi.org/10.1146/annurev.genet.36.050802.093940

Weiß, T. M., Zhu, X., Leiser, W. L., Li, D., Liu, W., Schipprack, W., Melchinger, A. E., Hahn, V., & Würschum, T. (2022). Unraveling the potential of phenomic selection within and among diverse breeding material of maize (*Zea mays* L.). *G3: Genes, Genomes, Genetics*, *12*(3), jkab445.

Wickham, H. (2016). *ggplot2 : Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Winn, Z. J., Amsberry, A. L., Haley, S. D., DeWitt, N. D., & Mason, R. E. (2023). Phenomic versus genomic prediction—A comparison of prediction accuracies for grain yield in hard winter wheat lines. *The Plant Phenome Journal*, *6*(1), e20084. https://doi.org/10.1002/ppj2.20084

Yu, X., Li, X., Guo, T., Zhu, C., Wu, Y., Mitchell, S. E., Roozeboom, K. L., Wang, D., Wang, M. L., Pederson, G. A., Tesso, T. T., Schnable, P. S., Bernardo, R., & Yu, J. (2016). Genomic prediction contributing to a promising global strategy to turbocharge gene banks. *Nature Plants*, *2*(10), 1–7. https://doi.org/10.1038/nplants.2016.150

Zhang, Z., Zhao, Y., Liao, X., Shi, W., Li, K., Zou, Q., & Peng, S. (2019). Deep learning in omics : A survey and guideline. *Briefings in Functional Genomics*, *18*(1), 41–57. https://doi.org/10.1093/bfgp/ely030

Zhu, X., Leiser, W. L., Hahn, V., & Würschum, T. (2021). Phenomic selection is competitive with genomic selection for breeding of complex traits. *The Plant Phenome Journal*, *4*(1), e20027. https://doi.org/10.1002/ppj2.20027

Zhu, X., Maurer, H. P., Jenz, M., Hahn, V., Ruckelshausen, A., Leiser, W. L., & Würschum, T. (2022). The performance of phenomic selection depends on the genetic architecture of the target trait. *Theoretical and Applied Genetics*, *135*(2), 653–665. https://doi.org/10.1007/s00122-021-03997-7

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bienvenu, C., Garin, V., Salas, N., Théra, K., Tekete, M. L., Sarathjith, M. C., Diallo, C., Berger, A., Calatayud, C., De Bellis, F., Rami, J.-F., Vaksmann, M., Segura, V., Pot, D., & de Verdal, H. (2025). Factors influencing phenomic prediction: A case study on a large sorghum back cross nested association mapping population. *The Plant Phenome Journal*, *8*, e70051. https://doi.org/10.1002/ppj2.70051