RESEARCH ARTICLE

Why Accuracy Metrics Fall Short in Comparing Phenomic and Genomic Prediction Models

Fangyi Wang[1] | Mitchell J Feldmann[1] | Daniel E Runcie[1,†]

[1]Department of Plant Sciences, University of California Davis, Davis, CA, 95616, USA

Correspondence
[†] Department of Plant Sciences, University of California Davis, Davis, CA, 95616, USA
Email: deruncie@ucdavis.edu

Funding information
United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) Specialty Crops Research Initiative (SCRI) (#2022-51181-38328); California Strawberry Commission; USDA NIFA Agriculture Genomes to Phenomes Initiative (AG2PI) (# 026418H); USDA NIFA AG2PI (#2024-70412-43648); USDA NIFA Agriculture and Food Research Initiative (AFRI) (#2024-67013-42591).

Phenomic Selection is a new paradigm in plant breeding that uses high-throughput phenotyping technologies and machine learning models to predict traits of new individuals and make selections. This can allow breeders to evaluate more plants in higher throughput more accurately, resulting in faster rates of gain and reduced labor costs. However, Phenomic Prediction models are frequently benchmarked against Genomic Prediction models using cross-validation to demonstrate their usefulness to breeders. We argue that this is inappropriate for two reasons: 1) Differences in the accuracy statistic measured by cross-validation do not reliably indicate differences in the accuracy parameter of the breeder's equation, and 2) Accuracy alone is insufficient to compare breeding schemes using Phenomic vs. Genomic Prediction because these tools differentially influence other parameters of the breeder's equation. We show analytically and through re-analysis of data from three representative Phenomic Prediction studies that conclusions about the superiority of Phenomic Prediction over Genomic Prediction change if compared using consistent methods. We conclude that Phenotypic Selection may be useful, but comparisons of accuracy between Genomic Prediction and Phenotypic Prediction models are not.

1 | INTRODUCTION

Breeders create new varieties by iterating through cycles of population creation, evaluation, and selection to create the next generation. The expected outcome of each cycle is predicted by the breeder's equation (Lynch et al., 1998):

$$\Delta g = \frac{i\rho\sigma_A}{L} \tag{1}$$

where Δg is the improvement in the average additive genetic value per year; i is the standardized selection intensity, the difference in the mean trait value of the selected individuals relative to the whole population, standardized by the standard deviation of trait values; ρ is the accuracy of selection, the Pearson Correlation Coefficient between the estimated and true additive genetic values among the selection candidates; σ_A is the standard deviation of additive genetic values; and L is the cycle length, the difference in years between when the population is created and when selections are made. Incorporating new procedures into breeding programs often results in changes to multiple of these parameters. Therefore, evaluating changes to a breeding program requires considering the

total effect on genetic gain.

In a traditional breeding program, selection decisions are based on phenotypic measurements of each candidate individual. In many species, collecting this data requires generating lines to grow in field trials and running replicated trials across multiple locations. Therefore, each selection cycle can take several years.

Meuwissen et al. (2001) introduced Genomic Selection as a tool to increase the rate of gain in breeding programs. Genomic Selection means selecting candidate individuals based on estimates of genetic values from a Genomic Prediction model using genetic marker data. Genomic Selection can increase the rate of gain (R) in a breeding program by increasing selection intensity (i), increasing selection accuracy (ρ), or reducing selection lengths (L). If genetic marker data are inexpensive, more candidates can be evaluated using a Genomic Prediction model than can be evaluated directly using field trials. Larger numbers of candidates allow breeders to impose a higher selection intensity while maintaining a constant number of selected individuals to conserve genetic variance (Bernardo et al., 2006; Cobb et al., 2019). Genomic Predictions of additive genetic values can be more accurate than direct estimates from field trials when candidates are genetically related at causal loci because the sample size per allele can be higher than the sample size per candidate (Zhong et al., 2009). Genomic Predictions of additive genetic values can be made as soon as genotype data are collected, even on seeds that have yet to germinate in some species (VanRaden, 2008). This can enable multiple cycles per year (Bhat et al., 2016). However, multiple factors continue to limit the use of Genomic Selection in plant breeding programs, including the cost of large-scale genotyping, the need to train and continuously update Genomic Prediction models, and the logistics of reorganizing existing breeding schemes.

Rincent et al. (2018) introduced Phenomic Selection as an alternative tool for breeders that can be less expensive and less logistically challenging than Genomic Selection. Like Genomic Selection, Phenomic Selection also relies on a statistical or Machine Learning-based model to make selections, called a "Phenomic Prediction" model, because instead of genetic markers, the features of the model are measured by high-throughput phenotyping technologies (HTP). For example, Rincent et al. (2018) used multispectral image data from wheat leaf and grains and poplar wood tissue, Winn et al. (2023) used spectral image data from a camera mounted on an aerial vehicle, and Krause et al. (2019) used hyperspectral image data from a camera mounted on an aircraft. Phenomic Selection has garnered interest in the breeding community for several reasons: i) HTP data can often be collected at lower cost and higher throughput than genotype data, enabling breeders to evaluate more candidates at once, or evaluate a common set of candidates more times to increase the heritability. ii) Logistically, Phenotypic Selection may require fewer changes to breeding schemes than Genomic Selection, and many of the same software packages used for implementing Genomic Prediction models also work for Phenomic Prediction models. iii) Phenomic Prediction models often achieve relative high predictive accuracy when evaluated using cross-validation. Since 2018, numerous papers have been published that demonstrate the predictive capabilities of Phenotypic Prediction models across crops and technologies (Rincent et al., 2018; Zhu et al., 2021; Winn et al., 2023; Robert et al., 2022b,a; Dallinger et al., 2023; Li et al., 2023; Lane et al., 2020; Lane and Murray, 2021; Weiß et al., 2022; Adak et al., 2023a,b,c; Herr et al., 2023; DeSalvio et al., 2022; Jackson et al., 2023; Brault et al., 2022; Roth et al., 2023; Van Tassel et al., 2022; Nguyen and Norton, 2020; Cuevas et al., 2019; Sandhu et al., 2021).

Both Genomic and Phenomic Selection schemes have the potential to accelerate plant breeding. But, since implementing either requires significant investment in equipment and changes to existing breeding schemes, there is considerable interest in determining which approach is best. Multiple recent papers have demonstrated the potential of Phenomic Selection by showing estimates of Prediction Accuracies of Phenomic Prediction models that are as high or higher than estimates of Prediction Accuracies of Genomic Prediction models made using the same datasets. However, our objective is to show that these comparisons *are not useful* for evaluating the relative benefits of implementing Phenotypic and Genomic Selection schemes in breeding programs. Our argument is based on the following two points:

1. The accuracy statistic reported by these studies is not an appropriate estimate of the selection accuracy parameter of the breeder's equation (ρ).
2. Genomic and Phenomic Selection schemes impact other

parameters of the breeder's equation differently, so comparisons need to consider the full effects on genetic gain.

We first demonstrate these points analytically. Then, using a more appropriate technique for estimating Prediction Accuracy we re-analyze published data from Rincent et al. (2018); Winn et al. (2023); Krause et al. (2019), to show that conclusions about the relative merits of Genomic and Phenomic change when compared using this method.

We conclude that Phenotypic Selection may be useful, but comparisons of accuracy between Genomic Prediction and Phenotypic Prediction models are not.

2 | THEORY

2.1 | Prediction Accuracy *vs* Predictive Ability

For simplicity, we assume that we have evaluated a set of n candidates by some method (*i.e.* phenotypic value, Genomic Prediction, or Phenomic Prediction), and that the candidates are unrelated and so independent and identically distributed. The accuracy of selecting among the candidates, represented by the parameter ρ in the breeder's equation, is the Pearson Correlation Coefficient between the estimated ($\hat{a}$) and true (a) breeding values (*i.e.* additive genetic values) of the candidates:

$$\rho = \text{cor}(\hat{a}, a) = \frac{\text{cov}(\hat{a}, a)}{\sqrt{\text{var}(\hat{a})}\sqrt{\text{var}(a)}}. \tag{2}$$

Unfortunately, it is not easy to create a consistent estimator of this parameter because additive genetic values (a) cannot be directly observed. Instead, we do observe *phenotypes* of candidates (y). The Pearson Correlation Coefficient between estimated breeding values and phenotypes (ρ_y) is:

$$\rho_y = \text{cor}(\hat{a}, y) = \frac{\text{cov}(\hat{a}, y)}{\sqrt{\text{var}(\hat{a})}\sqrt{var(y)}}. \tag{3}$$

We can easily form a consistent estimator of ρ_y as:

$$r_y = \widehat{\text{cor}}(\hat{a}, y) \tag{4}$$

using cross-validation (Legarra et al., 2008; Ould Estaghvirou et al., 2013), where $\widehat{\text{cor}}(\hat{a}, y)$ is the sample correlation. Legarra et al. (2008) called ρ_y "Predictive Ability" to distinguish it from "Prediction Accuracy" (ρ), but these terms are not used consistently in the plant breeding literature (*e.g.* Winn et al. (2023), Krause et al. (2019)). While r_y is a consistent estimator of ρ_y, it is not a consistent estimator of ρ. Thus, using r_y to predict responses to selection will usually not be accurate.

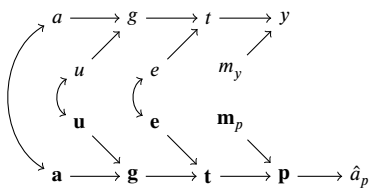
2.2 | Sources of bias

To understand the source of bias in r_y relative to ρ, consider a Phenotypic Selection scheme where $\hat{a} = y$, and a simple path diagram (Lynch et al., 1998; Dekkers, 2007) relating phenotypes (y) to additive genetic values (a):

$$\begin{tikzcd} & u \arrow[dr] & & e \arrow[dr] & & m \arrow[dr] & & \\ a \arrow[rr] & & g \arrow[rr] & & t \arrow[rr] & & y \arrow[rr] & & \hat{a} \end{tikzcd} \tag{5}$$

Here, arrows represent causal paths which are assumed to be linear. a and u are additive and non-additive genetic sources of variation, which together determine the total genetic value of a candidate (g). An organism's actual trait value t is caused by its genetic value plus an effect of its environment (e). We make an observation (y) which is subject to measurement error (m), and use this directly as our selection criterion ($\hat{a} = y$). Based on this diagram, $\text{cor}(\hat{a}, a) \neq \text{cor}(\hat{a}, y) = 1$ because there are exogenous variables that affect $\hat{a} = y$ but do not affect a (*i.e.*, u, e, and m), which we assume are uncorrelated with each other and with a. In particular, $\text{cov}(\hat{a}, y) \neq \text{cov}(\hat{a}, a)$ because u, e, and m are all common causes of y and $\hat{a}$, but not of a.

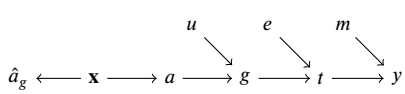
Next, consider a Phenomic Prediction model $f_p(\mathbf{p}) = \hat{a}_p$ with $\mathbf{p}$ a vector of phenomic features measured on a candidate. In a Phenomic Selection scheme, candidates will be ranked by $\hat{a}_p$, so this is the selection criterion: $\hat{a} = \hat{a}_p$. To see that $\rho_y \neq \rho$ for $f_p(\mathbf{p})$, we expand the path diagram to include causes of the phenomic features:



(6)

Here, we distinguish the vectors of additive genetic values (**a**), total genetic values (**g**), and underlying traits (**t**) for the phenomic features in **p**, as well as their vectors of non-additive genetic effects (**u**), environmental effects (**e**), and measurement errors ($\mathbf{m}_p$), from the corresponding values and effects underlying y. Additionally, various correlations among these effects are possible, the most likely ones of which are drawn with double-headed arrows between the additive genetic effects, non-additive genetic effects, and environmental effects. Thus, even though $f_p(\mathbf{p}) = \hat{a}_p$ doesn't involve y directly, **p** is caused by exogenous variables **u** and **e** that neither cause a nor are correlated with any causes of a, but are correlated with causes of y. This means that $\hat{a}_p \not\perp\!\!\!\perp y \mid a$, and thus $\text{cov}(\hat{a}_p, y) \neq \text{cov}(\hat{a}_p, a)$.

Finally, consider a Genomic Prediction model $f_g(\mathbf{x}) = \hat{a}_g$, with **x** a vector of genotypes for a candidate. We can expand the original path diagram to include the genetic features:



(7)

According to this path diagram, $\hat{a}_g \perp\!\!\!\perp y \mid a$ because there are no common exogenous variables impacting $\hat{a}_g$ and y that are not also shared with a. The only shared exogenous variable is **x** itself. Therefore, $\text{cov}(\hat{a}_g, a) = \text{cov}(\hat{a}_g, y)$. However, there are additional sources of variance in y that do not affect a (u, e, and m_y), so *i.e.* $\text{var}(y) \neq \text{var}(a)$. Thus, $\text{cor}(\hat{a}_g, a) \neq \text{cor}(\hat{a}_g, y)$.

Note that in this last diagram, we did not include a causal arrow from the genotypes (**x**) to the non-additive genetic effects (u). Non-additive genetic effects are caused by genetic features. But additive genetic values and dominance genetic values are uncorrelated (Su et al., 2012; Vitezica et al., 2013, 2016; Ertl et al., 2014; Muñoz et al., 2014; Aliloo et al., 2016, 2017; Xiang et al., 2016). If we use an additive encoding of these features that is orthogonal to dominance coefficients as in Vitezica et al. (2017), and use a Genomic Prediction model that is linear in the genetic features, like RRBLUP (Endelman, 2011), or Bayesian Alphabet models (Gianola, 2013), then we can assume approximately that $\text{cov}(f_g(\mathbf{x}), u) = 0$. In contrast, this independence will not hold for Genomic Prediction models with non-additive encodings of genotype information such as RKHS or Deep Learning models.

2.3 | Correction factors

In the case of an additive Genomic Prediction model, like RRBLUP, a simple analytical correction is possible to form a consistent estimator of ρ: Because $\text{cov}(\hat{a}_g, a) = \text{cov}(\hat{a}_g, y)$, ρ and ρ_y differ only due to the sources of variance in y that are not causes of a (*i.e.*, u, e, and m_y). Thus, $\rho = \rho_y\sqrt{\text{var}(y)}/\sqrt{\text{var}(a)}$, so r_y tends to *underestimate* ρ (Figure 1A). This is the basis for the estimator of ρ proposed by Legarra et al. (2008): $r = r_y/\sqrt{\widehat{h^2}}$, where $\widehat{h^2}$ is the estimate of the narrow-sense heritability of y.

This correction factor is not valid for Phenomic Prediction because $\text{cov}(\hat{a}_p, a) \neq \text{cov}(\hat{a}_p, y)$ as shown above. Instead, assuming the linear model $y = a + u + e + m_y$, we can show (see Appendix):

$$\rho = \frac{\rho_y - \rho_e\sqrt{1 - h^2}}{\sqrt{h^2}}, \tag{8}$$

where $\rho_e = \text{cor}(\hat{a}_p, u + e + m_y)$. Even if phenomic features (**p**) and the trait (y) are not measured on the same plots, so $\text{cov}(\hat{a}, e + m_y) = 0$, non-additive genetic variation (u) can still cause $\rho_e \neq 0$. The impact of ρ_e on the bias of $r = r_y/\sqrt{\widehat{h^2}}$ depends on the narrow-sense heritability (h^2), becoming stronger for low-heritability traits. Therefore, cross-validation-based calculations of r_y for a Phenomic Prediction model can either underestimate *or overestimate* the actual Prediction Accuracy (Figure 1B).

To demonstrate this, we highlight some examples:

1. If the heritability is low (*e.g.* $h^2 = 0.1$) and $\rho_y = 0.25$ for a Genomic Prediction model, its actual Prediction Accuracy

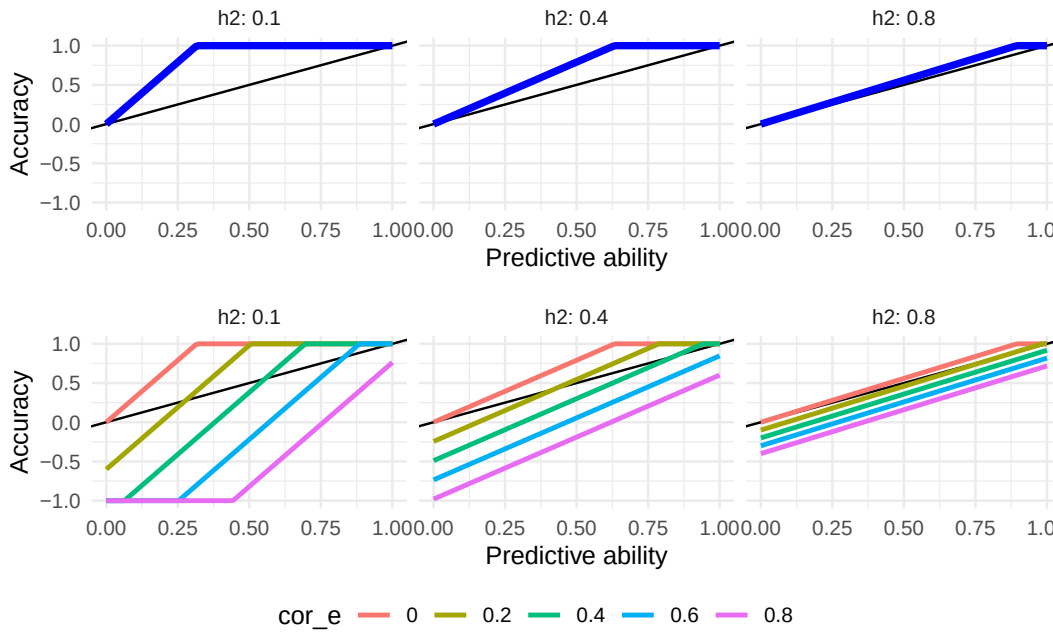


FIGURE 1 Predictive Ability (ρ_y) and Accuracy (ρ) for Genomic and Phenomic Prediction models

is much higher: $\rho = \rho_y/\sqrt{h^2} = 0.25/\sqrt{0.1} = 0.79$. If instead $\rho_y = 0.25$ for a Phenomic Prediction model, the actual Prediction Accuracy could be as high as $\rho = 0.79$ if $r_e = \text{cor}(\hat{a}_p, u + e + m_y) = 0$. But this is unlikely unless phenotypes (y) and phenomic features **p** are measured on different plots and σ_u^2 is very small. However, r could be lower than r_y if $\text{cor}(\hat{a}_p, u + e + m_y) > 0.18$. If the correlation of $\text{cor}(\hat{a})_p, a) = \text{cor}(\hat{a})_p, u + e + m_y)$, the actual accuracy would be $\rho = 0.04$.

2. If the heritability is moderate (*e.g.* $h^2 = 0.4$) and $\rho_y = 0.5$ for a Genomic Prediction model, its actual Prediction Accuracy is much higher: $\rho = \rho_y/\sqrt{h^2} = 0.5/\sqrt{0.4} = 0.79$. If instead $\rho_y = 0.5$ for a Phenomic Prediction model, the actual Prediction Accuracy could be as high as $\rho = 0.79$, but would equal $\rho = 0.2$ if the correlations of predictions with a and $u+e+m_y$ are equal, and could be as low as $\rho = 0$ if the phenomic data is driven more by environmentally-sensitive traits (so ρ_e is larger).
3. If the heritability is high (*e.g.* $h^2 = 0.8$) differences between ρ_y and ρ are smaller for both Genomic and Phenomic Prediction models. But even if ρ_y is very high for a Phenomic Prediction model (*e.g* $\rho_y = 0.8$), the actual accuracy can be $\rho < 0.6$.

While the difference between ρ_y and ρ have been known for many years (Legarra et al., 2008; Ould Estaghvirou et al., 2013), many studies continue to report r_y values when comparing Genomic Prediction models. Despite r_y tending to underestimate the actual Prediction Accuracy, comparing r_y values between two (additive) Genomic Prediction models (say model (1) is RRBLUP and model (2) is BayesCπ) is valid *if the comparison is based on the ratio* $\frac{r_y^{(1)}}{r_y^{(2)}}$ (Legarra et al., 2008):

$$\frac{r_y^{(1)}}{r_y^{(2)}} = \frac{r^{(1)}/\sqrt{h^2}}{r^{(2)}/\sqrt{h^2}} = \frac{r^{(1)}}{r^{(2)}}.$$

Therefore, the conclusion of the superiority of one model over the other would be the same whether r_y or r were used. The same is not true if models are compared based on the

difference of r_y estimates:

$$\begin{aligned} r_y^{(1)} - r_y^{(2)} &= r^{(1)}/\sqrt{h^2} - r^{(2)}/\sqrt{h^2} \\ &= (r^{(1)} - r^{(2)})/\sqrt{h^2} \\ &\neq r^{(1)} - r^{(2)}, \end{aligned}$$

although the *sign* of the comparison remains the same. However, as shown in Figure 1, if a Genomic Prediction and Phenomic Prediction model are compared either as a ratio (r_y^p/r_y^g) or a difference ($r_y^p - r_y^p$), both the sign and the magnitude of the result can differ from the true values (r^p/r^g or $r^p - r^g$, respectively). This means that conclusions about the superiority of Phenomic Prediction over Genomic Prediction based on comparisons of Predictive Ability (r_y) are unreliable.

2.4 | Alternative estimates of accuracy

Legarra and Reverter (2018) proposed an alternative estimator of Prediction Accuracy called the "LR method". Unfortunately this method is not appropriate for Phenomic Prediction models because the method assumes that $\text{cor}(\hat{a}, u + e) = 0$ which is generally violated for Phenomic Prediction models, as discussed above.

Runcie and Cheng (2019) demonstrated three alternative measures of accuracy more appropriate for Phenomic Prediction models, or models that combined phenomic and genomic features. Of these, the “parametric" approach is the most generally applicable, although it requires genetic or pedigree relationships among testing individuals to compute. Specifically, a consistent estimator of ρ can be calculated as (Lopez-Cruz et al., 2020):

$$r = \text{cor}_g(\hat{a}, y)\sqrt{h^2_{\hat{a}}}, \tag{9}$$

where $\text{cor}_g(\hat{a}, y)$ is an estimate of the genetic correlation between predictions ($\hat{a}$) and phenotypes (y) from a bivariate linear mixed model fit to testing candidates not used for model training, and $h^2_{\hat{a}}$ is the heritability of $\hat{a}$. Large sets of testing candidates are needed to estimate this parameter accurately. We use this method below to re-estimate accuracies of Phenomic Prediction models using data from three published studies because it is easier to estimate than equation 8.

3 | MATERIALS AND METHODS

We replicated the comparisons between Genomic and Phenomic Prediction Accuracies reported in three published studies to test how each set of conclusions would change if Prediction Accuracies were estimated using equation 9 instead of estimated Predictive Ability. In each case, we attempted to reproduce the Genomic and Phenomic Prediction models exactly as reported in each original paper.

3.1 | Case Study 1

Rincent et al. (2018) used two datasets, one from wheat and one from poplar, to compare the Predictive Abilities of Genomic and Phenomic Prediction models.

The wheat data consisted of 223 genotypes measured for heading date and grain yield in two treatments (dry and irrigated). Genomic data consisted of 84,529 SNPs. Phenomic data consisted of near-infrared spectroscopy scans from both leaf and grain tissues in each treatment. Phenotypic data were reported as best linear unbiased estimators (BLUEs) of genotype means correcting for random block effects. We considered the 4 trait-treatment combinations as 4 traits (heading date in the dry treatment, heading date in the irrigated treatment, grain yield in the dry treatment, and grain weight in the irrigated treatment), and used their corresponding phenomic data to train Genomic and Phenomic Prediction models following the methods in the source paper.

The poplar data consisted of 562 genotypes measured for height, circumference, bud flush, bud set, and rust resistance in two locations: Orleans, France and Savigliano, Italy. All 5 traits were measured at Orleans, but only circumference, bud flush and bud set were measured at Savigliano. Genomic data consisted of 7,808 SNPs. Phenomic data were collected from the wood samples in both locations. Phenotypic data were reported as BLUEs of genotype means correcting for random block effects. We considered each of the 8 trait-location combination as independent traits and used the corresponding phenomic data to train Genomic and Phenomic Prediction models following the methods in the source paper.

For each trait in each species, we fit the following Genomic Prediction model following the methods of Rincent et al. (2018):

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\alpha} + \epsilon_i \tag{10}$$

where y_i is the BLUE of the phenotypic value of genotype i, $\boldsymbol{x}_i$ is the SNP markers for genotype i, $\boldsymbol{\alpha} = \{\alpha_k\}$, $\alpha_k \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_a^2)$ are the random marker effects, and $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\epsilon^2)$ are residuals.

Also following Rincent et al. (2018), we fit the following Phenomic Prediction model for each trait in each species:

$$y_i = \boldsymbol{p}_i^T \boldsymbol{h} + e_i \tag{11}$$

where $\boldsymbol{p}_i$ is the phenomic features for genotype i, $\boldsymbol{h} = \{h_k\}$, $h_k \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_h^2)$ are the random phenomic feature effects, and $e_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2)$ are residuals.

We estimated Predictive Ability as sample Pearson's Correlation Coefficient between predictions and observed BLUEs using eight-fold cross-validation for the wheat data, and using five-fold cross-validation for the poplar data, following Rincent et al. (2018). We estimated Prediction Accuracies using Equation 9 by fitting a bivariate linear mixed model between the observed phenotypic traits and the predicted values from either the Genomic or Phenomic Prediction models, using the marker-based kinship matrix as the estimate of additive genetic relationships among individuals.

3.2 | Case Study 2

Winn et al. (2023) compared Genomic and Phenomic Predictive Abilities for grain yield in wheat. The dataset consisted of grain yield data for 3,442 genotypes measured across multiple replicated trials across multiple years and locations. Genotype data consisted of 23,130 SNPs. We removed SNPs with a minor allele frequency of or less than 5%.Phenomic data were derived from images taken by unoccupied aerial vehicles from the same genotypes across different years and different locations. We followed Scenario 1 in Winn et al. (2023) and calculated BLUEs of grain yield and of each spectral band at each genotype, correcting for the random effect of unique combinations of year and location. For Genomic Prediction, we used equation 10 as above, matching the Genomic Prediction model reported by Winn et al. (2023). For Phenomic Prediction, we followed the same methods of Winn et al. (2023) and used the following linear model which treats the phenomic features as fixed effects:

$$y_i = \boldsymbol{p}_i^T \boldsymbol{l} + \epsilon_i \tag{12}$$

where $\boldsymbol{l}$ is the vector of fixed effects of the phenomic features. We also used a linear mixed model, equivalent to Equation 11, which treats the phenomic features as mixed effects for Phenomic Prediction.

We used five-fold cross-validation to obtain the Pearson's correlation between the predicted values and the BLUEs. We estimated the Prediction Accuracy in the same way as in Case Study 1.

3.3 | Case Study 3

Krause et al. (2019) also compared Genomic and Phenomic Predictive Abilities for grain yield in wheat. The dataset consisted of grain yield data for 3,771 genotypes measured from multiple replicates on the same genotype across different breeding cycles and different management practices. The authors considered each combination of breeding cycle and management practice as a unique treatment. Genomic data were provided as a genomic relationship matrix calculated from 34,900 SNPs. Phenomic data consisted of hyperspectral reflectances collected with a hyperspectral camera mounted in an aircraft. The phenomic data were collected multiple times during the growing season for each treatment. The authors provided the BLUEs of grain yield and of each hyperspectral band for each genotype within each treatment, correcting for random effects of trial, replicate, and block. This was described as the **ALL** method in Krause et al. (2019).

For the grain yield of wheat in Krause et al. (2019), we fit a Genomic Prediction model for genotype i in treatment j as follows:

$$y_{i,j} = Trt_j + g_i + \epsilon_{i,j} \tag{13}$$

where Trt_j is the fixed effect for treatment j, $\boldsymbol{g} = \{g_i\}$, $\boldsymbol{g} \sim \mathrm{N}(\boldsymbol{0}, \sigma_g^2 \boldsymbol{K}_g)$ is the random additive genetic effects with

genomic relationship matrix $\boldsymbol{K}_g$ and additive genetic variance parameter σ_a^2. Residuals $\epsilon_{i,j} \overset{\text{iid}}{\sim} \text{N}(0, \sigma_\epsilon^2)$ were assumed independent and identically distributed among genotypes and treatments as in Krause et al. (2019).

Following Krause et al. (2019), we used the following Phenomic Prediction model for genotype i in treatment j:

$$y_{i,j} = Trt_j + Pm_{i,j} + \epsilon_{i,j} \tag{14}$$

where $y_{i,j}$ is the BLUE of phenotypic values of genotype i in treatment j, Trt_j is the fixed effect of treatment j, $Pm_{i,j} \sim \text{N}(\mathbf{0}, \sigma_H^2 \boldsymbol{K}_H)$ is the random effect of phenomic features on each genotype in each treatment with feature covariance matrix $\boldsymbol{K}_H = \frac{\boldsymbol{P}\boldsymbol{P}^T}{nbands}$, where $\boldsymbol{P}$ is the centered and standardized BLUEs of hyperspectral bands of genotypes in each treatment, and $nbands$ is the total number of hyperspectral bands. This Phenomic Prediction model is the same model described in Equation 5 in Krause et al. (2019).

We measured Predictive Ability using a leave-one-treatment-out strategy, training each model using data from all-but-one treatment and then measuring the Pearson's correlation between predictions and grain yield BLUEs in that remaining treatment. This was repeated for each treatment. We estimated the Prediction Accuracy in the same way as in Case Study 1.

In this case study, the genotypes were measured in multiple trials. The Genomic Prediction used the main effect of genotypes as the prediction, but Phenomic Prediction made trial-specific predictions, because genotypes were constant across trials, but phenomic features were not.

3.4 | Software

All analyses were conducted in R version 4.2.3 (R Core Team, 2024). Model 12 was fitted using the *lm* function in Base R. Model 10 and 11 were fitted using *mixed.solve* function from the *rrBLUP* package (Endelman, 2011). Model 13 and 14 were fitted using the *BGLR* function from *BGLR* package (Pérez and de Los Campos, 2014). Prediction Accuracies were estimated using the *estimate_gcor* function from the *MegaLMM* package (Runcie et al., 2021). BLUEs were calculated using the *lme4* and *emmeans* packages (Bates et al., 2015; Lenth, 2024). Graphs were made with the *ggplot2* and *cowplot* packages (Wickham, 2016; Wilke, 2020).

4 | RESULTS

We compared the estimated Predictive Abilities (r_y for $\rho_y = \text{cor}(\hat{a}, y)$) and Prediction Accuracies (r for $\rho = \text{cor}(\hat{a}, a)$) for both Genomic and Phenomic Prediction models using each of the datasets described above.

Rincent et al. (2018) reported two cases among the 4 wheat traits and 8 poplar traits with higher Predictive Abilities for Phenomic Prediction than Genomic Prediction (Rincent et al. (2018): poplar circumference in Orleans location and wheat grain yield in irrigated condition. They estimated Genomic and Phenomic Predictions for poplar circumference in Orleans location to be $\approx r_y = 0.6$ and $\approx r_y = 0.7$, respectively and those for wheat grain yield to be $\approx r_y = 0.35$ for Genomic Prediction and $\approx r_y = 0.55$ for Phenomic Prediction with grain samples, and $\approx r_y = 0.30$ with leaf samples. Winn et al. (2023), also reported higher Predictive Ability for wheat grain yield using the fixed-effect Phenomic Prediction model ($\approx r_y = 0.42$) vs. the Genomic Prediction model ($\approx r_y = 0.35$). We re-analyzed each data set using the same methods and replicated each of these results (Figure 2, a - c, left panels, Table 1). To be consistent with the other cases, we also fit a random-effect Phenomic Prediction model for the Winn et al. (2023) dataset (Figure 2 c, red) and estimated a similar Predictive Ability. Using the data from Krause et al. (2019), we observed similar, but slightly lower Predictive Ability for wheat grain yield using Phenomic Prediction than Genomic Prediction (0.15 vs 0.28).

Then, for each model in each dataset, we used Equation 9 to estimate Prediction Accuracy (ρ, Figure 2 a,b). None of the Phenomic Prediction models had significantly higher estimated Prediction Accuracy than the corresponding Genomic Prediction models. In all cases, estimates of Prediction Accuracy for Genomic Prediction models were higher than estimates of Predictive Ability, as expected because $h^2 < 1$. In 5/6 cases, estimates of Prediction Accuracy were slightly higher for the Phenomic Prediction models as well, but with less difference from Predictive Ability than for the Genomic Prediction models. Therefore, by reporting Predictive Ability

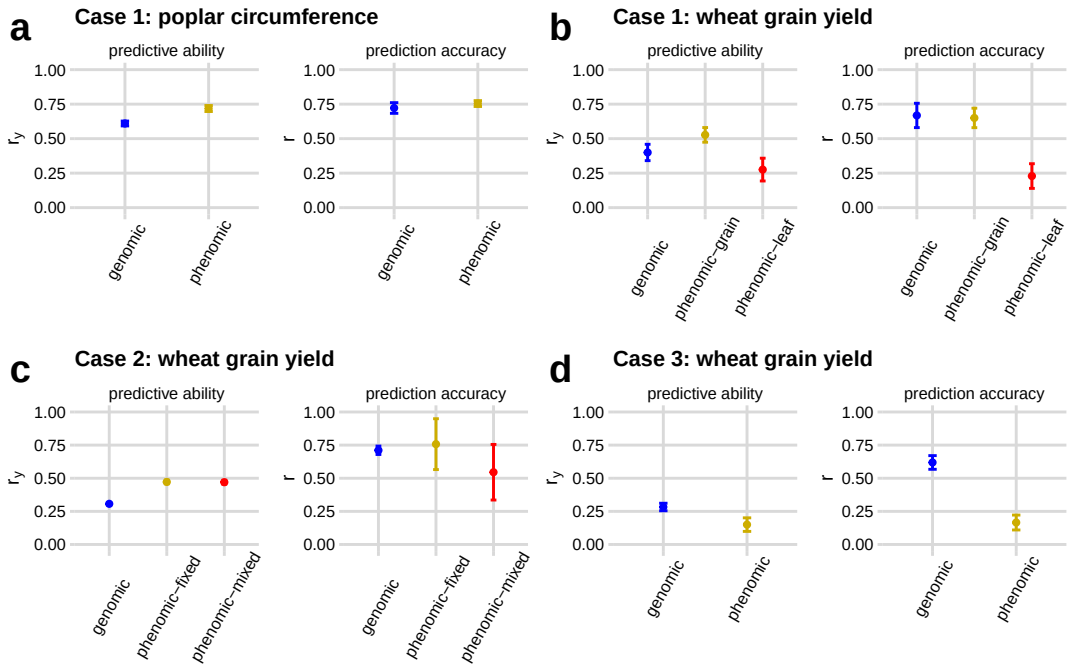


FIGURE 2 Panels show the mean and standard errors of Predictive Ability (r_y) and Prediction Accuracy (r) for Genomic (blue) and Phenomic (gold and red) Prediction models on poplar and wheat data from Rincent et al. (2018) (a, b), Winn et al. (2023) (c), and Krause et al. (2019) (d). The traits of interest are poplar circumference in Orleans location (a), wheat grain yield in irrigated condition (b) and wheat grain yield (c - d).

instead of predictive accuracy, the authors *underestimated* predictive accuracy for almost models, but the underestimation was greater for Genomic Prediction models, leading to the impression that Phenomic Prediction models had an advantage when they did not.

5 | DISCUSSION

Our main goal with this article is to clearly demonstrate that comparing *Predictive Ability* estimates (r_y, the Pearson Correlation Coefficient between predicted genetic values and observed traits) between Phenomic Prediction and the Genomic Prediction models is not an appropriate way to assess the relative utility of Phenomic Selection and Genomic Selection for breeding programs. As we show analytically and using previously published empirical results, Predictive Ability is a biased estimate of the *Prediction Accuracy* of genetic values (ρ), a key parameter of the Breeder's equation that determines the genetic gain after selection. While the Predictive Ability of an (additive) Genomic Prediction model is always less than its Prediction Accuracy, the Predictive Ability of a Phenomic Prediction model can be greater than its Prediction Accuracy. For example, under reasonable conditions, a predictable ability of $\rho_y = 0.5$ for a Genomic Prediction model can indicate a Prediction Accuracy of $\rho = 0.75$ (with $h^2 = 0.4$). But under the same conditions, a Predictive Ability of $\rho_y = 0.5$ for a Phenomic Prediction model might indicate a Prediction Accuracy of $\rho = 0.2$ (if residual and genetic correlations with the true genetic value are equal), or even lower (if predictions are dominated by non-genetic sources of variation). Therefore, comparisons of Predictive Abilities between Genomic and Phenomic Prediction models will tend to give the false impression that Phenomic Prediction models are more accurate, even when they are not. We highlighted four examples

TABLE 1 The estimated means and standard errors (in parentheses) of Predictive Abilities r_y, Prediction Accuracies r, and heritabilities h^2 of poplar and wheat traits across cross-folds.

Case Study	Crop	Trait	Prediction	r_y	r	h^2
Rincent et al. (2018)	Poplar	Orl_Height	Genomic	0.565(0.030)	0.684(0.032)	0.687(0.043)
		Orl_Height	Phenomic	0.656(0.010)	0.714(0.077)	0.687(0.043)
Rincent et al. (2018)	Poplar	Orl_BudFlush	Genomic	0.722(0.019)	0.801(0.026)	0.813(0.012)
		Orl_BudFlush	Phenomic	0.095(0.055)	0.017(0.021)	0.813(0.012)
Rincent et al. (2018)	Poplar	Sav_BudFlush	Genomic	0.729(0.020)	0.848(0.041)	0.748(0.043)
		Sav_BudFlush	Phenomic	0.333(0.039)	0.201(0.028)	0.748(0.043)
Rincent et al. (2018)	Poplar	Orl_BudSet	Genomic	0.751(0.031)	0.86(0.050)	0.745(0.049)
		Orl_BudSet	Phenomic	0.545(0.020)	0.55(0.028)	0.745(0.049)
Rincent et al. (2018)	Poplar	Sav_BudSet	Genomic	0.608(0.040)	0.669(0.043)	0.828(0.030)
		Sav_BudSet	Phenomic	0.337(0.017)	0.352(0.016)	0.828(0.030)
Rincent et al. (2018)	Poplar	Orl_Circumference	Genomic	0.61(0.018)	0.722(0.039)	0.726(0.048)
		Orl_Circumference	Phenomic	0.718(0.022)	0.754(0.021)	0.726(0.048)
Rincent et al. (2018)	Poplar	Sav_Circumference	Genomic	0.777(0.017)	0.987(0.013)	0.542(0.024)
		Sav_Circumference	Phenomic	0.809(0.014)	0.736(0.024)	0.542(0.024)
Rincent et al. (2018)	Poplar	Orl_RustResistance	Genomic	0.622(0.015)	0.681(0.024)	0.838(0.035)
		Orl_RustResistance	Phenomic	0.443(0.046)	0.452(0.044)	0.838(0.035)
Rincent et al. (2018)	Wheat	Dry_GrainYield	Genomic	0.238(0.067)	0.407(0.121)	0.616(0.128)
		Dry_GrainYield	Phenomic-grain	0.543(0.047)	0.641(0.084)	0.616(0.128)
		Dry_GrainYield	Phenomic-leaf	0.314(0.066)	0.334(0.077)	0.616(0.128)
Rincent et al. (2018)	Wheat	Dry_HeadingDate	Genomic	0.487(0.041)	0.657(0.066)	0.616(0.135)
		Dry_HeadingDate	Phenomic-grain	0.806(0.033)	0.855(0.013)	0.616(0.135)
		Dry_HeadingDate	Phenomic-leaf	0.848(0.020)	0.893(0.017)	0.616(0.135)
Rincent et al. (2018)	Wheat	Irr_GrainYield	Genomic	0.4(0.059)	0.668(0.088)	0.58(0.159)
		Irr_GrainYield	Phenomic-grain	0.527(0.053)	0.65(0.071)	0.58(0.159)
		Irr_GrainYield	Phenomic-leaf	0.276(0.083)	0.229(0.089)	0.58(0.159)
Rincent et al. (2018)	Wheat	Irr_HeadingDate	Genomic	0.511(0.033)	0.662(0.056)	0.682(0.096)
		Irr_HeadingDate	Phenomic-grain	0.618(0.027)	0.74(0.035)	0.682(0.096)
		Irr_HeadingDate	Phenomic-leaf	0.8(0.023)	0.862(0.027)	0.682(0.096)
Winn et al. (2023)	Wheat	GrainYield	Genomic	0.307(0.010)	0.711(0.032)	0.189(0.014)
		GrainYield	Phenomic-fixed	0.472(0.014)	0.757(0.192)	0.189(0.014)
		GrainYield	Phenomic-mixed	0.47(0.015)	0.545(0.21)	0.189(0.014)
Krause et al. (2019)	Wheat	GrainYield	Genomic	0.283(0.028)	0.619(0.052)	0.201(0.020)
		GrainYield	Phenomic	0.15(0.051)	0.166(0.056)	0.201(0.020)

of comparisons between Phenomic and Genomic Prediction models from the recent literature, using the same datasets to directly estimate Prediction Accuracy using an alternative method, and show that the conclusions about the superiority of one method over the other frequently change when the more appropriate statistic is used to compare models.

However, even if appropriate methods are used to estimate predictive accuracy (ρ) for both Genomic and Phenomic Prediction models, comparing these accuracy estimates directly is still not sufficient to compare Genomic and Phenomic Selec-

tion breeding schemes. Breeding schemes should be evaluated based on their expected genetic gains over a specific time interval, and using a specific budget. Comparisons between estimates of ρ alone are insufficient for this because Genomic and Phenomic Selection schemes are likely to impact the other three parameters of the Breeder's equation differently, which can overcome differences in accuracy. The other three parameters of the breeder's equation are L, the cycle length, σ_A, the additive genetic standard deviation, and i, the standardized selection intensity.

The most consequential parameter in the Breeder's equation is generally L, the cycle length. In many breeding programs, each cycle takes multiple years. For example, in Bhat et al. (2016); Tessema et al. (2020) (wheat), the cycle is $\approx 6-7$ years. In the University of California strawberry breeding program, we estimated the average cycle length to be 10 years (Pincot et al., 2021). In its most aggressive form, Genomic Selection can be used to shorten breeding cycles to $\ll 1$ year, using the two-part GS scheme proposed by Gaynor et al. (2017), in which candidates are selected based on seed or seedling genotypes without any phenotypic evaluation. This strategy is generally not possible for Phenomic Selection because most Phenomic Prediction models require plants to be grown to mature stages to collect phenomic data. Most apparently successful Phenomic Prediction models for yield and quality-related traits use features collected in field trials of candidates close in time to when yield measures are taken (Winn et al., 2023; Krause et al., 2019). Applied this way, Phenomic Selection would not shorten cycle lengths significantly. Thus, a Genomic Selection scheme like that proposed by Gaynor et al. (2017) would be preferred over a Phenomic Selection scheme even if the Genomic Prediction Accuracy was lower than the Phenomic Prediction Accuracy.

The aggressive two-part GS scheme proposed by Gaynor et al. (2017) is not currently widely used. More commonly, GS only slightly shortens breeding cycles, for example, using Genomic Prediction to select at the F5 or PYT stages of a wheat breeding scheme instead of the AYT stage (Tessema et al., 2020). In this case, differences in L may not be large between a Genomic Selection and Phenomic Selection scheme. However, one of the principle arguments for Phenomic Selection is that is that phenomic data can be collected more cheaply and in higher throughput than Genomic Data. If so,

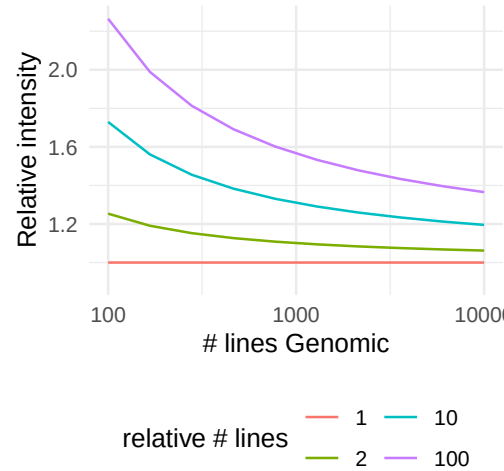


FIGURE 3 Relationship between population size and selection intensity

and if each field plot is sufficiently inexpensive, Phenomic Predictions could be made on much larger numbers of candidates, enabling breeders to increase selection intensity. The relationship between population size and selection intensity can be approximated as:

$$i \approx \varphi(\Phi^{-1}(1 - k/n))(n/k)$$

where n is the number of candidates, k is the target number of selections, and $\varphi(.)$ is the density function and $\Phi^{-1}(.)$ is the inverse of the cumulative density function of the standard normal distribution. Figure 3 shows that i is a function of both n and k/n, the proportion selected. Larger n increases i but at a relatively slow rate, as discussed by Cobb et al. (2019). Much larger numbers of candidates must be evaluated for the increase in i enabled by Phenomic Selection to compensate for decreased Prediction Accuracy. Also, population sizes are often limited by field space and field costs, so it may not be common for the number of selection candidates to be significantly limited when applying Genomic Selection. Therefore, Phenomic Selection schemes are unlikely to have a major advantage over Genomic Selection solely based on increased intensity.

The effects of Genomic Selection and Phenomic Selection

on the maintenance of genetic variance σ_A^2 are less clear than for the other parameters of the Breeder's equation. Jannink (2010) and Goddard et al. (2011) argued that Genomic Selection should maintain genetic variance longer than pedigree-based selction because genetic markers can be used to select individuals with good alleles at QTL but which are otherwise unrelated. However, others have argued that Genomic Selection can rapidly reduce genetic variance, and multiple strategies have been suggested to mitigate against rapid genetic variance loss during Genomic Selections. Phenomic Selection is really a form of phenotypic selection, thus the consequences of Phenomic Selction on genetic variance are likely similar. In either strategy, careful management of genetic variance will be necessary to maintain long-term gains.

Finally, comparisons of Phenomic and Genomic Prediction ability generally use training datasets that are the same size in an attempt to make the comparison fair. However, this choice itself likely biases the comparison between these approaches. Even if r could be estimated accurately, the actual Prediction Accuracy using either strategy will depend on the size and composition of the training data for each model. Since phenomic data are generally cheaper to acquire than genomic, an optimized Phenomic Selection scheme would likely permit larger training data sets than an optimized Genomic Selection scheme. Therefore, Phenomic and Genomic Prediction models should be made when each model is trained on an appropriately sized training dataset.

Our goal is not to suggest that Phenomic Selection cannot be useful for breeding. When Phenomic Prediction is used to increase the precision or throughput of phenomic evaluation of candidates, this can certainly improve breeding outcomes. In particular, if Phenomic Prediction is used to improve the phenotype data available to train Genomic Prediction models, this can greatly improve breeding outcomes. In this application, it is actually the *Predictive Ability* of a Phenomic Prediction model that matters - the ability of that model to produce high accuracy replacements for the trait data needed to train an accurate Genomic Prediction model. Therefore, we suggest that Phenomic Prediction studies *should* report Predictive Ability, while Genomic Prediction studies should report Prediction Accuracy. But these two numbers should not be directly compared - they measure different things.

6 | CONCLUSIONS

To summarize, we argue that comparisons between Predictive Abilities of Genomic and Phenomic Prediction models are largely uninformative for breeders. These comparisons use a statistic that is not directly relevant to breeding goals, and neglect the impact of each strategy on other equally important consequences of the implementation of Genomic and Phenomic selection schemes.

7 | DATA AVAILABILITY

The codes used in this study are available at `https://github.com/Faye-Wong-stat/breeder_equation`

8 | ACKNOWLEDGEMENTS

The authors would like to acknowledge helpful comments from Hans-Peter Piepho. UC Davis is located at Davis, California, which is located on land that was the home of the Patwin people for thousands of years.

9 | CONFLICT OF INTEREST

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

10 | AUTHOR CONTRIBUTIONS

FW, DER, and MJF contributed to the conception of this project. FW performed simulations and analyses. FW, DER, MJF wrote and edited the article.

REFERENCES

Adak, A., Anderson, S. L. and Murray, S. C. (2023a) Pedigree-management-flight interaction for temporal phenotype analysis and temporal phenomic prediction. *The Plant Phenome Journal*, **6**, e20057.

Adak, A., Kang, M., Anderson, S. L., Murray, S. C., Jarquin, D., Wong, R. K. and Katzfuß, M. (2023b) Phenomic data-

driven biological prediction of maize through field-based high throughput phenotyping integration with genomic data. *Journal of Experimental Botany*, erad216.

Adak, A., Murray, S. C. and Anderson, S. L. (2023c) Temporal phenomic predictions from unoccupied aerial systems can outperform genomic predictions. *G3*, **13**, jkac294.

Aliloo, H., Pryce, J. E., González-Recio, O., Cocks, B. G., Goddard, M. E. and Hayes, B. J. (2017) Including nonadditive genetic effects in mating programs to maximize dairy farm profitability. *Journal of Dairy Science*, **100**, 1203–1222. URL: https://www.journalofdairyscience.org/article/S0022-0302(16)30834-7/fulltext. Publisher: Elsevier.

Aliloo, H., Pryce, J. E., González-Recio, O., Cocks, B. G. and Hayes, B. J. (2016) Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. *Genetics Selection Evolution*, **48**, 8. URL: https://doi.org/10.1186/s12711-016-0186-0.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48.

Bernardo, R., Moreau, L. and Charcosset, A. (2006) Number and Fitness of Selected Individuals in Marker-Assisted and Phenotypic Recurrent Selection. *Crop Science*, **46**, 1972–1980. URL: https://onlinelibrary.wiley.com/doi/abs/10.2135/cropsci2006.01-0057. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2135/cropsci2006.01-0057.

Bhat, J. A., Ali, S., Salgotra, R. K., Mir, Z. A., Dutta, S., Jadon, V., Tyagi, A., Mushtaq, M., Jain, N., Singh, P. K., Singh, G. P. and Prabhu, K. V. (2016) Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding. *Frontiers in Genetics*, **7**, 221. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5186759/.

Brault, C., Lazerges, J., Doligez, A., Thomas, M., Ecarnot, M., Roumet, P., Bertrand, Y., Berger, G., Pons, T., François, P. et al. (2022) Interest of phenomic prediction as an alternative to genomic prediction in grapevine. *Plant Methods*, **18**, 108.

Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., Hagen, T., Quinn, M. and Ng, E. H. (2019) Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, **132**, 627–645. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6439161/.

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

Dallinger, H. G., Löschenberger, F., Bistrich, H., Ametz, C., Hetzendorfer, H., Michel, S. and Buerstmayr, H. (2023) Predictor bias in genomic and phenomic selection.

Dekkers, J. (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. *Journal of animal breeding and genetics*, **124**, 331–341.

DeSalvio, A. J., Adak, A., Murray, S. C., Wilde, S. C. and Isakeit, T. (2022) Phenomic data-facilitated rust and senescence prediction in maize using machine learning algorithms. *Scientific reports*, **12**, 7571.

Endelman, J. B. (2011) Ridge regression and other kernels for genomic selection with r package rrblup. *The plant genome*, **4**.

Ertl, J., Legarra, A., Vitezica, Z. G., Varona, L., Edel, C., Emmerling, R. and Götz, K.-U. (2014) Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. *Genetics Selection Evolution*, **46**, 40. URL: https://doi.org/10.1186/1297-9686-46-40.

Gaynor, R. C., Gorjanc, G., Bentley, A. R., Ober, E. S., Howell, P., Jackson, R., Mackay, I. J. and Hickey, J. M. (2017) A two-part strategy for using genomic selection to develop inbred lines. *Crop Science*, **57**, 2372–2386.

Gianola, D. (2013) Priors in Whole-Genome Regression: The Bayesian Alphabet Returns. *Genetics*, **194**, 573–596. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3697965/.

Goddard, M. E., Hayes, B. J. and Meuwissen, T. H. E. (2011) Using the genomic relationship matrix to predict the accuracy of genomic selection. *Journal of Animal Breeding and Genetics = Zeitschrift Fur Tierzuchtung Und Zuchtungsbiologie*, **128**, 409–421.

Herr, A. W., Adak, A., Carroll, M. E., Elango, D., Kar, S., Li, C., Jones, S. E., Carter, A. H., Murray, S. C., Paterson, A. et al. (2023) Unoccupied aerial systems imagery for phenotyping in cotton, maize, soybean, and wheat breeding. *Crop Science*, **63**.

Jackson, R., Buntjer, J. B., Bentley, A. R., Lage, J., Byrne, E., Burt, C., Jack, P., Berry, S., Flatman, E., Poupard, B. et al. (2023) Phenomic and genomic prediction of yield on multiple locations in winter wheat. *Frontiers in genetics*, **14**, 1164935.

Jannink, J.-L. (2010) Dynamics of long-term genomic selection. *Genetics Selection Evolution*, **42**, 35. URL: https://doi.org/10.1186/1297-9686-42-35.

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. and Mondal, S. (2019) Hyperspectral Reflectance-Derived Relationship Matrices for Genomic Prediction of Grain Yield in Wheat. *G3 Genes|Genomes|Genetics*, **9**, 1231–1247. URL: https://doi.org/10.1534/g3.118.200856.

Lane, H. M. and Murray, S. C. (2021) High throughput can produce better decisions than high accuracy when phenotyping plant populations. *Crop Science*, **61**, 3301–3313.

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

Legarra, A. and Reverter, A. (2018) Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the lr method. *Genetics Selection Evolution*, **50**, 1–18.

Legarra, A., Robert-Granié, C., Manfredi, E. and Elsen, J.-M. (2008) Performance of genomic selection in mice. *Genetics*, **180**, 611–618.

Lenth, R. V. (2024) *emmeans: Estimated Marginal Means, aka Least-Squares Means*. URL: https://CRAN.R-project.org/package=emmeans. R package version 1.10.5.

Li, Y., Yang, X., Tong, L., Wang, L., Xue, L., Luan, Q. and Jiang, J. (2023) Phenomic selection in slash pine multi-temporally using uavmultispectral imagery. *Frontiers in Plant Science*, **14**, 1156430.

Lopez-Cruz, M., Olson, E., Rovere, G., Crossa, J., Dreisigacker, S., Mondal, S., Singh, R. and Campos, G. d. l. (2020) Regularized selection indices for breeding value prediction using hyper-spectral image data. *Scientific Reports*, **10**, 8195.

Lynch, M., Walsh, B. et al. (1998) *Genetics and analysis of quantitative traits*, vol. 1. Sinauer Sunderland, MA.

Meuwissen, T. H. E., Hayes, B. J. and Goddard, M. E. (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, **157**, 1819–1829. URL: https://doi.org/10.1093/genetics/157.4.1819.

Muñoz, P. R., Resende, Jr, M. F. R., Gezan, S. A., Resende, M. D. V., de los Campos, G., Kirst, M., Huber, D. and Peter, G. F. (2014) Unraveling Additive from Nonadditive Effects Using Genomic Relationship Matrices. *Genetics*, **198**, 1759–1768. URL: https://doi.org/10.1534/genetics.114.171322.

Nguyen, G. N. and Norton, S. L. (2020) Genebank phenomics: A strategic approach to enhance value and utilization of crop germplasm. *Plants*, **9**, 817.

Ould Estaghvirou, S. B., Ogutu, J. O., Schulz-Streeck, T., Knaak, C., Ouzunova, M., Gordillo, A. and Piepho, H.-P. (2013) Evaluation of approaches for estimating the accuracy of genomic prediction in plant breeding. *BMC Genomics*, **14**, 860. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3879103/.

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

Pincot, D. D. A., Ledda, M., Feldmann, M. J., Hardigan, M. A., Poorten, T. J., Runcie, D. E., Heffelfinger, C., Dellaporta, S. L., Cole, G. S. and Knapp, S. J. (2021) Social network analysis of the genealogy of strawberry: retracing the wild roots of heirloom and modern cultivars. *G3 Genes|Genomes|Genetics*, **11**, jkab015. URL: https://doi.org/10.1093/g3journal/jkab015.

R Core Team (2024) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Rincent, R., Charpentier, J.-P., Faivre-Rampant, P., Paux, E., Le Gouis, J., Bastien, C. and 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**, 3961–3972.

Robert, P., Auzanneau, J., Goudemand, E., Oury, F.-X., Rolland, B., Heumez, E., Bouchet, S., Le Gouis, J. and Rincent, R. (2022a) Phenomic selection in wheat breeding: identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection. *Theoretical and Applied Genetics*, 1–20.

Robert, P., Brault, C., Rincent, R. and Segura, V. (2022b) Phenomic selection: A new and efficient alternative to genomic selection genomic selection (gs). In *Genomic Prediction of Complex Traits: Methods and Protocols*, 397–420. Springer.

Roth, L., Fossati, D., Krähenbühl, P., Walter, A. and Hund, A. (2023) Image-based phenomic prediction can provide valuable decision support in wheat breeding. *Theoretical and Applied Genetics*, **136**, 162.

Runcie, D. and Cheng, H. (2019) Pitfalls and remedies for cross validation with multi-trait genomic prediction methods. *G3: Genes, Genomes, Genetics*, **9**, 3727–3741.

Runcie, D. E., Qu, J., Cheng, H. and Crawford, L. (2021) Megalmm: mega-scale linear mixed models for genomic predictions with thousands of traits. *Genome biology*, **22**, 1–25.

Sandhu, K. S., Mihalyov, P. D., Lewien, M. J., Pumphrey, M. O. and 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.

Su, G., Christensen, O. F., Ostersen, T., Henryon, M. and Lund, M. S. (2012) Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. *PLOS ONE*, **7**, e45293. URL: `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0045293`. Publisher: Public Library of Science.

Tessema, B. B., Liu, H., Sørensen, A. C., Andersen, J. R. and Jensen, J. (2020) Strategies Using Genomic Selection to Increase Genetic Gain in Breeding Programs for Wheat. *Frontiers in Genetics*, **11**, 578123. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7748061/`.

Van Tassel, D. L., DeHaan, L. R., Diaz-Garcia, L., Hershberger, J., Rubin, M. J., Schlautman, B., Turner, K. and Miller, A. J. (2022) Re-imagining crop domestication in the era of high throughput phenomics. *Current Opinion in Plant Biology*, **65**, 102150.

VanRaden, P. M. (2008) Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, **91**, 4414–4423. URL: `https://www.sciencedirect.com/science/article/pii/S0022030208709901`.

Vitezica, Z. G., Legarra, A., Toro, M. A. and Varona, L. (2017) Orthogonal Estimates of Variances for Additive, Dominance, and Epistatic Effects in Populations. *Genetics*, **206**, 1297–1307.

Vitezica, Z. G., Varona, L., Elsen, J.-M., Misztal, I., Herring, W. and Legarra, A. (2016) Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. *Genetics Selection Evolution*, **48**, 6. URL: `https://doi.org/10.1186/s12711-016-0185-1`.

Vitezica, Z. G., Varona, L. and Legarra, A. (2013) On the Additive and Dominant Variance and Covariance of Individuals Within the Genomic Selection Scope. *Genetics*, **195**, 1223–1230. URL: `https://doi.org/10.1534/genetics.113.155176`.

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

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. URL: `https://ggplot2.tidyverse.org`.

Wilke, C. O. (2020) *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. URL: `https://CRAN.R-project.org/package=cowplot`. R package version 1.1.1.

Winn, Z. J., Amsberry, A. L., Haley, S. D., DeWitt, N. D. and 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**, e20084.

Xiang, T., Christensen, O. F., Vitezica, Z. G. and Legarra, A. (2016) Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. *Genetics Selection Evolution*, **48**, 92. URL: `https://doi.org/10.1186/s12711-016-0271-4`.

Zhong, S., Dekkers, J. C. M., Fernando, R. L. and Jannink, J.-L. (2009) Factors Affecting Accuracy From Genomic Selection in Populations Derived From Multiple Inbred Lines: A Barley Case Study. *Genetics*, **182**, 355–364. URL: `https://doi.org/10.1534/genetics.108.098277`.

Zhu, X., Leiser, W. L., Hahn, V. and Würschum, T. (2021) Phenomic selection is competitive with genomic selection for breeding of complex traits. *The Plant Phenome Journal*, **4**, e20027.

11 | APPENDIX

11.1 | Derivation of correction factor for ρ_y

Let $f(\mathbf{w}) = \hat{y}$ be a prediction model that takes the feature vector $\mathbf{w}$. Assuming the following decomposition of y into additive (a), nonadditive genetic (u) and residual (e) variation (where we combine e and m_y into the same term for simplicity):

$$y = a + u + e. \tag{15}$$

We find:

$$
\begin{aligned}
&\frac{\rho_y - \rho_e\sqrt{1-h^2}}{\sqrt{h^2}} = \\
&\frac{\mathrm{cov}(\hat{y}, y)}{\sqrt{\mathrm{var}(\hat{y})}\sqrt{\mathrm{var}(y)}} - \frac{\mathrm{cov}(\hat{y}, u+e)}{\sqrt{\mathrm{var}(\hat{y})}\sqrt{\mathrm{var}(u+e)}}\sqrt{\frac{\mathrm{var}(u+e)}{\mathrm{var}(y)}}\sqrt{\frac{\mathrm{var}(y)}{\mathrm{var}(a)}} \\
&\qquad = \frac{\mathrm{cov}(\hat{y}, y)}{\sqrt{\mathrm{var}(\hat{y})}\sqrt{\mathrm{var}(a)}} - \frac{\mathrm{cov}(\hat{y}, u+e)}{\sqrt{\mathrm{var}(\hat{y})}\sqrt{\mathrm{var}(a)}} \\
&\qquad = \frac{\mathrm{cov}(\hat{y}, a)}{\sqrt{\mathrm{var}(\hat{y})}\sqrt{\mathrm{var}(a)}} = \rho.
\end{aligned}
\tag{16}
$$

If we can assume that $\rho_e = \mathrm{cor}(\hat{y}, u+e) = 0$, then the quantity on the left simplifies to the familiar $\rho_y/\sqrt{h^2} = \rho$. This is often a reasonable assumption for additive Genomic Prediction models, but is not a reasonable assumption for any Phenomic Prediction model because $\mathrm{cor}(\hat{y}, u) \neq 0$. Furthermore, whenever the features $\mathbf{w}$ are measured on the same individuals as y, $\mathrm{cor}(\hat{y}, e) \neq 0$.

Runcie and Cheng (2019) provided a similar "semi-parametric accuracy estimate" for a prediction model $g(\mathbf{x}, \mathbf{w})$ that takes in both genetic ($\mathbf{x}$) and phenomic ($\mathbf{w}$) features. This estimate also accounts for non-independence among the selection candidates through a genomic relationship matrix $\mathbf{K}$.