Ensemble AnalySis with Interpretable Genomic Prediction (EasiGP): Computational Tool for Interpreting Ensembles of Genomic Prediction Models

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

[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

Core ideas

- A new computational tool, EasiGP, was created to interpret multiple genomic prediction models at the genomic level
- EasiGP visualises the inferred trait genetic architecture from multiple genomic prediction models with circos plots
- As a case study, EasiGP highlighted several well-known genes regulating the target trait, days to anthesis
- EasiGP can facilitate the discovery of novel genome regions underlying target traits for further investigation

ABSTRACT

Ensemble of multiple genomic prediction models have grown in popularity due to consistent prediction performance improvements in crop breeding. However, technical tools that analyse the predictive behaviour at the genome level are lacking. Here, we develop a computational tool called Ensemble AnalySis with Interpretable Genomic Prediction (EasiGP) that uses circos plots to visualise how different genomic prediction models quantify contributions of marker effects to trait phenotypes. As a demonstration of EasiGP, multiple genomic prediction models, spanning conventional statistical and machine learning algorithms, were used to infer the genetic architecture of days to anthesis (DTA) in a maize mapping population. 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, 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 reported for DTA. Finally, different views of trait genetic architecture were observed across sub-populations, highlighting challenges for between-population genomic prediction. A deeper understanding of genomic prediction models with enhanced interpretability using EasiGP 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.

Plain Language Summary

While an ensemble of genomic prediction models has been applied in crop breeding, the prediction mechanism has not been well-investigated due to the lack of a computational tool to interpret the predictive behaviour. It is critical to investigate prediction models at the genome level to understand how each model quantifies genomic marker effects contributing to the trait genetic architecture. Hence, we developed a computational tool, the Ensemble AnalySis with Interpretable Genomic Prediction (EasiGP), to investigate the genome features and predictive behaviours of the ensemble. Here, we demonstrate the utility of EasiGP using a maize breeding dataset. EasiGP visualised the genetic architecture from diverse interpretable genomic prediction models and identified several well-known key maize genes. EasiGP also revealed several potential new genomic regions for further investigation. EasiGP helps us investigate the trait genetic architecture that can be utilised to benefit crop breeding.

Abbreviations

AdamW: adaptive moment estimation with weight decay; DTA: days to anthesis; EasiGP: Ensemble AnalySis with Interpretable Genomic Prediction; F1: filial 1; GAT: graph attention network; GBLUP: genomic best linear unbiased prediction; GWAS: genome-wide association study; LD: linkage disequilibrium; MLP: multi-layer perceptron; NAM: nested association mapping; QTL: quantitative trait locus; RF: random forest; RIL: recombinant inbred line; RKHS: reproducing kernel Hilbert space; SAM: shoot apical meristem; SNP: single nucleotide polymorphism; SVR: support vector regression; TeoNAM: teosinte nested association mapping

1 INTRODUCTION

The popularity of an ensemble of multiple genomic prediction models (Meuwissen et al., 2001) has grown due to the consistent improvement in prediction performance for crop breeding applications (Wallach et al., 2018; McCormick et al., 2021; Heilmann et al., 2023; Kick and Washburn, 2023; Tomura et al., 2025; Washburn et al., 2025). The ensemble approach can potentially be an alternative option to investigations that focus on identifying the "best" individual genomic prediction models that can achieve the highest prediction performance across various prediction scenarios. This possibility aligns with the implication of the Diversity Prediction Theorem (Page, 2011, 2018; Messina et al., 2025) at the genome level, indicating that the prediction error of the ensemble becomes lower than the mean 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 for the analysis of diversity can be a critical concept in the research of genomic prediction.

Interpretability seeks an 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, such as standard genomic best linear unbiased prediction (GBLUP; VanRaden, 2008), 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, represented as the trait genetic architecture. The application of machine learning models for genomic prediction has also been investigated in crop breeding, considering their complex non-additive predictive algorithms. Genomic prediction algorithms capturing the interaction effects more effectively are expected to improve prediction performance (Hammer et al., 2006; Cooper et al., 2009; Messina et al., 2018; Yadav et al., 2021; Crossa et al., 2025; Tomura et al., 2025). While numerous studies have evaluated the prediction performance of machine learning models (Heslot et al., 2014; Wang et al., 2015; Montesinos-Lopez et al., 2024), the interpretability of genomic marker effects from machine learning algorithms has not been well-investigated in crop breeding due to the complex prediction mechanisms which are referred to as a "black-box" (Danilevicz et al., 2022; Escamilla et al., 2025). The combination of estimated marker effects from conventional and machine learning genomic prediction models as an ensemble needs to be interpretable at the genomic level to improve the characterisation of trait genetic architectures. Hence, the research community in genomic prediction for crop breeding can benefit from the development of a technical tool that can collectively investigate the predictive behaviour of multiple interpretable genomic prediction models at the genome level.

To compare the estimated genomic marker effects of genomic prediction models, circos plots (Krzywinski et al., 2009) can be constructed as a graphical view. 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 identified by genome-wide association studies (GWAS) in rice, with links representing epistatic interactions between genomic marker regions. Other research (Saade et al., 2016; Misra et al., 2017; Song et al., 2020) has also constructed circos

plots to effectively visualise GWAS results in crop breeding. The circos plot clearly illustrated the distribution of the key genes throughout the genome and their interactions with other genomic regions.

Here, we developed a computational tool called the Ensemble AnalySis with Interpretable Genomic Prediction (EasiGP) to analyse the genetic architectures identified by multiple interpretable genomic prediction models through circos plots. We use the Teosinte Nested Association Mapping (TeoNAM, Chen et al., 2019) experimental results for the days to anthesis (DTA) trait from Tomura et al. (2025) to demonstrate EasiGP and extend the interpretations to the genomic level. EasiGP was evaluated through three objectives: (1) Comparison of the genetic architectures with key genomic marker regions identified by previous studies, assessing the overlap with the known genetic architecture. (2) Comparison between the inferred 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 contributing to the TeoNAM. Together, we anticipate that the results obtained from pursuing these objectives will enhance the interpretability of genomic prediction models at the genome level using EasiGP.

2 MATERIALS AND METHODS

2.1 Overview of EasiGP

EasiGP has two main components, 1) an ensemble of genomic prediction models and 2) circos plots for visualisation of these models at the genetic level (Figure 1). For the genomic prediction function, N individual genomic prediction models and the ensemble of the N individual genomic prediction models were iteratively developed under M independent prediction scenarios. The prediction scenarios were

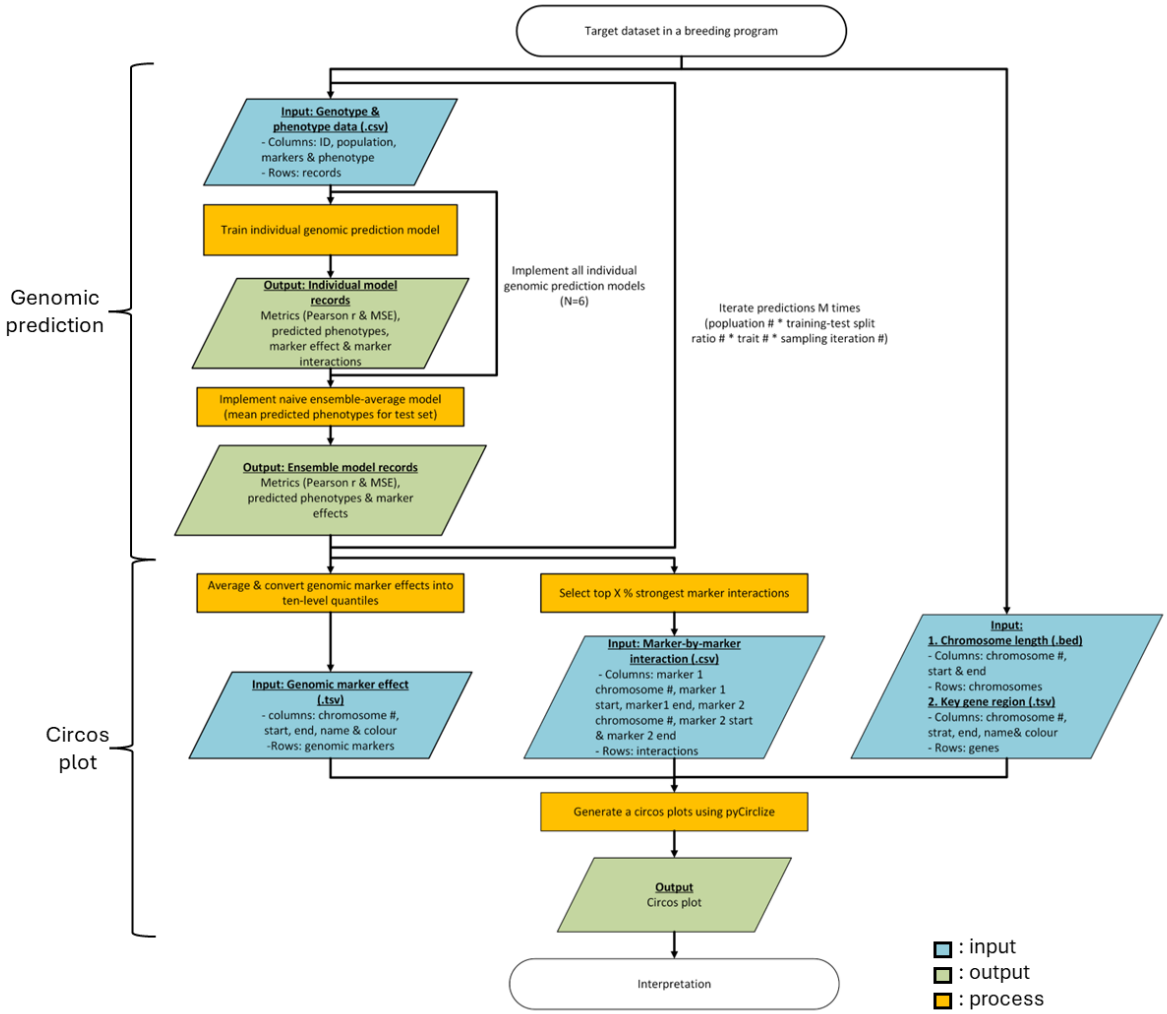


Figure 1: flowchart of EasiGP that consists of two functions; genomic prediction and circos plot development. The blue, green and orange colours represent input, output and process, respectively.

generated by the combination of the total number of sub-populations, the total number of training and test set ratios, the total number of target traits and the total number of random splits of data into the training and test sets. In each prediction scenario, the genotype and phenotype data containing the columns of identification, population, genomic markers and phenotypes with the rows of records were input into the individual genomic prediction models. After the model fitting and genomic prediction using training and test sets, respectively, the individual genomic models returned prediction performance metrics (Pearson correlation and mean square error (MSE)), predicted phenotypes for the training and test sets, genomic

133 marker effects and marker-by-marker interactions. For the ensemble model, the mean predicted phenotype
134 values of the individual genomic prediction models were calculated as the predicted phenotypes of the
135 ensemble model and returned as output of the ensemble model, accompanied by the prediction performance
136 metrics and genomic marker effects.

137 After finishing all the predictive iterations, EasiGP initiates the function for generating a circos
138 plot. The mean genomic marker effects in each prediction model were calculated and converted into the
139 ten-level quantiles. The mean marker-by-marker interaction values were also calculated, and the strongest
140 interactions were selected. Combined with chromosome length and key gene region datasets, a circos plot
141 was generated.

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

144 2.2 Genomic prediction models

145 Three classical genomic prediction models (ridge regression best linear unbiased prediction (rrBLUP),
146 BayesB and reproducing kernel Hilbert space (RKHS)) and three machine learning models (random forest
147 (RF), support vector regression (SVR) and graph attention network (GAT)) were applied as the individual
148 genomic prediction models in Tomura et al. (2025). The first five prediction models have been widely used
149 in numerous studies of genomic prediction models. Hence, genome level analysis of the five prediction
150 models may deepen the understanding of the predictive behaviour of commonly used genomic prediction
151 models for crop breeding. For GAT, the prediction mechanism is expected to differ considerably from the
152 other prediction models by incorporating datasets with a graphical representation, potentially generating a
153 different view of the trait genetic architecture, in contrast to the other selected prediction models. EasiGP
154 can be applied to any other possible genomic prediction models to help understand the predictive behaviour
155 at the genome level. The prediction mechanism of each selected genomic prediction model is explained in
156 the subsequent paragraphs.

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

$$\boldsymbol{y} = 1\boldsymbol{\mu} + \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{y} = \{y_1, y_1, \ldots, y_n\}$ is the set of predicted phenotypes, $\boldsymbol{\mu}$ is the intercept, $\boldsymbol{X}$ is the design matrix of genomic marker values, $\boldsymbol{\beta}$ is the allele substitution effect vector of the genomic markers in $\boldsymbol{X}$ and $\boldsymbol{\varepsilon} = \{\varepsilon_1, \varepsilon_1, \ldots, \varepsilon_n\}$ is the vector of residuals. In rrBLUP (Meuwissen et al., 2001), $\boldsymbol{\beta}$ is defined as $\boldsymbol{\beta} \sim N\left(0, \boldsymbol{I}, \sigma_\beta^2\right)$ where $\boldsymbol{I}$ is the identity matrix and σ_β^2 is the variance of genomic marker effects (Endelman, 2011), indicating that $\boldsymbol{\beta}$ is assumed to be normally distributed. For BayesB (Meuwissen et al., 2001), on the other hand, some values in $\boldsymbol{\beta}$ are diminished to zero with the assumption that not all the genomic markers J affect the trait phenotype. For RKHS regression (Gianola et al., 2008), the term $\boldsymbol{X\beta}$ in Equation (1) is replaced with a term $\boldsymbol{u}$ that explicitly captures the non-additive genomic marker effect by projecting the genomic markers into the Hilbert space using a kernel (del los Campos et al., 2009; Pérez and de los Campos, 2014):

$$\boldsymbol{y} = 1\boldsymbol{\mu} + \boldsymbol{u} + \boldsymbol{\varepsilon} \ \ with\ p(\boldsymbol{\mu}, \boldsymbol{u}, \boldsymbol{\varepsilon}) \propto N(\boldsymbol{u}|0, \boldsymbol{K\sigma_u^2})N(\boldsymbol{\varepsilon}|0, \boldsymbol{I\sigma_\varepsilon^2}) \tag{2}$$

where $\boldsymbol{u}$ are the vectors of random effects that are normally distributed with the variance based on the Gaussian kernel matrix ($\boldsymbol{K} = \{K(x_i, x_{i'})\}$). The kernel calculates the mean squared Euclidean distance between genotypes (x_i and $x_{i'}$):

$$K\left(x_i, x_{i'}\right) = \exp\left\{-h * \frac{\sum_{k=1}^{p}\left(x_{ik} - x_{i'k}\right)^2}{p}\right\} \tag{3}$$

where h is the bandwidth parameter that determines the speed of reduction in the covariance function as the distance between the genotype pair increases and p is the total number of genomic markers.

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

178 collection of decision trees that predict target values by tracing the binary decision mechanism using feature
179 values. The final prediction value is calculated by averaging the predicted values from each decision tree.
180 SVR (Drucker et al., 1996) determines the optimal location of a predictive hyperplane by including as many
181 data points as possible within the decision boundary assigned to each side of the hyperplane. The prediction
182 from the optimised hyperplane is expected to minimise the prediction error. GAT (Veličković et al., 2018)
183 is one of the deep learning approaches that leverages graph-structured data to predict target values. Each
184 genomic marker and each phenotype data point is represented as a node, and the genomic marker nodes
185 and phenotype nodes are connected with edges directed from the genomic marker nodes to the phenotype
186 nodes (Figure S1). The GAT application in Tomura et al. (2025) assigned no edges between the genomic
187 marker nodes. This graph structure indicates that each genomic marker affects phenotypes independently
188 and infinitesimally. GAT leverages the attention mechanism that allocates heavier weights to the genomic
189 marker nodes that strongly affect phenotypes, expecting to increase prediction performance.

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

193 The genomic prediction models were implemented in R (v1.1.4) and Python (v3.11.10) using
194 software and packages. For rrBLUP, BayesB and RKHS, the software tool BGLR (Pérez and de Los
195 Campos, 2014) was leveraged for the implementation in R. The iteration number and burn-in were set as
196 12,000 and 2,000, respectively. A Gaussian kernel was used for RKHS. The other parameters were set to
197 default. An increase in the values for the iteration number and burn-in did not change the prediction
198 performance of rrBLUP, BayesB and RKHS. Hence, the provided values were selected considering the
199 computational time. For RF and SVR, Sklearn (v1.2.2) was used to develop the models in Python. The total
200 number of decision trees was 1,000 in RF. All the other hyperparameters such as the maximum depth of
201 each tree, the maximum number of features and the maximum number of samples, were set to default. The
202 radial basis function was selected for the kernel in SVR. The rest of the parameters were set to default. For

GAT, PyTorch Geometric (v2.6.1) (Fey et al., 2019) 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 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 to default.

2.3 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 $\boldsymbol{\beta}$ 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):

$$\Phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{i\}}\left(x_{S \cup \{i\}}\right) - f_s(x_s) \right] \tag{4}$$

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. In this study, the cumulative effect of each genomic marker on the target phenotype is considered the additive genomic marker effect. Shapley values were calculated for every genotype data point (RIL record) and hence the mean 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. Shapley values were estimated from RKHS due to the inability to directly extract genomic marker effects in contrast to rrBLUP and BayesB. SHAP (v0.42.1) (Lundberg and Lee, 2017) was used for SVR in Python. For RF genomic

226 marker effects, feature importance was leveraged using the impurity-based approach (Ishwaran, 2015). This
227 approach calculates how well each feature (genomic marker) can separate data points containing similar
228 target values into the same nodes in the constituent decision trees. For GAT, Integrated Gradients was
229 leveraged as below (Sundararajan et al., 2017):

230
$$IG(x_i) \approx (x_i - x_i') * \sum_{k=1}^{m} \frac{\partial F\left(x_i' + \frac{k}{m} * (x_i - x_i')\right)}{\partial x_i} * \frac{1}{m} \quad (5)$$

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

238 Using the genomic marker effects estimated from all the individual genomic prediction models, the
239 genomic marker effects for the naïve ensemble-average model were calculated. The genomic marker effects
240 from each individual genomic prediction model were normalised per prediction scenario, and the arithmetic
241 mean of the normalised genomic marker effects across the individual genomic prediction models was used
242 as the genomic marker effects for the naïve ensemble-average model.

243 For the genomic marker-by-marker interaction effects, pairwise Shapley values were calculated
244 from RF (Equation (3)) using SHAP. Shapley values are calculated by summing all possible feature
245 combinations using conditional probability. To capture the non-additive interactions between genomic
246 markers (features), we combined the effect of a pairwise combination of markers on the target phenotype.
247 The mean genomic marker interaction effects across all prediction scenarios were used as the final genomic
248 marker interaction effects.

249 ## 2.4 Circos plot construction

250 The circos plots constructed in this study consist of three major sections: (1) key genomic marker region
251 rings representing prior knowledge from previous mapping studies, (2) genomic marker effect rings from
252 genomic prediction models, and (3) genomic marker-by-marker interaction effect links. Each ring consists
253 of segments for the chromosomes of the target species. For the maize TeoNAM (Chen et al., 2019) example
254 in this study, ten segments were developed to represent the total number of chromosomes in maize. The
255 segments were displayed in centimorgans (cM). Circos plots were generated using pyCirclize (v1.7.1) in
256 Python.

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

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

271 **Genomic Marker-by-Marker Interactions:** The top 0.01% of the most influential genomic
272 marker-by-marker pairwise Shapley scores were extracted from RF as selected interactions and displayed

273 as lines connecting the markers. The strength of the genomic marker interactions was indicated using the
274 width of the corresponding links; thicker links were assigned to stronger genomic marker interactions. The
275 genomic marker interactions were classified into two categories: genomic marker interactions within the
276 same chromosome and those between different chromosomes.

277 **2.5 Dataset**

278 The TeoNAM dataset (Chen et al., 2019) contains the genotype and phenotype records of recombinant
279 inbred lines (RILs) from the five hybrid sub-populations between the W22 maize (*Zea mays*) line and five
280 teosinte types (*Z. mays ssp. parviglumis* for TIL01, TIL03, TIL11 and TIL14 and *Z.mays ssp. Mexicana* for
281 TIL25). In each RIL sub-population, the F1 generation was developed by crossing W22 and the teosinte
282 type, backcrossed with W22 in the subsequent generation. Selfing was then conducted for the next four
283 generations. Each RIL sub-population was evaluated twice at the University of Wisconsin West Madison
284 Agricultural Research Station in a randomised complete block. The total number of genomic markers and
285 records from each of the five RIL sub-populations is summarised in Table S1.

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

291 The days to anthesis (DTA) trait was selected in this study primarily due to previous studies
292 revealing key genomic regions contributing to DTA (Buckler et al. 2009; Dong et al., 2012; Chen et al.,
293 2019; Wisser et al., 2019). This prior knowledge can be used to assess whether the genomic prediction
294 models used in this study can correctly detect known causative loci expected to segregate in the TeoNAM
295 populations. A deeper understanding of the genetic architecture of DTA using the interpretable genomic

296 prediction models can help improve the selection of genomic markers for predicting DTA, ultimately
297 benefiting prediction models used for crop yield improvement.

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

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

311 In this study, each sub-population was randomly split into a training and test set 500 times with
312 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
313 interaction effects were calculated for 7,500 different prediction scenarios (5 sub-populations * 3 ratios *
314 500 random split sampling iterations) for each genomic prediction model. For the circos plot development,
315 the mean genomic marker and interaction effects across all the 7,500 prediction scenarios were leveraged
316 to infer the genetic architecture by constructing the genomic marker effect rings and genomic marker-by-
317 marker interaction effect links in the circos plots.

318 For instance, the genomic prediction function in EasiGP may initiate from targeting the prediction
319 of DTA in population W22TIL01 with the training-test split ratio of 0.8-0.2. The preprocessed W22TIL01

sub-population data, containing 444 records and 274 genomic markers, is randomly split into training (355 records) and test sets (89 records) as the initial procedure of the phase. After training the genomic prediction models, the effect size of all 274 genomic markers and marker-by-marker interactions is estimated from all the prediction models using the respective genomic marker effect estimation methods described in the previous subsections. Such an estimation process iterates 500 times with different training-test set ratios in each prediction scenario. The mean effect size of each genomic marker and marker-by-marker interaction is calculated in the circos plot development function to represent overall genomic marker effects across different scenarios, subsequently linked to corresponding genomic marker regions to map the genomic marker effects as a circos plot.

3 RESULTS

3.1 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 2; Table 1). One of the strongest overlaps was observed in genomic marker regions containing genes involved in 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 were continuously in the top 10% of genomic marker effects across multiple prediction models. ZCN8 is a key gene for determining 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., 2018a). ZmCCT10 is another critical gene in the photoperiod pathway, which delays flowering 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

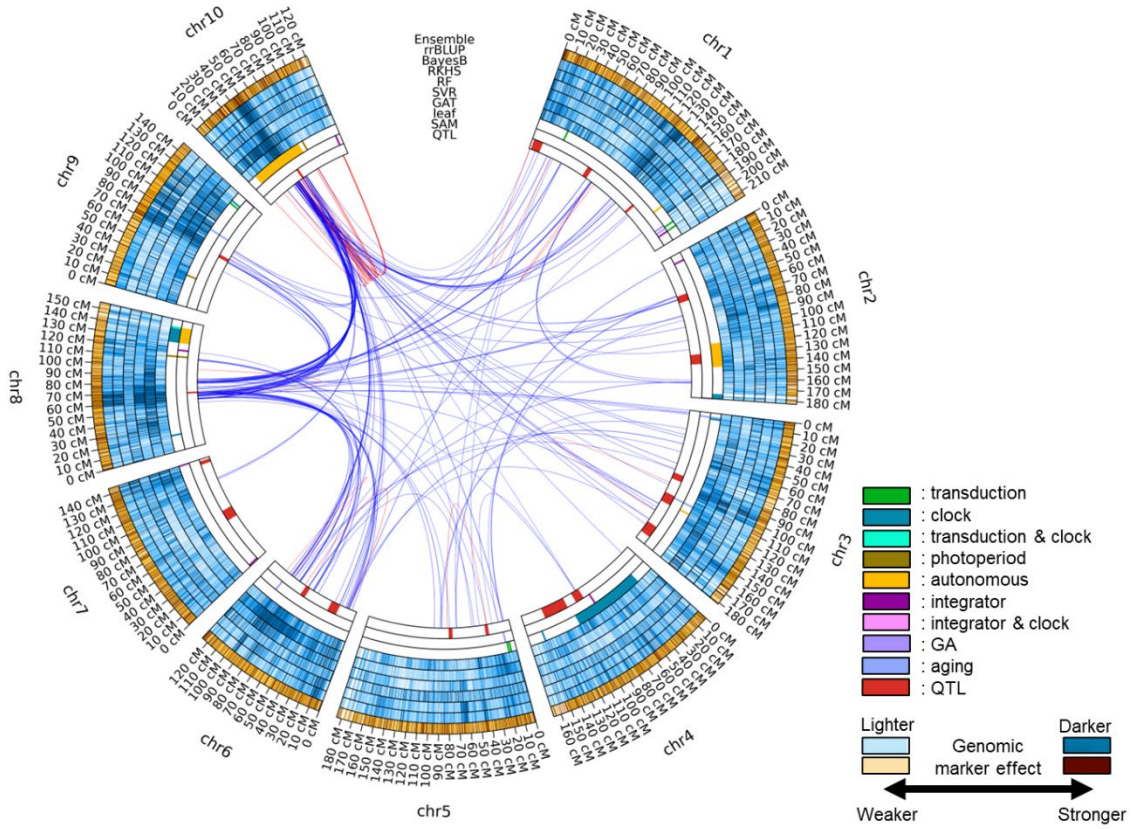


Figure 2: Circos plot for the trait days to anthesis (DTA) in the TeoNAM dataset across the five populations. The innermost ring (QTL) has the QTL genomic marker regions for DTA identified in Chen et al. (2019). The second and third innermost rings (SAM and leaf) represent the key genomic marker regions 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).

interactions between the regions of those genes, especially between chromosome 8 (ZCN8) and 10 (ZmCCT10) (Figure 2).

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

362 **Table 1:** A list of key genomic marker regions highlighted as strong features by at least two genomic prediction
363 models. The gene names and functions were based on Dong et al. (2012) and Chen et al. (2019). Genomic
364 prediction model indicates the models that continuously allocated the top 10% of genomic marker effects to the
365 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

366

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

374 Additionally, the genomic prediction models identified several key genomic marker regions in
375 chromosome 2, 7 and 10 that contain genes that function as integrators of the different pathways (ZFL1,

376 ZFL2 and dlf1). These integrators respond to the signals from the multiple pathways to determine DTA as
377 an output of the regulatory network. Hence, integrators may contain multiple different functions. ZFL1 and
378 ZFL2 are involved in the transition from the vegetative stage to the reproductive stage by upregulating the
379 ABC floral organ identity network of genes (Bomblies et al., 2003). In addition, ZFL2 is known to control
380 tassel branching and the traits relating to inflorescence structure in maize (Bomblies and Doebley, 2006).
381 dfl1 encodes a protein with a basic leucine zipper domain (Muszynski et al., 2006), expressed in the leaves
382 and regulating the timing of the transition to the flowering stage by sending signals from leaves to the SAM.

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

388 The genomic prediction models also highlighted several other genomic marker regions that have
389 not been well-studied previously (Figure 2). For instance, while the genomic marker region in chromosome
390 1 (from 125 cM to 135 cM) was highlighted by all the individual genomic prediction models besides RKHS,
391 no nearby key genomic marker regions were identified in the previous studies. Similarly, another genomic
392 marker region in chromosome 1 (from 185 cM to 190 cM) was highlighted in the inferred genetic
393 architecture, except for RKHS and SVR, but no key genomic marker region overlapped. Such discordant
394 genomic marker regions were observed in several other genomic marker regions (from 85 cM to 90 cM in
395 chromosome 2 and 45 cM and 65 cM in chromosome 5). Those highlighted genomic marker regions
396 indicate that the genomic prediction models heavily weighted the predictive information from several
397 genomic marker regions not identified in the previous studies.

398 **3.2 Each genomic prediction model captured different genetic architectures**

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

414 **3.3 Inferred genetic architectures differed among the TeoNAM sub-populations**

415 The patterns of the quantified genomic marker effects in each genomic marker region were distinctive in
416 each of the five sub-populations composing the TeoNAM dataset (Figure 3). For instance, while the
417 genomic marker region containing ZmCCT10 was highlighted in chromosome 10 across the sub-
418 populations, the range of the highlighted region varied among sub-populations. For W22TIL01 and
419 W22TIL25, the highlighted region was between 40 cM and 50 cM (Figure 3a, 3e), whereas the region was
420 extended in W22TIL03 (between 45 cM and 55 cM) and W22TIL14 (between 30 cM and 55 cM) (Figure
421 3b, 3d). The region for W22TIL11 was between 40cM and 55cM (Figure 3c) with a lower frequency of
422 highlights compared to the others. Similarly, the genomic marker region between 85 cM and 95 cM in

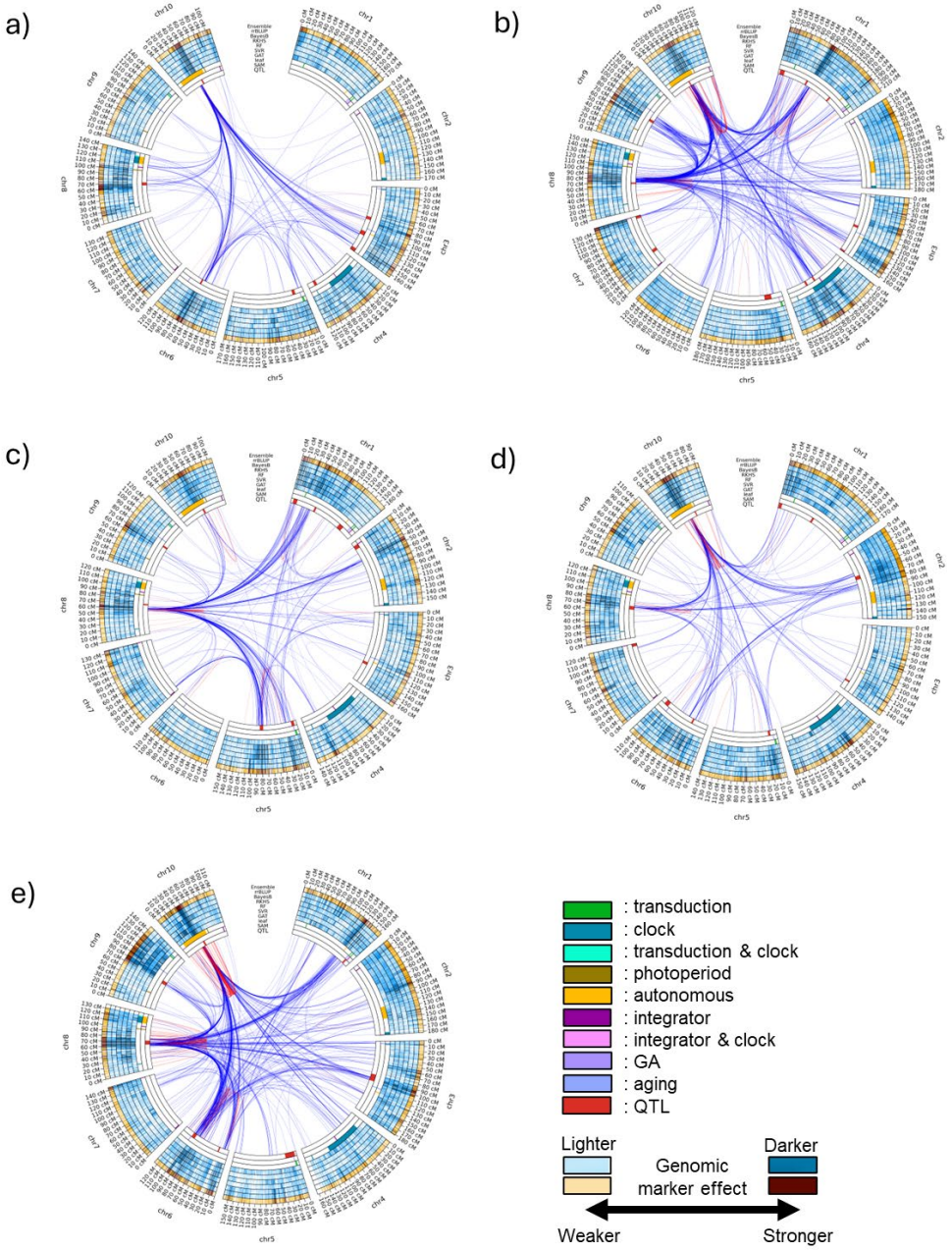


Figure 3: Circos plots for days to anthesis (DTA) in the TeoNAM dataset for the five sub-populations: a) W22TIL01, b) W22TIL03, c) W22TIL11, d) W22TIL14 and e) W22TIL25. The innermost ring (QTL) has the QTL genomic marker regions identified in Chen et al. (2019). The second and third innermost rings (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).

chromosome 1 was mainly allocated to the top 10% of the genomic marker effects by all the genomic prediction models in W22TIL03 (Figure 3b). However, the genomic prediction models in the other four sub-populations did not frequently allocate such high genomic marker effects 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 3b), the genomic prediction models did not clearly highlight the same genomic marker region in the four other sub-populations. A lack of consensus in the quantification patterns of the genomic marker effect among the five TeoNAM sub-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 sub-populations comprising the TeoNAM dataset.

Such diversity in the marker effect combinations among the six genomic prediction models for each sub-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 sub-populations, many of the detected marker-by-marker interaction regions differed among the five sub-populations. The region for chromosome 8 was located at 70 cM in W22TIL01, W22TIL14 and W22TIL25 (Figure 3a, 3d, 3e), whereas the region was located at 80 cM in W22TIL03 and 60 cM in W22TIL11 (Figure 3b, 3c). For chromosome 10, the region was located between 40 cM and 50 cM in W22TIL01 and W22TIL25 (Figure 3a, 3e), between 50 cM and 60 cM in W22TIL03 (Figure 3b), between 45 cM and 55 cM in W22TIL11 (Figure 3c) and between 30 cM and 45 cM in W22TIL14 (Figure 3d). Additionally, this region in chromosome 8 strongly interacted with other chromosomes in unique patterns identified across the five sub-populations. While all

458 the sub-populations showed interactions with chromosome 5, this was stronger than for the other sub-
459 populations in W22TIL11 (Figure 3c). Three sub-populations (W22TIL03, W22TIL11 and W22TIL25)
460 showed heavy marker-by-marker interactions associated with chromosome 1 and chromosome 3 (Figure
461 3b, 3c, 3e), whereas there were strong interactions associated with chromosome 2 for W22TIL11 and
462 W22TIL14 (Figure 3c, 3d). Such distinctive marker-by-marker interaction patterns were observed
463 throughout the chromosomes. Those results indicate that the inferred views of trait genetic architecture
464 were diverse across the five sub-populations.

465 **4 DISCUSSION**

466 **4.1 EasiGP helps investigate genomic prediction models at the genome level**

467 The circos plots generated by EasiGP provided a genome level view of how the six investigated genomic
468 prediction models estimated diverse genomic marker effects (Figure 2). This diversity is attributed to
469 algorithmic differences among the prediction models. The algorithms of rrBLUP and BayesB primarily
470 targeted the detection of the additive genomic marker effects (Meher et al., 2022), while RKHS also
471 included a kernel to explicitly identify several non-additive genomic marker-by-marker interaction effects
472 (Gianola et al., 2008). The machine learning models included components in their prediction algorithms to
473 identify more complex non-additive prediction patterns underlying the inclusion of the genetic markers
474 (Chen et al., 2023). The comparison of the estimated feature importance from logistic regression (additive
475 prediction model) and RF (non-additive prediction model) indicated that several identified key features
476 varied between the prediction models for the classification tasks in the cancer study (Saarela and Jauhiainen,
477 2021). A linear regression model was also compared with RF in regression tasks, showing a similar
478 conclusion that the level of importance in each feature identified by the linear regression was not necessarily
479 aligned with that of RF (Grömping, 2009). The estimated patterns in the feature importance were also
480 investigated within three machine learning models (multilayer perceptron (MLP), naïve Bayes and
481 Bagging), with the observation that each prediction model weighted feature importance distinctively

482 (Štrumbelj and Kononenko, 2014). Hence, the different prediction strategies in each prediction algorithm
483 quantified the genomic marker effects uniquely throughout the chromosomes.

484 Such differences in genome profiles of the captured genomic marker effects were represented as
485 diverse views of the genetic architecture of the trait DTA in our study using the TeoNAM dataset (Figure
486 2, Table 1). The circos plot enabled the visualisation of diverse patterns in the colour intensity across the
487 chromosomes that were not identical to any other views. The combination of such diverse views in the
488 ensemble approach provides a novel view of marker weights, taking into consideration the consistent
489 performance improvement of the ensemble approach observed in Tomura et al. (2025). The weaknesses of
490 individual prediction algorithms can potentially be offset by aggregating the diversely quantified genomic
491 marker effects from each model (Kick and Washburn, 2023; Washburn et al., 2025). The interpretation of
492 the visualised genetic architecture for the ensemble is supported by the Diversity Prediction Theorem (Page,
493 2018), implying that diversified predicted phenotypes through differently quantified genomic marker
494 effects contributed to the reduction of the ensemble error for the naïve ensemble-average model compared
495 to the mean model error. This implication is also aligned with the explanation of the No Free Lunch
496 Theorem (Wolpert and Macready, 1997), stating that no individual prediction model can consistently reach
497 the highest prediction performance across the prediction scenarios. The investigation of the genomic
498 prediction models at the genomic level deepens our understanding of the possible reasons for the prediction
499 performance improvement by the ensemble. The circos plot visualises how an ensemble can leverage the
500 diversity of prediction algorithms to model the high dimensionality of the genotype-to-phenotype state
501 space for complex traits.

502 The application of both interpretable prediction models and circos plots has been reported in several
503 studies, primarily in studies of the human genome, using the self-attention mechanism and statistical
504 analysis (Lee et al., 2023; Peng et al., 2023; Prakash and Banerjee, 2023). However, this approach has not
505 been widely applied to the genome-level study of genomic prediction models for crop breeding applications.

506 Hence, we have developed EasiGP to enable the construction of circos plots that visualise genomic marker
507 effects from interpretable genomic prediction models and their ensemble.

508 **4.2 EasiGP may reveal potential new key genomic marker regions**

509 The circos plot view of the six genomic prediction models and the ensemble highlighted several genomic
510 marker regions that were not previously identified as key genomic regions (Figure 2). Those regions did
511 not overlap with key gene regions reported from previous studies and could potentially be leveraged as new
512 genomic marker regions that can contribute to improving target traits. There are multiple possible reasons
513 why these new regions were identified by the genomic prediction models rather than by traditional QTL
514 mapping. Those reasons can be attributed to the fundamental distinction that traditional QTL analysis
515 leverages different approaches for QTL detection than the interpretable genomic prediction (Holland, 2004;
516 Alonso et al., 2006). This distinction is largely driven by QTL analysis focusing on testing individual loci
517 for significance, whereas interpretable genomic prediction models the entire genome simultaneously.

518 The use of traditional and new approaches can contribute to identifying more comprehensive sets
519 of key genomic markers for target traits. The combination of both QTL analysis and interpretable prediction
520 models (machine learning) has started being evaluated in crop breeding programs targeting soybeans (Gill
521 et al., 2022; Yoosefzadeh et al., 2022). Since the use of interpretable genomic prediction for identifying key
522 genomic marker regions has been a relatively recent approach in crop breeding, it is vital to verify the
523 reliability through empirical studies and simulations. Nevertheless, there is potential for the collective
524 analysis of key genomic marker regions to enhance the understanding of trait genetic architecture that can
525 support prediction applications for breeding.

526 **4.3 Distinctive genetic architecture in each population may hinder accurate between-**
527 **population prediction**

528 The genomic prediction models identified different key patterns in the genomic marker and marker-by-
529 marker interaction effects across the five sub-populations comprising the TeoNAM dataset (Figure 3).

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

544 The differences in the genetic architecture for each sub-population may hinder the between-sub-
545 population prediction with high accuracy. Each genomic prediction model learns key prediction patterns
546 from the genetic architecture of the training set. However, the learnt prediction patterns may not be aligned
547 with the key prediction patterns in the test set from another sub-population due to differences in the standing
548 genetic variation. For example, in wheat, the prediction performance of the between-population prediction
549 for 19 traits was considerably lower than the within-population prediction (Huang et al., 2018b). This poor
550 prediction performance was attributed to a lack of lines closely related to the test set. The performance of
551 the between-population prediction for sorghum depended on the combination of the populations for the
552 training and test sets (Sapkota et al., 2020). These experimental results from previous studies indicate that
553 the differences in the genetic architecture between the training and test sets considerably affect the
554 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.

5 CONCLUSIONS

EasiGP supports visualisation of the inferred genetic architectures from the six interpretable individual genomic prediction models reported in Tomura et al. (2025) through an application of circos plots. The generated 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 TeoNAM sub-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. Through application to appropriate training sets, EasiGP can deepen the understanding of each interpretable 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

EasiGP is available at https://github.com/ShunichiroT/EasiGP. The original TeoNAM dataset (Chen et al., 2019), leveraged to infer the genomic marker and interaction effects, is available at https://datacommons.cyverse.org/brow-se/iplant/home/shared/panzea/genotypes/GB-S/TeosinteNAM for the genotypes and https//gsajournals.figshare.com/articles/dataset/Supplemental_Material_for_Chen_et_al_2019/9250682 for the phenotypes.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Shunichiro Tomura: Formal analysis, Investigation, Methodology, Software, Validation, Visualisation, Writing – original draft and Writing – review & editing; Melanie J. Wilkinson: Validation, Supervision and Writing – review & editing; Owen Powell: Investigation, Supervision and Writing – review & editing; Mark Cooper: Conceptualization, Investigation, Methodology, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review & editing

ORCID

Shunichiro Tomura: 0009-0005-9050-9963; Melanie J. Wilkinson: 0000-0003-1400-3176; Owen Powell: 0000-0001-5131-4286; Mark Cooper: 0000-0002-9418-3359

ACKNOWLEDGEMENTS

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. We also thank Mr. Greg McLean for constructive feedback to improve the implementation of EasiGP on GitHub.

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).

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.

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.

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.

Danilevicz MF, Gill M, Anderson R, Batley J, Bennamoun M, Bayer PE, Edwards D. (2022). Plant genotype to phenotype prediction using machine learning. Frontiers in Genetics. 13:822173. https://doi.org/10.3389/fgene.2022.822173.

de los Campos G, Gianola D, Rosa GJ. (2009). Reproducing kernel hilbert spaces regression: a general framework for genetic evaluation. Journal of animal science. 87:1883–1887. https://doi.org/10.2527/jas.2008-1259.

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.

Escamilla DM, Li D, Negus KL, Kappelmann KL, Kusmec A, Vanous AE, Schnable PS, Li X, Yu J. (2025). Genomic selection: Essence, applications, and prospects. The Plant Genome. 18:e70053. https://doi.org/10.1002/tpg2.70053.

Fey M, Lenssen JE. (2019). Fast graph representation learning with pytorch geometric. arXiv preprint arXiv:1903.02428. https://doi.org/10.48550/arXiv.1903.02428.

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.

677 Guo L, Wang X, Zhao M, Huang C, Li C, Li D, Yang CJ, York AM, Xue W, Xu G *et al*. (2018). Stepwise
678 cis-regulatory changes in zcn8 contribute to maize flowering-time adaptation. Current biology.
679 28:3005–3015. https://doi.org/10.1016/j.cub.2018.07.029.

680 Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S, Podlich D (2006).
681 Models for navigating biological complexity in breeding improved crop plants. Trends in Plant Science
682 11(12):587-593. https://doi.org/10.1016/j.tplants.2006.10.006.

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

686 Heilmann PG, Frisch M, Abbadi A, Kox T, Herzog E. (2023). Stacked ensembles on basis of parentage
687 information can predict hybrid performance with an accuracy comparable to marker-based gblup.
688 Frontiers. https://doi.org/10.3389/fpls.2023.1178902.

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

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

693 Huang C, Sun H, Xu D, Chen Q, Liang Y, Wang X, Xu G, Tian J, Wang C, Li D *et al*. (2018a). Zmcct9
694 enhances maize adaptation to higher latitudes. Proceedings of the National Academy of Sciences.
695 115:E334–E341. https://doi.org/10.1073/pnas.1718058115.

696 Huang M, Ward B, Griffey C, Van Sanford D, McKendry A, Brown-Guedira G, Tyagi P, Sneller C.
697 (2018b). The accuracy of genomic prediction between environments and populations for soft wheat traits.
698 Crop Science. 58:2274–2288. https://doi.org/10.2135/cropsci2017.10.0638.

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

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

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

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

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

713 Loshchilov I, Hutter F. (2017). Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.
714 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.

McCormick RF, Truong SK, Rotundo J, Gaspar AP, Kyle D, Van Eeuwijk F, Messina CD. (2021). Intercontinental prediction of soybean phenology via hybrid ensemble of knowledge-based and datadriven models. in silico Plants. 3:diab004. https://doi.org/10.1093/insilicoplants/diab004.

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. Theoretical and Applied Genetics. https://doi.org/10.1007/s00122-025-04928-6.

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.

Misra G, Badoni S, Anacleto R, Graner A, Alexandrov N, Sreenivasulu N. (2017). Whole genome sequencing-based association study to unravel genetic architecture of cooked grain width and length traits in rice. Scientific Reports. 7:12478. https://doi.org/10.1038/s41598-017-12778-6.

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 (2011). *Diversity and Complexity*. Princeton University Press.

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.

Saade S, Maurer A, Shahid M, Oakey H, Schmöckel SM, Negrão S, Pillen K, Tester M. (2016). Yield-related salinity tolerance traits identified in a nested association mapping (nam) population of wild barley. Scientific reports. 6:32586. https://doi.org/10.1038/srep32586.

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.

Song JM, Guan Z, Hu J, Guo C, Yang Z, Wang S, Liu D, Wang B, Lu S, Zhou R *et al*. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of brassica napus. Nature plants. 6:34–45. https://doi.org/10.1038/s41477-019-0577-7.

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. 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.

VanRaden PM. (2008). Efficient methods to compute genomic predictions. Journal of dairy science. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.

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.

Wallach D, Martre P, Liu B, Asseng S, Ewert F, Thorburn PJ, van Ittersum M, Aggarwal PK, Ahmed M, Basso B *et al*. (2018). Multimodel ensembles improve predictions of crop–environment–management interactions. Global Change Biology. 24. https://doi.org/10.1111/gcb.14411.

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. https://doi.org/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.

825 **Supplementary materials**

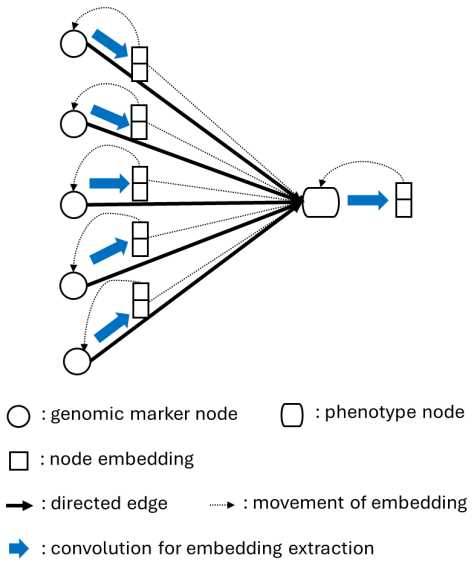
826

827 **Table S1:** The total number of genomic markers (SNPs) and the total number of records obtained from
828 recombinant inbred lines (RILs) scored twice in each population, created from Tomura et al. (2025).
829 "Imputation" indicates the total number of SNPs after excluding SNPs containing more than 10% of missing
830 marker calls during the missing allele imputation process with the most frequent alleles. Subsequently,
831 remaining SNPs were filtered through linkage disequilibrium (LD) with a threshold of 0.8. The total number of
832 final SNPs after the "LD filtering" depends on the combination of RILs included in the training set in each
833 prediction scenario.

834

Population name	SNPs			Total record
	Original	Imputation	LD filtering	
W22TIL01	13,089	13,042	274 - 322	444
W22TIL03	16,110	16,076	295 - 342	540
W22TIL11	13,188	13,153	268 - 314	438
W22TIL14	11,396	11,375	270 - 320	464
W22TIL25	14,885	14,857	294 - 341	618

835
836
837
838



839

840 **Figure S1**: Simplified illustrative view of a graph attention network (GAT) in this study. Each genomic marker
841 node was connected to the phenotype node with an edge directed from the marker nodes to the phenotype node.
842 Key predictive information was extracted from each marker node as an embedding in a vector format through
843 convolution, subsequently gathered with weights as attentions to calculate the embedding of the phenotype node.
844 The embedding of each node was reassigned to the corresponding node as the initial node information for the
845 next embedding extraction (hidden) layer. The embedding size was determined by the number of channels in

each hidden layer (2 in this example). At the last (outer) layer, the size of the embedding was set as 1. The embedding of the phenotype node at the last layer is returned from GAT as a predicted phenotype.