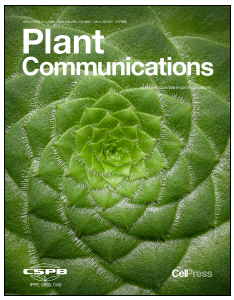
Journal Pre-proof



TOPlus: An algorithm integrating prior genetic and environmental information for multi-trait synergic selection in maize hybrid breeding

Wenjie Fang, Bingjie Wu, Xiaoyuan Hao, Zhenwei Zhang, Yong Peng, Nannan Liu, Wenqiang Li, Tingting Guo, Jianbing Yan, Yingjie Xiao, Wenyu Yang

PII: S2590-3462(26)00397-4

DOI: https://doi.org/10.1016/j.xplc.2026.102089

Reference: XPLC 102089

To appear in: *Plant Communications*

Received Date: 29 April 2026

Revised Date: 25 July 2026

Accepted Date: 31 August 2026

Please cite this article as: Fang, W., Wu, B., Hao, X., Zhang, Z., Peng, Y., Liu, N., Li, W., Guo, T., Yan, J., Xiao, Y., Yang, W., TOPlus: An algorithm integrating prior genetic and environmental information for multi-trait synergic selection in maize hybrid breeding, *Plant Communications* (2026), doi: https://doi.org/10.1016/j.xplc.2026.102089.

TOPlus: An algorithm integrating prior genetic and environmental information for multi-trait synergic selection in maize hybrid breeding

Wenjie Fang[1,3], Bingjie Wu[1,3], Xiaoyuan Hao[1,3], Zhenwei Zhang[1,3], Yong Peng[1], Nannan Liu[4], Wenqiang Li[1,3], Tingting Guo[1,3], Jianbing Yan[1,3], Yingjie Xiao[1,3*] and Wenyu Yang[1,2,3*]

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

[2] College of Science, Huazhong Agricultural University, Wuhan 430070

[3] Hubei Hongshan Laboratory, Wuhan 430070, China

[4] College of Life Sciences, Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University, Fuzhou 350002, China

[*]Correspondence: Wenyu Yang (yangwenyu@mail.hzau.edu.cn) and Yingjie Xiao (yxiao25@mail.hzau.edu.cn).

Short summary

TOPlus combines phenotype prediction based on environmental information with multi-trait synergic selection incorporating functional gene priors for maize hybrid breeding. Across diverse validation scenarios, TOPlus improves cross-environment prediction and hybrid prioritization and supports region-level suitability assessment, providing a useful framework for genomic breeding in maize.

Abstract

In maize hybrid breeding, synergic multi-trait selection of elite hybrids in specific target environments remains a major challenge. Enviromic data and functional gene knowledge are rapidly increasing; however, they have not been effectively integrated into crop breeding decisions. Here, we present TOPlus, a multi-trait hybrid prioritization framework comprising predictive and selective modules. The predictive module uses environmental information to improve phenotype prediction for untested genotypes and environments, whereas the selective

module incorporates functional gene priors and multiple predicted traits to prioritize hybrids for target environments. Across 17 agronomic traits, TOPlus improved average cross-environment prediction accuracy by 12% over JGRA and 3% over EADW+GW. Incorporating functional gene priors further improved hybrid prioritization: hybrids selected by TOPlus showed 5.90–19.64% higher yield than those selected by the original TOP method while maintaining comparable performance for other traits. The TOPlus algorithm internally clustered the functional genes into environmentally stable and plastic gene groups, with stable genes associated with core plant developmental processes and plastic genes enriched in stress-response and environmental-adaptation pathways. Independent validation in commercial hybrid panels demonstrated that TOPlus supports region-level suitability assessment and extrapolative deployment across diverse agroecological zones for specific candidate maize hybrid varieties. Overall, by integrating enviromic data and functional gene priors within an interpretable framework, TOPlus provides a biologically grounded and data-driven approach for cross-environment prediction and multi-trait synergic selection of hybrids for specific regions in maize breeding.

Keywords: Maize breeding, Environmental adaptation, Functional gene priors, Multiple traits, Genomic selection

Introduction

Current crop yield improvement is insufficient to meet global food demands by 2050 (FAO, 2009). Challenges, including land degradation, climate change, and depletion of natural resources, further threaten food security worldwide (Voss-Fels et al., 2019). Enhancing crop productivity per unit area, yield stability, and environmental adaptability therefore remain the central goals of modern plant breeding (Bailey-Serres et al., 2019; Farooq et al., 2024). Traditional breeding strategies rely on phenotype-based selection across multiple generations and environments, which is time-consuming and resource-intensive (Sharma et al., 2002). With advances in molecular markers and statistical genetics, genomic selection (GS) has become an important breeding tool by predicting genomic estimated breeding values (GEBV), allowing the lines with the highest GEBV to be prioritized for the next breeding cycle (Bernardo and Yu, 2007; Jonas and de Koning, 2013; Meuwissen et al., 2001). This predictive framework enables early-stage selection before field trials, thereby shortening the breeding cycle, reducing costs,

and improving resource-use efficiency (Crossa et al., 2017; Wang et al., 2023; Xu et al., 2018). However, many existing GS frameworks have been developed and evaluated primarily in diverse natural populations for genetic studies, rather than in the populations derived directly from elite breeding materials, resulting in a potential application gap between methodological innovations and breeding practices (Bassi et al., 2016; Hickey et al., 2019).

In maize breeding, recurrent selection based on population improvement is an efficient approach for gradually elevating the favorable allelic spectrum across long-term cycles (Bernardo, 2014). This breeding-oriented population provides a basis for enhancing genetic gain while maintaining broad genetic diversity for future breeding. A representative example is the Illinois long-term maize selection experiment for kernel protein and oil content, initiated in 1896, in which continuous directional selection over more than a century led to persistent and cumulative phenotypic divergence between selection lines, demonstrating the capacity of recurrent selection to pyramid favorable alleles underlying complex quantitative traits (Dudley and Lambert, 2003; Laurie et al., 2004). Recently, the Germplasm Advancement and Innovation Network (GAIN) initiative has been proposed for long-term maize population improvement in China, aiming to accelerate genetic gain through genomics- and phenomics-based recurrent selection (Wang et al., 2025). Previously, we developed a multiparent synthetic improvement population, named the Complete-diallel plus Unbalanced Breeding-derived Inter-Cross (CUBIC) population, by intercrossing 24 elite inbred lines widely used in the Huang-Huai-Hai maize production region of China from four endogenous breeding pedigrees (LvDaHongGu, ZI330, SiPingTou and Yugoslavia-improved), finally producing 1404 multiparent recombinant inbred lines. This synthetic population accumulated abundant genetic and phenotypic variation, allowing the systematic genetic study of agriculturally important traits (Liu et al., 2020a), further providing an unprecedented opportunity for linking predictive breeding models to realistic breeding.

To facilitate hybrid breeding, the CUBIC population was crossed with 30 additional elite inbred lines derived from foreign heterotic groups following a North Carolina II (NCII) mating design (Xiao et al., 2021). A 20% subset of this large hybrid population was generated and phenotyped in multi-environment trials. Using these data to train genomic prediction models, the remaining 80% untested hybrids were reliably predicted for multiple agronomic traits, enabling large-scale hybrid evaluation without exhaustive field phenotyping. This hybrid dataset not only

enabled a comprehensive exploration of the genetic basis of maize heterosis utilization, but also provided a robust training resource from a breeding perspective. Based on this dataset, the machine-learning method LightGBM was developed for hybrid predictive breeding by modeling nonlinear allelic interactions (Yan et al., 2021).

Plant breeders generally aim to improve multiple traits simultaneously, including yield, quality, and resistance to biotic and abiotic stresses, in elite lines or hybrid combinations. However, these important traits often exhibit antagonistic relationships; for example, high-yielding varieties may be more vulnerable to pathogen attacks (Derbyshire et al., 2024), making simultaneous multi-trait selection a major breeding challenge. Thus, the target-oriented prioritization (TOP) approach based on the maximum entropy algorithm has been proposed to perform multi-trait selection rather than prediction *per se*, allowing breeders to select the best candidate line or hybrid with globally similar performance to a benchmark variety but higher yield for entering the next breeding cycle (Yang et al., 2022). However, the current TOP algorithm primarily focuses on selection based on GEBV across multiple traits, and it remains challenging to extend this framework to selection under specific environmental conditions amid future extreme climates (Des Marais et al., 2013; Napier et al., 2022; Scheiner, 1993).

Genotype-by-environment interactions (G×E) have therefore been increasingly incorporated into statistical and machine learning models for predictive breeding (Millet et al., 2019; Sabir et al., 2023; Tester and Langridge, 2010). Early statistical approaches generally utilized the environmental gradient regression for modeling G×E effect. Joint regression analysis modeled the genotype performance across environments as a linear function of environmental means to quantify plasticity and adaptation (Yates and Cochran, 1938). The Finlay-Wilkinson framework further generalized this approach, quantifying genotype response to environment using regression coefficients (Finlay and Wilkinson, 1963). Furthermore, multivariate statistical models have been proposed to address more complex G×E structures beyond linear interactions. AMMI (additive main effects and multiplicative interaction) decomposed main and interaction effects via matrix factorization, improving the resolution of interaction analysis (Zobel et al., 1988). With the increasing availability of genomic and enviromic data, prediction models have continued to advance. Two major classes have emerged: joint genomic regression analysis (JGRA) (Li et al., 2021) and the main-effect-based EADW model (Costa-Neto et al., 2021). JGRA explicitly incorporated environmental covariates into genomic prediction models,

improving predictive performance particularly for untested or novel environments by modeling genotype-specific environmental sensitivity, although its predictive gain depends on the relevance and interpretability of selected environmental variables. In contrast, EADW+GW utilized model kernels integrating envirotype-based environmental similarity and genotype-by-environment interactions, achieving consistently higher prediction accuracy in multi-environment trials.

Crop functional genomics has made substantial progress so far, greatly enhancing the understanding of genetic architecture for agriculturally important traits and further facilitating gene-based editing and molecular design through gene pyramiding for crop improvement. In rice, high-quality reference genomes and pangenomes have facilitated large-scale dissection of the genetic mechanisms underlying quantitative traits and functional gene networks supporting molecular breeding design (Qin et al., 2021; Wang and Han, 2022). In cotton, rapid expansion of genomic resources and fine-scale functional element analyses have demonstrated feasible routes for identifying superior alleles and improving elite germplasm (Qi et al., 2025; Yang et al., 2023). Recently, the maize community has reported unprecedented progress in genomic and genetic studies and functional gene discovery for yield, quality, plant architecture, biotic and abiotic stresses, providing a comprehensive atlas for maize improvement via key genes (Li et al., 2025). These advances further highlight the potential to reform breeding algorithms by integrating functional gene information as key model features.

Here, we present TOPlus, an algorithm that combines environment-driven prediction with functional-gene-informed selection to enable multi-trait breeding selection with environment-specificity, improved predictive accuracy, and enhanced selection performance (https://github.com/Lee-Shell/TOPlus). TOPlus consists of two sequential analytic modules: the predictive module and the selective module (Figure 1). The predictive module implements an environment-driven iterative variable selection strategy for predicting the phenotypes of untested genotypes in novel environments (Supplemental Figure 1A, B); then the selective module integrates prior knowledge of maize functional genes and predicted traits to learn selection weight parameters for multiple traits and incorporated genes, and prioritizes candidate hybrids that show high similarity to benchmark cultivars while exhibiting improved yield performance in target environments from a breeding perspective (Supplemental Figure 1C, D). Through validation on three independent datasets (Supplemental Figure 2), TOPlus provides a reliable and

explainable solution that enables accurate cross-environment prediction and hybrid prioritization across variable climatic environments in maize and other crop species.

Results

TOPlus improves the prediction ability of maize agronomic traits across environments

In the predictive module, to evaluate the TOPlus performance of trait prediction in untested genotypes and untested environments, we used 569 representative hybrids as the training population and the remaining 5,251 hybrids as the testing population, with 17 agronomic traits evaluated across five environments in 2015, following the analytical scheme for genomic prediction previously reported for this maize NCII population (Yang et al., 2022) (Figure 2A).

The data from the training population in four environments were used to build the TOPlus model, aiming to predict the trait performance of the testing population in the fifth environment (Supplemental Figure 3). The process was iterated five times, with each environment serving as the prediction target in turn. In total, 25 environmental factors including day length (DL), all sky surface PAR total (APAR) and wind speed at 2 meters (WS) were considered to be enviromic variables contributing to trait prediction (Supplemental Table 1). The collected environments exhibited broad environmental variation and diversity, but the environments in the NCII training set showed a distribution distinct from that of the environments in the association mapping population (AMP) validation set of 507 maize inbred lines (Supplemental Figure 4A). Similar patterns were observed in the pairwise environmental-distance analysis and representative seasonal environmental profiles (Supplemental Figure 4B, C). For each environmental factor, a sliding-window approach was applied across the growing season, generating 10,731 time-resolved environmental variables.

To extract biologically relevant environmental variables from high-dimensional enviromic data, we first applied the Critical Environmental Regressor through Informed Search (CERIS) algorithm as an initial variable screening step, thereby reducing dimensionality and retaining candidate environmental factors potentially associated with target traits. Based on this filtered set of relevant environmental variables, the predictive module of TOPlus employs an environment-driven iterative factor selection strategy, in which environmental variables are dynamically optimized across training environments to improve prediction in novel environments. Rather than relying on fixed covariates, this framework evaluates environment-specific variable-trait

relationships and refines factor selection in a cross-environment manner. Within this framework, for relatively simple traits such as flowering time, environmental variables were selected based on their direct association with the target trait. In contrast, for complex traits such as yield (e.g., ear weight), environmental variables associated not only with the target trait itself but also with key developmental traits, such as flowering time and plant architecture (e.g., plant height), were incorporated, reflecting that yield is a composite trait influenced by multiple environmentally responsive developmental processes. In the predictive module, the frequencies of environmental variables selected across multiple training environments indicated their relative contributions to cross-environment trait prediction (Supplemental Figure 5). We found that agronomic traits exhibited distinct contribution patterns based on model fitting environmental factors (Figure 2B). Notably, primary and derived environmental factors contributed relatively evenly to the prediction of flowering traits, whereas their contributions differed more markedly for plant architecture and yield-related traits. For example, primary environmental factors such as ultraviolet A (UVA) showed relatively lower contributions, whereas derived environmental factors such as growing degree days (GDD) showed greater importance in model fitting.

Compared with JGRA and EADW+GW as benchmark methods, the predictive performance of TOPlus on 17 agronomic traits across environments was evaluated. TOPlus achieved an average prediction accuracy of 0.66, ranging from 0.24 to 0.88, which significantly outperformed the two benchmark methods for most traits, with 12% improvement over JGRA and 3% improvement over EADW+GW on average (Figure 2C; Supplemental Table 2). We used ear weight to exemplify the TOPlus performance for cross-environment prediction. For the Liaoning environment, TOPlus showed comparable accuracy to EADW+GW but significantly higher than JGRA. In other environments, TOPlus achieved the highest predictive performance for ear weight, followed by EADW+GW and JGRA (Figure 2D). Consistent relative improvements were also observed for the remaining 16 traits (Supplemental Figure 6). To evaluate the robustness of TOPlus to incomplete phenotypic records under realistic breeding scenarios, we simulated phenotypic sparsity by randomly masking 10%–50% of the phenotypic records across the five environments in the fixed 569-hybrid training set and evaluated three representative traits: DTA, PH, and EW. Prediction accuracy gradually decreased as the proportion of missing phenotypic records increased but remained broadly comparable to that obtained using the complete training data even when 50% of the records were masked, indicating

that TOPlus maintained robust predictive performance under moderate phenotypic sparsity (Supplemental Figure 7).

To further assess the utility of TOPlus on trait cross-environment prediction in the independent population, we validated the TOPlus model in the association mapping panel with 507 genetically diverse maize inbred lines (AMP population) showing genetic differentiation from the training set (NCII population) (Figure 2A). Two prediction scenarios were evaluated using days to silking (DTS), a vital trait characterizing vegetative-to-reproductive transition during plant development, for which complete phenotypic records were available in both the past (2011–2012) and future (2019–2020) environments relative to the environments of training data (2015). Compared with GBLUP, which served as the internal trait-prediction module in the original TOP method, TOPlus consistently achieved higher prediction accuracy across all five past environments (mean prediction accuracy: 0.37 vs. 0.34) and showed comparable overall performance in future environments (mean prediction accuracy: 0.40 vs. 0.39), with improved accuracy in four of the five validation environments (Figure 2E). We observed apparent variation in prediction accuracy across environments, which may be explained by the enviromic heterogeneity between the training and validation datasets (Supplemental Figure 4A). Furthermore, the potential of TOPlus for breeding applications through hybrid ranking was evaluated. We found that the top 5% of candidate hybrids selected based on predicted DTS generally showed earlier observed silking dates than the hybrids in 10 randomly selected sets across multiple environments, indicating that TOPlus, at its current level of prediction accuracy, has practical utility for selecting earlier-maturing candidates, which is important for balancing yield performance and breeding objectives (Supplemental Figure 8). Together, these results demonstrate that TOPlus maintains robust predictive performance across years, environments and populations, supporting its utility for practical breeding applications.

Hybrid selection performance of TOPlus using functional gene priors

The rapid expansion of crop functional genomics is reshaping the design principles of modern breeding algorithms. Thus, we incorporated functional genes into the TOPlus selective module, aiming to further improve hybrid selection performance.

To explicitly evaluate the contribution of functional genes as prior information, we incorporated 50, 100, and 180 functional genes into TOPlus (Supplemental Table 3), using equal

numbers of randomly selected genes as controls. We found that the inclusion of 50 functional genes resulted in modest and occasionally environment-dependent effects, while including 180 functional genes led to a clear improvement in selecting superior hybrids, with an average 3.18% yield improvement relative to using random genes (Supplemental Figure 9). Taking the Hebei environment as an example, the model fitting 50–180 functional genes enabled us to select hybrids with mean yield ranging from 280.91–290.49g, whereas the model incorporating random genes selected hybrids with apparently lower yield (275.05–278.47g). It indicated that the biological priors of functional genes had the potential to mathematically guide breeding practices.

To compare the performance of TOPlus with that of the original TOP model, which did not incorporate functional genes, a two-stage approach was applied, consisting of model-based pre-selection followed by phenotype-driven prioritization. The two methods identified different sets of 25 hybrids superior to "Zhengdan958" (Lai et al., 2010) as the benchmark variety from the testing population of 5251 hybrids based on algorithm prioritization with the same parameters. When validated using the observed phenotypes of the selected hybrids, it was found that the hybrids selected by TOPlus exhibited significantly higher yield than those selected by the original TOP method across environments ($P < 0.01$), with the yield improvements of 5.90–19.64%, while maintaining comparable global trait similarity to the benchmark variety between the two methods (Figure 3A). In the Hebei environment, for example, the superior hybrids prioritized by TOPlus exhibited a pronounced shift toward higher yield compared with those selected by the TOP method. In the pool of selected hybrids, the TOPlus-based hybrids resulted in a 9.54% mean genetic gain of yield relative to the benchmark, compared with a 4.09% gain from TOP-based selection (Figure 3B). In Hebei, the best hybrid selected by TOPlus showed a clear improvement in most yield-related traits relative to TOP *per se*, including ear row number (ERN), kernel number per row (KNPR), kernel number per ear (KNPE), and kernel weight per ear (KWPE), resulting in an increase in ear weight from 308.75 g to 347.66 g in the top-selected hybrid (Figure 3C). Similar performance was observed for the TOPlus method across the other four environments (Supplemental Figure 10).

Biological interpretation of model-inferred gene contributions to cross-environment hybrid prioritization

In real breeding programs, hundreds or thousands of candidate hybrids are tested in multi-environment trials, with elite commercial varieties used as field checks. Based on field phenotyping, breeders focus on selecting superior hybrids with higher yields than the check varieties and assessing the environmental stability and specificity of candidate hybrids. Following this paradigm, we tested the performance of TOPlus in identifying superior hybrids with environmental sensitivity and stability.

We designated Jilin, Liaoning, Beijing, Hebei, and Henan as five independent testing environments. In each environment, TOPlus was used to predict the phenotypes of untested hybrids and to select the top 50 candidates relative to the target variety. The majority of the selected hybrids exhibited pronounced environmental specificity (Figure 4A), with 56 hybrids selected in only a single environment, reflecting genotype-by-environment effects on target-related trait expression. In contrast, 13 hybrids were selected in all five environments.

To further explore the biological patterns underlying this environment-dependent selection, we examined the importance of functional genes within the TOPlus selective module. Model-derived gene weights were treated as indicators of feature importance. Among the 180 prior genes, those with null model weights across all environments were excluded from subsequent analyses, resulting in 109 informative genes for hybrid prioritization in at least one environment. Hierarchical clustering of environment-specific contribution profiles partitioned the 109 genes into two distinct classes: TOPlus-derived stable genes (n = 42), whose importance values remained relatively consistent across environments, and TOPlus-derived plastic genes (n = 67), whose importance values were highly sensitive to environmental variation (Figure 4B). The numbers of stable and plastic genes identified in different environments are shown in Supplemental Figure 11.

To examine the link between model-derived gene contribution patterns and hybrid environmental sensitivity, we compared the proportions of stable and plastic genes among hybrid groups with different environmental response patterns. Hybrids selected in a specific environment showed comparable contributions from stable and plastic genes. Interestingly, stable genes gradually became dominant over plastic genes among hybrids shared across multiple environments (Figure 4C). These results indicate that the model-inferred features may provide a useful explanation for the performance of TOPlus in identifying broadly adaptable superior hybrids across environments.

To better understand the biological implications of model-derived genes, we performed Gene Ontology (GO) enrichment analysis, which suggested functional differences between the two gene classes. Stable genes were predominantly associated with core developmental processes, whereas plastic genes were enriched in pathways related to stress responses and environmental adaptation (Figure 4D). For instance, *UB2*, previously reported to control ear row number (Du et al., 2020), was classified as an environmentally stable gene by the predictive model, possibly reflecting its consistent additive effect on ear row number across environments (Figure 4E). In contrast, *ZmTPS14.1*, previously reported to regulate flowering time (Jin et al., 2023), was classified as a plastic gene by the predictive model and exhibited highly dynamic genetic effects across environments (Figure 4E). These results suggest that TOPlus captured biologically meaningful patterns from gene-based model features, providing interpretable guidance for selecting superior hybrids with environment-specific responses.

Independent validation of the predictive and selective potential of TOPlus in commercial hybrids

It is important to test the applicability of the developed method in contemporary breeding populations. Thus, we selected 110 recently released commercial maize hybrid varieties as an independent validation population, comprising two subsets of 81 and 29 hybrids, to evaluate the performance of TOPlus trained on 569 hybrids from the NCII population in 2015.

PCA analysis revealed genetic similarity between the training population and two testing populations, indicating the feasibility of extrapolative prediction (Figure 5A). To provide geographic context for the subsequent validation and regional adaptation analyses, 14 representative sites spanning four major maize production regions in China were considered (Supplemental Table 4). To validate the predictive ability, two subsets comprising 110 commercial hybrids were evaluated in independent field trials. The subset of 81 hybrids was tested in Xiangyang in 2023, where the prediction accuracy for EW reached 0.49 (Figure 5B), while the subset of 29 hybrids was tested across 6 environments in 2022, with prediction accuracy for EW ranging from 0.254 to 0.427, with an average of 0.341 (Figure 5C), demonstrating the stable predictive performance of TOPlus across diverse agroecological conditions.

The 110 commercial hybrid varieties had received national variety approval in China, with recommended ecological production regions for each hybrid based on agronomic and yield performance from comprehensive multi-environment trials (Supplemental Table 5). TOPlus has the ability to prioritize hybrids superior to the benchmark variety in different environments, thus providing the potential to predict suitable production regions before extensive multi-environment trials. To test the feasibility, we predicted the most suitable production region for 110 hybrid varieties (see methods), and evaluated the agreement between the prediction results and national variety approval records. Given the limited number of varieties, we focused on the results for the Northeast and Huang-Huai-Hai regions, two major maize production regions in China with distinct climatic conditions. The TOPlus algorithm correctly identified 71 of the 76 varieties approved for the Northeast region and 15 of the 19 varieties approved for the Huang-Huai-Hai region (Figure 5D). The commercial hybrid Denghai1810, which was officially approved for cultivation in the Huang-Huai-Hai region, was predicted by TOPlus to exhibit favorable adaptability and yield stability in this region (Figure 5E). Interestingly, the predictive model indicated that this variety has the potential to expand its production area into the Northwest and Northeast regions, potentially providing a new cost-effective paradigm for future variety production trials aided by artificial intelligence and agricultural big data.

Discussion

We propose the TOPlus algorithm, which integrates an interpretable prior-gene knowledge layer into a multi-trait synergistic selection strategy that jointly leverages genetic and environmental information to support more biologically informed decision-making in breeding. Our comprehensive evaluation demonstrated the reliable breeding potential for multi-trait hybrid selection based on accurate cross-environment prediction. In practical breeding, parent selection, crossing design, and field evaluation remain fundamental steps that define the genetic potential and practical value of a breeding population. TOPlus is therefore best viewed as a decision-support framework that complements these upstream decisions by improving early prediction and prioritization among candidate hybrids generated from existing breeding materials. By ranking candidate hybrids before extensive field testing, TOPlus may help enrich promising materials and allocate field-testing resources more efficiently.

For cross-environment prediction, JGRA and EADW+GW are classic analytic approaches (Costa-Neto et al., 2021; Li et al., 2021). JGRA applies a linear model to all traits and may be limited in identifying or recommending specific environmental variables for prediction. EADW+GW relies on overall inter-environment similarity but may fail to distinguish diverse contributions of individual environmental factors. The predictive module of TOPlus addresses these limitations through a flexible modeling strategy: a linear model is applied to flowering traits, whereas quadratic models are used for other complex traits. Despite these improvements, cross-environment prediction in independent populations remains intrinsically challenging. Prediction accuracy can be constrained by trait complexity, genotype-by-environment effects, genetic divergence between training and validation populations, and the environmental distance between training and target environments. This limitation is especially relevant when target environments fall outside the environmental range represented by the training data, where prediction becomes partly extrapolative. Therefore, the selected environmental variables, model-derived gene contribution patterns, and hybrid rankings should be interpreted within the genetic and environmental scope of the training data. Nevertheless, the current prediction accuracy may still provide practical value when it consistently improves candidate ranking and helps reduce the scale of downstream field evaluation.

Increasing evidence suggests that noncoding variants contribute substantially to complex traits in plants. For example, open chromatin regions have been shown to explain a large proportion of phenotypic variation (Rodgers-Melnick et al., 2016), and sequence variation within promoters or enhancers can alter key agronomic traits (Qiu et al., 2025; Studer et al., 2011). Consequently, the current implementation of TOPlus may not fully capture prior functional knowledge for complex traits. Recent studies have demonstrated that models incorporating trait-associated markers outperform those relying on genome-wide or randomly selected SNPs with prediction accuracy improvements exceeding 20% for several agronomic traits (He et al., 2025; Xu et al., 2024). It may be feasible to incorporate more functional elements identified from ATAC-seq, chromatin interaction maps, and regulatory maps into both the predictive and selective modules of TOPlus to further improve model performance. The biological patterns of the stable and plastic genes were inferred from model-derived weights across environmental contexts; thus, they should be interpreted jointly with the environmental conditions and trait responses represented in the analysis, rather than as fixed biological attributes of individual

genes. Broader sampling across multiple years and stress scenarios will help evaluate the consistency of these patterns and refine their biological interpretation.

Across multiple datasets, TOPlus demonstrates consistent predictive and selective capacity across environments, populations, and breeding contexts, supporting the identification of superior hybrids, the achievement of measurable genetic gain relative to elite benchmarks, and reliable suitability assessment across agroecological regions. These properties suggest that the model could capture transferable genotype-environment signals and support deployment-oriented breeding decisions. An important direction for future research is the development of a scalable and transferable model trained on large historical datasets that integrate multi-environment trials, environmental variables, and diverse functional elements (Cooper et al., 2021; Gogna et al., 2026), with the flexibility to tailor training populations and environmental representations to specific testing scenarios, thereby enhancing model generalizability, reducing the need for extensive phenotyping, and enabling more efficient extrapolative prediction across populations and target environments.

The fundamental goal of a breeding program is to identify, in advance, core inbred lines and candidate hybrids that exhibit superior performance under uncertain future environments. As extreme climate events become more frequent, data from a single year or a single test environment are insufficient to capture the true capacity of genotypes to respond to environmental variation (Khaipho-Burch et al., 2023; Snowdon et al., 2021). By integrating environmental fluctuations over nine years (2015–2023), the model could systematically evaluate hybrid performance across a broader distribution of environmental conditions at specific locations over consecutive years, which is often difficult to achieve through field phenotyping because of resource and labor constraints (Supplemental Figure 12). This may provide a reliable route for selecting lines and hybrids with superior and stable long-term performance in silico before field trials in future breeding programs.

Elite hybrid varieties are typically defined as high-yielding, high-quality, and stress- and disease-resistant in commercial breeding programs. Traditionally, breeders perform numerous trials across multiple environments and years and select the hybrids that outperform the check variety, with a primary focus on yield performance. Essentially, this is a pragmatic paradigm for multi-trait selection by exposing candidate hybrids to diverse natural stresses. For example, trials affected by storm threats probably provide natural opportunities for selecting hybrids with

lodging resistance; environments with prevalent pest and disease outbreaks are beneficial for selecting hybrids with disease and pest resistance. Previously, we proposed the target-oriented prioritization (TOP) approach, demonstrating a novel breeding route for selecting multiple traits based on breeding values rather than selecting yield per se across dozens of environments. In this study, we advance TOP to TOPlus by considering the resource trade-off between multi-trait phenotyping and multi-environment trials, enabling the selection of elite hybrids by simultaneously considering multi-trait performance and across-environment stability (Supplemental Figure 13). Given the climatic heterogeneity across maize production areas in China, maize breeding requires the development of elite hybrid varieties suitable for specific ecological production regions. Interestingly, the proof-of-concept test was performed using the 110 commercial maize hybrid varieties, which had been developed through traditional breeding approaches involving yield-based multi-environment selection. Notably, TOPlus, trained on data from a limited number of environments and several key agronomic traits, demonstrated the potential to independently predict suitable production regions, especially for varieties certified for the Northeast and Huang–Huai–Hai ecological zones. Our proposed method and analytical results demonstrate a new breeding paradigm driven by agricultural big data and advanced algorithms, facilitating the optimization of resource use and the acceleration of breeding cycles in maize and other crops.

Methods

Genotype and phenotype datasets

The maize NCII population consisted of 5,820 F_1 hybrids derived from crosses between 194 maternal inbred lines, which represent a subset of the Complete-diallel plus Unbalanced Breeding-derived Inter-Cross (CUBIC) population, and 30 diverse elite paternal lines, with a total of 13.8 million single nucleotide polymorphisms (SNPs) identified in this population (Xiao et al., 2021). The genotypes of the NCII hybrids were inferred from the genotypes of their corresponding maternal and paternal inbred parents according to the crossing design. For the hybrid panel, SNPs with a minor allele frequency (MAF) < 0.05 were excluded using PLINK v1.90 (Chang et al., 2015), followed by linkage disequilibrium (LD) pruning with a threshold of 0.2. From the remaining variants, 25,454 SNPs were randomly and uniformly sampled across the

maize genome to construct a sparse yet representative marker set for downstream analyses. Missing hybrid genotypes resulting from parental heterozygosity were imputed using Beagle v4.1 (Browning and Browning, 2007).

According to the original description of this public dataset, the NCII hybrids were evaluated using an augmented field design. At each location, each hybrid was planted as a single-row plot containing 17 individual plants. The field trials were conducted at a planting density of 4,500 plants per mu, equivalent to approximately 67,500 plants ha^{-1}, which represents a conventional maize production density in China. Plants with multiple ears were treated as missing phenotypic records. The commercial hybrid cultivar Zhengdan958 was included as a control and planted once every 50 rows, and its phenotypic data were used to correct for field trends and spatial heterogeneity. Field trials were conducted across five locations in 2015 for phenotypic evaluation: Yushu in Jilin Province (43°42′N, 125°18′E), Shenyang in Liaoning Province (42°03′N, 123°33′E), Beijing (40°10′N, 116°21′E), Baoding in Hebei Province (38°39′N, 115°51′E), and Xinxiang in Henan Province (35°27′N, 114°01′E).

Seventeen agronomic traits were evaluated across all environments, including three flowering time traits: days to tasseling (DTT), days to anthesis (DTA), and days to silking (DTS), each defined as the interval from sowing to the day when approximately 50% of plants within a line reached tassel emergence, pollen shedding, and silk emergence, respectively. Six plant architecture traits were recorded, including plant height (PH, measured as the vertical distance from the soil surface to the top of the tassel), ear height (EH, measured as the distance from the soil surface to the node bearing the uppermost ear), ear leaf length (ELL, measured as the straight length of the primary ear leaf), ear leaf width (ELW, measured at the widest point of the ear leaf), tassel length (TL, measured as the length of the main tassel axis), and tassel branch number (TBN, counted as the number of primary tassel branches). For these traits, measurements were taken from five consecutive plants in the center of each row, and mean values were used for subsequent analyses. In addition, eight yield-related traits were measured, including cob weight (CW), ear weight (EW), ear diameter (ED), ear length (EL), ear row number (ERN), kernel number per ear (KNPE), kernel number per row (KNPR), and kernel weight per ear (KWPE). For yield traits, five representative ears were randomly sampled from each line and used to obtain averaged measurements.

To characterize phenotypic variation and cross-environment consistency in the NCII population, we examined the phenotypic distributions of DTA, PH, and EW (Supplemental Figure 14), calculated pairwise correlations among the 17 agronomic traits within each environment (Supplemental Figure 15) and cross-environment correlations for each trait (Supplemental Figure 16), and estimated across-environment broad-sense heritability by treating the five NCII environments as repeated observations (Supplemental Figure 17).

The maize AMP population constitutes a globally diverse panel of 513 inbred lines that has been extensively used over the past 15 years (Fu et al., 2013; Jin et al., 2017; Li et al., 2013; Wen et al., 2014; Yang et al., 2011). Based on whole-genome resequencing data from the AMP population (Chen et al., 2022), SNP filtering and marker selection were conducted using the same pipeline applied to the NCII training population. Following marker harmonization, 25,143 SNP loci shared between the AMP and NCII datasets were retained, ensuring concordance in genomic coordinates and allele coding across the training and testing populations. This marker concordance enabled direct application of the NCII-trained TOPlus model to the AMP population without additional harmonization.

Field evaluations for the AMP population were carried out during 2011 and 2012 at five Chinese locations, specifically Hebi in Henan Province (35°44'N, 114°18'E), Wuhan in Hubei Province (30°22'N, 114°19'E), Chongqing (29°30'N, 106°31'E), Kunming in Yunnan Province (25°02'N, 102°42'E), and Sanya in Hainan Province (18°21'N, 109°10'E). Multiple agronomic traits, including flowering-time traits, were evaluated following the measurement procedures described above (Liu et al., 2020b). Furthermore, the AMP inbred population was cultivated across five distinct environments in 2019 and 2020 for comprehensive phenotypic characterization, including trial sites at Gongzhuling in Jilin Province (43°30'N, 124°49'E), Beijing (39°54'N, 116°28'E), Zhangye in Gansu Province (38°55'N, 100°27'E), Wuhan in Hubei Province (30°37'N, 114°21'E), and Sanya in Hainan Province (18°12'N, 109°30'E). DTS and ELW were measured using the same protocols described above. For independent validation of the TOPlus prediction framework, DTS was used as the representative trait because complete phenotypic records were consistently available across both past (2011–2012) and future (2019–2020) environments relative to the 2015 training environments of the NCII population.

Environmental dataset

Daily environmental data were retrieved for each trial using the Critical Environmental Regressor through Informed Search (CERIS) method (Li et al., 2021), based on the geographic coordinates of experimental sites and the corresponding crop growing periods. These data were used to represent the dynamic environmental conditions experienced during crop development. The growing period was defined from sowing to physiological maturity.

A total of 14 primary environmental factors were collected from two public data sources. Day length (DL) was obtained from the National Oceanic and Atmospheric Administration (NOAA; http://www.ncdc.noaa.gov), while the remaining 13 factors were retrieved from the NASA POWER platform (https://power.larc.nasa.gov/data-access-viewer/), including all-sky surface photosynthetically active radiation (APAR), clear-sky surface photosynthetically active radiation (CPAR), precipitation (PR), surface pressure (PS), relative humidity at 2 m (RH), profile soil moisture (SM), root zone soil wetness (SW), maximum temperature at 2 m (TMAX), minimum temperature at 2 m (TMIN), all-sky surface UVA irradiance (UVA), all-sky surface UVB irradiance (UVB), wind direction at 2 m (WD), and wind speed at 2 m (WS). To better capture environmentally driven crop responses, 11 additional derived environmental factors were generated, including diurnal temperature range, growing degree days, and other composite indices integrating multiple factors. In total, 25 environmental factors were obtained (Supplemental Table 1).

The TOPlus algorithm

TOPlus is an advanced, flexible framework that extends the TOP algorithm by integrating prior genetic knowledge and environmental information to identify optimal breeding candidates for specific environments. This method selects individuals through the following computational workflow: 1) predicts phenotypes in untested environments and fits values for functional genes within the training population; 2) applies the TOP algorithm to derive environment-specific trait weights and gene-specific weights from the training population; and 3) uses the similarity function to select individuals with the highest overall phenotypic and genotypic similarity to the target and superior performance in target traits (Supplemental Figure 1).

1. Phenotype prediction

In the phenotype prediction module, environment-specific phenotypic values were used as response variables to retain genotype-by-environment information. Functional gene priors were

not used as predictors in this phenotype prediction module; they were introduced only in the subsequent selective module for hybrid prioritization.

For the NCII analyses, we followed the fixed partition used in the previous TOP study (Yang et al., 2022): 569 hybrids from the three diagonal strips of the NCII mating design were used as the training set, and the remaining 5,251 off-diagonal hybrids were used as the testing set. This same partition was applied to TOPlus and all benchmark methods.

TOPlus implements an environment-driven iterative factor selection strategy, in which each training environment is sequentially designated as the selection environment to identify key environmental variables that correspond to the environmental factors filtered using the CERIS algorithm and that are significantly associated with the target trait, while the remaining environments are harnessed to fit trait-specific prediction models, and phenotypes in untested environments are predicted by aggregating the contributions of these pre-selected environmental variables.

(1) Environmental variable selection

Environmental factors represent broad categories of environmental measurements (e.g., temperature-, radiation-, or moisture-related factors), whereas environmental variables denote window-specific values extracted from defined temporal intervals within each factor.

Given that different agronomic traits exhibit distinct sensitivities to environmental conditions across developmental stages, trait-specific temporal ranges were defined based on biological relevance. For flowering traits, analyses were restricted to the period from sowing to pre-flowering stages, whereas for plant architecture and yield-related traits, environmental conditions across the entire growing period were considered.

To characterize the temporal dynamics of environmental factors during crop development, a sliding window approach was applied to the daily environmental factor time series within the defined temporal ranges. Windows of varying lengths, ranging from 7 days to the full length of the growing period, were generated and sequentially moved along the timeline with a step size of one day. The mean value of each environmental factor within a given window was calculated and used as the corresponding environmental variable. For each training environment, the correlation between window-specific environmental variables and the mean phenotypic value of the corresponding population was calculated. The window with the maximum absolute correlation coefficient was selected as the representative window for each environmental factor.

In cases where multiple windows showed identical maximum absolute correlations, the window with the longest duration was prioritized. If ties persisted, the earliest window in the growing period was selected, favoring environmental signals with longer cumulative effects on subsequent development (Supplemental Figure 1A).

(2) Phenotype prediction for training set individuals

The function $y = f_{i,j}^*(x)$ is required to be fitted based on the training dataset $\{(x_{i,l}, y_{l,j}),\ l \in \{1,2,\cdots,m-1\}\ and\ l \neq c\}$ such that $f_{i,j}^*(x) = \arg\min_{f_{i,j}(x)} \sum_{\substack{l=1 \\ l\neq c}}^{m-1} (f_{i,j}(x_{i,l}) - y_{l,j})^2$, where i represents i-th environmental variable, l represents l-th environmental and $j \in \{1,2,\cdots,J\}$ represents j-th individual. For flowering traits $f_{i,j}(x) = a_{0,i,j} + a_{1,i,j}x$ and for all other traits $f_{i,j}(x) = a_{0,i,j} + a_{1,i,j}x + a_{2,i,j}x^2$. The selected set of environmental variables, denoted as $S_c = \{i | corr(\boldsymbol{f_i^*}, \boldsymbol{y_c}) > \theta\}$, is determined by screening through the selection environment $c \in \{1,2,\cdots,m-1\}$, where $\boldsymbol{f_i^*} = \left(f_{i,1}^*(x_{i,c}), f_{i,2}^*(x_{i,c}), \cdots, f_{i,J}^*(x_{i,c})\right)^T$, $\boldsymbol{y_c} = (y_{c,1}, y_{c,2}, \cdots, y_{c,J})^T$ and θ is a given threshold set to 0.25. By iterating through the selection environment c, the predicted phenotypic value for the j-th individual in the m-th environment is obtained as

$$\tilde{y}_{m,j} = \frac{1}{m-1}\sum_{c=1}^{m-1} \frac{1}{|S_c|} \sum_{i \in S_c} f_{i,j}^*(x_{i,m}) \qquad (1)$$

where $|S_c|$ represents the cardinality of S_c (number of selected variables). In the training set, we replaced the observed values $y_{l,j}$ with their predicted values $\hat{y}_{l,j}$ obtained from 10-fold cross-validation within a single environment. Then, the predicted phenotypic value for the j-th individual in the m-th environment $\hat{y}_{m,j}$ is derived using S_c via Equation (1).

(3) Phenotype prediction for testing set individuals

When predicting the phenotype $\hat{y}_{m,k}$ ($k \in \{J+1, J+2, \cdots, J+K\}$) for individuals in the testing set within the m-th environment, the observed values $y_{l,j}$ in the training set were replaced with the predicted phenotype values $\hat{y}_{l,k}$ of the testing set individuals from the first $m-1$ environments to obtain $f_{i,k}^*(x)$. Subsequently, utilizing the selected set of environmental variables S_c from the training set, $\hat{y}_{m,k}$ can be calculated through Equation (1) (Supplemental Figure 1B).

2. Genotype fitting

TOPlus utilizes functional genes as prior information and estimates their contribution values based on phenotypic data from the training population. We used a mixed linear model incorporating the phenotypic kinship matrix to estimate the contribution value of each functional gene. The mixed linear model is specified as follows: $g = X\beta + \xi + \varepsilon$, where: g is an $n \times 1$ vector of genotypic values for the functional gene across n individuals, X is an $n \times p$ design matrix, β is a $p \times 1$ vector of fixed effects, $\xi \sim N(0, K_p\sigma_p^2)$ is an $n \times 1$ vector of random effects, $\varepsilon \sim N(0, I\sigma_e^2)$ is an $n \times 1$ vector of residual errors, σ_p^2 and σ_e^2 are variance components associated with random effects and residual errors, respectively, I is an identity matrix. K_p is a kinship matrix constructed from multi-trait phenotypic data of individuals, using the single-view kinship matrix method from MVBLUP (Wu et al., 2025). The random effects vector is derived using the following equation: $\hat{\xi} = \sigma_p^2 K_p V^{-1}(g - X\hat{\beta})$, where $V = K_p\sigma_p^2 + I\sigma_e^2$ and $\hat{\beta} = (X^T V^{-1} X)^{-1}(X^T V^{-1} g)$ (Supplemental Figure 1C).

3. Model training

(1) Target determination

TOPlus algorithm takes as its core targets the multi-trait phenotypic profiles of elite hybrid lines and their favorable allelic genotypes at functional genes under specific environmental conditions. These target phenotypic values can be either directly observed or predicted from the training population. SNP variants located within each of the 180 genes were tested for association with ear weight in the training population. Specifically, for each gene-associated locus, individuals carrying the candidate favorable genotype were compared with the remaining individuals using a Student's *t*-test. Genes containing loci showing significant phenotypic superiority in the favorable direction were retained as gene priors for TOPlus model training and selection. The gene set used in this study consisted of 180 genes, including 143 genes collected from previous studies and public functional gene resources and 37 candidate genes identified through in-house QTL mapping and GWAS based on associations with ear weight. For each gene, the gene ID, associated trait category, genomic position, source, and TOPlus-derived gene class are provided in Supplemental Table 3. Notably, although the prior gene set used in this study is dominated by maize yield-associated genes, this analytical framework is equally amenable to the integration of genes associated with other agronomic traits or breeding objectives in other crops.

(2) Model construction and solution

TOPlus constructs a model using the pseudo-observation vector $\tilde{\boldsymbol{y}}_{\boldsymbol{m,j}}$ and predicted value vector $\hat{\boldsymbol{y}}_{\boldsymbol{m,j}}$ in the m-th environment of the training population, along with the genotype vector $\boldsymbol{g_j}$ of the target hybrid carrying the favorable allelic genotype of yield-related genes and the prediction vector $\hat{\boldsymbol{g}}_{\boldsymbol{j}}$ of these genes. The TOP method is employed to derive the optimal weight vector: $\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}} \prod_{j=1}^{J} \exp(-\boldsymbol{w_p^T}|\tilde{\boldsymbol{y}}_{\boldsymbol{m,j}} - \hat{\boldsymbol{y}}_{\boldsymbol{m,j}}| - \boldsymbol{w_g^T}|\boldsymbol{g_j} - \hat{\boldsymbol{g}}_{\boldsymbol{j}}|)$, where $\boldsymbol{w_p}$ represents the weight vector for multi-trait phenotypes, $\boldsymbol{w_g}$ denotes the weight vector for incorporated functional genes and $\boldsymbol{w} = (\boldsymbol{w_p^T}, \boldsymbol{w_g^T})^{\boldsymbol{T}}$ (Supplemental Figure 1D).

4. Selection in specific environments

TOPlus employs a probabilistic similarity function, defined as follows:

$$P(\hat{\boldsymbol{y}}_{\boldsymbol{m,k}}, \boldsymbol{y_t}) = \frac{\exp(-\boldsymbol{w_p^{*T}}|\boldsymbol{y_t} - \hat{\boldsymbol{y}}_{\boldsymbol{m,k}}| - \boldsymbol{w_g^{*T}}|\boldsymbol{g_t} - \boldsymbol{g_k}|)}{\sum_{k=J+1}^{J+K} \exp(-\boldsymbol{w_p^{*T}}|\boldsymbol{y_t} - \hat{\boldsymbol{y}}_{\boldsymbol{m,k}}| - \boldsymbol{w_g^{*T}}|\boldsymbol{g_t} - \boldsymbol{g_k}|)}$$

This function quantifies the integrated phenotypic-genotypic similarity between the k-th test hybrid and the target ideotype in the m-th environment. Here, $\boldsymbol{y_t}$ and $\hat{\boldsymbol{y}}_{\boldsymbol{m,k}}$ represent the phenotype vector of the target ideotype and the phenotype value vector of the k-th individual within the testing population, respectively. Similarly, $\boldsymbol{g_t}$ and $\boldsymbol{g_k}$ refer to the favorable allelic genotypes of yield-related functional genes in the target and the corresponding genotypes of the k-th test hybrid, respectively (Supplemental Figure 1D).

Field validation of TOPlus model

We conducted a systematic evaluation of multi-year and multi-location performance for 110 commercial maize hybrids using the TOPlus model. Genotypic data for the 110 commercial hybrids were generated using a maize 45K genotyping-by-target-sequencing (GBTS) panel based on liquid-phase probe capture by MolBreeding Biotechnology Co., Ltd. The panel was designed to detect 44,637 predefined SNP loci. Captured libraries were sequenced on the DNBSEQ-T7 platform using paired-end 150-bp reads, with an average target sequencing depth of approximately 30×. Specifically, genotypes at 24,822 SNP loci overlapping those defined in the NCII training population were retained for subsequent analyses, ensuring consistency in genomic coordinates and allele coding between the training and validation datasets. The TOPlus model was trained on 569 NCII F_1 hybrids combined with environmental data collected from 2015 to 2023 across 14 representative sites spanning four major production regions. The Northwest region included Changji (44°01′N, 87°19′E), Bayannaoer (40°45′N, 107°25′E) and Zhangye

(38°51′N, 100°23′E), while the Northeast region comprised Suihua (46°38′N, 127°00′E), Gongzhuling (43°30′N, 124°49′E), Tieling (42°17′N, 123°51′E) and Xinzhou (38°25′N, 112°44′E). The Southwest region covered Chengdu (30°43′N, 103°52′E) and Shannan (29°14′N, 91°20′E), with the Huang-Huai-Hai region encompassing Weifang (36°43′N, 119°06′E), Xinxiang (35°18′N, 113°53′E), Fuyang (32°54′N, 115°49′E), Xiangyang (32°18′N, 112°04′E) and Changsha (28°11′N, 113°04′E). For performance assessment, we established a dual-reference system utilizing Zhengdan958 as the nationwide standard and Xianyu335 as the regional benchmark specifically for Northeast China.

Our study implemented field trials across seven representative environments (Changji, Zhangye, Xinzhou, Xiangyang, Chengdu, Shannan, and Changsha) during 2022–2023. In 2022, 29 commercial hybrids were cultivated across six locations (Changji, Zhangye, Xinzhou, Chengdu, Shannan and Changsha). In 2023, 81 additional commercial hybrids were cultivated at Xiangyang. EW was recorded for all 110 hybrids following the measurement procedures described above.

For each commercial hybrid, the TOPlus model predicted its comprehensive performance across 126 test environments comprising 9 years and 14 locations. Selection criteria required hybrids to rank within the top 50% across test environments in terms of overall similarity with the target variety. By correlating these predictions with China's national maize variety registration data, which were obtained from the Seed Management Bureau database of the Ministry of Agriculture and Rural Affairs of China (http://202.127.42.47:6009/Home/BigDataIndex, accessed February 2026) by querying each hybrid using its official variety name, we systematically classified the 110 commercial hybrids into two distinct categories. The consistent category included varieties that received both national registration approval and model recommendation for cultivation in their designated production zones. The inconsistent category contained varieties that, while nationally registered, failed to meet the model's recommendation criteria for cultivation in their approved production zones.

Data and code availability

The source code of TOPlus and the demonstration datasets required to reproduce the example workflow are publicly and freely available on GitHub (Fang et al., 2026; https://github.com/Lee-

Shell/TOPlus) under the GNU General Public License v3.0. Detailed instructions for using TOPlus are provided in the Supplemental Information and in the TOPlus GitHub repository. The genotype and phenotype datasets used in this study, including those from the maize NCII population, the AMP population, and the commercial hybrid populations, are publicly available on Zenodo (https://zenodo.org/records/18605163).

Funding

This work was supported by the National Key Research and Development Program of China (2022YFD1201503) and the National Natural Science Foundation of China (32321005, 32201855).

Author Contributions

W.Y., Y.X. and J.Y. designed and supervised this study. W.F. developed the TOPlus algorithm and conducted model validation and performance evaluation. B.W. conducted comparative analyses against previously published models. W.F., X.H., Z.Z., Y.P., N.L. and W.L. participated in data preparation. T.G. provided critical input on model development and validation. W.F., J.Y., Y.X. and W.Y. prepared the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The computations in this paper were run on the bioinformatics computing platform of the National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University. We thank all reviewers for providing critical and constructive comments that helped improve the manuscript.

Supplemental Information

Document S1. Figures S1-S17

Document S2. Supplemental Tables 1-5.

Declaration of Interests

The authors have no conflicts of interest to declare.

References

Bailey-Serres, J., Parker, J.E., Ainsworth, E.A., Oldroyd, G.E.D., and Schroeder, J.I. (2019). Genetic strategies for improving crop yields. Nature **575**:109-118.

Bassi, F.M., Bentley, A.R., Charmet, G., Ortiz, R., and Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). Plant Science **242**:23-36.

Bernardo, R. (2014). Genomewide Selection when Major Genes Are Known. Crop Science **54**:68-75.

Bernardo, R., and Yu, J. (2007). Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Science **47**:1082-1090.

Browning, S.R., and Browning, B.L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. The American Journal of Human Genetics **81**:1084-1097.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience **4**10.1186/s13742-015-0047-8.

Chen, L., Luo, J., Jin, M., Yang, N., Liu, X., Peng, Y., Li, W., Phillips, A., Cameron, B., Bernal, J.S., et al. (2022). Genome sequencing reveals evidence of adaptive variation in the genus Zea. Nature Genetics **54**:1736-1745.

Cooper, M., Voss-Fels, K.P., Messina, C.D., Tang, T., and Hammer, G.L. (2021). Tackling G × E × M interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. Theoretical and Applied Genetics **134**:1625-1644.

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

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J.M., Pérez-Elizalde, S., Beyene, Y., et al. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends in Plant Science **22**:961-975.

Derbyshire, M.C., Newman, T.E., Thomas, W.J.W., Batley, J., and Edwards, D. (2024). The complex relationship between disease resistance and yield in crops. Plant Biotechnology Journal **22**:2612-2623.

Des Marais, D.L., Hernandez, K.M., and Juenger, T.E. (2013). Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment. Annual Review of Ecology, Evolution, and Systematics **44**:5-29.

Du, Y., Liu, L., Peng, Y., Li, M., Li, Y., Liu, D., Li, X., and Zhang, Z. (2020). UNBRANCHED3 Expression and Inflorescence Development is Mediated by UNBRANCHED2 and the Distal Enhancer, KRN4, in Maize. PLOS Genetics **16**:e1008764.

Dudley, J.W., and Lambert, R.J. (2003). 100 Generations of Selection for Oil and Protein in Corn. In Plant Breeding Reviews, pp. 79-110.

Fang, W., Yan, J., Xiao, Y., and Yang, W. (2026). TOPlus: An algorithm integrating prior genetic and environmental information for multi-trait synergic selection in maize hybrid breeding. GitHub. https://github.com/Lee-Shell/TOPlus. (2026.08.27).

FAO. (2009). How to Feed the World in 2050: Proceedings of the Expert Meeting. https://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf.

Farooq, M.A., Gao, S., Hassan, M.A., Huang, Z., Rasheed, A., Hearne, S., Prasanna, B., Li, X., and Li, H. (2024). Artificial intelligence in plant breeding. Trends in Genetics **40**:891-908.

Finlay, K., and Wilkinson, G. (1963). The analysis of adaptation in a plant-breeding programme. Australian Journal of Agricultural Research **14**:742-754.

Fu, J., Cheng, Y., Linghu, J., Yang, X., Kang, L., Zhang, Z., Zhang, J., He, C., Du, X., Peng, Z., et al. (2013). RNA sequencing reveals the complex regulatory network in the maize kernel. Nature Communications **4**:2832.

Gogna, A., Kamali, B., Wimmer, V., Schmidt, R.H., Rezaei, E.E., Eckhoff, W.M., Reif, J.C., and Zhao, Y. (2026). Predicting enviromically adapted varieties with big data. Genome Biology **27**:3.

He, K., Yu, T., Gao, S., Chen, S., Li, L., Zhang, X., Huang, C., Xu, Y., Wang, J., Prasanna, B.M., et al. (2025). Leveraging Automated Machine Learning for Environmental Data-Driven Genetic Analysis and Genomic Prediction in Maize Hybrids. Advanced Science **12**:2412423.

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. Nature Biotechnology **37**:744-754.

Jin, M., Zhang, X., Zhao, M., Deng, M., Du, Y., Zhou, Y., Wang, S., Tohge, T., Fernie, A.R., Willmitzer, L., et al. (2017). Integrated genomics-based mapping reveals the genetics underlying maize flavonoid biosynthesis. BMC Plant Biology **17**:17.

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 Communications **4**.

Jonas, E., and de Koning, D.-J. (2013). Does genomic selection have a future in plant breeding? Trends in Biotechnology **31**:497-504.

Khaipho-Burch, M., Cooper, M., Crossa, J., de Leon, N., Holland, J., Lewis, R., McCouch, S., Murray, S.C., Rabbi, I., Ronald, P., et al. (2023). Genetic modification can improve crop yields - but stop overselling it. Nature **621**:470-473.

Lai, J., Li, R., Xu, X., Jin, W., Xu, M., Zhao, H., Xiang, Z., Song, W., Ying, K., Zhang, M., et al. (2010). Genome-wide patterns of genetic variation among elite maize inbred lines. Nature Genetics **42**:1027-1030.

Laurie, C.C., Chasalow, S.D., LeDeaux, J.R., McCarroll, R., Bush, D., Hauge, B., Lai, C., Clark, D., Rocheford, T.R., and Dudley, J.W. (2004). The Genetic Architecture of Response to Long-Term Artificial Selection for Oil Concentration in the Maize Kernel. Genetics **168**:2141-2155.

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nature Genetics **45**:43-50.

Li, Q., Lai, J., Chen, J., Li, L., Song, W., Xin, B., Zhao, H., Xiao, Y., Tian, F., Li, G., et al. (2025). Decades' progress and prospects on maize functional genomics and molecular breeding. Science China Life Sciences **68**:3509-3574.

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. Molecular Plant **14**:874-887.

Liu, H.-J., Wang, X., Xiao, Y., Luo, J., Qiao, F., Yang, W., Zhang, R., Meng, Y., Sun, J., Yan, S., et al. (2020a). CUBIC: an atlas of genetic architecture promises directed maize improvement. Genome Biology **21**:20.

Liu, N., Du, Y., Warburton, M.L., Xiao, Y., and Yan, J. (2020b). Phenotypic Plasticity Contributes to Maize Adaptation and Heterosis. Molecular Biology and Evolution **38**:1262-1275.

Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics **157**:1819-1829.

Millet, E.J., Kruijer, W., Coupel-Ledru, A., Alvarez Prado, S., Cabrera-Bosquet, L., Lacube, S., Charcosset, A., Welcker, C., van Eeuwijk, F., and Tardieu, F. (2019). Genomic prediction of maize yield across European environmental conditions. Nature Genetics **51**:952-956.

Napier, J.D., Heckman, R.W., and Juenger, T.E. (2022). Gene-by-environment interactions in plants: Molecular mechanisms, environmental drivers, and adaptive plasticity. The Plant Cell **35**:109-124.

Qi, G., Li, Y., Zhang, W., Han, Z., Chen, J., Zhang, Z., Xuan, L., Chen, R., Fang, L., Hu, Y., et al. (2025). Reveal genomic insights into cotton domestication and improvement using gene level functional haplotype-based GWAS. Nature Communications **16**:4734.
Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., He, Q., Ou, S., Zhang, H., Li, X., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. Cell **184**:3542-3558.e3516.
Qiu, Y., Liu, L., Yan, J., Xiang, X., Wang, S., Luo, Y., Deng, K., Xu, J., Jin, M., Wu, X., et al. (2025). Precise engineering of gene expression by editing plasticity. Genome Biology **26**:51.
Rodgers-Melnick, E., Vera, D.L., Bass, H.W., and Buckler, E.S. (2016). Open chromatin reveals the functional maize genome. Proceedings of the National Academy of Sciences **113**:E3177-E3184.
Sabir, K., Rose, T., Wittkop, B., Stahl, A., Snowdon, R.J., Ballvora, A., Friedt, W., Kage, H., Léon, J., Ordon, F., et al. (2023). Stage-specific genotype-by-environment interactions determine yield components in wheat. Nature Plants **9**:1688-1696.
Scheiner, S.M. (1993). Genetics and Evolution of Phenotypic Plasticity. Annual Review of Ecology, Evolution, and Systematics **24**:35-68.
Sharma, H.C., Crouch, J.H., Sharma, K.K., Seetharama, N., and Hash, C.T. (2002). Applications of biotechnology for crop improvement: prospects and constraints. Plant Science **163**:381-395.
Snowdon, R.J., Wittkop, B., Chen, T.-W., and Stahl, A. (2021). Crop adaptation to climate change as a consequence of long-term breeding. Theoretical and Applied Genetics **134**:1613-1623.
Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. Nature Genetics **43**:1160-1163.
Tester, M., and Langridge, P. (2010). Breeding Technologies to Increase Crop Production in a Changing World. Science **327**:818-822.
Voss-Fels, K.P., Stahl, A., and Hickey, L.T. (2019). Q&A: modern crop breeding for future food security. BMC Biology **17**:18.
Wang, C., and Han, B. (2022). Twenty years of rice genomics research: From sequencing and functional genomics to quantitative genomics. Molecular Plant **15**:593-619.
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. Molecular Plant **16**:279-293.
Wang, Y., Zhang, P., Xiao, Y., Liu, H.-J., Ding, J., Qu, Y., Gao, J., Li, W., Guo, X., Wang, A., et al. (2025). GAIN: A long-term maize germplasm innovation plan. Molecular Plant **18**:915-919.
Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J., et al. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nature Communications **5**:3438.
Wu, B., Xiong, H., Zhuo, L., Xiao, Y., Yan, J., and Yang, W. (2025). Multi-view BLUP: a promising solution for post-omics data integrative prediction. Journal of Genetics and Genomics **52**:839-847.
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 Biology **22**:148.
Xu, Y., Wang, X., Ding, X., Zheng, X., Yang, Z., Xu, C., and Hu, Z. (2018). Genomic selection of agronomic traits in hybrid rice using an NCII population. Rice **11**:32.
Xu, Y., Zhang, Y., Cui, Y., Zhou, K., Yu, G., Yang, W., Wang, X., Li, F., Guan, X., Zhang, X., et al. (2024). GA-GBLUP: leveraging the genetic algorithm to improve the predictability of genomic selection. Briefings in Bioinformatics **25**.
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 Biology **22**:271.
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 Biology **23**:80.

Yang, X., Gao, S., Xu, S., Zhang, Z., Prasanna, B.M., Li, L., Li, J., and Yan, J. (2011). Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. Molecular Breeding **28**:511-526.

Yang, Z., Gao, C., Zhang, Y., Yan, Q., Hu, W., Yang, L., Wang, Z., and Li, F. (2023). Recent progression and future perspectives in cotton genomic breeding. Journal of Integrative Plant Biology **65**:548-569.

Yates, F., and Cochran, W.G. (1938). The analysis of groups of experiments. The Journal of Agricultural Science **28**:556-580.

Zobel, R.W., Wright, M.J., and Gauch Jr., H.G. (1988). Statistical Analysis of a Yield Trial. Agronomy Journal **80**:388-393.

Figure legends

Figure 1. The schematic overview of the TOPlus algorithm.

TOPlus consists of a predictive module and a selective module. The predictive module leverages environmental information to predict the performance of untested genotypes in target environments by learning specific environmental response factors in a statistical model. Guided by the hypothesis of pleiotropy across multiple traits, the selective module incorporates functional gene priors to advance the multi-trait selection of breeding lines and hybrids that outperform benchmark cultivars in specific target environments. Trait abbreviations: DTA, days to anthesis; DTS, days to silking; PH, plant height; EH, ear height; EW, ear weight; and KWPE, kernel weight per ear.

Figure 2. Environment-informed TOPlus enables robust cross-environment prediction and generalization across maize populations.

(A) Principal component analysis (PCA) of the training population (TrP; NCII; 569 hybrids), testing population 0 (TeP0; NCII; 5,251 hybrids) and testing population 1 (TeP1; AMP; 507 inbred lines).

(B) Trait-specific environmental variables identified by TOPlus and their clustering patterns across environments. Environmental-factor abbreviations and their definitions, including those of formula-based derived indices, are provided in Supplemental Table 1.

(C) Prediction accuracy of 17 agronomic traits across five environments for joint genomic regression analysis (JGRA), EADW+GW, and TOPlus. Relative improvements were calculated using JGRA as the reference method: (method − JGRA) / JGRA. Significance was assessed by two-sided paired t-tests comparing each method with JGRA across the five environments for each trait. ns, $P \geq 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

(D) EW prediction accuracy in the NCII population across five environments. The black error bars denote the standard errors.

(E) Prediction accuracy for DTS in the AMP population across past environments (2011–2012; left) and future environments (2019–2020; right). The training data of the NCII population were collected in 2015. GBLUP represents the predictive module of the original TOP model.

Prediction accuracy in (C–E) was quantified as the Pearson correlation coefficient between observed and predicted phenotypes. Trait abbreviations: DTT, days to tasseling; DTA, days to anthesis; DTS, days to silking; PH, plant height; EH, ear height; ELL, ear leaf length; ELW, ear leaf width; TBN, tassel branch number; TL, tassel length; CW, cob weight; EW, ear weight; ED, ear diameter; EL, ear length; ERN, ear row number; KNPE, kernel number per ear; KNPR, kernel number per row; and KWPE, kernel weight per ear.

Figure 3. Functional gene-informed TOPlus enhances target-oriented hybrid selection relative to the original target-oriented prioritization (TOP) method.

(A) Distribution of ear weight (EW) and phenotypic similarity to the target variety among selected hybrids across five environments. Phenotypic similarity between selected hybrids and Zhengdan958 was quantified as the mean squared error (MSE) across all traits without yield-related traits for each hybrid. Significance was assessed by two-sided t-tests comparing hybrids selected by TOP and TOPlus within each environment. ns, $P \geq 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

(B) Distribution of EW among the candidate hybrids selected by TOP and TOPlus in the Hebei environment. Each method selected 25 hybrids; the union of selected hybrids is shown, with shared hybrids displayed only once. Yellow, red, and gray bars indicate hybrids selected only by TOP, only by TOPlus, and by both methods, respectively. The dashed line indicates the EW of the target.

(C) Relative multi-trait profiles of the best-performing hybrid selected by each method in the Hebei environment. Trait values were normalized by the corresponding phenotypes of the target variety such that the target is set to 1 for all traits, and values greater than 1 indicate trait performance exceeding the target. Dashed lines represent the multi-trait profile of the target after normalization. Trait abbreviations are as defined in Figure 2.

Figure 4. Environment-specific hybrid prioritization and functional gene contribution patterns inferred by TOPlus.

(A) Overlap of hybrids selected across five independent testing environments (Jilin, Liaoning, Beijing, Hebei, and Henan), based on predicted phenotypes.

(B) Heatmap of environment-specific contribution weights for informative functional genes integrated in TOPlus. Color gradients indicate the magnitude and direction of gene contributions (red, positive; blue, negative). Hierarchical clustering based on contribution profiles partitions genes into environmentally stable (red) and environmentally plastic (blue) classes.
(C) Enrichment patterns of environmentally stable and plastic genes among TOPlus-selected hybrids as a function of the number of environments in which a hybrid was selected. Enrichment ratios were calculated relative to the corresponding gene set for each category. Significance was assessed by two-sided paired t-tests comparing the enrichment ratios of environmentally stable and plastic genes within each hybrid-selection frequency category. ns, $P \geq 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.
(D) Gene Ontology (GO) enrichment analysis of environmentally stable and plastic gene classes based on contribution-weight clustering. GO enrichment was performed using TBtools-II v2.225 with the maize B73 RefGen v4 GO annotation. GO terms with $P \leq 0.05$ and an enrichment score ≥ 2 were retained.
(E) Additive effects on ear weight (EW) for haplotypes of the environmentally stable gene *UB2* and the environmentally plastic gene *ZmTPS14.1*.

Figure 5. Predictive validation and regional adaptation assessment of TOPlus in commercial maize hybrids.
(A) PCA of the training set (TrP; NCII; 569 hybrids), testing population 2 (TeP2; 81 hybrids), and testing population 3 (TeP3; 29 hybrids).
(B) Observed versus predicted ear weight (EW) of commercial hybrids evaluated in Xiangyang in 2023.
(C) Prediction accuracy of EW for commercial hybrids across seven environments. Prediction accuracy in (B) and (C) was quantified as the Pearson correlation coefficient between observed and predicted EW.
(D) Consistency between TOPlus-based regional suitability classifications and national variety approval records across the four major production regions.
(E) Performance of Denghai1810. Standardized EW improvement relative to the target is shown across nine years and four production zones. The solid line represents TOPlus-predicted EW improvement of Denghai1810, while the black dashed line indicates the baseline performance of

the target, and the gray dashed line denotes a 5% yield improvement over the target. Filled circles indicate environments (years and locations) where Denghai1810 met the criteria for similarity to the target, whereas open circles indicate environments where it did not.

PH
DTA
ZmTPS14.1 ...
DTS
KWPE
UB2 ...
EW
EH
Br2 ...

TOP+

Yield

Past
2015
2022
2024
Future

Training environments

Env.	Training set
1	Env1
2	Env2
3	Env3
...	...
n	EnvN

Step1 Environmental variable selection

Environmental factors

Selected factors

Step2 Cross-env. prediction

EnvN+1

Predictive module

Step3 Hypothesis-guided joint weight learning

Weight

W_{Phe1} W_{Phe2} ... W_{Phei}

Polygenicity

W_{Gene1} W_{Gene2} ... W_{Genej}

Pleiotropy

Step4 Similarity score-based hybrid prioritization

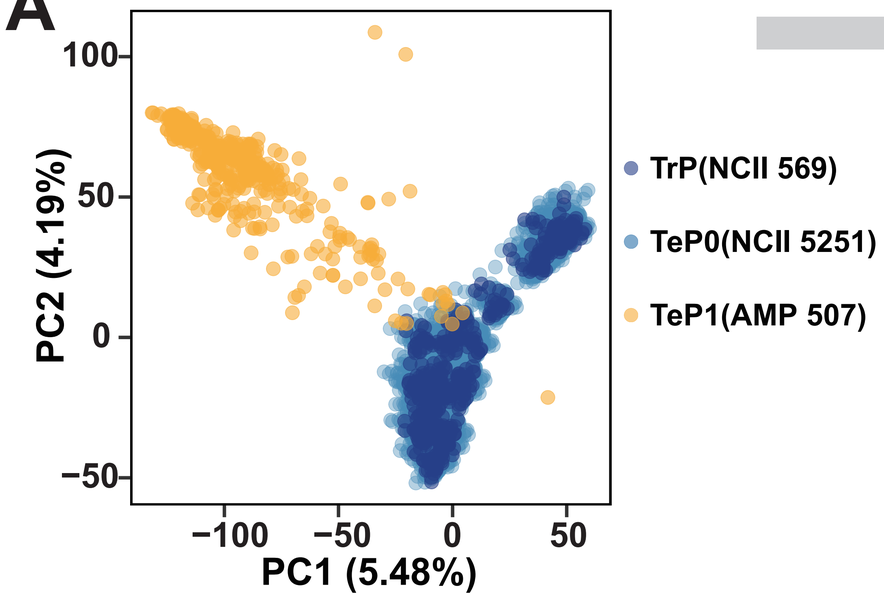
Similarity score

Hybrid ranking in target environment

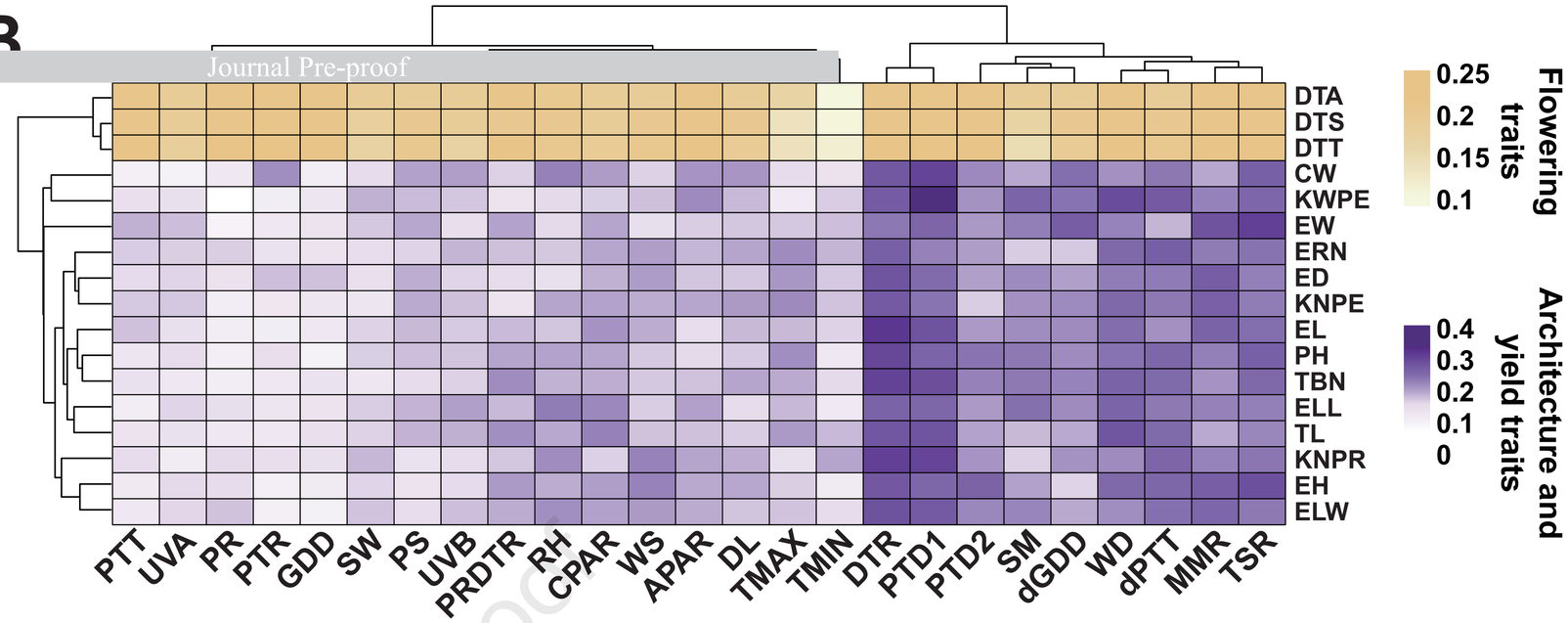
2 1 3

Selective module

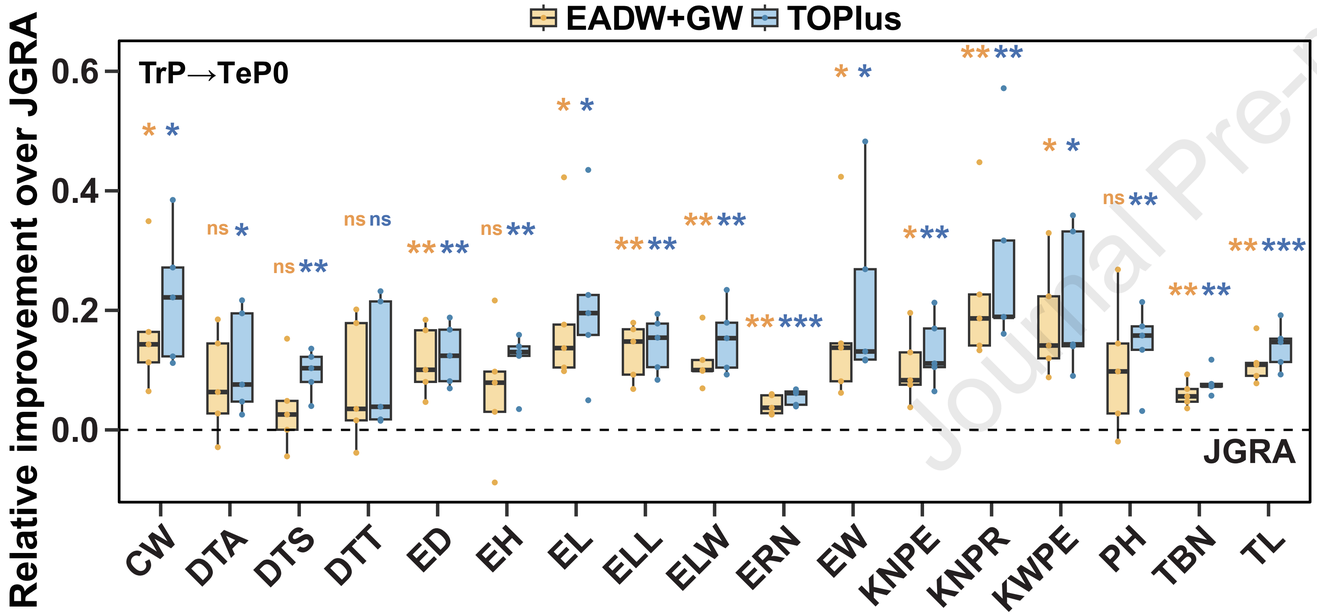
A



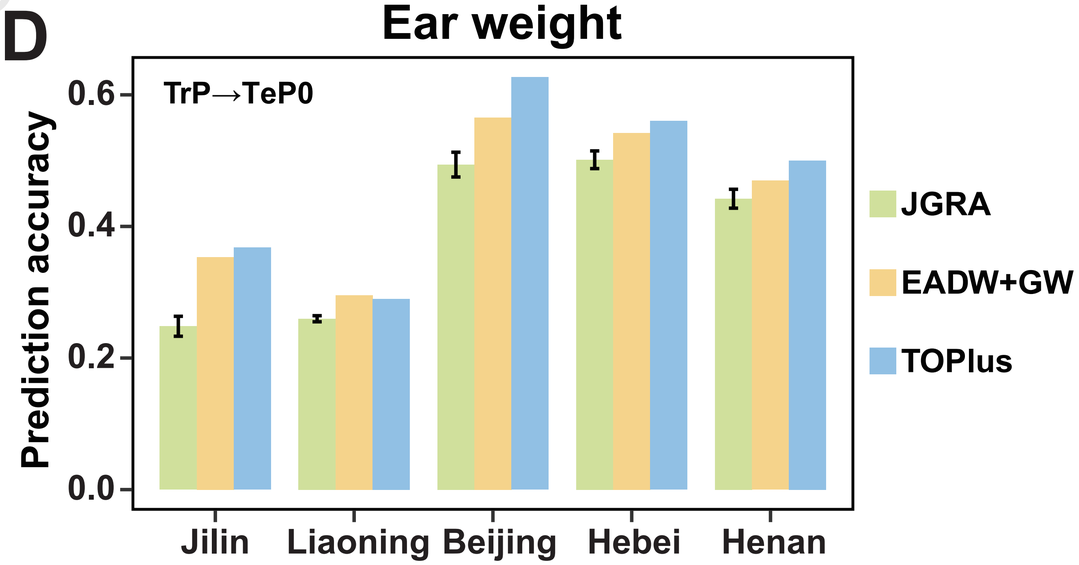
B



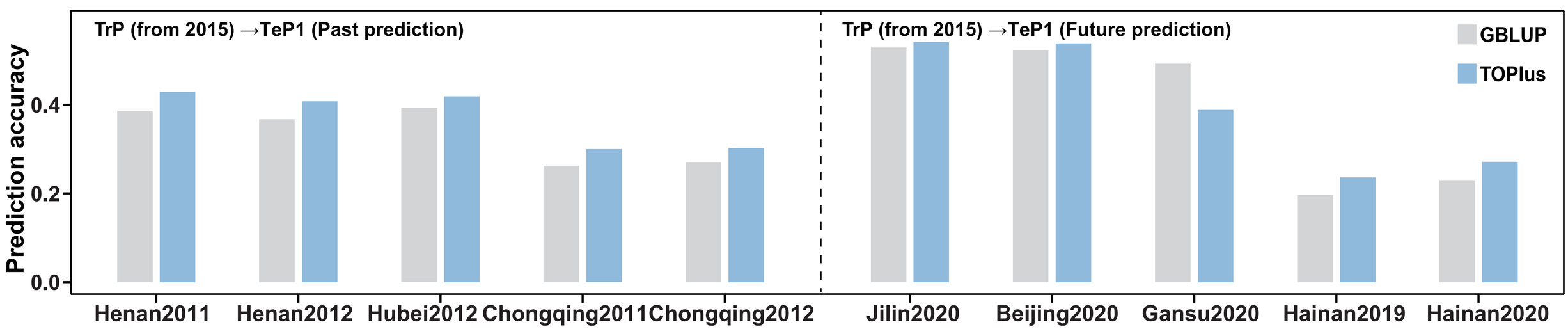
C

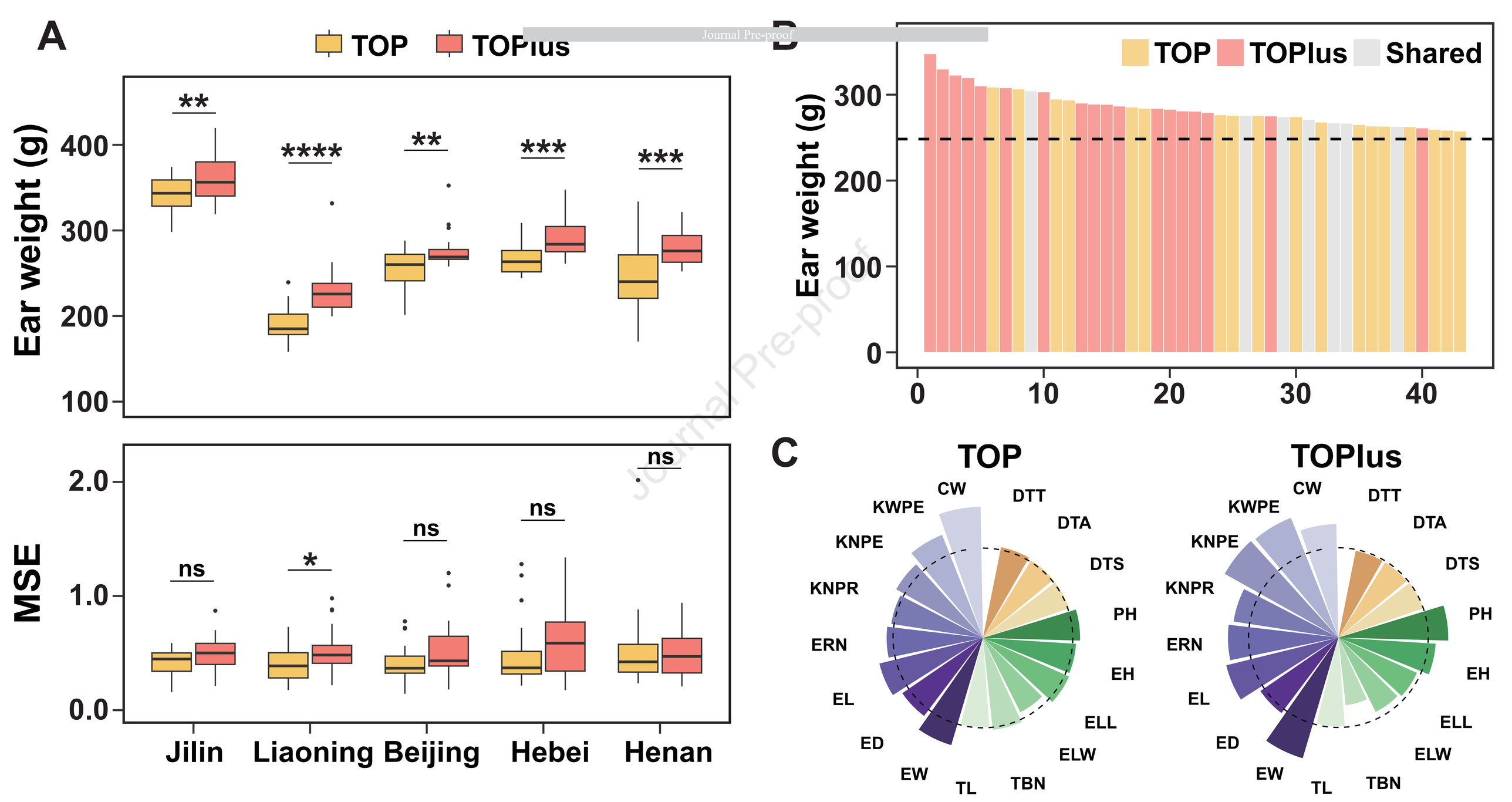


D

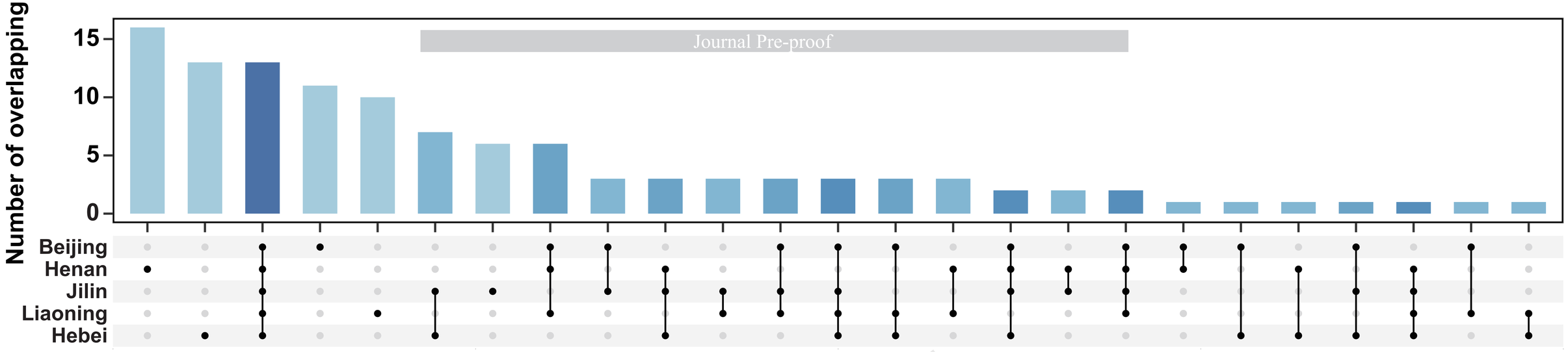


E

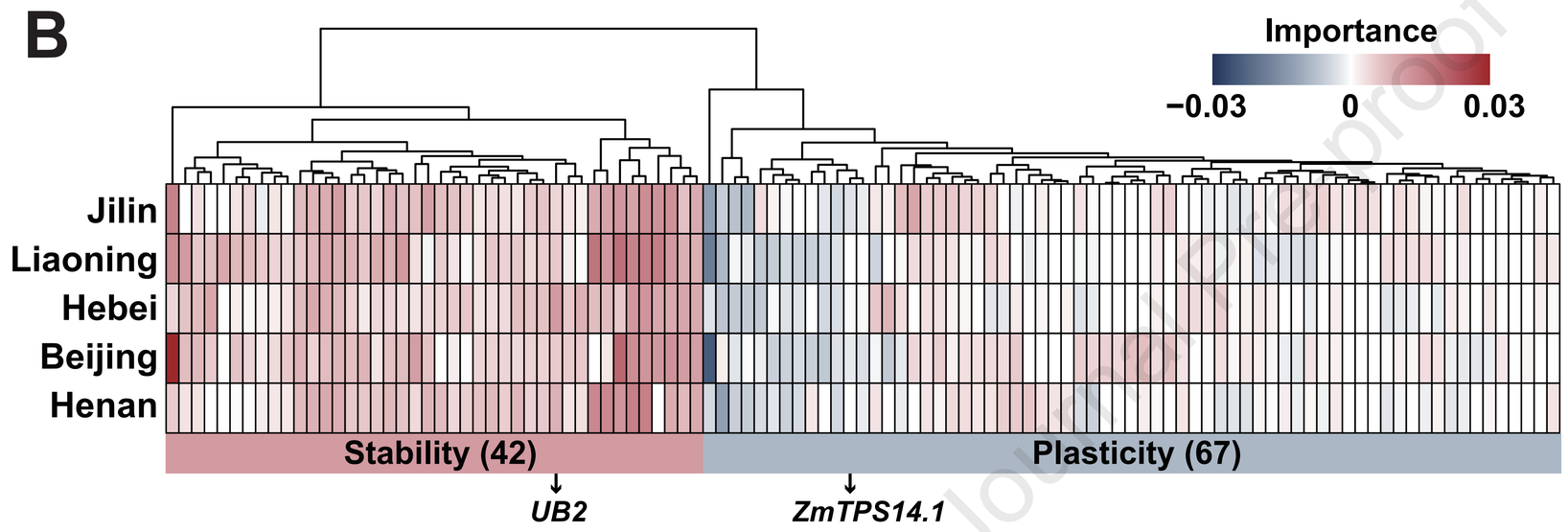




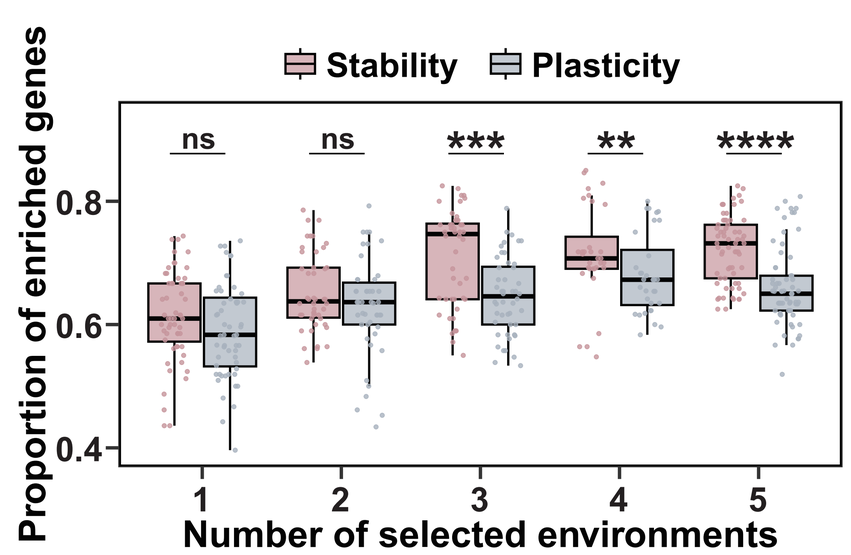
A



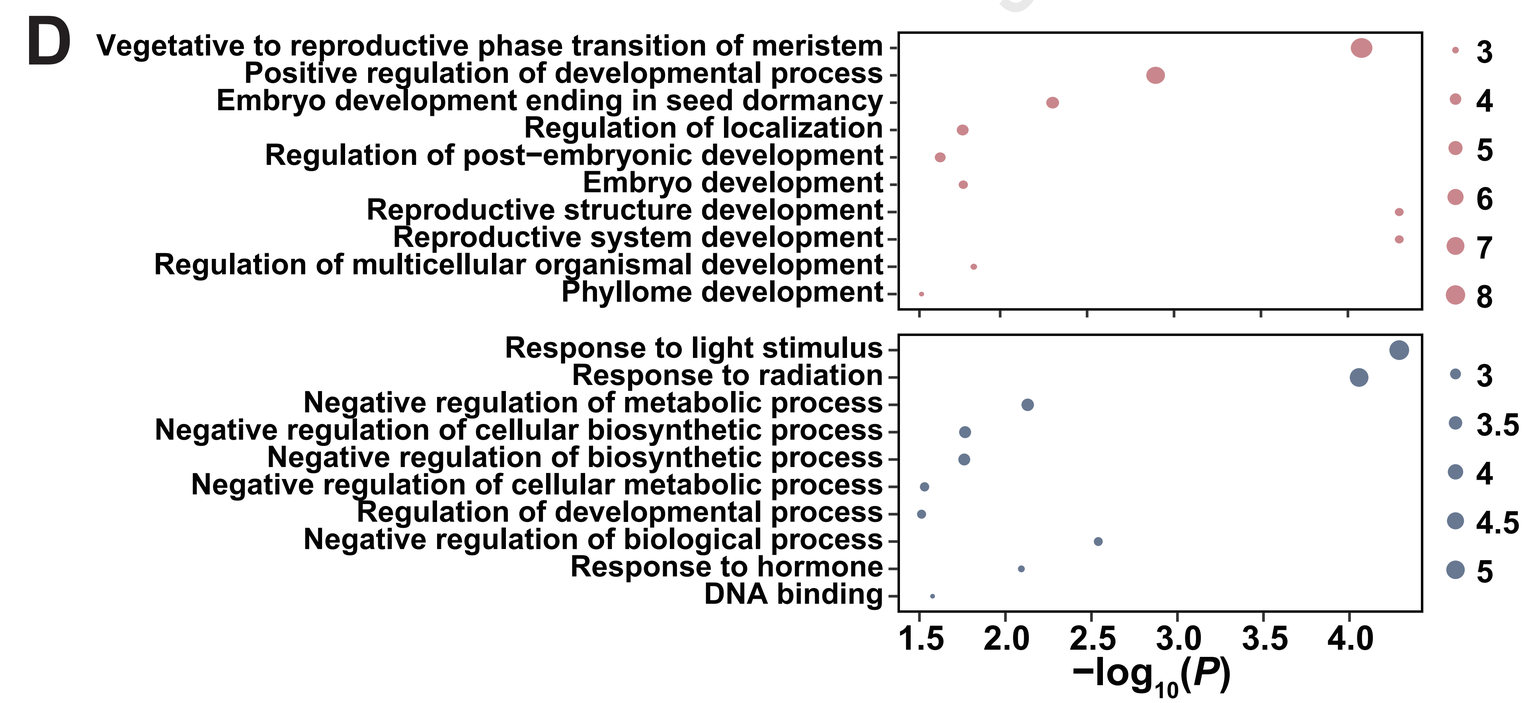
B



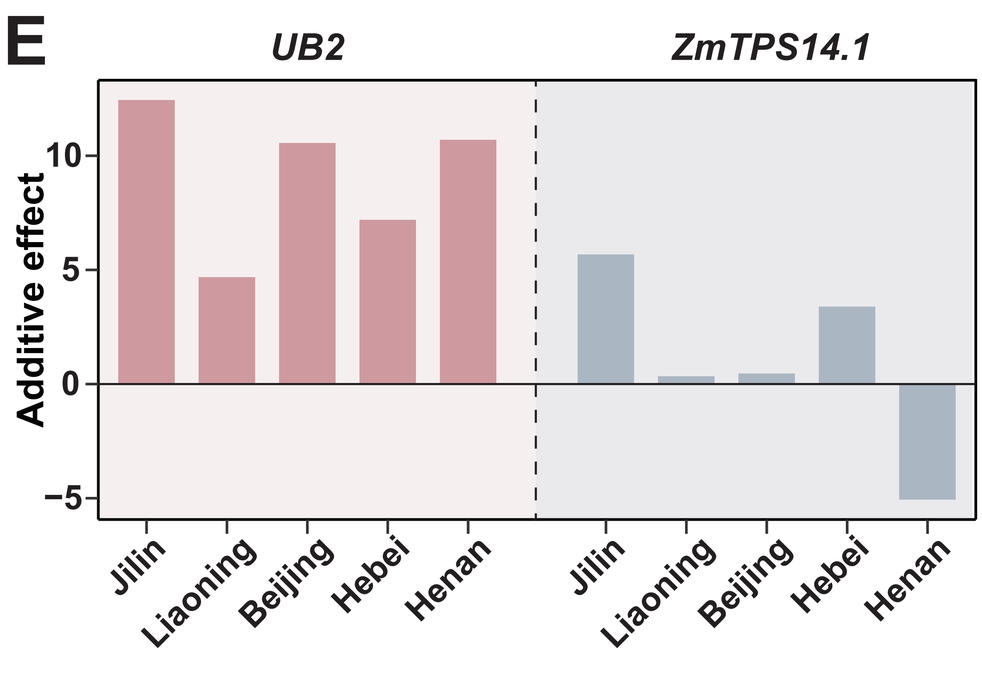
C



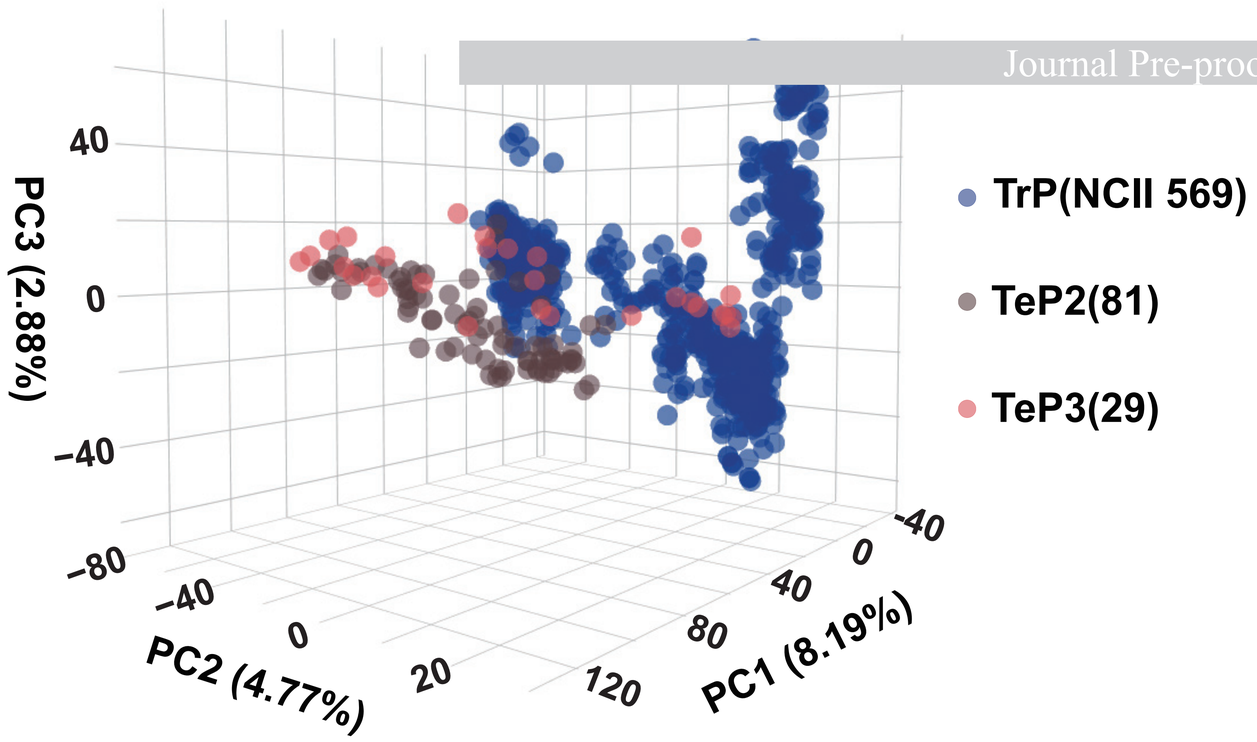
D



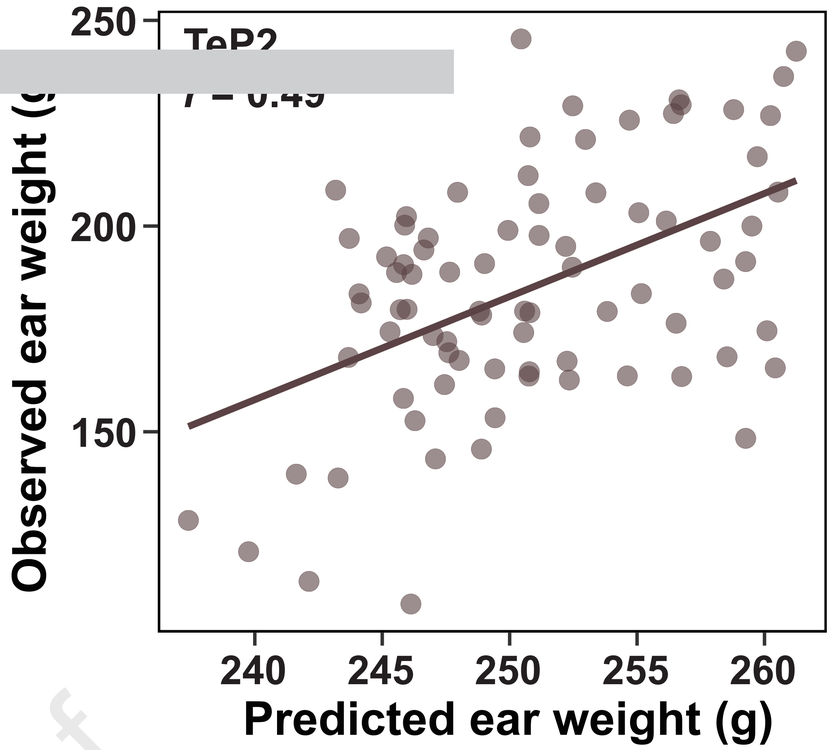
E



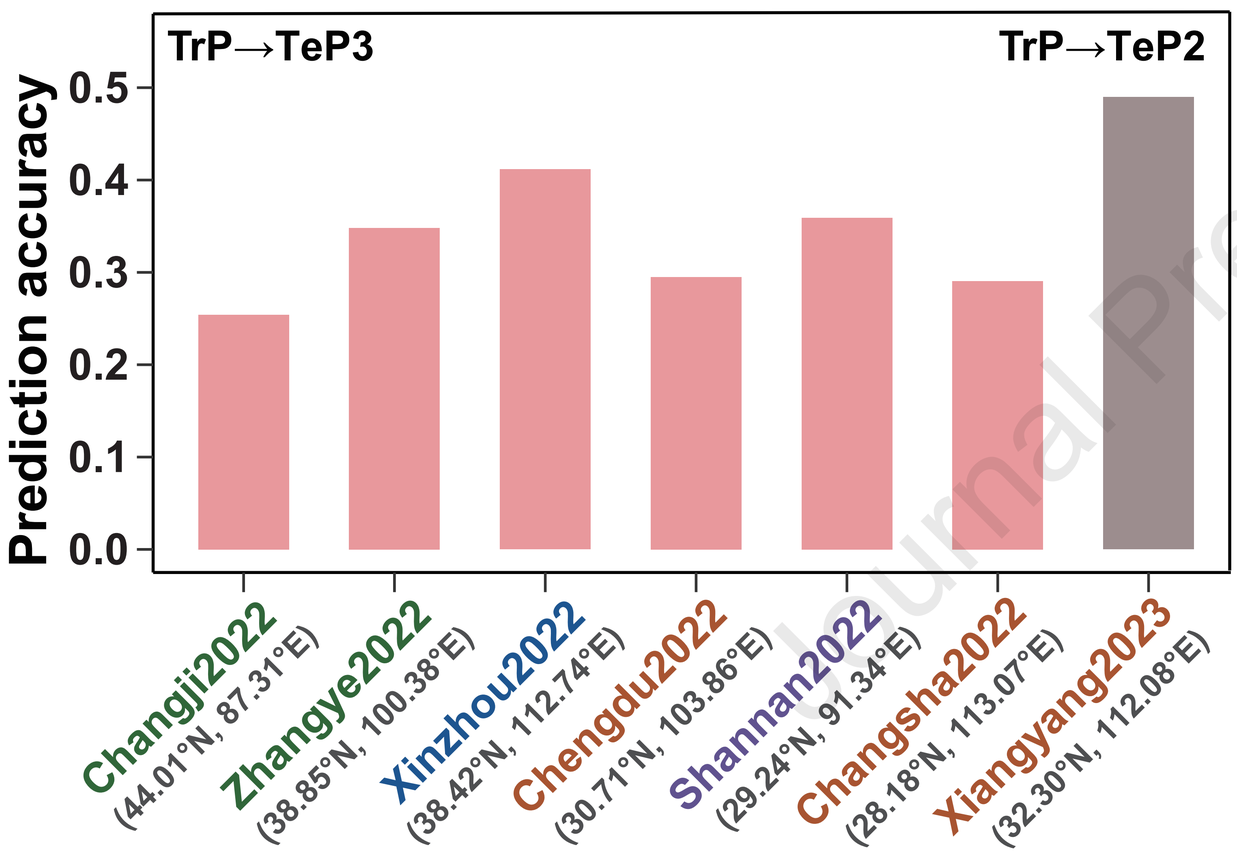
A



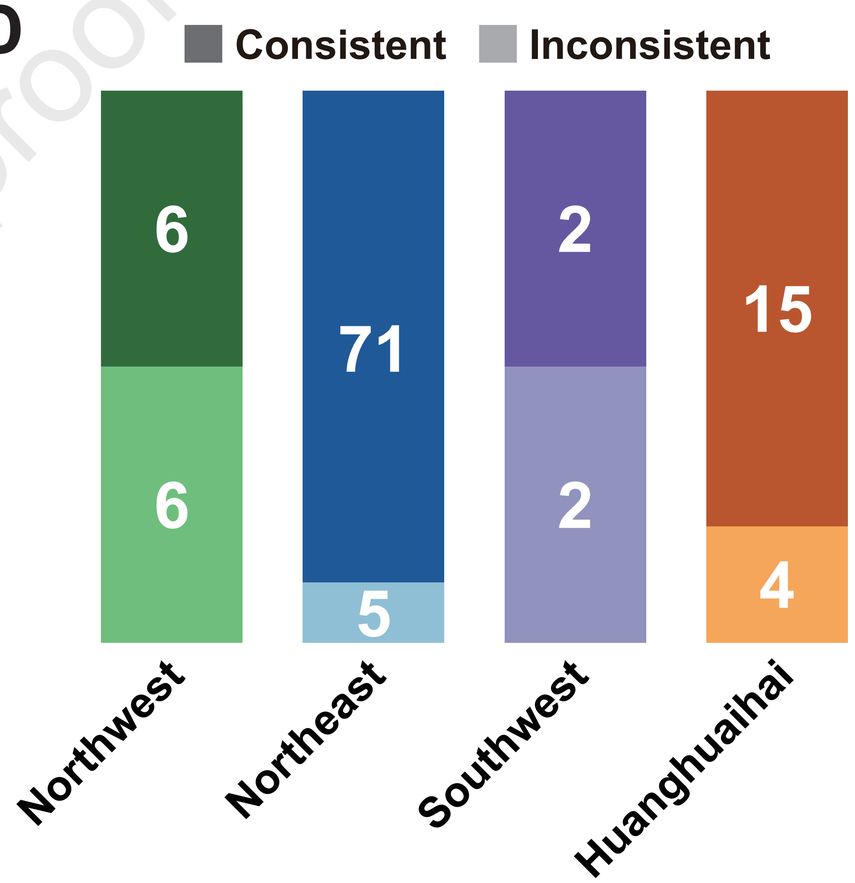
B



C



D



E