Molecular Plant
Resource article

CellPress
Partner Journal

GEFormer: A genotype-environment interaction-based genomic prediction method that integrates the gating multilayer perceptron and linear attention mechanisms

Zhou Yao[1,2,3,5], Mengting Yao[3,5], Chuang Wang[1,2,3], Ke Li[3], Junhao Guo[3], Yingjie Xiao[1,4], Jianbing Yan[1,4] and Jianxiao Liu[1,2,3,4,*]

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

[2]Hubei Key Laboratory of Agricultural Bioinformatics, Huazhong Agricultural University, Wuhan 430070, China

[3]College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

[4]Hubei Hongshan Laboratory, Wuhan 430070, China

[5]These authors contributed equally to this article.

***Correspondence: Jianxiao Liu (liujianxiao@mail.hzau.edu.cn)**

https://doi.org/10.1016/j.molp.2025.01.020

ABSTRACT

The integration of genotypic and environmental data can enhance genomic prediction accuracy for crop field traits. Existing genomic prediction methods fail to consider environmental factors and the real growth environments of crops, resulting in low genomic prediction accuracy. In this work, we developed GEFormer, a genotype–environment interaction genomic prediction method that integrates gating multilayer perceptron (gMLP) and linear attention mechanisms. First, GEFormer uses gMLP to extract local and global features among SNPs. Then, Omni-dimensional Dynamic Convolution is used to extract the dynamic and comprehensive features of multiple environmental factors within each day, taking into consideration the real growth pattern of crops. A linear attention mechanism is used to capture the temporal features of environmental changes. Finally, GEFormer uses a gating mechanism to effectively fuse the genomic and environmental features. We examined the accuracy of GEFormer for predicting important agronomic traits of maize, rice, and wheat under three experimental scenarios: untested genotypes in tested environments, tested genotypes in untested environments, and untested genotypes in untested environments. The results showed that GEFormer outperforms six cutting-edge statistical learning methods and four machine learning methods, especially with great advantages under the scenario of untested genotypes in untested environments. In addition, we used GEFormer for three real-world breeding applications: phenotype prediction in unknown environments, hybrid phenotype prediction using an inbred population, and cross-population phenotype prediction. The results showed that GEFormer had better prediction performance in actual breeding scenarios and could be used to assist in crop breeding.

Key words: genomic prediction, crop growth environment, genotype–environment interactions, gated MLP, linear attention mechanism

Yao Z., Yao M., Wang C., Li K., Guo J., Xiao Y., Yan J., and Liu J. (2025). GEFormer: A genotype-environment interaction-based genomic prediction method that integrates the gating multilayer perceptron and linear attention mechanisms. Mol. Plant. **18**, 527–549.

INTRODUCTION

Global climate change poses great challenges to food security and sustainable agricultural development (Abberton et al., 2016). Breeders face pressure to improve and develop crop varieties, and crops will need to be more productive and capable of adapting to dynamically changing environments in the future (Hickey et al. 2019). The development of novel crop

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

varieties adapted to diverse environmental conditions can help to further alleviate food shortages caused by biotic and abiotic stresses (Xiong et al., 2022).

The field phenotypes of crops are the result of interactions between genotype and environmental factors. Environmental factors refer to factors that affect crop growth, including soil, moisture, light, and temperature (Xu, 2016). Genotype–environment interactions (Crossa et al., 2004) can lead to variance differences and rank changes among SNPs (Cooper and DeLacy, 1994). Genomic prediction (GP) refers to the use of genotypic data to predict the field phenotypes of crops (including yield, plant height [PH], and flowering time). Breeders can quickly screen varieties on the basis of predicted phenotype values. Therefore, GP could help to shorten the breeding cycle and improve resource utilization efficiency in animal and plant breeding. However, the performance of GP needs to be improved in multi-environment trials (Burgueño et al., 2012; Jarquín et al., 2017; Gillberg et al., 2019). Therefore, it is necessary to consider genotype–environment interactions and construct genotype–environment–phenotype (Xu et al., 2022) GP models (G × E) to improve the accuracy of phenotype prediction.

In recent years, researchers have developed a number of GP methods that consider genotype–environment interactions to improve the prediction accuracy of complex traits. Some researchers convert the function of environmental data response norms into environmental covariates (ECs), which are the linear responses of genotype data to the target environmental gradient. They then construct a kinship matrix of environmental factors based on the correlations between ECs. The corresponding GP methods related to environmental feature kernels include benchmark genomic best unbiased predictor (GB), Gaussian kernel (GK), and deep kernel (DK) methods (Costa-Neto et al., 2021b; Resende et al., 2021; Araújo et al., 2024). In addition, some studies first extract genotypic features using the genomic relationship matrix (GRM) based on additive effects, the GRM based on dominance effects, and the GRM based on epistatic effects (Hayes et al., 2009; Vitezica et al., 2013, 2017; Martini et al., 2017). Then, a mixed linear model is used to integrate the genotype and environmental features and achieve multi-environment GP in maize (Granato et al., 2018; Costa-Neto et al., 2021b). However, exploring the complex nonlinear relationships between genotype data and environmental factors remains the key to improving prediction accuracy. Recently, researchers have used machine-learning methods, including random forest (RF), eXtreme gradient boosting (XGBoost), stacking ensemble model (Stacking), and deep learning (DL), to capture high-order interactions between genotype and environmental factors (Westhues et al., 2021, 2022). Such machine-learning methods outperformed linear random-effect methods in predicting maize grain yield (Westhues et al., 2021), demonstrating the great potential in the field of G × E GP. However, machine-learning methods often use principal-component analysis to reduce the dimensionality of the original genotype data, resulting in the loss of nonlinear features (dominant and epistatic effect features) embedded within the original genotype data. That is to say, machine-learning methods often input dimensionally reduced genotypic features and the covariate of the phenological period into models for prediction. However, environmental features change dynamically over the whole crop growth period. Relying on only the covariate characteristics of the phenological period will decrease the temporal features of the whole growth period, thus affecting prediction accuracy.

Current methods for extraction of environmental features rely mainly on the similarity of ECs or phenological ECs to construct environmental features. However, these approaches are unable to capture high-order interactions among environmental factors and dynamically changing features throughout the whole growth period. Currently, DL methods have shown great advantages in the fields of computer vision, natural language processing, biological data analysis, and multi-modal feature fusion (Achiam et al., 2023). DeepGS (Ma et al., 2018), Galiana (Raimondi et al., 2022), DNNGP (Wang et al., 2023), TrG2P (Li et al., 2024b), SoyDNGP (Gao et al., 2023), GPformer (Wu et al., 2024), DeepCCR (Ma et al., 2024), and other methods have also verified that DL has superior GP effects. In addition, some studies have used multi-omics data for GP (Wang et al. 2023; Ren et al., 2024). The focus of these works was not to incorporate environmental data from the entire crop growth period into GPs. DL methods can undoubtedly capture complex nonlinear relationships among genotypes, environmental factors, and genotype–environment interaction features. Regarding genotype and environmental data as different modal features, the present work proposes a G × E genome-wide prediction method called GEFormer based on multi-modal DL.

(1) GEFormer consists of gated multilayer perceptron (gMLP) layer, TimeFeatureBlock layer, and CrossGatedMLP layer. The gMLP layer is used to extract genotype features from both local and global perspectives, thereby capturing the long-range dependencies among SNPs. In consideration of the patterns of the real crop growth environment, the TimeFeatureBlock layer uses Omni-dimensional Dynamic Convolution (ODConv) to fuse the features among different environmental factors within each day. The linear attention mechanism is used to extract the temporal features of environmental factors during the crop growth period. Based on the idea of cross-modal feature fusion, the CrossGatedMLP layer integrates the extracted genotype and environmental features to capture the complex nonlinear relationships hidden in different omics data.

(2) We validated the accuracy of GEFormer in predicting important agronomic traits using multiple populations of maize, rice, and wheat. We performed experiments using three scenarios: tested genotypes in untested environments, untested genotypes in tested environments, and untested genotypes in untested environments. The experimental results of the three scenarios demonstrated that GEFormer outperformed six cutting-edge statistical learning methods and four machine-learning methods.

(3) We used three scenarios to analyze the application potential of GEFormer for crop breeding. The first application was phenotype prediction in unknown environments. GEFormer could accurately predict phenotype in unknown environments in the scenario of tested genotypes in untested environments. The second application was phenotype prediction for a hybrid population using inbred lines. Incorporation of maternal and paternal inbred materials improved the prediction accuracy of the hybrid population. The third application was cross-population phenotype

prediction. The incorporation of some inbred (hybrid) materials to the training set of the hybrid (inbred) population significantly boosted prediction accuracy for the hybrid (inbred) population.

RESULTS

GEFormer model and experimental setup

GEFormer extracts genotype features (G*) and environmental data features (E*) in order to predict phenotypes through integration of G* and E*. GEFormer includes five layers: the input layer, gMLP layer, TimeFeatureBlock layer, CrossGatedMLP layer, and output layer (Figure 1). It first preprocesses the original genotype and environmental data, which are input into the gMLP layer and TimeFeatureBlock layer, respectively. The gMLP layer (Liu et al., 2021a) is used to extract both local and global features among SNPs, thus obtaining the genotype features G*. The long-range dependencies among SNPs are captured through the gating mechanism and nonlinear activation function. To capture the holistic impact of environmental factors on crop growth processes, the TimeFeatureBlock module is used to extract features (E*) of 25 environmental factors, including day length (DayL), relative humidity (RH), and surface air pressure (PS). As we know, there are interactions among different environmental factors during the real growth process of crops, and each environmental factor exhibits temporal characteristics that vary over time. The TimeFeatureBlock module first uses ODConvto extract the comprehensive effects among the environmental factors within each day. It then uses linear attention mechanism to capture the temporal features of the environmental factors between different days. ODConv can adjust the parameters of convolutional kernels dynamically during the training process, thus enhancing the ability to capture complex features (Li et al., 2022a). Linear attention mechanisms can determine the importance of different positions in the input sequence by introducing learnable weights, thereby improving the ability to capture local key features in the sequence (Zhou et al., 2021). By fusing the features of G*, E*, and G* × E*, the CrossGatedMLP layer uses the learning gating signal to adaptively determine the contribution of each feature to the fusion feature and enhance the representation ability of the model (Tu et al., 2022). The output layer uses MLP to extract the fused features and maps them to the phenotype values. The architecture of GEFormer is shown in Figure 1.

We used three experimental scenarios (M1–M3) to assess the prediction performance of GEFormer, as shown in Supplemental Figure 1. In M1, 10-fold cross-validation was used to evaluate the prediction accuracy for untested genotypes in tested environments (Supplemental Figure 1A). In M2, leave-one-environment-out cross-validation was used to evaluate the prediction accuracy for tested genotypes in untested environments (Supplemental Figure 1B). In M3, leave-one-environment-out and 10-fold cross-validation were used to evaluate the prediction accuracy for untested genotypes in untested environments. In scenario M3, we divided the population materials into 10 groups. One group of materials in one environment was used as the validation set, and the remaining nine groups of materials in all other environments were used as the training set (Supplemental Figure 1C). We used the average prediction accuracy to compare GEFormer with four machine-learning methods and six statistical learning methods (Materials and methods). The four machine-learning methods were RF, XGBoost, Stacking, and MLPs (DL) (Westhues et al., 2022). The six statistical learning methods were the main-effect EA plus reaction norm for GE based on genomic best unbiased predictor (GBLUP) (EA + GW(GB)), the main-effect EADW plus reaction norm for GE based on GBLUP (EADW + GW(GB)), the main-effect EA plus reaction norm for GE based on GK (EA + GW(GK)), the main-effect EADW plus reaction norm for GE based on GK (EADW + GW(GK)), the main-effect EA plus reaction norm for GE based on DK (EA + GW(DK)), and the main-effect EADW plus reaction norm for GE based on deep kernel (EADW + GW(DK)) (Granato et al., 2018). We performed these experiments using datasets from four maize populations: the CUBIC1404 inbred population (the Complete-diallel design plus Unbalanced Breeding-like Inter-Cross, CUBIC) (Liu et al., 2020), the CUBIC1404*JING724 hybrid population (Xiao et al., 2021), the CUBIC1404*ZHENG58 hybrid population (Xiao et al., 2021), and the Maize976 hybrid population (Liu et al., 2021b). We also used the datasets from rice and wheat to validate the effectiveness of GEFormer (Materials and methods).

Evaluation of phenotype prediction in scenario M1

We compared the prediction accuracy of GEFormer with that of four machine-learning methods and six statistical learning methods using three maize populations (the CUBIC1404 inbred population, the CUBIC1404*JING724 hybrid population, and the CUBIC1404*ZHENG58 hybrid population) in five environments (JiLin (JL), LiaoNing (LN), BeiJing (BJ), HeBei (HB), HeNan(HN)) (Figure 2 and Supplemental Figures 2 and 3). As shown in Figure 2A–2C, GEFormer achieved higher accuracy in PH prediction compared with the other ten methods in most environments. GEFormer performed best in the CUBIC1404 inbred population (0.5893) and the CUBIC1404*JING724 hybrid population (0.4683), with accuracies at least 2.20% and 2.87% higher than those of the comparison methods. For the CUBIC1404*ZHENG58 population, the average prediction accuracy of GEFormer was 0.5392, slightly lower (by 0.12%) than that of the best method (EA + GW(GK)). As shown in Figure 2D–2F, GEFormer outperformed the comparison methods in ear weight (EW) prediction accuracy across all environments in the different populations. Specifically, its average EW prediction accuracies were 0.4013, 0.3185, and 0.4161, which were at least 3.57%, 3.99%, and 2.56% higher than those of the comparison methods. In addition, GEFormer had the highest average accuracy for prediction of days to anther (DTA) in the three populations (Supplemental Figure 2).

We also compared GEFormer, RF, XGBoost, Stacking, DL, EA + GW(GB), EADW + GW(GB), EA + GW(GK), and EADW + GW(GK) using the Maize976 hybrid population (Supplemental Figure 3). GEFormer outperformed all comparison methods for prediction of PH and EW in all environments. The average prediction accuracies for PH and EW were 0.7618 and 0.4453, surpassing those of the best comparison methods by 4.34% (EA + GW(GK)) and 10.26% (EA + GW(GK)).

Overall, the prediction accuracy of the DL method was significantly lower than that of the other methods. Performance of the other three machine-learning methods (RF, XGBoost, and Stacking) was similar to that of the six statistical learning methods. GEFormer outperformed the comparison methods for EW prediction across different populations and environments. GEFormer also had the

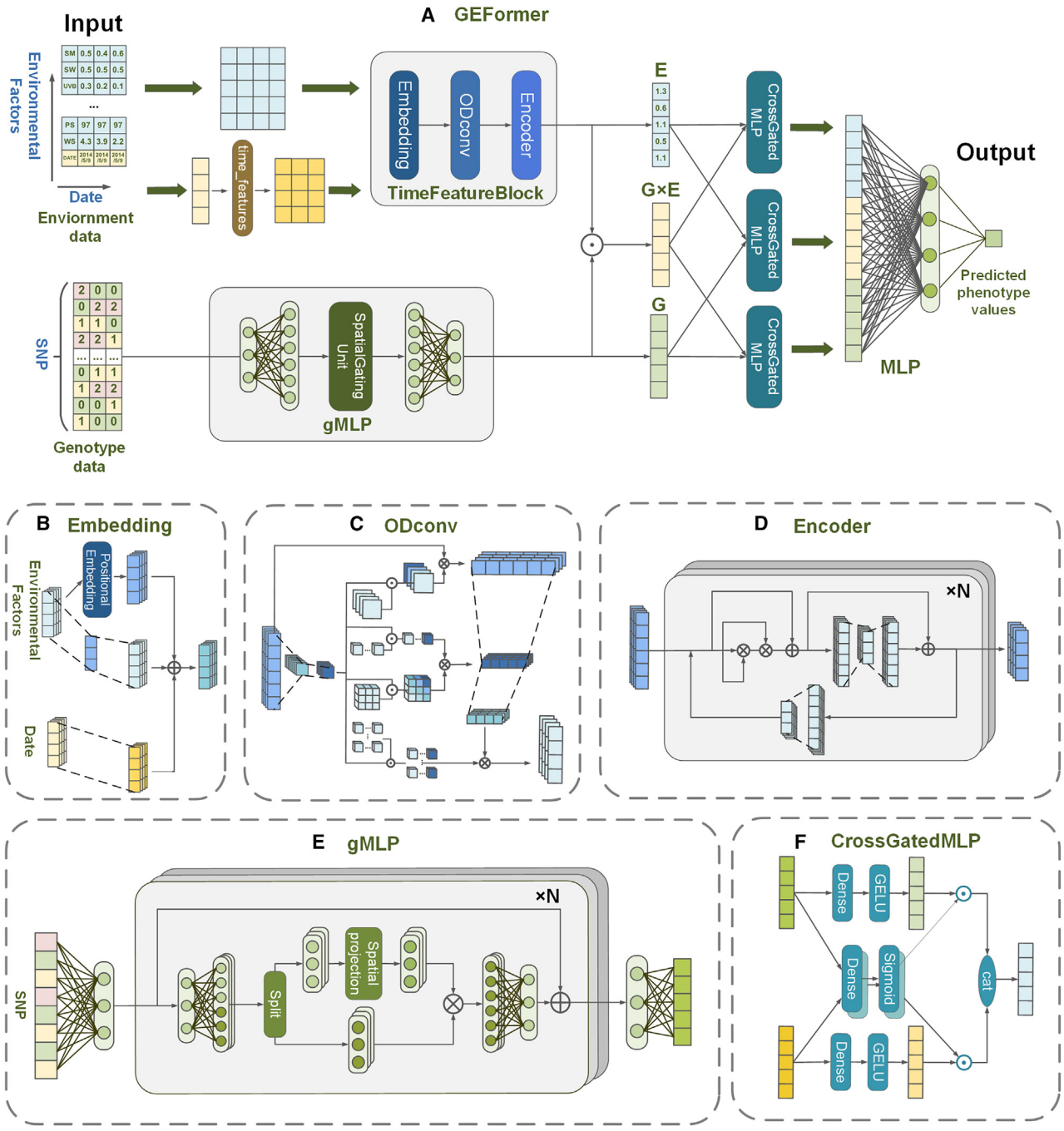

Figure 1. The architecture of GEFormer.

(A) GEFormer consists of five layers: the input layer, gMLP layer, TimeFeatureBlock layer, CrossGatedMLP layer, and output layer.

(B) The input layer is used to standardize the input data.

(C) Omni-dimensional Dynamic Convolution (ODConv) is used to obtain the comprehensive effects of environmental factors in each day.

(D) The linear attention mechanism is used to capture temporal features among the environmental factors between different days.

(E) The gMLP layer extracts the genotype features G*. The TimeFeatureBlock layer captures the environmental features E*.

(F) The CrossGatedMLP layer integrates the features of G*, E*, and G* × E* using cross-gating mechanisms. The output layer uses MLP to further extract the fused features and maps them to the phenotype values.

highest prediction accuracy for PH and DTA in most environments. These results indicate that GEFormer shows better performance in capturing the features of complex traits (i.e., EW) and also shows good prediction effects for simple traits such as PH and DTA.

Evaluation of phenotype prediction in scenario M2

In the scenario of tested genotypes in untested environments, we compared GEFormer with four machine-learning methods and six statistical learning methods using three maize populations.

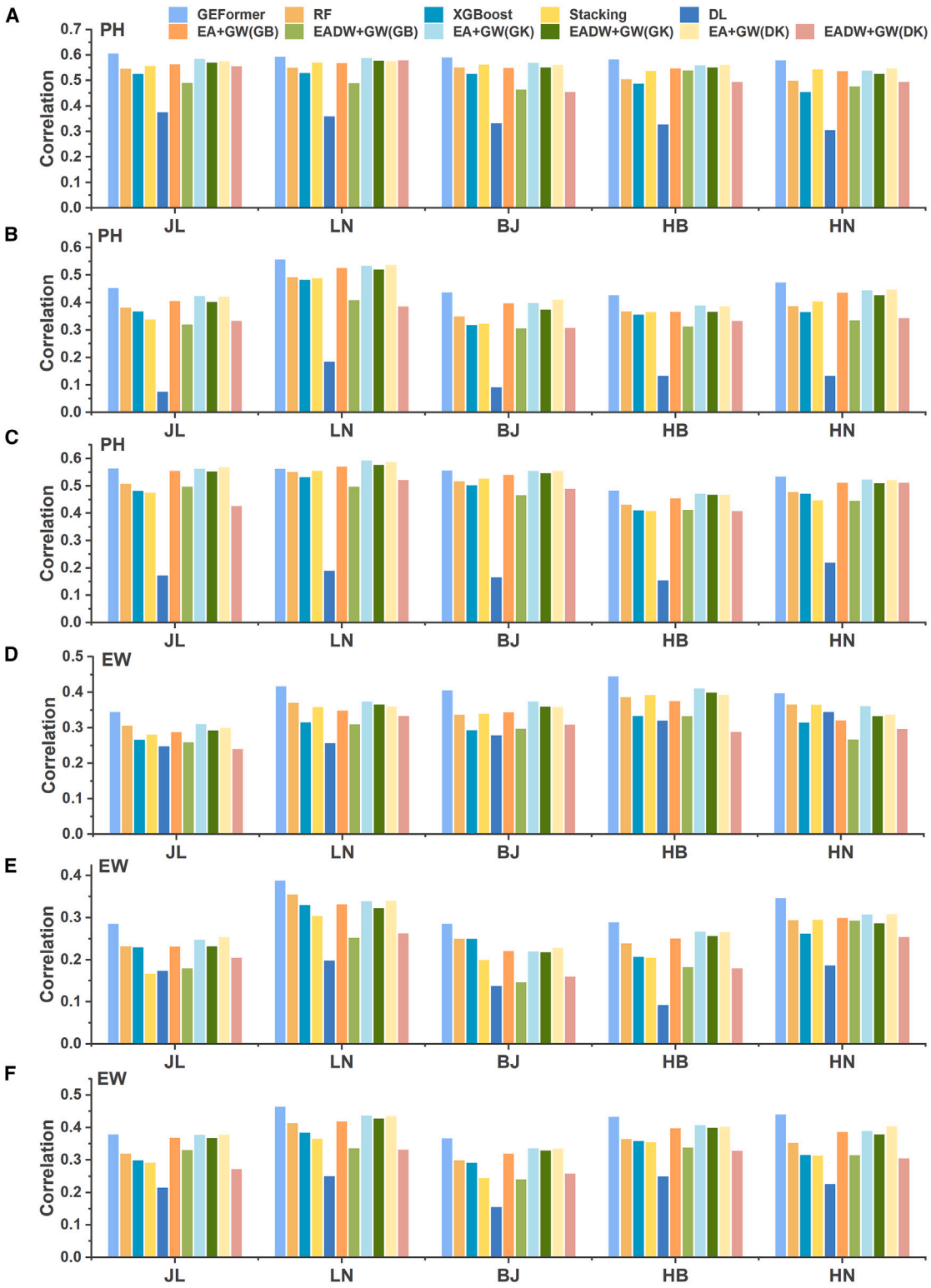

Figure 2. Comparison of prediction accuracy of GEFormer with 10 representative methods for three maize datasets in the scenario of untested genotypes in tested environments (M1).

(A–C) PH prediction accuracy for the CUBIC1404 inbred population, the CUBIC1404*JING724 hybrid population, and the CUBIC1404*ZHENG58 hybrid population.

(D–F) EW prediction accuracy of the 11 methods for the three populations.

For PH, EW, and DTA in different populations, DL, EADW + GW(GB), and EADW + GW(DK) had lower average prediction accuracy across different environments (Figure 3, Supplemental Figure 4). GEFormer and eight methods (RF, XGBoost, Stacking, EA + GW(GB), EA + GW(GK), EADW + GW(GK), and EA + GW(DK)) had stable performance across different environments. For EW, GEFormer had the highest average prediction accuracy for the three populations, at least 0.46%, 2.47%, and 0.24% higher than that of the comparison methods. The prediction accuracy for PH and EW in the Maize976 hybrid population is shown in Supplemental Figure 5. For PH in 2011, GEFormer had the highest average prediction accuracy (0.8156), 0.56% higher than that of the best comparison method (RF). For PH in 2012, the average prediction accuracy of GEFormer was 0.8496, slightly lower (by 0.12%) than that of the best method (XGBoost). The performance of different methods for EW varied greatly in most environments. GEFormer had an average prediction accuracy of 0.4773 for EW in 2011, outperforming RF, XGBoost, Stacking, DL, and EADW + GW(GB). Furthermore, the average prediction accuracy of GEFormer for EW in 2012 was 0.5048, slightly lower (by 0.69%) than that of the best comparison method (EA + GW(GK)). Overall, GEFormer had a higher average prediction accuracy and showed relatively stable performance for the scenario of tested genotypes in untested environments.

Evaluation of phenotype prediction in scenario M3

We next compared the phenotype prediction accuracy of GEFormer with that of four classic machine-learning methods and six statistical learning methods for the scenario of untested genotypes in untested environments (Figure 4, Supplemental Figure 6). GEFormer had the highest PH prediction accuracy in all environments for three maize populations. The average accuracies for the CUBIC1404 inbred line population, CUBIC1404*JING724 hybrid population, and CUBIC1404*ZHENG58 hybrid population were 0.5887, 0.4812, and 0.5516, respectively, and were at least 1.57%, 2.57%, and 2.14% higher than those of the comparison methods. GEFormer also had the highest EW prediction accuracy in all environments. The average prediction accuracies of GEFormer for the three populations were 0.3627, 0.2652, and 0.3668, respectively, and were 2.57% (EA + GW(GK)), 5.70% (EADW + GW(GK)), and 1.96% (EA + GW(GK)) higher than those of the best comparison method. In addition, the average prediction accuracy of GEFormer for DTA in the three populations was higher than that of other methods by at least 3.01%, 3.26%, and 3.44%.

Overall, the average prediction accuracy of GEFormer was better than that of the other methods in five environments for three populations. These findings indicate that GEFormer can effectively capture the genotypic and environmental features of different populations and shows better generalization ability in the scenario of untested genotypes in untested environments.

Breeding application analysis of GEFormer

Phenotype prediction of GEFormer in unknown environments

For phenotype prediction in untested environments, we partitioned the known environments into different training and test sets to train the models. There are two strategies for exploring the optimal number of environments for training populations. The first is to exhaust all combinations, but this strategy consumes a lot of time and computational resources. The second is the iterative optimization strategy used in our experiment. If there are n environments for the training population, the first strategy requires training $\sum_{i=1}^{n-2} A_n^i$ combinations. The second strategy requires training a number of combinations between $\sum_{i=1}^{2} C_n^i$ and $\sum_{i=1}^{n-2} C_n^i$ to obtain the relatively optimal combination. When the optimal training population consists of two environments, the maximum number of combinations that need to be trained is $\sum_{i=1}^{n-2} C_n^i$. When the optimal accuracy of the next subset after iterative optimization is lower than that of the previous subset, the number of training combinations is $\sum_{i=1}^{2} C_n^i$. Therefore, the combination number of the second strategy is much smaller than that of the first strategy. For this reason, we used the second strategy to obtain the optimal number of environments for the training populations.

Taking the CUBIC1404 population in the JL environment as an example, we selected one environment from the known environments of LN, BJ, HB, and HN for model validation and used the remaining three environments for model training. We then used the four trained models to predict the phenotype in the JL environment independently, and we used three environmental training datasets with the best accuracy for further analysis. We next selected one environment for model validation and used the other two for model training in these three environments. Similarly, we selected the model with the highest prediction accuracy and compared it with the previous optimal accuracy. If the optimal accuracy of the two-environment training set was lower than that of the three-environment training set, this indicated that the three-environment training set was required for prediction of the target environment. This also meant that the optimal environment training set for predicting the phenotype in the JL environment had been found. Otherwise, we continued to analyze the training environments with the optimal accuracy. Specifically, we used one of the two environments for model validation and the other for model training. By evaluating the prediction accuracy of the two models and comparing them with the optimal accuracy obtained in the previous step until the highest prediction accuracy was obtained, the optimal training environment for predicting the phenotype in the JL environment was identified. The above strategy helped to identify the relatively optimal training environment for predicting unknown environmental phenotypes with fewer experiment combinations, instead of exhaustively judging all combinations of known environments. For the CUBIC1404 inbred population, the independent prediction accuracies of the models trained with different combinations of known environments are shown in Supplemental Figures 7 and 8.

When the environments of JL, LN, BJ, HB, and HN were unknown, the PH and EW prediction accuracy of the models trained on three environments was generally better than that of the models trained on two environments, with their average optimal prediction accuracies being 0.7721 and 0.5219, respectively. When the DTAs of the JL, HB, and HN environments were unknown, the models trained on two environments outperformed the models trained on three environments by 2.73%, 1.33%,

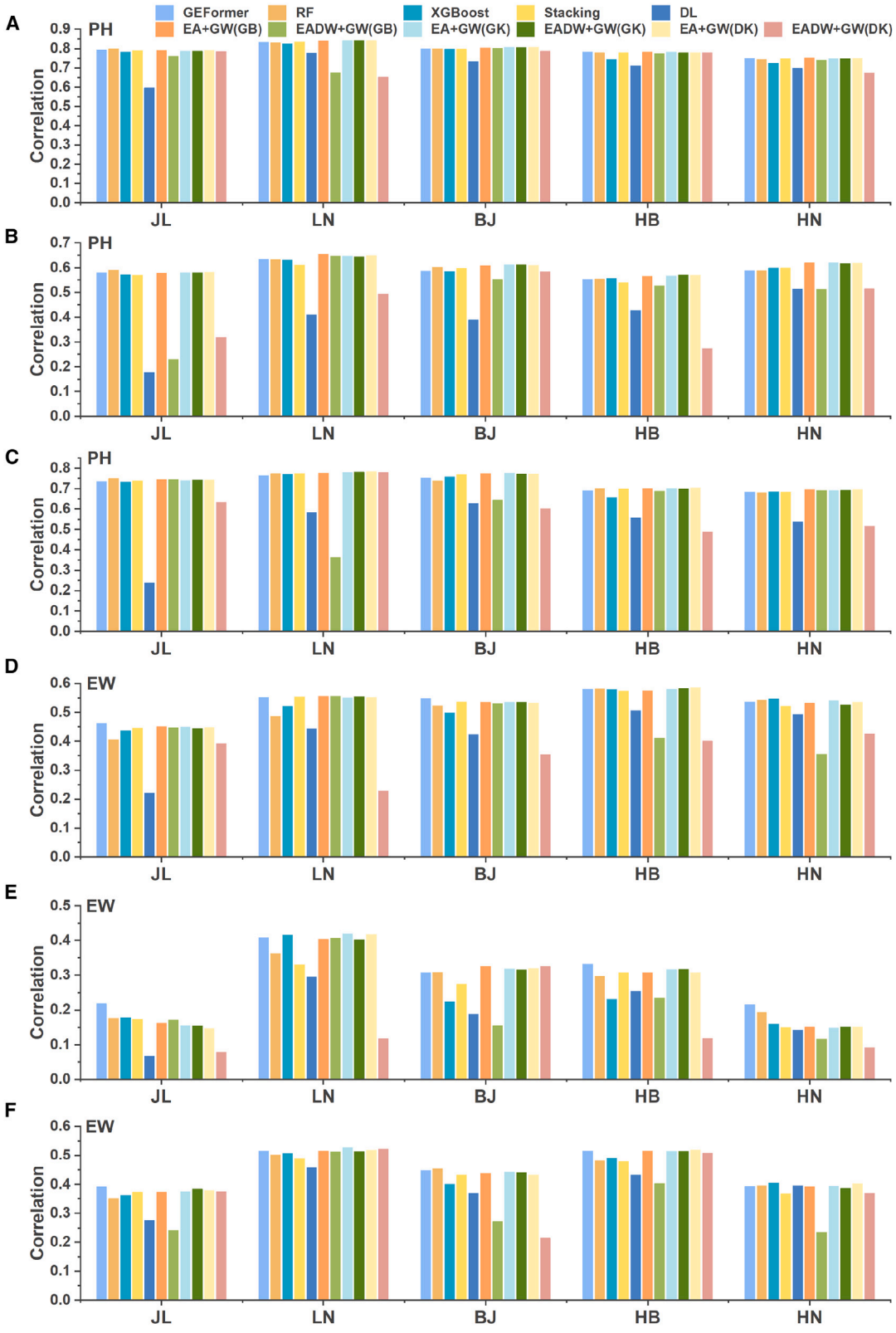

Figure 3. Comparison of prediction accuracy of GEFormer with 10 representative methods for three maize populations in the scenario of tested genotypes in untested environments (M2).

(A–C) PH prediction accuracy for the CUBIC1404 inbred population, the CUBIC1404*JING724 hybrid population, and the CUBIC1404*ZHENG58 hybrid population.

(D–F) EW prediction accuracy for the three populations.

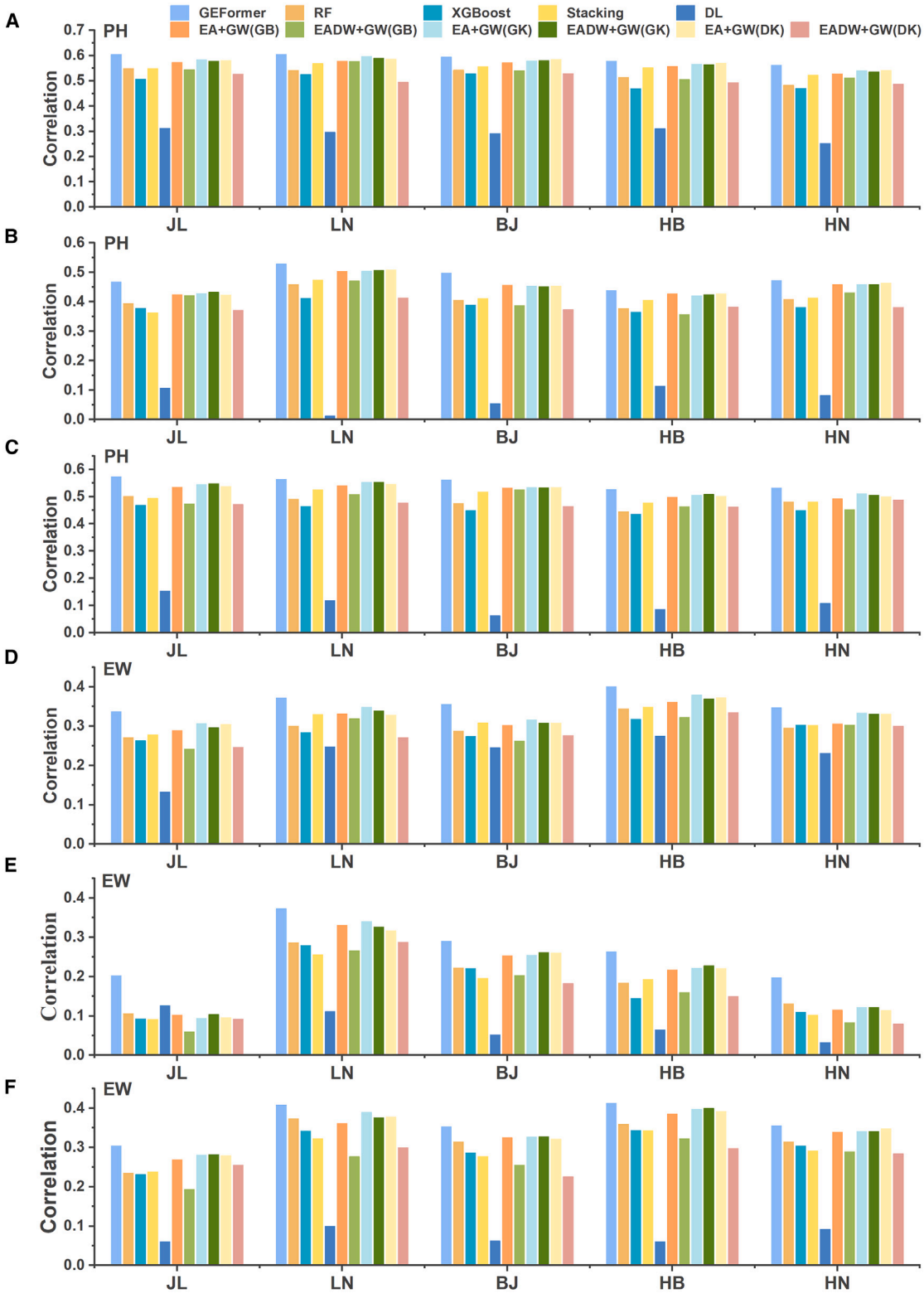

Figure 4. Comparison of prediction accuracy of GEFormer with 10 representative methods in the scenario of untested genotypes in an untested environment (M3).

(A–C) PH prediction accuracy for the CUBIC1404 inbred population, the CUBIC1404*JING724 hybrid population, and the CUBIC1404*ZHENG58 hybrid population in maize.

(D–F) EW prediction accuracy for the three maize populations.

and 1.75%, respectively, although these differences were relatively small. Therefore, the models trained on more environments generally demonstrated better prediction ability in unknown environments.

We used maize hybrid populations to confirm that the prediction models trained on more environmental data had better accuracy. We trained the models using data from three environments for independent prediction in the CUBIC1404*JING724 and CUBIC1404*ZHENG58 hybrid populations (Supplemental Table 1). For the CUBIC1404*JING724 population, the optimal average prediction accuracies for PH, EW, and DTA were 0.576, 0.2819, and 0.4704, respectively. For the CUBIC1404*ZHENG58 population, the optimal average prediction accuracies of PH, EW, and DTA were 0.7110, 0.4314, and 0.5347, respectively. The experimental results also showed that different divisions of the training and validation sets could lead to variations in prediction accuracy. Notably, the optimal training environments obtained using the hybrid populations were generally the same as the results obtained using the CUBIC1404 population. This indicates that GEFormer can better predict the phenotypes of the two hybrid populations in unknown environments and has good generalization ability. The experimental results also showed that models trained on data from more environments were more accurate than those trained on fewer environments.

To evaluate the stable and high-yielding materials recommended using the optimal prediction model for unknown environments, we calculated the overlap (at least one shared material) between the top 5% of high-yielding materials predicted by GEFormer and the actual top 5% materials. We also randomly selected 5% of the materials from the CUBIC inbred populations, calculated their overlap with the actual top 5% high-yielding materials over 100 repetitions, and then compared these results with the accuracy of the GEFormer recommendations, as shown in Supplemental Table 2 and Figure 5A–5E. When BJ was the unknown environment, GEFormer recommended 27 materials that could achieve high yields when planted in BJ, with an accuracy of 38.57%; this was higher than the accuracy of the random recommendations (4.87%). For HB, GEFormer recommended 22 materials with 31.43% accuracy, much higher than the 5.51% accuracy of the random recommendations. For HN, GEFormer recommended 28 materials with 38.57% accuracy, much higher than the 5.06% accuracy of the random recommendations. For JL, GEFormer recommended 16 materials with 22.86% accuracy, much higher than the 5.40% accuracy of the random recommendations. For LN, GEFormer recommended 22 materials with 31.43% accuracy, much higher than the 5.51% accuracy of the random recommendations. Three of the recommended materials (MG_739, MG_1000, and MG_1232) were stably high yielding in all five environments (Figure 5F). Similarly, we analyzed the stable and high-yielding materials recommended in the CUBIC1404*JING724 and CUBIC1404*ZHENG58 hybrid populations, as shown in Supplemental Figures 9 and 10. The results indicated that the inbred populations contained materials that were relatively stable and high yielding across multiple environments. However, the hybrids may be more susceptible to environmental factors, and it may be challenging to identify materials that consistently achieve high yields in most environments from hybrid populations.

Improving the prediction accuracy of hybrids using inbred lines

The 6210 hybrid population was obtained by crossing 207 materials from the CUBIC1404 inbred population and 30 testing materials (Liu et al., 2020; Yang et al., 2022). We trained the GEFormer model under the scenario of untested genotypes in untested environments using the 207 maternal materials or both the maternal and paternal materials (207 + 30) as the training population. We used the trained GEFormer models to predict the PH, EW, and DTA phenotypes of the 6210 hybrid population, and the results are shown in Supplemental Figure 11. The average prediction accuracies for PH, EW, and DTA were 0.4644, 0.2430, and 0.5771; these were 31.25%, 16.99%, and 51.62% higher than those of the model trained using only the maternal materials. These results demonstrate that incorporating both maternal and paternal inbred materials improved the prediction accuracy for the hybrid population.

Because EA + GW(GK) performed the best in the M3 prediction scenario, we compared EA + GW (GK) with GEFormer, and the results are shown in Supplemental Table 3. The prediction accuracy when the maternal and paternal materials (207 + 30) were used as the training population was significantly better than that obtained using the 207 maternal materials alone. For PH and DTA from the 6210 population, the prediction accuracy of GEFormer was at least 20% higher than that of EA + GW(GK). For EW, the prediction accuracy of GEFormer was slightly lower than that of EA + GW(GK) in HN and higher than that of EA + GW(GK) in the remaining four regions. In summary, the prediction accuracy was significantly improved for all methods when the training population contained both the maternal and paternal materials. In addition, GEFormer showed the greatest improvement and the most stable performance.

We next used general combining ability (GCA) and special combining ability (SCA) to analyze whether the usage of parental materials could predict phenotypes more accurately and further assist in obtaining high-yield materials (Schrag et al., 2009). We calculated the GCA and SCA for each environment based on the EWs of the CUBIC6210 population predicted from the 207 and 237 training populations. For GCA, we calculated the overlap between the top 50% high-GCA materials predicted using the 237 training population and the actual top 50% high-GCA materials (top50%_237_GCA), as well as the overlap between the top 50% high-GCA materials predicted using the 207 training population and the actual top 50% high-GCA materials (top50%_207_GCA) (Table S4). For all parent materials in different environments, the number of high-GCA materials screened by the top50%_237_GCA was higher than that screened by the top50%_207_GCA. Among them, the number of overlapping materials accounted for half of the screened materials with high combining ability separately. These materials were relatively conserved and were insensitive to the data of the training population (Supplemental Figure 12). We denoted the unique materials in top50%_237 and top50%_207 as unique_237_GCA and unique_207_GCA, respectively. The results showed that the average EW of unique_237_GCA was higher than that of unique_207_GCA (Supplemental Table 5). At the same time, the average standard error of unique_237_GCA was lower than that of unique_207_GCA, indicating that the unique_237_GCA material has strong stability. These results further demonstrate

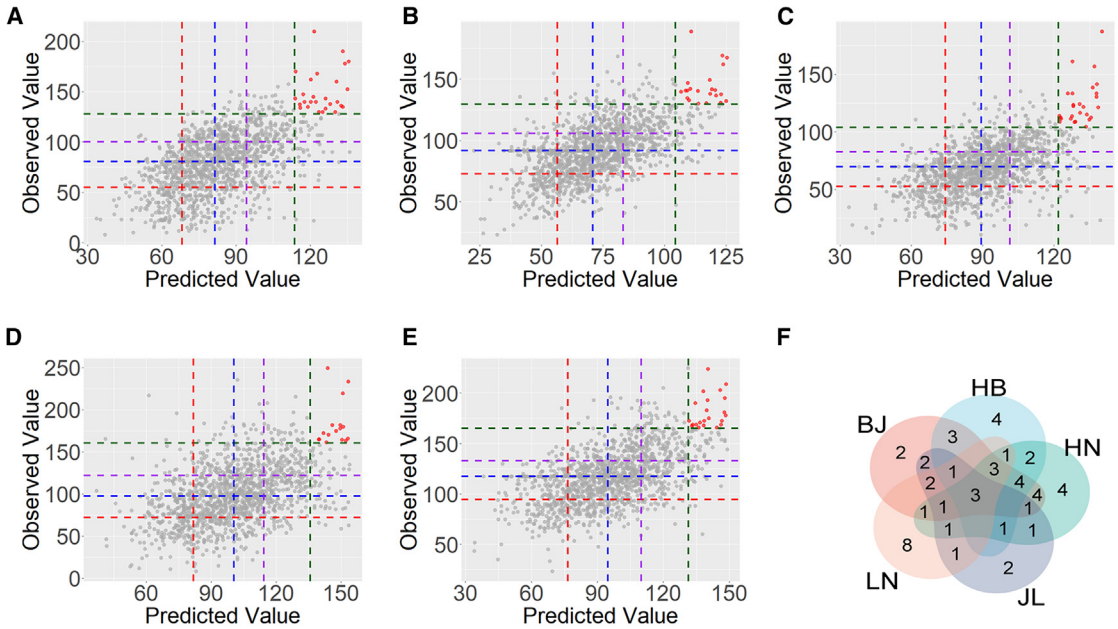

Figure 5. Prediction of EW for the CUBIC1404 inbred population in unknown environments using GEFormer.

(A–E) Overlap between the EW phenotypes predicted by GEFormer and the actual yields in the unknown environments of BJ, HB, HN, JL, and LN. The upper right corner of the green line represents materials for which the actual yield and the predicted EW value were both in the top 5%. The upper right corner of the purple line represents materials for which the actual yield and the predicted EW value were both in the top 25%. The upper right corner of the blue line represents materials for which the actual yield and the predicted EW value were both in the top 50%. The upper right corner of the red line represents materials for which the actual yield and the predicted EW value were both in the top 75%.

(F) Number of overlapping materials for which actual and predicted EW values from the five environments (BJ, HB, HN, JL, and LN) were both in the top 5%.

that the 237 training population is more conducive to capturing materials with higher GCA of EW than the 207 training population. In addition, the results of the parent material in all environments were basically consistent with the results of the parent material (Supplemental Figure 13 and Supplemental Table 5).

For SCA, we calculated the overlap between the top 50% high-SCA materials predicted using the 237 training population and the actual top 50% high-SCA materials (top50%_237_SCA), as well as the overlap between the top 50% high-SCA materials predicted using the 207 training population and the actual top 50% high-SCA materials (top50%_207_SCA) (Supplemental Table 6). For all hybrid materials in all environments, the number of high-SCA material combinations screened by the top50%_237_SCA was higher than that screened by the top50%_207_SCA. Among them, the number of overlapping materials accounted for more than half of the screened materials with high combining ability separately. These materials were relatively conserved and were insensitive to the data of the training population (Supplemental Figure 14). We denoted the unique materials in top50%_237 and top50%_207 as unique_237_SCA and unique_207_SCA, respectively. The results showed that the average EW of unique_237_SCA was higher than that of unique_207_SCA (Supplemental Table 7). The average standard error of unique_237_SCA was lower than that of unique_207_SCA, indicating that the unique_237_SCA material has strong stability. These results further demonstrate that the 237 training population is more conducive to selecting materials with higher SCA of EW than the 207 training population.

In view of the low-dimensional and high-dimensional relationships between the phenotypes and the features captured by the specific module in GEFormer, we analyzed the reasons for the high prediction accuracy of GEFormer trained using both maternal and paternal materials. Taking the EW prediction results in the LN environment for the 6210 hybrid population as an example, the two-dimensional visualization of material features and the significance analysis results of the features extracted using the GEFormer_M3_M model and the GEFormer_M3_P model are shown in Figure 6 (Malik and Piepho, 2018). We used the t-SNE (t-distributed stochastic neighbor embedding technology) feature visualization method (Van der Matten and Hinton, 2008; Platzer, 2013) to reduce the high-dimensional features extracted by each module in GEFormer to a two-dimensional plane feature (Chatzimparmpas et al., 2020). Figures 6A and 6G illustrate the two-dimensional visualization results of the genotype features extracted by the gMLP layer in GEFormer_M3_M and GEFormer_M3_P, respectively. In addition, we used *t*-tests to analyze the significance (see the section "Statistical analysis") of correlations about the phenotypes and features extracted by the gMLP layer based on the GEFormer_M3_M model (Figure 6D) and the GEFormer_M3_P model (Figure 6J). The results showed that 99.21% and 30.95% of the features extracted by GEFormer_M3_P and GEFormer_M3_M were significant at $p < 0.001$, and 100% and 49.21% were significant at $p < 0.05$. Furthermore, 10 significant features out of the top 10% of significant features captured by GEFormer_M3_P were not considered significant in the GEFormer_M3_M model, accounting for 76.92% of the top 10% of significant features. We then used two-way analysis of variance (ANOVA) (Wei et al.,

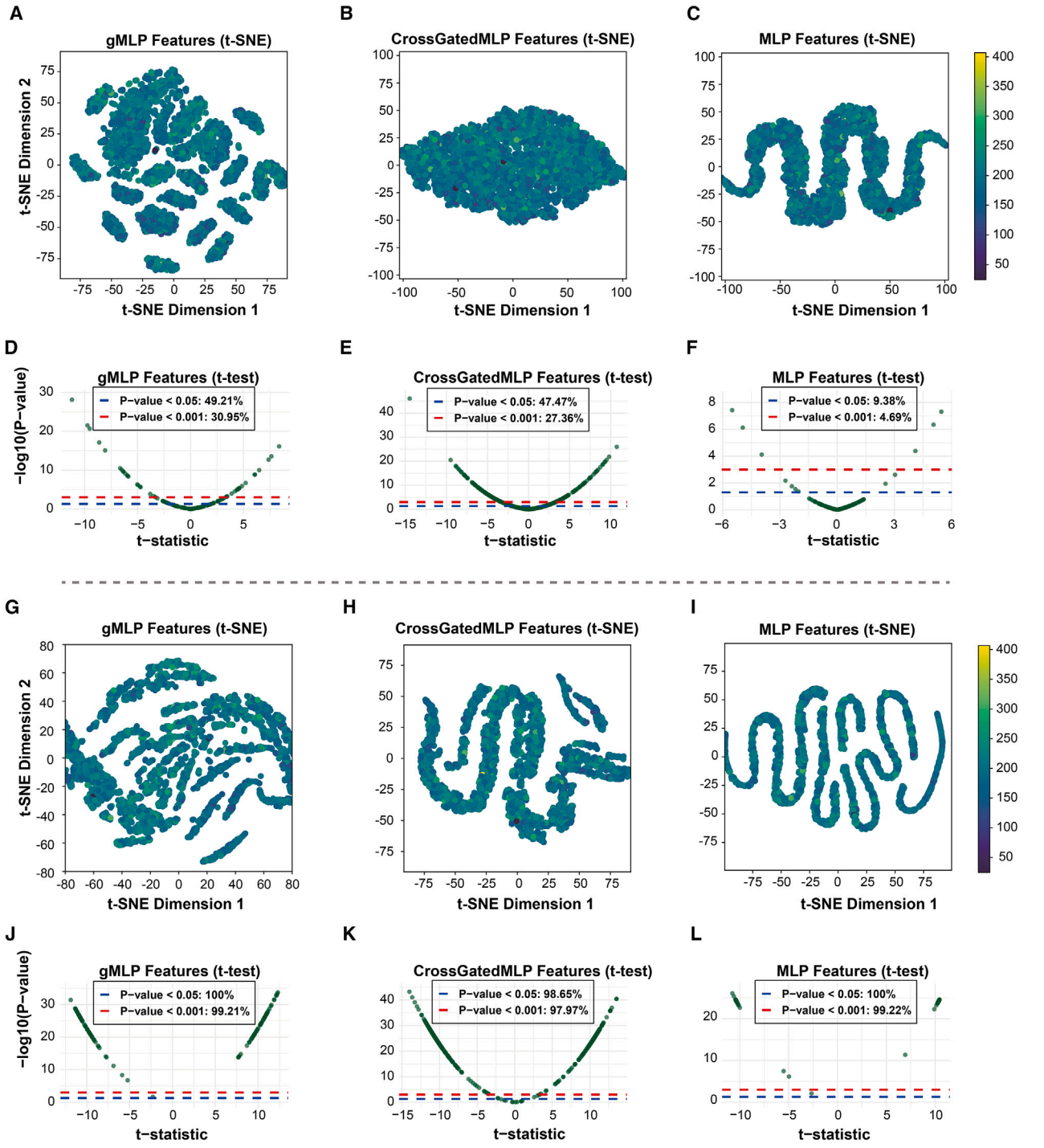

Figure 6. Two-dimensional visualization of material features and significance analysis results of the features extracted from the GEFormer_M3_M model trained using only maternal materials and the GEFormer_M3_P model trained using both maternal and paternal materials for prediction of EW in the 6210 hybrid population in the LN environment.

(A and G) Two-dimensional visualization results of the genotype features extracted by the gMLP layer of GEFormer_M3_M and GEFormer_M3_P based on t-SNE. Each node represents the feature after dimensionality reduction of each material using t-SNE, and the color of each node represents the phenotype value of the material.

(D and J) The *t* test significance analysis results for the phenotypes and genotype features extracted by the gMLP layer of GEFormer_M3_M and GEFormer_M3_P.

(legend continued on next page)

2024) to obtain the epistatic interaction features related to these 10 features ($p < 0.05$), accounting for 74.94% of the epistatic interactions related to the top 10% of significant features. Therefore, GEFormer_M3_P could capture more complex nonlinear features related to epistatic effects. Figures 6B and 6H show the two-dimensional visualization results of the features obtained by the fusion of genotype and environmental features in the CrossGatedMLP layer in GEFormer_M3_M and GEFormer_M3_P, respectively. We performed *t*-test correlation analysis on the phenotypes and features extracted by the CrossGatedMLP layer based on GEFormer_M3_M (Figure 6E) and GEFormer_M3_P (Figure 6K). We found that 97.97% and 27.36% of the features extracted by GEFormer_M3_P and GEFormer_M3_M were significant at $p < 0.001$, and 98.65% and 47.47% were significant at $p < 0.05$. Furthermore, 34 significant features out of the top 10% of significant features captured by GEFormer_M3_P were not considered significant in the GEFormer_M3_M model, accounting for 46.57% of the top 10% of significant features. We then used two-way ANOVA (Wei et al., 2024) to obtain the epistatic interaction features related to these 34 features ($p < 0.05$), accounting for 76.68% of the epistatic interactions related to the top 10% of significant features. Compared with the genotypic and environmental factor features before fusion, richer features could be captured using the fusion operation. This means that GEFormer can capture complex and nonlinear fusion features and that GEFormer_M3_P captured more complex variant features than GEFormer_M3_M. Figures 6C and 6I show the two-dimensional visualization results of the further fused features using MLP. We performed *t*-test correlation analysis on the phenotypes and features extracted by the MLP layer based on GEFormer_M3_M (Figure 6F) and GEFormer_M3_P (Figure 6L). The results showed that 99.22% and 4.69% of features extracted by GEFormer_M3_P and GEFormer_M3_M were significant at $p < 0.001$, and 100% and 9.38% were significant at $p < 0.05$. Thus, GEFormer_M3_P can capture more complex nonlinear relationships and richer variation features than GEFormer_M3_M. The GEFormer_M3_P model trained using both maternal and paternal data captured more complex nonlinear features of genotypes, environmental factors, and genotype–environment interactions. Therefore, the prediction accuracy of GEFormer_M3_P was higher than that of GEFormer_M3_M trained using only maternal data. The feature visualization results (Supplemental Figures 15–28) for other environments and traits were similar to those obtained for EW in the LN environment.

We also used the GEFormer_M3 model trained on three populations (the CUBIC1404 inbred population, 1404 + 1404*ZHENG58 hybrid population, and 1404 + 1404*JING724 hybrid population) to predict the phenotypes of two hybrid populations (1404*JING724 and 1404*ZHENG58). The prediction accuracy results are shown in Supplemental Table 8. The prediction accuracy increased after addition of the hybrid population to the training set of inbred lines, as shown in Supplemental Table 9. For the 1404*JING724 hybrid population, the average prediction accuracy for PH, EW, and DTA in different environments was increased by 44.09%, 11.77%, and 34.77% upon addition of the 1404*ZHENG58 hybrid population to the training set of the CUBIC1404 population. For the 1404*ZHENG58 hybrid population, the average prediction accuracy for the three traits was increased by 47.49%, 23.07%, and 43.52% after addition of the 1404*JING724 hybrid population to the training set of the CUBIC1404 population. In all, the addition of one hybrid population to the training set of the inbred population significantly enhanced the prediction accuracy of another hybrid population.

Predicting phenotypes of hybrids across populations using GEFormer

In this experiment, we randomly selected 100 materials from the Maize976 hybrid population as the test set and denoted the other materials in this population as Maize976*. We constructed GEFormer_M2 and GEFormer_M3 for the M2 and M3 scenarios using the Maize976* population, the CUBIC1404 inbred population, and the CUBIC1404 + Maize976* population. Table 1 illustrates the PH and EW prediction results obtained using these models on the test set.

Compared with their results when trained on the Maize976* population, GEFormer_M2 and GEFormer_M3 showed average (HN2011 and HN2012) reductions in prediction accuracy of 28.0% and 45.62% for PH and 52.11% and 26.21% for EW when trained on the CUBIC1404 population. However, the two models trained on the CUBIC1404 + Maize976* dataset achieved the highest prediction accuracy for most phenotypes. This indicates that incorporation of some inbred materials into the training set of the hybrid population significantly boosted the prediction accuracy for the hybrid population. Similarly, we randomly selected 100 materials from the CUBIC1404 inbred population as the test set and denoted the other materials in the CUBIC1404 population as CUBIC1404*. We performed experiments using the strategy described above for Maize976*, and the results are shown in Table 1. The GEFormer_M2 and GEFormer_M3 models trained on the CUBIC1404* + Maize976 dataset achieved the highest prediction accuracy for most phenotypes. Therefore, combining the inbred and hybrid populations could improve the accuracy of phenotype prediction for the inbred and hybrid populations.

Ablation experiment

We used general convolution and fully connected layers to replace the ODconv and CrossGatedMLP layer in the GEFormer architecture, respectively, and we performed ablation experiments using different combinations of genotype and environmental factors. The ablation experiment method for GEFormer is shown in Supplemental Table 10. In GEFormer-1, ODconv is replaced with general convolution (Conv2d). In GEFormer-2,

(B and H) Two-dimensional visualization results of the features obtained by the fusion of genotype and environmental features in the CrossGatedMLP layer of GEFormer_M3_M and GEFormer_M3_P based on t-SNE.
(E and K) The *t* test significance analysis results of the phenotypes and genotype features extracted by the CrossGatedMLP layer of GEFormer_M3_M and GEFormer_M3_P.
(C and I) Two-dimensional visualization results of the further fused features using the MLP of GEFormer_M3_M and GEFormer_M3_P based on t-SNE.
(F and L) The *t* test significance analysis results of phenotypes and the further fused features using the MLP of GEFormer_M3_M and GEFormer_M3_P.

Environment	Model	Training set	PH	EW
HN2011	GEFormer_M2	976*	0.7364	0.2842
		1404	0.4978	−0.3680
		976* + 1404	0.8039	0.3126
	GEFormer_M3	976*	0.7726	0.0760
		1404	0.2672	−0.2377
		976* + 1404	0.7968	−0.0426
HN2012	GEFormer_M2	976*	0.7168	0.2127
		1404	0.3954	−0.1775
		976* + 1404	0.7258	0.1593
	GEFormer_M3	976*	0.6942	0.1469
		1404	0.2872	−0.0638
		976* + 1404	0.6921	0.1802
HN2014	GEFormer_M2	1404*	0.5336	0.3117
	–	976	0.1732	0.0978
	–	1404* + 976	0.5593	0.2896
	GEFormer_M3	1404*	0.5171	0.2638
	–	976	0.1670	0.0809
	–	1404* + 976	0.5373	0.3095

Table 1. PH and EW prediction results for cross populations.
976* represents the population after removal of 100 test materials from the Maize976 population.
1404* represents the population after removal of 100 test materials from the CUBIC1404 population.

CrossGatedMLP is replaced with MLP. The GEFormer-3 model excludes the environmental effect and the comprehensive effect of genotype and environmental factors. The GEFormer-4 model excludes the comprehensive effect of genotype and environmental factors. The GEFormer-5 model excludes the environmental effect, and the GEFormer-6 model excludes both the genotype and the environmental effects.

The results of the GEFormer ablation experiment are shown in Supplemental Figure 29. The average prediction accuracies of GEFormer_M1 for PH and DTA were 58.93% and 57.18%, at least 0.2% and 0.2% higher than those of the comparison methods, demonstrating better stability (Supplemental Figure 29A and 29C). For EW, the prediction accuracy of GEFormer_M1 and most other models (GEFormer-1, GEFormer-2, GEFormer-4, GEFormer-5, GEFormer-6) ranged from 40.0% to 40.81% (Supplemental Figure 29B). The GEFormer-3 model that considered only the genotype effect had poor stability and lower prediction accuracy than the other models for all phenotypes. As shown in Supplemental Figure 29D, GEFormer_M2 demonstrated the best performance (79.20%) and stability for PH prediction. The EW prediction accuracy of all models fell within the range of 52.61%–53.88% (Supplemental Figure 29E), and GEFormer_M2 had the best stability. As shown in Supplemental Figure 29F, GEFormer_M2 performed the best (68.51%) for DTA prediction. The results of GEFormer_M3 were similar to those of GEFormer_M1 and GEFormer_M2. These findings confirm the necessity of the ODconv layer and CrossGatedMLP layer in GEFormer. The fusion of genotype, environment, and the comprehensive effects of genotype and environmental factors can strengthen prediction accuracy. In addition, we performed ablation experiments of the statistical method (Supplemental Table 11). The ablation experiment results showed that the prediction accuracy of GEFormer was superior to that of statistical learning methods in most cases (Supplemental Table 12), highlighting the strong robustness of GEFormer.

Use of GEFormer to capture key environmental features

Different environmental factors have a comprehensive impact on crop growth and development. The TimeFeatureBlock module in GEFormer uses ODConv to dynamically extract the comprehensive effects of multiple environmental factors within each day. This module also extracts temporal features among days through a linear attention mechanism. We used the DL interpretable algorithm of the saliency map (Simonyan, 2013) to calculate the impact of all environmental factors and individual environmental factors on the phenotype of each day during the entire growth period of maize. We used CERIS (the Critical Environmental Regressor through Informed Search) (Li et al., 2021) to identify the key growth periods affecting three maize traits and compared them with the results of GEFormer.

For the key environmental factors affecting maize EW, the overlap between the top 50% of environmental factors identified by CERIS (Supplemental Figures 33 and 34) and the top 50% of environmental factors identified by GEFormer (Supplemental Figure 35 and Supplemental Table 13) was 53.85% (Supplemental Figure 37A). Specifically, two of the top three environmental factors were the same. The top three environmental factors captured by CERIS were all related to

temperature, whereas two of the top three environmental factors captured by GEFormer were related to temperature and the other to water. Temperature is known to be closely related to the metabolism, development, and growth of crops (Kerbler and Wigge, 2023). Both methods identified environmental factors related to temperature, further demonstrating the feasibility of the two methods for detection of key environmental factors. Owing to the biophysical interdependence among temperature, water, and crop physiology, these factors jointly affect crop yield (Lesk et al., 2022). Lobell et al. reported that there are significant correlations between temperature, rainfall (water related), and yield in maize (Lobell et al., 2011). Akpalu et al. reported that temperature increase and rainfall have a greater effect on crop yield and that the effect of rainfall on crop yield is also more important than that of temperature (Akpalu et al., 2011). The key environmental factors identified by GEFormer as affecting PH and DTA were similar to those identified for EW (section Discussion and Supplemental Figures 37–45).

For the key growth periods affecting maize EW, the optimal growth period identified by CERIS (Supplemental Figure 34) was the depression and maturity period. These two periods are mainly related to water and nutrient accumulation. The optimal growth periods identified by GEFormer (Supplemental Figure 36) were the milk stage, V5–V6 stage, and maturity stage. These periods basically cover the vegetative growth and reproductive growth stages of maize, and they are mainly related to water and temperature. These results are consistent with the important environmental factors detected using GEFormer. Although the importance of the environmental factors detected in different periods varied, the contribution values of the environmental factors of 81.9% of periods were not significantly different ($p > 0.05$, t-test). The key growth periods identified as affecting PH and DTA by GEFormer were similar to those identified for EW (section Discussion and Supplemental Figure 37–45).

In summary, GEFormer can capture more realistic crop growth patterns, as well as key environmental factors and temporal features at different growth stages. The experimental results indicated that GEFormer had better prediction accuracy than CERIS in all environments, as shown in Supplemental Table 14. GEFormer can detect more comprehensive key environmental factors and growth periods than CERIS.

Exploring the potential of GEFormer for G × E GP in different species

To explore the generalization ability of GEFormer, we compared GEFormer with other methods using datasets from wheat and rice under scenarios M1, M2, and M3. For scenario M1 in rice (Supplemental Figure 48), GEFormer achieved higher prediction accuracy for grain yield (GY) and rice percentage of head rice recovery (PHR) than other methods in most environments; its accuracy was at least 3.13% (GY) and 12.76% (PHR) higher than that of the comparison methods. For the wheat dataset, the average prediction accuracies of GEFormer for yield, height, and heading date (DTH) were 0.5287, 0.6197, and 0.5020 and were at least 4.85%, 1.99%, and 4.3% higher than those of the comparison methods. For scenario M2 (Supplemental Figure 49), the prediction results of all methods were unstable for all phenotypes. In the rice dataset, the average prediction accuracies of GEFormer for GY and PHR were 0.4883 and 0.5584, at least 2.21% and 2.84% higher than those of the comparison methods. In the wheat dataset, the average prediction accuracies for yield and height did not differ significantly among the methods. Under scenario M3 (Supplemental Figure 50), the prediction accuracy of GEFormer was significantly better than that of other methods for all traits in rice and wheat. In the rice dataset, the average prediction accuracies of GEFormer for GY and PHR were at least 8.43% and 27.95% higher than those of the comparison methods. In the wheat dataset, the average prediction accuracies of GEFormer for yield, height, and DTH were at least 6.65%, 7.19%, and 5.27% higher than those of the comparison methods. In summary, GEFormer showed significant advantages with the rice and wheat datasets in most environments.

DISCUSSION

We developed a GP method (GEFormer) that integrates two modalities of genotype and environmental data. GEFormer can effectively mine the hidden complex nonlinear relationships between genotype and environmental data, thereby achieving multi-environment phenotype prediction. GEFormer contains agMLP layer, TimeFeatureBlock layer, and CrossGatedMLP layer. The gMLP layer automatically extracts features among SNPs through a gating mechanism with a fully connected neural network. Drawing on the real growth patterns of crops, the TimeFeatureBlock layer exploits the principle that crops respond to the environment by integrating the effects of all environmental factors. First, ODConv is used to fuse the comprehensive effect of different environmental factors within each day. Then, the linear attention mechanism is used to extract the temporal features of environmental factors during the crop growth period. The CrossGatedMLP layer fuses the extracted cross-modal features of genotype and environment. Compared with the classical G × E model, GEFormer uses DL technology to automatically capture the complex nonlinear features of genotype and environmental data.

Model performance

We compared the prediction performance of GEFormer under different experimental scenarios. (M1) Untested genotypes in the tested environments. GEFormer significantly outperformed classical six statistical learning methods and four machine-learning methods for predicting PH and DTA in most environments. For the complex trait of EW, GEFormer had significantly better prediction performance than the statistical learning and machine-learning methods in all the environments. In all, the results showed that GEFormer can capture complex nonlinear features and demonstrates great superiority in the prediction of complex traits. The trend of prediction accuracy of all the GP methods was correlated with phenotypic heritability (Yang et al., 2011), as shown in Supplemental Figure 51. The higher the phenotypic heritability, the better the prediction accuracy of the various methods. In addition, the accuracy of XGBoost was greater than that of the Stacking model in some cases. This is related to the prediction results of the base learners and the model used to combine the outputs of the base learners in the meta learner of the Stacking model. (M2) Tested genotypes in

an untested environment. The accuracy of GEFormer for PH, DTA, and EW prediction in multiple populations was slightly better than that of the six statistical learning methods and four machine-learning methods. (M3) Untested genotypes in an untested environment. This experimental scenario was extremely challenging for the GP algorithms. The experimental results showed that the prediction accuracy of all methods in this scenario was significantly lower than that in the above two scenarios. Interestingly, GEFormer exhibited notably superior prediction accuracy across all traits, populations, species, and environments compared with the other ten methods.

Three application scenarios

We examined the use of GEFormer for breeding applications in three scenarios and analyzed the biological significance. The first application was phenotype prediction in unknown environments. From the predicted phenotype values in different environments, breeders can quickly select target materials with specific phenotypes (e.g., high yield) in different planting environments and areas. GP could therefore help to accelerate the breeding process, improve efficiency, and save costs in animal and plant breeding. The second application was phenotype prediction for a hybrid population based on an inbred population. When using hybrid materials as independent testing populations, leveraging both the paternal and maternal materials as the training data resulted in significantly higher prediction accuracy than using only the maternal materials. In practical breeding applications, the parental materials (paternal and maternal materials) can be used as training populations to predict the field phenotypes of their hybrid offspring, thus saving planting costs in the field. The third application was phenotype prediction of cross populations. A larger number of training populations helped to improve the prediction accuracy for cross populations. This result demonstrates that the GEFormer model can integrate information from multiple populations with different genetic backgrounds in maize. In the future, we can integrate more and more populations from breeders to break through bottlenecks of populations with different genetic backgrounds.

The impact of environmental factors on GP

Incorporating environmental data into GP can help to accurately predict crop phenotypes in the field under complex and ever-changing climatic conditions. It can also help researchers analyze the complex relationships among genetic and non-genetic factors and phenotypes. In this study, we considered the environmental factors throughout the whole growth period, all phenological periods, and the temporal relationships among phenological periods of crops. As we know, genotype–environment interactions occur every day during the crop growth period. It is difficult to simulate the actual growth environment of crops using only the mean of each environmental factor or EC simulations (Jin et al., 2023) throughout the whole growth period. Some studies have divided the whole reproductive period into multiple phenological stages, performing GP by establishing the interaction relationship between EC and genotype in each phenological stage. However, the positive and negative correlations between EC and phenotype vary at different phenological stages. Therefore, it is necessary to comprehensively consider the interactions between genotype and environmental factors throughout the whole crop growth period rather than using only the environmental features of key phenological periods (Westhues et al., 2021). Current research on genotype–environment interactions can identify key phenological periods and environmental factors (Li et al., 2018, 2021; Guo et al., 2020; Fu and Wang, 2023). Some studies have integrated dynamic crop growth models (CGMs) into GP (Technow et al., 2015; Cooper et al., 2016). CGMs can simulate specific crop phenotypes using environmental data, but they need to assume the distribution of environmental features. However, the interactions between genotype and environmental factors are not only related to a few key environmental factors but also to the combined effects of all non-genetic environmental factors. Therefore, it is necessary to consider all environmental factors when constructing a G × E model, thus simulating the crop growth process more realistically and improving the prediction accuracy.

The construction of genotype–environment interactions is crucial for capturing the complex nonlinear relationships between genotype and environmental factors. In the field of quantitative genetics, some researchers have captured the interactions between genotype and environment by calculating the covariance. However, this method cannot capture new genotype–environment covariance if new environmental information is not added to the model (Li et al., 2018; Jarquin et al., 2020; Costa-Neto et al., 2021c). In the field of machine learning, XGBoost uses a tree structure to explore the potential interactions between genotype and environmental factors (Westhues et al., 2021). However, it requires additional feature extraction engineering to process features with larger dimensions. DL methods can flexibly process features with different dimensions without assuming data distributions in advance and can automatically capture complex nonlinear interactions between genotypic and environmental features.

Future prospects

GP models can accelerate the selection of breeding materials, thereby improving breeding efficiency. With the continuous accumulation of crop genotype, phenotype, and environmental data, the advantages of DL are expected to be further improved in GP (Washburn et al., 2020; Hayes et al., 2023). GP models based on DL can enable the selection of high-quality germplasm resources for different environments, thereby improving land utilization efficiency. Researchers can incorporate GEFormer into breeding toolkits for large-scale field breeding, thereby improving the genetic gain of traits (Wang et al., 2023). Data remain the foundation, and accurate prediction models are the key for future breeding research. With the continued accumulation of genotype data, differences in allele frequency and linkage imbalance may limit the accuracy of GP. For future phenotype predictions in specific regions and years, CGM (Kochkov et al., 2024) and GraphCast (Lam et al., 2023) models can be used to predict environmental factors. We can then use GEFormer to predict phenotypes from the predicted environmental factors and genotype data. In addition, we will build an online prediction web platform for GEFormer to better serve crop breeding (Li et al., 2024a, 2024b; Zhu et al., 2024). With the rapid development of multi-omics sequencing technologies, it is necessary to collect more omics data (such as transcriptome, metabolome, and ionome data) to obtain features that cannot be captured by genotype and environmental data (Hu et al., 2019; Yang et al., 2021; Guo and Li, 2023). In addition, the crop

growth process is influenced by multiple physiological and ecological environmental factors. Consequently, it would be advisable to collect more types of environmental data (Xu, 2016) to further mine genotype–environment interactions during crop growth. We will also integrate GEFormer and federated learning (Xiong et al., 2021) to jointly train models of multiple populations, thereby breaking through genetic bottlenecks among different populations and enhancing the model's generalization ability. GEFormer based on DL shows good prediction accuracy, but its high complexity can affect computational efficiency. Compared with other models, GEFormer requires longer training times (Supplemental Table 14), and training time has a greater impact on parameters such as epoch and batch size. Finally, we will explore the application of interpretable techniques within GEFormer to uncover new genes that interact with environmental factors and provide broader genetic resources to support intelligent-design breeding (Wang et al., 2024).

METHODS

Genotype and phenotype datasets

The maize CUBIC1404 inbred line population is composed of 1404 materials derived from complete alleles plus unbalanced breeding (Liu et al., 2020). The Illumina HiSeq 2500 platform was used to perform whole-genome sequencing, generating 5 Tb of sequenced base pairs for each material. We then used the BaseRecalibrator and PrintReads tools from GATK to recalibrate the base quality scores and correct sequencing errors and other potential experimental artifacts (Liu et al. 2020). Each material contained more than 14 million SNPs. We removed SNPs with a minor-allele frequency <0.05 and obtained a total of 7 736 104 high-quality SNPs for each material. We then selected 32 559 SNPs from the original SNP set according to Yan et al. (2021). We obtained 32 336 SNPs by mapping the 32 559 SNPs to the B73V4 reference genome. The CUBIC1404 inbred lines were planted in five cities in northern China in 2014: Changchun City, Jilin Province (N 43°88′, E 125°35′); Shenyang City, Liaoning Province (N 41°48′, E 123°38′); Shunyi District, Beijing City (N 40°13′, E 116°13′); Baoding City, Hebei Province (N 38°85′, E 115°48′); and Xinxiang City, Henan Province (N 35°31′, E 113°85′). The sowing dates for the five cities were May 9, 2014; May 11, 2014; May 13, 2014; June 11, 2014; and June 12, 2014, respectively. The planting method for the CUBIC1404 population was 5 m long, with about 17 plants in each row. CHANG7-2 was planted as a control for 50 rows of materials to check whether there were systematic errors and for standardization. PH, DTA, and EW data were collected from the CUBIC1404 population at the five locations for use in subsequent experiments. DTA was measured by taking into account all surviving individuals of the same genotype, with the time at which 50% of the individuals exhibited a specific trait being used as the criterion. For PH, the five consecutive individuals in the middle of each family were measured. For EW, five representative and relatively consistent ears were selected for measurement after harvest. The raw sequencing data for the CUBIC1404 population are available at NCBI under Bioproject accession number PRJNA597703 (https://www.ncbi.nlm.nih.gov/bioproject/597703).

The maize CUBIC1404*JING724 hybrid population originated from crossing 1404 inbred line materials with JING724 (Xiao et al., 2021). Genotype data for the CUBIC1404*JING724 hybrid population was processed as described above for the CUBIC1404 population, yielding a total of 32 336 high-quality SNPs based on the V4 version of the reference genome. The sowing dates, planting locations, planting method, and phenotype measurements for the CUBIC1404*JING724 population were the same as those for the CUBIC1404 population. PH, DTA, and EW data from the CUBIC1404*JING724 hybrid population measured in the five locations were used for subsequent experiments. The maize CUBIC1404*ZHENG58 hybrid population (Xiao et al., 2021) originated from crossing 1404 inbred line materials with ZHENG58. Genotype and phenotype data for the CUBIC1404*ZHENG58 population were obtained and processed as described for the CUBIC1404 population. The Maize6210 F1 hybrid population originated from crossing 207 maternal lines in the CUBIC1404 inbred lines with 30 elite male lines (Yang et al., 2022). Genotype data for this population were processed as described above. The planting locations, planting methods, and phenotype measurements for the Maize6210 F1 hybrid population were the same as those for the CUBIC1404 population. The sowing dates in the five cities were May 13, 2015 (Changchun City); May 12, 2015 (Shenyang City); May 14, 2015 (Beijing City); June 10, 2015 (Baoding City); and June 11, 2015 (Xinxiang City). PH and EW data from the 6210 F1 hybrid population in five locations were used in subsequent experiments.

The Maize976 hybrid population originated from crossing 488 female lines from the association mapping population with two male lines, MO17 and Zheng58 (Liu et al., 2021b). Genotype data for this population were processed as described above. The 976 hybrid population was planted in seven cities in central and southern China in 2011: Sanya City, Hainan Province (109.17997 E, 18.36078 N); Nanning City, Guangxi Province (108.27949 E, 22.78704 N); Honghe Autonomous Prefecture (102.42755 E, 23.37449 N); Ya'an City, Sichuan Province (102.95836 E, 29.99362 N); Chongqing City (106.51756 E, 29.50793 N); Wuhan City, Hubei Province (114.32855 E, 30.38156 N); and Hebi City, Henan Province (114.30559 E, 35.74695 N). The sowing dates for the seven cities were October 24, 2011; April 7, 2011; April 19, 2011; April 6, 2011; April 4, 2011; March 31, 2011; and June 14, 2011, respectively. The 976 hybrid population was also planted in four cities in southern China in 2012: Honghe Autonomous Prefecture (102.42755 E, 23.37449 N); Chongqing City (106.51756 E, 29.50793 N); Wuhan City, Hubei Province (114.32855 E, 30.38156 N); and Hebi City, Henan Province (114.30559 E, 35.74695 N). The sowing dates for the four cities were April 22, 2012; April 5, 2012; April 6, 2012; and June 13, 2012, respectively. Maize976 populations were planted in the field with a plant spacing of 0.25 m. For each hybrid, five healthy and uniform plants were selected for measurement of PH and yield traits. PH was measured from the tip of the tassel to the ground. Yield was measured as the weight of dried ear collected from the uppermost ear of the plant. After removal of materials with missing phenotype values, PH (598) and EW (480) data for the 976 hybrid population in seven locations in 2011 and four locations in 2012 were used in subsequent experiments.

The rice dataset was obtained from INIA-Uruguay National Institute of Agricultural Research (INIA-Uruguay) and consisted of 327 elite *indica* rice lines. Genotype data for each line were obtained

through genotyping by sequencing (GBS). SNP data were generated using the TASSEL 3.0 GBS pipeline and aligned to the Nipponbare reference genome version 7.0. We used TASSEL5.0 to impute missing SNP data and finally obtained 92 430 SNPs for use in subsequent experiments. The rice population was planted in Paso de la Laguna Experimental Station (UEPL), Treinta y Tres, Uruguay (33°15′ S, 54°25′ W) in 2010, 2011, and 2012. The population was planted in six-row plots using an augmented randomized complete block design with two or three replications. Two Uruguayan cultivars (El Paso 144, INIA Olimar) were included as control varieties (Monteverde et al., 2018; Xu et al., 2018). PHR and GY data from the rice population at three locations (years) were used for subsequent experiments.

The wheat dataset contained 635 recombinant inbred lines (RILs) from 26 nested association mapping families. The wheat population was genotyped using the Illumina 90 K SNP array (Wang et al., 2014) and GBS. We deleted SNP markers with more than 20% missing data, minor allele frequencies <0.10, and RIL missing rate >10% of genotype data, resulting in 44 768 SNPs for the experiment. The wheat population was planted in 2014, 2015, and 2016 at the Spillman Agronomy Farm near Pullman (WA, United States). A modified augmented field design was used in each trial, with three replicated check cultivars (Berkut, McNeal, and Thatcher) (Sandhu et al., 2021). PH (height), DTH, and yield data for the wheat population at three locations (years) were used in subsequent experiments.

Environmental dataset

The combined effects of the interactions among multiple environmental factors within each day, as well as the dynamic changes in environmental factors between days during the crop growth period, collectively influence the outcomes of GP. We downloaded the daily environmental factor data for all populations from planting to harvest. Specifically, we used the EnvRtype package (Costa-Neto et al., 2021a) to collect environmental factor data from NASAPower for five populations planted in different locations at different times. We collected data related to five types of environmental factors: temperature, water, light, air, and soil. There were seven temperature-related environmental factors: maximum temperature at 2 m above the surface of the earth (T_{max}), minimum temperature at 2 m above the surface of the earth (T_{min}), DTR ($DTR = T_{max} - T_{min}$), TSR ($TSR = T_{max}^2 - T_{min}^2$), MMR ($MMR = T_{min}/T_{max}$), GDD ($GDD = (T_{max} + T_{min})/2 - T_{base}$), and dGDD ($dGDD = |GDD(day\_n) - GDD(day\_n - 1)|$). T_{base} was 10°C (50°F). If the average temperature was equal to or lower than T_{base}, GDD was set to zero. If the average temperature was higher than T_{base}, GDD was defined as the average temperature minus T_{base}. There were three water-related environmental factors: relative air humidity at 2 m above the surface of the earth (RH), precipitation corrected (PR), and PRDTR ($PRDTR = PR/DTR$) (Li et al., 2022b). There were 10 light-related environmental factors: DayL (day length), APAR (all sky surface Photosynthetically Active Radiation (PAR) total), CPAR (clear sky surface PAR total), UVA (all sky surface UVA irradiance), UVB (all sky surface UVB irradiance), PTT ($PTT = GDD \times DayL$), dPTT ($dPTT = |PTT(day\_n) - PTT(day\_n - 1)|$), PTR ($PTR = GDD/DayL$), PTD1 ($PTD1 = DTR \times DayL$), and PTD2 ($PTD2 = DTR/DayL$). There were three air-related environmental factors: surface pressure (PS), wind speed at 2 m above the surface of the earth (WS2M), and wind direction at 2 m above the surface of the earth (WD). There were two soil-related environmental factors: root-zone soil wetness (SW) and profile soil moisture (SM). The DayL-related environmental factor data were downloaded using the geosphere package (Robert, 2023).

Methods

By extracting and integrating genetic features (G*) and environmental features (E*), GEFormer performs phenotype prediction from the perspective of gene–environment interactions in maize. The GEFormer model consists of five parts: the input layer, gMLP layer, TimeFeatureBlock layer, CrossGatedMLP layer, and output layer. The implementation process of the five layers is shown in Figure 1. The input layer preprocesses the original data and inputs the genotypic and the environmental data into the gMLP and TimeFeatureBlock layers, respectively. The gMLP layer (Liu et al., 2021a) extracts genotype features from both local and global perspectives, obtaining the genotypic features G*. By introducing gating mechanism and nonlinear activation functions, gMLP can capture the long-range dependencies among SNPs. To simulate the comprehensive effects of environmental factors on crop growth, we used the TimeFeatureBlock module to extract features (E*) of 25 environmental factors. During the actual crop growth process, environmental factors are interconnected within the same day, and environmental factors also exhibit sequential features between days. This module first uses ODConv (Li et al., 2022a) to extract the comprehensive effects of each environmental factor within a day and then uses the linear attention mechanism to obtain the features of environmental factors changing between days. ODConv can dynamically adjust the convolution kernel parameters during the training process and helps to capture complex features. The linear attention mechanism can determine the importance of different positions in the input sequence by introducing learnable weights, thereby helping to improve the ability to capture local key features in the sequence (Zhou et al., 2021). The CrossGatedMLP layer (Tu et al., 2022) integrates the features of G*, E*, and G* × E*. It adaptively determines the contribution of each feature to the fused features through gating signals, thereby improving the model's representation ability. The output layer uses MLP to further extract the fused features and maps them to the phenotype values. The computational process of capturing relationships among features and fusing features in the construction of the G × E model is based on the Transformer algorithm, and we called this model GEFormer.

Input layer

The genotype data input into the model are a sequence of SNPs with length l, denoted as $G = \{g_1, g_2 \ldots g_l\}$, $g_i \in \{0,1,2\}$. The environmental data are a two-dimensional matrix $A (A \in R^{d \times (e+1)})$ composed of d days and e environmental factors per day during crop growth. The SNP sequence G is input into the gMLP layer. The two-dimensional matrix A is divided into the date vector $D (D \in R^d)$ and the environmental factor vector $E (E \in R^{d \times e})$. Then, $D^{(1)}$ and $E^{(1)}$ are generated and input into the TimeFeatureBlock layer.

gMLP layer

The gMLP layer captures the features of genotype data through nonlinear transformation and spatial gating unit (SGU). The SNP sequence G is input into the gMLP layer. It then maps G to a

higher-dimensional space through linear transformation to enhance the feature representation capability and obtains $G^{(1)}$. The definition of the linear transformation is shown in Equation 1.

$$G^{(1)} = W_G G + bias_G \quad \text{(Equation 1)}$$

In Equation 1, W_G is the weight parameter of the linear transformation, and $bias_G$ is the bias parameter of the linear transformation.

$G^{(1)}$ is then input into the GELU activation function (Hendrycks and Gimpel, 2016) to obtain the output $G^{(2)}$. The GELU activation function refers to a smooth nonlinear transformation. It approximates a linear mapping for small input values, and it saturates for large input values. Therefore, it helps the model quickly propagate gradients, prevents overfitting, and makes the model more stable in the process of gradient calculation and backpropagation. The definition of the GELU activation function is shown in Equation 2.

$$G^{(2)} = GELU(g) = g * P\left(G^{(1)} \le g\right) = g \int_{-\infty}^{g} \frac{e^{-\frac{\left(G^{(1)} - \mu\right)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dG^{(1)} \quad \text{(Equation 2)}$$

In Equation 2, $g \in G^{(1)}$, μ represents the mean of the normal distribution, and σ represents the standard deviation of the normal distribution.

The output $G^{(2)}$ of the activation function is then input to the SGU. $G^{(2)}$ is split into two parts, *res* and *gate*, with the same shape; *res* is used to transmit the original information, and *gate* is used in the gating mechanism that regulates the flow information.

The feature dimensions of *gate* are partitioned into multiple heads to obtain $gate^{(1)}$. Equation 3 is then used to perform spatial projection on $gate^{(1)}$ to obtain $gate^{(2)}$. This operation helps the model to capture features from different perspectives. In addition, different weights are used to weigh the gating tensors of each head, enabling the model to learn features more flexibly.

$$gate^{(2)} = \sum_n gate^{(1)} * W + bias \quad \text{(Equation 3)}$$

In Equation 3, $gate^{(2)} \in R^{b \times h \times m \times d}$, $gate^{(1)} \in R^{b \times h \times n \times d}$, $W \in R^{h \times m \times n}$, and $bias \in R^{1\text{fn}1 \times h \times n \times 1}$; b is the batch size, h is the number of heads, d is the feature dimension, m is the output feature dimension, n is the length of the sample sequence, and W and *bias* are the model parameters.

The output $gate^{(2)}$ of the spatial projection is multiplied by the corresponding output of the linear transformation to obtain $G^{(3)}$. $G^{(3)}$ retains the long-range dependencies while preserving key features and attenuating less important ones. The definition of $G^{(3)}$ is shown in Equation 4.

$$G^{(3)} = res \odot gate^{(2)} \quad \text{(Equation 4)}$$

Finally, layer normalization, global average pooling, and linear transformation are performed on the output $G^{(3)}$ of SGU to extract the global information from the feature map and project it to the specified dimension, obtaining the genotypic feature G^*.

TimeFeatureBlock layer

The TimeFeatureBlock layer is used to extract features from environmental data; it contains an embedding layer, dynamic convolution (ODconv) layer, and Encoder layer. The embedding layer consists of three parts: value embedding, position embedding, and time embedding. The ODconv layer captures the comprehensive effects among various environmental factors within each day. The linear attention mechanism extracts the features of environmental factors between days. The input of TimeFeatureBlock is the date vector $D^{(1)}$ and the environmental factor vector $E^{(1)}$. $R^{d \times 3}$ is the encoded date vector of d days, $R^{d \times e}$ is the tensor of e normalized environmental factors in each day. $D^{(1)} \in R^{d \times 3}$, and $E^{(1)} \in R^{d \times e}$.

First, $D^{(1)}$ and $E^{(1)}$ are input into the embedding layer. Value embedding is used to convert the values in $E^{(1)}$ into a fixed-dimensional representation $E^{(2)}$, which helps the model to learn the relationships among the values in the input tensor. The method for calculation of value embedding is shown in Equation 5.

$$Output\left(E^{(1)}\right) = \sum_{n=0}^{c-1} W(c', n) * input(E) + bias(c') \quad \text{(Equation 5)}$$

In Equation 5, $E \in R^n$, $E^{(1)} \in R^{c'}$, n is the index of the input channel, c is the number of input channels, and c' is the number of output channels.

$E^{(1)}$ is then input into the position embedding module, which employs sine and cosine functions to encode the position information of the input sequence into a vector. This operation enables the model to learn the relative position relationships among elements in the sequence and produces a vector $E^{(3)}$ that contains internal positional relationships within the sequence. The weight calculation process of the position embedding is shown in Equation 6.

$$Pos\_Embedding(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d}}\right)$$
$$Pos\_Embedding(pos, 2i+1) = \cos\left(\frac{pos}{10000^{2i/d}}\right) \quad \text{(Equation 6)}$$

In Equation 6, *pos* is the position of the input sequence, i is the dimension index, and d is the embedding dimension. *Pos_Embedding*(*pos*, 2*i*) and *Pos_Embedding*(*pos*, 2*i*+1) are the even and odd dimensions of position *pos* in the position embedding matrix.

$D^{(1)}$ is then input into the time embedding module to obtain $D^{(2)}$. This model projects the time features to a higher dimension to better understand the time information. The definition of time embedding is shown in Equation 7.

$$Embedding_{temporal}\left(D^{(1)}\right) = \sum_{i=0}^{3} Embedding_i\left(D^{(1)}[:, :, i]\right) \quad \text{(Equation 7)}$$

In Equation 7, i represents the time dimension, and $Embedding_i$ denotes the embedding function of the i-th time dimension.

Next, the output $E^{(2)}$ of the value embedding, the output $E^{(3)}$ of the position embedding, and the output $D^{(2)}$ of the time embedding are combined to obtain $E^{(4)}$. By integrating the embedding of time sequences and environmental factors in the above manner, the model can form a richer semantic representation.

The output $E^{(4)}$ of the embedding layer is input into the ODconv layer. The multi-dimensional attention mechanism is used to obtain channel attention $A_{channel}$, spatial attention $A_{spatial}$, kernel attention A_{kernel}, and filter attention A_{filter} (Equations 8–11). The weights of $A_{channel}$, $A_{spatial}$, and A_{filter} are adjusted locally through the sigmoid activation function (Equation 12). A_{kernel} uses the softmax activation function to perform normalization, improving the calculation stability.

$$A_{channel} = \sigma\left(Conv1d_{channel}\left(E^{(4)}\right)\right) \quad \text{(Equation 8)}$$

$$A_{spatial} = \sigma\left(Conv1d_{spatial}\left(E^{(4)}\right)\right) \quad \text{(Equation 9)}$$

$$A_{kernel} = soft\ \max\left(Conv1d_{ker\ nel}\left(E^{(4)}\right)\right) \quad \text{(Equation 10)}$$

$$A_{filter} = \sigma\left(Conv1d_{filter}\left(E^{(4)}\right)\right) \quad \text{(Equation 11)}$$

In Equations 8–11, $\sigma(\cdot)$ represents the sigmoid activation function; $A_{channel} \in R^{b\times c\times 1}$, $A_{spatial} \in R^{b\times 1\times 1\times 1\times k}$, $A_{kernel} \in R^{b\times n\times 1\times 1\times 1}$, $A_{filter} \in R^{b\times c'\times 1}$, and $E^{(4)} \in R^{b\times c\times 1}$; b is the batch size, c is the number of input channels, c' is the number of output channels, k is the convolution kernel size, and n is the number of kernels.

$$Sigmoid(z) = \frac{1}{1+e^{-z}} \quad \text{(Equation 12)}$$

In Equation 12, e is the base of the natural logarithm. The sigmoid function maps the output of the linear layer to a gating weight with a value between 0 and 1 and is used to filter and transmit specific features.

$A_{channel}$ is then applied to each channel of $E^{(4)}$ to obtain $E^{(5)}$, and a convolution operation is performed on $E^{(5)}$. Meanwhile, $A_{spatial}$ and A_{kernel} are multiplied by the initial weight to obtain the weights of the convolution, resulting in the output $E^{(6)}$. Each channel of $E^{(4)}$ corresponds to different environmental factors (daylength, RH, surface pressure, etc.) and time features. The channel attention assigns different weights to these channels, enabling the model to focus more on environmental factors with greater importance. Using the spatial attention to adjust the importance of different time steps in the sequence could help the model to capture crucial time information. The kernel attention is used to adjust the model's reliance flexibly on each convolution kernel, thus enabling different input data to be processed more finely and enhancing the model's expression and generalization capabilities.

Finally, A_{filter} is applied to the feature vector $E^{(6)}$ of the convolution output to obtain $E^{(7)}$. By adjusting the weights of all filters, more critical environmental features can be extracted and feature representation capabilities can be improved.

The output $E^{(7)}$ of the ODconv layer is input into the linear attention layer to obtain $E^{(8)}$. This operation can capture the dependence relationships among environmental factors for each day. In the linear attention layer, a linear transformation operation on $E^{(7)}$ is first performed to obtain the query vector Q, the key vector K, and the value vector V. Then, the query number L_Q and key number L_K are used to calculate the random sampling number U_part and the queried sample number u (Equations 13 and 14). U_part keys are randomly sampled to obtain K_sample. Subsequently, the sparsity measurement value M is calculated based on the attention score of query Q and the sampled key K_sample (Equation 15). The sparsity measurement value M reflects the divergence between the weight distribution of each query and the uniform distribution, with larger values indicating more sparse weight distribution. The query vector Q_{reduce} of the top u queries is extracted based on M, and the attention scores between K and Q_{reduce} are calculated. The attention score represents the correlation between each time step and other time steps. Finally, the attention scores are applied to the value vector V to obtain the representative sequence feature $E^{(8)}$ (Equation 16).

$$U\_part = \max(c \times \lceil \log(L_K) \rceil, L_K) \quad \text{(Equation 13)}$$

$$u = \max(c \times \lceil \log(L_Q) \rceil, L_Q) \quad \text{(Equation 14)}$$

In Equations 13 and 14, the constant c is a scaling factor, L_K is the length of the key sequence, and L_Q is the length of the query sequence.

$$M(q_m, K\_sample) = \max_n\left\{\frac{q_m k_n^{\mathrm{T}}}{\sqrt{dim}}\right\} - \frac{1}{L_K}\sum_{n=1}^{L_K}\frac{q_m k_n^{\mathrm{T}}}{\sqrt{dim}} \quad \text{(Equation 15)}$$

In Equation 15, $q_m \in Q$, q_m is the m-th query vector, $k_n \in K\_sample$, k_n is the n-th key vector, dim is the dimension of the query vector and key vector, and L_K is the length of the key sequence.

$$E^{(8)} = Attention(Q, K, V) = Soft\ \max\left(\frac{Q_{reduce} \times K^{\mathrm{T}}}{\sqrt{dim}}\right) \times V \quad \text{(Equation 16)}$$

In Equation 16, Q_{reduce} is the selected top u queries based on the sparsity metric M.

The linear attention mechanism uses a sparsity metric to select important key–query pairs for attention computation. The linear complexity is approximately (O(d*lnd)), which can effectively save space and time compared with the traditional self-attention mechanism whose complexity is (O(d^2)). At the same time, this mechanism retains the information most relevant to the query, maintaining good model performance. Applying the linear attention mechanism to the multi-head attention mechanism can further strengthen the computational efficiency of the model.

The linear transformation is then applied to $E^{(8)}$ to integrate the comprehensive information from multiple heads and obtain $E^{(9)}$. The definition of $E^{(9)}$ is shown in Equation 17.

$$E^{(9)} = E^{(8)} * W_{proj}^{T} + bias_{proj} \quad \text{(Equation 17)}$$

In Equation 17, W_{proj} is the weight matrix of the linear transformation, and $bias_{proj}$ is the bias.

Feature extraction is then performed on $E^{(9)}$ through the operations of convolution and max pooling to obtain the environmental feature E^{*}. The convolution uses circular padding, which maintains the periodicity of the data when processing the sequence boundaries. The max pooling operation selects key features in the sequence. This operation benefits the modeling of long-term dependencies in the sequence while suppressing noise and improving the generalization ability.

The linear attention mechanism is capable of capturing dependencies between any two elements within a sequence, enabling learning of the overall sequence structure and improving attention to key information through weight-allocation operations. The convolutional layer is used to extract local features, and combining convolution with the linear attention mechanism helps to enhance understanding of the sequence information.

CrossGatedMLP layer

The CrossGatedMLP layer introduces a cross-gating mechanism that applies the gating weights to the hidden representations of the input features. The CrossGatedMLP layer fuses the genotype features G^{*} and the environmental features E^{*} by passing and integrating the two kinds of features. This layer consists of two MLPs. One MLP (hidden MLP) with the GELU activation function (Equation 2) is used to calculate the hidden representations of the inputs. The other MLP (gate MLP) with the sigmoid activation function (Equation 12) is used to calculate the gating weights.

First, the output G^{*} of the gMLP layer and the output E^{*} of the TimeFeatureBlock layer are input into the CrossGatedMLP layer. The hidden MLP calculates the hidden representations $G^{(4)}$ and $E^{(10)}$ of G^{*} and E^{*}. Meanwhile, the gate MLP is used to calculate the gating weights $G^{(5)}$ and $E^{(11)}$ of G^{*} and E^{*}.

The gating weights $G^{(5)}$ and $E^{(11)}$ are then applied to the hidden representations $G^{(4)}$ and $E^{(10)}$ (Equation 18) to obtain $C^{(1)}$ and $C^{(2)}$. Next, $C^{(1)}$ and $C^{(2)}$ are concatenated to obtain the fused feature $C^{(3)}$. The cross-gating mechanism regulates the interplay among input features by modulating the gating weights. This operation enables the model to learn the intrinsic characteristics of the inputs and to capture the interdependencies among features, thereby enhancing feature expression capability.

$$\begin{aligned} C^{(1)} &= \left(1 - G^{(5)}\right) \times G^{(4)} + E^{(11)} \times E^{(10)} \\ C^{(2)} &= \left(1 - E^{(11)}\right) \times E^{(10)} + G^{(5)} \times G^{(4)} \end{aligned} \quad \text{(Equation 18)}$$

Finally, $G^{*} \times E^{*}$ is obtained by multiplying G^{*} and E^{*}. Similarly, the fused features $C^{(4)}$ and $C^{(5)}$ are obtained by fusing G^{*} and $G^{*} \times E^{*}$, as well as E^{*} and $G^{*} \times E^{*}$, in the CrossGatedMLP layer. Then, the features of $C^{(3)}$, $C^{(4)}$, and $C^{(5)}$ are concatenated to obtain the interaction feature $C^{(6)}$ between genotype and environment.

Output layer

The output layer uses MLP to further extract the fused features and maps them to the phenotype values. MLP can integrate features from different levels, which helps to obtain more meaningful and robust feature representations.

First, the interaction features $C^{(6)}$ between genotype and environment are input into a linear layer, and $C^{(7)}$ is obtained through linear mapping (Equation 17). This operation reduces the dimension of feature vectors and extracts key features.

$C^{(7)}$ is then input into the LeakyReLU activation function (Xu et al. 2020). LeakyReLU introduces a non-zero slope, and it solves the problem of inability to update neurons due to the zero gradient inthe negative regions. The output $C^{(8)}$ of the activation function fits the input features non-linearly and has stronger feature expression capability. The definition of the LeakyReLU activation function is shown in Equation 19, and $c \in C^{(7)}$.

$$Leaky\mathrm{ReLU}(c) = \max(\alpha c, c) \quad \text{(Equation 19)}$$

The output $C^{(8)}$ of the activation function randomly sets the output of some neurons to 0 using Dropout to obtain $C^{(9)}$. This approach weakens the dependencies among neurons in the network, prevents overfitting, and improves the generalization ability.

Finally, the linear transformation is used to map $C^{(9)}$ to a single value, which is the predicted phenotype value $C^{(10)}$.

Comparison methods and experimental environment

This work compared GEFormer with commonly used GP methods of genotype–environment interaction, including RF, XGBoost, Stacking, MLP (DL), the main-effect EA plus reaction norm for GE based on GBLUP (EA + GW(GB)), the main-effect EADW plus reaction norm for GE based on GBLUP (EADW + GW(GB)), the main-effect EA plus reaction norm for GE based on GK (EA + GW(GK)), the main-effect EADW plus reaction norm for GE based on GK (EADW + GW(GK)), the main-effect EA plus reaction norm for GE based on DK (EA + GW(DK)), and the main-effect EADW plus reaction norm for GE based on DK (EADW + GW(DK)). The EA + GW(GB), EADW + GW(GB), EA + GW(GK), EADW + GW(GK), EA + GW(DK), and EADW + GW(DK) methods were implemented using the BGGE package (Granato et al. 2018). RF, XGBoost, Stacking, and DL were implemented using the learnMET package (Westhues et al. 2022). The DL method refers to the classical MLP algorithm. The implementation framework of GEFormer is python 3.9, pytorch 1.7.1, cuda 11.4, torchvision 0.8.2, and torchaudio 0.7.2. The hyperparameters of GEFormer include learning rate, batch size, dropout, and number of neurons. The hyperparameters were obtained using the Optuna package for automatic optimization (Akiba et al. 2019). The combination of optimal hyperparameters for each trait is shown in the supplemental information. The experimental operating environment was a Supermicro server, the CPU was an Intel Xeon Platinum 8255C CPU @ 2.50 GHz, the memory was 377.3 GB, and the GPU was an NVIDIA GeForce RTX3090*4.

Abbreviations are defined as follows: EA, the main additive effects; GE, genotype × environment interaction; EA + GW, main-effect EA plus reaction norm for GE; EADW, main-effect EAD with main envirotype information; GW, reaction-norm variation based on the genomic × envirotype effects; EADW + GW, main-effect EADW plus reaction norm for GE; GB, GBLUP; Stacking, stacking ensemble model.

Statistical analysis

Heritability is used to evaluate the contribution of genetic factors to phenotypic trait variation. The heritability of phenotypic traits was calculated as the proportion of genetic variance to total variance (the sum of genetic variance and environmental variance). The variance was estimated by a mixed linear model in the GCTA 1.94.1 package.

Student's *t*-test was used to analyze the significance of correlations between the phenotypes and the features extracted using GEFormer. We divided the phenotypes into two groups: high phenotype value and low phenotype value. The mean phenotype values of two groups were considered to differ significantly at $p < 0.05$. In addition, feature interactions (e.g., genetic interactions) were detected using two-way ANOVA for specific features. When $p < 0.05$, these interaction pairs were considered to be statistically significant.

DATA AND CODE AVAILABILITY

The GEFormer scripts are available in the release package on GitHub (https://github.com/Deep-Breeding/GEFormer/tree/main/GEFormerV1.0). The datasets used in this study are available from https://github.com/Deep-Breeding/GEFormer/tree/main/GEFormerV1.0/data. The detailed user manual is available from https://github.com/Deep-Breeding/GEFormer/blob/main/GEFormerV1.0/GEFormerV1_%20usermannual.pdf.

FUNDING

This work has been supported by the Biological Breeding-National Science and Technology Major Project (2023ZD04076), the Hubei Provincial Natural Science Foundation (2023AFB832), the Natural Science Foundation of Guizhou Province Science and Technology Agency (ZK[2025]096), the Major Project of Hubei Hongshan Laboratory (2022HSZD031), and the Yingzi Tech & Huazhong Agricultural University Intelligent Research Institute of Food Health (IRIFH202209).

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive comments, which helped us to substantially improve the manuscript. We thank the experimental teaching center of the College of Informatics at Huazhong Agricultural University for providing the experimental environment and computational resources. No conflict of interest is declared.

AUTHOR CONTRIBUTIONS

J.Y., J.L., and Y.X. designed the study. Z.Y., M.Y., C.W., K.L., and J.G. analyzed the data and performed the experiments. Z.Y., M.Y., and J.L. wrote the manuscript. All authors read and approved the final manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

Received: June 20, 2024
Revised: December 8, 2024
Accepted: January 25, 2025
Published: January 28, 2025

REFERENCES

Abberton, M., Batley, J., Bentley, A., Bryant, J., Cai, H., Cockram, J., de Oliveira, A.C., Cseke, L.J., Dempewolf, H., De Pace, C., et al. (2016). Global agricultural intensification during climate change: a role for genomics. Plant Biotechnol. J. **14**:1095–1098.

Achiam, J., Adler, S., and Agarwal, S. (2023). Gpt-4 technical report. Preprint at arXiv. https://doi.org/10.48550/arXiv.2303.08774.

Akiba, T., Sano, S., and Yanase, T. (2019). Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2623–2631.

Akpalu, W., Rashid, H.M., and Ringler, C. (2011). Climate variability and maize yield in the Limpopo region of South Africa: Results from GME and MELE methods. Clim. Dev. **3**:114–122.

Araújo, M.S., Chaves, S.F.S., and Dias, L.A.S. (2024). GIS-FA: an approach to integrating thematic maps, factor-analytic, and envirotyping for cultivar targeting. Theor. Appl. Genet. **137**:1–23.

Burgueño, J., Campos, G., and Weigel, K. (2012). Genomic prediction of breeding values when modeling genotype× environment interaction using pedigree and dense molecular markers. Crop Sci. **52**:707–719.

Chatzimparmpas, A., Martins, R.M., and Kerren, A. (2020). t-visne: Interactive assessment and interpretation of t-sne projections. IEEE Trans. Vis. Comput. Graph. **26**:2696–2714.

Cooper, M., and DeLacy, I.H. (1994). Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. Theor. Appl. Genet. **88**:561–572.

Cooper, M., Technow, F., Messina, C., Gho, C., and Totir, L.R. (2016). Use of crop growth models with whole-genome prediction: application to a maize multienvironment trial. Crop Sci. **56**:2141–2156.

Costa-Neto, G., Galli, G., Carvalho, H.F., Crossa, J., and Fritsche-Neto, R. (2021a). EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. G3 (Bethesda). **11**:jkab040.

Costa-Neto, G., Fritsche-Neto, R., and Crossa, J. (2021b). Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. Heredity **126**:92–106.

Costa-Neto, G., Crossa, J., and Fritsche-Neto, R. (2021c). Enviromic assembly increases accuracy and reduces costs of the genomic prediction for yield plasticity in maize. Front. Plant Sci. **12**:717552.

Crossa, J., Yang, R.C., and Cornelius, P.L. (2004). Studying crossover genotype × environment interaction using linear-bilinear models and mixed models. J. Agric. Biol. Environ. Stat. **9**:362–380.

Fu, R., and Wang, X. (2023). Modeling the influence of phenotypic plasticity on maize hybrid performance. Plant Commun. **4**:100548.

Gao, P., Zhao, H., Luo, Z., Lin, Y., Feng, W., Li, Y., Kong, F., Li, X., Fang, C., and Wang, X. (2023). SoyDNGP: a web-accessible deep learning framework for genomic prediction in soybean breeding. Briefings Bioinf. **24**:bbad349.

Gillberg, J., Marttinen, P., Mamitsuka, H., and Kaski, S. (2019). Modelling G×E with historical weather information improves genomic prediction in new environments. Bioinformatics **35**:4045–4052.

Granato, I., Cuevas, J., Luna-Vázquez, F., Crossa, J., Montesinos-López, O., Burgueño, J., and Fritsche-Neto, R. (2018). BGGE: a new package for genomic-enabled prediction incorporating genotype× environment interaction models. G3 **8**:3039–3047.

Guo, T., Mu, Q., Wang, J., Vanous, A.E., Onogi, A., Iwata, H., Li, X., and Yu, J. (2020). Dynamic effects of interacting genes underlying rice flowering-time phenotypic plasticity and global adaptation. Genome Res. **30**:673–683.

Guo, T., and Li, X. (2023). Machine learning for predicting phenotype from genotype and environment. Curr. Opin. Biotechnol. **79**:102853.

Hayes, B.J., Visscher, P.M., and Goddard, M.E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. **91**:47–60.

Hayes, B.J., Chen, C., Powell, O., Dinglasan, E., Villiers, K., Kemper, K.E., and Hickey, L.T. (2023). Advancing artificial intelligence to help feed the world. Nat. Biotechnol. **41**:1188–1189.

Hendrycks, D., and Gimpel, K. (2016). Gaussian error linear units (gelus). Preprint at arXiv. https://doi.org/10.48550/arXiv.1606.08415.

Hickey, L.T., N Hafeez, A., Robinson, H., Jackson, S.A., Leal-Bertioli, S.C.M., Tester, M., Gao, C., Godwin, I.D., Hayes, B.J., and Wulff, B.B.H. (2019). Breeding crops to feed 10 billion. Nat. Biotechnol. **37**:744–754.

Hu, X., Xie, W., Wu, C., and Xu, S. (2019). A directed learning strategy integrating multiple omic data improves genomic prediction. Plant Biotechnol. J. **17**:2011–2020.

Jarquín, D., Lemes da Silva, C., and Gaynor, R.C. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype×environment interactions in Kansas wheat. Plant Genome **10**:1–15.

Jarquin, D., Kajiya-Kanegae, H., Taishen, C., Yabe, S., Persa, R., Yu, J., Nakagawa, H., Yamasaki, M., and Iwata, H. (2020). Coupling day length data and genomic prediction tools for predicting time-related traits under complex scenarios. Sci. Rep. **10**:13382.

Jin, M., Liu, H., Liu, X., Guo, T., Guo, J., Yin, Y., Ji, Y., Li, Z., Zhang, J., Wang, X., et al. (2023). Complex genetic architecture underlying the plasticity of maize agronomic traits. Plant Commun. **4**:100473.

Kerbler, S.M., and Wigge, P.A. (2023). Temperature sensing in plants. Annu. Rev. Plant Biol. **74**:341–366.

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., et al. (2024). Neural general circulation models for weather and climate. Nature **632**:1060–1066.

Lesk, C., Anderson, W., Rigden, A., Coast, O., Jägermeyr, J., McDermid, S., Davis, K.F., and Konar, M. (2022). Compound heat and moisture extreme impacts on global crop yields under climate change. Nat. Rev. Earth Environ. **3**:872–889.

Li, C., Zhou, A., and Yao, A. (2022a). Omni-dimensional Dynamic Convolution. Preprint at arXiv. https://doi.org/10.48550/arXiv.2209.07947.

Li, H., Li, X., Zhang, P., Feng, Y., Mi, J., Gao, S., Sheng, L., Ali, M., Yang, Z., Li, L., et al. (2024a). Smart Breeding Platform: a web-based tool for high-throughput population genetics, phenomics, and genomic selection. Mol. Plant **17**:677–681.

Liu, H.J., Wang, X., Xiao, Y., Luo, J., Qiao, F., Yang, W., Zhang, R., Meng, Y., Sun, J., Yan, S., et al. (2020). CUBIC: an atlas of genetic architecture promises directed maize improvement. Genome Biol. **21**:1–17.

Liu, H., Dai, Z., and So, D. (2021a). Pay attention to mlps. Adv. Neural Inf. Process. Syst. **34**:9204–9215.

Li, J., Zhang, D., Yang, F., Zhang, Q., Pan, S., Zhao, X., Zhang, Q., Han, Y., Yang, J., Wang, K., et al. (2024b). TrG2P: A transfer learning-based tool integrating multi-trait data for accurate prediction of crop yield. Plant Commun. **5**:100975.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al. (2023). Learning skillful medium-range global weather forecasting. Science **382**:1416–1421.

Li, X., Guo, T., Mu, Q., Li, X., and Yu, J. (2018). Genomic and environmental determinants and their interplay underlying phenotypic plasticity. Proc. Natl. Acad. Sci. USA **115**:6679–6684.

Li, X., Guo, T., Wang, J., Bekele, W.A., Sukumaran, S., Vanous, A.E., McNellie, J.P., Tibbs-Cortes, L.E., Lopes, M.S., Lamkey, K.R., et al. (2021). An integrated framework reinstating the environmental dimension for GWAS and genomic selection in crops. Mol. Plant **14**:874–887.

Li, X., Guo, T., Bai, G., Zhang, Z., See, D., Marshall, J., Garland-Campbell, K.A., and Yu, J. (2022b). Genetics-inspired data-driven approaches explain and predict crop performance fluctuations attributed to changing climatic conditions. Mol. Plant **15**:203–206.

Liu, N., Du, Y., Warburton, M.L., Xiao, Y., and Yan, J. (2021b). Phenotypic plasticity contributes to maize adaptation and heterosis. Mol. Biol. Evol. **38**:1262–1275.

Lobell, D.B., Schlenker, W., and Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. Science **333**:616–620.

Ma, W., Qiu, Z., Song, J., Li, J., Cheng, Q., Zhai, J., and Ma, C. (2018). A deep convolutional neural network approach for predicting phenotypes from genotypes. Planta **248**:1307–1318.

Ma, X., Wang, H., Wu, S., Han, B., Cui, D., Liu, J., Zhang, Q., Xia, X., Song, P., Tang, C., et al. (2024). DeepCCR: large-scale genomics-based deep learning method for improving rice breeding. Plant Biotechnol. J. **22**:2691–2693.

Malik, W.A., and Piepho, H.P. (2018). Biplots: Do not stretch them. Crop Sci. **58**:1061–1069.

Martini, J.W.R., Gao, N., Cardoso, D.F., Wimmer, V., Erbe, M., Cantet, R.J.C., and Simianer, H. (2017). Genomic prediction with epistasis models: on the marker-coding-dependent performance of the extended GBLUP and properties of the categorical epistasis model (CE). BMC Bioinf. **18**:1–16.

Monteverde, E., Rosas, J.E., Blanco, P., Pérez de Vida, F., Bonnecarrère, V., Quero, G., Gutierrez, L., and McCouch, S. (2018). Multienvironment models increase prediction accuracy of complex traits in advanced breeding lines of rice. Crop Sci. **58**:1519–1530.

Platzer, A. (2013). Visualization of SNPs with t-SNE. PLoS One **8**:e56883.

Raimondi, D., Corso, M., Fariselli, P., and Moreau, Y. (2022). From genotype to phenotype in Arabidopsis thaliana: in-silico genome interpretation predicts 288 phenotypes from sequencing data. Nucleic Acids Res. **50**:e16.

Ren, Y., Wu, C., Zhou, H., Hu, X., and Miao, Z. (2024). Dual-extraction modeling: A multi-modal deep-learning architecture for phenotypic prediction and functional gene mining of complex traits. Plant Commun. **5**:101002.

Resende, R.T., Piepho, H.P., Rosa, G.J.M., Silva-Junior, O.B., E Silva, F.F., de Resende, M.D.V., and Grattapaglia, D. (2021). Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. Theor. Appl. Genet. **134**:95–112.

Robert, J. H. (2023). geosphere: Spherical Trigonometry. Rpackage version 1.5-19.

Sandhu, K.S., Lozada, D.N., Zhang, Z., Pumphrey, M.O., and Carter, A.H. (2020). Deep learning for predicting complex traits in spring wheat breeding program. Front. Plant Sci. **11**:613325.

Schrag, T.A., Möhring, J., Maurer, H.P., Dhillon, B.S., Melchinger, A.E., Piepho, H.P., Sørensen, A.P., and Frisch, M. (2009). Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. Theor. Appl. Genet. **118**:741–751.

Simonyan, K. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. Preprint at arxiv. https://doi.org/10.48550/arXiv.1312.6034.

Technow, F., Messina, C.D., Totir, L.R., and Cooper, M. (2015). Integrating crop growth models with whole genome prediction through approximate Bayesian computation. PLoS One **10**:e0130855.

Tu, Z., Talebi, H., and Zhang, H. (2022). Maxim: Multi-axis mlp for image processing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5769–5780.

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. **9**.

Vitezica, Z.G., Varona, L., and Legarra, A. (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics **195**:1223–1230.

Vitezica, Z.G., Legarra, A., Toro, M.A., and Varona, L. (2017). Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. Genetics **206**:1297–1307.

Wang, K., Abid, M.A., Rasheed, A., Crossa, J., Hearne, S., and Li, H. (2023). DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants. Mol. Plant **16**:279–293.

Wang, P., Li, Z., and Li, H. (2024). SMART CROPs. New Crops **1**:100007.

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., Maccaferri, M., Salvi, S., Milner, S.G., Cattivelli, L., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol. J. **12**:787–796.

Washburn, J.D., Burch, M.B., and Franco, J.A.V. (2020). Predictive breeding for maize: Making use of molecular phenotypes, machine learning, and physiological crop models. Crop Sci. **60**:622–638.

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.

Westhues, C.C., Mahone, G.S., da Silva, S., Thorwarth, P., Schmidt, M., Richter, J.C., Simianer, H., and Beissinger, T.M. (2021). Prediction of maize phenotypic traits with genomic and environmental predictors using gradient boosting frameworks. Front. Plant Sci. **12**:699589.

Westhues, C.C., Simianer, H., and Beissinger, T.M. (2022). learnMET: an R package to apply machine learning methods for genomic prediction using multi-environment trial data. G3 **12**:jkac226.

Wu, C., Zhang, Y., Ying, Z., Li, L., Wang, J., Yu, H., Zhang, M., Feng, X., Wei, X., and Xu, X. (2023). A transformer-based genomic prediction method fused with knowledge-guided module. Briefings Bioinf. **25**:bbad438.

Xiao, Y., Jiang, S., Cheng, Q., Wang, X., Yan, J., Zhang, R., Qiao, F., Ma, C., Luo, J., Li, W., et al. (2021). The genetic mechanism of heterosis utilization in maize improvement. Genome Biol. **22**:148.

Xiong, W., Reynolds, M., and Xu, Y. (2022). Climate change challenges plant breeding. Curr. Opin. Plant Biol. **70**:102308.

Xiong, Z., Cheng, Z., Lin, X., Xu, C., Liu, X., Wang, D., Luo, X., Zhang, Y., Jiang, H., Qiao, N., et al. (2022). Facing small and biased data dilemma in drug discovery with enhanced federated learning approaches. Sci. China Life Sci. **65**:529–539.

Xu, J., Li, Z., and Du, B. (2020). Reluplex made more practical: Leaky ReLU. 2020 IEEE Symposium on Computers and Communications (ISCC) (IEEE), pp. 1–7.

Xu, Y. (2016). Envirotyping for deciphering environmental impacts on crop plants. Theor. Appl. Genet. **129**:653–673.

Xu, X., Sharma, R., Tondelli, A., Russell, J., Comadran, J., Schnaithmann, F., Pillen, K., Kilian, B., Cattivelli, L., Thomas, W.T.B., et al. (2018). Genome-Wide Association Analysis of Grain Yield-Associated Traits in a Pan-European Barley Cultivar Collection. Plant Genome **11**:170073.

Xu, Y., Zhang, X., Li, H., Zheng, H., Zhang, J., Olsen, M.S., Varshney, R.K., Prasanna, B.M., and Qian, Q. (2022). Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction. Mol. Plant **15**:1664–1695.

Yan, J., Xu, Y., Cheng, Q., Jiang, S., Wang, Q., Xiao, Y., Ma, C., Yan, J., and Wang, X. (2021). LightGBM: accelerated genomically designed crop breeding through ensemble learning. Genome Biol. **22**:1–24.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. **88**:76–82.

Yang, W., Guo, T., Luo, J., Zhang, R., Zhao, J., Warburton, M.L., Xiao, Y., and Yan, J. (2022). Target-oriented prioritization: targeted selection strategy by integrating organismal and molecular traits through predictive analytics in breeding. Genome Biol. **23**:80.

Yang, Y., Saand, M.A., Huang, L., Abdelaal, W.B., Zhang, J., Wu, Y., Li, J., Sirohi, M.H., and Wang, F. (2021). Applications of multi-omics technologies for crop improvement. Front. Plant Sci. **12**:563953.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. Proc. AAAI Conf. Artif. Intell. **35**:11106–11115.

Zhu, W., Han, R., Shang, X., Zhou, T., Liang, C., Qin, X., Chen, H., Feng, Z., Zhang, H., Fan, X., et al. (2024). The CropGPT project: Call for a global, coordinated effort in precision design breeding driven by AI using biological big data. Mol. Plant **17**:215–218.