Circos Plots for Genome Level Interpretation of Genomic Prediction Models

Shunichiro Tomura[1,2,*], Owen Powell[1,2], Melanie J. Wilkinson[1,2,3], Mark Cooper[1,2]

[1]Queensland Alliance for Agriculture and Food Innovation (QAAFI), Centre for Crop Science, the University of Queensland, St Lucia, QLD 4072, Australia

[2]ARC Centre of Excellence for Plant Success in Nature and Agriculture, the University of Queensland, St Lucia, QLD 4072, Australia

[3]ARC Training Centre in Predictive Breeding, the University of Queensland, St Lucia, QLD 4072, Australia

[*]Corresponding author: s.tomura@uq.edu.au

Abstract

Investigations of genomic prediction models have focused primarily on the level of metrics summarising the accuracy and error of predicted phenotypes. Analysis of the genomic regions contributing to the genomic prediction models can deepen our understanding of the trait genetic architecture. Here, we develop circos plots to visualise how different genomic prediction models quantify marker effects. The genetic architecture was inferred using interpretable genomic prediction models to characterise each method at the genome level. Application is demonstrated for the trait days to anthesis (DTA) in the TeoNAM dataset. The results indicate that genomic prediction models can capture different views of trait genetic architecture, even when their overall profiles of prediction accuracy are similar. Combinations of diverse views of the genetic architecture for the DTA trait in the TeoNAM study might explain the improved prediction performance achieved by ensembles of prediction models, aligned with the implication of the Diversity Prediction Theorem. In addition to identifying well-known genomic regions contributing to the genetic architecture of DTA in maize, the ensemble of genomic prediction models highlighted several new genomic regions that have not been previously identified for DTA. Finally, different views of trait genetic architecture were observed across populations, within the TeoNAM design, highlighting challenges for between-population genomic prediction. A deeper understanding of genomic prediction models with enhanced interpretability can reveal several critical findings at the genome level from the inferred genetic architecture, providing insights into the improvement of genomic prediction for crop breeding programs.

Introduction

Research into genomic selection (Meuwissen et al., 2001) for crop breeding has primarily focused on the accuracy of genomic prediction models (Crossa et al., 2017; Alemu et al., 2024; Kumar et al., 2024). Beyond the standard genomic best linear unbiased predictor (GBLUP, Meuwissen et al., 2001) that only considers additive genomic marker effects explicitly, the improvement of the prediction accuracy has motivated the development of a range of genomic prediction models that can accurately capture both additive and non-additive (epistatic) genomic marker effects (Bose et al., 2024). Capturing both additive and non-additive effects provides us with a deeper understanding of the genetic architecture of the target traits, enabling more precise predictions and uncovering complex genetic interactions that contribute to trait variation (Cooper et al., 2009; Messina et al., 2018; Yadav et al., 2021; Crossa et al., 2025; Tomura et al., 2025).

Genomic prediction models have been evaluated under a diverse range of prediction scenarios (Heslot et al., 2014; Wang et al., 2015; Montesinos-Lopez et al., 2024). However, investigations of genomic prediction models have mainly focused on the outcomes at the trait level through breeding values. Prediction performance comparison using metrics, such as prediction accuracy and error, evaluates only one facet of the genomic prediction models. The implications of the different genomic prediction models remain unclear at the genomic level. The investigation of genomic prediction models at the genome level may deepen the understanding of the trait genetic architecture, which may reveal potential new gene regions that have not been identified previously. The comparison of the quantified genomic markers across the genomic prediction models may show the diversity in the effect sizes of each marker due to the properties of each distinctive algorithm. Differing effect sizes among genomic prediction models indicate that each model captures a unique view of the underlying trait genetic architecture. A collection of such diverse genetic architectures might explain the prediction performance improvement achieved by the ensemble of individual models (Messina et al., 2025; Tomura et al., 2025). This possibility aligns with the implication of the Diversity Prediction Theorem (Page, 2018) at the genome level, stating that the prediction error of the ensemble becomes lower than the average error of the individual prediction models, provided that the predicted values achieved by the individual prediction models are diverse. Hence, we propose that the interpretation of estimated genomic marker effects from multiple prediction models can be another critical concept in the research of genomic prediction.

Interpretability seeks explanation of the output values from prediction models by estimating the effect of each feature and the interactions among the features (Molnar, 2020). Interpretable prediction models have been leveraged in a wide range of research fields, such as healthcare (Stiglic et al., 2020; Abdullah et al., 2021; Baniecki et al., 2025), chemistry (Oviedo et al., 2022; Tavakoli et al., 2023) and physics (Wetzel et al., 2025). In crop breeding, conventional genomic prediction models can reveal the effect of each feature as the additive effect of each genomic marker and the interaction effects as non-additive effects towards target phenotypes. However, the interpretability of genomic marker effects from machine learning algorithms has not been well-investigated in crop breeding. The combination of estimated marker effects from conventional and machine learning genomic prediction models can deepen the investigation of the interpretability of genomic prediction models at the genomic level and improve the characterisation of trait genetic architectures.

To compare the estimated genomic marker effects of genomic prediction models, we constructed a new graphical view using circos plots (Krzywinski et al., 2009). Circos plots can be applied to visualise trait genetic architecture by mapping the estimated marker effects to the corresponding genomic regions and highlighting their interactions (Tessele et al., 2025). For example, the circos plot constructed by Wei et al. (2024) visualised the chromosome locations of key genes for 16 traits in rice, with links representing epistatic interactions between genomic marker regions. The circos plot clearly illustrated the distribution of the key genes throughout the genome and their interactions with other genomic marker regions.

Here, we develop circos plots to analyse the genetic architectures identified by interpretable genomic prediction models. We use the Teosinte Nested Association Mapping (TeoNAM, Chen et al., 2019) experimental results from Tomura et al. (2025) to extend the interpretations to the genomic level. Our development of circos plots was guided by the pursuit of three objectives: (1) Comparison of the genetic architectures with key genomic marker regions identified by previous studies to assess the overlap with the known genetic architecture. (2) Comparison between the genetic architectures of the different genomic prediction models to investigate how genomic marker effects were quantified differently. (3) Comparison of the circos plots for each of the five TeoNAM populations to understand the characteristics of the genetic architecture revealed for each population. Together, we anticipate that these objectives will enhance the interpretability of genomic prediction models at the genome level.

Materials and Methods

1. Circos plot construction

The circos plots were developed using pyCirclize (v1.7.1) in Python (v3.11.10). The results from the TeoNAM experiment were used as an illustrative example (Chen et al., 2019; Tomura et al., 2025). The circos plots constructed in this study consist of three major sections: (1) key genomic marker region rings representing prior knowledge from previous mapping studies, (2) genomic marker effect rings from genomic prediction

models, and (3) genomic marker-by-marker interaction effect links. As with previous applications, each ring consists of segments for the chromosomes of the target species. For the maize TeoNAM example, ten segments were developed to represent the total number of chromosomes in maize. The segments were displayed in centimorgans (cM).

Prior knowledge: Three ring components were constructed using information from two previous studies. One ring represented QTL regions identified by Chen et al. (2019). The QTL regions relevant to DTA were inferred and mapped to their corresponding genomic marker regions on the ring. The second and third rings indicated the genomic regions of key genes that operate in the leaf and shoot apical meristem (SAM) tissues to determine the flowering time in maize, respectively (Dong et al., 2012).

Genomic Marker effects: Seven ring components were developed to illustrate the genetic architecture of each genomic prediction model. The inner six rings represented the estimated genomic marker effects from the individual genomic prediction models mapped to the corresponding genome regions. The outermost ring displayed the estimated genomic marker effects for the naïve ensemble-average model (Tomura et al., 2025). The effect sizes of the genomic markers were ranked based on the strength of the corresponding genomic marker effects and split into ten quantiles. The colour intensity indicated marker strength, where darker colours represented larger marker effects. The displayed marker regions on the circos plots were extended by adding 0.2 cM to each side of the actual genomic marker regions for visualisation.

Genomic Marker-by-Marker Interactions: The genomic marker pairs with the highest pairwise RF interactions (top 0.01% of Shapley values) were selected and displayed as lines connecting the markers. The strength of the genomic marker interactions was indicated using the width of the corresponding links; thicker links were assigned to stronger genomic marker interactions. The genomic marker interactions were classified into two categories; genomic marker interactions within the same chromosome and between different chromosomes.

Below, we describe the TeoNAM dataset (Chen et al., 2019) and the methods used to estimate the genomic prediction model marker effects and identify the genomic marker-by-marker interactions.

2. Dataset

The TeoNAM dataset (Chen et al., 2019) contains the genotype and phenotype records of recombinant inbred lines (RILs) from the five hybrid populations between the W22 maize (*Zea mays*) line and five teosinte types (*Z. mays ssp. parviglumis* for TIL01, TIL03, TIL11 and TIL14 and *Z.mays ssp. Mexicana* for *TIL25*). In each RIL population, the F1 generation was developed by crossing W22 and the teosinte type, backcrossed with W22 in the subsequent generation. Selfing process followed afterwards in the next four generations. Each RIL population was evaluated twice at the University of Wisconsin West Madison Agricultural Research Station in a randomised complete block. The W22TIL01, W22TIL03 and W22TIL11 populations were evaluated in the summer of 2015 and 2016. The W22TIL14 population was evaluated in the summer of 2016 and 2017. The W22TIL25 population was evaluated in two blocks in the summer of 2017. The total number of genomic markers (SNPs) and RILs from each of the five RIL populations is summarised in Table 1.

The diverse crosses with Teosinte motivated the selection of the TeoNAM dataset in this study to demonstrate the application of the circos plot. The crosses with an ancestral species of maize can diversify the fixed gene networks for traits in the W22 domesticated maize line. The increased genetic diversity within the crosses may help the genomic prediction models identify key genomic marker regions and pathways.

The days to anthesis (DTA) trait was selected in this study primarily due to previous studies revealing key genomic regions contributing to DTA (Buckler et al. 2009; Dong et al., 2012; Chen et al., 2019; Wisser et al., 2019). This prior knowledge can be used to assess whether the genomic prediction models used in this study can correctly detect known causative loci expected to segregate in the populations. A deeper understanding of the genetic architecture of DTA using the interpretable genomic prediction models can help improve genomic markers for DTA, ultimately benefiting prediction models used for crop yield improvement.

Each RIL contains two phenotype scores for DTA, since each population was evaluated twice, and hence the TeoNAM dataset consists of two sub-datasets in each population record. Two sub-datasets were concatenated into one within each population record, with a factor explicitly representing different evaluations with the value of 0 or 1.

Table 1: The total number of genomic markers (SNPs) before and after linkage disequilibrium (LD) filtering and the total number of recombinant inbred lines (RILs) in each population based on Tomura et al. (2025).

Population name	SNPs			RIL number
	Original	Imputation	LD filtering	
W22TIL01	13,089	13,042	274 – 322	222
W22TIL03	16,110	16,076	295 – 342	270
W22TIL11	13,188	13,153	268 – 314	219
W22TIL14	11,396	11,375	270 – 320	230
W22TIL25	14,885	14,857	294 - 341	308

3. Data preprocessing

Three data preprocessing methods were applied to the dataset prior to training the genomic prediction models in Tomura et al. (2025): missing allele imputation for genomic markers, missing trait phenotype RIL removal and the reduction of genomic markers based on linkage disequilibrium relationships. The missing alleles were imputed using the most frequent allele of the population at that site after excluding genomic markers containing more than 10% of missing marker calls. All the RILs with a missing phenotype value were excluded from the dataset. Linkage disequilibrium (LD) relationships between markers were leveraged to mitigate the effect of the curse of dimensionality (Bellman, 1957; Ramstein et al., 2019) by reducing the total number of genomic markers. LD filtering was performed using PLINK (v1.9) (Chang et al., 2015) with a window size of 30,000, a step size of 5 and a squared correlation (r^2) of greater than 0.8.

4. Genomic prediction models

Three classical genomic prediction models (ridge regression best linear unbiased prediction (rrBLUP), BayesB and reproducing kernel Hilbert space (RKHS)) and three machine learning models (random forest (RF), support vector regression (SVR) and graph attention network (GAT)) were applied as the individual genomic prediction models in Tomura et al. (2025).

For the classical genomic prediction models, the base linear mixed model can be defined as follows (Pérez and de Los Campos, 2014):

$$\boldsymbol{\eta} = 1\boldsymbol{\mu} + \sum_{j=1}^{J} \boldsymbol{X}_j \boldsymbol{\beta}_j + \sum_{l=1}^{L} \boldsymbol{u}_l \tag{1}$$

where $\boldsymbol{\eta} = \{\eta_1, \eta_1, \ldots, \eta_n\}$ is the set of predicted phenotypes corresponding to the observed phenotypes $\boldsymbol{y} = \{y_1, y_1, \ldots, y_n\}$, $\boldsymbol{\mu}$ is the intercept, $\boldsymbol{X}_j$ are the design matrices of genomic marker values for the genomic markers J, $\boldsymbol{\beta}_j$ are the allele substitution effect matrices of the genomic markers in $\boldsymbol{X}_j$ and $\boldsymbol{u}_l = \{u_{l1}, u_{l2}, \ldots, u_{ln}\}$ are the vector of random effects. In rrBLUP (Meuwissen et al., 2001), $\boldsymbol{\beta}_j$ is defined as $\boldsymbol{\beta}_j \sim N\left(0, \boldsymbol{I}, \sigma^2_{\beta_j}\right)$ where $\boldsymbol{I}$ is the identity matrix and $\sigma^2_{\beta_j}$ is the variance of genomic marker effects (Endelman, 2011), indicating that $\boldsymbol{\beta}_j$ is assumed to be normally distributed. For BayesB (Meuwissen et al., 2001), on the other hand, some values in $\boldsymbol{\beta}_j$ are diminished to zero with the assumption that not all the genomic markers J affect the trait phenotype. For RKHS (Gianola et al., 2008), Equation (1) is extended to explicitly capture the non-additive genomic marker effect by projecting the genomic markers into another dimensional space called the Hilbert space (Howard et al., 2014):

$$\boldsymbol{\eta} = 1\boldsymbol{\mu} + \sum_{j=1}^{J} \boldsymbol{X}_j \boldsymbol{\beta}_j + \sum_{j=1}^{J} g\left(\boldsymbol{X}_j\right) + \sum_{l=1}^{L} \boldsymbol{u}_l \tag{2}$$

where $g\left(\boldsymbol{X}_j\right)$ is the genomic marker effect from genomic markers mapped to the Hilbert Space using a kernel.

For the machine learning models, each aimed to capture both additive and non-additive genomic marker effects through the training process to learn key predictive patterns. RF (Breiman, 2001) is a collection of decision trees that predict target values by tracing the binary decision mechanism using feature values. The final prediction value is calculated by averaging the predicted values from each decision tree. SVR (Drucker et al., 1996) determines the optimal location of a predictive hyperplane by including as many data points as possible within the decision boundary assigned to each side of the hyperplane. The prediction from the optimised hyperplane is expected to minimise the prediction error. GAT (Veličković et al., 2018) is one of the deep learning approaches that leverages graph-structured data to predict target values. Each genomic marker and each phenotype data point is represented as a node, and the genomic marker nodes and phenotype nodes are connected with edges directed from the genomic marker nodes to the phenotype nodes. The GAT

application in Tomura et al. (2025) assigned no edges between the genomic marker nodes. This graph structure indicates that each genomic marker affects phenotypes independently and infinitesimally. GAT leverages the attention mechanism that allocates heavier weights to the genomic marker nodes that strongly affect phenotypes, expecting to increase prediction performance.

The naïve ensemble-average model (Tomura et al., 2025) averaged the predicted phenotypes from each individual genomic prediction model with equal weight. This ensemble model was implemented after each individual genomic prediction model predicted phenotypes.

For the classical genomic prediction models, BGLR (Pérez and de Los Campos, 2014) was leveraged for the implementation in R (v1.1.4). The iteration number and burn-in were set as 12,000 and 2,000, respectively. A Gaussian kernel was used for RKHS. The other parameters were set as default. For RF and SVR, Sklearn (v1.2.2) was used to develop the models in Python. The total number of decision trees was 1,000 in RF. The radial basis function (RBF) was selected for the kernel in SVR. The rest of the parameters were set as default. For GAT, Pytorch Geometric (v2.3.0) was employed for the implementation in Python. GAT was constructed with one hidden layer containing 20 channels and a dropout of 0. The Exponential Linear Unit (ELU) function was leveraged as the activation function. GAT was trained iteratively 50 times with the number of batches set as 8. The Adaptive Moment Estimation with Weight Decay (AdamW, Loshchilov and Hutter, 2017) was adopted as the optimiser. The learning rate and weight decay for the optimiser were 0.005 and 0, respectively. The other hyperparameters were set as default.

5. Genomic marker and marker-by-marker interaction effect inference

Genomic marker effects were inferred from each genomic prediction model to analyse the estimated additive genomic marker effects. For rrBLUP and BayesB, the allele substitution effect $\boldsymbol{\beta}_j$ was used as the genomic marker effect. For RKHS and SVR, the genomic marker effect was estimated using Shapley values (Shapley, 1953) as below (Lundberg and Lee, 2017):

$$\Phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{i\}}\left(x_{S \cup \{i\}}\right) - f_s(x_s) \right] \tag{3}$$

where Φ_i is the Shapley value of the feature i (genomic marker), S is a subset of the feature collection F, $f_{S\cup\{i\}}$ is the prediction from a model with a feature subset including i, f_s is the prediction from the model without including the feature subset i. Shapley values are the cumulative effect of a particular feature, considering all possible combinations based on the conditional probability. Shapley values were calculated for every genotype data point (RIL) and hence the average Shapley values for the RILs in the test set were calculated as the final Shapley value. For the implementation, iml (v0.11.3) (Molnar, 2020) was used in R to calculate Shapley values for RKHS. SHAP (v0.42.1) (Lundberg and Lee, 2017) was used for SVR. For RF genomic marker effects, feature importance was leveraged using the impurity-based approach (Ishwaran, 2015). This approach calculates how well each feature (genomic marker) can separate data points containing similar target values into the same nodes in the constituent decision trees. For GAT, Integrated Gradients was leveraged as below (Sundararajan et al., 2017):

$$IG(x_i) \approx (x_i - x_i') * \sum_{k=1}^{m} \frac{\partial F\left(x_i' + \frac{k}{m} * (x_i - x_i')\right)}{\partial x_i} * \frac{1}{m} \tag{4}$$

where x is a feature (genomic marker), x' is the baseline value of x and m is the total number of interpolation steps for the integral. Integrated Gradients calculates the effect of features based on the changes in predicted values between the inclusion and exclusion of the features in the model. Integrated Gradients was calculated per data point (RIL) and the average values from the test set were estimated as the final Integrated Gradients. Pyg (v2.4.0) was leveraged for the implementation. The genomic marker effects were inferred in each prediction scenario for every genomic marker. Hence, the average of the genomic marker effects across all the prediction scenarios was leveraged as the final genomic marker effect.

Using the genomic marker effects estimated from all the individual genomic prediction models, the genomic marker effects for the naïve ensemble-average model were calculated. The genomic marker effects from each individual genomic prediction model were normalised per prediction scenario (the details are explained in "Experimental design"), and the arithmetic average of the normalised genomic marker effects across the individual genomic prediction models was used as the genomic marker effects for the naïve ensemble-average model.

For the genomic marker-by-marker interaction effects, pairwise Shapley values were calculated from RF (Equation (3)) using SHAP. Since Shapley values are the summation of Shapley values based on possible feature combinations with the conditional probability, the value can be decomposed at the pairwise level. The genomic marker-by-marker interaction effects were also returned per prediction scenario. The average genomic marker interaction effects across all prediction scenarios were used as the final genomic marker interaction effects.

6. Experimental design

The genomic prediction models were iteratively developed independently under different prediction scenarios, and the genomic marker and marker-by-marker interaction effects were inferred for each scenario to reduce the risk of obtaining biased results. Each population was randomly split into a training and test set 500 times with three different training and test set ratios (0.8-0.2, 0.65-0.25 and 0.5-0.5). Hence, the genomic marker and interaction effects were inferred 7,500 times (5 populations * 3 ratios * 500 random splits) for each genomic prediction model. The averaged genomic marker and interaction effects across all the 7,500 prediction scenarios were leveraged to infer the genetic architecture by constructing the genomic marker effect rings and genomic marker-by-marker interaction effect links in the circos plots.

Results

1. Genomic prediction models identified several key genomic marker regions

The genomic prediction models identified several genomic marker regions that collocate with previously discovered influential genes for DTA in maize (Figure 1; Table 2). One of the strongest overlaps was observed in genomic marker regions containing genes controlling the photoperiod pathways, specifically ZmCCT9, ZmCCT10 and ZCN8. Almost all the genomic prediction models highlighted the genomic marker regions corresponding to these three genes. The word “highlight” here is defined as the genomic marker regions that

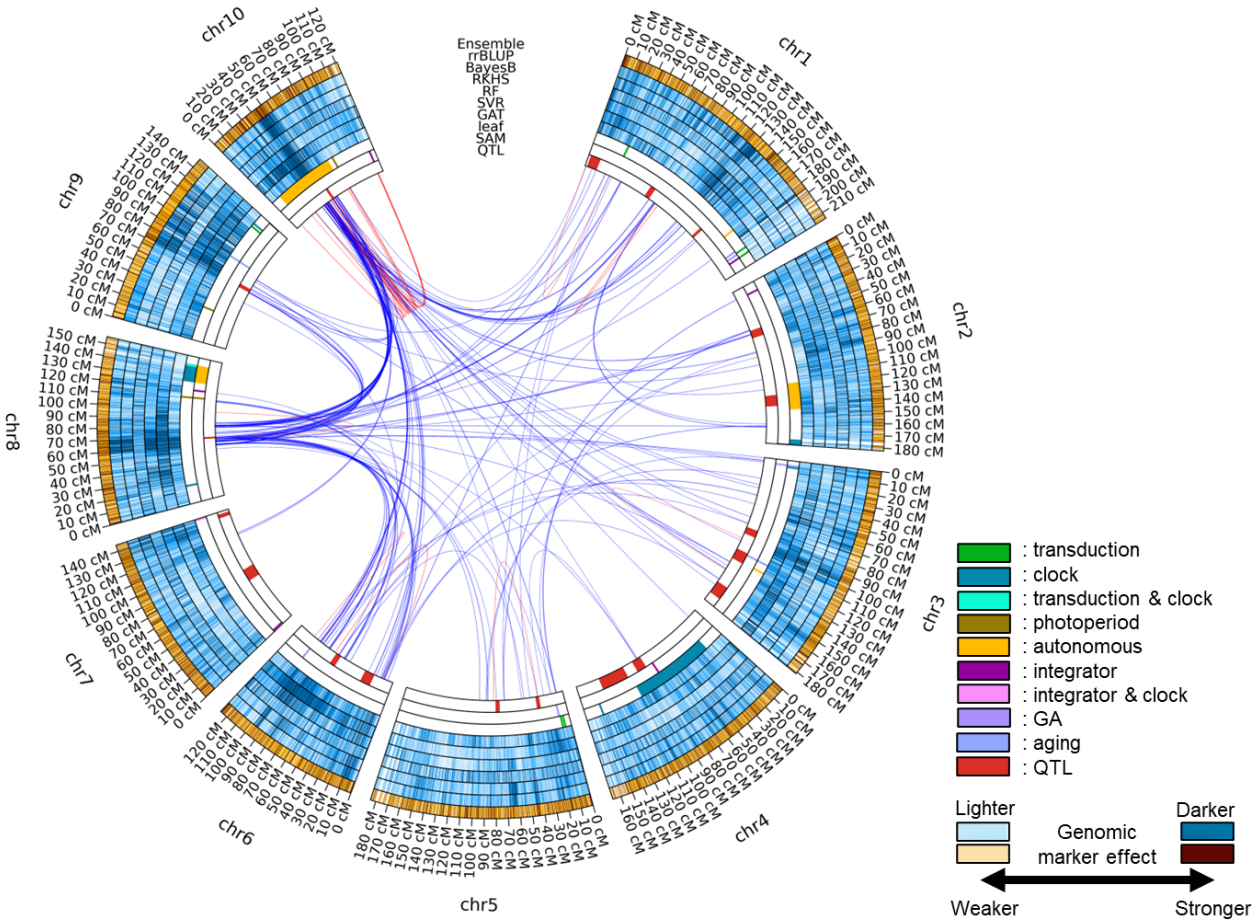

Figure 1: Circos plot for the trait days to anthesis (DTA) in the TeoNAM dataset across the five populations. The innermost ring (QTL) has the QTL genomic marker regions for DTA identified in Chen et al. (2019). The second and third innermost rings (SAM and leaf) represent the key genomic marker regions that affect the shoot apical meristem (SAM) and leaf, respectively, from Dong et al. (2012). These key genomic marker regions affect light transduction (transduction), circadian clock (clock), photoperiod, autonomous, integrator, gibberellin (GA) and aging pathways. The subsequent blue rings (fourth to ninth) represent the genomic marker effects of each genomic marker region inferred from the GAT, SVR, RF, RKHS, BayesB and rrBLUP genomic prediction models, respectively. The outermost orange ring is the genomic marker effect estimated for the naïve ensemble-average model. The intensity of blue and orange colours shows the strength of the genomic marker effect classified into ten quantiles. Darker colours are allocated to the genomic marker regions with higher genomic marker effect levels. The red and blue lines between genomic marker regions are the genomic marker-by-marker interaction effects estimated by pairwise Shapley values from RF (top 0.01%; red = within chromosome and blue = between chromosomes).

were continuously in the top 10% of genomic marker effects across prediction models. ZCN8 is a key gene for flowering time in maize, controlling the transition from the vegetative to reproductive stage by functioning in both the leaf and SAM (Meng et al. 2011; Dong et al., 2012; Guo et al., 2018; Aiyesa et al., 2025). ZmCCT9 promotes flowering time when daylight hours become longer at high-latitude regions by regulating ZCN8 (Huang et al., 2018). ZmCCT10 is another critical gene in the photoperiod pathway, which delays flowering

Table 2: A list of key genomic marker regions identified by at least two genomic prediction models. The gene names and functions were based on Dong et al. (2012) and Chen et al. (2019). Genomic prediction model indicates the models that continuously allocated the top 10% of genomic marker effects to the corresponding key genomic marker regions.

Chromosome	Region (cM)	Gene name	Function	Genomic prediction model
1	33 - 36	PhyB1	Light transduction	rrBLUP, BayesB, RKHS, GAT and Ensemble
1	80 - 85	-	-	rrBLUP, BayesB and RF
2	10 - 15	ZFL2	Integrator	rrBLUP, BayesB, and GAT
2	133 - 134	FCA	Autonomous	RKHS, GAT and Ensemble
3	5 - 10	miR156, GIGZ1B	Aging, Clock	rrBLUP, BayesB, RKHS, SVR, GAT and Ensemble
3	85 - 90	ZmMADS69	Adaptation	All
4	55 - 60	ZmLHY2	Clock	All
6	75 - 76	ZAG1	Flower development	BayesB, RKHS, RF, SVR, GAT and Ensemble
7	68 - 70	-	Early flowering	RKHS, GAT and Ensemble
7	140 - 145	dlf1, ZmPRR37	Integrator, Clock	All
8	65 - 70	ZCN8	Photoperiod	rrBLUP, BayesB, RF, SVR, GAT and Ensemble
9	70 - 75	ZmCCT9	Photoperiod	rrBLUP, BayesB, RF, SVR, GAT and Ensemble
10	40 - 55	ZmCCT10	Photoperiod	rrBLUP, BayesB, RF, SVR, GAT and Ensemble
10	110 -115	ZFL1	Integrator	RKHS, RF and Ensemble

under long days through interactions between ZCN8 and ZmCCT9 (Dong et al., 2012; Hung et al., 2012; Wisser et al., 2019). Such close interactions between the genes in the photoperiod pathway were captured in the circos plot by showing heavy marker-by-marker interactions between the regions of those genes, especially between chromosome 8 (ZCN8) and 10 (ZmCCT10) (Figure 1).

Several colocation overlaps were also identified at the genomic marker regions regulating the circadian clock (ZmLHY2, ZmPRR37 and GIGZ1B). ZmLHY2 is a transcription factor that regulates the morning loop of the circadian clock (Hayes et al., 2010). ZmPRR37 also regulates the circadian clock in maize by signalling diurnal outputs, which is a homolog to PRR genes in rice (Hayes et al., 2010). GIGZ1B is expressed at its peak level during the daylight hours, controlling the circadian clock, regardless of the length of daylight (Miller et al., 2008). These colocation overlaps of regions identified by the genomic prediction models with regions containing known genes involved in controlling maize DTA indicate that the genomic prediction models captured several key features of genomic marker information relating to the circadian clock.

Additionally, the genomic prediction models identified several key genomic marker regions in chromosome 2, 7 and 10 that contain genes that function as integrators of the different pathways (ZFL1, ZFL2 and dlf1). These integrators respond to the signals from the multiple pathways to determine DTA as an output of the regulatory network. Hence, integrators may contain multiple different functions. ZFL1 and ZFL2 are involved in the transition from the vegetative stage to the reproductive stage by upregulating the ABC floral organ identity network of genes (Bomblies et al., 2003). In addition, ZFL2 is known to control tassel branching and the traits relating to inflorescence structure in maize (Bomblies and Doebley, 2006). dfl1 encodes a protein with a basic leucine zipper domain (Muszynski et al., 2006), expressed in the leaves and regulating the timing of the transition to the flowering stage by sending signals from leaves to the SAM.

Other key genomic marker regions identified by the genomic prediction models, such as PhyB1, FCA, miR156, ZmMADS69 and ZAG1, influence the flowering time in maize by affecting light transduction, aging, adaptation and flower development pathways. Collectively, those results indicate that the genomic prediction models captured several key genomic marker regions and interactions involved in the genetic architecture of DTA in maize.

The genomic prediction models also highlighted several other genomic marker regions that have not been well-studied previously (Figure 1). For instance, while the genomic marker region in chromosome 1 (from 125 cM to 135 cM) was highlighted by all the individual genomic prediction models besides RKHS, no nearby key genomic marker regions were identified in the previous studies. Similarly, another genomic marker region in chromosome 1 (from 185 cM to 190 cM) was highlighted in the inferred genetic architecture, except for RKHS and SVR, but no key genomic marker region overlapped. Such discordant genomic marker regions were observed in several other genomic marker regions (from 85 cM to 90 cM in chromosome 2 and 45 cM and 65 cM in chromosome 5). Those highlighted genomic marker regions indicate that the genomic prediction models heavily weighted the predictive information from several genomic marker regions not identified in the previous studies.

2. Each genomic prediction model captured different genetic architectures

While the genomic prediction models identified several key genomic marker regions, their relative magnitudes differed among the genomic prediction models. The different relative values are indicated by unique patterns of colour intensity within each ring (Figure 1). For example, while the genomic marker region between 30 cM and 60 cM in chromosome 10 showed a similar pattern in the colour intensity by rrBLUP, BayesB, RF and GAT, the genetic architecture of RKHS and SVR in this region showed a unique pattern. RKHS did not specifically highlight the region as highly influential, while SVR emphasised the genomic marker effects on the first half of chromosome 10 more than the other genomic prediction models. Such diversity was also observed in the other highlighted genomic marker regions that overlapped with key genomic marker regions identified by the previous studies (Table 2). While several key genomic marker regions were identified across all the genomic prediction models as contributing strong marker effects, such regions were mainly highlighted by a fraction of the genomic prediction models. The lack of a clear consensus in the highlighted genomic marker regions for all six genomic prediction models considered also showed that each genomic prediction model quantified genomic marker effects differently. Such uniqueness led to different views of the genetic architecture of DTA in the TeoNAM dataset for the different genomic prediction models.

3. Inferred genetic architectures differed among the TeoNAM populations

The patterns of the quantified genomic marker effects in each genomic marker region were distinctive in each of the five populations composing the TeoNAM dataset. For instance, while the genomic marker region containing ZmCCT10 was highlighted in chromosome 10 across the populations, the range of the highlighted region varied among populations. For W22TIL01 and W22TIL25, the highlighted region was between 40 cM and 50 cM (Figure 2a, 2e), whereas the region was extended in W22TIL03 (between 45 cM and 55 cM) and W22TIL14 (between 30 cM and 55 cM) (Figure 2b, 2d). The region for W22TIL11 was between 40cM and 55cM (Figure 2c) with a lower frequency of highlights compared to the others. Similarly, the genomic marker region between 85 cM and 95 cM in chromosome 1 was mainly allocated to the top 10% of the genomic marker effects by all the genomic prediction models in W22TIL03 (Figure 2b). However, the genomic prediction models in the other four populations did not frequently allocate such high genomic marker effects

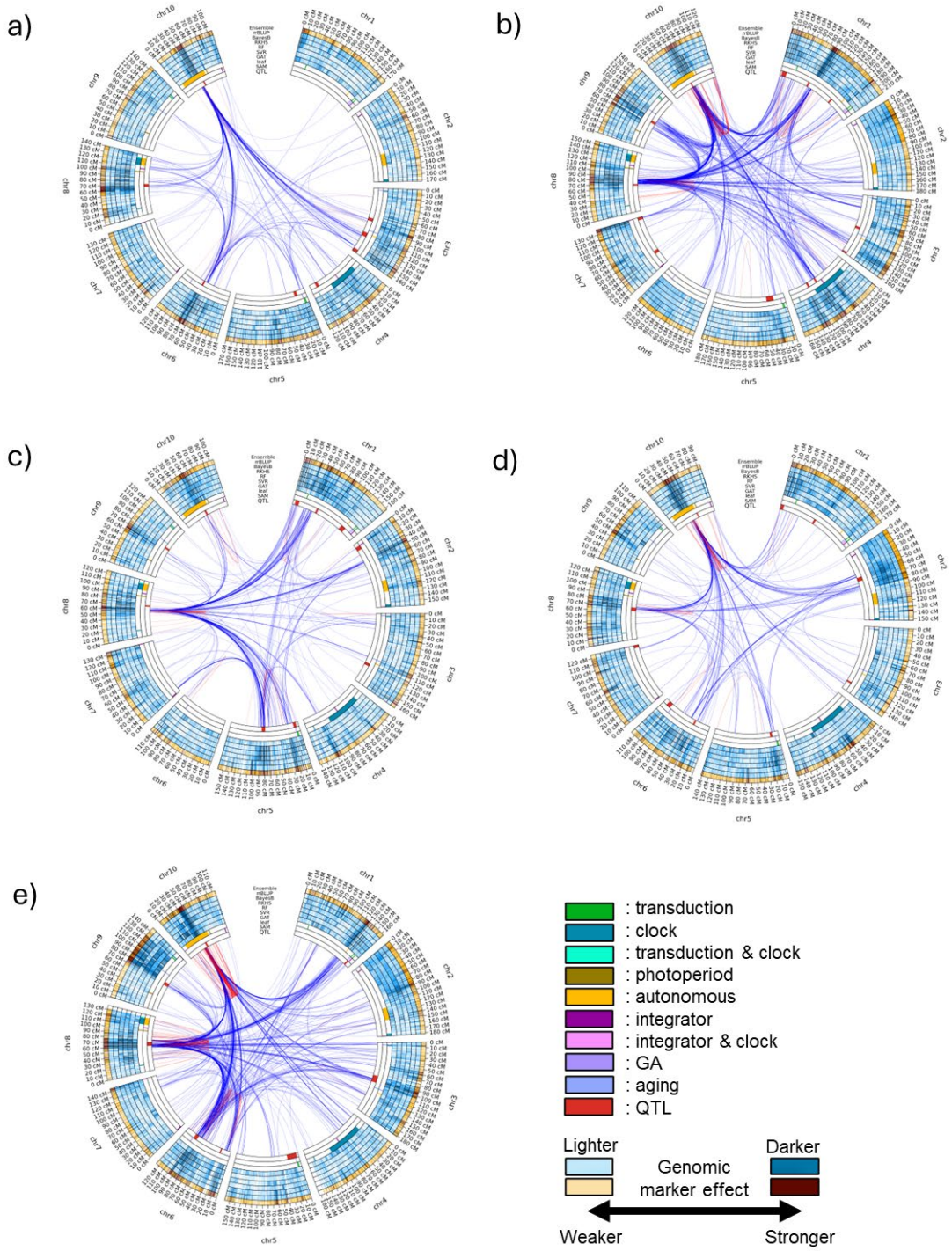

Figure 2: Circos plots for days to anthesis (DTA) in the TeoNAM dataset for the five populations: a) W22TIL01, b) W22TIL03, c) W22TIL11, d) W22TIL14 and e) W22TIL25. The innermost ring (QTL) has the QTL genomic marker regions identified in Chen et al. (2019). The second and third innermost rings (SAM and leaf) represent the key genomic marker regions that affect the shoot apical meristem (SAM) and leaf, respectively, from Dong et al. (2012). These key genomic marker regions affect light transduction (transduction), circadian clock (clock), photoperiod, autonomous, integrator, gibberellin (GA) and aging pathways. The subsequent blue rings (fourth to ninth) represent the genomic marker effects of each genomic marker region inferred from GAT, SVR, RF, RKHS, BayesB and rrBLUP, respectively. The outermost orange ring is the genomic marker effect estimated from the naïve ensemble-average model. The intensity of blue and orange colours shows the strength of the genomic marker effect classified into ten levels based on the quantiles. Darker colours are allocated to the genomic marker regions with higher genomic marker effect levels. The red and blue lines between genomic marker regions are the genomic marker interaction effect estimated by pairwise Shapley values from RF (top 0.01%; red = within chromosome and blue = between chromosomes).

to those regions of the genome. Instead, the genomic prediction models more frequently assigned higher genomic marker effects to other genomic marker regions. Further, while all the genomic prediction models frequently assigned the top 10% of the genomic marker effects to the genomic marker region between 60 cM and 80 cM in chromosome 9 of W22TIL03 (Figure 2b), the genomic prediction models did not clearly highlight the same genomic marker region in the four other populations. A lack of consensus in the quantification patterns of the genomic marker effect among the five TeoNAM populations was observed for other genomic marker regions across the chromosomes. These diverse patterns indicate that the genomic prediction models recognise different genomic marker regions as influential in each of the five populations comprising the TeoNAM dataset.

Such diversity in the marker effect combinations among the six genomic prediction models for each population was also observed in the inferred genomic marker-by-marker interaction effects. While the strong interactions between chromosomes 8 and 10 were conserved across the populations, many of the marker-by-marker interaction regions differed among the five populations. The region for chromosome 8 was located at 70 cM in W22TIL01, W22TIL14 and W22TIL25 (Figure 2a, 2d, 2f), whereas the region was located at 80 cM in W22TIL03 and 60 cM in W22TIL11 (Figure 2b, 2c). For chromosome 10, the region was located between 40 cM and 50 cM in W22TIL01 and W22TIL25 (Figure 2a, 2e), between 50 cM and 60 cM in W22TIL03 (Figure 2b), between 45 cM and 55 cM in W22TIL11 (Figure 2c) and between 30 cM and 45 cM in W22TIL14 (Figure 2d). Additionally, this region in chromosome 8 strongly interacted with other chromosomes in unique patterns identified across the five populations. While all the populations showed interactions with chromosome 5, this was stronger than for the other populations in W22TIL11 (Figure 2c). Three populations (W22TIL03, W22TIL11 and W22TIL25) showed heavy marker-by-marker interactions between chromosome 1 and chromosome 3 (Figure 2b, 2c, 2e), whereas there were strong interactions in chromosome 2 for W22TIL11 and W22TIL14 (Figure 2c, 2d). Such distinctive marker-by-marker interaction patterns were observed throughout the chromosomes. Those results indicate that the inferred views of trait genetic architecture were diverse across the five populations.

Discussion

1. Diverse quantification of genomic marker effects improves the prediction performance of the ensemble

The circos plot provided a genome level view of how the six investigated genomic prediction models estimated diverse genomic marker effects (Figure 1). This diversity is attributed to algorithmic differences among the prediction models. The algorithms of rrBLUP and BayesB primarily targeted the detection of the additive genomic marker effects (Meher et al., 2022), while RKHS also included a kernel to explicitly identify several non-additive genomic marker-by-marker interaction effects (Gianola et al., 2008). The RF, SVR and MLP included components in their prediction algorithms to identify more complex non-additive prediction patterns underlying the inclusion of the genetic markers (Chen et al., 2023). Several studies have investigated the diversity in the quantified feature importance (genomic marker effects in our study) using different prediction models. The comparison of the estimated feature importance from logistic regression (additive prediction model) and RF (non-additive prediction model) indicated that several identified key features varied from the prediction models for the classification tasks in the cancer study (Saarela and Jauhiainen, 2021). A linear regression model was also compared with RF in regression tasks, showing a similar conclusion that the level of importance in each feature identified by the linear regression was not necessarily aligned with that of RF (Grömping, 2009). The estimated patterns in the feature importance were also investigated within three machine learning models (multi-layer perception, naïve Bayes and Bagging), with the observation that each prediction model weighted feature importance distinctively (Štrumbelj and Kononenko, 2014). Hence, the different prediction strategies in each prediction algorithm quantified the genomic marker effects uniquely throughout the chromosomes.

Such uniqueness in the captured genomic marker effects was represented as diverse views of the genetic architecture of the trait DTA in our study using the TeoNAM dataset (Figure 1, Table 2). The circos plot enabled the visualisation of diverse patterns in the colour intensity across the chromosomes that were not identical to any other views. The combination of such diverse views might have generated another view through the ensemble approach that can be more comprehensive, considering the consistent performance improvement of the ensemble approach observed in Tomura et al. (2025). The weaknesses of prediction algorithms can potentially be offset by aggregating the diversely quantified genomic marker effects from each model (Kick and Washburn, 2023; Washburn et al., 2025). The interpretation of the visualised genetic

architecture for the ensemble supports the Diversity Prediction Theorem (Page, 2018), implying that diversified predicted phenotypes through differently quantified genomic marker effects contributed to the reduction of the ensemble error for the naïve ensemble-average model compared to the average model error. This implication is also aligned with the explanation of the No Free Lunch Theorem (Wolpert and Macready, 1997), stating that no individual prediction model can consistently reach the highest prediction performance across the prediction scenarios. The investigation of the genomic prediction models at the genomic level deepens our understanding of the possible reasons for the prediction performance improvement by the ensemble through the concept of diversity.

The application of both interpretable prediction models and circos plots has been observed in several studies, primarily in the human genome, using the self-attention mechanism and statistical analysis (Lee et al., 2023; Peng et al., 2023; Prakash and Banerjee, 2023). However, this approach has not been widely applied to the genome-level study of genomic prediction models for crop breeding applications. Hence, we have developed a novel genomic tool - the construction of circos plots that visualise genomic marker effects from interpretable genomic prediction models.

2. Circos plots may reveal potential new key genomic marker regions

The circos plot view of the six genomic prediction models and the ensemble highlighted several genomic marker regions that were not previously identified as key genomic regions (Figure 1). Those regions did not overlap with key gene regions reported from previous studies and could potentially be leveraged as new genomic marker regions that can contribute to improving target traits. There are multiple possible reasons why these new regions were identified by the genomic prediction models rather than by traditional QTL mapping. Those reasons can be attributed to the fundamental distinction that traditional QTL analysis leverages different approaches for QTL detection than the interpretable genomic prediction (Holland et al. 2004; Alonso et al. 2006).

The use of traditional and new approaches can contribute to identifying more comprehensive sets of key genomic markers for target traits. The combination of both QTL analysis and interpretable prediction

models (machine learning) has started being evaluated in crop breeding programs targeting soybeans (Gill et al., 2022; Yoosefzadeh et al., 2022). Since the use of interpretable genomic prediction for identifying key genomic marker regions has been a relatively recent approach in crop breeding, it is vital to verify the reliability through empirical studies and simulations. Nevertheless, there is potential for the collective analysis of key genomic marker regions to enhance the understanding of trait genetic architecture that can support prediction applications for breeding.

3. Distinctive genetic architecture in each population may hinder accurate between-population prediction

The genomic prediction models identified different key patterns in the genomic marker and marker-by-marker interaction effects across the five populations comprising the TeoNAM dataset (Figure 2). While similar genomic marker regions were highlighted with heavy interactions for some regions, the majority of the genetic architectures identified by the genomic prediction models showed unique patterns by quantifying different genomic marker regions with high genomic marker and marker-by-marker interaction effects. The diversity in the genetic architecture at the population level might have been attributed to the structure of the TeoNAM populations. The development of NAM populations was proposed to improve resolution through diversifying the allele combinations within the mapping study while maintaining common linkage information for QTL analysis, achieved by leveraging a common parent and numerous donor parents (Gage et al., 2020). Each population is designed to be genetically diverse from the others. Consequently, the allele combination becomes unique with different allele frequency patterns in each population (Gireesh et al., 2021). In the TeoNAM dataset, five populations were developed from one of the two teosinte types crossed with the W22 maize line used as the common parent, followed by backcrossing and a series of selfing (Chen et al., 2019). Thus, the five populations were characterised by distinctive allele combination patterns. Consequently, the genetic architecture of each population showed a distinctive structure.

The differences in the genetic architecture for each population may hinder the between-population prediction with high accuracy. Each genomic prediction model learns key prediction patterns from the genetic architecture of the training set. However, the learnt prediction patterns may not be aligned with the key

prediction patterns in the test set from another population due to differences in the standing genetic variation. For example, in wheat, the prediction performance of the between-population prediction for 19 traits was considerably lower than the within-population prediction for the elite population and the yield population (Huang et al., 2018). This poor prediction performance was attributed to a lack of lines closely related to the test set. The performance of the between-population prediction for sorghum depended on the combination of the populations for the training and test sets (Sapkota et al., 2020). These experimental results from previous studies indicate that the differences in the genetic architecture between the training and test sets considerably affect the performance of genomic prediction models for the between-population scenarios. Higher prediction performance was achieved when the population structure between the training and test sets was similar. De Roos et al. (2009) concluded that the performance of between-population prediction increases when samples from all populations are included in the training set. The inclusion of samples from different populations allows genomic prediction models to capture multiple views of the trait genetic architecture.

Conclusion

Circos plots were developed to visualise the inferred genetic architectures from the six individual genomic prediction models reported in Tomura et al. (2025). The circos plots revealed important diversity in the trait genetic architectures identified for the different genomic prediction models by quantifying genomic marker effects differently while still capturing several key genomic marker regions. The aggregation of such diverse views can potentially enhance the prediction performance using the ensemble, as indicated by the Diversity Prediction Theorem. The diverse views of the trait genetic architecture across populations indicate a source of difficulty in achieving high accuracy for between-population predictions. There is also potential for the interpretable genomic prediction model approach to discover novel key genomic marker regions that have not been identified previously. The interpretability can deepen the understanding of each genomic prediction model at the genome level and thus benefit crop breeding programs by deepening our understanding of trait genetic architecture and the prediction performance improvement using the ensembles of genomic prediction models.

Data and code availability

The code and data used in this study are available at https://github.com/ShunichiroT/Circos_plot. The original TeoNAM dataset (Chen et al., 2019), leveraged to infer the genomic marker and interaction effects, is available at https://datacommons.cyverse.org/browse/iplant/home/shared/panzea/genotypes/GBS/TeosinteNAM for the genotypes and https://gsajournals.figshare.com/articles/dataset/Supplemental_Material_for_Chen_et_al_2019/9250682 for the phenotypes.

Acknowledgments

We thank the National Computing Infrastructure (NCI) and Research Computing Centre (RCC) at the University of Queensland for providing access to use High Performance Computing (HPC) machines.

Funding

This study was funded by the Australian Research Council through the support of the Australian Research Council Centre of Excellence for Plant Success in Nature and Agriculture (CE200100015).

Conflicts of interest

The authors declare no conflicts of interest.

References

Abdullah T, Zahid M, Ali W. 2021. A review of interpretable ml in healthcare: Taxonomy, applications, challenges, and future directions. Symmetry. 13. https://doi.org/10.3390/sym13122439.

Aiyesa LV, Beissinger T, Scholten S, Link W, Zumbach B, Kaufmann D. 2025. Individual plant genetics reveal the control of local adaption in european maize landraces. BMC Biology. 23. https://doi.org/10.1186/s12915-025-02241-8.

Alemu A, Åstrand J, Montesinos-Lopez OA, y Sanchez JI, Fernandez-Gonzalez J, Tadesse W, Vetukuri RR, Carlsson AS, Ceplitis A, Crossa J *et al*. 2024. Genomic selection in plant breeding: Key factors shaping two decades of progress. Molecular Plant. 17:552–578. https://doi.org/10.1016/j.molp.2024.03.007.

Alonso-Blanco C, Koornneef M, van Ooijen JW. 2006. Qtl analysis. Arabidopsis protocols. pp. 79–99. https://doi.org/10.1385/1-59745-003-0:79.

Baniecki H, Sobieski B, Szatkowski P, Bombinski P, Biecek P. 2025. Interpretable machine learning for time-to-event prediction in medicine and healthcare. Artificial Intelligence in Medicine. 159:103026. https://doi.org/10.1016/j.artmed.2024.103026.

Bellman R. 1957. Dynamic programming. Princeton University Press. pp. 24–73. https://doi.org/10.1126/science.153.3731.34.

Bomblies K, Doebley JF. 2006. Pleiotropic effects of the duplicate maize floricaula/leafy genes zfl1 and zfl2 on traits under selection during maize domestication. Genetics. 172:519–531. https://doi.org/10.1534/genetics.105.048595.

Bomblies K, Wang RL, Ambrose BA, Schmidt RJ, Meeley RB, Doebley J. 2003. Duplicate floricaula/leafy homologs zfl1 and zfl2 control inflorescence architecture and flower patterning in maize. Development. 130. https://doi.org/10.1242/dev.00457.

Bose S, Banerjee S, Kumar S, Saha A, Nandy D, Hazra S. 2024. Review of applications of artificial intelligence (ai) methods in crop research. Journal of Applied Genetics. 65:225–240. https://doi.org/10.1007/s13353-023-00826-z.

Breiman L. 2001. Random forests. Machine Learning. 45:5–32. https://doi.org/10.1023/A:1010933404324.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC *et al*. 2009. The genetic architecture of maize flowering time. Science. 325:714–718. https://doi.org/10.1126/science.1174276.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation plink: rising to the challenge of larger and richer datasets. GigaScience. 4. https://doi.org/10.1186/s13742015-0047-8.

Chen C, Powell O, Dinglasan E, Ross EM, Yadav S, Wei X, Atkin F, Deomano E, Hayes BJ. 2023. Genomic prediction with machine learning in sugarcane, a complex highly polyploid clonally propagated crop with substantial non-additive variation for key traits. The Plant Genome. 16:e20390. https://doi.org/10.1002/tpg2.20390.

Chen Q, Yang CJ, York AM, Xue W, Daskalska LL, DeValk CA, Krueger KW, Lawton SB, Spiegelberg BG, Schnell JM *et al*. 2019. Teonam: A nested association mapping population for domestication and agronomic trait analysis in maize. Genetics. 213:1065–1078. https://doi.org/10.1534/genetics.119.302594.

Cooper M, van Eeuwijk FA, Hammer GL, Podlich DW, Messina C. 2009. Modeling qtl for complex traits: detection and context for plant breeding. Current opinion in plant biology. 12:231–240. https://doi.org/10.1016/j.pbi.2009.01.006.

Crossa J, Montesinos-Lopez OA, Costa-Neto G, Vitale P, Martini JW, Runcie D, FritscheNeto R, Montesinos-Lopez A, Pérez-Rodríguez P, Gerard G *et al*. 2025. Machine learning algorithms translate big data into predictive breeding accuracy. Trends in Plant Science. https://doi.org/10.1016/j.tplants.2024.09.011.

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, De Los Campos G, Burguenõ J, González-Camacho JM, Pérez-Elizalde S, Beyene Y *et al*. 2017. Genomic selection in plant breeding: methods, models, and perspectives. Trends in plant science. 22:961–975. http://dx.doi.org/10.1016/j.tplants.2017.08.011.

De Roos A, Hayes BJ, Goddard M. 2009. Reliability of genomic predictions across multiple populations. Genetics. 183:1545–1553. https://doi.org/10.1534/genetics.109.104935.

Dong Z, Danilevskaya O, Abadie T, Messina C, Coles N, Cooper M. 2012. A gene regulatory network model for floral transition of the shoot apex in maize and its dynamic modeling. Plos one. https://doi.org/10.1371/journal.pone.0043450.

Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. 1996. Support vector regression machines. Advances in neural information processing systems. 9.

Endelman JB. 2011. Ridge regression and other kernels for genomic selection with r package rrblup. The Plant Genome. 4:250–255. https://doi.org/10.3835/plantgenome2011.08.0024.

Gage JL, Monier B, Giri A, Buckler ES. 2020. Ten years of the maize nested association mapping population: impact, limitations, and future directions. The Plant Cell. 32:2083–2093. https://doi.org/10.1105/tpc.19.00951.

Gianola D, Van Kaam JB. 2008. Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics. 178:2289–2303. https://doi.org/10.1534/genetics.107.084285.

Gill M, Anderson R, Hu H, Bennamoun M, Petereit J, Valliyodan B, Nguyen HT, Batley J, Bayer PE, Edwards D. 2022. Machine learning models outperform deep learning models, provide interpretation and facilitate feature selection for soybean trait prediction. BMC plant biology. 22:180. https://doi.org/10.1186/s12870-022-03559-z.

Gireesh C, Sundaram RM, Anantha SM, Pandey MK, Madhav MS, Rathod S, Kondajji R. Yathish P, Senguttuvel, Kalyani BM, Ranjith E *et al*. 2021. Nested association mapping (nam) populations: Present status and future prospects in the genomics era. Critical Reviews in Plant Sciences. https://doi.org/10.1080/07352689.2021.1880019.

Grömping U. 2009. Variable importance assessment in regression: linear regression versus random forest. The American Statistician. 63:308–319. https://doi.org/10.1198/tast.2009.08199.

Guo L, Wang X, Zhao M, Huang C, Li C, Li D, Yang CJ, York AM, Xue W, Xu G *et al*. 2018. Stepwise cis-regulatory changes in zcn8 contribute to maize flowering-time adaptation. Current biology. 28:3005–3015. https://doi.org/10.1016/j.cub.2018.07.029.

Hayes KR, Beatty M, Meng X, Simmons CR, Habben JE, Danilevskaya ON. 2010. Maize global transcriptomics reveals pervasive leaf diurnal rhythms but rhythms in developing ears are largely limited to the core oscillator. PloS one. 5:e12887. https://doi.org/10.1371/journal.pone.0012887.

Heslot N, Yang HP, Sorrells ME, Jannink JL. 2012. Genomic selection in plant breeding: a comparison of models. Crop science. 52:146–160. https://doi.org/10.2135/cropsci2011.06.0297.

Holland JB. 2004. Implementation of molecular markers for quantitative traits in breeding programs—challenges and opportunities. volume 26. pp. 1–13. Regional Institute Gosford, Australia.

Howard R, Carriquiry AL, Beavis WD. 2014. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. G3 Genes—Genomes—Genetics. 4:1027–1046. https://doi.org/10.1534/g3.114.010298.

Huang C, Sun H, Xu D, Chen Q, Liang Y, Wang X, Xu G, Tian J, Wang C, Li D *et al*. 2018a. Zmcct9 enhances maize adaptation to higher latitudes. Proceedings of the National Academy of Sciences. 115:E334–E341. https://doi.org/10.1073/pnas.1718058115.

Huang M, Ward B, Griffey C, Van Sanford D, McKendry A, Brown-Guedira G, Tyagi P, Sneller C. 2018b. The accuracy of genomic prediction between environments and populations for soft wheat traits. Crop Science. 58:2274–2288. https://doi.org/10.2135/cropsci2017.10.0638.

Hung HY, Shannon LM, Tian F, Bradbury PJ, Chen C, Flint-Garcia SA, McMullen MD, Ware D, Buckler ES, Doebley JF *et al*. 2012. Zmcct and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proceedings of the National Academy of Sciences. 109:E1913–E1921. https://doi.org/10.1073/pnas.1203189109.

Ishwaran H. 2015. The effect of splitting on random forests. Machine learning. 99:75–118. https://doi.org/10.1007/s10994-014-5451-2.

Kick DR, Washburn JD. 2023. Ensemble of best linear unbiased predictor, machine learning and deep learning models predict maize yield better than each model alone. in silico Plants. 5:diad015. https://doi.org/10.1093/insilicoplants/diad015.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. Genome research. 19:1639–1645. https://doi.org/10.1101/gr.092759.109.

Kumar R, Das SP, Choudhury BU, Kumar A, Prakash NR, Verma R, Chakraborti M, Devi AG, Bhattacharjee B, Das R *et al*. 2024. Advances in genomic tools for plant breeding: harnessing dna molecular markers, genomic selection, and genome editing. Biological Research. 57:80. https://doi.org/10.1186/s40659-024-00562-6.

Lee H, Lee S, Lee I, Nam H. 2023. Amp-bert: Prediction of antimicrobial peptide function based on a bert model. Protein Science. 32:e4529. https://doi.org/10.1002/pro.4529.

Loshchilov I, Hutter F. 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101. https://doi.org/10.48550/arXiv.1711.05101.

Lundberg SM, Lee SI. 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems. 30.

Meher PK, Rustgi S, Kumar A. 2022. Performance of bayesian and blup alphabets for genomic prediction: analysis, comparison and results. Heredity. 128:519–530. https://doi.org/10.1038/s41437-02200539-9.

Meng X, Muszynski MG, Danilevskaya ON. 2011. The ft-like zcn8 gene functions as a floral activator and is involved in photoperiod sensitivity in maize. The Plant Cell. 23:942–960. https://doi.org/10.1105/tpc.110.081406.

Messina C, Garcia-Abadillo J, Powell O, Tomura S, Zare A, Ganapathysubramanian B, Cooper M. 2025. Towards a general framework for ai enabled prediction in crop improvement. Submitted to Theoretical and Applied Genetics.

Messina CD, Technow F, Tang T, Totir R, Gho C, Cooper M. 2018. Leveraging biological insight and environmental variation to improve phenotypic prediction: Integrating crop growth models (cgm) with whole genome prediction (wgp). European Journal of Agronomy. 100:151–162. https://doi.org/10.1016/j.eja.2018.01.007.

Meuwissen TH, Hayes BJ, Goddard M. 2001. Prediction of total genetic value using genome-wide dense marker maps. genetics. 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819.

Miller TA, Muslin EH, Dorweiler JE. 2008. A maize constans-like gene, conz1, exhibits distinct diurnal expression patterns in varied photoperiods. Planta. 227:1377–1388. https://doi.org/10.1007/s00425008-0709-1.

Molnar C. 2020. *Interpretable machine learning*. Lulu. com.

Montesinos-Lopez A, Crespo-Herrera L, Dreisigacker S, Gerard G, Vitale P, Saint Pierre C, Govindan V, Tarekegn ZT, Flores MC, Pérez-Rodríguez P *et al*. 2024. Deep learning methods improve genomic prediction of wheat breeding. Frontiers in Plant Science. 15:1324090. https://doi.org/10.3389/fpls.2024.1324090.

Muszynski MG, Dam T, Li B, Shirbroun DM, Hou Z, Bruggemann E, Archibald R, Ananiev EV, Danilevskaya ON. 2006. Delayed flowering1 encodes a basic leucine zipper protein that mediates floral inductive signals at the shoot apex in maize. Plant physiology. 142:1523–1536. https://doi.org/10.1104/pp.106.088815.

Oviedo F, Ferres JL, Buonassisi T, Butler KT. 2022. Interpretable and explainable machine learning for materials science and chemistry. Accounts of Materials Research. 3:597–607. https://doi.org/10.1021/accountsmr.1c00244.

Page SE. 2018. *The model thinker: What you need to know to make data work for you*. Hachette UK.

Peng L, Yuan R, Han C, Han G, Tan J, Wang Z, Chen M, Chen X. 2023. Cellenboost: a boostingbased ligand-receptor interaction identification model for cell-to-cell communication inference. IEEE Transactions on NanoBioscience. 22:705–715. https://doi.org/10.1109/TNB.2023.3278685.

Pérez P, de Los Campos G. 2014. Genome-wide regression and prediction with the bglr statistical package. Genetics. 198:483–495. https://doi.org/10.1534/genetics.114.164442.

Prakash A, Banerjee M. 2023. An interpretable block-attention network for identifying regulatory feature interactions. Briefings in Bioinformatics. 24:bbad250. https://doi.org/10.1093/bib/bbad250.

Ramstein GP, Jensen SE, Buckler ES. 2019. Breaking the curse of dimensionality to identify causal variants in breeding. Theoretical and Applied Genetics. 32:559–567. https://doi.org/10.1007/s00122018-3267-3.

Saarela M, Jauhiainen S. 2021. Comparison of feature importance measures as explanations for classification models. SN Applied Sciences. 3:272. https://doi.org/10.1007/s42452-021-04148-9.

Sapkota S, Boyles R, Cooper E, Brenton Z, Myers M, Kresovich S. 2020. Impact of sorghum racial structure and diversity on genomic prediction of grain yield components. Crop Science. 60:132–148. https://doi.org/10.1002/csc2.20060.

Shapley LS. 1953. Stochastic games. Proceedings of the national academy of sciences. 39:1095–1100. https://doi.org/10.1073/pnas.39.10.1095.

Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L. 2020. Interpretability of machine learning-based prediction models in healthcare. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 10:e1379. https://doi.org/10.1002/widm.1379.

Štrumbelj E, Kononenko I. 2014. Explaining prediction models and individual predictions with feature contributions. Knowledge and information systems. 41:647–665. https://doi.org/10.1007/s10115013-0679-x.

Sundararajan M, Taly A, Yan Q. 2017. Axiomatic attribution for deep networks. In: . pp. 3319–3328. PMLR.

Tavakoli M, Baldi P, Carlton AM, Chiu YT, Shmakov A, Van Vranken D. 2023. Ai for interpretable chemistry: Predicting radical mechanistic pathways via contrastive learning. Advances in Neural Information Processing Systems. 36:4080–4096.

Tessele A, Morris GP, Akhunov E, Johnson BE, Clinesmith M, Fritz AK. 2025. Interchromosomal linkage disequilibrium analysis reveals strong indications of sign epistasis in wheat breeding families. bioRxiv. pp. 2025–05. https://doi.org/10.1101/2025.05.16.654524.

Tomura S, Wilkinson MJ, Cooper M, Powell O. 2025. Improved genomic prediction performance with ensembles of diverse models. G3: Genes, Genomes, Genetics. p. jkaf048. https://doi.org/10.1093/g3journal/jkaf048.

Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. 2018. Graph attention networks. arXiv:1710.10903. https://doi.org/10.48550/arXiv.1710.10903.

Wang X, Yang Z, Xu C. 2015. A comparison of genomic selection methods for breeding value prediction. Science Bulletin. 60:925–935. https://doi.org/10.1007/s11434-015-0791-2.

Washburn JD, Varela JI, Xavier A, Chen Q, Ertl D, Gage JL, Holland JB, Lima DC, Romay MC, Lopez-Cruz M *et al*. 2025. Global genotype by environment prediction competition reveals that diverse modeling strategies can deliver satisfactory maize yield estimates. Genetics. 229:iyae195. https://doi.org/10.1093/genetics/iyae195.

Wei X, Chen M, Zhang Q, Gong J, Liu J, Yong K, Wang Q, Fan J, Chen S, Hua H *et al*. 2024. Genomic investigation of 18,421 lines reveals the genetic architecture of rice. Science. 385:eadm8762. DOI: 10.1126/science.adm8762.

Wetzel SJ, Ha S, Iten R, Klopotek M, Liu Z. 2025. Interpretable machine learning in physics: A review. arXiv preprint arXiv:2503.23616. https://doi.org/10.48550/arXiv.2503.23616.

Wisser RJ, Fang Z, Holland JB, Teixeira JE, Dougherty J, Weldekidan T, de Leon N, Flint-Garcia S, Lauter N, Murray SC *et al*. 2019. The genomic basis for short-term evolution of environmental adaptation in maize. Genetics. 213:1479–1494. https://doi.org/10.1534/genetics.119.302780.

Wolpert DH, Macready WG. 1997. No free lunch theorems for optimization. IEEE transactions on evolutionary computation. 1:67–82. https://doi.org/10.1109/4235.585893.

Yadav S, Wei X, Joyce P, Atkin F, Deomano E, Sun Y, Nguyen LT, Ross EM, Cavallaro T, Aitken KS *et al*. 2021. Improved genomic prediction of clonal performance in sugarcane by exploiting non-additive genetic effects. Theoretical and Applied Genetics. 134:2235–2252. https://doi.org/10.1007/s00122-021-03822-1.

Yoosefzadeh-Najafabadi M, Eskandari M, Torabi S, Torkamaneh D, Tulpan D, Rajcan I. 2022. Machine-learning-based genome-wide association studies for uncovering qtl underlying soybean yield and its components. International Journal of Molecular Sciences. 23:5538. https://doi.org/10.3390/ijms23105538.