

WISER: an innovative and efficient method for correcting population structure in omics-based prediction and selection

Laval Jacquin¹

`laval.jacquin@inrae.fr`

Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV

Walter Guerra

Research Centre Laimburg

Mariusz Lewandowski

The National Institute of Horticultural Research

Andrea Patocchi

Agroscope

Marijn Rymenants

Better3fruit N.V

Charles-Eric Durel¹

Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV

François Laurens¹

Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV

Maria José Aranzana

Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB

Lidia Lozano

IRTA (Institut de Recerca i Tecnologia Agroalimentàries)

Hélène Muranty¹

Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV

Research Article

Keywords:

Posted Date: August 20th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7091927/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

WISER: an innovative and efficient method for correcting population structure in omics-based prediction and selection

Laval Jacquin^{1*}, Walter Guerra², Mariusz Lewandowski³, Andrea Patocchi⁴, Marijn Rymenants⁵, Charles-Eric Durel¹, François Laurens¹, Maria José Aranzana^{6,7}, Lidia Lozano⁷, H       Muranty¹

*Correspondence:

Laval Jacquin

laval.jacquin@inrae.fr, jacquin.julien@gmail.com

H       Muranty

helene.muranty@inrae.fr

Abstract

This work introduces WISER (whitening and successive least squares estimation refinement), an innovative and efficient method designed to enhance phenotype estimation by addressing population structure. WISER outperforms traditional methods such as least squares (LS) means and best linear unbiased prediction (BLUP) in phenotype estimation, offering a more accurate approach for omics-based selection and having the potential to improve association studies. Unlike existing approaches that correct for population structure, WISER provides a generalized framework applicable across diverse experimental setups, species, and omics datasets, including single nucleotide polymorphisms (SNPs), metabolomics, and near-infrared spectroscopy (NIRS) used as phenomic predictors. Central to WISER is the concept of whitening, a statistical transformation that removes correlations between variables and standardizes their variances. Within its framework, WISER extends classical methods that use eigen-information as fixed-effect covariates to correct for population structure, by relaxing their assumptions and implementing a true whitening matrix instead of a pseudo-whitening matrix. This approach corrects fixed effects (e.g., environmental effects) for the genetic covariance structure embedded within the experimental design, thereby minimizing confounding factors between fixed and genetic effects. To support its practical application, a user-friendly R package named `wiser` has been developed. The WISER method has been employed in analyses for genomic prediction and heritability estimation across four species and 33 traits using multiple datasets, including rice, maize, apple, and Scots pine. Results indicate that genomic predictive abilities based on WISER-estimated phenotypes consistently outperform the LS-means and BLUP approaches for phenotype estimation, regardless of the predictive model applied. This underscores WISER's potential to advance omics analyses and related research fields by capturing stronger genetic signals.

1. Introduction

In the realm of modern quantitative genetics, accurate phenotype estimation relies heavily on advanced statistical models that analyze the vast and complex data generated from experimental designs. These models are crucial for effectively capturing the nuances of phenotypic variation and ensuring reliable assessments of traits across diverse environments and conditions. A significant challenge emerges when population structure—which can be simple, with the occurrence of genetically different subgroups within a sample or more complex, with a continuum of relatedness between individuals of the sample—is not adequately addressed. Failure to consider population structure can confound the relationship between genetic variants and phenotypic traits [1–3]. This oversight may ultimately lead to spurious associations and inaccurate estimates of genetic effects [4, 5]. Correcting for population structure is, therefore, a critical step in both genomic selection and

genome-wide association studies (GWAS), as these approaches depend on accurate estimates of the genetic marker effects associated with traits of interest [6–12].

In cases where phenotypic measurements are not repeated for the same individual, population structure can be simply considered using e.g. the classical kinship matrix [13]. In this article, we address the problem of considering the population structure when individuals are repeated in an experimental design.

To address the confounding effects of population structure, numerous methodologies have been developed, in human, animal and plant genomics. Tools such as principal component analysis [10] and linear mixed models [8] have been implemented to correct for unequal genetic relatedness within population in GWAS. Similarly, methods that incorporate eigen-information as fixed-effect covariates—derived from the spectral decomposition of kinship or genomic covariance matrices—within the framework of genomic best linear unbiased prediction (GBLUP) and its derivatives have proven essential in mitigating these confounding factors [6, 7, 11, 14, 15].

Effectively addressing population structure is particularly important in plant breeding populations, where variations in relatedness among individuals, arising from the sampling of genetic resources in a structured population, can act as a significant confounding factor. For instance, in crops such as rice and maize, population structure can be due to geographic origins [16] or selective breeding practices [17]. Despite its potential impact, population structure is often overlooked, especially when estimated to be weak [18–20]. In a two-step approach, phenotype estimation for each individual is typically conducted using a least-squares (LS) means approach [19, 21–29], which does not account for the genetic covariance structure among individuals as the first step. GWAS and/or genomic prediction are then performed as a second step with one phenotype per individual, eventually considering population structure in various ways. However, unbalanced designs—where individuals do not have the same number of replications in each level of each environmental factor—can introduce significant confounding between fixed environmental effects and genetic effects of interest in the first step. This, in turn, can lead to less accurate estimates of the underlying genetic effects.

The best linear unbiased predictor (BLUP) is another frequently used method in the first step, for phenotype estimation per individual, [30–33]. In this approach, principal components derived from PCA on omic data could be included as fixed-effect covariates to account for population structure. However, there is no consensus on the effectiveness of this method for consistently accounting for population structure in the GWAS step, as findings have been mixed. Some studies argue that principal components may not reliably capture population structure [34–38], noting limitations such as subjective selection of the number of components and low variance explained by the first few components in omic data, which may poorly represent the data structure [34, 35]. Additionally, in some association studies, principal component correction for population structure has been shown to fail in controlling the false positive rate [35, 37] and can reduce the detection of true positives [35, 39].

To address this need, we introduce WISER (whitening and successive least squares estimation refinement), a novel approach specifically developed to efficiently model the genetic covariance structure among individuals within experimental designs. This approach takes advantage of the experimental design population structure during the first step, when estimating a phenotype per individual, before genomic prediction or GWAS. It is based on the premise that accounting for population structure in a single-stage manner, rather than a two-stage approach, yields better predictive ability, as described in [40].

At the core of WISER is the concept of whitening—a statistical transformation that removes correlations between variables and standardizes their variances. For instance, in image analysis, whitening techniques are effectively used to preprocess non-random vectors, such as images [41, 42]. Importantly, WISER is not limited to specific species (e.g., animals) or experimental configurations;

rather, it offers a flexible and generalizable framework applicable across a wide range of designs (e.g., alpha lattice, multi-environment trials, multi-environment provenance-progeny trials), species, and omics datasets—including single nucleotide polymorphisms (SNPs), metabolomics, and near-infrared spectroscopy (NIRS) used as phenomic predictors. This versatility makes WISER especially valuable inbreeding, where diverse experimental setups and integrated omics data are commonly employed to improve the accuracy of omics-based selection and marker-trait associations.

A key feature of WISER is its extension of eigen-information as fixed-effect covariates—a widely used approach for correcting for population structure [6, 7, 11, 43]. By whitening fixed-effect variables, WISER corrects for genetic covariance among individuals, reducing confounding between fixed and genetic effects and improving phenotype estimation accuracy. This is especially valuable in plant breeding, where precise omics-based selection and marker-trait associations support the development of high-yield, disease-resistant, and climate-resilient crops [44]. WISER is available as a user-friendly R package, featuring automated tools for optimizing key parameters like whitening matrices—simplifying use and reducing potential for error. Its flexibility also extends to longitudinal datasets in animal breeding, broadening its utility across agricultural sciences [44, 45]. Beyond accessibility, WISER integrates seamlessly into existing omics-based selection and GWAS pipelines. By accounting for population structure across diverse experimental designs, species, and omics data types, it supports more comprehensive analyses—enabling integration of genomics, transcriptomics, and metabolomics to better understand the genetic basis of complex traits [46].

The rest of this article is organized as follows: Sections 2.1 and 2.2 present the linear mixed model and its connection to the WISER statistical framework, explaining how the whitening process within WISER minimizes the confounding effects of population structure on fixed-effect variables, thereby improving phenotype estimation.

Supplementary File 1 demonstrates that previous approaches, particularly the covariance analysis via eigenvectors (EVG) approach introduced by [15] and [6], can be viewed as specific cases of the WISER framework, where WISER extends EVG by relaxing its underlying assumptions and implementing a true whitening matrix instead of a pseudo-whitening matrix. The EVG approach adjusts fixed-effect variables for population structure and has shown superior genomic predictive ability compared to several other methods, as highlighted by [6]. This provides a solid foundation for illustrating how WISER extends and enhances this methodology, as well as other models that use principal components (PC) as fixed effects. Notably, this section shows that, under relaxed assumptions, the ordinary least squares (OLS) estimate of fixed effects in the EVG approach is equal to that of the PC model—demonstrating that WISER not only encompasses but also improves upon models that use PCs as fixed effects by offering a more robust correction for population structure.

Section 2.3 describes four datasets—rice, maize, apple, and Scots pine—used to compare the following phenotype estimation methods: WISER, LS-means and BLUP. To ensure a fair comparison with WISER, BLUP estimates used as phenotypes were also calculated using principal components from genomic PCA as fixed-effect covariates. The section further outlines a simulation approach based on these datasets, designed to evaluate the estimation accuracy of simulated genetic values under two conditions: with and without whitening of fixed-effect variables. It also details the predictive models employed to evaluate each phenotype estimation method across the traits of the four datasets. Additionally, it describes how genomic heritability estimates were calculated for phenotypes derived from the WISER, LS-means, and BLUP methods. Finally, it outlines different computed statistics aimed at explaining the improved genomic predictive ability when using phenotypes derived from WISER.

Section 3 (Results) provides a comparative analysis of the genomic predictive abilities and heritabilities associated to phenotypes derived from WISER, LS-means, and BLUP approaches, along with an exploration of the factors driving the differences in these results. In this section, we demonstrate that

WISER's correction for the genetic covariance structure between individuals in an experimental design enhances genomic predictive abilities and heritability estimates, thereby improving genomic selection and potentially advancing association studies through the capture of stronger genetic signals.

The R package `wiser` can be easily installed from GitHub at <https://github.com/ljacquin/wiser>. To fully align with the FAIR principles, all repositories containing the datasets and R scripts used in this study's analyses are publicly available on GitHub: <https://github.com/ljacquin?tab=repositories>.

2. Materials and methods

2.1 The linear mixed model and the impact of whitening fixed-effect covariates

In the following, we will use the term *genotype* exclusively to mean cultivar, accession or animal. Consider the following linear mixed model:

$$Y = X\beta + Zu + \varepsilon$$

where:

- $Y(n \times 1)$ is the vector of raw phenotypic measurements, with values possibly repeated for each genotype across one or more environments.
- $X(n \times l)$ is the design matrix linking fixed environmental effects to raw phenotypic measurements.
- $\beta(l \times 1)$ is the vector of fixed environmental effects.
- $u(q \times 1)$ is the vector of random genetic values for q genotypes, where $u \sim N_q(0, \sigma_u^2 K)$ and K denotes the genetic covariance matrix between genotypes estimated from omic data.
- $Z(n \times q)$ is the design matrix linking genetic values to raw phenotypic measurements in the experimental design.
- $\varepsilon(n \times 1)$ is the vector of residuals, where $\varepsilon \sim N_n(0, \sigma_\varepsilon^2 I_n)$.

In this model, the estimation of u can be biased if the structure of the experimental design—hereafter referred to as the experimental design population structure or simply population structure—introduces significant confounding between fixed and genetic effects. One common source of such structure is imbalance in the experimental design.

Whitening is a statistical transformation that eliminates correlations between variables and standardizes their variances. In the linear mixed model, a whitening matrix W can be derived from the genetic covariance structure among individuals, captured by $\Sigma_u = \text{Var}(Zu) = Z\text{Var}(u)Z' = \sigma_u^2 ZKZ'$. Applying the matrix W to the fixed-effect design matrix X helps reduce confounding between fixed and random effects. By decorrelating these components, whitening can enable more accurate estimation of the underlying genetic effects. A whitening matrix W satisfies the following condition, known as the whitening property [47–49]:

$$W\Sigma_u W' = I_n$$

where I_n is the identity matrix of size n .

This property ensures that the transformed variables are uncorrelated and have unit variance. Importantly, for any covariance matrix Σ_u , there exists a (generally non-unique) matrix W that fulfills this condition.

2.2 The WISER statistical framework

Using the same definitions as for the components of the linear mixed model, the WISER statistical framework solves the following model to estimate a vector $v = (v_1, \dots, v_q)'$ of genetic values, which are treated as fixed effects:

$$Y = WX\beta + Zv + \varepsilon = \tilde{X}\beta + \xi$$

where $\tilde{X} = WX$ and $\xi = Zv + \varepsilon$. Note that v is modeled as a distinct vector of fixed effects relative to β , and is not treated as a vector of random effects with a genetic covariance matrix K estimated from omic data. This ensures that the estimation of v is not derived as a linear or non-linear function of omic effects (e.g., marker effects), as it typically is in genomic BLUP in the linear case. Instead, v is modeled and estimated separately from these effects. As a result, the WISER-estimated genetic values v remain structurally distinct from omic-derived components and can therefore be used as phenotypes in downstream omic-based selection and association analyses. Within this framework, the estimated genetic values v will be referred to as WISER-estimated phenotypes

The whitening matrix W is derived from the genetic covariance matrix $\Sigma_u = \sigma_u^2 KZ'Z$. It plays a pivotal role as it transforms the fixed environmental variables to minimize the confounding between environmental and genetic effects. There are several methods for computing a whitening matrix W , including zero-phase component correlation analysis (ZCA-cor), principal component correlation analysis (PCA-cor) and Cholesky matrix decomposition (Cholesky) among others [47]. ZCA-cor and PCA-cor whitening matrices are calculated using the spectral decomposition of $R_u = V^{-\frac{1}{2}}\Sigma_u V^{-\frac{1}{2}}$, the genetic correlation matrix associated with Σ_u , where $V = \text{diag}(\Sigma_u)$ is a diagonal matrix with the variances of Σ_u on its diagonal. Since R_u is symmetric and positive semi-definite, it has the following spectral decomposition: $R_u = G\Theta G'$, where G is the orthogonal eigenvector matrix for R_u , while Θ is the diagonal matrix of positive eigenvalues for R_u . The inverse square root matrix of R_u is given by $R_u^{-\frac{1}{2}} = G\Theta^{-\frac{1}{2}}G'$, which satisfies $R_u^{-1} = R_u^{-\frac{1}{2}}R_u^{-\frac{1}{2}}$ since G is an orthogonal matrix. According to [47], the whitening matrices associated to ZCA-cor, PCA-cor and Cholesky are given by:

$$\begin{cases} W^{ZCA-cor} = R_u^{-\frac{1}{2}}V^{-\frac{1}{2}} = G\Theta^{-\frac{1}{2}}G'V^{-\frac{1}{2}} = G\left[\Theta^{-\frac{1}{2}}G'V^{-\frac{1}{2}}\right] \\ W^{PCA-cor} = \Theta^{-\frac{1}{2}}G'V^{-\frac{1}{2}} \\ W^{Cholesky} = L^{-1} \end{cases}$$

where L is a full-rank lower triangular matrix derived from the following Cholesky decomposition $\Sigma_u = \text{Var}(Zu) = LL'$. The PCA-cor whitening procedure can be seen as standardizing variables using $V^{-\frac{1}{2}}$, followed by a rotation using the transposed correlation eigenmatrix G' , and then scaling using the inverted correlation singular values matrix $\Theta^{-\frac{1}{2}}$. ZCA-cor whitening extends this by applying an additional rotation G to revert to the original basis of the standardized variables. While [47] recommends both ZCA-cor and PCA-cor, each whitening method is optimal under distinct criteria. WISER implements these two approaches alongside the Cholesky procedure. Notably, ZCA-cor uniquely ensures that the whitened variables maintain the highest correlation with the original variables, [47] provides a detailed discussion of these criteria and the optimality of each method.

Note that constructing the whitening matrix W depends on $\Sigma_u = \sigma_u^2 ZKZ'$, which, in turn, requires estimating the variance component σ_u^2 . In WISER, this estimation, along with that of σ_ε^2 , is performed using a parallelized approximate Bayesian computation (ABC) algorithm to ensure fast and stable variance component estimation for large datasets associated with complex experimental designs. Additionally, because Σ_u may approach singularity in practical applications, and thus prevent the computation of the inverse square root matrix of R_u and Θ , WISER incorporates established shrinkage and regularization procedures to ensure positive-definiteness [50–55]. Further details on the ABC algorithm and regularization procedures are provided in Supplementary File 2.

To automate the choice of the method used to construct the whitening matrix and of regularization parameters, WISER also includes a dedicated optimization function. This function performs a grid search across predefined combinations of whitening methods, regularization parameters, and various prediction methods, such as genomic best linear unbiased prediction (GBLUP), least absolute shrinkage and selection operator (LASSO), support vector regression (SVR), reproducing kernel Hilbert space (RKHS) regression, and random forest. Using k-fold cross-validation (CV), the function evaluates the performance of each combination and identifies the optimal whitening method and regularization parameter that minimize the mean squared error based on a majority vote across the k-fold CV results. By default, the optimization is performed on a subset of the dataset, limited at a maximum of 5,000 raw phenotypic measurements to ensure computational efficiency.

Once the whitening matrix is computed, estimates for β and v are obtained using the following successive OLS procedure:

$$\begin{cases} \hat{\beta} = \underset{\beta \in \mathbb{R}^l}{\text{argmin}} ||Y - \tilde{X}\beta||_2^2 = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y = (X'W'WX)^{-1}X'W'Y \\ \hat{v} = \underset{v \in \mathbb{R}^q}{\text{argmin}} ||\hat{\xi} - Zv||_2^2 = (Z'Z)^{-1}Z'\hat{\xi} \end{cases}$$

where $\tilde{X} = WX$, $\hat{\xi} = Y - \tilde{X}\hat{\beta}$ and $||\cdot||_2^2$ represents the squared l_2 norm. Note that $\hat{\xi}$ corresponds to the part of Y that is not explained by the transformed fixed-effect variables, and is orthogonal to these variables. As with any linear model, this orthogonality arises because $\hat{\xi}$ is estimated by projecting Y onto the space orthogonal to that spanned by the whitened fixed-effect variables, ensuring orthogonality by design:

$$\tilde{X}'\hat{\xi} = \tilde{X}'(Y - \tilde{X}\hat{\beta}) = \tilde{X}'(Y - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y) = \tilde{X}'Y - \tilde{X}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y = \tilde{X}'Y - \tilde{X}'Y = 0$$

Thus, the vector $\hat{\xi}$ can be used to estimate the vector v . It is also important to note that $\hat{\beta}$ is computed using W , which is derived from Σ_u and thus depends on an estimate of σ_u^2 . Furthermore, when no repeated phenotypic measurements per genotype are available, the genetic covariance matrix Σ_u associated with individuals in the experimental design becomes equal to the genetic covariance matrix K associated with genotypes. In this case, the population structure is represented by $\Sigma_u = K$. To mitigate issues of numerical singularity of $\tilde{X}'\tilde{X}$ or $Z'Z$ in certain practical applications, WISER also incorporates the Moore-Penrose generalized pseudo-inverse method for estimating β and v .

2.3 Analyzed datasets, compared phenotype estimation methods, predictive models, and computed statistics

2.3.1. Analyzed datasets

Four datasets from experimental designs involving rice [18], maize [20], apple [56], and Scots pine [57] were used to compare genomic predictive ability (PA) and heritability estimation across 33 traits. In these datasets, the omic data were SNP data. The rice dataset contained 4,308 raw phenotypic measurements and 9,928 SNPs associated with 334 genotypes. The 334 rice genotypes originated from a synthetic population developed from 35 tropical japonica (*Oryza sativa* L.) families, following three recombination cycles and advancement to $S_{0.2}$ and $S_{0.3}$ generations [18]. The maize dataset included 18,970 raw phenotypic measurements and 41,722 SNPs associated with 246 genotypes, corresponding to a diversity panel of dent maize hybrids [20]. The apple reference population (REFPOP) dataset, after preprocessing to remove phenotypic outliers, consisted of 42,522 raw phenotypic measurements and 303,240 SNPs associated with 534 genotypes. These genotypes included both a set of accessions and a set of progenies. The accessions derived from 10 geographical origins, while the progenies belong to 27 distinct families. Among the SNPs, only 50,000 were used, selected through uniform sampling to reduce computation time and resource usage, as [19] demonstrated that a PA plateau is typically reached with as few as 10,000 SNPs for this dataset. Additionally, the apple dataset included two traits—scab and powdery mildew—not covered in [56]. Finally, the Scots pine dataset included 1,248 raw phenotypic measurements and 20,795 SNPs for 208 genotypes from 40 families across 5 native *P. sylvestris* populations, evaluated at two sites [57].

The traits predicted and analyzed for each species were as follows:

- Rice: days to flowering (FL), plant height (PH), grain yield (YLD), and grain zinc concentration (ZN).
- Maize: anthesis, anthesis-silking interval, ear height, grain number, grain weight, grain yield, plant height, silking, and tassel height.
- Apple: color over, flowering start, flowering end, full flowering, flowering intensity, fruit number, single fruit weight, fruit weight, harvest date, powdery mildew, russet frequency, scab, trunk diameter, and trunk increment.
- Scots pine: height (H), annual height increment (I), duration of budburst (D)—the time taken to progress from stage 4 (scales open along the length of the shoot, no needles) to stage 6 (green-tipped needles visible), and time taken to reach individual budburst stages: stage 4 (T4), stage 5 (T5), and stage 6 (T6).

For each dataset, the environmental variable was defined as a combination of available factors such as site (or country), year, management type, block, and genotype generation. In the rice dataset, it was defined by site, year, block, and genotype generation, resulting in 48 unique environments. For maize, it was defined by site, year, management type (rainfed or irrigated), and block, yielding 1,082 levels. In the apple dataset, the environmental variable was defined as the combination of country, year, and

management type (standard irrigation and pesticide use, standard irrigation with reduced pesticide use, or reduced irrigation with standard pesticide use), with 75 distinct levels. For pine, the environmental variable was based on site, year, and block, also resulting in 48 levels.

For each dataset, population structure was visually assessed using three complementary approaches applied to the SNP data: a heatmap of the genomic covariance matrix, principal component analysis (PCA), and uniform manifold approximation and projection (UMAP) [58]. These analyses were implemented using the R packages `pheatmap` [59], `mixOmics` [60] and `umap` [61].

Raw phenotypes for a generic trait were simulated for each dataset using fixed effects, random genetic effects, and random genotype-by-environment interaction effects. The trait heritability was set to a moderate value of 0.5, and the total phenotypic variance was fixed at 1. The simulation procedure was as follows:

$$Y = X\beta + Zu + Z_{ge}u_{ge} + \varepsilon$$

where:

- $Y(n \times 1)$ is the vector of simulated raw phenotypic measurements, with values possibly repeated for each genotype across one or more environments.
- $X(n \times l)$ is the design matrix linking fixed effects, the overall mean and the $l - 1$ environmental effects, to raw phenotypic measurements.
- $\beta = (\mu, \beta_1, \dots, \beta_j, \dots, \beta_{l-1})'$ is the vector of fixed effects where $\mu \sim U(0,1)$ and $\beta_j \sim N(0,3)$.
- $u(q \times 1)$ is the vector of random genetic values for q genotypes, where $u \sim N_q(0, \sigma_u^2 K)$ and K denotes the genetic covariance matrix between genotypes estimated from omic data.
- $Z(n \times q)$ is the design matrix linking genetic values to raw simulated phenotypic values in the experimental design.
- $u_{ge}(q(l - 1) \times 1)$ is the vector of random genotype by environment interaction effects, where $u_{ge} \sim N_q(0, \sigma_{ge}^2 I)$ and $\sigma_{ge}^2 = 0.5 \times \sigma_u^2$.
- $Z_{ge}(n \times q(l - 1))$ is the design matrix linking random interaction effects to raw simulated phenotypic values in the experimental design.
- $\varepsilon(n \times 1)$ is the vector of residuals, where $\varepsilon \sim N_n(0, \sigma_\varepsilon^2 I_n)$.

This simulation approach was used to assess the accuracy of estimating the vector u under two conditions: with and without whitening of the fixed-effect variables, in order to evaluate the impact of this transformation. To estimate the distribution of estimation accuracy (EA) under each condition, raw phenotypes were simulated 100 times. EA was defined as the Pearson correlation between the simulated genetic values and their estimates, obtained with and without whitening of the fixed-effect variables.

2.3.2. Compared phenotype estimation methods

Four phenotype estimation approaches were compared: LS-means, two BLUP approaches, and WISER.

To account for spatial variation prior to phenotype estimation using the LS-means and BLUP approaches, a spatial heterogeneity correction was applied to each combination of trait and

environment, resulting in spatially adjusted phenotypes. This correction was implemented when spatial location information was available, which was the case for the maize, apple, and pine datasets, where row and column, row and position, and GPS coordinates were respectively available. For the rice dataset, spatial location data was unavailable thus no correction could be applied. The spatial heterogeneity correction was performed using the spatial analysis of field trials with splines method [62], following the same approach as described in [19]. The implementation used to account for spatial variation in the WISER framework will be detailed below. The spatial independence of the resulting residuals was then visually assessed using residual spatial plots and variograms. The spatial plots showed no distinct color clusters, and the variograms were generally flat, both indicating minimal remaining spatial structure.

For the LS-means approach, the following multiple linear regression model was initially fitted:

$$y_{ijk} = \mu + g_i + l_j + \varepsilon_{ijk}$$

where y_{ijk} was the k^{th} spatially adjusted raw phenotype associated to genotype i in environment j , μ was the fixed overall mean, g_i was the fixed effect of genotype i ($1 \leq i \leq q$), l_j was the fixed effect of the j^{th} environment, and ε_{ijk} was the error term. The error term ε_{ijk} was assumed to follow a normal distribution, $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$, with its associated vector ε considered to follow a multivariate normal distribution with covariance matrix $\sigma_\varepsilon^2 I_n$. For the rice dataset, y_{ijk} corresponded to raw phenotypic measurements. Phenotypic LS-means for the genotypes were then computed across all environments using the R package `lsmeans` [63] to estimate a single phenotypic value per genotype. The Moore-Penrose generalized pseudo-inverse was used during the multiple linear regression procedure to address issues related to aliased variables (i.e., perfect multicollinearity) for certain traits. This adjustment was particularly necessary for traits D, H, and I in the pine dataset.

For the BLUP approaches, the following linear mixed model, was fitted using the R package `lme4` [64]:

$$y_{ijk} = \mu + \sum_{s=1}^S c_{is} \gamma_s + r_i + l_j + \varepsilon_{ijk}$$

In this model, y_{ijk} , μ , l_j and ε_{ijk} were defined as in the LS-means approach. Here, r_i represents the random effect of genotype i ($1 \leq i \leq q$), with $r_i \sim N(0, \sigma_r^2)$, and its associated vector r was assumed to follow a multivariate normal distribution with an isotropic covariance matrix, $\sigma_r^2 I_q$. Additionally, to account for population structure, principal component (PC) coordinates for each genotype i , denoted c_{is} ($1 \leq s \leq S$) were introduced in the model and treated as fixed effects. These PC coordinates were computed using genomic data from each dataset with the R package `mixOmics` [60]. In this context, γ_s represented the effect associated with the s^{th} PC. Population structure is generally regarded as a low-dimensional process embedded within a high-dimensional space [9, 65]. Consequently, a relatively small number of principal components, rarely exceeding 10, is assumed to effectively capture the underlying genetic structure of the population [10]. The number S of principal components (PCs) selected for computing the BLUP values of r was chosen to minimize the Akaike information criterion (AIC), obtained via maximum likelihood estimation, within the range of values tested (1 to 10), as AIC is widely regarded as one of the most commonly used criteria for model comparison [66]. The BLUP

was then computed using restricted maximum likelihood (REML) to obtain unbiased variance components estimates for use in the BLUP expression. The computed BLUP values using this linear mixed model were named BLUP-PCA values. This model was also fitted without the component $\sum_{s=1}^S c_{is} \gamma_s$, using the REML procedure, to obtain BLUP values which do not account for population structure.

In the `wiser` framework, the overall mean (included by default) and the environmental variable were modeled as fixed-effect factors for the rice dataset. For the apple dataset, the overall mean, the environment (comprising country, year, and management), and rows and positions for spatial adjustment were modeled as fixed-effect factors, with rows and positions treated as environment-specific factors. For the maize and pine datasets, the overall mean and the environmental variable were included as a fixed-effect factor, while row and column numbers (for maize) and GPS coordinates (for pine) were treated as fixed-effect quantitative variables specific to each environment, to account for spatial variation. This approach minimized the number of fixed-effect variables, thereby reducing the risk of inestimability. For example, the maize dataset had 1,061 levels for the environment variable related to anthesis, resulting in 2,122 fixed-effect quantitative variables when accounting for row and column within each environment. Thus, the number of fixed-effect variables would have been considerably larger if row and column numbers had been modeled as factors within each environment. Using rows and positions for apple, row and column number for maize, and GPS coordinates for pine, directly in the model enabled to skip the pre-processing step of spatial heterogeneity correction applied to each combination of trait and environment used with the LS-means and BLUP approaches. The three whitening procedures implemented in `wiser` are based on the R package `whitening` [67]. For each trait, the whitening procedure and regularization parameter were automatically selected using the `optimize_whitening_and_regularization` function from the `wiser` package, which was specifically designed for automated selection. As described in Section 2.2, this function performs a grid search over a subset of the dataset to identify the optimal combination of whitening procedure and regularization parameter that minimizes the mean square error.

The WISER approach was also applied to the simulated datasets, both with and without its default whitening matrix (ZCA-cor), to evaluate the estimation accuracy of the simulated vector u of genetic values. Notably, using WISER without a whitening matrix in this context is equivalent to fitting a classical linear model in which both environmental and genetic effects are treated as fixed. In this case, the resulting genetic effect estimates correspond to classical best linear unbiased estimates (BLUE) and will be referred to as BLUE throughout.

2.3.3. Compared predictive models

Five predictive models were employed for genomic prediction: random forest (RF), support vector regression (SVR), genomic best linear unbiased prediction (GBLUP), reproducing kernel Hilbert space (RKHS) regression, and least absolute shrinkage and selection operator (LASSO). These models were selected to represent diverse genetic architectures and to evaluate whether the genomic predictive ability (PA) rankings across the different phenotype estimation methods and traits remained consistent, regardless of the prediction approach. The RF, SVR, and LASSO models were fitted using the R packages `ranger` [68], `kernlab` [69], and `glmnet` [70], respectively. GBLUP and RKHS regression were fitted using the R package `KRMM` [71]. For RF, the number of randomly selected candidate variables (`mtry`) for testing splits at each tree node was set to one-third of the total number of SNPs, a common choice for regression tasks, and the number of trees was fixed at 1,000. For SVR, LASSO, GBLUP, and RKHS regression, hyperparameters were estimated or selected following the approach described in [71]. Specifically, for SVR, a Gaussian kernel was applied using the `ksvm` function, and its bandwidth parameter was estimated automatically using the heuristic in the `sigest` function. The regularization parameter C for SVR was computed as $C = \max(|\bar{Y} + 3\sigma_Y|, |\bar{Y} - 3\sigma_Y|)$,

where $\bar{Y}$ and σ_Y are the mean and standard deviation of phenotypes respectively, as recommended by [72]. The ε -insensitive loss parameter in SVR was set to its default value of 0.1, as tested values ranging from 0.01 to 0.5 produced similar results, which were comparable to those of other prediction models. For LASSO, the regularization parameter was determined using K-fold cross-validation implemented in the `cv.glmnet` function, with the default value of `nfolds` set to 10. For GBLUP, the `krmm` function was applied with a linear kernel, using the default initial guess values for the variance component estimates required by the EM-REML algorithm. The same settings were used for RKHS regression, except that a Gaussian kernel was applied. The Gaussian kernel's bandwidth parameter referred to as the rate of decay in `krmm` was set to its default value of 0.1, which provided comparable results to the other prediction models.

A 5-fold cross-validation (CV) scheme with 20 random shufflings of the datasets was employed to estimate the distributions of predictive ability and genomic heritability (h^2) across various prediction models and phenotype estimation methods for all traits and species. For each shuffle, a new seed was generated by multiplying the shuffle number by 100, thereby introducing substantial variation in the randomized genotype distribution across shuffles. This approach generated a distribution of 100 estimated PA values for each combination of species, trait, prediction model, and phenotype estimation method. Similarly, a distribution of 100 estimated h^2 values was generated with the GBLUP model using this approach. For each phenotype estimation method, PA was evaluated for each validation fold by computing the Pearson correlation between the genomically predicted phenotypes and those estimated using the phenotype estimation method. In contrast, h^2 was computed from the estimated variance components of the following GBLUP model, which was fitted on the training folds:

$$y = 1_q \mu + u + \varepsilon$$

In this model, $y = (y_1, \dots, y_q)'$ represents the vector of q phenotypes derived from either WISER, LS-means, or BLUP approaches. The vector $1_q = (1, \dots, 1)'$ consists of q ones, and μ corresponds to the fixed overall mean. The vector $u = (u_1, \dots, u_q)'$ represents the genetic values associated with the q genotypes, where $u \sim N_q(0, \sigma_u^2 K)$, and $K = MM'$ is the genomic covariance matrix. Here, $M = [m_i^j]_{1 \leq i \leq q}^{1 \leq j \leq p}$ is the centered SNP marker matrix of size $q \times p$, where $m_i^j = w_i^j - \mathbb{E}[w^j]$, with w_i^j coded as 0, 1, or 2. The error vector $\varepsilon = (\varepsilon_1, \dots, \varepsilon_q)'$ follows a normal distribution, $\varepsilon \sim N_q(0, \sigma_\varepsilon^2 I_q)$. From this model, the genomic heritability h^2 was defined as:

$$h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}$$

2.3.4. Computed statistics

From the generated distributions of predictive ability (PA) and genomic heritability (h^2), median values and deviations, computed for each specific phenotype estimation method, were calculated to facilitate graphical representations and subsequent analyses. To explore the relationships between PA and h^2 , deviations in the PA medians were compared to those in the h^2 medians. Additionally, the deviations in the PA medians were compared to the Shannon entropy of genotype frequencies across all trait-species combinations. For each trait-species combination, the Shannon entropy of genotype frequencies was defined as:

466

467

$$H(Genotype) = - \sum_{i=1}^q p_i \log_2(p_i)$$

468

469

470

471

472

473

474

where p_i is the relative frequency of observations of the i^{th} genotype for the trait in the dataset, and q is the total number of genotypes. This comparison was intended to assess the impact of imbalanced genotype frequencies across all trait-species combinations on the deviations observed in the PA medians. A similar comparison was made between these deviations and the Shannon joint entropies of genotype and site, as well as genotype and environment. For each trait-species combination, these joint entropies were computed as follows:

475

476

$$H(Genotype, V) = - \sum_{i=1}^q \sum_{j=1}^x p_{ij} \log_2(p_{ij})$$

477

478

479

480

481

where V is either the site or environment variable, p_{ij} represents the joint relative frequency of observations of genotype i and site (or environment) j in the dataset, and x is the number of sites (or environments). These joint entropies were used to quantify the confounding degree between genotypes and their respective sites or environments.

482

483

484

485

To assess the degree of association between genotype and site or environment, and identify potential multicollinearity among their dummy variables, Cramér's V statistic was computed. It was particularly reported when PA median values were low for a trait-species combination, helping to identify possible causes for these low values. This statistic was computed using the R package `vcd` [73].

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

To explore the relationship between genotype population structure and deviations in PA medians, dimensionality reduction and clustering techniques were employed. UMAP was used to generate five-dimensional representations of the genomic data for each species. The choice of five dimensions was a trade-off between preserving enough genetic variability for effective clustering and mitigating the "curse of dimensionality", which can reduce the accuracy of distance-based methods in higher-dimensional spaces. The UMAP approach was implemented using the R package `umap` [61]. The resulting five-dimensional representations were clustered using the K -means algorithm for each trait-species combination, with the optimal number of clusters determined as the value of K that maximized the Calinski-Harabasz (CH) index. Note that the clustering was performed separately for each trait-species combination using only genotypes with available phenotypic measurements, as such data were not consistently available across all genotypes for each trait. Moreover, PCA was not used for dimensionality reduction prior to K -means, as it failed to distinguish known accessions and 27 progeny families in the apple REFPOP dataset (see Supplementary Figures 1 to 4). In contrast, UMAP, without any labeling information, was able to effectively separate the accessions and most of the families, making it better suited for use prior to the clustering algorithm.

501

The CH index was evaluated over a range of K values from 2 to 30 and is defined as:

502

503

$$CH = \frac{B_K}{K-1} \bigg/ \frac{W_K}{n-K},$$

where

- $B_K = \sum_{k=1}^K |C_k| \|\mu_k - \mu\|_2^2$ is the between-cluster sum of squares (BCSS) for the SNP marker data. Here, $\mu_k = (\mathbb{E}_{C_k}[m^1], \dots, \mathbb{E}_{C_k}[m^p])'$ represents the centroid vector for cluster C_k , $\mu = (\mathbb{E}[m^1], \dots, \mathbb{E}[m^p])'$ is the mean vector for the entire dataset, and $|C_k|$ is the number of data points (i.e., vectors) in cluster C_k .
- $W_K = \sum_{k=1}^K \sum_{m \in C_k} \|m - \mu_k\|_2^2$ is the within-cluster sum of squares (WCSS), quantifying the compactness of data points around their respective cluster centroids. Here, $m = (m^1, \dots, m^p)'$ represents an SNP vector with values coded as 0, 1 or 2.

The CH index generalizes Fisher's test statistic from a one-way ANOVA with K levels to the multidimensional setting. While this index is not a test statistic and lacks a formal statistical distribution, it serves as a robust metric to balance the compactness of clusters with their separation, providing an effective measure for determining the optimal clustering structure. Note that K -means clustering aims to find the optimal partitioning of data into K clusters that minimizes W_K . Therefore, the CH index is a natural choice for selecting K , as both the CH index and K -means are related to the total sum of squares (TSS) decomposition, which is expressed as:

$$\sum_{k=1}^K \sum_{m \in C_k} \|m - \mu\|_2^2 = \sum_{k=1}^K |C_k| \|\mu_k - \mu\|_2^2 + \sum_{k=1}^K \sum_{m \in C_k} \|m - \mu_k\|_2^2 = B_K + W_K$$

The CH index was computed using the R package `fpc` [74]. Finally, to quantify genetic differentiation between genotypes, the average fixation index F_{ST} was computed using genomic data associated with the K -estimated clusters for each trait-species combination. Notably, for each trait-species combination, the genomic data used accounted for genotype incidence relative to the available phenotypic measures in the experimental design. To improve computational efficiency and minimize resource usage, a stratified sampling approach was employed to maintain the data structure, while limiting the genomic dataset to 5,000 rows per trait-species combination. The F_{ST} values were calculated using the R packages `adegenet` [75] and `hierfstat` [76].

3. Results

3.1. Population structure at the genotype level

Supplementary Figures 5 to 16 present a visual assessment of population structure based on the complementary use of heatmaps, PCA and UMAP applied to the SNP data. The apple and pine datasets revealed pronounced population structuring, characterized by distinct clusters corresponding to family groups. Similarly, the maize dataset displayed some clustering patterns despite the absence of prior family information. In contrast, the rice dataset showed minimal clustering, consistent with the findings of [18], who reported no detectable structure in this dataset.

3.2. Genomic predictive ability (PA) and heritability (h^2) estimates

Tables 1 and 2 summarize the median PA and h^2 values for phenotypes estimated using WISER, LS-means, and BLUP approaches across all species and traits, for the GBLUP model. Supplementary File 3

presents median PA values for additional prediction models (random forest, SVR, RKHS, and LASSO), which exhibit, for WISER, LS-means, and BLUP approaches, consistent rankings, similar deviations in median PA values and comparable PA ranges across most models. Therefore, only the GBLUP results are presented here for conciseness. Supplementary Figures 17 to 20 also present a graphical summary of the GBLUP PA values, organized by species, for phenotypes estimated using WISER, LS-means, and BLUP approaches.

Table 1 highlights that WISER consistently achieved the highest GBLUP median PA values across all species and traits. LS-means and BLUP showed similar median PA for most species and traits, except for the Pine dataset, and generally outperformed BLUP-PCA, except in a few isolated cases. BLUP-PCA consistently showed the lowest median PA values in most cases. It is worth noting that, the GBLUP median PA values, for LS-means and BLUP in Table 1, are consistent with the genomic PA values for the corresponding traits reported in [18], [20], [56], and [57], despite differences in CV schemes and associated scenarios in these studies. This emphasizes the improvement that WISER offers over currently used phenotype estimation methods.

Table 1 – GBLUP median PA values for phenotype estimation methods (WISER, LS-means, and BLUP approaches) across all species and traits, with interquartile ranges (IQR) in parentheses. Values are based on 100 cross-validation (CV) scenarios.

Dataset	Trait	GBLUP median PA values			
		WISER	LS-means	BLUP	BLUP-PCA
Rice dataset 334 genotypes 9,929 SNPs	days to flowering (FL)	0.74 (0.05)	0.41 (0.12)	0.41 (0.11)	0.21 (0.11)
	plant height (PH)	0.73 (0.08)	0.51 (0.09)	0.51 (0.09)	0.42 (0.09)
	grain yield (YLD)	0.62 (0.08)	0.49 (0.11)	0.49 (0.11)	0.42 (0.10)
	grain zinc concentration (ZN)	0.69 (0.05)	0.36 (0.16)	0.36 (0.16)	0.26 (0.19)
Maize dataset 246 genotypes 41,723 SNPs	Anthesis	0.74 (0.07)	0.69 (0.07)	0.69 (0.07)	0.49 (0.11)
	anthesis-silking interval	0.66 (0.07)	0.37 (0.18)	0.37 (0.18)	0.11 (0.20)
	ear height	0.67 (0.07)	0.24 (0.15)	0.24 (0.13)	0.11 (0.19)
	grain number	0.68 (0.09)	0.53 (0.14)	0.52 (0.14)	0.34 (0.16)
	grain weight	0.61 (0.10)	0.54 (0.13)	0.54 (0.13)	0.15 (0.18)
	grain yield	0.77 (0.08)	0.66 (0.10)	0.66 (0.11)	0.32 (0.16)
	plant height	0.67 (0.08)	0.22 (0.15)	0.22 (0.15)	0.18 (0.12)
	Silking	0.70 (0.08)	0.67 (0.06)	0.66 (0.07)	0.40 (0.13)
	tassel height	0.67 (0.07)	0.30 (0.14)	0.30 (0.14)	0.16 (0.16)
Apple dataset 534 genotypes 50,000 SNPs	color over	0.78 (0.05)	0.70 (0.05)	0.70 (0.05)	0.48 (0.07)
	flowering begin	0.92 (0.02)	0.64 (0.09)	0.64 (0.09)	0.55 (0.09)
	flowering end	0.86 (0.05)	0.44 (0.10)	0.44 (0.10)	0.36 (0.10)
	flowering full	0.80 (0.04)	0.40 (0.10)	0.42 (0.09)	0.33 (0.09)
	flowering intensity	0.66 (0.09)	0.52 (0.09)	0.52 (0.09)	0.32 (0.11)
	fruit number	0.78 (0.04)	0.67 (0.06)	0.67 (0.06)	0.34 (0.10)
	fruit weight	0.78 (0.04)	0.65 (0.06)	0.65 (0.05)	0.30 (0.10)
	fruit weight single	0.73 (0.05)	0.58 (0.07)	0.58 (0.07)	0.52 (0.09)
	harvest date	0.90 (0.03)	0.76 (0.06)	0.76 (0.06)	0.59 (0.07)
	powdery mildew	0.56 (0.08)	0.40 (0.10)	0.40 (0.10)	0.28 (0.10)
	russet frequency	0.60 (0.07)	0.46 (0.09)	0.46 (0.09)	0.36 (0.10)
	Scab	0.56 (0.09)	0.33 (0.10)	0.34 (0.09)	0.21 (0.12)
	trunk diameter	0.66 (0.06)	0.50 (0.06)	0.50 (0.06)	0.28 (0.09)
	trunk increment	0.51 (0.07)	0.49 (0.06)	0.48 (0.06)	0.25 (0.12)
Pine dataset 208 genotypes 20,796 SNPs	D	0.66 (0.10)	0.08 (0.19)	0.14 (0.17)	0.11 (0.19)
	H	0.72 (0.08)	0.18 (0.17)	0.31 (0.16)	0.23 (0.18)
	I	0.35 (0.25)	0.10 (0.21)	0.31 (0.20)	0.20 (0.20)
	T4	0.39 (0.22)	0.24 (0.18)	0.32 (0.20)	0.30 (0.23)
	T5	0.26 (0.25)	0.14 (0.22)	0.26 (0.24)	0.12 (0.24)

For the pine dataset, GBLUP median PA values were notably lower, especially when using phenotypes estimated via the LS-means and BLUP-PCA. For example, traits D, H, and I exhibited a strong association between genotypes and environments, as indicated by high Cramér's V statistics (0.43 for D, 1.00 for H, and 0.41 for I). This association resulted in confounding due to the interrelationship between genotypes and environments, leading to perfect multicollinearity among their dummy variables. For example, the variance inflation factor (VIF) for the linear regression used by LS-means was uncomputable for these traits, and similarly, the restricted maximum likelihood (REML) algorithm in BLUP-PCA computation failed to converge to identifiable models. In contrast, WISER mitigated these issues through its whitening process, which reduced the effect of confounding between genotypes and environment. This resulted in more stable and reliable phenotypic estimates. In terms of interquartile ranges (IQR) for GBLUP PA values, WISER typically produced smaller ranges compared to LS-means and BLUP, with a few exceptions, most notably in the pine dataset. Across all species and traits, the average IQR values were 0.09 for WISER, 0.12 for LS-means, 0.12 for BLUP, and 0.14 for BLUP-PCA, highlighting WISER's greater robustness in most scenarios.

To evaluate the potential advantages of WISER over existing phenotype estimation methods—and given the superior performance of LS-means and BLUP compared to BLUP-PCA—only the differences between the GBLUP median PA values of WISER and those of LS-means and BLUP (referred to as the median PA gain) were calculated for the 33 traits. This yielded 66 median PA gain values. Similarly, differences in GBLUP median h^2 values between WISER and LS-means and BLUP were computed, resulting in 66 median h^2 deviations. The average median PA gain and median h^2 deviation across all species and traits were 0.21 and 0.17, respectively, highlighting WISER's improved accuracy in phenotype estimation.

Table 2 demonstrates similar trends for GBLUP median h^2 values, with WISER consistently outperforming LS-means and BLUP approaches across all species and traits, except for two isolated cases: grain weight and silking. However, it is important to note that a higher h^2 estimate does not always indicate a better estimation, as it may result from overfitting of the GBLUP model, which can lead to an overestimation of the trait's heritability. For instance, while the h^2 values estimated using LS-means and BLUP phenotypes were higher than those obtained with WISER for grain weight and silking, the corresponding GBLUP median PA values based on LS-means phenotypes were lower than those derived from WISER phenotypes.

For traits D and I in the pine dataset, both LS-means and BLUP-PCA exhibited exceptionally low h^2 values close to zero due to numerical instability caused by strong confounding between genotypes and environments. Specifically, the estimated values were 0.0035 (0.17) and 0.0042 (0.073) for D, and 0.00092 (0.0086) and 0.0011 (0.0027) for I. When rounded to two decimal places in Table 2, these values appear as zero, but they instead reflect challenges in model convergence rather than a true lack of heritability. For trait D, BLUP also exhibited an exceptionally low h^2 value due to the same challenges. Similarly, convergence difficulties were observed with BLUP-PCA phenotypes for several other traits across the rice, maize, and pine datasets, where the estimated h^2 was also exceptionally low. In contrast, these issues with both LS-means and BLUP approaches were not observed in the REFPOP dataset, which represented a larger population of 534 genotypes compared to the other datasets: 334 in rice, 246 in maize, and 208 in pine. This larger sample size likely contributed to improved model convergence and more reliable estimation.

Table 2 – GBLUP median h^2 values for phenotype estimation methods (WISER, LS-means, and BLUP approaches) across all species and traits, with interquartile ranges (IQR) in parentheses. Values are based on 100 cross-validation (CV) scenarios.

Dataset	Trait	GBLUP median h^2 values			
		WISER	LS-means	BLUP	BLUP-PCA
Rice dataset 334 genotypes 9,929 SNPs	days to flowering (FL)	0.87 (0.02)	0.53 (0.06)	0.53 (0.06)	0.00 (0.00)
	plant height (PH)	0.86 (0.03)	0.76 (0.03)	0.76 (0.03)	0.70 (0.05)
	grain yield (YLD)	0.82 (0.04)	0.73 (0.05)	0.73 (0.05)	0.68 (0.06)
	grain zinc concentration (ZN)	0.84 (0.03)	0.55 (0.10)	0.56 (0.10)	0.42 (0.17)
Maize dataset 246 genotypes 41,723 SNPs	Anthesis	0.96 (0.02)	0.93 (0.03)	0.93 (0.03)	0.88 (0.04)
	anthesis-silking interval	0.82 (0.05)	0.66 (0.11)	0.66 (0.11)	0.00 (0.06)
	ear height	0.69 (0.05)	0.60 (0.14)	0.55 (0.15)	0.00 (0.01)
	grain number	0.96 (0.02)	0.86 (0.06)	0.85 (0.05)	0.78 (0.09)
	grain weight	0.65 (0.06)	0.83 (0.06)	0.83 (0.05)	0.00 (0.00)
	grain yield	0.97 (0.02)	0.93 (0.04)	0.92 (0.05)	0.80 (0.10)
	plant height	0.68 (0.05)	0.34 (0.15)	0.33 (0.15)	0.00 (0.01)
	Silking	0.71 (0.04)	0.92 (0.04)	0.92 (0.03)	0.85 (0.06)
	tassel height	0.68 (0.04)	0.54 (0.12)	0.54 (0.12)	0.00 (0.00)
Apple dataset 534 genotypes 50,000 SNPs	color over	0.90 (0.03)	0.84 (0.03)	0.84 (0.03)	0.76 (0.05)
	flowering begin	0.99 (0.00)	0.94 (0.02)	0.94 (0.02)	0.93 (0.03)
	flowering end	1.00 (0.00)	0.73 (0.03)	0.73 (0.03)	0.69 (0.04)
	flowering full	0.94 (0.02)	0.67 (0.06)	0.69 (0.05)	0.64 (0.05)
	flowering intensity	0.77 (0.03)	0.66 (0.04)	0.67 (0.04)	0.56 (0.07)
	fruit number	0.76 (0.03)	0.65 (0.04)	0.65 (0.04)	0.42 (0.08)
	fruit weight	0.75 (0.03)	0.59 (0.03)	0.59 (0.04)	0.35 (0.08)
	fruit weight single	0.93 (0.02)	0.81 (0.04)	0.81 (0.04)	0.79 (0.04)
	harvest date	0.97 (0.01)	0.92 (0.02)	0.92 (0.02)	0.88 (0.02)
	powdery mildew	0.74 (0.04)	0.51 (0.05)	0.52 (0.05)	0.36 (0.09)
	russet frequency	0.81 (0.04)	0.67 (0.06)	0.67 (0.05)	0.54 (0.10)
	Scab	0.68 (0.04)	0.44 (0.05)	0.45 (0.04)	0.29 (0.06)
	trunk diameter	0.82 (0.03)	0.59 (0.04)	0.59 (0.04)	0.35 (0.11)
	trunk increment	0.59 (0.04)	0.53 (0.04)	0.53 (0.05)	0.25 (0.14)
Pine dataset 208 genotypes 20,796 SNPs	D	0.95 (0.03)	0.00 (0.17)	0.03 (0.22)	0.00 (0.07)
	H	0.97 (0.01)	0.30 (0.39)	0.56 (0.16)	0.44 (0.52)
	I	0.73 (0.16)	0.00 (0.01)	0.62 (0.22)	0.00 (0.00)
	T4	0.74 (0.14)	0.54 (0.22)	0.62 (0.22)	0.59 (0.24)
	T5	0.52 (0.23)	0.30 (0.44)	0.48 (0.28)	0.00 (0.23)
	T6	0.28 (0.34)	0.08 (0.31)	0.30 (0.35)	0.00 (0.11)

3.3. Factors contributing to the variations in median PA gain

To further investigate the factors contributing to the variations in median PA gain across species and traits, correlations were computed between the median PA gain and the median h^2 differences (deviations), as well as with several additional statistics defined in Section 2.3: entropy of genotype frequencies in datasets, Shannon joint entropy of genotype and site (or environment) in datasets, and the F_{ST} calculated from the clusters estimated through the combination of UMAP and K -means applied to SNP marker data. These analyses provide valuable insights into the underlying mechanisms driving WISER's improved performance, particularly its capacity to address confounding relationships.

Figure 1 presents correlation plots between the 66 median PA gain values and various statistics, as defined in Section 2.3.4, for the 33 traits-species combinations. To enable correlation with the 66 PA gain values, each of the 33 statistic values was duplicated. A strong positive correlation of 0.71 was observed between the median PA gain values and the median h^2 deviations. This result aligns with expectations, as an increase in h^2 estimates typically reflects higher genetic variance estimates and

better predictive ability. In contrast, weak negative correlations were found between the median PA gain values and the Shannon entropy of genotype frequencies in datasets (-0.16), as well as the Shannon joint entropies of genotype and site (-0.25) and genotype and environment (-0.29) in datasets. Although these correlations were not strong, they suggested a trend where the median PA gain achieved by WISER tended to increase in scenarios with higher genotype-site or genotype-environment confounding. Notably, when considering the median h^2 deviations rather than the median PA gain, stronger trends emerged. More pronounced negative correlations with the Shannon joint entropies of genotype and site (-0.54) and genotype and environment (-0.56) were observed, highlighting the significant impact of the confounding between genotype and site or environment on genetic variance estimation. On the other hand, correlations close to zero were observed between the median PA gain and F_{ST} , as well as between the median h^2 deviations and F_{ST} , making the relationship inconclusive.

The weak negative or near-zero correlations observed in Figure 1 are likely unreliable and may reflect species-specific factors such as marker density, linkage disequilibrium patterns, and allele or genotype frequencies, which show little to no variation across traits within a given species. Table 3 presents the median values and interquartile ranges (IQR) of Shannon entropy for genotype frequencies ($H(Genotype)$) and F_{ST} values across traits for each species. Notably, it shows limited or no variation in these statistics among traits within each species, which likely contributed to the weak negative or near-zero correlations.

Figure 1 – Correlation plots between GBLUP median PA gain, median h^2 deviations, Shannon entropies, and F_{ST} values across all species and traits.

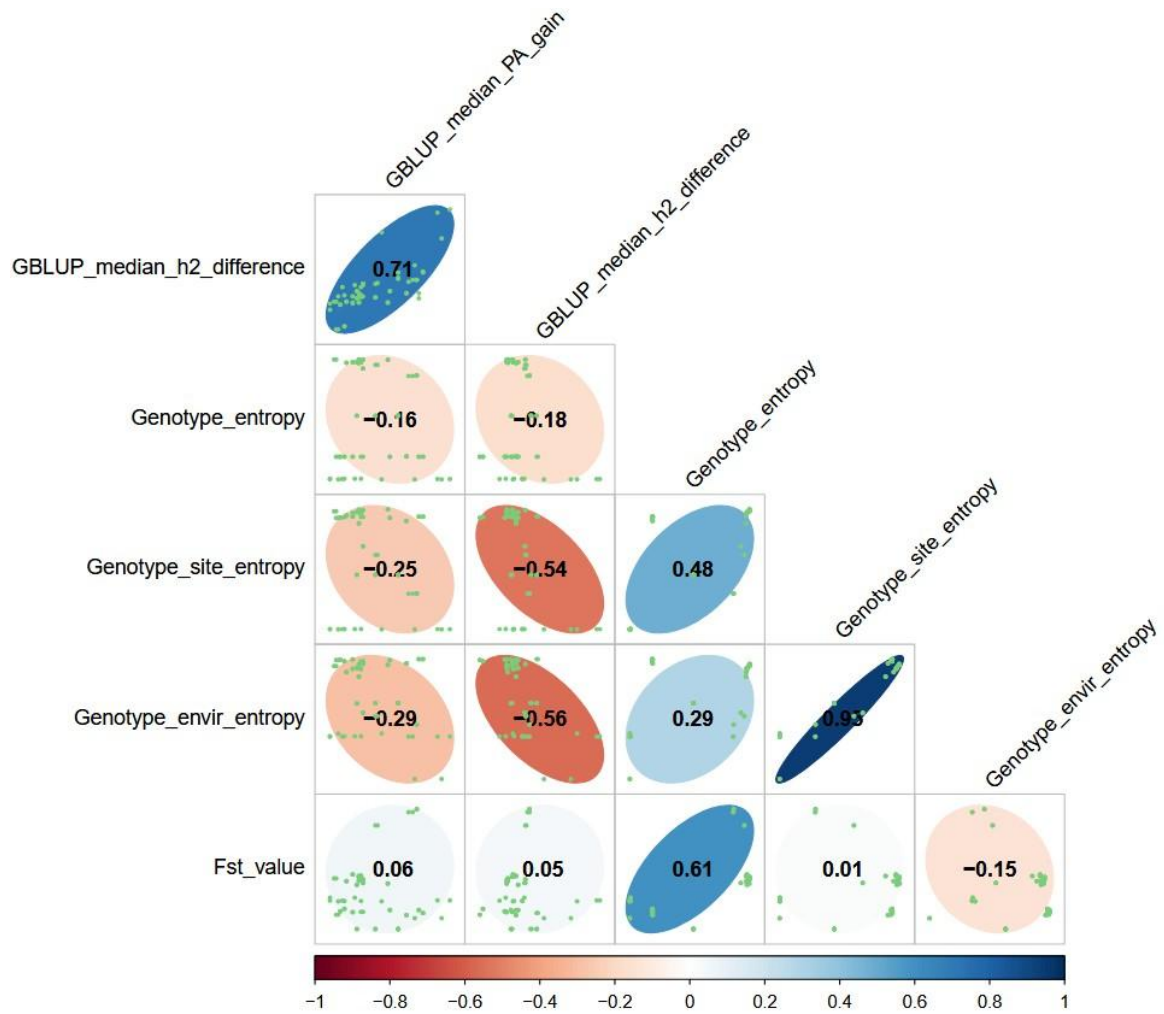
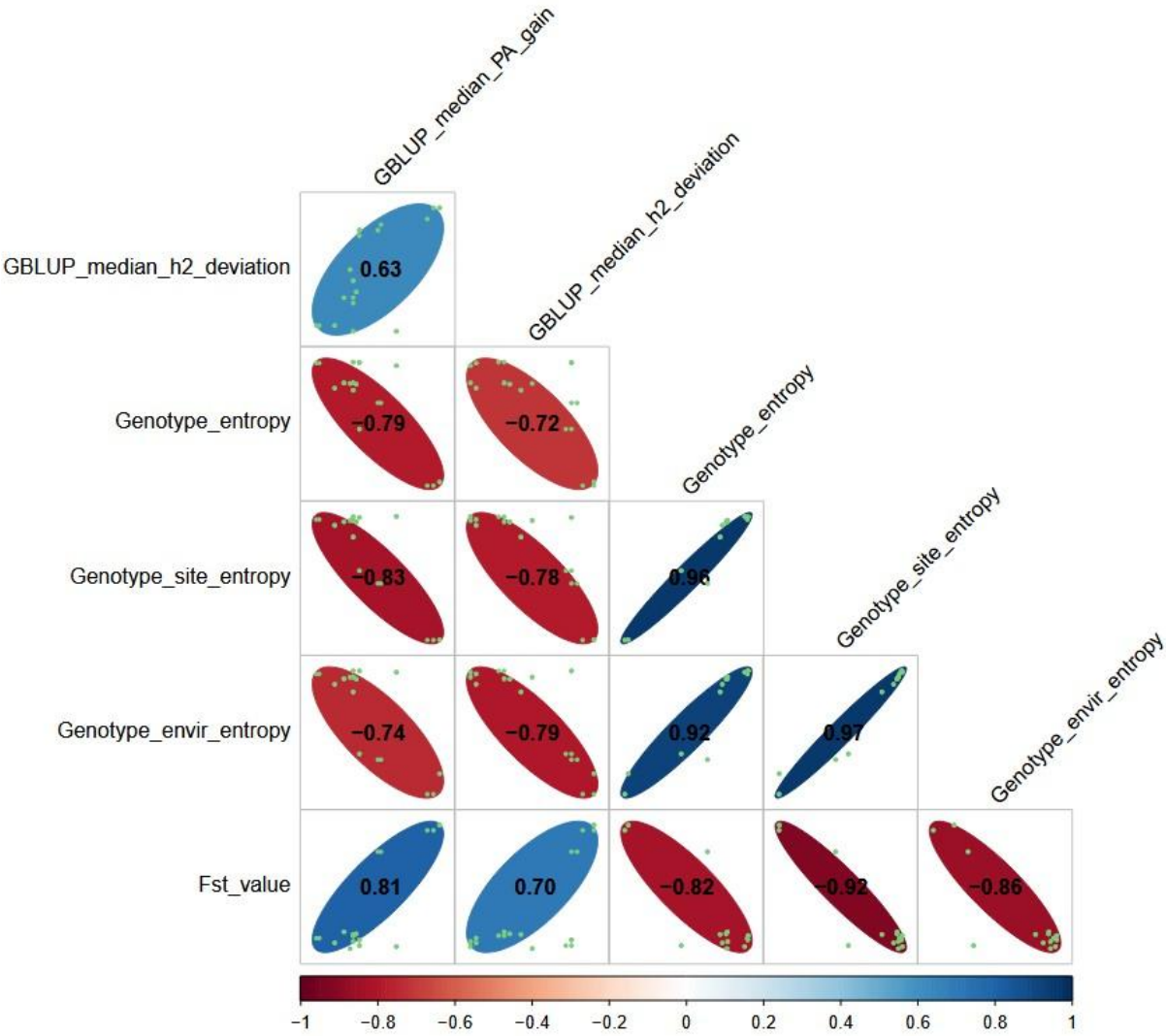


Table 3 – Median and interquartile range (IQR) values for $H(\text{Genotype})$ and F_{ST} across traits per species

Species	$median(H(\text{Genotype}))$	$IQR(H(\text{Genotype}))$	$median(F_{ST})$	$IQR(F_{ST})$
Rice	8.365	0.000	0.010	0.000
Maize	7.932	0.001	0.025	0.003
Apple	8.929	0.042	0.056	0.006
Pine	7.695	0.005	0.035	0.001

Figure 2 presents correlation plots for the apple dataset, generated using the same approach as in Figure 1. This dataset included the highest number of traits (14) among all species. The plots are based on 28 values for the median PA gain and median h^2 deviation, calculated from the differences between median values of WISER and BLUP, as well as WISER and LS-means. This enabled an exploration of intra-species relationships within a single species. The observed trends were consistent with those from the inter-species analysis, particularly regarding correlations between median PA gain, median h^2 deviation, and the Shannon entropies. A positive correlation between median PA gain and F_{ST} was also observed, however, this may be coincidental given the limited variability and dispersion seen in the corresponding plot.

Figure 2 – Correlation plots between GBLUP median PA gain, median h^2 deviations, Shannon entropies, and F_{ST} values for apple traits.



In summary, the negative correlations observed between the median PA gain values (or the median h^2 deviations) and the Shannon entropies highlight WISER's ability to improve phenotype estimation in the presence of confounding factors and population complexity.

3.4. Impact of whitening versus non-whitening of fixed-effect variables on genetic value estimation

Table 4 presents the medians and interquartile ranges (IQR, shown in parentheses) of estimation accuracy (EA) and mean squared error (MSE) by species for the simulated trait with moderate heritability (0.5), as described in Section 2.3.1. Results are reported for two estimation methods: WISER (with whitening of fixed-effect variables) and BLUE (without whitening), as detailed in Section 2.3.2. The results show that whitening generally improves genetic value estimation: EA medians were higher for WISER than for BLUE with the maize and pine datasets, and similar with the rice and apple datasets; on the contrary, MSE medians lower under WISER compared to BLUE in all species. These findings are consistent with trends observed in the real datasets and further highlight the advantages of whitening fixed-effect variables prior to estimation. Supplementary Figures 21 to 28 present the graphical summary of these results organized by species.

Table 4 – Median and interquartile range (IQR) of estimation accuracy (EA) and mean squared error (MSE) for simulated genetic values by species, under whitening (WISER) and non-whitening (BLUE) of fixed-effect variables.

Dataset	Statistic	WISER	BLUE
Rice	EA	0.71 (0.06)	0.72 (0.05)
	MSE	0.29 (0.31)	0.43 (0.5)
Maize	EA	0.45 (0.06)	0.21 (0.08)
	MSE	0.30 (0.46)	3.34 (1.77)
Apple	EA	0.66 (0.05)	0.66 (0.04)
	MSE	0.43 (0.27)	0.79 (1.3)
Pine	EA	0.46 (0.08)	0.07 (0.08)
	MSE	0.31 (0.57)	12.22 (15.86)

4. Discussion

This study demonstrates the significant improvement in phenotype estimation offered by the WISER approach compared to traditional methods like LS-means and BLUP, across four datasets and 33 traits. The results presented here highlight WISER's ability to handle genotype-environment confounding, genotype frequency imbalances, and complex population structures, which are major challenges in omics-based selection and association studies.

4.1. Comparison of WISER, LS-means, and BLUP

The findings demonstrate that WISER consistently outperforms both LS-means and BLUP approaches in terms of median PA and h^2 estimates. Across all species and traits, WISER achieved the highest GBLUP median PA values, followed by BLUP and LS-means, while BLUP-PCA often exhibited the lowest accuracy. Within the WISER framework, the vector v of phenotypes is not assumed to be a random vector drawn from a distribution with a specified covariance matrix. This design avoids imposing an unnecessary covariance structure during the estimation of v , which could lead to undesirable consequences. For example, assuming $v \sim N_q(0, \sigma_v^2 I_q)$ is often unrealistic and results in a decorrelated covariance structure in the best linear unbiased predictor (BLUP) of v . This is particularly problematic due to BLUP's inherent properties as a shrinkage estimator, especially in the context of imbalanced datasets. For instance, [77] emphasizes that BLUP estimates, when used as phenotypes, are not recommended if they are derived under the assumption that genetic values are I.I.D. when the data are unbalanced—a very frequent situation due to missing data or incomplete block experimental designs. For example, the experimental designs of the rice, maize, and pine datasets all included incomplete blocks, whereas the apple dataset featured management groups in which not all genotypes were represented (see Supplementary File 4). In such situations, the combination of the I.I.D. assumption and data imbalance leads to shrinkage of genotype estimates toward the population mean, with the degree of shrinkage being inversely proportional to the number of replicates for each genotype. This effect is particularly problematic when many genotypes have few replicates, as their BLUP values will disproportionately shrink toward the mean, distorting genetic rankings and undermining the interpretability and utility of these estimates for subsequent analyses.

The successive ordinary least squares (OLS) procedure implemented in WISER addresses these issues by operating without such assumptions. By treating v as a distinct vector of fixed effects, the OLS procedure ensures that its estimation is not derived as a linear or non-linear function of omic effects, and that v is estimated separately from them. This approach aligns with the recommendations of [77], providing reliable phenotypic estimates and preserving genetic value rankings, even in unbalanced experimental designs. Moreover, by ensuring that the estimation of v is not a linear or non-linear combination of marker effects, the approach preserves the genetic signals essential for association

analyses, ensuring that no valuable information is inadvertently filtered out—an issue that would arise if v was estimated using a polygenic background based on omic data like in GBLUP.

On the other hand, LS-means adjust for unequal group sizes by estimating marginal means as if all groups were equally represented, which can lead to values that differ from the actual group means. In unbalanced data, this may distort comparisons and reduce the reliability of downstream analyses. As noted by [78], such model-dependent estimators may not reflect the actual structure of the data. In contrast, WISER's avoidance of such model-dependent adjustments provides more stable and interpretable estimates across varying levels of data imbalance.

While LS-means and BLUP were generally more robust than BLUP-PCA, they still underperformed compared to WISER, especially in datasets where genotype-environment associations were strong and multicollinearity was severe. For only two traits—trunk increment (apple) and silking (maize)—the improvements in median PA for WISER were slight or negligible compared to LS-means. WISER's superior performance can be attributed to its whitening process, which reduces the effect of confounding between genotypes and environment. By mitigating such confounding, WISER yields more reliable and stable estimates, especially for datasets where traditional approaches for phenotype estimation struggle with multicollinearity, as seen in the pine dataset for traits with strong genotype-environment associations.

Furthermore, incorporating PC coordinates derived from genomic data as fixed-effect covariates did not improve BLUP-based phenotype estimation, as the GBLUP median PA for BLUP-PCA was often the lowest. The reduced GBLUP median PA observed in BLUP-PCA is potentially attributable to the inclusion of principal component coordinates as fixed effects, which captures genetic variation which is then missing in the BLUP estimates. This suggests that, depending on the phenotype estimation method used, using omic data to adjust fixed effects for confounding due to population structure does not systematically guarantee a notable improvement in the predictive ability of the estimated phenotypes. Unlike whitening methods that remove genetic correlations and confounding, the use of PC coordinates as fixed effects captures only a fraction of the total genetic variation and does not account for the complete genetic covariance structure (see Supplementary File 1). This highlights a key advantage of WISER, which effectively incorporates the entire genetic covariance structure without being constrained by the choice of the number of PCs, while improving phenotype estimation.

WISER consistently outperformed both LS-means and BLUP approaches in terms of median predictive ability (PA) and heritability (h^2) estimates, despite relying on spatial heterogeneity correction based solely on one-dimensional variables (e.g., row and column positions or GPS coordinates), modeled as fixed effects. In contrast, the other phenotype estimation methods incorporated both one- and two-dimensional components in their spatial correction. A potential improvement for WISER would be to adopt the approach proposed by [40], which involves modeling spatial heterogeneity as a random effect. This allows for the efficient inclusion of two-dimensional spatial components, which would be impractical to model as fixed effects due to the high number of levels. Furthermore, [40] demonstrated that fitting all relevant effects within a single model generally improves the estimation of genetic values.

Finally, note that one drawback of the WISER framework is the reduced interpretability of fixed-effect estimates. This can be limiting when environmental effects are of primary interest, as it complicates their use in downstream analyses or practical applications such as trial design or management recommendations. Although WISER performs well in estimating phenotypes and maintaining genetic rankings, it is less well-suited for situations requiring detailed interpretation of fixed environmental effects.

4.2. Impact of confounding effects and imbalanced frequencies

The study also explored how confounding effects and imbalanced frequencies of observations of genotypes in the environments explored in a dataset influence the accuracy of phenotype estimation. WISER was found to be particularly effective in mitigating these challenges. For instance, in the pine dataset, where traits such as D, H, and I exhibited strong genotype-environment associations and confounding, WISER provided more reliable phenotype estimates compared to LS-means and BLUP-PCA. This was because the REML algorithm in BLUP-PCA and the linear regression (prior to LS-means) encountered issues such as non-optimal convergence and multicollinearity, respectively. This finding is critical, as genotype-environment confounding, often exacerbated by high variability in genotype frequencies across environments, can severely compromise the accuracy of phenotype estimations and predictions. Moreover, imbalances in genotype frequencies, along with their variability across environments, inevitably affect the genetic covariance structure at the level of a whole experimental design. This underscores the complex interdependence of these factors and their combined impact on phenotype estimation. WISER's whitening process addresses these issues by simultaneously mitigating confounding, multicollinearity, and imbalances in genotype frequencies, ultimately improving the accuracy and robustness of phenotype estimates.

Notably, some degree of confounding, imbalance in genotype frequencies, and population structure are nearly inevitable, even in carefully balanced experimental designs. This is exemplified by the apple REFPOP dataset, which, despite being a reference population for apple breeding, still displayed imbalances and confounding. This highlights the pervasive nature of these challenges in real-world datasets and underscores the importance of methods like WISER, which can effectively address such complexities.

4.3. Impact of the degree of population structure

The study further explored the relationship between improvements in phenotype estimation for each trait using WISER and the degree of population structure, quantified through genetic differentiation. However, the relationship remained inconclusive due to species-specific factors and limited to no variation in the genomic data across traits for each species. Nonetheless, WISER proved effective in minimizing the confounding between population structure and environmental effects. Among the various computed statistics, the strongest negative correlations were observed between both the median PA gain and the median h^2 deviations, and the Shannon joint entropy of genotype and environment. Hence, these findings point out the critical importance of considering population structure when selecting methodologies for phenotype estimation. As demonstrated in this study, population structure significantly impacts genomic predictive ability through confounding between genotypes and environments, and it is a well-established factor influencing association studies [6–12].

4.4. Phenotype estimation, genetic variance, and predictive ability

The positive correlation between the median PA gain and the median h^2 deviations underscores the relationships between phenotype estimation, genetic variance, and predictive ability. Across all species and traits, a strong positive correlation (0.71) was observed between the median PA gain and the median h^2 deviations. This demonstrates that WISER-estimated phenotypes improve the capture of genetic variance when using GBLUP, thereby enhancing predictions. Interestingly, the median h^2 deviations exhibited moderate negative correlations with the Shannon joint entropies of genotype and site or environment. Similarly, weaker negative correlations were observed between the median PA gain values and these Shannon joint entropies. These trends were also observed in the apple dataset, where they appeared more pronounced. These findings suggest that WISER is particularly effective in addressing scenarios characterized by a high degree of genotype-site and/or genotype-environment confounding, thereby improving phenotype estimation in these complex situations, as well as improving genetic variance estimation and predictive ability.

4.5. Potential for genome-wide association studies (GWAS)

Although not explicitly demonstrated in this study, WISER shows potential for improving genome-wide association studies (GWAS). Preliminary analyses (results not shown) indicate that several known quantitative trait loci (QTLs) in apple are consistently detected using both LS-means and WISER-estimated phenotypes as response variables, while additional distinct QTLs are identified depending on the phenotype used. This suggests that incorporating WISER-estimated phenotypes into association studies, alongside traditional methods such as LS-means may reinforce confidence in QTLs identified using different phenotype estimation approaches. Additionally, these preliminary findings also revealed that WISER-estimated phenotypes could uncover additional QTLs that explain a significant portion of phenotypic variation. However, a more detailed investigation into the full potential of WISER-estimated phenotypes in association studies will be the subject of future research, as exploring this in-depth is beyond the scope of the current study.

4.6. Limitations, computational efficiency, and population size considerations

WISER's strengths are most apparent when multiple fixed-effect variables are considered in defining the environmental variable. However, its utility is potentially limited in experimental designs with very few fixed-effect variables (e.g., only an overall mean and one additional factor), as these do not permit a more precise definition of the environmental variable. This highlights the importance of comprehensive experimental designs to fully harness WISER's potential. While computational efficiency may have been a concern in the past, the current availability of high-performance computing (HPC) clusters has largely mitigated this issue. For example, processing times for the rice and pine datasets were under five hours, depending on resource availability and user demand on the HPC cluster used. Furthermore, WISER employs parallelized algorithms, such as an ABC approach, which ensures rapid and stable variance component estimation, even for large datasets linked to complex experimental designs. This approach is preferable to iterative algorithms, such as expectation-maximization (EM), which can be slow, produce inaccurate results, or converge to suboptimal solutions, especially when handling large datasets [79–81]. Nonetheless, it is important to note that WISER may encounter scalability challenges when applied to datasets with populations significantly larger than those in the apple REFPOP dataset. Future work should explore WISER's performance on larger datasets.

4.7. Historical context, implications for breeding programs, versatility of the whitening approach and future directions

Historically, tools for managing confounding effects due to population structure and imbalanced genotype frequencies in phenotype estimation were limited. This led to the widespread use of methods like BLUP. BLUP remains a popular choice due to its simplicity and the availability of user-friendly R packages like `lme4` [64]. However, BLUP's reliance on the I.I.D. assumption in this context and its sensitivity to genotype frequency imbalances result in known suboptimal performance, as pointed out by [77]. The `wiser` R package addresses this gap by offering a more robust framework for phenotype estimation, specifically tailored for experimental setups with population structure, confounding, and imbalances in genotype frequencies. By overcoming these challenges, WISER has significant implications for breeding and genetic selection programs. Accurate phenotype estimation is critical for predicting complex traits and selecting individuals with desirable genetic profiles. The reduced interquartile ranges (IQR) for GBLUP PA values observed with WISER demonstrate its ability to deliver more consistent predictions across cross-validation scenarios, a feature invaluable for breeders seeking to improve the reliability of selection decisions. Looking ahead, the capacity of WISER to handle more complex datasets with additional environmental covariates represents an exciting avenue for further research. Future studies could also explore integrating WISER with advanced predictive modeling techniques, such as deep learning, to enhance phenotype predictive accuracy.

Furthermore, incorporating genotype-environment interactions into WISER's estimation process represents a promising direction for refining phenotype estimates. These advancements have the potential to unlock even greater possibilities for phenotype estimation and omic-based selection, driving innovation in breeding programs.

5. Conclusion

In conclusion, the results of this study provide strong evidence that WISER offers significant advantages in phenotype estimation, particularly in datasets with complex genotype-environment confounding, genotype frequency imbalances, and population structure. By addressing these challenges, WISER enables more accurate and stable phenotype estimations, and predictions, which could lead to improved selection strategies in crop and livestock breeding programs.

Abbreviations

ABC	Approximate Bayesian computation
AIC	Akaike information criterion
ANOVA	Analysis of variance
BLUE	Best linear unbiased estimation
BLUP	Best linear unbiased prediction
CH index	Calinski-Harabasz index
CV	Cross-validation
D	Duration of budburst
EA	Estimation accuracy
EM	Expectation-maximization
EVG	covariance analysis via eigenvectors
FL	Days to flowering
GBLUP	Genomic best linear unbiased prediction
GPS	Global positioning system
GWAS	Genome-wide association studies
H	Height
I	Annual height increment
IQR	Interquartile range
LASSO	Least absolute shrinkage and selection operator
LS-means	Least squares means
NIRS	Near-infrared spectroscopy

889 OLS Ordinary least squares

890 PA Predictive ability

891 PC Principal component

892 PCA Principal component analysis

893 PCA-cor Principal component correlation analysis

894 PH Plant height

895 QTL Quantitative trait locus

896 REFPOP Reference population

897 REML Restricted maximum likelihood

898 RF Random forest

899 RKHS Reproducing kernel Hilbert space

900 SNP Single nucleotide polymorphism

901 SVR Support vector regression

902 T4 Time taken to reach individual budburst stages: stage 4

903 T5 Time taken to reach individual budburst stages: stage 5

904 T6 Time taken to reach individual budburst stages: stage 6

905 UMAP Uniform manifold approximation and projection

906 VIF Variance inflation factor

907 WISER Whitening and successive least squares estimation refinement

908 YLD Grain yield

909 ZCA-cor Zero-phase component correlation analysis

910 ZN Zinc concentration

911 Supplementary information

912 Supplementary File 1

913 Mathematical proofs demonstrating that WISER extends models that use eigen-information as fixed-
 914 effect covariates to correct for population structure, by relaxing their assumptions and employing a
 915 true whitening matrix instead of a pseudo-whitening matrix.

916 Supplementary File 2

917 Presents a detailed overview of the ABC algorithm implemented within the WISER framework for
 918 variance components (i.e., σ_u^2 and σ_ε^2) estimation, along with the regularization and shrinkage
 919 techniques used to ensure positive-definiteness for the genetic covariance matrix Σ_u .

920 Supplementary File 3

921 Presents median predictive ability (PA) values and interquartile ranges (IQR) in parentheses, across all
 922 species and traits, for several prediction models including random forest (RF), support vector

923 regression (SVR), genomic best linear unbiased prediction (GBLUP), reproducing kernel Hilbert space
924 (RKHS) regression, and least absolute shrinkage and selection operator (LASSO).

925 Supplementary Figures 1–4

926 PCA and UMAP 2D projections of the apple REFPOP genomic data, shown with and without family
927 cluster labels to visualize population structure.

928 Supplementary Figures 5–16

929 Heatmaps, PCA, and UMAP 2D plots illustrating population structure in the rice, maize, apple, and pine
930 genomic datasets.

931 Supplementary Figures 17–20

932 Boxplots and violin plots of the GBLUP predictive ability (PA) values for traits of each species, for
933 phenotypes estimated using WISER, LS-means, and BLUP approaches.

934 Supplementary Figures 21–28

935 Boxplots and violin plots of mean squared error (MSE) and estimation accuracy (EA) between
936 simulated and estimated genetic values for WISER (with whitening of fixed-effect variables) and BLUE
937 (without whitening), based on a generic trait simulated with moderate heritability (0.5) under each
938 species' experimental design.

939

940 Authors' contributions

941 LJ derived the algebra for the analytical results, created the GitHub repositories, developed the
942 associated R scripts, developed the WISER R package and conducted all dataset analyses. LJ also
943 authored the manuscript. WG, ML, AP, MR, CED, FL, LL, MJA, and HM reviewed and approved the final
944 version of the manuscript.

945

946 Authors' information

947 The authors declare no conflict of interests.

948

949 Funding

950 Phenotypic data collection was partially supported by the Horizon 2020 Framework Program of the
951 European Union under grant agreement No 817970 (project INVITE: "Innovations in plant variety
952 testing in Europe to foster the introduction of new varieties better adapted to varying biotic and abiotic
953 conditions and to more sustainable crop management practices") and by the SusCrop Agrobiodiversity
954 project Apple BIOME ("Microbiome and genomic analysis in apple germplasm towards broadening
955 genetic resources to breed for resilient varieties"). LJ was supported by the ANR grant for the Apple
956 BIOME project (ANR-22-SUSC-0001-05).

957

958 Acknowledgments

The authors thank the field technicians and staff at INRAE IRHS and Experimental Unit (UE Horti), Angers, France, the Fruit Breeding Group at Agroscope in Wädenswil, Switzerland, and technical staff at all apple REFPOP sites for the maintenance of the orchards and phenotypic data collection.

Availability of data and materials

All datasets and scripts used in this study are publicly available on GitHub:

- https://github.com/ljacquin/wiser_genomic_prediction_rice
- https://github.com/ljacquin/wiser_genomic_prediction_maize
- <https://github.com/ljacquin/refpop>
- https://github.com/ljacquin/wiser_genomic_prediction_pine
- https://github.com/ljacquin/compute_stats_wiser_results

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, F-49000 Angers, France. ²Research Centre Laimburg, Laimburg 1, 39040 Auer, Italy. ³The National Institute of Horticultural Research, Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland. ⁴Agroscope, Fruit Breeding, Mueller-Thurgau-Strasse 29, 8820 Wädenswil, Switzerland. ⁵Better3fruit N.V., 3202 Rillaar, Belgium. ⁶Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, 08193 Bellaterra, Barcelona, Spain. ⁷IRTA (Institut de Recerca i Tecnologia Agroalimentàries), 08140 Caldes de Montbui, Barcelona, Spain.

References

1. Byrne RP, van Rheenen W, van den Berg LH, Veldink JH, McLaughlin RL. Dutch population structure across space, time and GWAS design. *Nat Commun.* 2020;11:4556.
2. Gloss AD, Steiner MC, Novembre J, Bergelson J. The design of mapping populations: Impacts of geographic scale on genetic architecture and mapping efficacy for defense and immunity. *Current Opinion in Plant Biology.* 2023;74:102399.
3. Sul JH, Martin LS, Eskin E. Population structure in genetic studies: Confounding factors and mixed models. *PLOS Genetics.* 2018;14:e1007309.
4. Astle W, Balding DJ. Population Structure and Cryptic Relatedness in Genetic Association Studies. *Statistical Science.* 2009;24:451–71.

996 5. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association Mapping in Structured
997 Populations. *The American Journal of Human Genetics*. 2000;67:170–81.

998 6. Azevedo CF, Resende MDV de, Silva FF e, Nascimento M, Viana JMS, Valente MSF. Population
999 structure correction for genomic selection through eigenvector covariates. *Crop Breed Appl*
1000 *Biotechnol*. 2017;17:350–8.

1001 7. Daetwyler HD, Kemper KE, van der Werf JHJ, Hayes BJ. Components of the accuracy of genomic
1002 prediction in a multi-breed sheep population. *J Anim Sci*. 2012;90:3375–84.

1003 8. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, et al. Efficient control of
1004 population structure in model organism association mapping. *Genetics*. 2008;178:1709–23.

1005 9. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. *PLOS Genetics*.
1006 2006;2:e190.

1007 10. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components
1008 analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904–9.

1009 11. Riedelsheimer C, Technow F, Melchinger AE. Comparison of whole-genome prediction models for
1010 traits with contrasting genetic architecture in a diversity panel of maize inbred lines. *BMC Genomics*.
1011 2012;13:452.

1012 12. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model
1013 method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*.
1014 2006;38:203–8.

1015 13. Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. An efficient multi-locus mixed
1016 model approach for genome-wide association studies in structured populations. *Nat Genet*.
1017 2012;44:825–30.

1018 14. Fritsche-Neto R, Galli G, Borges KLR, Costa-Neto G, Alves FC, Sabadin F, et al. Optimizing
1019 Genomic-Enabled Prediction in Small-Scale Maize Hybrid Breeding Programs: A Roadmap Review.
1020 *Front Plant Sci*. 2021;12:658267.

1021 15. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis.
1022 *Am J Hum Genet*. 2011;88:76–82.

1023 16. Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, et al. Genome-wide association
1024 mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun*.
1025 2011;2:467.

1026 17. Pritchard JK. Deconstructing maize population structure. *Nat Genet*. 2001;28:203–4.

1027 18. Baertschi C, Cao T-V, Bartholomé J, Ospina Y, Quintero C, Frouin J, et al. Impact of early genomic
1028 prediction for recurrent selection in an upland rice synthetic population. *G3*
1029 *Genes|Genomes|Genetics*. 2021;11:jkab320.

1030 19. Jung M, Roth M, Aranzana MJ, Auwerkerken A, Bink M, Denancé C, et al. The apple REFPOP—a
1031 reference population for genomics-assisted breeding in apple. *Hortic Res*. 2020;7:1–16.

1032 20. Millet EJ, Kruijer W, Coupel-Ledru A, Alvarez Prado S, Cabrera-Bosquet L, Lacube S, et al. Genomic
1033 prediction of maize yield across European environmental conditions. *Nat Genet*. 2019;51:952–6.

1034 21. Amadeu RR, Ferrão LFV, Oliveira I de B, Benevenuto J, Endelman JB, Munoz PR. Impact of
1035 dominance effects on autotetraploid genomic prediction. *Crop Science*. 2020;60:656–65.

1036 22. Lorenzi A, Bauland C, Mary-Huard T, Pin S, Palaffre C, Guillaume C, et al. Genomic prediction of
1037 hybrid performance: comparison of the efficiency of factorial and tester designs used as training sets
1038 in a multiparental connected reciprocal design for maize silage. *Theor Appl Genet*. 2022;135:3143–
1039 60.

1040 23. Lorenzi A, Bauland C, Pin S, Madur D, Combes V, Palaffre C, et al. Portability of genomic
1041 predictions trained on sparse factorial designs across two maize silage breeding cycles. *Theor Appl*
1042 *Genet*. 2024;137:75.

1043 24. Njuguna JN, Clark LV, Lipka AE, Anzoua KG, Bagmet L, Chebukin P, et al. Genome-wide association
1044 and genomic prediction for yield and component traits of *Miscanthus sacchariflorus*. *GCB Bioenergy*.
1045 2023;15:1355–72.

1046 25. Rio S, Gallego-Sánchez L, Montilla-Bascón G, Canales FJ, Isidro Y Sánchez J, Prats E. Genomic
1047 prediction and training set optimization in a structured Mediterranean oat population. *Theor Appl*
1048 *Genet*. 2021;134:3595–609.

1049 26. Sakhale SA, Yadav S, Clark LV, Lipka AE, Kumar A, Sacks EJ. Genome-wide association analysis for
1050 emergence of deeply sown rice (*Oryza sativa*) reveals novel aus-specific phytohormone candidate
1051 genes for adaptation to dry-direct seeding in the field. *Front Plant Sci*. 2023;14:1172816.

1052 27. Samira R, Kimball JA, Samayoa LF, Holland JB, Jamann TM, Brown PJ, et al. Genome-wide
1053 association analysis of the strength of the MAMP-elicited defense response and resistance to target
1054 leaf spot in sorghum. *Sci Rep*. 2020;10:20817.

1055 28. Tucker JR, Brûlé-Babel AL, Hiebert CW, Larios R, Legge WG, Badea A, et al. Genetic structure and
1056 genome-wide association study of a genomic panel of two-row, spring barley (*Hordeum vulgare* L.)
1057 with differential reaction to *Fusarium* head blight (*Fusarium graminearum* Schwabe) and
1058 deoxynivalenol production. *Canadian Journal of Plant Pathology*. 2022;44:874–91.

1059 29. Vogel G, Giles G, Robbins KR, Gore MA, Smart CD. Quantitative Genetic Analysis of Interactions in
1060 the Pepper-*Phytophthora capsici* Pathosystem. *Mol Plant Microbe Interact*. 2022;35:1018–33.

1061 30. Chen Y, Gao Y, Chen P, Zhou J, Zhang C, Song Z, et al. Genome-wide association study reveals
1062 novel quantitative trait loci and candidate genes of lint percentage in upland cotton based on the
1063 CottonSNP80K array. *Theor Appl Genet*. 2022;135:2279–95.

1064 31. Li H, Cheng X, Zhang L, Hu J, Zhang F, Chen B, et al. An Integration of Genome-Wide Association
1065 Study and Gene Co-expression Network Analysis Identifies Candidate Genes of Stem Lodging-Related
1066 Traits in *Brassica napus*. *Front Plant Sci*. 2018;9.

1067 32. McLeod L, Barchi L, Tumino G, Tripodi P, Salinier J, Gros C, et al. Multi-environment association
1068 study highlights candidate genes for robust agronomic quantitative trait loci in a novel worldwide
1069 *Capsicum* core collection. *The Plant Journal*. 2023;116:1508–28.

1070 33. Zhang J, Song Q, Cregan PB, Jiang G-L. Genome-wide association study, genomic prediction and
1071 marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor Appl Genet*.
1072 2016;129:117–30.

1073 34. Carress H, Lawson DJ, Elhaik E. Population genetic considerations for using biobanks as
1074 international resources in the pandemic era and beyond. *BMC Genomics*. 2021;22:351.

1075 35. Elhaik E. Principal Component Analyses (PCA)-based findings in population genetic studies are
1076 highly biased and must be reevaluated. *Sci Rep.* 2022;12:14683.

1077 36. Lawson DJ, Davies NM, Haworth S, Ashraf B, Howe L, Crawford A, et al. Is population structure in
1078 the genetic biobank era irrelevant, a challenge, or an opportunity? *Hum Genet.* 2020;139:23–41.

1079 37. Liu L, Zhang D, Liu H, Arendt C. Robust methods for population stratification in genome wide
1080 association studies. *BMC Bioinformatics.* 2013;14:132.

1081 38. Yao Y, Ochoa A. Limitations of principal components in quantitative genetic association models
1082 for human studies. *Elife.* 2023;12:e79238.

1083 39. Wang M, Fang Z, Yoo B, Bejerano G, Peltz G. The Effect of Population Structure on Murine
1084 Genome-Wide Association Studies. *Front Genet.* 2021;12.

1085 40. Fernández-González J, Isidro y Sánchez J. Optimizing fully-efficient two-stage models for genomic
1086 selection using open-source software. *Plant Methods.* 2025;21:9.

1087 41. Ermolov A, Siarohin A, Sangineto E, Sebe N. Whitening for Self-Supervised Representation
1088 Learning. In: *Proceedings of the 38th International Conference on Machine Learning.* PMLR; 2021. p.
1089 3015–24.

1090 42. Gluckman J. Higher order whitening of natural images. In: *2005 IEEE Computer Society*
1091 *Conference on Computer Vision and Pattern Recognition (CVPR'05).* 2005. p. 354–60 vol. 2.

1092 43. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat*
1093 *Genet.* 2012;44:821–4.

1094 44. Henderson CR. *Applications of linear models in animal breeding.* Guelph, Ont.: University of
1095 Guelph; 1984.

1096 45. VanRaden PM. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science.*
1097 2008;91:4414–23.

1098 46. Hu H, Campbell MT, Yeats TH, Zheng X, Runcie DE, Covarrubias-Pazaran G, et al. Multi-omics
1099 prediction of oat agronomic and seed nutritional traits across environments and in distantly related
1100 populations. *Theor Appl Genet.* 2021;134:4043–54.

1101 47. Kessy A, Lewin A, Strimmer K. Optimal Whitening and Decorrelation. *The American Statistician.*
1102 2018;72:309–14.

1103 48. Podosinnikova A. On the method of moments for estimation in latent linear models. phdthesis.
1104 Université Paris sciences et lettres; 2016.

1105 49. Wang C, Tindemans SH, Palensky P. Improved Anomaly Detection and Localization Using
1106 Whitening-Enhanced Autoencoders. *IEEE Trans Ind Inf.* 2024;20:659–68.

1107 50. Kwan CCY. *An Introduction to Shrinkage Estimation of the Covariance Matrix: A Pedagogic*
1108 *Illustration. Spreadsheets in Education.* 2011.

1109 51. Nikovski D, Byadarhaly K. Regularized covariance matrix estimation with high dimensional data
1110 for supervised anomaly detection problems. In: *2016 International Joint Conference on Neural*
1111 *Networks (IJCNN).* Vancouver, BC, Canada: IEEE; 2016. p. 2811–8.

1112 52. Schäfer J, Strimmer K. A shrinkage approach to large-scale covariance matrix estimation and
1113 implications for functional genomics. *Stat Appl Genet Mol Biol*. 2005;4:Article32.

1114 53. Steland A. Shrinkage for covariance estimation: asymptotics, confidence intervals, bounds and
1115 applications in sensor monitoring and finance. *Stat Papers*. 2018;59:1441–62.

1116 54. Theiler J. The incredible shrinking covariance estimator. Baltimore, Maryland; 2012. p. 83910P-
1117 83910P – 12.

1118 55. Touloumis A. Nonparametric Stein-type shrinkage covariance matrix estimators in high-
1119 dimensional settings. *Computational Statistics & Data Analysis*. 2015;83:251–61.

1120 56. Jung M, Keller B, Roth M, Aranzana MJ, Auwerkerken A, Guerra W, et al. Genetic architecture and
1121 genomic predictive ability of apple quantitative traits across environments. *Horticulture Research*.
1122 2022;9:uhac028.

1123 57. Perry A, Wachowiak W, Beaton J, Iason G, Cottrell J, Cavers S. Identifying and testing marker–trait
1124 associations for growth and phenology in three pine species: Implications for genomic prediction.
1125 *Evolutionary Applications*. 2022;15:330–48.

1126 58. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and
1127 Projection. *Journal of Open Source Software*. 2018;3:861.

1128 59. Kolde R. pheatmap: Pretty Heatmaps. 2010;:1.0.13.

1129 60. Rohart F, Gautier B, Singh A, Cao K-AL. mixOmics: An R package for ‘omics feature selection and
1130 multiple data integration. *PLOS Computational Biology*. 2017;13:e1005752.

1131 61. Konopka T. umap: Uniform Manifold Approximation and Projection. 2023.

1132 62. Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PHC. Correcting for spatial heterogeneity
1133 in plant breeding experiments with P-splines. *Spatial Statistics*. 2018;23:52–71.

1134 63. Lenth RV. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*.
1135 2016;69:1–33.

1136 64. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal*
1137 *of Statistical Software*. 2015;67:1–48.

1138 65. Hoffman GE. Correcting for Population Structure and Kinship Using the Linear Mixed Model:
1139 Theory and Extensions. *PLOS ONE*. 2013;8:e75707.

1140 66. Beugin M-P, Gayet T, Pontier D, Devillard S, Jombart T. A fast likelihood solution to the genetic
1141 clustering problem. *Methods in Ecology and Evolution*. 2018;9:1006–16.

1142 67. Korbinian Strimmer, Takoua Jendoubi, Agnan Kessy, Alex Lewin. whitening: Whitening and High-
1143 Dimensional Canonical Correlation Analysis. 2018;:1.4.0.

1144 68. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional
1145 Data in C++ and R. *Journal of Statistical Software*. 2017;77:1–17.

1146 69. Karatzoglou A, Smola A, Hornik K, Zeileis A. kernlab - An S4 Package for Kernel Methods in R.
1147 *Journal of Statistical Software*. 2004;11:1–20.

1148 70. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via
1149 Coordinate Descent. *J Stat Softw.* 2010;33:1–22.

1150 71. Jacquin L, Cao T-V, Ahmadi N. A Unified and Comprehensible View of Parametric and Kernel
1151 Methods for Genomic Prediction with Application to Rice. *Frontiers in Genetics.* 2016;7.

1152 72. Cherkassky V, Ma Y. Practical selection of SVM parameters and noise estimation for SVM
1153 regression. *Neural Networks.* 2004;17:113–26.

1154 73. Meyer D, Zeileis A, Hornik K, Friendly M. vcd: Visualizing Categorical Data. 2002;;1.4-13.

1155 74. Hennig C. fpc: Flexible Procedures for Clustering. 2003;;2.2-13.

1156 75. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics.*
1157 2008;24:1403–5.

1158 76. Goudet J, Jombart T. hierfstat: Estimation and Tests of Hierarchical F-Statistics. 2004;;0.5-11.

1159 77. Holland JB, Piepho H-P. Don't BLUP Twice. *G3 Genes|Genomes|Genetics.* 2024;14:jkae250.

1160 78. Searle SR, Speed FM, Milliken GA. Population Marginal Means in the Linear Model: An Alternative
1161 to Least Squares Means. *The American Statistician.* 1980;34:216–21.

1162 79. Bradley P, Fayyad U, Reina C. Scaling EM (Expectation Maximization) Clustering to Large
1163 Databases. 1998.

1164 80. Ordonez C, Omiecinski E. FREM: fast and robust EM clustering for large data sets. In: *Proceedings*
1165 *of the eleventh international conference on Information and knowledge management.* McLean
1166 Virginia USA: ACM; 2002. p. 590–9.

1167 81. Xu L, Jordan MI. On Convergence Properties of the EM Algorithm for Gaussian Mixtures. *Neural*
1168 *Computation.* 1996;8:129–51.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure2.png](#)
- [SupplementaryFigure1.png](#)
- [SupplementaryFigure6.png](#)
- [SupplementaryFigure4.png](#)
- [SupplementaryFigure5.png](#)
- [SupplementaryFigure7.png](#)
- [SupplementaryFigure10.png](#)
- [SupplementaryFigure3.png](#)
- [SupplementaryFigure8.png](#)
- [SupplementaryFigure13.png](#)
- [SupplementaryFigure15.png](#)
- [SupplementaryFigure14.png](#)
- [SupplementaryFigure16.png](#)
- [SupplementaryFigure12.png](#)
- [SupplementaryFigure9.png](#)
- [SupplementaryFigure11.png](#)
- [SupplementaryFigure21.png](#)
- [SupplementaryFigure23.png](#)
- [SupplementaryFigure17.png](#)
- [SupplementaryFigure22.png](#)
- [SupplementaryFigure27.png](#)
- [SupplementaryFile3.xlsx](#)
- [SupplementaryFigure28.png](#)
- [SupplementaryFigure25.png](#)
- [SupplementaryFigure18.png](#)
- [SupplementaryFigure19.png](#)
- [SupplementaryFile4.xlsx](#)
- [SupplementaryFigure26.png](#)
- [SupplementaryFile2.docx](#)
- [SupplementaryFigure24.png](#)
- [SupplementaryFigure20.png](#)

- [SupplementaryFile1.docx](#)