

**Robustly enhancing crop genomic prediction accuracy through ensemble
learning and iterative optimization**

Yao et al.

Supplementary Note 1. Performance evaluation based on dataset partitioning by population structure

We further evaluated the performance of GEG2P under a data partitioning scheme stratified by population structure, and we selected rrBLUP as the comparison method. Following Liu *et al.*¹, we divided the CUBIC population into one reference population (Reference) and four subpopulations (L1, L2, L3, and L4), containing 1,140, 67, 64, 63, and 70 lines, respectively. The genetic distances between the four subpopulations and the reference population, ranked from greatest to smallest, were L4, L3, L2, and L1. Each subpopulation was sequentially used as the test set, while the remaining populations were used as the training set to evaluate the prediction accuracy of GEG2P and rrBLUP for DTA, PH, and EW (Supplementary Data 3). The results showed that prediction accuracy of the subpopulations of L4, L3, L2, and L1 was positively correlated with their genetic relatedness to the reference population. In addition, GEG2P achieved higher prediction accuracy than rrBLUP and showed stronger stability for all the traits.

Supplementary Note 2. Pseudocode of three-step iterative optimization strategy

Algorithm 1. Three-step iterative optimization strategy of GEG2P

Input: Base Learner Set B , GenoType Data G , Trait Name T , Genomic Algorithm Parameters C ,
Phenotype Data P , Base Learner Labels L

Output: GEG2Pv3Model M , Saved Base Learners S , Prediction Values R ,
Weights of Base Learners W

Step 1: Constructing GEG2P(v1) to break category boundary.

- 1: $BLModels \leftarrow \text{TrainBaseLearners}(B, G, P, T)$
- 2: $GEG2Pv1Model, BLWeights \leftarrow \text{CreateGEG2PwithGA}(BLModels, G, P, T, C)$
- 3: $Predv1 \leftarrow \text{EnsemblePredictions}(GEG2Pv1Model, BLWeights, G, P, T)$

Step 2: Constructing GEG2P(v2) to use composite base learners.

- 4: $DL \leftarrow \text{SelectDeepLearningBL}(B, L)$
- 5: $ML \leftarrow \text{SelectMachineLearningBL}(B, L)$
- 6: $SS \leftarrow \text{SelectStatisticalBL}(B, L)$
- 7: $DLMoels \leftarrow \text{TrainBaseLearners}(DL, G, P, T)$
- 8: $GEG2PDLModel, DLWeights \leftarrow \text{CreateGEG2PwithGA}(DLMoels, G, P, T, C)$
- 9: $MLModels \leftarrow \text{TrainBaseLearners}(ML, G, P, T)$
- 10: $GEG2PMLModel, MLWeights \leftarrow \text{CreateGEG2PwithGA}(MLModels, G, P, T, C)$
- 11: $SSModels \leftarrow \text{TrainBaseLearners}(SS, G, P, T)$
- 12: $GEG2PSSModel, SSWeights \leftarrow \text{CreateGEG2PwithGA}(SSModels, G, P, T, C)$
- 13: $BLModelsv2 \leftarrow \{ BLModels, GEG2Pv1Model, GEG2PDLModel, GEG2PMLModel, GEG2PSSModel \}$
- 14: $GEG2Pv2Model, BLWeightsv2 \leftarrow \text{CreateGEG2PwithGA}(BLModelsv2, G, P, T, C)$
- 15: $Predv2 \leftarrow \text{EnsemblePredictions}(GEG2Pv2Model, BLWeightsv2, G, P, T)$

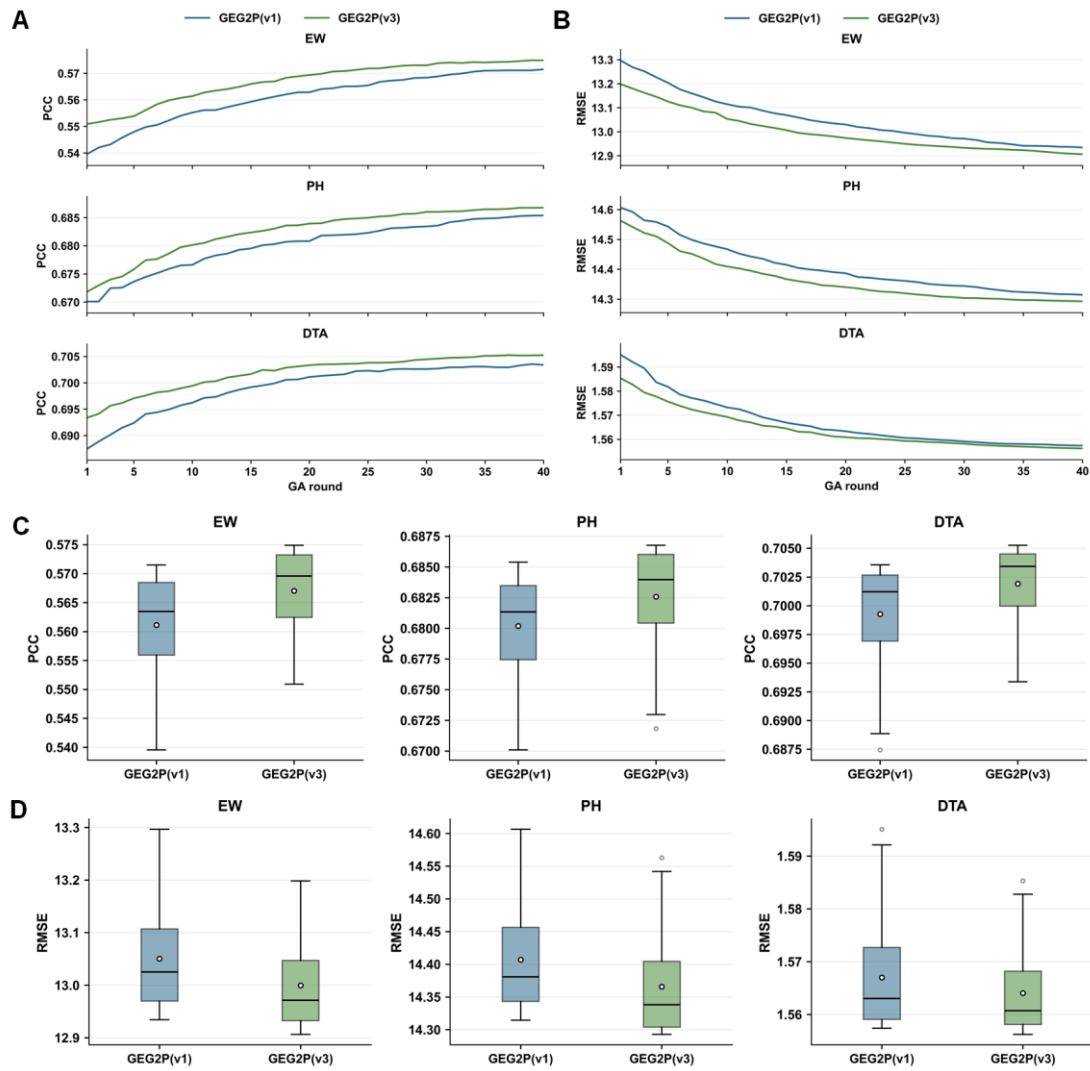
Step 3: Constructing GEG2P(v3) to remove base learners with poor accuracy.

- 16: $CandidateBL \leftarrow \{ BLModels, GEG2Pv1Model, GEG2PDLModel, GEG2PMLModel, GEG2PSSModel \}$
 - 17: $BestScore \leftarrow -\infty$; $SavedBL \leftarrow CandidateBL$; $isImproved \leftarrow \text{True}$
 - 18: **while** $isImproved$ and $|SavedBL| \geq 2$ **do**
 - 19: $GEG2Pv3Model, BLWeightsv3 \leftarrow \text{CreateGEG2PwithGA}(SavedBL, G, P, T, C)$
 - 20: $Predv3 \leftarrow \text{EnsemblePredictions}(GEG2Pv3Model, BLWeightsv3, G, P, T)$
 - 21: $Score \leftarrow \text{Metric}(Predv3, P)$
 - 22: **if** $Score > BestScore$ **then**
 - 23: $BestScore \leftarrow Score$
-

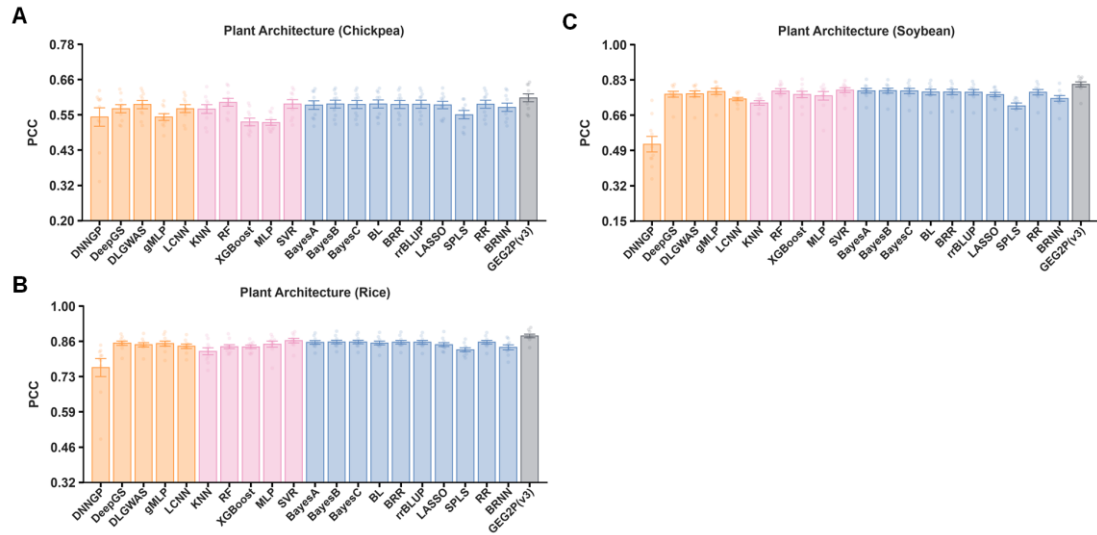
```

24:         SavedBL  $\leftarrow$  CandidateBL
25:     else
26:         isImproved  $\leftarrow$  False
27:         break
28:     end if
29:     Contrib  $\leftarrow$  EvaluateSingleBL(CandidateBL,  $G, P, T$ )
30:     WorstBL  $\leftarrow$  SelectWorstBL(Contrib)
31:     CandidateBL  $\leftarrow$  RemoveBL(SavedBL, WorstBL)
32: end while
33: GEG2Pv3Model, BLWeightsv3  $\leftarrow$  CreateGEG2PwithGA(SavedBL,  $G, P, T, C$ )
34: Predv3  $\leftarrow$  EnsemblePredictions(GEG2Pv3Model, BLWeightsv3,  $G, P, T$ )
35:  $M \leftarrow$  GEG2Pv3Model;  $S \leftarrow$  SavedBL;  $R \leftarrow$  Predv3;  $W \leftarrow$  BLWeightsv3
36: return  $M, S, R, W$ 

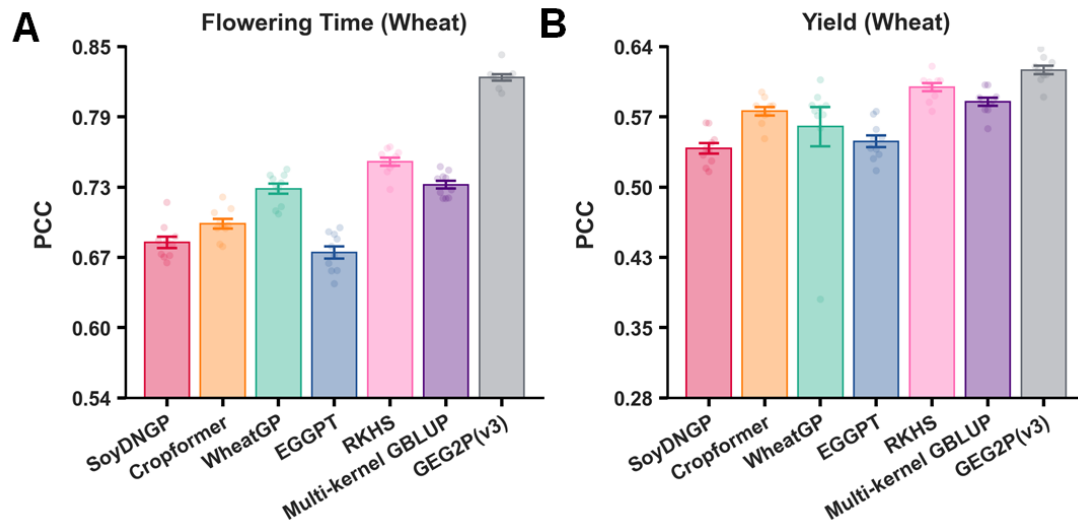
```



Supplementary Figure 1. Comparison of the predictive performance of GEG2P(v1) and GEG2P(v3) across different iteration rounds of genetic algorithm. (A) Trends in PCC of GEG2P(v1) and GEG2P(v3) across iteration rounds for EW, PH, and DTA in maize. (B) Trends in RMSE of GEG2P(v1) and GEG2P(v3) across iteration rounds for EW, PH, and DTA in maize. (C) PCC distributions of GEG2P(v1) and GEG2P(v3) for EW, PH, and DTA in maize. (D) RMSE distributions of GEG2P(v1) and GEG2P(v3) for EW, PH, and DTA in maize. Colors denote GEG2P(v1) (blue) and GEG2P(v2) (green). For each box, middle horizontal lines indicate medians, and white dots indicate means; the upper and lower boundaries indicate the interquartile range (IQR, 25th-75th percentiles); whiskers extend to the smallest and largest non-outlier values within $1.5 \times \text{IQR}$ from the box boundaries; and points represent outliers. Source data are provided as a Source Data file.

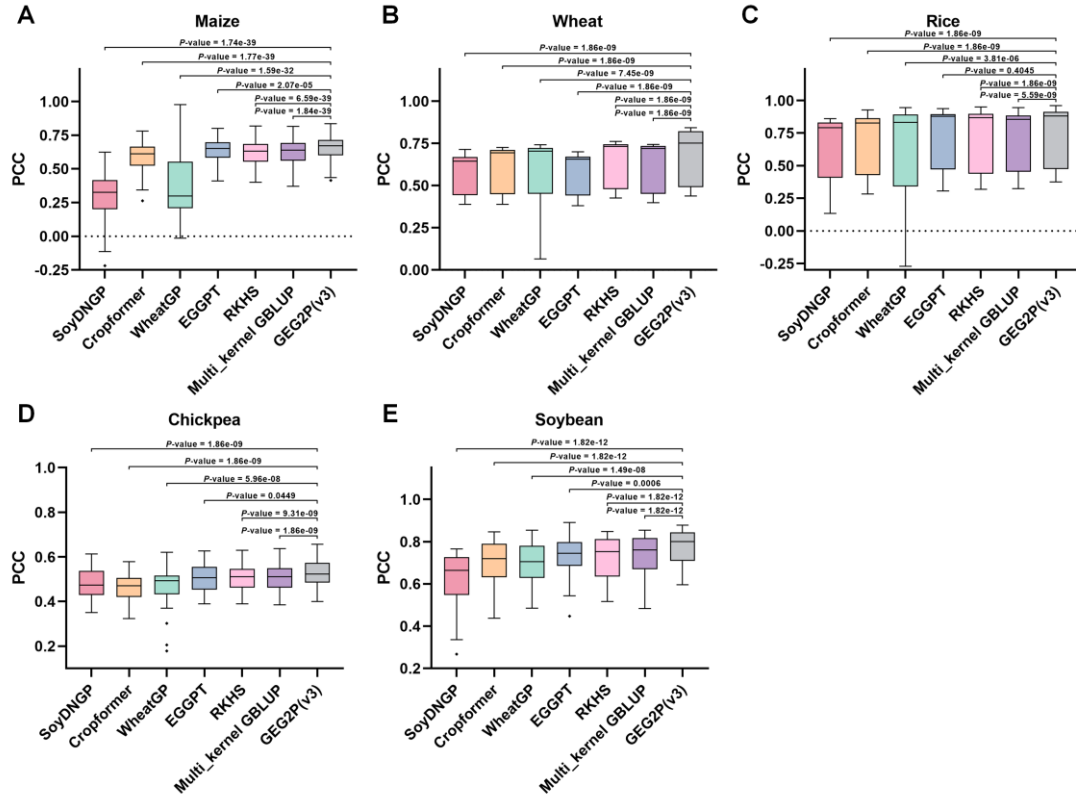


Supplementary Figure 2. Phenotype prediction accuracy of GEG2P and twenty base learners for plant architecture-related traits in different species. (A) Chickpea. (B) Rice. (C) Soybean. Bars and error bars indicate average Pearson correlation coefficients (PCCs) and standard errors, respectively (n=10). Bar colors denote deep-learning base learners (orange), machine-learning base learners (pink), statistical base learners (blue), and GEG2P (grey). Source data are provided as a Source Data file.

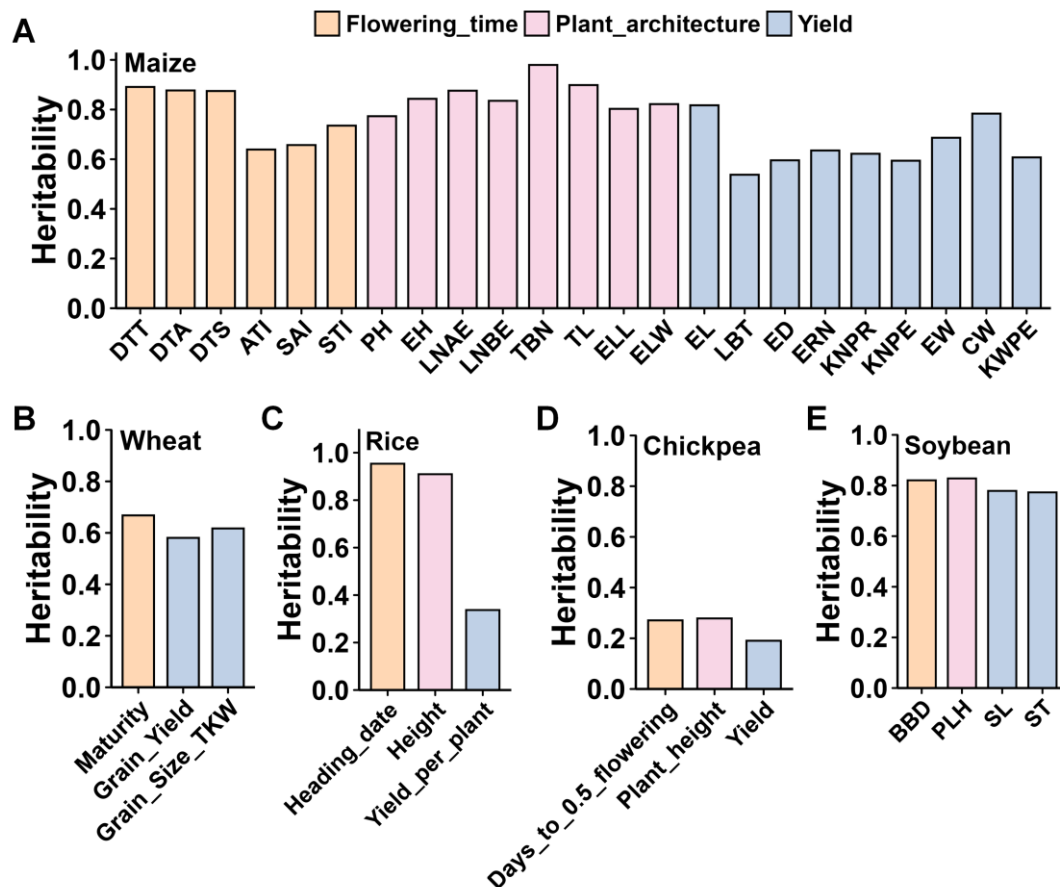


Supplementary Figure 3. Phenotype prediction accuracy of GEG2P and six SOTA methods in wheat.

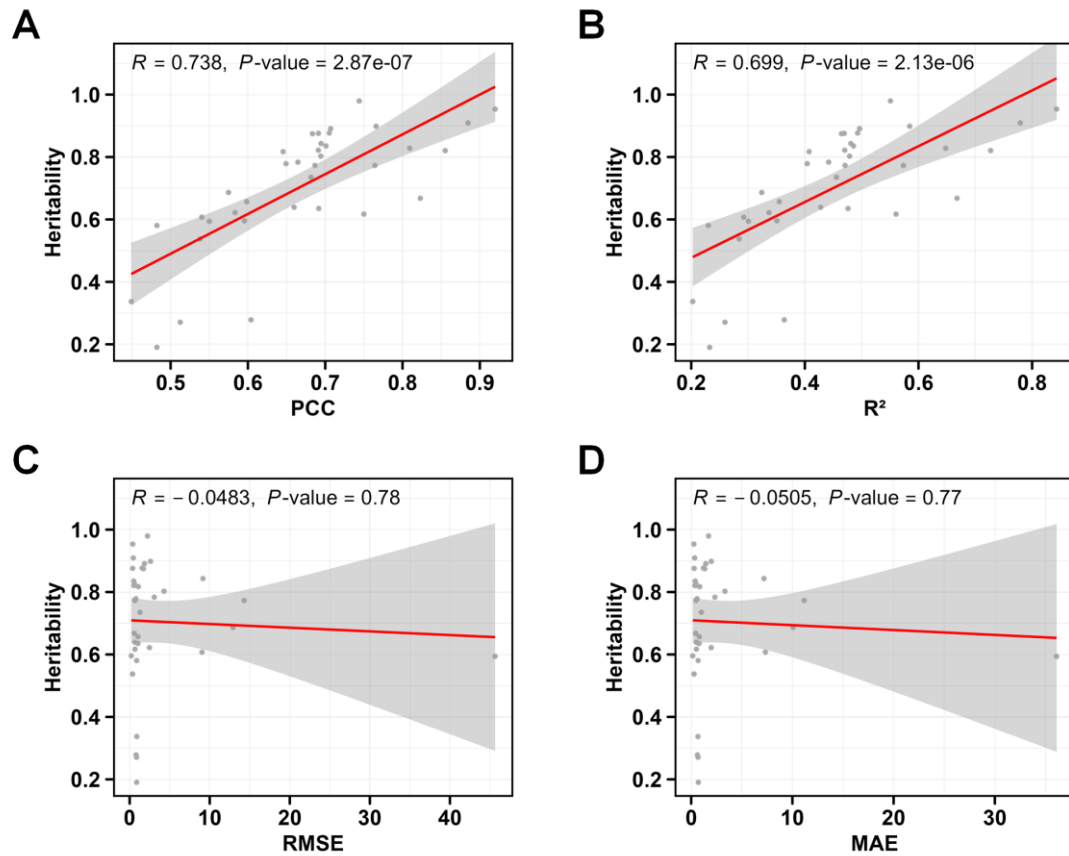
(A) Flowering time-related traits. (B) Yield-related traits. Bars and error bars indicate average Pearson correlation coefficients (PCCs) and standard errors, respectively (n=10). Bar colors denote SoyDNGP (red), Cropformer (orange), WheatGP (green), EGGPT (blue), RKHS (pink), multi-kernel GBLUP (purple), and GEG2P(v3) (grey). Source data are provided as a Source Data file.



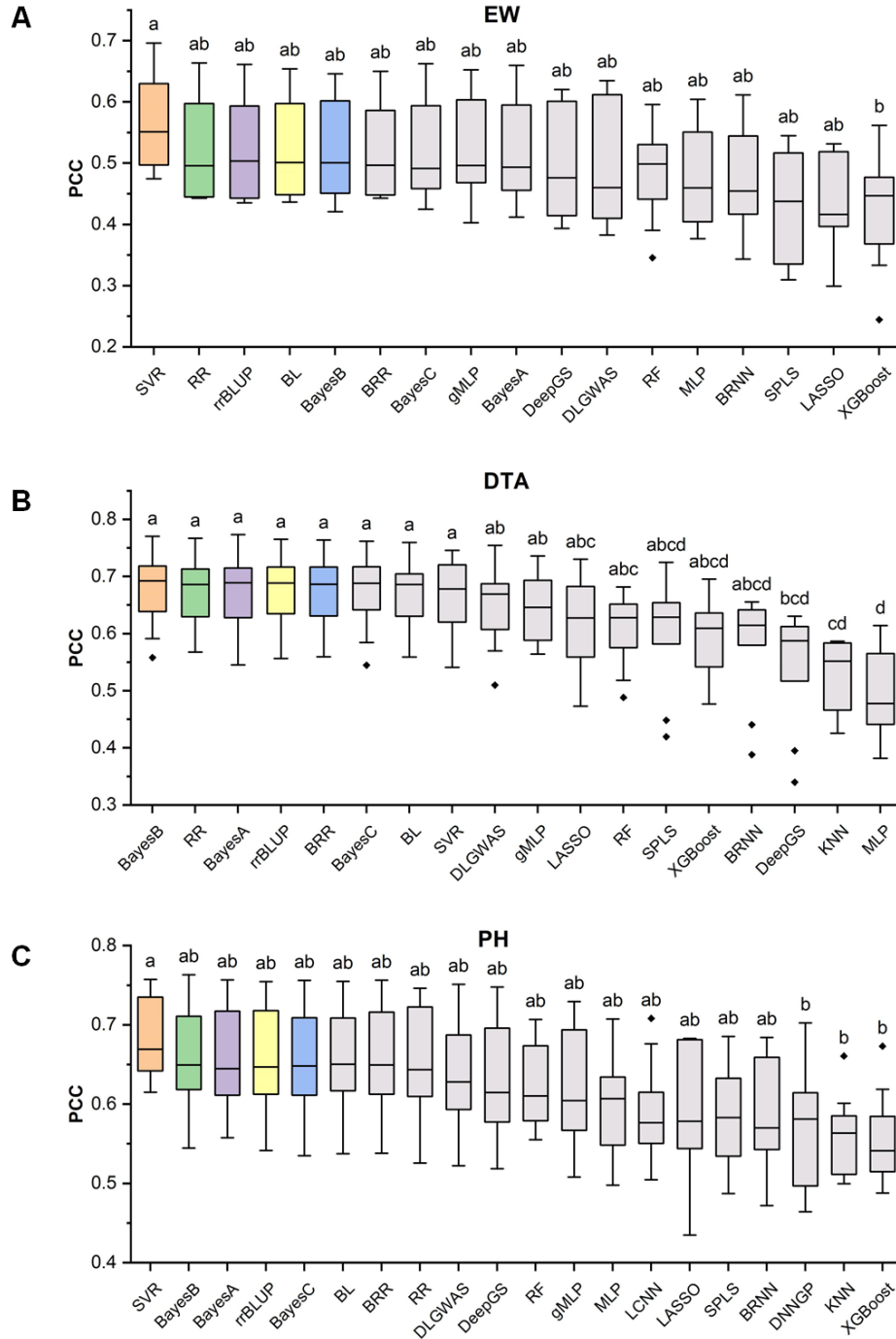
Supplementary Figure 4. Statistical comparison results of prediction accuracy between GEG2P and six SOTA methods across all the traits of five species. PCC comparisons under 10-fold cross-validation for (A) maize (23 traits, $n = 230$), (B) wheat (3 traits, $n = 30$), (C) rice (3 traits, $n = 30$), (D) chickpea (3 traits, $n = 30$), and (E) soybean (4 traits, $n = 40$). P -values are from two-sided Wilcoxon signed-rank test. For each box, middle horizontal lines indicate medians; the upper and lower boundaries indicate the interquartile range (IQR, 25th-75th percentiles); whiskers extend to $1.5 \times \text{IQR}$ from the box boundaries; and points represent outliers. Box plot colors: rose, SoyDNGP; orange, Cropformer; green, WheatGP; blue, EGGPT; pink, RKHS; purple, Multi_kernel GBLUP; grey, GEG2P(v3). Source data are provided as a Source Data file.



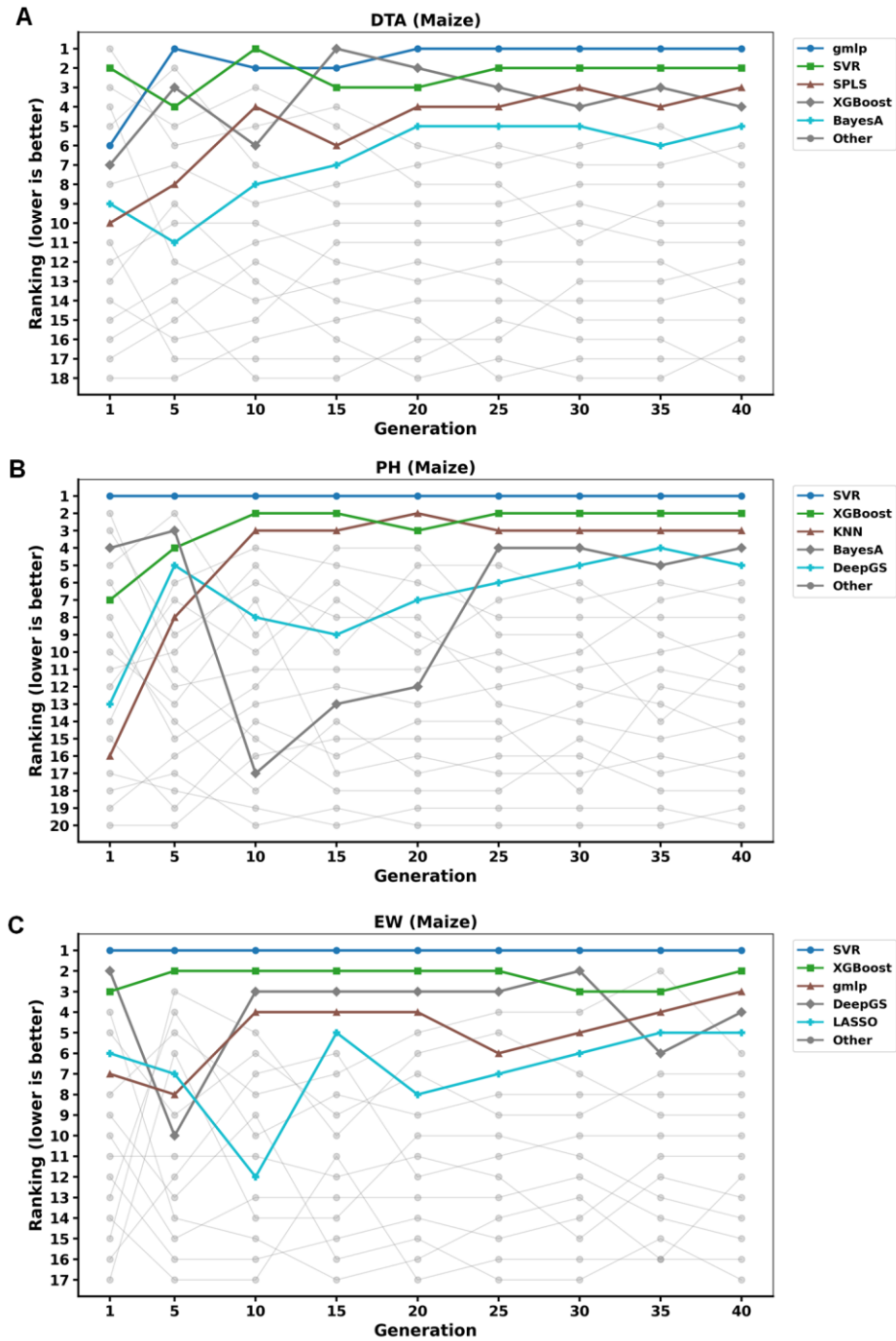
Supplementary Figure 5. Heritability of 36 traits in maize, wheat, rice, chickpea, and soybean. (A) Heritability of 23 traits related to flowering time, plant architecture, and yield in maize. (B) Heritability of 3 traits related to flowering time and yield in wheat. (C) Heritability of traits related to flowering time, plant architecture, and yield in rice. (D) Heritability of traits related to flowering time, plant architecture, and yield in chickpea. (E) Heritability of 4 traits related to flowering time, plant architecture, and yield in soybean. Bar plot colors: orange, flowering time-related traits; pink, plant architecture-related traits; blue, yield-related traits. Source data are provided as a Source Data file.



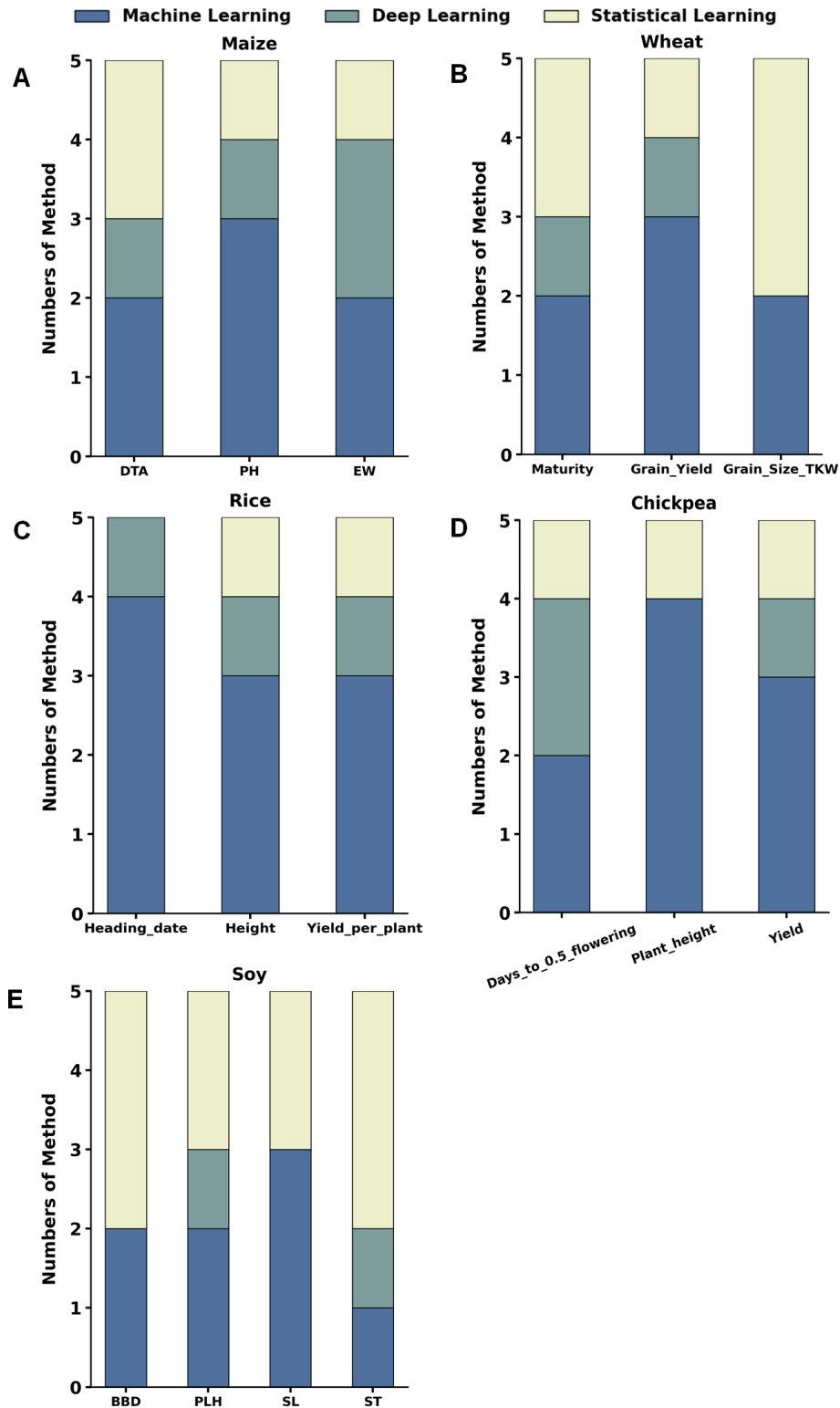
Supplementary Figure 6. Correlation analysis results between the PCC, R^2 , RMSE, and MAE of GEG2P and narrow-sense heritability across 36 traits. (A) PCC. (B) R^2 . (C) RMSE. (D) MAE. Correlation analyses were conducted using Pearson's correlation test. Each point represents one trait. The solid red line indicates the fitted linear regression, and the shaded area represents the 95% confidence interval. Source data are provided as a Source Data file.



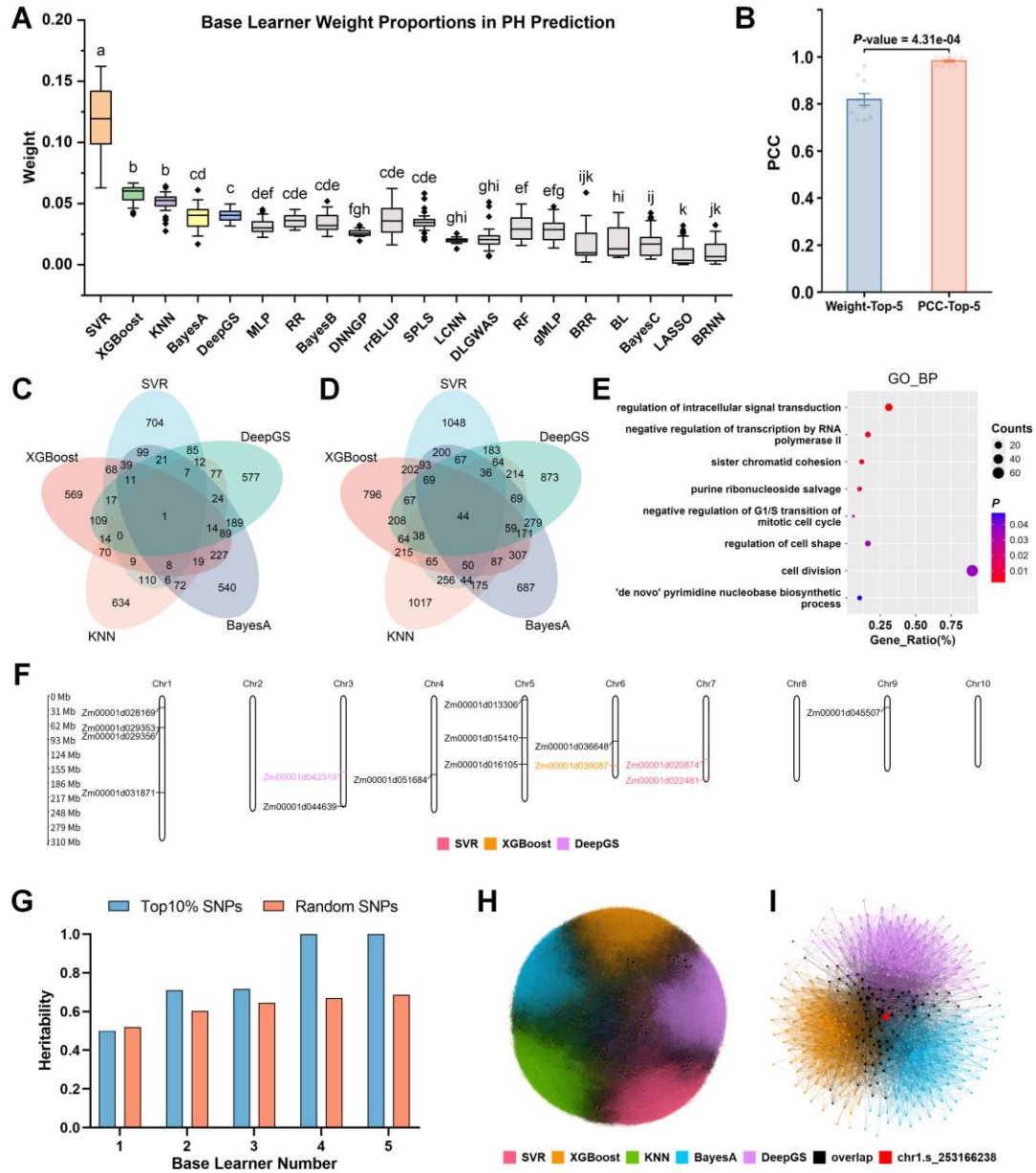
Supplementary Figure 7. Accuracy of different base learners in predicting three maize traits (ranked by PCC from high to low). (A) EW. (B) DTA. (C) PH. For each box, middle horizontal lines indicate medians; the upper and lower boundaries indicate the interquartile range (IQR, 25th-75th percentiles); whiskers extend to $1.5 \times \text{IQR}$ from the box boundaries; and points represent outliers. Different letters above the boxes indicate significant differences among groups as determined by one-way ANOVA followed by Tukey's multiple-comparisons test based on PCC values across ten-fold cross-validation ($n = 10$; P -value < 0.05). Orange, green, purple, yellow, and blue boxes indicate the five base learners with the highest PCC values. Grey boxes indicate all remaining base learners. Source data are provided as a Source Data file.



Supplementary Figure 8. Weight changes of the integrated base learners for predicting different maize traits during the optimization process of genetic algorithm in GEG2P. (A) DTA. (B) PH. (C) EW. The x-axis represents the iteration number of the genetic algorithm, and the y-axis represents the weight ranking of each base learner. Line colors denote gmlp (blue), SVR (green), SPLS (brown), XGBoost (grey), BayesA (cyan), and other base learners (light grey) in (A); SVR (blue), XGBoost (green), KNN (brown), BayesA (grey), DeepGS (cyan), and other base learners (light grey) in (B); and SVR (blue), XGBoost (green), gmlp (brown), DeepGS (grey), LASSO (cyan), and other base learners (light grey) in (C). Source data are provided as a Source Data file.

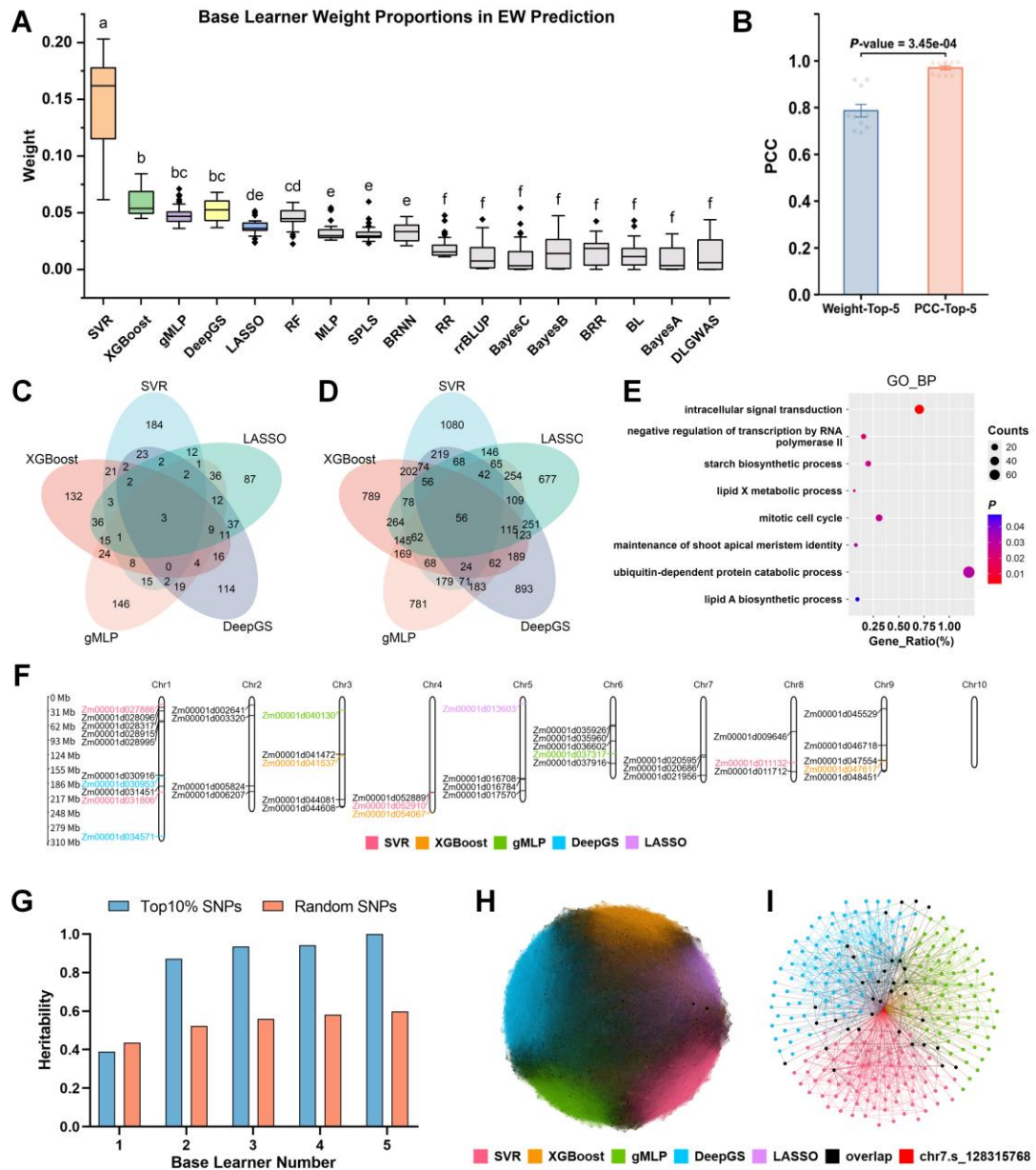


Supplementary Figure 9. Categories of genomic prediction methods corresponding to the top five base learners with the highest weights in GEG2P. (A) Maize. (B) Wheat. (C) Rice. (D) Chickpea. (E) Soybean. Bar colors denote machine learning (blue), deep learning (blue-grey), and statistical learning (light yellow). Source data are provided as a Source Data file.



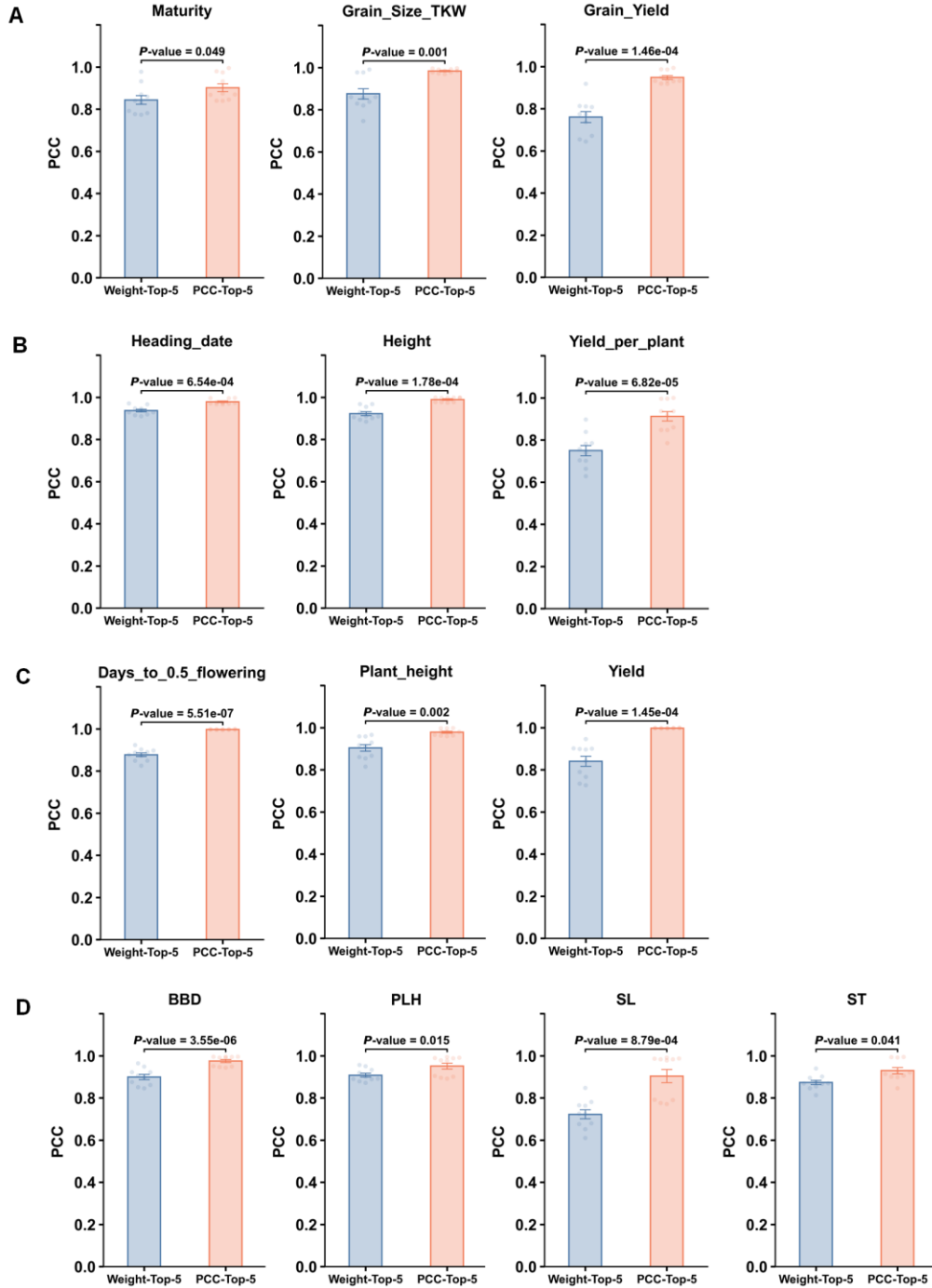
Supplementary Figure 10. Analysis of functional differences and complementarity of PH-Weight-Top-5. (A) The weight proportions of all base learners during 40 iterative optimization rounds of GEG2P. For each box, middle horizontal lines indicate medians; the upper and lower boundaries indicate the interquartile range (IQR, 25th-75th percentiles); whiskers extend to $1.5 \times \text{IQR}$ from the box boundaries; and points represent outliers. Different letters above the boxes indicate significant differences among groups as determined by one-way ANOVA followed by Tukey's multiple-comparisons test based on 40 iterative optimization rounds ($n = 40$; $P\text{-value} < 0.05$). Box plot colors: orange, SVR; green, XGBoost; purple, KNN; yellow, BayesA; blue, DeepGS; grey, all remaining base learners. (B) The average correlations of the predicted phenotype values among the top five base learners ranked by weight in GEG2P (Weight-Top-5), and among the top five base learners ranked by prediction accuracy (PCC-Top-5). Bars and error

bars indicate average Pearson correlation coefficients (PCCs) and standard errors, respectively ($n = 10$). P -values are from two-sided t -test. (C) Colocalization results of the Top10% SNPs of PH-Weight-Top-5 with plant architecture-related sQTLs. (D) The intersection of genes mapped by the Top10% SNPs of the five base learners in PH-Weight-Top-5. (E) GO enrichment analysis results of the union of all genes mapped by the Top10% SNPs of the five base learners in PH-Weight-Top-5. Point size indicates the number of enriched genes. Point color indicates the enrichment significance. P -values are from Fisher's exact test implemented in DAVID website. (F) The chromosomal locations of the cloned genes identified in (D). Genes mapped exclusively by the Top10% SNPs of SVR, XGBoost, and DeepGS are indicated in rose, yellow, and purple, respectively. (G) The narrow-sense heritability of successively accumulating the Top10% SNPs from the base learners in PH-Weight-Top-5. X-axis represents the cumulative number of base learners. Bar plot colors: blue, union of Top10% SNPs from PH-Weight-Top-5; orange, equal-number random SNPs. (H) Potential epistasis network of the Top10% SNPs of base learners in PH-Weight-Top-5, with rose, yellow, green, blue, and purple nodes representing the specific Top10% SNPs of SVR, XGBoost, KNN, BayesA, and DeepGS, respectively. Black nodes represent the Top10% SNPs shared by two or more base learners. (I) Potential epistasis network of SNPs that significantly interact with chr1.s_253166238 of the Top10% SNPs of three base learners (XGBoost, BayesA, and DeepGS). Yellow, blue, and purple nodes represent the specific Top10% SNPs that significantly interact with chr1.s_253166238 of XGBoost, BayesA, and DeepGS, respectively. Black nodes represent the Top10% SNPs shared by two or more base learners that significantly interact with chr1.s_253166238. Red node represents chr1.s_253166238. P -values for SNP interactions are from two-way ANOVA. Significant SNP interactions are selected at P -value < 0.001 . Source data are provided as a Source Data file.

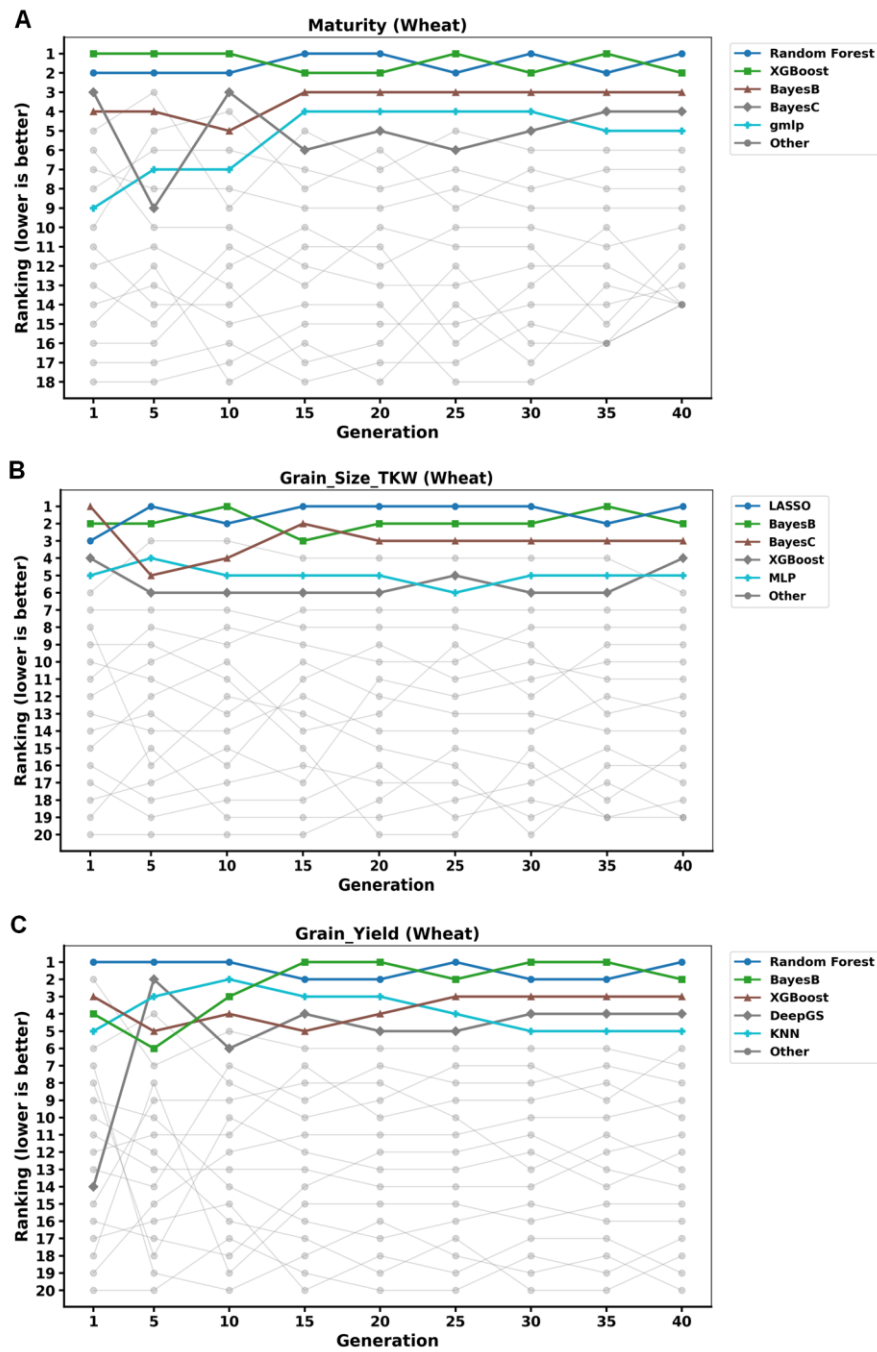


Supplementary Figure 11. Analysis of functional differences and complementarity of EW-Weight-Top-5. (A) The weight proportions of all base learners during 40 iterative optimization rounds of GEG2P. For each box, middle horizontal lines indicate medians; the upper and lower boundaries indicate the interquartile range (IQR, 25th-75th percentiles); whiskers extend to $1.5 \times \text{IQR}$ from the box boundaries; and points represent outliers. Different letters above the boxes indicate significant differences among groups as determined by one-way ANOVA followed by Tukey's multiple-comparisons test based on 40 iterative optimization rounds ($n = 40$; $P\text{-value} < 0.05$). Box plot colors: orange, SVR; green, XGBoost; purple, gMLP; yellow, DeepGS; blue, LASSO; grey, all remaining base learners. (B) The average correlations of the predicted phenotype values among the top five base learners ranked by weight in GEG2P (Weight-Top-5), and among the top five base learners ranked by prediction accuracy (PCC-Top-5). Bars and error bars indicate average PCCs and standard errors, respectively ($n = 10$). P -values are from two-sided

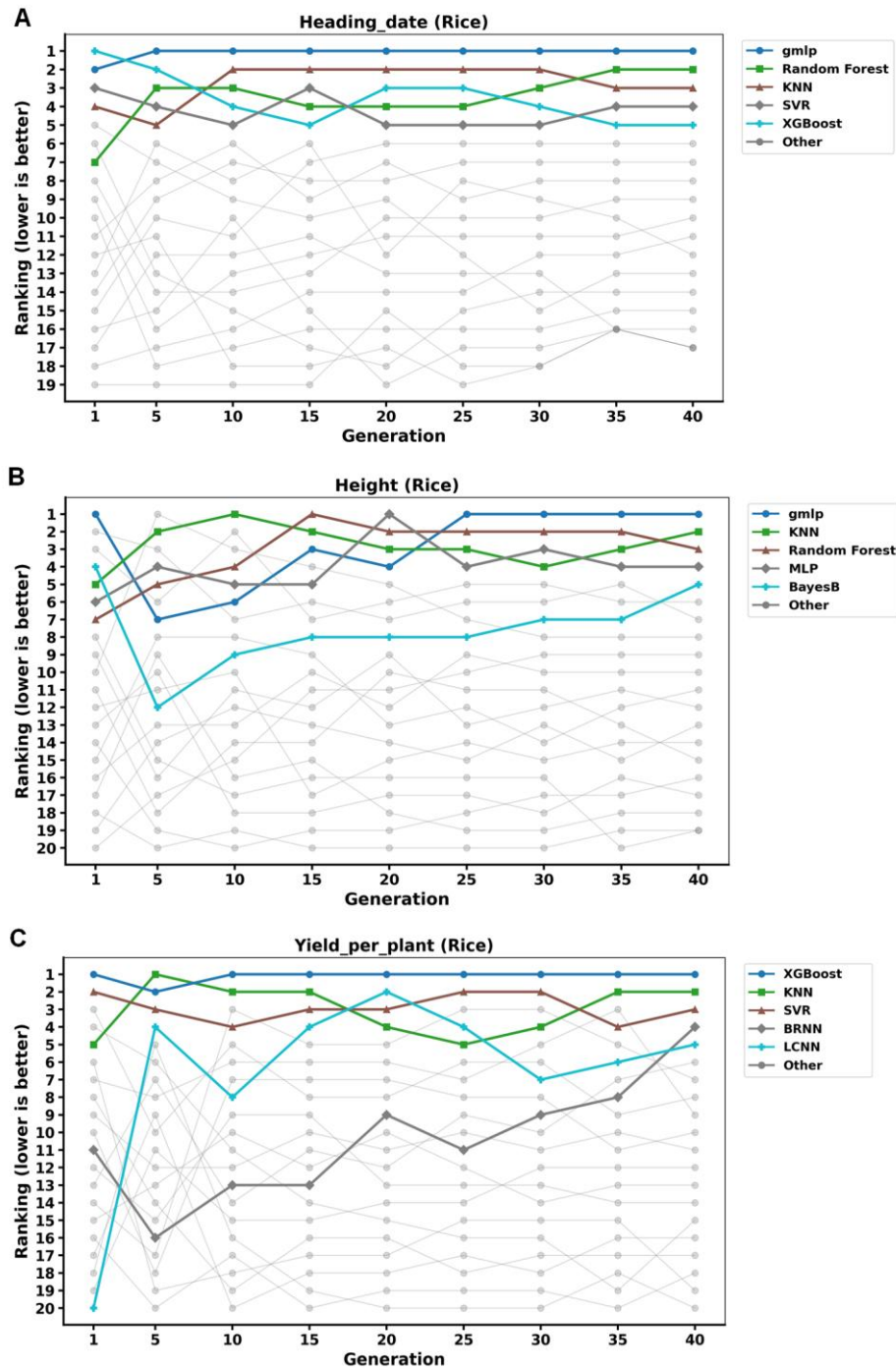
t-test. (C) Colocalization results of the Top10% SNPs of EW-Weight-Top-5 with yield-related sQTLs. (D) The intersection of genes mapped by the Top10% SNPs of the five base learners in EW-Weight-Top-5. (E) GO enrichment analysis results of the union of all genes mapped by the Top10% SNPs of the five base learners in EW-Weight-Top-5. Point size indicates the number of enriched genes. Point color indicates the enrichment significance. *P*-values are from Fisher's exact test implemented in DAVID website. (F) The chromosomal locations of the cloned genes identified in (D). Genes mapped exclusively by the Top10% SNPs of SVR, XGBoost, gMLP, DeepGS, and LASSO are indicated in rose, yellow, green, blue, and purple, respectively. (G) The narrow-sense heritability of successively accumulating the Top10% SNPs from the base learners in EW-Weight-Top-5. X-axis represents the cumulative number of base learners. Bar plot colors: blue, union of Top10% SNPs from EW-Weight-Top-5; orange, equal-number random SNPs. (H) Potential epistasis network of the Top10% SNPs of base learners in EW-Weight-Top-5, with rose, yellow, green, blue, and purple nodes representing the specific Top10% SNPs of SVR, XGBoost, gMLP, DeepGS, and LASSO, respectively. Black nodes represent the Top10% SNPs shared by two or more base learners. (I) Potential epistasis network of SNPs that significantly interact with chr7.s_128315768 of the Top10% SNPs of three base learners (SVR, gMLP and DeepGS). Rose, green and blue nodes represent the specific Top10% SNPs that significantly interact with chr7.s_128315768 of SVR, gMLP and DeepGS, respectively. Black nodes represent the Top10% SNPs shared by two or more base learners that significantly interact with chr7.s_128315768. Red node represents chr7.s_128315768. *P*-values for SNP interactions are from two-way ANOVA. Significant SNP interactions are selected at *P*-value < 0.001. Source data are provided as a Source Data file.



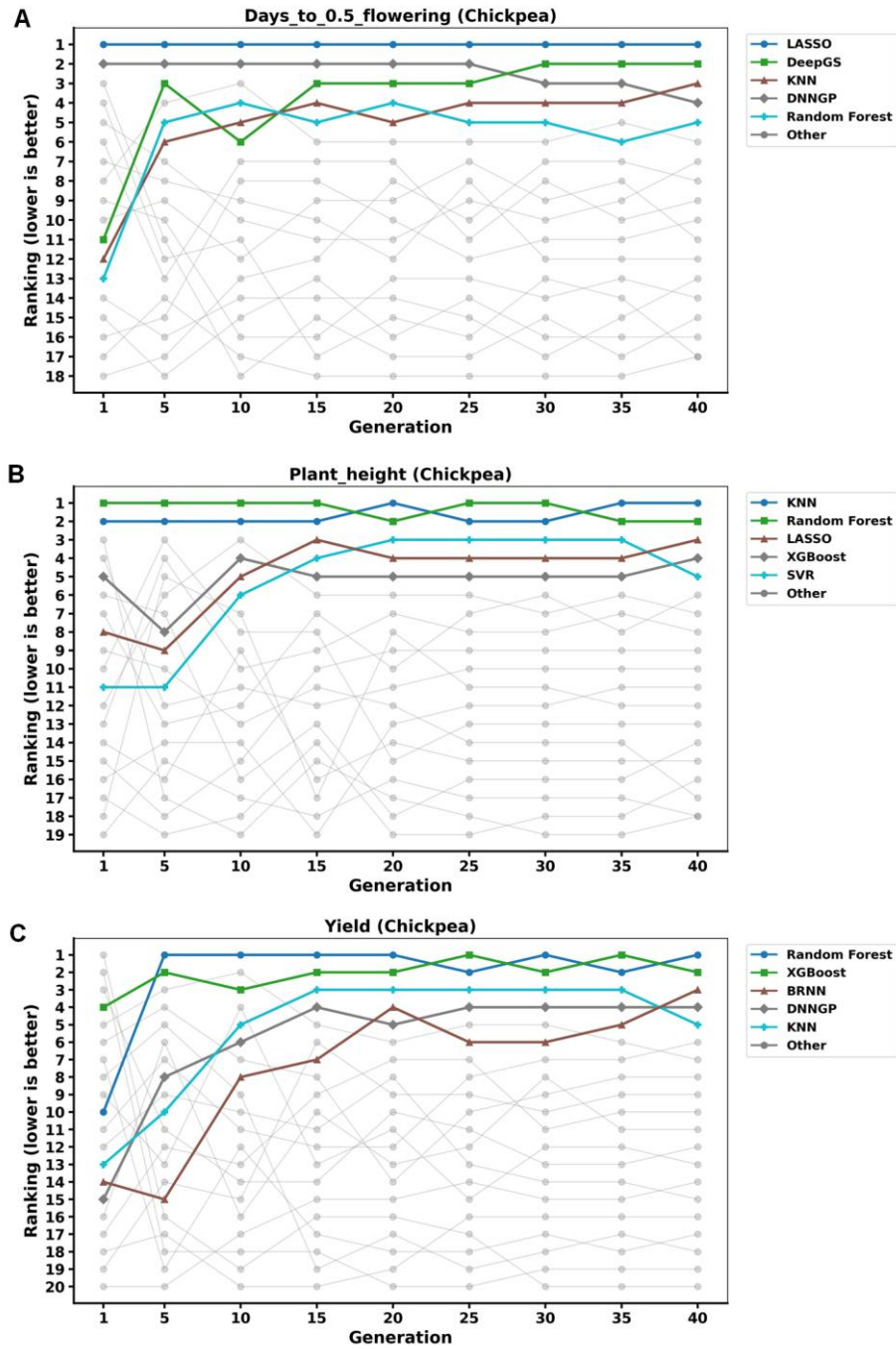
Supplementary Figure 12. Comparison results of average correlations between phenotype values predicted by base learners in GEG2P. Weight-Top-5 represents the top five base learners with the highest weights in GEG2P, while PCC-Top-5 represents the top five base learners ranked by PCC. (A) Maturity, Grain_Size_TKW, and Grain_Yield in wheat. (B) BLUP_Heading_date, BLUP_Height, and BLUP_Yield_per_plant in rice. (C) Days_to_0.5_flowering, Plant_height, Yield in chickpea. (D) BBD_BLUP, PLH_BLUP, SL_BLUP, ST_BLUP in soybean. Bars and error bars indicate average PCCs and standard errors, respectively (n = 10). P-values are from two-sided *t*-test. Bar colors denote Weight-Top-5 (light blue) and PCC-Top-5 (light red). Source data are provided as a Source Data file.



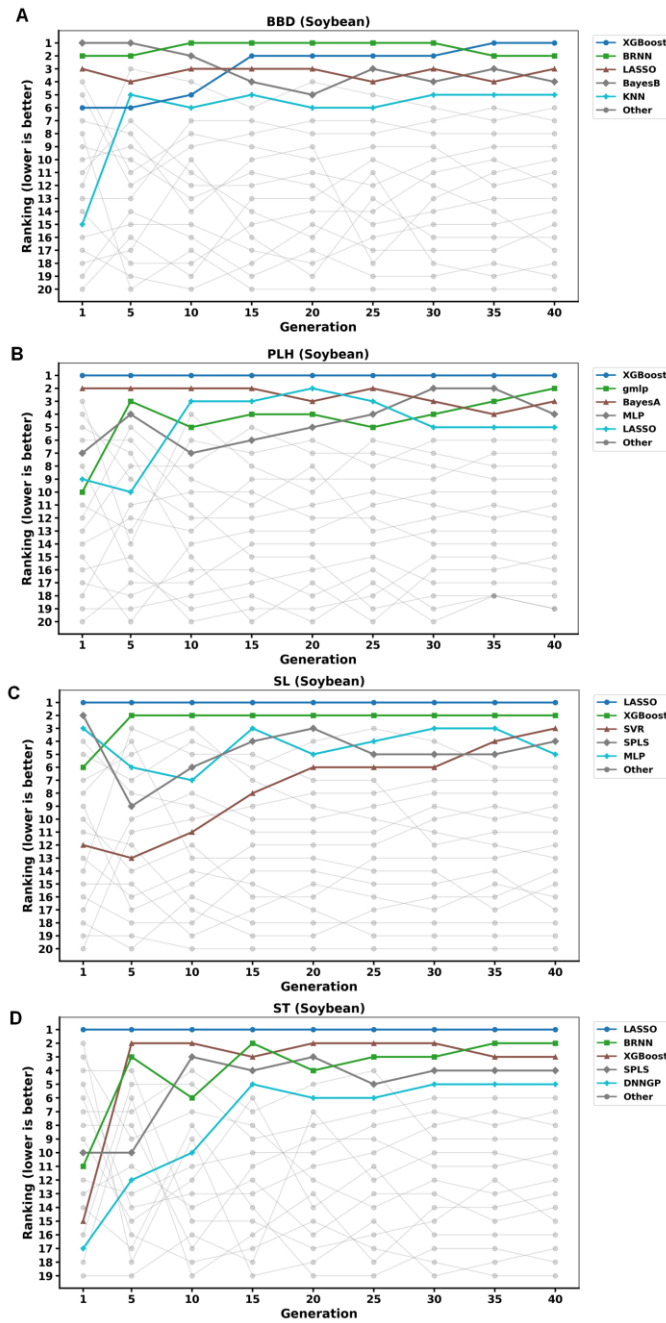
Supplementary Figure 13. Weight changes of the integrated base learners for predicting different wheat traits during the optimization process of genetic algorithm in GEG2P. (A) Maturity. (B) Grain_Size_TKW. (C) Grain_Yield. The x -axis represents the iteration number of the genetic algorithm, and the y -axis represents the weight ranking of each base learner. Line colors denote Random Forest (blue), XGBoost (green), BayesB (brown), BayesC (grey), gMLP (cyan), and other base learners (light grey) in (A); LASSO (blue), BayesB (green), BayesC (brown), XGBoost (grey), MLP (cyan), and other base learners (light grey) in (B); and Random Forest (blue), BayesB (green), XGBoost (brown), DeepGS (grey), KNN (cyan), and other base learners (light grey) in (C). Source data are provided as a Source Data file.



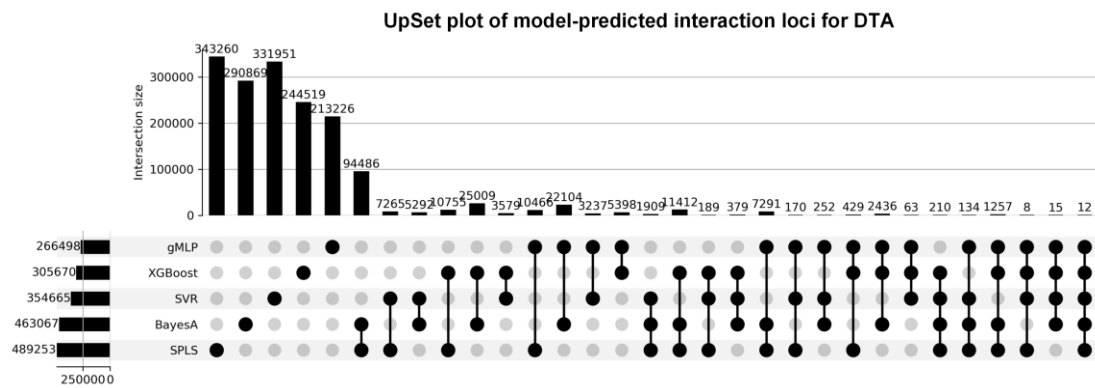
Supplementary Figure 14. Weight changes of the integrated base learners for predicting different rice traits during the optimization process of genetic algorithm in GEG2P. (A) Heading_date. (B) Height. (C) Yield_per_plant. The x-axis represents the iteration number of the genetic algorithm, and the y-axis represents the weight ranking of each base learner. Line colors denote gmlp (blue), Random Forest (green), KNN (brown), SVR (grey), XGBoost (cyan), and other base learners (light grey) in (A); gmlp (blue), KNN (green), Random Forest (brown), MLP (grey), BayesB (cyan) in (B); and XGBoost (blue), KNN (green), SVR (brown), BRNN (grey), LCNN (cyan) in (C). Source data are provided as a Source Data file.



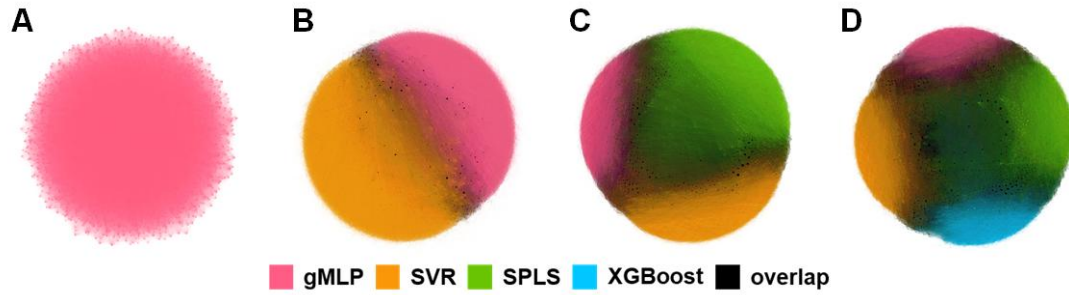
Supplementary Figure 15. Weight changes of the integrated base learners for predicting different chickpea traits during the optimization process of genetic algorithm in GEG2P. (A) Days_to_0.5_flowering. (B) Plant_height. (C) Yield. The x -axis represents the iteration number of the genetic algorithm, and the y -axis represents the weight ranking of each base learner. Line colors denote LASSO (blue), DeepGS (green), KNN (brown), DNNGP (grey), Random Forest (cyan), and other base learners (light grey) in (A); KNN (blue), Random Forest (green), LASSO (brown), XGBoost (grey), SVR (cyan), and other base learners (light grey) in (B); and Random Forest (blue), XGBoost (green), BRNN (brown), DNNGP (grey), KNN (cyan), and other base learners (light grey) in (C). Source data are provided as a Source Data file.



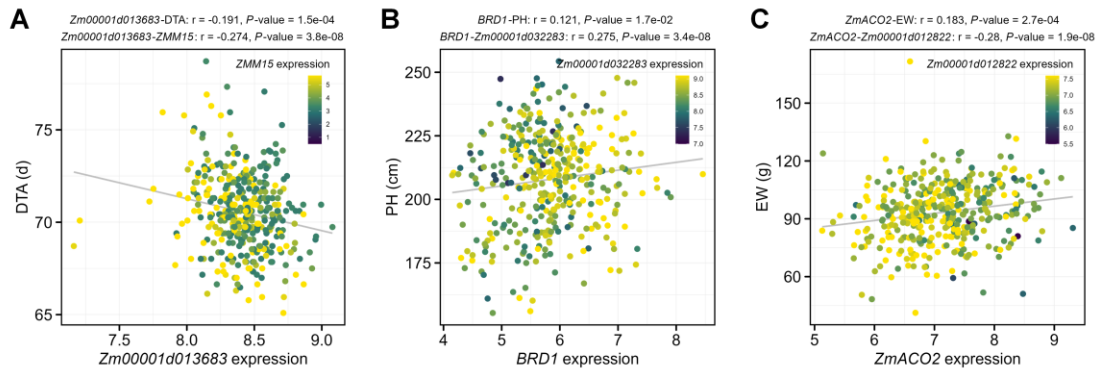
Supplementary Figure 16. Weight changes of the integrated base learners for predicting different soybean traits during the optimization process of genetic algorithm in GEG2P. (A) BBD. (B) PLH. (C) ST. (D) ST. The *x*-axis represents the iteration number of the genetic algorithm, and the *y*-axis represents the weight ranking of each base learner. Line colors denote XGBoost (blue), BRNN (green), LASSO (brown), BayesB (grey), KNN (cyan), and other base learners (light grey) in (A); XGBoost (blue), gmlp (green), BayesA (brown), MLP (grey), LASSO (cyan), and other base learners (light grey) in (B); LASSO (blue), XGBoost (green), SVR (brown), SPLS (grey), MLP (cyan), and other base learners (light grey) in (C); and LASSO (blue), BRNN (green), XGBoost (brown), SPLS (grey), DNNGP (cyan), and other base learners (light grey) in (D). Source data are provided as a Source Data file.



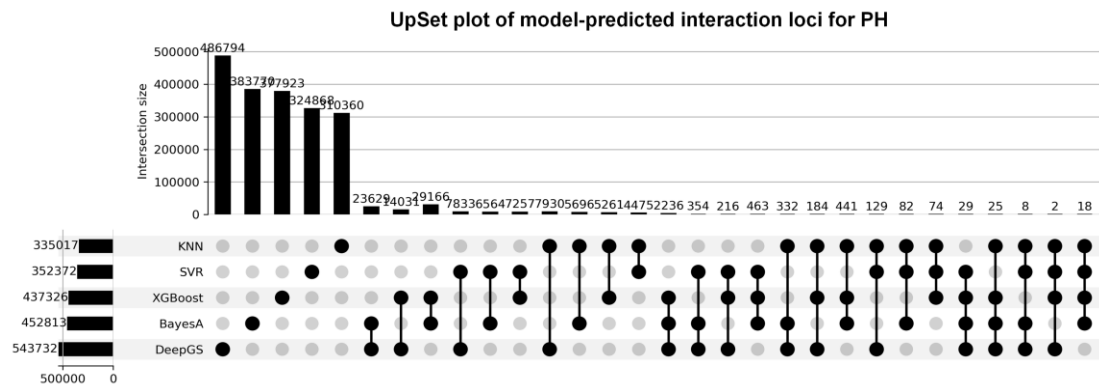
Supplementary Figure 17. The number of significant SNP-SNP interactions among the Top10% SNPs of DTA-Weight-Top-5. The first five bars represent the number of significant SNP-SNP interactions within the Top10% SNPs of individual base learners, whereas the remaining bars indicate the number of significant SNP-SNP interactions shared among different base learners. *P*-values for SNP interactions are from two-way ANOVA. Significant SNP interactions are selected at *P*-value < 0.001. Source data are provided as a Source Data file.



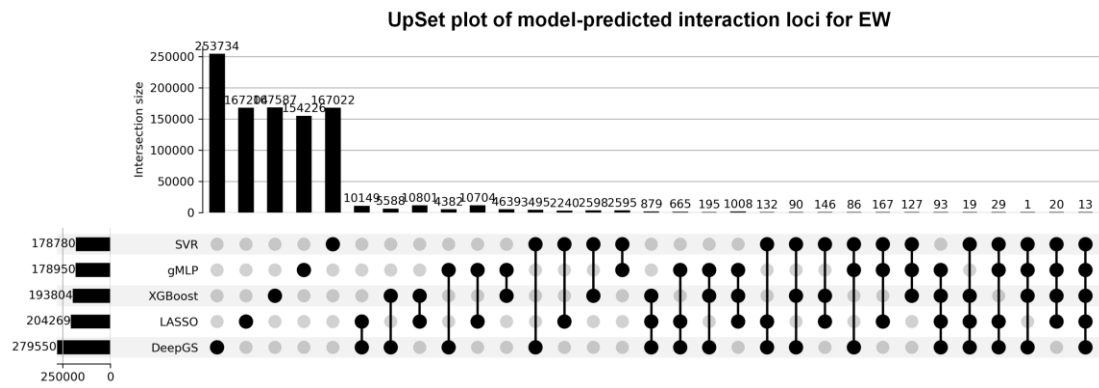
Supplementary Figure 18. Epistasis network of Top10% SNPs of top four base learners ranked by weight in GEG2P of maize DTA. (A) Epistasis network of the significant SNP-SNP interactions among the Top10% SNPs identified by gMLP. (B)-(D) The epistasis networks of the Top10% SNPs from multiple base learners. The networks were constructed based on (A) by sequentially incorporating the significant SNP-SNP interactions of the Top10% SNPs from SVR (B), SPLS (C), and XGBoost (D), according to the descending order of the base learner weights for DTA prediction. Rose, yellow, green, and blue nodes represent the specific Top10% SNPs of gMLP, SVR, SPLS, and XGBoost, respectively. Black nodes represent the Top10% SNPs shared by two or more base learners. *P*-values for SNP interactions are from two-way ANOVA. Significant SNP interactions are selected at *P*-value < 0.001. Source data are provided as a Source Data file.



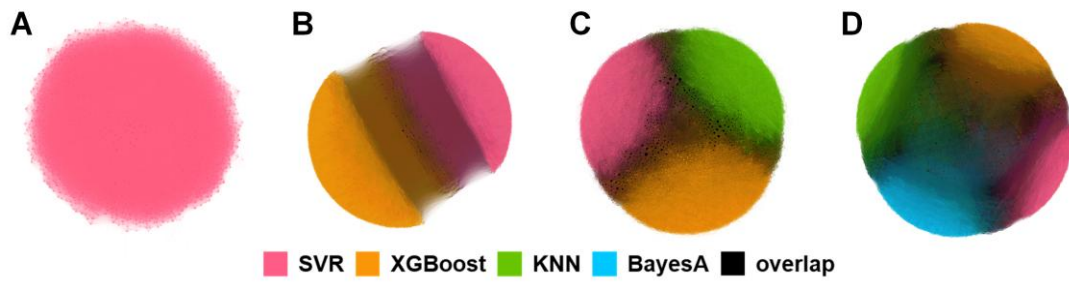
Supplementary Figure 19. Correlation between expression values of interacting gene pairs and phenotypic variation. (A) Pearson correlation between the expression value of *Zm00001d013683* (x-axis) and DTA (y-axis). Color scale indicates *ZMM15* expression value. (B) Pearson correlation between the expression value of *BRD1* (x-axis) and PH (y-axis). Color scale indicates *Zm00001d032283* expression value. (C) Pearson correlation between the expression value of *ZmACO2* (x-axis) and EW (y-axis). Color scale indicates *Zm00001d012822* expression value. Each point represents one accession. The solid grey line indicates the fitted linear regression. Source data are provided as a Source Data file.



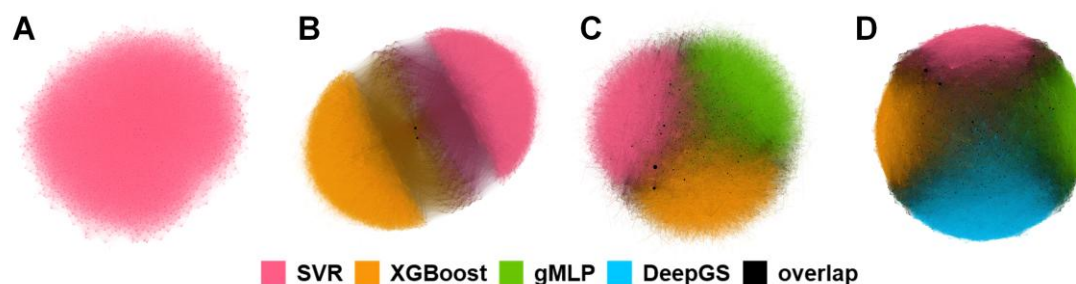
Supplementary Figure 20. The number of significant SNP-SNP interactions among the Top10% SNPs of PH-Weight-Top-5. The first five bars represent the number of significant SNP-SNP interactions within the Top10% SNPs of individual base learners, whereas the remaining bars indicate the number of significant SNP-SNP interactions shared among different base learners. *P*-values for SNP interactions are from two-way ANOVA. Significant SNP interactions are selected at *P*-value < 0.001. Source data are provided as a Source Data file.



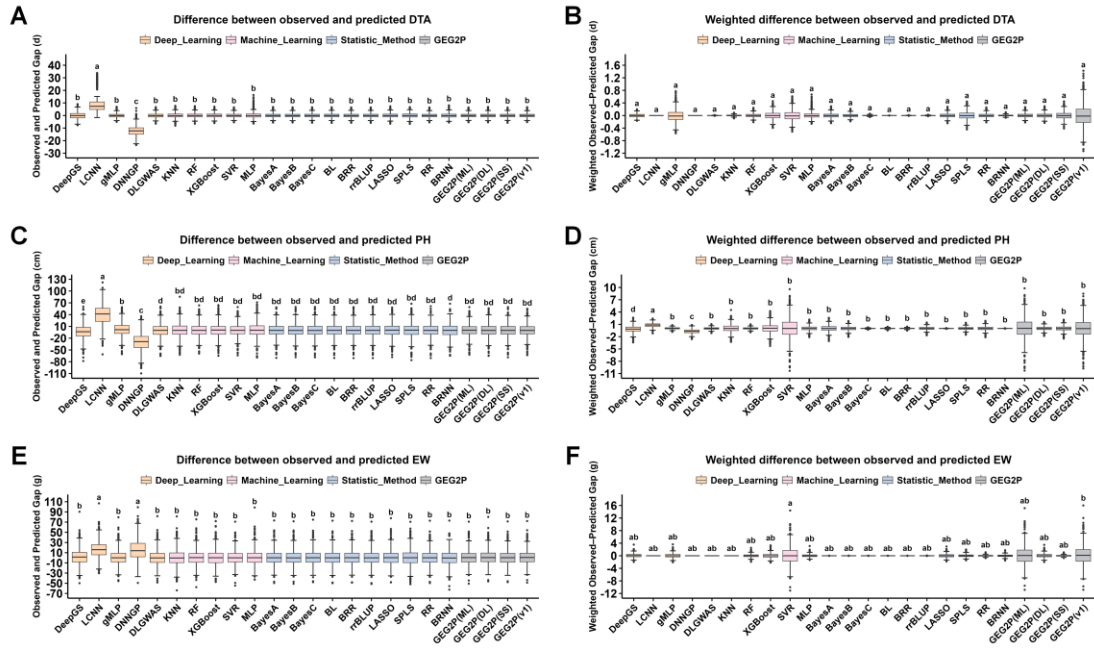
Supplementary Figure 21. The number of significant SNP-SNP interactions among the Top10% SNPs of EW-Weight-Top-5. The first five bars represent the number of significant SNP-SNP interactions within the Top10% SNPs of individual base learners, whereas the remaining bars indicate the number of significant SNP-SNP interactions shared among different base learners. *P*-values for SNP interactions are from two-way ANOVA. Significant SNP interactions are selected at *P*-value < 0.001. Source data are provided as a Source Data file.



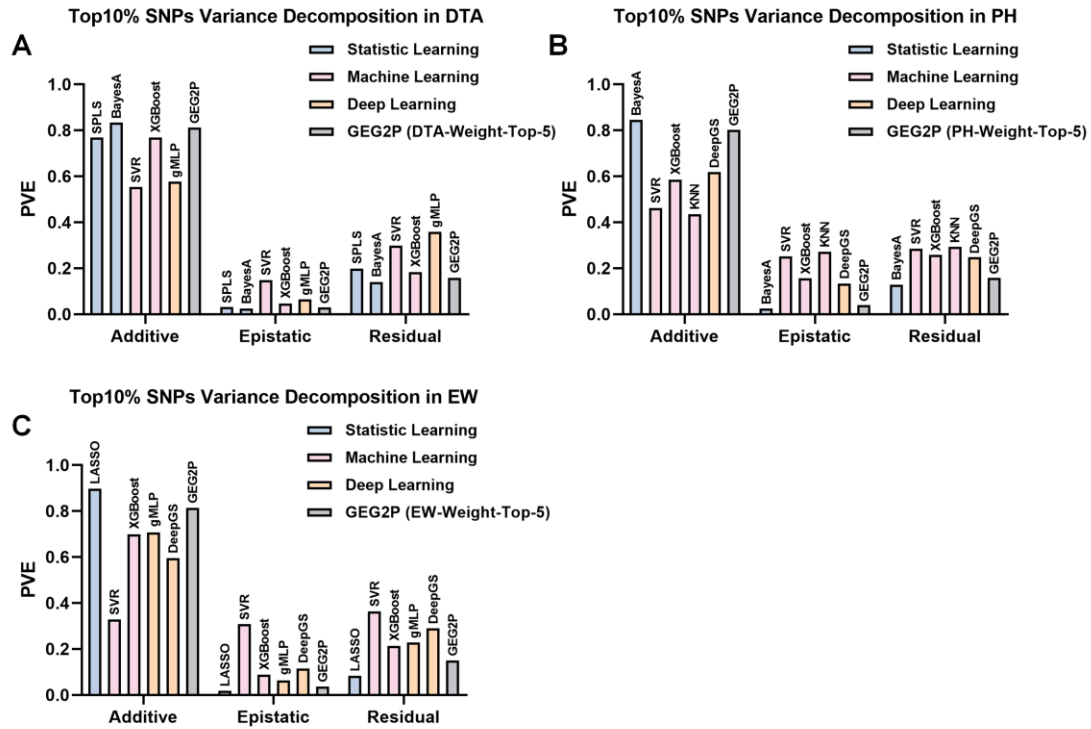
Supplementary Figure 22. Epistasis network of Top10% SNPs from top four base learners ranked by weight in GEG2P for maize PH. (A) Epistasis network of the significant SNP-SNP interactions among the Top10% SNPs identified by SVR. (B)-(D) The epistasis networks of the Top10% SNPs from multiple base learners. The networks were constructed based on (A) by sequentially incorporating the significant SNP-SNP interactions of the Top10% SNPs from XGBoost (B), KNN (C), and BayesA (D), according to the descending order of the base learner weights for PH prediction. Rose, yellow, green, and blue nodes represent the specific Top10% SNPs of SVR, XGBoost, KNN and BayesA, respectively. Black nodes represent the Top10% SNPs shared by two or more base learners. *P*-values for SNP interactions are from two-way ANOVA. Significant SNP interactions are selected at *P*-value < 0.001. Source data are provided as a Source Data file.



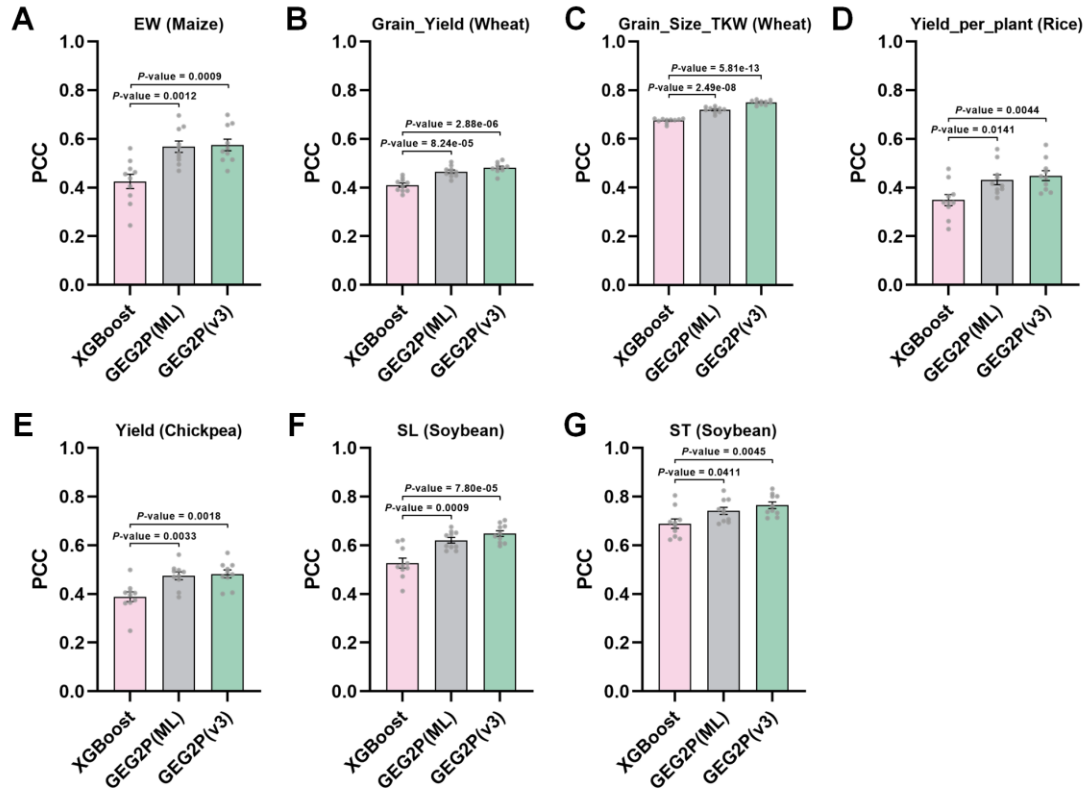
Supplementary Figure 23. Epistasis network of Top10% SNPs from top four base learners ranked by weight in GEG2P for maize EW. (A) Epistasis network of the significant SNP-SNP interactions among the Top10% SNPs identified by SVR. (B)-(D) The epistasis networks of the Top10% SNPs from multiple base learners. The networks were constructed based on (A) by sequentially incorporating the significant SNP-SNP interactions of the Top10% SNPs from XGBoost (B), gMLP (C), and DeepGS (D), according to the descending order of the base learner weights for EW prediction. Rose, yellow, green, and blue nodes represent the specific Top10% SNPs of SVR, XGBoost, gMLP and DeepGS, respectively. Black nodes represent the Top10% SNPs shared by two or more base learners. *P*-values for SNP interactions are from two-way ANOVA. Significant SNP interactions are selected at *P*-value < 0.001. Source data are provided as a Source Data file.



Supplementary Figure 24. Weight-based ensemble correction among prediction results of different base learners. (A, C, E) Prediction deviations (observed minus predicted values) of base learners for the DTA, PH, and EW, respectively. (B, D, F) The weighted results of these deviations derived from base learner weights. For each box, middle horizontal lines indicate medians; the upper and lower boundaries indicate the interquartile range (IQR, 25th-75th percentiles); whiskers extend to $1.5 \times \text{IQR}$ from the box boundaries; and points represent outliers. Different letters above the boxes indicate significant differences among groups as determined by one-way ANOVA followed by Tukey's multiple-comparisons test based on 1,404 maize accessions ($n = 1404$; $P\text{-value} < 0.05$). Box plot colors: light orange, deep learning methods; pink, machine learning methods; blue, statistical learning methods; grey, GEG2P. Source data are provided as a Source Data file.



Supplementary Figure 25. Variance decomposition of genetic effects for the Top10% SNPs identified by individual base learners and their union in DTA-Weight-Top-5 (A), PH-Weight-Top-5 (B), and EW-Weight-Top-5 (C). Bar plot colors: blue, statistical learning methods in Weight-Top-5; pink, machine learning methods in Weight-Top-5; light orange, deep learning methods in Weight-Top-5; grey, union of SNPs from GEG2P Weight-Top-5 base learners. Source data are provided as a Source Data file.



Supplementary Figure 26. Prediction accuracy of maize EW (A), wheat Grain_Yield (B), wheat Grain_Size_TKW (C), rice Yield_per_plant (D), chickpea Yield (E), soybean SL (F) and soybean ST (G) using XGBoost, GEG2P (ML), and GEG2P (v3). Bars and error bars indicate average Pearson correlation coefficients (PCCs) and standard errors, respectively (n=10). *P*-values are from two-sided *t*-test. Bar plot colors: pink, prediction accuracy of XGBoost; grey, prediction accuracy of GEG2P(ML); green, prediction accuracy of GEG2P(v3). Source data are provided as a Source Data file.

Supplementary reference

1. Liu, H. *et al.* CUBIC: An atlas of genetic architecture promises directed maize improvement. *Genome Biol.* **21**, 20 (2020).