Statistics Innovation

ARTICLE

Open Access

https://doi.org/10.48130/stati-0026-0012

Statistics Innovation **2026**, 3: e011

dynGBLUP: a unified framework for genomic prediction and association mapping of dynamic phenotypes

Xinxin Yang[1], Ruiqi Zhao[1], Xiaolin Li[1] and Meixia Ye[2*]

[1] *School of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China*
[2] *School of Grassland Science, Beijing Forestry University, Beijing 100083, China*
* Correspondence: yemeixia123@bjfu.edu.cn (Ye M)

Abstract

Genomic best linear unbiased prediction (GBLUP) serves as a cornerstone of genomic selection (GS) by integrating genomic relationships into a mixed model to estimate breeding values. While powerful, conventional GBLUP frameworks are primarily designed for static traits, limiting their ability to capture developmental dynamics. To address this limitation, we introduce dynamic genomic best linear unbiased prediction (dynGBLUP), a unified model that jointly predicts dynamic phenotypes and estimates genome-wide SNP effects across multiple time points. dynGBLUP captures developmental dynamics by integrating flexible growth curves and further accounts for population structure and kinship while estimating variance-covariance parameters via restricted maximum likelihood. By simultaneously modeling all SNPs as random effects, it captures the polygenic architecture underlying complex traits, a key distinction from single-marker association methods. Evaluated on stem diameter of *Populus szechuanica* var. *tibetica*, dynGBLUP achieved a predictive accuracy of 0.85, compared with 0.72 for rrBLUP. Among machine learning methods, random forest, elastic net, and SVM achieved predictive accuracies of 0.71, 0.73, and 0.80, respectively. Simulations further confirmed controlled false-positive rates, with quantitative trait locus (QTL) detection rates reaching up to 77% as sample size and heritability increased. By integrating growth curves with polygenic modeling, dynGBLUP provides a unified framework that bridges dynamic phenotyping and genomic prediction, establishing a methodological foundation for dissecting the genetic architecture of dynamic traits and advancing the integration of high-dimensional phenotyping with polygenic modeling.

Keywords: Dynamic phenotypes, Best linear unbiased prediction, Genomic selection, Genomic prediction, Genome-wide association study

Citation: Yang X, Zhao R, Li X, Ye M. 2026. dynGBLUP: a unified framework for genomic prediction and association mapping of dynamic phenotypes. *Statistics Innovation* 3: e011 https://doi.org/10.48130/stati-0026-0012

Introduction

As a foundational method in quantitative genetics, best linear unbiased prediction (BLUP) predicts breeding values using a mixed linear model. Its ability to minimize prediction variance while ensuring unbiased estimates has underpinned its widespread application in the genetic dissection of key traits, shortening breeding cycles in animal and plant improvement[1]. Central to BLUP is the mixed model, a framework that has driven key methodological advances in genome-wide association studies (GWAS) while simultaneously ushering in the era of genomic selection (GS) and genomic prediction (GP) in breeding[2,3]. As a foundational method within GS, genomic best linear unbiased prediction (GBLUP) estimates marker effects via the genomic relationship matrix, serving as a critical bridge between GWAS and GS. By providing a unified and robust statistical framework for genomic data analysis, GBLUP establishes a solid foundation for genome-enabled selection and stands as a theoretical cornerstone of modern breeding systems[4,5].

GBLUP translates the principles of quantitative genetics into a practical statistical framework for genomic prediction. Rooted in the infinitesimal model, BLUP assumes an additive genetic architecture in which complex traits are governed by numerous genes of small effect[4]. This additive assumption provides the theoretical clarity that sets GBLUP apart from current artificial intelligence (AI) approaches, which excel at capturing complex relationships between genotypes and phenotypes but operate as 'black boxes' with opaque internal workings[6]. For breeders seeking not only predictive accuracy but also mechanistic insight, GBLUP offers a transparent framework in which the influence of relatedness and variance components on random effects can be directly interpreted. In doing so, it transforms genomic prediction from a purely computational exercise into a tool for biological discovery[7]. GBLUP also aligns with the practical realities of breeding programs: while AI models often require large training datasets and careful hyperparameter tuning to avoid overfitting, GBLUP delivers low overfitting risk, modest computational demands, and statistical robustness in high-dimensional settings[8–10]. These attributes have secured its central role in contemporary genetic evaluation systems. With the proliferation of AI-based methods, GBLUP has become a widely adopted benchmark for assessing methodological progress. Moreover, the conceptual issues it has raised, such as genotype-by-environment interactions, have actively shaped the development of AI breeding models toward more refined and interpretable directions[11].

The theoretical value of GBLUP has driven its expansion beyond prediction into broader analytical domains. A key early extension came from Wang et al.[12], who showed that GBLUP outputs could be transformed into GWAS results by deriving individual SNP effects from genomic estimated breeding values (GEBVs), effectively bridging genomic prediction and association mapping. This translational capacity has since been reinforced by a growing ecosystem of tools, including rrBLUP, ssBLUP, Bayesian models, and BGLR, that support single-trait breeding applications[13–16], alongside computational packages such as BLUPF90 and DMU that enable industrial-scale genetic evaluation[17]. Recognizing that complex traits are rarely governed by single or static traits, the field has progressively moved toward multi-trait frameworks. The HIBLUP model, for instance, accommodates multi-trait genomic selection alongside environmental and genetic interactions[17], while Jung et al. have advanced multi-environment prediction by explicitly modeling

https://doi.org/10.48130/stati-0026-0012

genotype-by-environment interactions and envirotyping effects[18]. The logical next step has been to embrace the temporal dimension of phenotypic expression. Hobby et al. integrated traditional genomic prediction with dynamic mode decomposition (DMD), overcoming the single-time-point limitation of conventional GS to enable dynamic prediction across developmental trajectories[19]. Collectively, these developments reflect a unifying trajectory: the progressive extension of BLUP models to handle increasingly complex phenotypic data across traits, environments, and time, with the ultimate goal of improving prediction accuracy and biological interpretability throughout entire growth cycles.

The continuous phenotypic variation observed over time represents a critical data source for dissecting the genetic architecture of complex traits. Developing GBLUP models for such data holds the potential to drive transformative advances in GWAS and GS in the era of high-dimensional markers and high-dimensional phenotypes[20,21]. Unlike single-trait analyses, dynamic traits involve temporal regulation of gene expression. Extending GBLUP to this data type would enable not only genomic prediction across developmental and environmental contexts but also the back-solving of genome-wide SNP effects, thereby integrating prediction with candidate marker discovery in a unified framework[22,23]. Although research in this area is still nascent, pioneering efforts have established important conceptual foundations. Functional mapping, introduced by Ma and colleagues, first demonstrated the value of fitting growth curves to temporal dynamics or allometric equations to environmental responses, offering breakthrough insights into the genetic basis of development[24–27]. Subsequent work by Hou et al. advanced this line of inquiry by proposing a nonlinear mixed model that separates random effects from residuals, yet it stopped short of incorporating a genome-wide relationship matrix, thereby limiting its capacity for unbiased estimation of marker effects or breeding values[28]. More recently, Ning et al. addressed this gap by integrating a genomic relationship matrix into their GMA framework, but they relied on a single non-parametric curve (Legendre orthogonal polynomials, LOPs) to capture phenotypic dynamics, leaving the approach ill-equipped to represent the diverse patterns exhibited by complex dynamic traits[29]. Collectively, these efforts underscore the need for a GBLUP-based solution that simultaneously models all SNP effects as random effects while accommodating flexible phenotypic trajectories.

To enable dynamic prediction across multiple time points and to facilitate rapid estimation of SNP effects from breeding values, this study integrates the biological principles underlying developmental dynamic gradients. We propose a dynamic GBLUP model (dynGBLUP), designed specifically for dynamic phenotypes. The model accounts for the confounding effects of covariates and relatedness on dynamic phenotypes within a mixed model framework. Building on this foundation, dynGBLUP incorporates growth curves derived from functional mapping theory, including logistic, Gompertz, and Richards[30–32]. By leveraging the dynamic patterns of phenotypic change over time, dynGBLUP enables systematic dissection of the genetic architecture and dynamic regulatory mechanisms underlying complex high-dimensional traits, thereby advancing breeding decisions from static to dynamic paradigms.

Model construction

Construction of the dynGBLUP model

Within the mixed model framework, dynGBLUP is formulated as:

$$\boldsymbol{y}_i = \boldsymbol{W}_i\boldsymbol{b} + \boldsymbol{Q}_i\sum_{j=1}^{k}\left(s_{ij}\otimes \mathrm{I}_{nr_1+1}\right)\boldsymbol{u}_j + \boldsymbol{Z}_i\boldsymbol{p}_i + \boldsymbol{e}_i \tag{1}$$

where, $\boldsymbol{y}_i$ represents the vector of phenotypic observations for individual i, measured across multiple time points or environments. $\boldsymbol{W}_i$, $\boldsymbol{Q}_i$, and $\boldsymbol{Z}_i$ are design matrices corresponding to fixed effects, SNP random effects, and general random effects, respectively. $\boldsymbol{b}$ is the vector of fixed regression coefficients, $\boldsymbol{u}_j$ denotes the random regression coefficient for SNP_j, $\boldsymbol{p}_i$ represents the coefficients of the general random regression polynomials, and $\boldsymbol{e}_i$ denotes the residual error. The genotype of individual i at SNP_j is indicated by s_{ij}. The orders of the LOPs are nr_1 for the SNP random effects and nr_2 for the general random effects. In Eq. (1), the second term represents the breeding value obtained by summing the effects of all k genome-wide markers, while the third term accounts for general random effects, such as permanent environmental effects. The choice of LOP order was informed by the work of Ning et al.[29], who developed a Q + K model for dynamic longitudinal phenotypes. Although that study used LOPs as the only curve fitting approach, their strategy for modeling random effects is directly relevant to ours. Following their successful implementation, we adopted an LOP order of four for random effects. A fourth-order LOP with five parameters provides sufficient flexibility to capture the diversity of random effect trajectories without over-parameterization.

For phenotypic data that follow specific growth patterns, such as logistic or Gompertz functions, these parametric growth curves can be used to model the fixed effects. For phenotypes whose growth trajectories are less well defined, nonparametric LOP curves are employed for fixed-effect modeling. In both cases, LOP curves are also used to model the random effects. The random effects are assumed to follow:

$$\boldsymbol{u} \sim N(0, \mathbf{I}\otimes\boldsymbol{\Sigma}_{\boldsymbol{u}}); \boldsymbol{p} \sim N\left(0, \mathbf{I}\otimes\boldsymbol{\Sigma}_{\boldsymbol{p}}\right); \boldsymbol{e} \sim N(0, \boldsymbol{R}) \tag{2}$$

where, $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{p}}$ denote the variance–covariance matrices of the SNP random regression coefficients and the general random regression coefficients, respectively. The residual variance-covariance structure $\boldsymbol{R}$ is modeled using a first-order structured antedependence model (SAD [1][33] with $\mathbf{I}$ denotes the identity matrix.

According to Wang et al.[30], Eq. (1) was linearized using a first-order Taylor expansion, yielding:

$$\tilde{\boldsymbol{y}}_i = \boldsymbol{X}_i\boldsymbol{b} + \boldsymbol{Q}_i\sum_{j=1}^{k}(s_{ij}\otimes\mathbf{I}_{nr_1+1})\boldsymbol{u}_j + \boldsymbol{Z}_i\boldsymbol{p}_i \tag{3}$$

where, $\tilde{\boldsymbol{y}}_i$ denotes the phenotypic values after transformation, $\boldsymbol{X}_i$, $\boldsymbol{Q}_i$, $\boldsymbol{Z}_i$ represent the matrix of first-order partial derivatives of the fixed effects with respect to each parameter of the target growth curve and the two types of random effect function.

Solution of dynGBLUP

Estimation of dynGBLUP relies on the mixed model equations (MME). For Eq. (3), the system is:

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Q} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Q}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Q}'\boldsymbol{R}^{-1}\boldsymbol{Q}+\boldsymbol{K}^{-1}\otimes\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} & \boldsymbol{Q}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Q} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}+I\otimes\boldsymbol{\Sigma}_{\boldsymbol{P}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{P}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\tilde{\boldsymbol{y}} \\ \boldsymbol{Q}'\boldsymbol{R}^{-1}\tilde{\boldsymbol{y}} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\tilde{\boldsymbol{y}} \end{bmatrix} \tag{4}$$

The polygenic random effects for individual i, denoted as $\boldsymbol{a}_i$ is considered as the whole of genome-wide marker effect at individual level, is expressed as:

$$\boldsymbol{a}_i = \sum_{j=1}^{k}\left(s_{ij}\otimes\mathbf{I}_{nr_1+1}\right)\boldsymbol{u}_j \tag{5}$$

Define $\boldsymbol{C}$ as the coefficient matrix on the left-hand side of Eq. (4), with its inverse given by:

$$\boldsymbol{C}^{-1} = \begin{bmatrix} \boldsymbol{C}^{\mathrm{XX}} & \boldsymbol{C}^{\mathrm{XQ}} & \boldsymbol{C}^{\mathrm{XZ}} \\ \boldsymbol{C}^{\mathrm{QX}} & \boldsymbol{C}^{\mathrm{QQ}} & \boldsymbol{C}^{\mathrm{QZ}} \\ \boldsymbol{C}^{\mathrm{ZX}} & \boldsymbol{C}^{\mathrm{ZQ}} & \boldsymbol{C}^{\mathrm{ZZ}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{bmatrix} \tag{6}$$

where $\begin{pmatrix} \boldsymbol{C}_{22}^{(1)} & \boldsymbol{C}_{22}^{(2)} \\ \boldsymbol{C}_{22}^{(3)} & \boldsymbol{C}_{22}^{(4)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{C}^{\mathrm{QQ}} & \boldsymbol{C}^{\mathrm{QZ}} \\ \boldsymbol{C}^{\mathrm{ZQ}} & \boldsymbol{C}^{\mathrm{ZZ}} \end{pmatrix}$.

For random regression coefficient for SNP$_j$, denoted $\hat{\boldsymbol{u}}_j$, and its variance are expressed as:

$$\hat{\boldsymbol{u}}_j = \left(\boldsymbol{s}_j' \boldsymbol{K}^{-1} \otimes \mathbf{I}\right)\left(\boldsymbol{C}_{21}^{(1)} \boldsymbol{X}' \boldsymbol{R}^{-1} \hat{\boldsymbol{y}} + \boldsymbol{C}_{22}^{(1)} \boldsymbol{Q}' \boldsymbol{R}^{-1} \hat{\boldsymbol{y}} + \boldsymbol{C}_{22}^{(2)} \boldsymbol{Z} \boldsymbol{R}^{-1} \hat{\boldsymbol{y}}\right) \tag{7}$$

$$\mathrm{var}\left(\hat{\boldsymbol{u}}_j\right) = \left(\boldsymbol{s}_j' \boldsymbol{K}^{-1} \otimes \mathbf{I}\right)\left(\boldsymbol{K} \otimes \boldsymbol{\Sigma}_a - \boldsymbol{C}_{22}^{(1)}\right)\left(\boldsymbol{s}_j' \boldsymbol{K}^{-1} \otimes \mathbf{I}\right)' \tag{8}$$

where, $\boldsymbol{\Sigma}_a$ denotes the variance–covariance matrix of the polygenic random effects, and $\boldsymbol{K}$ is the marker-derived relationship matrix constructed following the approach proposed by VanRaden[4]. Based on the solutions for fixed and random effects, with the latter denoting the solved or estimated marker effect, the phenotypic values at the individual level along a temporal sequence can be predicted using Eq. (1). For the statistical significance, the Wald χ^2 statistic for each SNP is constructed as follows:

$$\hat{\boldsymbol{u}}_j \left[\mathrm{var}\left(\hat{\boldsymbol{u}}_j\right)\right]^{-1} \hat{\boldsymbol{u}}_j \sim \chi^2 \left(nr_1 + 1\right) \tag{9}$$

p-values for each SNP are subsequently calculated based on the degrees of freedom (1 + nr_1) to identify significant SNPs. Following Verbyla[34], all derivations were performed in the framework of the restricted maximum-likelihood (REML). Variance components within the mixed model were estimated, following the iterative algorithm developed by Jensen et al.[35].

Model application

Simulation tests

Assessing the type I error rate is critical for evaluating model reliability. For the Q + K model, which accounts for population structure and kinship, realistic simulation of kinship and random effects presents a challenge. To address this, we conducted simulations based on the procedure described in Zhou & Stephens[36], with the goal of evaluating the model's ability to control false positives under varying sample sizes and heritability levels. Specifically, phenotypes were simulated using real genotype data from the empirical datasets, with quantitative trait loci (QTLs) randomly assigned to explain a proportion of the phenotypic variance. This approach enables a rigorous assessment of the model's type I error rate and statistical power across realistic genetic architectures.

Application to real data

To evaluate the predictive performance of dynGBLUP and its efficacy in QTL detection via back-solved SNP effects, we applied the model to empirical datasets. The dataset was derived from Wei[37] and consists of a natural population of *Populus szechuanica* var. *tibetica* sampled from high- and low-altitude sites in the Sejila Mountains of Tibet, comprising 200 genotypic lines. *Populus szechuanica* var. *tibetica* is one of the few tree species that survive on the Qinghai-Tibetan Plateau. Generally, species diversity declines sharply with increasing altitude, and trees rarely occur at very high elevations. The natural distribution of *Populus szechuanica* under harsh conditions strongly suggests the presence of genetic adaptations that enable survival in such extreme environments. Therefore, our sampling included both high- and low-altitude populations to capture this adaptive genetic variation. Given this design, it is essential to account for potential population stratification in downstream modeling. To prevent this stratification from confounding genetic parameter estimation, we explicitly incorporated the Q matrix as a fixed-effect covariate in the dynGBLUP model[38]. In 2014, cuttings from this population were propagated in a greenhouse at Beijing Forestry University, and stem diameter was measured throughout the growing season. SNP genotypes were obtained from whole-genome resequencing; after quality control, 24,949 high-quality SNPs were retained. This dataset enables assessment of dynGBLUP's ability to predict dynamic growth traits across multiple time points and to map QTLs underlying temporal trajectories.

To evaluate the predictive performance of dynGBLUP, we compared it with rrBLUP and three machine learning algorithms, i.e., random forest, elastic net, and support vector machine (SVM), following a unified analytical pipeline, including phenotypic dimensionality reduction, coefficient prediction, and curve reconstruction. Specifically, a fourth-order LOP was fitted for each individual according to their growth trajectories. SNPs were then retained as predictors, and separate regression models were trained to predict each coefficient. The predicted dynamic phenotypes were subsequently recovered using the LOP functions. All models were assessed via 10 repeats of 10-fold cross-validation. Pearson correlation coefficient was performed between observed and predicted values.

Results

We evaluated the predictive performance of the dynGBLUP model by performing 10 repetitions of 10-fold cross-validation, using stem diameter data of *Populus szechuanica* var. *tibetica*. The model consistently achieved strong predictive accuracy across the dataset. Figure 1a shows the dynamic changes in stem diameter of cuttings from high- and low-altitude populations of *Populus szechuanica* var. *tibetica* throughout the growing season. Phenotypic predictions from dynGBLUP yielded an average Pearson correlation of 0.85 between predicted and observed values, outperforming the 0.72 achieved by rrBLUP. We additionally tested three machine learning methods: random forest, elastic net, and SVM, which produced predictive correlations of 0.71, 0.73, and 0.80, respectively. Figure 1b–f further reveals that dynGBLUP better preserved the dynamic characteristics of the Gompertz growth curve, with predicted phenotypic variance across 10 time points ranging from 0.49 to 0.55. In contrast, while rrBLUP captured the overall growth trend, its predictions were largely confined to the mean trajectories of the two subpopulations, with predicted phenotypic variance reaching only 0.04–0.14. The predicted phenotypic variance was 0.027–0.03 for random forest, 0.02–0.03 for elastic net, and 0.28–0.45 for SVM, all of which were lower than that of dynGBLUP and thus failed to capture the full phenotypic diversity of the population.

To compare model performance in capturing temporal dynamics, we used the Pearson correlation between observed and predicted values as the evaluation metric. Figure 2a shows that dynGBLUP exhibited consistently strong predictive performance across the dataset. For the dynamic growth of *Populus szechuanica* var. *tibetica* cuttings, the prediction correlations of dynGBLUP across 10 time points remained around 0.75, with small standard errors ranging from 0.0272 to 0.0308 (Fig. 2a). In contrast, the prediction correlation of SVM was around 0.64, while the correlations achieved by

https://doi.org/10.48130/stati-0026-0012

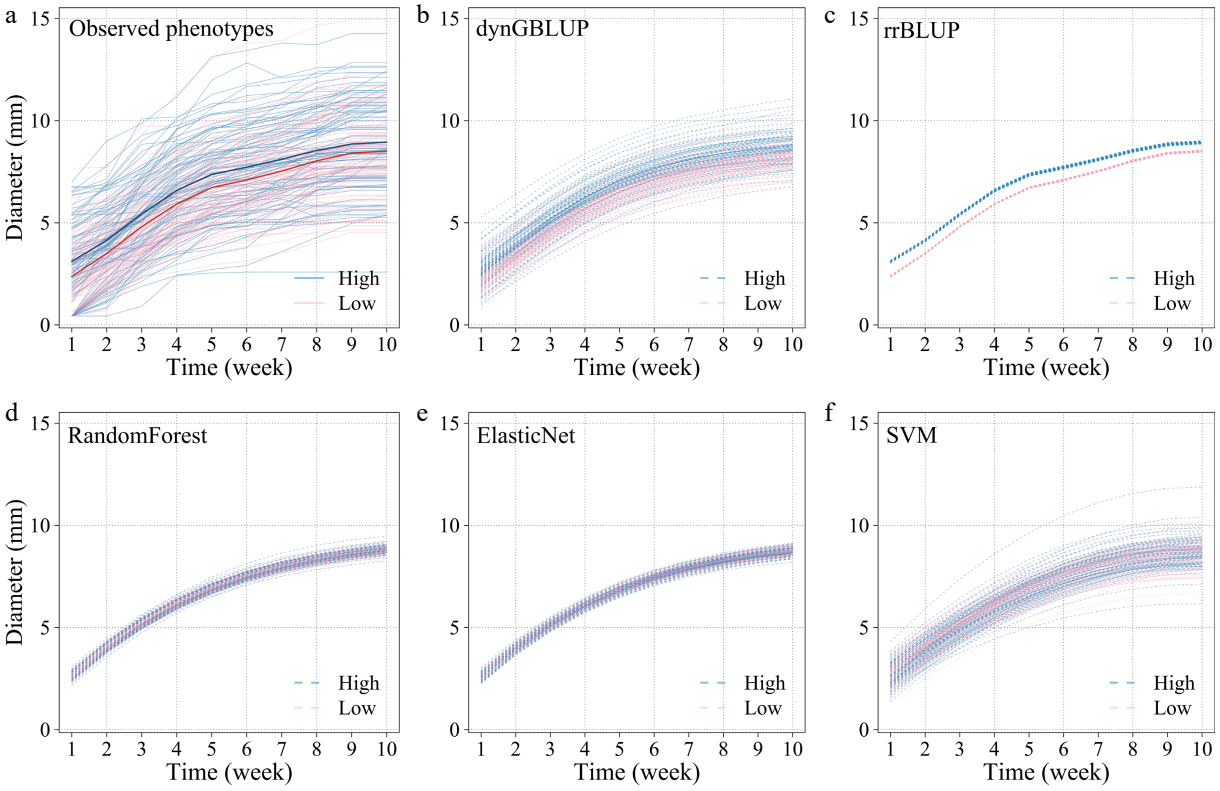

Fig. 1 (a) Observed stem diameter phenotypes of *Populus szechuanica* var. *tibetica* across 10 time points. Average predicted values from 10 repetitions of 10-fold cross-validation using (b) dynGBLUP, (c) rrBLUP, (d) random forest, (e) elastic net, and (f) SVM, for stem diameter data.

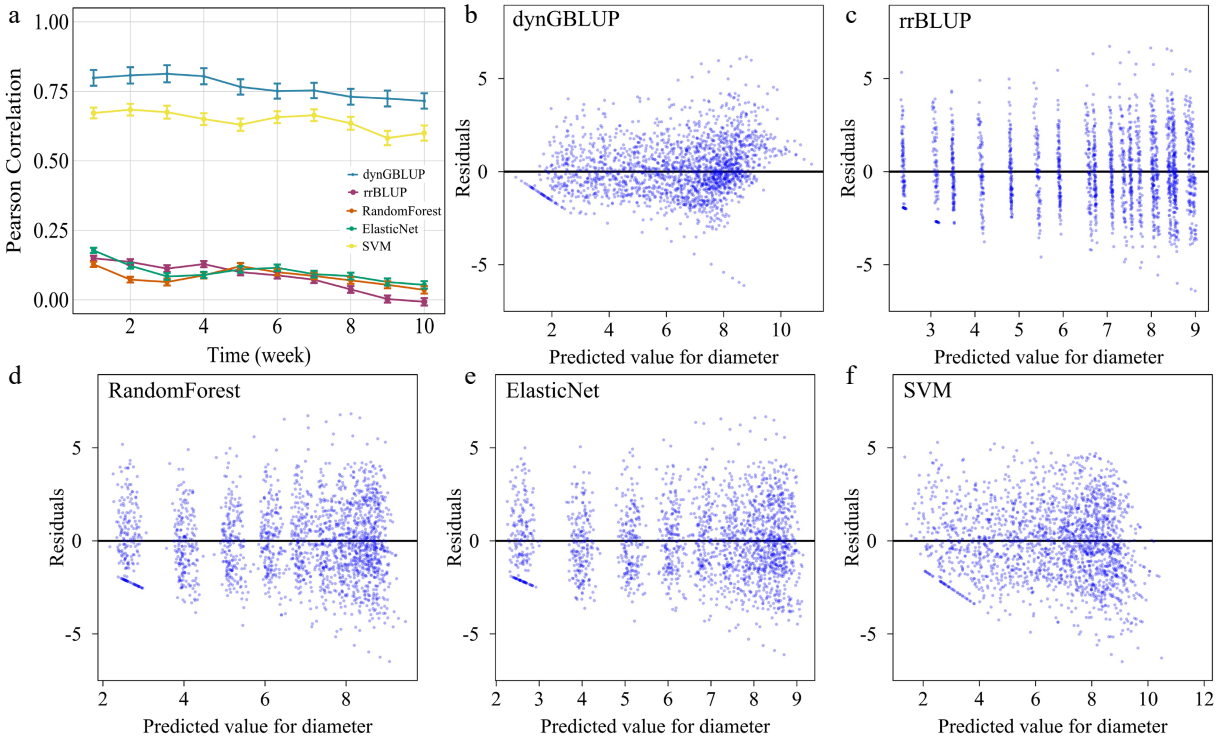

Fig. 2 (a) Comparison of the predictive performance of dynGBLUP and rrBLUP models across individual time points. Predicted residuals from (b) dynGBLUP, (c) rrBLUP, (d) random forest, (e) elastic net, and (f) SVM, for stem diameter growth in *Populus szechuanica* var. *tibetica* cuttings.

rrBLUP, random forest, and elastic net consistently remained below 0.25. The residual distributions of different models for the stem diameter data of *Populus szechuanica* var. *tibetica* are shown in Fig. 2b–f. The residuals of dynGBLUP were approximately randomly and uniformly distributed around zero across the range of predicted values, with small and stable fluctuations and no obvious systematic bias (Fig. 2b). In contrast, the predicted values of rrBLUP were concentrated with small variation, causing the residuals to cluster

vertically at a limited number of predicted values and form clear vertical bands, with the residuals tending to increase with higher predicted values (Fig. 2c). The predicted phenotypes of random forest and elastic net were similarly concentrated, and their residual plots also exhibited distinct vertical banding at early growth stages (Fig. 2d, e), indicating that these three models likewise failed to adequately capture the true phenotypic variation in the population. The residuals of SVM showed a relatively random distribution, similar to that of dynGBLUP (Fig. 2f).

By estimating genome-wide marker effects and providing a framework for evaluating SNP significance via Eq. (9), dynGBLUP inherently supports GWAS. To evaluate its performance, we employed a simulation strategy in which the phenotype-genotype correspondence was randomly shuffled to reconstruct phenotypic datasets. Figure 3a shows that when heritability was set to zero, the type I error rate was controlled between 9.40% and 16.91%. Figure 3b indicates that detection power strongly depends on heritability, increasing substantially as heritability rises. When the sample size exceeded 800, the gain in power from additional samples was smaller than the increase attributable to higher heritability. Under a sample size of 1,000 and a heritability of 3%, the model achieved a detection power of 77%. These results demonstrate that dynGBLUP provides high power for GWAS when applied to multi-time phenotypic data.

To further evaluate the GWAS utility of dynGBLUP, we computed SNP *p*-values from the genetic effects estimated by the model using empirical data. The stem diameter dataset yielded robust association mapping results. The Manhattan plot showed that significant SNPs were widely distributed across the genome (Fig. 4a). Based on functional annotation, *THESEUS1* encodes a brassinosteroid-regulated receptor kinase known to regulate overall plant growth by mediating cell elongation during the vegetative growth stage; LOC7490611 with its homology of *At5g40410* in *Arabidopsis* is highly expressed during young leaf growth, suggesting it may play a positive regulatory role in leaf development and overall plant growth; *SRG1* responds to intracellular sugar content at the transcriptional level[39]. Because sugar signaling plays a central role in regulating plant growth, energy metabolism, and senescence, this gene may influence overall plant growth by mediating sugar homeostasis. The QQ plot indicated that 99.4% of SNPs conformed to the expected distribution, with upward deviation observed only in the extreme tail corresponding to a *p*-value threshold of 0.0038 (Fig. 4b). Collectively, these results demonstrate that estimating SNP effects within the dynGBLUP framework provides a unified approach for simultaneous GWAS of dynamic traits.

Discussion

As a core method in genomic selection and prediction, GBLUP has driven a paradigm shift in modern breeding, transitioning from posterior-based phenotypic selection to prior-based genomic prediction[40,41]. This transition has overcome key limitations in leveraging molecular marker data to estimate GEBVs, enabling breeding decisions to be made at early stages with greater precision and efficiency. With the rise of intelligent breeding, there is growing recognition that developmental stages critically influence key phenotypic traits in both plants and animals, spurring the development of dynamic genomic prediction[42,43]. To address this challenge, we propose the dynGBLUP model, which predicts dynamic phenotypes at specific developmental stages while simultaneously estimating genome-wide marker effects. This framework offers a novel strategy for predicting dynamic traits across multiple time points using genome-wide markers, enabling breeders to select for target traits while optimizing growth trajectories encoded in the underlying genetic architecture.

By integrating time-series, dynGBLUP incorporates the developmental dynamics of biological traits through a range of growth curves. Capturing the dynamic genetic regulatory mechanisms underlying growth and development, this framework enables prediction of phenotypes at specific developmental stages. Simultaneously, its statistical structure allows estimation of genetic effects for individual SNPs within a 'one-step GWAS' framework. When estimating SNP effects, researchers often encounter the 'curse of dimensionality', in which the number of SNPs (*p*) greatly exceeds the sample size (*n*). To mitigate the risks of overfitting and overparameterization arising from a rank-deficient design matrix in least-squares methods, dynGBLUP employs the BLUP estimation approach. In Eq. (1), the random effects component encompasses genome-wide SNPs. When translating the output into GWAS results, if we define the polygenic effect as the sum of individual marker effects, it can be equivalently expressed in a form that satisfies the Q + K model commonly used in GWAS. Taken together, dynGBLUP provides a robust bridge between GS and GWAS for traits exhibiting temporal dynamics, highlighting the consistency of statistical frameworks for analyzing dynamic phenotypes within both GBLUP and GWAS.

Given the critical role of dynamic traits in elucidating the genetic architecture of complex biological traits, Ning et al. developed a GMA-based association analysis algorithm for multi-time-point phenotypes[29]. However, their approach relies solely on non-parametric LOP curves, limiting its ability to accurately capture diverse

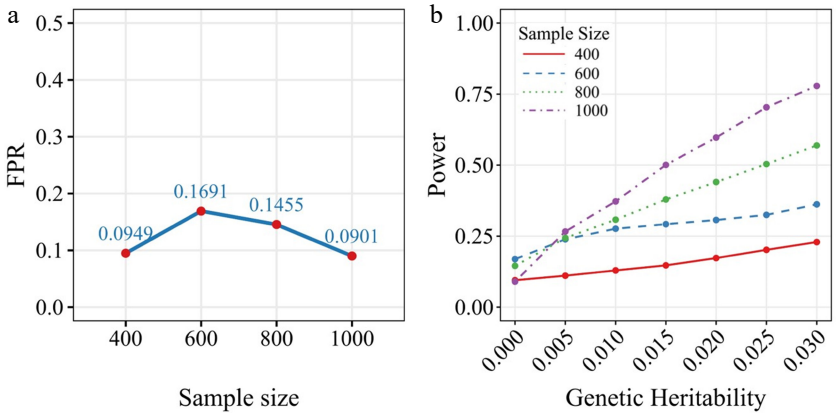

Fig. 3 (a) False positive rate and (b) statistical power of the dynGBLUP model.

https://doi.org/10.48130/stati-0026-0012

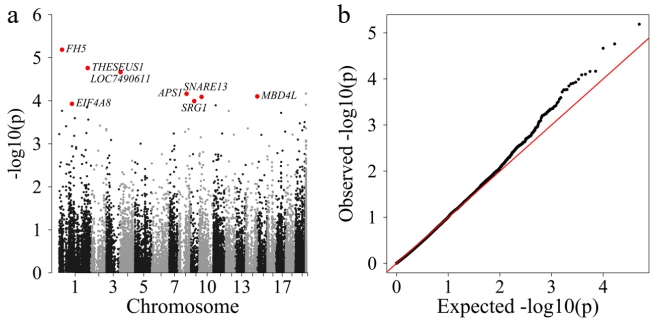

Fig. 4 (a) Manhattan and (b) QQ plots for stem diameter growth in *Populus szechuanica* var. *tibetica* cuttings.

biological growth patterns and developmental trajectories. In contrast, the present study builds upon the theoretical framework of functional mapping and integrates multiple types of growth curves to systematically characterize dynamic and quasi-dynamic patterns of phenotypic variation over time or along environmental gradients. We developed a GBLUP framework tailored for dynamic phenotypes by employing SAD(1) to model the variance–covariance structure of dynamic phenotypic residuals. The successful use of Gompertz equations enables more biologically meaningful fits for fixed effects and more accurately reflects the principles of development. This approach overcomes the limitations of traditional nonparametric basis functions in capturing complex developmental traits, thereby enhancing analytical flexibility for intricate dynamic phenotypes. The assumption of a shared baseline growth curve is indeed a common and biologically starting point for many dynamic models. For a given trait, the underlying growth pattern (e.g., a logistic or Gompertz curve) is generally similar across individuals, which is consistent with basic biological principles. Individual differences manifest as deviations from this common baseline, and our model explicitly captures such heterogeneity through multiple components: (1) random regression coefficients that allow individual-specific curve adjustments; (2) the Q matrix (population structure) as fixed effects, which accounts for systematic differences between subpopulations; and (3) heterogeneous residual variances. This modeling strategy is conceptually analogous to how a standard regression framework handles the overall mean μ, a shared baseline (akin to μ) plus systematic differences captured by fixed effects (such as the Q matrix for population stratification) and additional individual-specific random deviations. Therefore, we believe the assumption is reasonable and sufficiently flexible for the data at hand. The dynGBLUP model was validated using the real dataset and computer simulations, with comparisons to rrBLUP and multiple machine learning methods. It consistently achieved high Pearson correlation coefficients in genomic phenotype prediction across multiple time points, demonstrating both feasibility and accuracy for complex dynamic traits.

Notably, Hobby et al. combined DMD with genomic prediction to construct the dynamicGP model, which is capable of predicting the dynamic growth of multiple traits[19]. However, dynamicGP relies on the preprocessing of high-dimensional time-series phenotypic data, and its performance depends on the number of input traits and time points, making it best suited for physiologically related traits. In contrast, dynGBLUP directly targets key agronomic traits, such as plant height and stem diameter, without relying on dynamic mathematical decomposition or additional traits, serving as a complementary rather than competitive approach to existing genomic prediction algorithms. A key assumption underlying dynGBLUP is that SNP effects follow a homoscedastic random-effects distribution, consistent with the infinitesimal model in which each SNP contributes a small effect. While this assumption provides a computationally tractable and interpretable framework, future extensions could incorporate Bayesian sparse models to enhance the accuracy of effect estimation for loci with larger phenotypic contributions. For multi-environment phenotypic data, the current use of population means as environmental indices could be refined by integrating more biologically meaningful environmental covariates, as exemplified by recent advances in envirotyping[44]. Collectively, these extensions point toward a broader vision: transitioning genomic selection from predicting static, singular genetic potential toward a framework that embraces dynamic, plastic, and stable selection.

Author contributions

The authors confirm their contributions to the paper as follows: study conception and design: Yang X, Ye M; data collection: Yang X, Zhao R, Li X; analysis and interpretation of results: Yang X, manuscript preparation: Yang X, Ye M; All authors reviewed the results and approved the final version of the manuscript.

Data availability

The *Populus szechuanica* var. *tibetica* stem diameter data and the model code generated during this study are both available on GitHub (https://github.com/xsxsxaa1/DynGBLUP-).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 32071796). The authors sincerely thank the editor and the referees for their constructive comments and helpful suggestions.

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 26 March 2026; Revised 17 May 2026; Accepted 29 May 2026; Published online 31 July 2026

References

[1] Henderson CR. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447

[2] Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829

[3] Meuwissen T, Boerner V. 2025. Multitrait genome-wide association best linear unbiased prediction of genetic values. *Genetics Selection Evolution* 57:15

[4] VanRaden PM. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91:4414–4423

[5] Li H, Su G, Jiang L, Bao Z. 2017. An efficient unified model for genome-wide association studies and genomic selection. *Genetics Selection Evolution* 49:64

[6] van Hilten A, Katz S, Saccenti E, Niessen WJ, Roshchupkin GV. 2024. Designing interpretable deep learning applications for functional genomics: a quantitative analysis. *Briefings in Bioinformatics* 25:bbae449

[7] Habier D, Fernando RL, Garrick DJ. 2013. Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics* 194:597–607

[8] Sendrowski J, Bataillon T, Ramstein GP. 2025. *In silico* prediction of variant effects: promises and limitations for precision plant breeding. *Theoretical and Applied Genetics* 138:193

[9] Montesinos-López OA, Montesinos-López A, Hernandez-Suarez CM, Barrón-López JA, Crossa J. 2021. Deep-learning power and perspectives for genomic selection. *The Plant Genome* 14:e20122

[10] Wang X, Zhang Z, Du H, Pfeiffer C, Mészáros G, et al. 2024. Predictive ability of multi-population genomic prediction methods of phenotypes for reproduction traits in Chinese and Austrian pigs. *Genetics Selection Evolution* 56:49

[11] Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, et al. 2017. Genomic selection in plant breeding: methods, models, and perspectives. *Trends in Plant Science* 22:961–975

[12] Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research* 94:73–83

[13] Endelman JB. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* 4:250–255

[14] Pérez P, de los Campos G. 2014. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483–495

[15] Teissier M, Larroque H, Robert-Granié C. 2018. Weighted single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: a quantitative trait influenced by a major gene. *Genetics Selection Evolution* 50:31

[16] Crossa J, Martini JWR, Vitale P, Pérez-Rodríguez P, Costa-Neto G, et al. 2025. Expanding genomic prediction in plant breeding: harnessing big data, machine learning, and advanced software. *Trends in Plant Science* 30:756–774

[17] Yin L, Zhang H, Tang Z, Yin D, Fu Y, et al. 2023. HIBLUP: an integration of statistical models on the BLUP framework for efficient genetic evaluation using big genomic data. *Nucleic Acids Research* 51:3501–3512

[18] Jung M, Quesada-Traver C, Roth M, Aranzana MJ, Muranty H, et al. 2025. Integrative multi-environmental genomic prediction in apple. *Horticulture Research* 12:uhae319

[19] Hobby D, Tong H, Heuermann M, Mbebi AJ, Laitinen RAE, et al. 2025. Predicting plant trait dynamics from genetic markers. *Nature Plants* 11:1018–1027

[20] Bouidghaghen J, Moreau L, Beauchêne K, Chapuis R, Mangel N, et al. 2023. Robotized indoor phenotyping allows genomic prediction of adaptive traits in the field. *Nature Communications* 14:6603

[21] Wu X, Feng H, Wu D, Yan S, Zhang P, et al. 2021. Using high-throughput multiple optical phenotyping to decipher the genetic architecture of maize drought tolerance. *Genome Biology* 22:185

[22] Osterman J, Gutiérrez L, Öhlund L, Ortiz R, Hammenhag C, et al. 2024. Comparison of single-trait and multi-trait GBLUP models for genomic prediction in red clover. *Agronomy* 14:2445

[23] Yang S, Moshelion M, Chen H, Xu P. 2026. From phenomics to post-phenomics: multidisciplinary integration driving autonomous agricultural systems. *Plant Communications* 7:101532

[24] Ma CX, Casella G, Wu R. 2002. Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 161:1751–1762

[25] Sun L, Wu R. 2015. Mapping complex traits as a dynamic system. *Physics of Life Reviews* 13:155–185

[26] Wu R, Lin M. 2006. Functional mapping – how to map and study the genetic architecture of dynamic complex traits. *Nature Reviews Genetics* 7:229–237

[27] Li Z, Sillanpää MJ. 2015. Dynamic quantitative trait locus analysis of plant phenomic data. *Trends in Plant Science* 20:822–833

[28] Hou W, Li H, Zhang B, Huang M, Wu R. 2008. A nonlinear mixed-effect mixture model for functional mapping of dynamic traits. *Heredity* 101:321–328

[29] Ning C, Wang D, Zhou L, Wei J, Liu Y, et al. 2019. Efficient multivariate analysis algorithms for longitudinal genome-wide association studies. *Bioinformatics* 35:4879–4885

[30] Wang ZJ, Ye MX, Zhang H, Wu RL. 2024. 功能作图的混合效应模型开发 [Mixed-effect model development for functional mapping]. *北京林业大学学报 [Journal of Beijing Forestry University]* 46:163–172 (in Chinese)

[31] Tsoularis A, Wallace J. 2002. Analysis of logistic growth models. *Mathematical Biosciences* 179:21–55

[32] Richards FJ. 1959. A flexible growth function for empirical use. *Journal of Experimental Botany* 10:290–301

[33] Zhao W, Hou W, Littell RC, Wu R. 2005. Structured antedependence models for functional mapping of multiple longitudinal traits. *Statistical Applications in Genetics and Molecular Biology* 4:Article33

[34] Verbyla AP. 1990. A conditional derivation of residual maximum likelihood. *Australian Journal of Statistics* 32:227–230

[35] Jensen J, Mäntysaari EA, Madsen P, Thompson R. 1997. Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information. *Journal of the Indian Society of Agricultural Statistics* 49:215–236

[36] Zhou X, Stephens M. 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods* 11:407–409

[37] Wei XD. 2023. 基于异速生长的系统作图解析藏川杨生长的遗传调控机制 *[Analysis of genetic regulation mechanism of growth of Populus szechuanica var. tibetica based on allometric system mapping]*. Thesis. Beijing Forestry University, Beijing, China. doi: 10.26949/d.cnki.gblyu.2023.000697 (in Chinese)

[38] Ye M, Zhu X, Gao P, Jiang L, Wu R, et al. 2020. Identification of quantitative trait loci for altitude adaptation of tree leaf shape with *Populus szechuanica* in the Qinghai-Tibetan Plateau. *Frontiers in Plant Science* 11:632

[39] Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA, et al. 2016. A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant Journal* 88:1058–1070

[40] Alemu A, Åstrand J, Montesinos-López OA, Isidro y Sánchez J, Fernández-Gónzalez J, et al. 2024. Genomic selection in plant breeding: Key factors shaping two decades of progress. *Molecular Plant* 17:552–578

[41] Raimondi D, Verplaetse N, Passemiers A, Jans DS, Cleynen I, et al. 2025. Genomic prediction with kinship-based multiple kernel learning produces hypothesis on the underlying inheritance mechanisms of phenotypic traits. *Genome Biology* 26:84

[42] Toda Y, Sasaki G, Ohmori Y, Yamasaki Y, Takahashi H, et al. 2024. Reaction norm for genomic prediction of plant growth: modeling drought stress response in soybean. *Theoretical and Applied Genetics* 137:77

[43] Wang J, Liu L, He K, Gebrewahid TW, Gao S, et al. 2025. Accurate genomic prediction for grain yield and grain moisture content of maize hybrids using multi-environment data. *Journal of Integrative Plant Biology* 67:1379–1394

[44] Li X, Guo T, Wang J, Bekele WA, Sukumaran S, et al. 2021. An integrated framework reinstating the environmental dimension for GWAS and genomic selection in crops. *Molecular Plant* 14:874–887