Li *et al. BMC Genomics* (2026) 27:487
https://doi.org/10.1186/s12864-026-12850-4

BMC Genomics

RESEARCH Open Access

Divergent genetic and phenotypic trajectories in China's maize hybrids

Guoji Li[1], Hanyue Mao[1,2,3], Sen Li[1] and Hai Wang[1,2,3*]

Abstract

Maize germplasm is fundamental for breeding progress, yet the genetic and phenotypic trajectories of China's modern commercial hybrids are poorly characterized. Through whole-genome resequencing of the hybrids themselves—which directly represent commercially deployed varieties— and multi-trait phenotyping of 61 elite hybrids, we delineated two major germplasm groups: a broad-based V1 group and a V2 group dominated by Pioneer ancestry, with moderate genetic differentiation (Fst = 0.0555). Temporal analysis showed V1 maintained stable genetic diversity ($\pi = 0.00488$), while V2 exhibited a marked diversity increase in its later phase (from $\pi = 0.00443$ to 0.00484), indicating intensified germplasm admixture. Phenotypically, V1 achieved successful decoupling of the tassel dry weight–leaf length trade-off, reducing tassel size while increasing leaf length—a notable optimization of plant architecture. In contrast, V2 showed phenotypic stability. Our findings clarify the distinct roles of the two pools: V1 represents a valuable template for breaking trade-offs to coordinately improve architecture, whereas the V2 pool shows that its genetic base has been substantially broadened without leading to significant phenotypic optimization. This suggests that selection on this germplasm under current breeding objectives may be approaching a plateau, indicating that future substantial improvements may require the introduction of novel genetic variation or a shift in selection paradigms.

Keywords Maize hybrids, Germplasm evolution, Phenotypic trade-offs, Genetic convergence, Trait decoupling

Introduction

Since its introduction from the New World to Europe by Columbus, Maize (*Zea mays L.*) has achieved global dissemination, emerging as the highest-yielding cereal crop worldwide and playing a pivotal role in global food security [1]. To meet the growing demand under population growth and limited arable land, substantial and sustainable increases in maize production are imperative [2, 3]. However, the recent slowdown in the growth rate of global maize productivity has raised concerns [4, 5]. As the world's second-largest producer, China plays a critical role in global maize supply. Therefore, continuous genetic improvement of maize hybrids is essential to enhance yield potential and adaptability.

Under the pressures of population growth, economic development, and limited arable land, enhancing the

*Correspondence:
Hai Wang
wanghai@cau.edu.cn
[1]State Key Laboratory of Maize Bio-breeding, Frontiers Science Center for Molecular Design Breeding, Joint International Research Laboratory of Crop Molecular Breeding, National Maize Improvement Center, College of Agronomy and Biotechnology, China Agricultural University, Beijing 100193, People's Republic of China
[2]Sanya Institute of China Agricultural University, Sanya 572025, People's Republic of China
[3]Hainan Yazhou Bay Seed Laboratory, Sanya 572025, People's Republic of China

yield of maize—as one of China's primary staple crops—has become increasingly imperative [2, 3]. China's modern maize breeding efforts span a century, beginning with Zhao Lianfang's pioneering work on hybrid breeding in 1925. Subsequent milestones include Li Jingxiong's development of early breakthrough hybrids like "Nongda 3" and "Nongda 4", the shift toward single-cross hybrids marked by "Xindan 1", the widespread adoption of historically significant hybrids such as "Danyu 13" and "Yedan 13", and finally the broad influence of landmark modern varieties "Zhengdan 958" and "Xianyu 335". This progression reflects continuous innovation in germplasm introduction and breeding techniques within China [6–8].

The past two decades have witnessed a dynamic shift in maize breeding and varietal deployment in China. The widespread adoption of hybrids derived from introduced elite germplasm, exemplified by the commercially successful Pioneer hybrid Xianyu 335, marked a significant influx of new genetic material. In more recent years, a parallel trend has been the commercialization of hybrids with less publicly traceable parental information. For representative varieties such as Weiyu 618, Zhengyuanyu 432, NK815, and Jiushenghe 2468, the parental lines are often not fully specified in available records, making complete pedigree tracing challenging [9–12]. This trend contributes to an increasingly complex and opaque genetic landscape for China's commercial maize hybrids. However, the precise population genetic architecture arising from these diverse germplasm inputs, the tempo and mode of their genomic interactions over time, and the consequent phenotypic trajectories remain systematically uncharted. Moreover, the perennial challenge of overcoming intrinsic phenotypic trade-offs continues to constrain coordinated trait improvement.

To address these questions, we performed an integrated population genomic and phenomic analysis of 61 widely cultivated Chinese maize hybrids, representing major varieties approved for commercial use between 1998 and 2021. This approach enabled a systematic investigation that revealed: (1) the population genetic structure and differentiation of China's primary maize germplasm pools, (2) how the genetic diversity within each pool and the genetic differentiation between them have changed over time, and (3) the long-term phenotypic trends and the dynamics of key trait correlations.

Materials and methods

Experimental materials

This study utilized 61 representative maize hybrids, commercially approved and widely cultivated between 1998 and 2021. These varieties were among the most extensively planted in the Northeast Spring, Huang-Huai-Hai Summer, and Southwest Spring maize regions of China, including: NK815, Anzao 10, Dafeng 30, Demeiya 1/2/3, Denghai 11/605, Dika 159/653, Dongdan 1331, Dongnong 254, Heyu 25, Heyu 187, Heyu 808, Jidan 7, Jiayu 538, Jinduyu 808, Jingke 968/999, Jingnongke 728, Jiushenghe 2468, Jundan 20, Liyu 16, Lianchuang 808, Liangyu 99, Linao 1, Longdan 76, Longyu 10, Longping 206, Nendan 18, Nongda 108/372, Nonghua 101, Suyu 29, Suiyu 19/20, Tiannong 9, Tongyu 808, Wansheng 68, Weike 702, Xianyu 335, Xiangyu 998, Xinxin 1, Yayu 889, Yufeng 303, Yunrui 21/408/8/88/999, Zhengda 999, Zhengdan 958, Zhengyuanyu 432, Zhongdan 808/909, Zhongke 11, Zhongkeyu 505, Zhongnongda 787, Qiandan 15, and Weiyu 618. Field trials were conducted at the Shangzhuang Experimental Station of China Agricultural University during the spring seasons of 2022 and 2025. The 2022 trial was established in two adjacent experimental blocks, while the 2025 trial was located in the southern district. For all cultivars, 15 agronomic traits—encompassing both in-growth and post-harvest ear/kernel characteristics—were evaluated. The final dataset comprised observations from 23 plants per cultivar in 2022 and 10 plants per cultivar in 2025.

Statistics of initial approval year and approval number

The approval numbers and the earliest approval years for the 61 hybrids were retrieved from Baidu Baike. In this study, the approval level and crop type within the approval numbers are presented in a concise English abbreviation format. This abbreviation system is structured as follows: the initial letter(s) denote the approval authority-'GSY' for Guoshenyu (National Approval), 'LSY' for Liaoshenyu (Liaoning Approval), 'JSY' for Jishenyu (Jilin Approval), 'HSY' for Heishenyu (Heilongjiang Approval), 'JiSY' for Jishenyu (Hebei Approval), 'QSY' for Qianshenyu (Guizhou Approval), 'GuSY' for Guishenyu (Guangxi Approval), 'DSYM' for Dianshenyumi (Yunnan Approval), 'NSY' for Ningshenyu (Ningxia Approval), 'WPS' for Wanpinshen (Anhui Approval), 'JinSY' for Jinshenyu (Shanxi Approval), 'YSY' for Yushenyu (Henan Approval). The subsequent approval year and serial number remain unchanged. This set of abbreviations corresponds precisely to the official Chinese approval codes. We recorded the earliest approval year for each hybrid. For hybrids with multiple approval numbers, the national approval number was prioritized; if no national number existed, the earliest provincial approval number was recorded.

Phenotypic measurement methods

Fifteen phenotypic indicators were measured for maize plants using the following methods: Plant architecture-related traits: Plant height (PH) was measured as the vertical distance from ground level to the tip of the tassel using a measuring staff; ear height (EH) was measured from the ground to the node of the primary ear.

The maximum expanded leaf length (LL) and width (LW) above the primary ear were measured with a ruler. The leaf angle (LA) between the leaf above the ear and the main stem was measured with a protractor. The number of leaves above the primary ear (LAE) was counted manually. Ear and yield-related traits: The primary ear was selected. Kernel row number (KRN) was counted manually. Ear length (EL) was measured vertically from the base to the tip, and ear width (EW) was measured at the maximum diameter using a ruler. Hundred-kernel weight (HKW) was determined by weighing 100 air-dried kernels using an electronic balance. Kernel traits: Typical ears from each plot at the Shangzhuang site were selected. Six middle-position kernels (3 per ear) were measured. Kernel length (KL), width (KW), and thickness (KT) were measured using a digital caliper (accuracy 0.01 mm). Tassel traits: Tassel length (TL) was measured linearly from the base to the tip of the main axis. Tassel dry weight (TDW) was measured using an electronic balance after samples were oven-dried to a constant weight.

Analysis of phenotypic trends and correlations

Phenotypic data analysis was performed in R (version 4.3.1) (https://www.R-project.org/) using the followin g packages: ggplot2 (v3.5.2) [13], GGally (v2.2.1) [14], stringr (v1.5.1) [15], rlang (v1.1.4) [16], dplyr (v1.1.4) [17], tidyr (v1.3.1) [18], readr (v2.1.5) [19], lme4 (v1.1-35) [20], and lmerTest (v3.1-3) [21].

To account for environmental variation across planting years (2022 and 2025), we estimated best linear unbiased predictions (BLUP) for each hybrid using mixed linear models. For each phenotypic trait, we fitted the following model using the 'lmer' function from the lme4 package:

'Trait ~ Planting_Year + (1 | Hybrid)'.

where 'Planting_Year' was treated as a fixed effect to account for year-to-year environmental variation, and `Hybrid` was treated as a random effect. The model was fitted using restricted maximum likelihood (REML). The random effect estimates (BLUPs) for each hybrid were extracted and used as the adjusted genotypic values for all downstream analyses.

To examine phenotypic changes over the breeding period, we regressed the BLUP values of each trait against the hybrid approval year (ranging from 1998 to 2021). For each trait, we fitted a linear regression model using the 'lm' function in R:

'BLUP ~ Approval_year'.

where 'BLUP' represents the BLUP-adjusted genotypic value of a given hybrid, and 'Approval_year' denotes the earliest approval year of that hybrid. For each trait, we extracted the regression coefficient (slope) as an indicator of the rate of phenotypic change per year, the Pearson correlation coefficient (r) between BLUP value and approval year, and the associated p-value to evaluate statistical significance. The analysis was performed separately for the two subpopulations (V1 and V2) identified by population structure analysis to account for potential differences in breeding trajectories. Temporal trends were visualized as scatter plots with fitted regression lines and 95% confidence intervals using ggplot2, where each point represents an individual hybrid, and the x-axis represents the approval year. For each plot, the Pearson correlation coefficient (r) and p-value were displayed as annotations.

After obtaining BLUP values for each hybrid, traits were categorized into two groups: plant architecture-related traits (PH, EH, LL, LW, LA, LAE, TL, TDW) and yield-related traits (EL, EW, KRN, HKW, KL, KW, KT). Pairwise Pearson correlation coefficients (r) and their statistical significance (p-values) were computed within each trait group using the BLUP values. Correlation matrices were visualized as combined scatter-correlation plots using the 'ggpairs' function from the GGally package, with correlation coefficients displayed in the upper triangle (accompanied by significance levels: *** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$, ns for $p \geq 0.05$), scatter plots with fitted regression lines in the lower triangle, and density distributions on the diagonal.

To minimize potential bias caused by geo-adaptive differences between the two germplasm groups, we restricted the temporal trend analysis (Sect. 3.5) to varieties that are exclusively adapted to the Northern Spring Corn Region, the Huang-Huai-Hai Summer Corn Region, or the Northwest Irrigated Corn Region. Specifically, a variety was included if it was suitable for cultivation in at least one of these three regions and was not suitable for cultivation in the Southwestern Mountain Corn Region or the Southern Hilly Corn Region. This definition excludes varieties with broad ecological adaptability that span both northern and southern regions, ensuring that the selected subset shares similar ecological conditions (e.g., accumulated temperature, growing season length, and photoperiod). This approach minimizes the confounding effects of inherent adaptive divergence between groups and ensures that observed phenotype–year relationships reflect genuine breeding trends within the same target ecoregion. Following this definition, the selected subset comprised 26 V1 varieties and 16 V2 varieties.

DNA extraction, library construction, and sequencing

At the seedling stage, healthy plants per hybrid were selected, and fresh leaf samples (6–7 cm) were collected, flash-frozen in liquid nitrogen, and stored at -80 °C. Genomic DNA was extracted using the CTAB method [22, 23]. The specific protocol was as follows: (1) Add 7.5 mL CTAB extraction buffer (containing 15 μL β-mercaptoethanol) to a 50 mL centrifuge tube. Grind frozen leaf tissue in a pre-cooled mortar with liquid

nitrogen to a fine powder and transfer to the lysis buffer. (2) Incubate at 65 °C for 30 min, vortexing every 10 min. After equilibrating to room temperature, add chloroform to a final volume of 15 mL, mix thoroughly, and centrifuge at 4000 rpm for 20 min. (3) Transfer 6–7 mL of the upper aqueous phase to a new tube. After a second chloroform purification, adjust the volume to 10 mL and perform phase separation under the same centrifugation conditions. (4) Transfer 750 µL of the upper aqueous phase to a 2 mL microcentrifuge tube, add an equal volume of isopropanol to precipitate nucleic acids, incubate at -20 °C for 30 min, then centrifuge at 12,000 rpm for 10 min to pellet the DNA. Wash the pellet twice with 75% ethanol (12,000 rpm, 10 min each), air-dry at room temperature, and dissolve the DNA in 50 µL of ultra-pure water. (5) Quality control: Measure DNA concentration using 2 µL. Assess DNA quality by 1% agarose gel electrophoresis (mixing 4 µL DNA with 1 µL 6× loading buffer). Samples passing quality control were used for library construction and paired-end sequencing (2 × 150 bp) on the MGI DNBSEQ-T7 platform.

Genomic data basic analysis

Approximately 40 Gb of sequencing data was obtained per hybrid. Bioinformatics analysis was performed based on the maize Zea mays B73 RefGen_v5 reference genome (Zm-B73-REFERENCE-NAM-5.0), which was downloaded from Ensembl Plants (https://plants.ensembl.org) at the FTP address: https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-62/fasta/zea_mays/dna/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.dna.toplevel.fa.gz. The analysis pipeline was deployed on our laboratory server and the High-Performance Computing cluster of China Agricultural University. Specific steps included: (1) Quality trimming of raw FASTQ files using Sickle (v1.33) [24], truncating low-quality bases (Phred score < 20) and removing short sequences (< 50 bp). (2) Alignment of sequencing reads to the reference genome using the BWA-MEM algorithm (v0.7.17-r1188) [25], generating SAM files. (3) Format conversion and coordinate sorting of SAM files to BAM files using SAMtools (v1.12) [26]. (4) Marking PCR duplicates and rebuilding BAM indexes using Picard (v3.0.0) MarkDuplicates and BuildBamIndex. (5) Variant calling using GATK (v4.4.0.0) [27] HaplotypeCaller module with GVCF output, followed by joint genotyping using GenotypeGVCFs to produce a VCF file. (6) Extraction of SNP sets using GATK SelectVariants. To obtain high-quality variants, we filtered out those meeting any of the following criteria: QUAL < 30.0, MQ < 40.0, FS > 60.0, SOR > 3.0, MQRankSum < -12.5, ReadPosRankSum < -8.0. (7) Standard population genetics filtering using PLINK (v1.90b6.20) [28] to remove variants with: minor allele frequency (MAF) < 0.05, genotype missing rate > 0.1, Hardy-Weinberg equilibrium $P < 1 \times 10^{-6}$, and sample missing rate > 0.2. The final dataset was converted to binary BED format for downstream analysis, resulting in 39,179,346 high-quality SNPs. The raw sequence data generated in this study have been deposited in the Genome Sequence Archive (GSA) at the National Genomics Data Center (NGDC) under the accession number CRA038291, which is publicly accessible at https://ngdc.cncb.ac.cn/gsa/browse/CRA038291[29, 30].

Population genetic structure, principal component analysis and kinship analysis

Based on the binary BED format genotype data, Bayesian clustering analysis was performed using ADMIXTURE (v1.3.0) [31] for K = 1 to K = 10. Population structure visualization was achieved by integrating numpy (v1.21.6) [32], pandas (v1.3.5) (https://doi.org/10.5281/zenodo.17229934), matplotlib (v3.5.3) [33] with its pyplot/gridspec modules, and the hierarchical clustering function from scipy (v1.7.3) [34] in a Python3 (v3.6.8) (https://docs.python.org/3/) environment. The cross-validation error curve for selecting the optimal K was plotted using seaborn (v0.12.2) [35] and glob (v0.7) (https://docs.python.org/3/). A genetic relationship matrix was calculated using PLINK, and the first 20 principal components were extracted (eigenvec file). A two-dimensional PCA scatter plot was constructed in R (v4.3.1) (https://www.R-project.org/) using ggplot2 (v3.5.2), ggrepel (v0.9.6) [36], and RColorBrewer (v1.1.3) [37], with subgroups V1 and V2 color-coded. A kinship matrix was calculated by PLINK using the genome-wide kinship coefficient (genome file). A heatmap was generated using the pheatmap (v1.0.13) [38], reshape2 (v1.4.4) [39], grid (v4.3.1) (https://www.R-project.org/), and RColorBrewer (v1.1.3) packages in R, with subgroups V1 and V2 color-coded.

All population genetic analyses were performed using diploid genotype calls that retained the natural heterozygous state of the F1 maize hybrids. No modification, filtering, or conversion of heterozygous sites was performed; heterozygous genotypes were directly retained in the variant calling format (VCF) and processed using standard parameters for diploid data in all software tools. For PCA, principal component analysis was performed using PLINK (v1.90b6.20) with the '--pca' option, which internally encodes genotypes as the number of alternative alleles (0 = homozygous reference, 1 = heterozygous, 2 = homozygous alternative). For ADMIXTURE, population structure analysis was performed using ADMIXTURE (v1.3.0) with the binary BED format generated by PLINK, which retains heterozygous genotypes (coded as 2 in the .bed file), and ADMIXTURE was run under its default diploid mode. For kinship analysis, PLINK with the '--genome' option was used, which estimates IBD

coefficients using the observed allele frequencies and accounts for heterozygosity in the calculation.

Fixation index (Fst) and nucleotide diversity (π) analysis

We employed population genetics methods to quantify genomic differentiation and nucleotide diversity. First, based on the previously obtained BED format genotype file, PLINK (v1.9) was used for format conversion to generate a standard VCF file. Subsequently, VCFtools (v0.1.16) [40] was used to calculate Fst based on the Weir-Cockerham method, assessing the level of genetic differentiation between groups. To visually represent genomic differentiation patterns, a genome-wide Fst Manhattan plot was constructed using the Python3 environment (pandas, numpy libraries) with matplotlib (v3.5.3) for visualization, specifically using the matplotlib.pyplot module and a LinearSegmentedColormap for custom color mapping. For nucleotide diversity analysis, the π statistic was calculated using VCFtools to characterize the level of genetic polymorphism within groups. Finally, the distribution of π values was statistically analyzed and visualized using the R platform (tidyverse 2.0.0, patchwork 1.3.0, RColorBrewer 1.1.3, and scales 1.4.0 packages) [41, 42], integrating multi-dimensional visual elements for comprehensive result presentation.

For Fst analysis, we used VCFtools (v0.1.16) with the `--weir-fst-pop` option, which directly accounts for heterozygosity by incorporating observed heterozygote frequencies into the variance components. The interpretation of Fst values followed the criteria originally proposed by Wright (1978) [43] and subsequently adopted by Hartl and Clark (1997) [44] and Suarez-Kurtz and de Araújo (2022) [45], where Fst < 0.05 indicates little genetic differentiation, 0.05–0.15 indicates moderate differentiation, 0.15–0.25 indicates great differentiation, and > 0.25 indicates very great differentiation. For nucleotide diversity (π) analysis, we used VCFtools with the `--window-pi` option, which calculates π as the average number of pairwise differences per site across all individuals in the population, naturally incorporating heterozygous sites into the calculation.

Results

Population genetic structure delineates two major germplasm groups

Population structure analysis of the 61 major Chinese maize hybrids using ADMIXTURE showed the lowest cross-validation error at K = 2 (CV error = 0.43785), indicating that the genetic structure is most significantly divided into two ancestral groups (Fig. 1-A).

This division was further supported by ancestry coefficients: 37.7% (23 hybrids) showed high genetic differentiation with a Q-value exceeding 0.9 for one ancestral group. For instance, 11 hybrids, including Zhengdan 958, Yayu 889, and Zhongke 11, had Q-values > 0.9999 for ancestral group I (Cluster I); six hybrids, including Xianyu 335, Xiangyu 998, and Dafeng 30, were strongly biased towards ancestral group II (Cluster II). The remaining 62.3% (38 hybrids) exhibited varying degrees of genetic admixture (Q-values between 0.5 and 0.9). Examples include Zhongnongda 787 (Cluster I Q = 0.682), Heyu 25 (Cluster I Q = 0.687), and Jiayu 538 (Cluster II Q = 0.687), showing moderate admixture; hybrids like Qiandan 15, Zhengyuanyu 432, and Jingke 968 displayed nearly equal genetic backgrounds (Q-value ~ 0.5), suggesting historical gene flow between subgroups (Fig. 1-B). Using a Q-value threshold of 0.5, the population was divided into V1 (corresponding to Cluster I, 43 hybrids) and V2 (corresponding to Cluster II, 18 hybrids) for subsequent analysis.

For analysis, we compiled information on 61 hybrids, including name, parental lines, population classification, year of first registration, and primary registration number, sorted by year of first registration (Supplementary Table 1). The V1 group is based primarily on Huanggai lines (e.g., Chang 7 – 2) and improved Reid germplasm (e.g., Zheng 58), incorporating resources from tropical, European, and U.S. backgrounds. The V2 group originated with Xianyu 335, registered in 2004. Parents in this group mainly trace to Pioneer germplasms PH4CV and PH6WC, as well as materials derived from them, such as the X-line X1132X (PH4CV × PH6WC). V2 also includes parents like JCD122BRDan15 - a parent of Zhengyuanyu 432—which was developed from a double-cross hybrid using Xianyu 696 (female parent PH6WC, male parent PHB1M) as the female and Dika 517 as the male, followed by two generations of selfing. It is noteworthy that the clarity of parental origins varied over time. For varieties registered particularly after 2016, such as Wansheng 68, Heyu 187, and Jingke 999, one or both parents were often recorded as "company-bred lines" or "materials introduced from abroad," making complete pedigree tracing more challenging compared to earlier releases.

Germplasm divergence, spatial distribution, and genetic convergence

Combining population structure (K = 2) with varietal planting regions, we generated a composite pie chart (Fig. 2-A) integrating group classification with all suitable cultivation areas.

The percentages represent the proportion of varieties within each group adapted to each region, calculated as the number of varieties in V1 or V2 suitable for a given region divided by the total number of varieties in the corresponding group (43 for V1, 18 for V2). Due to the broad adaptability of some hybrids, a single variety may be suitable for multiple ecological regions, resulting in percentages that sum to more than 100% across regions.

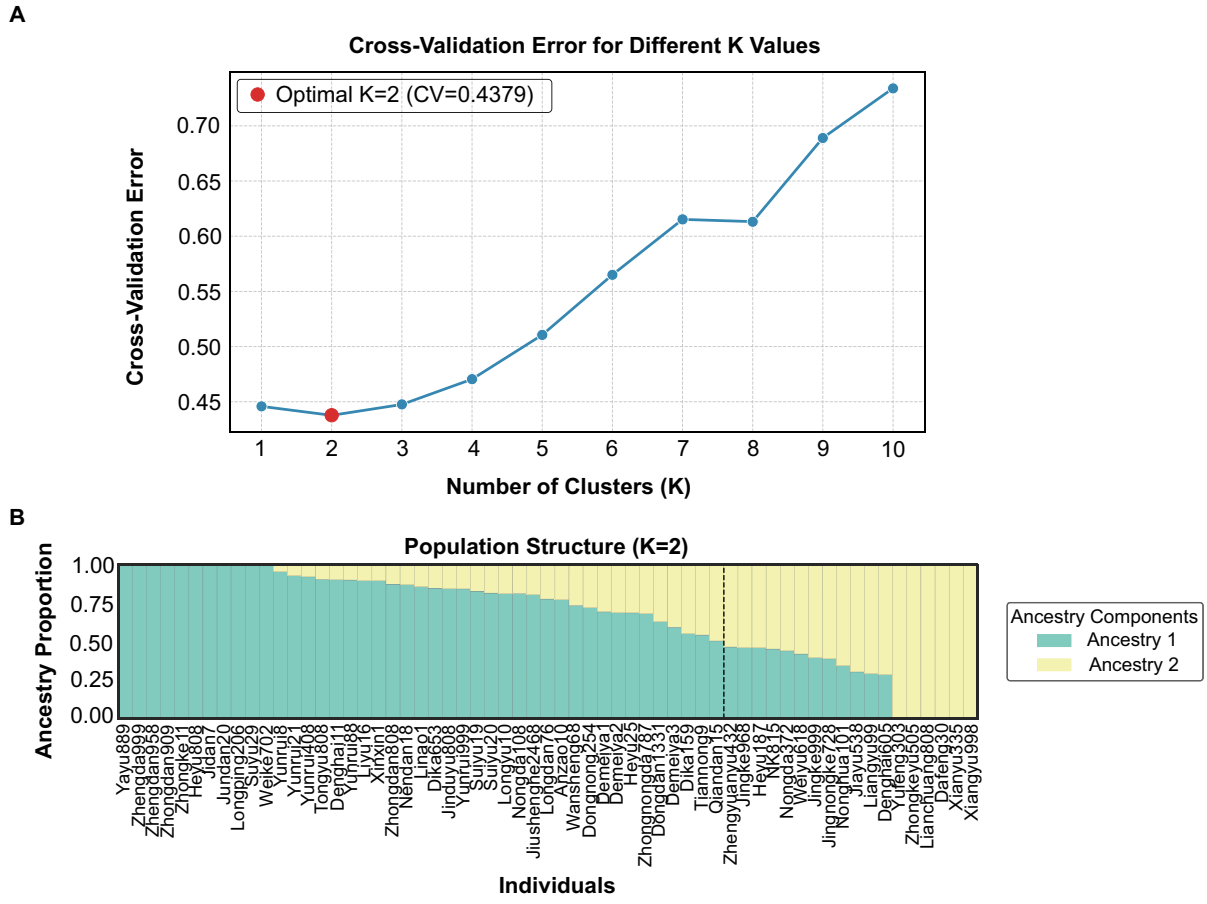

Fig. 1 Population genetic structure analysis of maize hybrids. **A** Determination of optimal ancestral population number (K) by cross-validation. **B** Genetic structure distribution of 61 varieties at K = 2: green indicates Cluster I, yellow indicates Cluster II

As shown in Fig. 2-A, the V1 group demonstrates broad ecological adaptability. It accounts for 51.2% (22 varieties) in the Northern Spring Corn Region, 44.2% (19 varieties) in the Huang-Huai-Hai Summer Corn Region, and 16.3%, 32.6%, and 9.3% in the Northwest, Southwest, and Southern regions, respectively. Representative varieties such as Nongda 108 and Liyu 16 are adapted to multiple major production regions. In contrast, the V2 group (18 varieties), characterized by its prevalent Pioneer (PH4CV/PH6WC) ancestry, exhibits a more concentrated distribution. It is predominantly planted in the Huang-Huai-Hai region (88.9%, 16 varieties) and the Northern Spring Corn Region (55.6%, 10 varieties), with minimal presence in the Northwest, Southwest, and Southern regions. This distribution aligns with the ecological adaptation of its core germplasm, which is optimized for conditions similar to the U.S. Corn Belt.

In the PC1-PC2 genetic space (cumulative variance explained: 27.0%), the V1 (dark green) and V2 (dark orange) groups form a V-shaped distribution, reflecting their distinct breeding trajectories over time (Fig. 2-B). Early V1 varieties (approved 1998–2009) are scattered across the genetic space: for example, Nongda 108 (0.049, 0.122) and Zhengdan 958 (0.195, -0.333) lie at opposite ends of the PC2 axis, illustrating the diverse use of germplasm prior to the widespread adoption of Pioneer materials. The registration of Xianyu 335 (-0.282, -0.015) in 2004 introduced a variety positioned at the negative extreme of PC1. Subsequent breeding efforts led to an expansion of genetic resources and diversification of parental origins, which gradually blurred the genetic boundaries between the two major germplasm groups. This is exemplified by varieties such as Jingke 968 (-0.088, -0.181) and Denghai 605 (-0.148, 0.023), which reflect recombination between Pioneer and local (or other exotic) germplasm. In varieties registered after 2016, this blurring of genetic boundaries became more pronounced, as seen in the positions of Weiyu 618 (-0.105, -0.060) and Zhengyuanyu 432 (-0.089, 0.056).

Based on kinship parameters (PI_HAT and DST), the populations were further divided into five subgroups (Fig. 2-C). These include the Core Pioneer-derived subgroup (Xianyu 335-Dafeng 30, PI_HAT = 0.73, DST = 0.93) and the Pioneer-introgressed subgroup (e.g.,

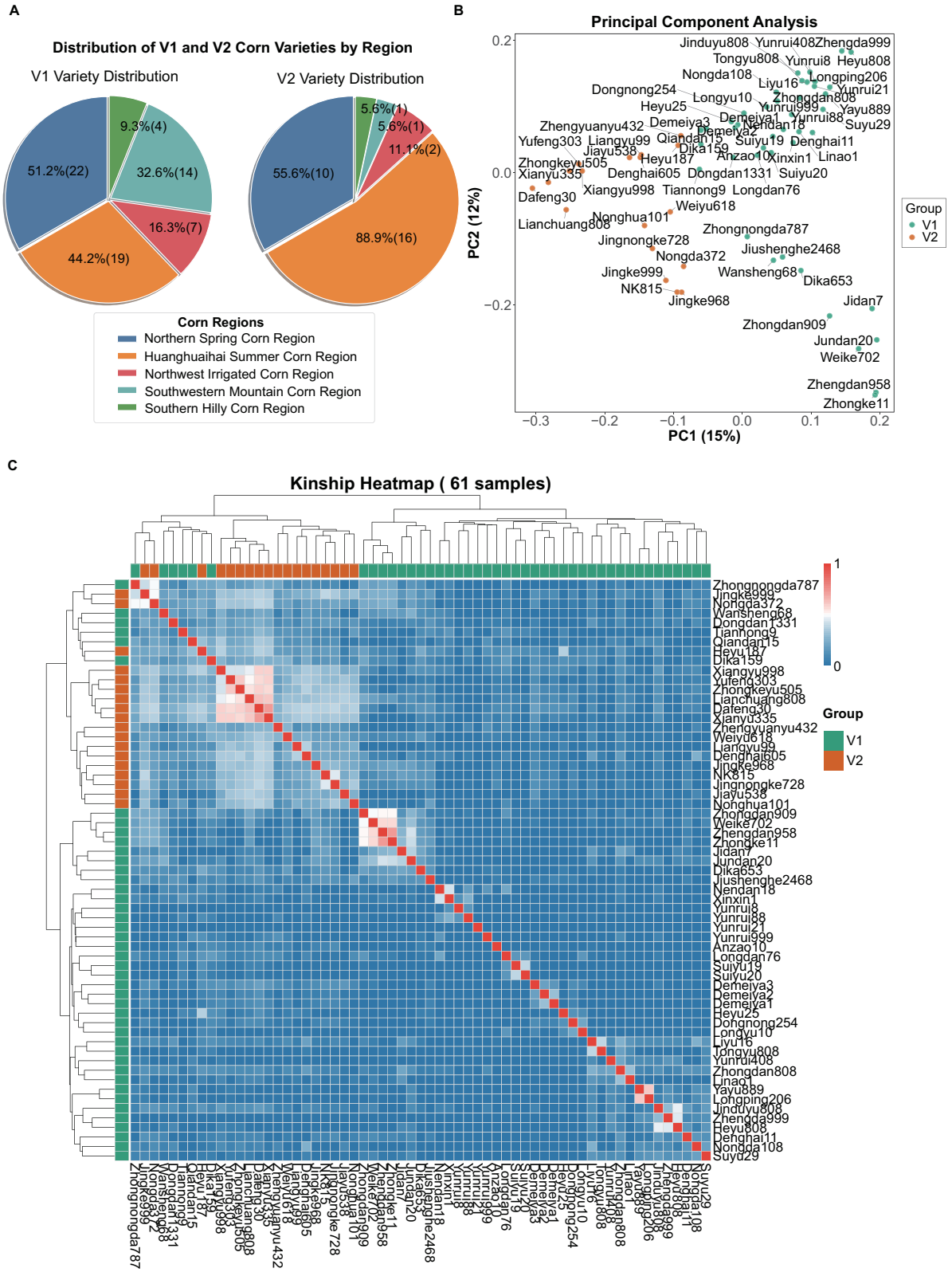

Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Population differentiation and admixture characteristics of maize hybrid germplasm resources. **A** Geographical distribution of the two major genetic groups (V1 and V2) across China's principal maize production regions. Pie charts show the proportion of varieties within each group that are adapted to each region, calculated as the number of varieties in V1 or V2 adapted to the region divided by the total number of varieties within the corresponding group (43 for V1, 18 for V2). Due to the broad adaptability of some hybrids, a single variety may be suitable for multiple ecological regions, causing the percentages across regions to sum to more than 100%. **B** Topological configuration of V1 (dark green) and V2 (dark orange) groups in PC1-PC2 genetic space (27.0% cumulative variance explained). **C** Heatmap of population genetic structure based on kinship parameters (PI_HAT/DST), where hierarchical clustering delineates the two major groups (V1: dark green, V2: dark orange)

NK815-Jingnongke 728, PI_HAT = 0.45, DST = 0.86). The Core Pioneer-derived subgroup consists entirely of V2 varieties, centered on Pioneer germplasm such as PH4CV and PH6WC. The Pioneer-introgressed subgroup, also composed of V2 varieties, shows initial blurring of boundaries with V1, reflecting extensive incorporation of non-Pioneer germplasm—particularly through lineages such as X1132X. The Transitional subgroup (e.g., Zhongnongda 787-Jingke 999, PI_HAT = 0.41, DST = 0.85) includes varieties located at the interface between V1 and V2 in the PCA plot, comprising accessions from both groups. Two additional subgroups belong to V1: the Huanggai Core (Zhengdan 958-Zhongke 11, PI_HAT = 0.72, DST = 0.93) and the Complex Genetic Background subgroup (e.g., Demeiya 1-Demeiya 2, PI_HAT = 0.31, DST = 0.83). This kinship structure corroborates the genetic distribution observed in PCA and illustrates a clear temporal pattern of genetic convergence between the two germplasm groups.

Temporal dynamics of genetic differentiation between V1 and V2 groups

Analysis of genetic differentiation index (Fst) between groups revealed a value of 0.0555 (CV = 0.505) between V1 and V2, exceeding the threshold for moderate genetic differentiation (Fst = 0.05) defined by Wright (1978) (Fig. 3-A).

The coefficient of variation (CV = 0.505) indicates heterogeneity in Fst across the genome, suggesting that genetic differentiation between V1 and V2 is not uniformly distributed. To investigate temporal changes, we subdivided the groups based on the year of registration into early (pre-2015; designated V1-1 and V2-1) and late (post-2016; designated V1-2 and V2-2) subgroups. The Fst between the early subgroups V1-1 and V2-1 was 0.0654 (CV = 0.529), while that between V1-2 and V2-1 was 0.0611 (CV = 0.727), also indicating moderate to marked differentiation (Fig. 3-B-C). The elevated CV values (0.529 and 0.727) reflect increased genomic heterogeneity in these comparisons, suggesting that differentiation varies across the genome. Notably, Fst values between the late subgroup V2-2 and both V1-1 and V1-2 dropped to 0.0244 (CV = 1.2) and 0.0141 (CV = 2.9), respectively (Fig. 3-D-E). According to Wright (1978), these values fall below the 0.05 threshold, indicating little genetic differentiation. The high CV values (1.2 and 2.9) indicate that while genome-wide differentiation is low, the degree of differentiation varies considerably across genomic regions. Within the V1 group, differentiation between temporal subgroups was minimal (V1-1 vs. V1-2 Fst = 0.0006), reflecting its structural stability over time (Fig. 3-F). In contrast, the V2 group showed an internal Fst of 0.0158 (CV = 2.11) between its early and late subgroups, indicating increased genetic variation in the later period (Fig. 3-G). The high CV (2.11) in this within-group comparison suggests that the increased genetic diversity in late V2 is not uniformly distributed across the genome. Overall, moderate to marked genetic differentiation was primarily observed between the early V2 subgroup (V2-1) and both the V1 group and the late V2 subgroup (V2-2). This temporal pattern suggests that the high proportion of Pioneer germplasm in V2-1 may have contributed to a distinct genetic identity during the early period, whereas subsequent germplasm admixture after 2016 may have contributed to the reduced differentiation observed between V2-2 and the V1 groups.

Temporal shifts in nucleotide diversity within V1 and V2 groups

Genome-wide nucleotide diversity (π) analysis revealed higher overall diversity in the V1 group (π = 0.00488) with a stable coefficient of variation (CV = 0.42) (Fig. 4-A).

The relatively low CV value indicates that genetic diversity is distributed relatively evenly across the genome in this group. Genetic diversity within each V1 subgroup remained consistent between the early (pre-2015) and late (post-2016) periods (V1-1: π = 0.00488, CV = 0.42; V1-2: π = 0.00490, CV = 0.43), reflecting its stable genetic structure over time (Fig. 4-C-D). In contrast, the V2 group showed lower overall diversity (π = 0.00454, CV = 0.429), particularly in its early phase (V2-1: π = 0.00443, CV = 0.436). A notable increase in diversity was observed in the later phase (V2-2: π = 0.00484, CV = 0.438), with a maximum π value reaching 0.018482, indicating enhanced nucleotide diversity through multi-source recombination (Fig. 4-B, E-F). The CV values in V2 remain moderate (0.429–0.438), suggesting that the increased diversity is distributed across the genome rather than concentrated in specific regions. These findings demonstrate a clear temporal contrast: while the V1 group maintained stable diversity with even genomic distribution, the V2 group underwent a notable expansion of its genetic diversity in the later period, consistent with increased germplasm admixture. The stable CV values in

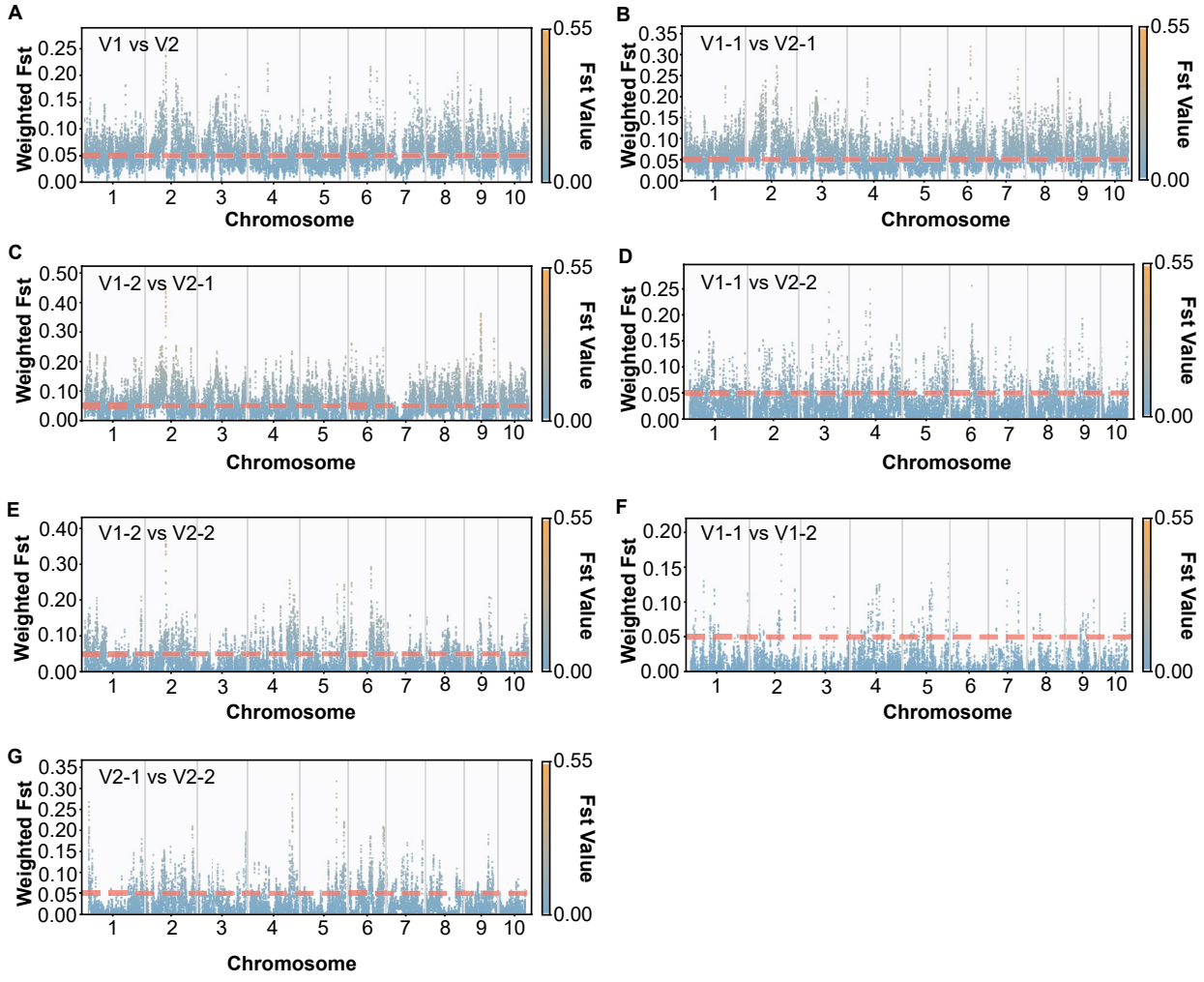

Fig. 3 Genetic differentiation characteristics of V1 and V2 populations. **A** Distribution of Fst between V1 and V2 populations (Fst = 0.0555, coefficient of variation CV = 0.505). **B**-**E** Comparisons of Fst values across subpopulations from distinct periods (V1-1 vs. V2-1: Fst = 0.0654, CV = 0.529; V1-2 vs.V2-1: Fst = 0.0611, CV = 0.727; V1-1 vs. V2-2: Fst = 0.0244, CV = 1.2; V1-2 vs. V2-2: Fst = 0.0141, CV = 2.9). **F**-**G** Analysis of within-group variation coefficients for V1 and V2 populations (V1-1 vs. V1-2: Fst = 0.0006, CV = 32.17; V2-1 vs. V2-2: Fst = 0.0158, CV = 2.11)

V2 further suggest that this expansion occurred broadly across the genome, reflecting the introduction of genetic material from diverse sources.

Phenotypic trends under policy influence

To elucidate long-term trends in breeding selection, we assessed 15 plant architecture and yield-related traits across 61 major cultivars over the 2022 and 2025 growing seasons. Best Linear Unbiased Prediction (BLUP) was applied to integrate the two-year phenotypic data, generating stabilized genetic potential estimates for each cultivar. These BLUP values were then subjected to linear regression against the year of variety registration (Figs. 5-A-H and 6-A-G).

To minimize potential bias caused by geo-adaptive differences between the two groups (V2 varieties are predominantly adapted to northern regions, while V1 varieties have broader ecological distribution), we restricted the temporal trend analysis to varieties suitable for cultivation exclusively in the Northern Spring Corn Region, the Huang-Huai-Hai Summer Corn Region, or the Northwest Irrigated Corn Region, and not suitable for cultivation in the Southwestern Mountain Corn Region or the Southern Hilly Corn Region. Following this definition, the selected subset comprised 26 V1 varieties and 16 V2 varieties. This approach ensures that observed phenotype–year relationships reflect genuine breeding trends within the same target ecoregion rather than inherent adaptive divergence between groups.

Among plant architecture traits in the V1 group, tassel dry weight significantly decreased (r = -0.447, p = 0.022), while leaf length increased (r = 0.52, p = 0.006) over registration years. This coordinated shift suggests a systematic optimization of plant architecture, likely favoring

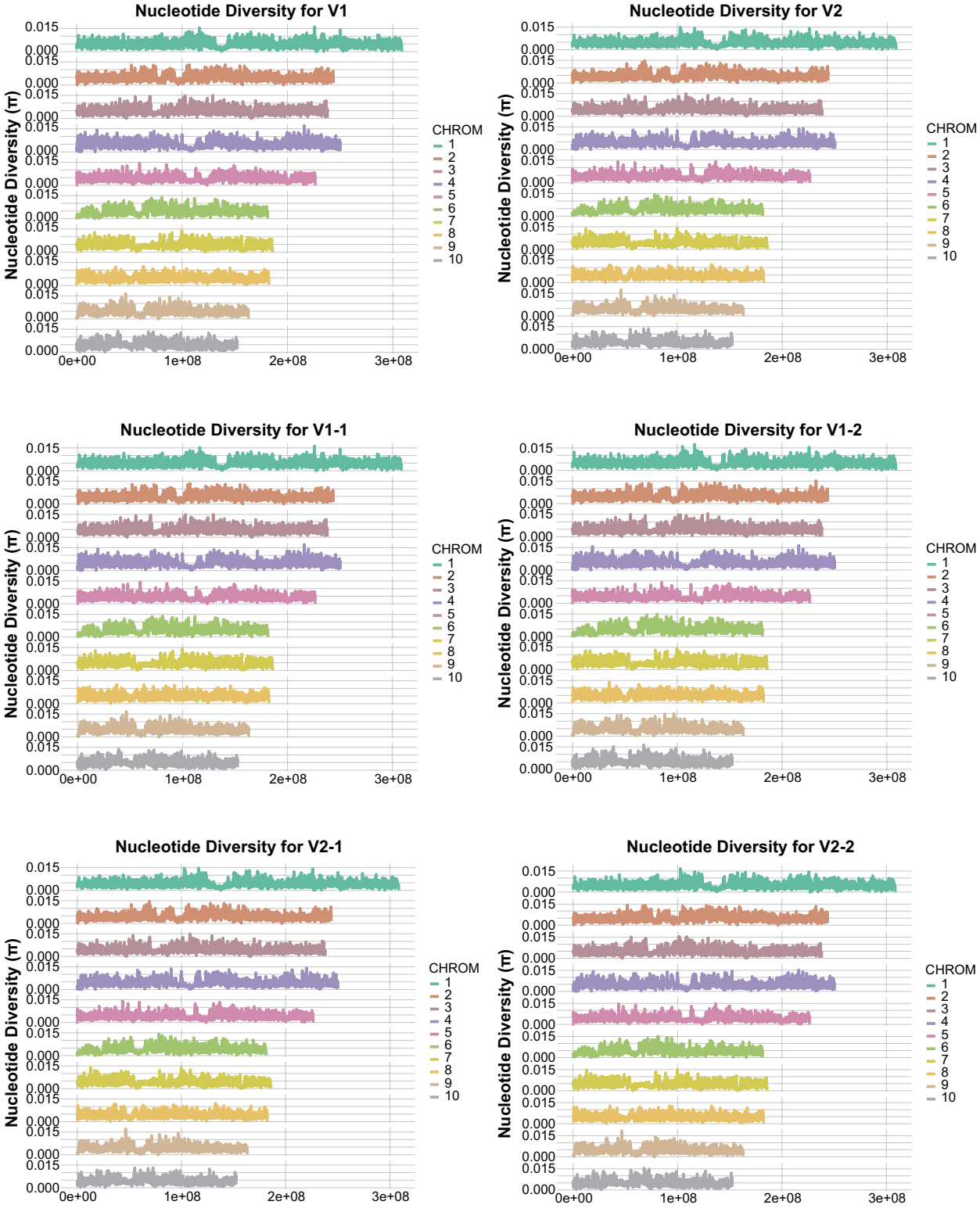

Fig. 4 Genome-wide nucleotide diversity (π) in V1 and V2 groups. Analyzed with 100-kb sliding windows (50-kb step size). **A** Genome-wide π distribution landscape in V1 group. **B** Genome-wide π distribution landscape in V2 group. **C** Subgroup-specific π distribution in V1-1. **D** Subgroup-specific π distribution in V1-2. **E** Subgroup-specific π distribution in V2-1. **F** Subgroup-specific π distribution in V2-2

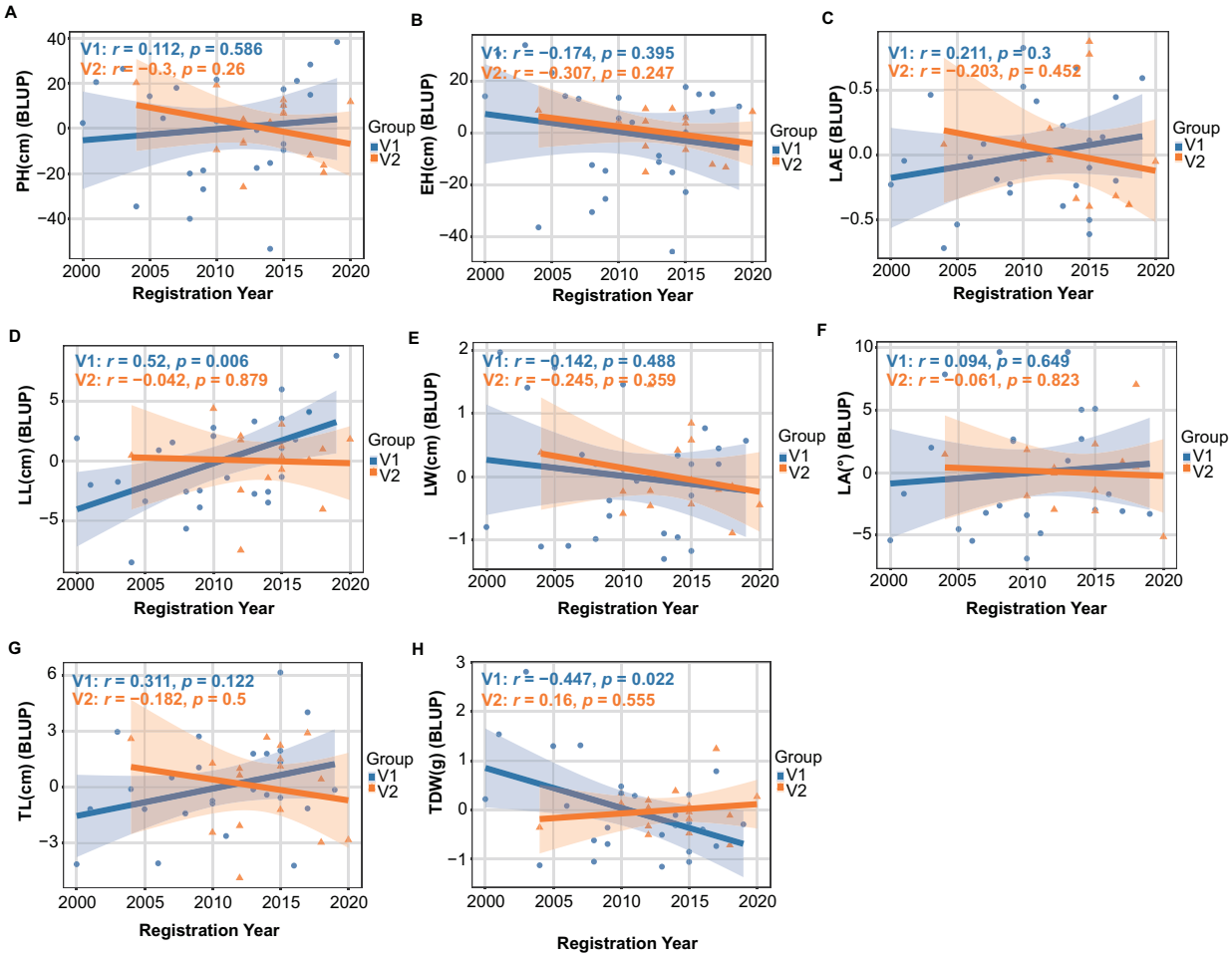

Fig. 5 Temporal dynamics of plant architecture traits in V1 and V2 groups. **A** Plant height (PH, cm). **B** Ear height (EH, cm). **C** Leaves above ear (LAE). **D** Leaf length (LL, cm) (V1: $r = 0.52$, $p = 0.006$). **E** Leaf width (LW, cm). **F** Leaf angle (LA, °). **G** Tassel length (TL, cm). **H** Tassel dry weight (TDW, g) (V1: $r = -0.447$, $p = 0.022$). *r-values = Pearson's correlation coefficient; p-values indicate significance

improved light interception and resource allocation efficiency (Fig. 5-D, H). For grain traits, kernel width decreased (r = -0.466, $p = 0.016$), indicating more compact grain arrangement, though other yield-related traits showed no significant trends. In contrast, the V2 group exhibited no significant temporal trends in any measured traits, reflecting phenotypic stability since its initial introduction. This phenotypic contrast suggests that the diverse V1 germplasm pool exhibited greater responsiveness to long-term breeding selection for architectural traits, while the V2 group maintained a stable phenotype throughout the period studied.

Analysis of phenotypic correlations and the decoupling of trait trade-offs

We applied the BLUP method to integrate phenotypic data from the 2022 and 2025 growing seasons, followed by separate phenotypic correlation analyses for plant architecture and yield-related traits. For plant architecture traits, plant height and ear height showed a strong positive correlation ($r = 0.88$). Both traits were also highly positively correlated with leaf length and tassel dry weight ($r = 0.56$–0.75), and moderately correlated with the number of leaves above the ear, leaf width, and tassel length ($r = 0.31$–0.51), but only weakly correlated with leaf angle. Other significant positive correlations were observed between the number of leaves above the ear and leaf length/leaf width, leaf length and tassel length/tassel dry weight, leaf width and tassel dry weight, leaf angle and tassel length, as well as tassel length and tassel dry weight, with coefficients ranging from 0.27 to 0.47 (Fig. 7-A).

For yield-related traits, the number of kernel rows per ear was strongly positively correlated with kernel width ($r = 0.63$). Hundred-kernel weight was positively correlated with kernel length and kernel width ($r = 0.64$ and 0.43, respectively). Ear width was also positively correlated with kernel length ($r = 0.50$). Additional significant positive correlations were detected between ear length and ear width/number of kernel rows, ear width and

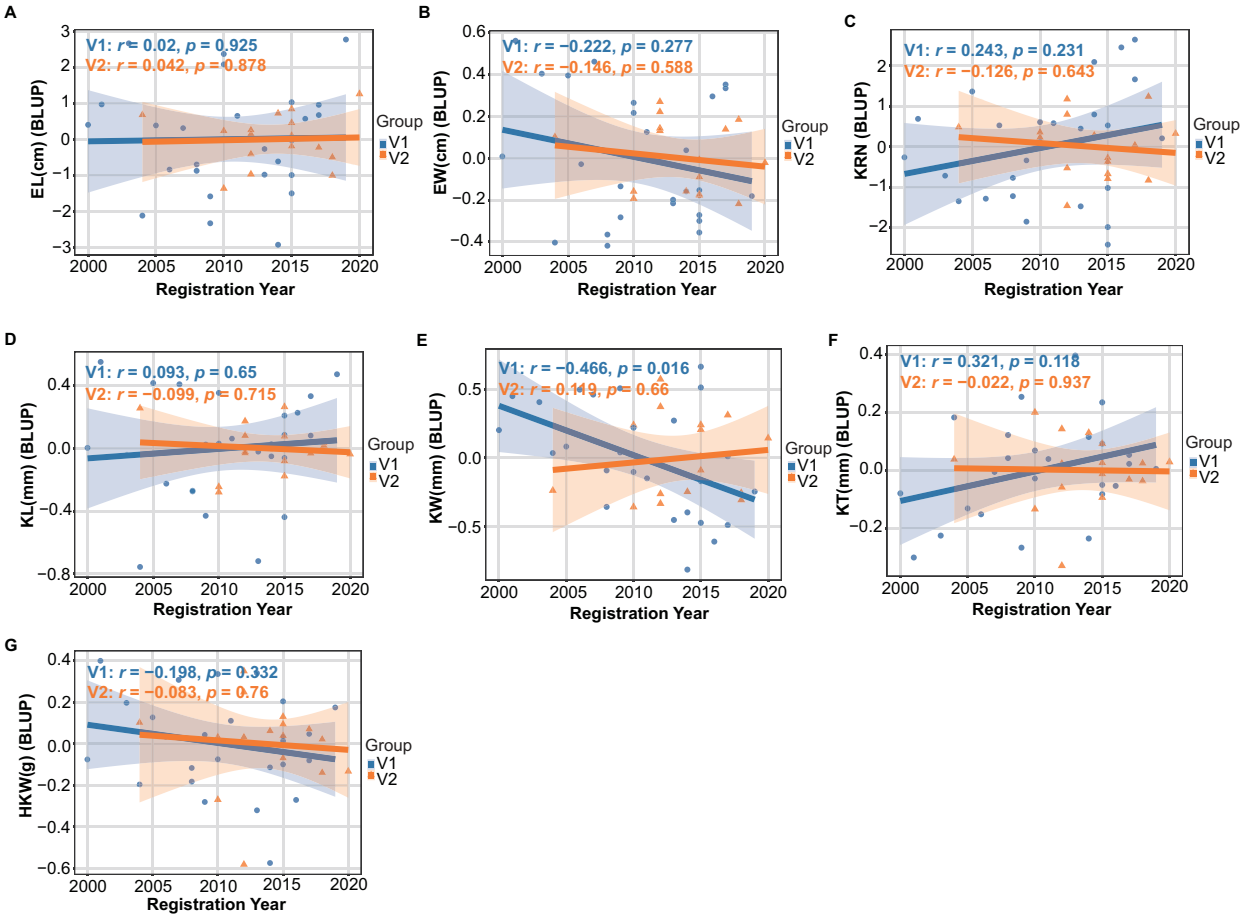

Fig. 6 Temporal dynamics of yield-related traits in V1 and V2 groups. **A** Ear length (EL, cm). **B** Ear width (EW, cm). **C** Kernel row number (KRN). **D** Kernel length (KL, mm). **E** Kernel width (KW, mm) (V1: r = -0.466, p = 0.016). **F** Kernel thickness (KT, mm). **G** Hundred-kernel weight (HKW, g). *r-values = Pearson's correlation coefficient; p-values indicate significance

hundred-kernel weight, kernel length and number of kernel rows, and kernel width and hundred-kernel weight, with coefficients ranging from 0.29 to 0.43. In contrast, kernel width showed a strong negative correlation with the number of kernel rows (r = -0.53), and ear width was negatively correlated with kernel thickness (r = -0.27) (Fig. 7-B).

Critically, within the V1 group, we observed a coordinated change that defied a major phenotypic correlation: against the background of a significant positive correlation between tassel dry weight (TDW) and leaf length (LL) (r = 0.37, p < 0.001), modern breeding has successfully achieved a decrease in TDW coupled with an increase in LL over registration years (Fig. 7-A). This decoupling of a key trait trade-off indicates a breakthrough in the optimization of plant architecture, suggesting a strategic reassignment of photosynthetic products. In contrast, the V2 group, despite its increased genetic diversity in the later period (Sect. 3.4), maintained stable inter-trait relationships with no significant shifts in the correlated traits over time. This absence of temporal change within the existing correlation framework is consistent with the overall phenotypic stability observed in V2 (Sect. 3.5).

Discussion

Genetic architecture and evolutionary trajectories of major maize germplasm pools in China

Our integrated genomic analysis delineates two principal germplasm pools, V1 and V2, which form the genetic foundation of China's elite maize hybrids. The significant genetic differentiation (Fst = 0.0555) between them confirms they represent distinct evolutionary lineages. The V1 pool is characterized by a broad genetic base, amalgamating domestic landraces (e.g., Huanggai), improved Reid germplasm, and diverse exotic non-Pioneer resources [46]. This composition reflects a century-long breeding history of introgressing and adapting global germplasm for local production systems. In contrast, the V2 pool is predominantly derived from a specific stream of U.S. Pioneer germplasm (PH4CV/PH6WC),

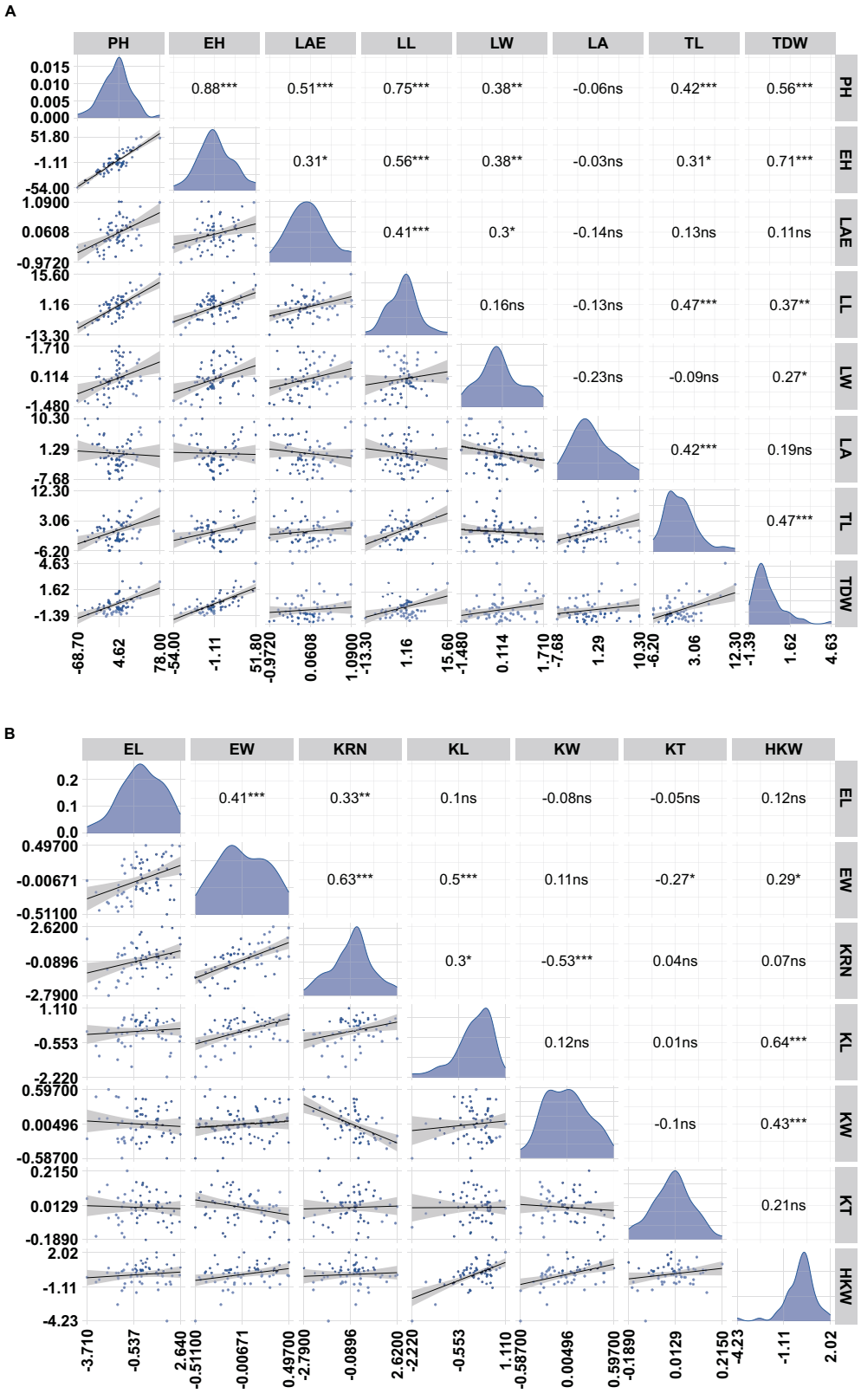

Fig. 7 (See legend on next page.)

(See figure on previous page.)

Fig. 7 Phenotypic correlation network analysis. Note: Diagonal shows univariate distribution; Lower triangle: scatter plots with regression lines; Upper triangle: correlation coefficients (R) with significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). **A** Plant architecture traits correlation matrix Traits: Plant height (PH), Ear height (EH), Leaves above ear (LAE), Leaf length (LL), Leaf width (LW), Leaf angle (LA), Tassel length (TL), Tassel dry weight (TDW). Key correlation: TDW-LL ($r = 0.37$**). **B** Yield-related traits correlation matrix. Traits: Ear length (EL), Ear width (EW), Kernel row number (KRN), Kernel length (KL), Kernel width (KW), Kernel thickness (KT), Hundred-kernel weight (HKW)

introduced and widely disseminated following the commercial success of the hybrid Xianyu 335 in the early 2000s [47].

Beyond characterizing this static structure, our temporal analysis revealed a dynamic process of genetic convergence. The significant reduction in genetic differentiation between V1 and V2 over time, coupled with a marked increase in nucleotide diversity within V2 in its later phase, is consistent with active and ongoing germplasm admixture. This blurring of genetic boundaries is likely a consequence of strategic breeding efforts to introgress favorable alleles from the diverse V1 pool into the high-performing V2 background, and vice versa, to broaden adaptation and mitigate genetic vulnerability. Beyond characterizing this static structure, our temporal analysis revealed a dynamic process of genetic convergence. The significant reduction in genetic differentiation between V1 and V2 over time, coupled with a marked increase in nucleotide diversity within V2 in its later phase, is consistent with active and ongoing germplasm admixture. This blurring of genetic boundaries is likely a consequence of strategic breeding efforts to introgress favorable alleles from the diverse V1 pool into the high-performing V2 background, and vice versa, to broaden adaptation and mitigate genetic vulnerability. This observed genetic convergence reflects a broader trend within China's maize breeding landscape. The continuous expansion of germplasm resources and the diversification of genetic improvement approaches have progressively blurred the genetic boundaries between traditional heterotic groups, such as the Huanggai, Improved Reid, and the derived PA, PB, and X groups [48–52]. Our genomic evidence quantifies this ongoing process, demonstrating that the initially distinct, Pioneer-derived pool (V2) and the broad-based domestic pool (V1) are actively integrating. This measurable admixture is reshaping the foundational genetic architecture available for future breeding in China.

The divergent yet intermingling trajectories of V1 and V2 underscore a complex breeding ecosystem in China. It is one where historically accumulated, diverse genetic resources (V1) coexist and interact with intensively selected, performance-driven germplasm blocks (V2). This interaction, evidenced by the convergence pattern, is not merely a passive genetic drift but likely an active driver of contemporary genetic gains, creating a novel and complex genetic landscape for future selection.

Decoupling of a key trade-off: a blueprint for ideotype breeding

The most significant phenotypic finding of this study is the successful decoupling of the positive correlation between tassel dry weight (TDW) and leaf length (LL) within the V1 germplasm pool. This represents a notable achievement in plant architecture optimization. In maize, tassel size and vegetative growth often compete for photosynthetic assimilates; a strong, positive correlation between these traits typically poses a fundamental constraint for breeders aiming to simultaneously reduce unproductive tassel investment and enhance the photosynthetic canopy. Our results demonstrate that modern breeding selection within the V1 lineage has overcome this constraint, achieving a synergistic improvement—reducing TDW while increasing LL.

The agronomic importance of this architectural shift must be understood within the context of contemporary Chinese maize production. Recent nationwide analyses confirm that increased planting density is the principal driver of yield gains, with the national average density rising significantly over the past decade [53]. Crucially, this study highlights that the yield response to higher density exhibits strong regional heterogeneity, and that smallholder land fragmentation presents a major barrier to adopting the precision management required for such systems [53]. This context underscores the value of our finding: the V1 pool does not merely offer a static "dense-planting" ideotype, but provides a flexible genetic blueprint for re-architecting plants.

Therefore, the V1 germplasm provides a valuable foundation for "adaptive ideotype breeding." It demonstrates that the complex, polygenic potential to dismantle key yield-limiting trade-offs resides within this historically cultivated Chinese germplasm. This genetic blueprint empowers breeders to engineer plants with optimized resource allocation—a trait complex critical for higher productivity ceilings. Crucially, this blueprint offers flexibility, allowing for the development of hybrids suited to diverse regional conditions and management scales, thereby directly addressing the heterogeneous challenges of yield intensification in China.

Phenotypic stability amidst genetic diversification: implications for breeding strategy

In contrast to V1, the V2 germplasm pool presents a scenario of genetic expansion without phenotypic shift. Our data show that while V2 underwent significant

genetic diversification through introgressive blending (as reflected in its increased nucleotide diversity), it maintained phenotypic stability across all measured traits over time. This uncoupling suggests that under the prevailing breeding objectives and selection pressures, the V2 pool may be approaching a selection plateau for its core agronomic phenotype. This interpretation is presented as a hypothesis that warrants further investigation.

Several alternative interpretations of the phenotypic stability observed in the V2 pool should be considered. First, this stability could reflect a genuine selection plateau, where current breeding objectives have exhausted the additive genetic variation for yield-architecture traits within this germplasm. Second, it is possible that selection has been successful for other traits not measured in this study, such as stress tolerance, grain quality, or disease resistance, which may have been the primary targets of breeding efforts in this pool. Third, the phenotypic panel employed here, while comprehensive for agronomic architecture, may have been insufficient to capture more subtle or non-linear phenotypic changes that occurred over the study period. These alternative explanations highlight the need for complementary approaches, including more detailed phenotyping and direct assessment of breeder selection records, to fully understand the evolutionary trajectory of the V2 pool.

This finding aligns with and informs forward-looking breeding strategies. The recently proposed "Maize2035" vision for intelligent breeding outlines two complementary paths to break such plateaus and accelerate genetic gain: deep mechanistic understanding for designing specific traits, and data-driven prediction systems for complex trait improvement [54]. The V2 case exemplifies the need for the latter. Its broadened genetic base represents a valuable reservoir of latent variation. To convert this variation into tangible phenotypic gains for next-generation objectives (e.g., nutrient efficiency, climate resilience), future efforts may require a paradigm shift: leveraging large-scale, high-throughput phenotyping and genomic prediction models within this elite genetic background to identify novel trait associations and guide targeted selection, thereby moving beyond its current phenotypic equilibrium.

Strategic germplasm utilization and future breeding perspectives

Our analysis synthesizes a dual-pathway framework for strategically utilizing China's foundational maize germplasm. The V1 pool serves as a validated genetic blueprint for targeted trait improvement, having demonstrably broken a key architectural trade-off. This makes it a prime resource for knowledge-driven "design breeding" aimed at coordinated phenotypic innovation. In contrast, the V2 pool exemplifies strategic genetic diversification, where substantial base broadening has enhanced its evolutionary potential while maintaining a stable, high-performing phenotype. However, its lack of concomitant phenotypic gains also signals a potential selection plateau for current breeding objectives, indicating that its greatest future value may lie as a superior, diversified platform for integrating novel traits.

Therefore, future breeding efforts should adopt a synergistic strategy leveraging these complementary roles. The V1 pool should be intensively mined to identify the causal genetic mechanisms underlying trade-off decoupling. Concurrently, the genetically broadened V2 pool should be prioritized as the premier background for targeted introgression of these and other novel alleles (e.g., for nutrient efficiency or stress resilience), catalyzing gains under redefined breeding objectives. This integrated approach provides a clear roadmap for advancing yield, resilience, and efficiency in Chinese maize breeding.

Limitations and future directions

While this study provides a comprehensive analysis of the genetic and phenotypic trajectories of China's major maize hybrids, several limitations should be acknowledged. First, although we identified significant phenotypic trends-particularly the decoupling of the tassel dry weight-leaf length trade-off in the V1 pool-we did not identify the underlying genetic variants responsible for this decoupling. Future studies employing quantitative trait locus (QTL) mapping, genome-wide association studies (GWAS), or selective sweep analyses are needed to pinpoint the causal genes or regulatory elements driving this coordinated architectural shift.

Second, our analysis focused primarily on agronomic architecture traits (e.g., plant height, leaf traits, yield components) that are routinely measured in breeding programs. This focus may overlook other important selection targets in modern breeding, such as stress tolerance (drought, heat, cold), nutrient use efficiency, grain quality (protein, oil, starch content), and disease resistance. A more comprehensive phenotyping approach that incorporates these traits would provide a fuller picture of the selection pressures acting on each germplasm pool.

Third, our temporal analysis spanned approximately two decades (1998–2021), which may not fully capture longer-term evolutionary trends. Continued monitoring of these germplasm pools as breeding programs evolve will be essential to validate the hypothesized trajectories and to detect emerging patterns.

Addressing these limitations through integrated genetic mapping, expanded phenotyping, and continued temporal monitoring will be critical for translating the insights from this study into actionable strategies for future maize breeding.

Conclusions

Collectively, our analyses reveal a dual narrative in the evolution of China's elite maize hybrids. The V1 germplasm pool exemplifies successful phenotypic innovation, where modern breeding has effectively broken a key architectural trade-off to unlock coordinated improvement. In parallel, the V2 pool presents a contrasting yet informative scenario of genetic expansion without immediate phenotypic returns; its broadened genetic base alongside stable performance suggests that selection under prevailing objectives may be nearing a limit. These complementary findings provide a coherent framework for future strategy: they validate the V1 approach as a replicable paradigm for trait uncoupling, while framing the V2 case as a diagnostic signal that may call for novel genetic inputs or redefined breeding goals to catalyze the next cycle of gain.

Abbreviations

PH	Plant height
EH	Ear height
LL	Leaf length
LW	Leaf width
LA	Leaf angle
LAE	Leaf above ear
KRN	Kernel row number
EL	Ear length
EW	Ear width
HKW	Hundred-kernel weight
KL	Kernel length
KW	Kernel width
KT	Kernel thickness
TL	Tassel length
TDW	Tassel dry weight
CTAB	Cetyltrimethylammonium bromide
BLUP	Best linear unbiased prediction

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12850-4.

Supplementary Material 1.

Acknowledgements

Not applicable.

Authors' contributions

**Guoji Li: ** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization, Writing - review and editing. **Hanyue Mao: ** Investigation, Data curation, Formal analysis, Writing - review and editing. **Sen Li: ** Investigation, Writing - review and editing. **Hai Wang: ** Supervision, Resources, Funding acquisition, Writing - review & editing.

Funding

This work was supported by the National Key Research and Development Program of China (2022YFD1201500), Special Guidance Fund for Central Universities (2025AC030) and the 2115 Talent Development Program of China Agricultural University.

Data availability

Raw sequence data that support the findings of this study have been deposited in the Genome Sequence Archive (GSA) at the National Genomics Data Center (NGDC), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences, under the accession number CRA038291 (publicly accessible at: https://ngdc.cncb.ac.cn/gsa/browse/CRA038291). The maize Zea mays B73 RefGen_v5 reference genome (Zm-B73-REFERENCE-NAM-5.0) used for bioinformatics analysis was retrieved from the Ensembl Plants FTP repository (https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-62/fasta/zea_mays/dna/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.dna.toplevel.fa.gz). All other data generated or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 31 January 2026 / Accepted: 7 April 2026
Published online: 14 April 2026

References

1. Nunn N, Qian N. The Columbian exchange: a history of disease, food, and ideas. J Economic Perspect. 2010;24:163–88. https://doi.org/10.1257/jep.24.2.163. https://pubs.aeaweb.org/doi/pdfplus/.
2. Tilman D, Balzer C, Hill J, Befort BL. Global food demand and the sustainable intensification of agriculture. Proc Natl Acad Sci U S A. 2011;108:20260–4. https://doi.org/10.1073/pnas.1116437108.
3. Bodirsky BL, Dietrich JP, Martinelli E, Stenstad A, Pradhan P, Gabrysch S, Mishra A, Weindl I, Le Mouël C, Rolinski S, Baumstark L, Wang X, Waid JL, Lotze-Campen H, Popp A. The ongoing nutrition transition thwarts long-term targets for food security, public health and environmental protection. Sci Rep. 2020;10:19778. https://doi.org/10.1038/s41598-020-75213-3.
4. Ray DK, Ramankutty N, Mueller ND, West PC, Foley JA. Recent patterns of crop yield growth and stagnation. Nat Commun. 2012;3:1293. https://doi.org/10.1038/ncomms2296.
5. Agnolucci P, Rapti C, Alexander P, De Lipsis V, Holland RA, Ekins F. Impacts of rising temperatures and farm management practices on global yields of 18 crops. Nat Food. 2020;1:562–71. https://doi.org/10.1038/s43016-020-00148-x.
6. Zhu R. Contemporary China's Crop Industry. China Social Sci Press. 1988; pp. 11–8. (in Chinese).
7. Tong PY. History of Maize Science and Technology in China. China Agricultural Science and Technology Press, Beijing. 2000:60–80. (in Chinese).
8. Dai J. Scientific and technological innovation of maize breeding in China. J Maize Sci. 2010;18:1–5. https://doi.org/10.13597/j.cnki.maize.science.2010.01.021. (in Chinese with English abstract).
9. Xue JB, Wang CH, Wei GY, Zhang DM, Jiang H, Zhao HJ, Chang BX, Gao X, Hou XS, Liu Y. Breeding of new maize variety Jiushenghe 2468. China Seed Ind. 2017;8:67–8. https://doi.org/10.19462/j.cnki.1671-895x.2017.08.024. (in Chinese). https://link.cnki.net/doi/.
10. Chen HJ, Ma WZ, Zhao ZH, Yuan CC, Wan BM, Zhang YD, Zhang ZG. A new machine-harvestable maize variety Zhengyuanyu 432. China Seed Ind. 2019;7:89–91. https://doi.org/10.19462/j.cnki.1671-895x.20190627.018. (in Chinese).
11. Wang XT, Gao HW, Hao JJ, Zhang J, Zhang GF, Wang ZP. Molecular marker-assisted selection of maize variety 'Weiyu618' with resistance to southern corn rust. Mol Plant Breed. 2024;22:4998–5006. https://doi.org/10.13271/j.mpb.022.004998. (in Chinese with English abstract).
12. Feng PY, Song RL, Yan HP, Xing CJ, Zhou F, Wang XG, Shi GQ. Analysis of High Yield Technology Model of NK815 Seed Production in 2023-To Explore New Ways of Large-Scale Maize Seed Production with High Quality, High Yield and High Efficiency. China Seed Ind. 2024;3:39–44. https://doi.org/10.19462/j.cnki.zgzy.20231215004. (in Chinese with English abstract).
13. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-; 2016.

14. Schloerke B, Cook D, Larmarange J, Briatte F, Marbach M, Thoen E, Elberg A, Crowley J, GGally. Extension to 'ggplot2' (Version 2.2.1). 2024. https://CRAN.R-project.org/package=GGally.
15. Wickham Hstringr. Simple, Consistent Wrappers for Common String Operations (Version 1.5.1). 2023. https://CRAN.R-project.org/package=stringr.
16. Henry L, Wickham H. rlang: Functions for Base Types and Core R and 'Tidyverse' Features (Version 1.1.4). 2024. https://CRAN.R-project.org/package=rlang.
17. Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: A Grammar of Data Manipulation (Version 1.1.4). 2023. https://CRAN.R-project.org/package=dplyr.
18. Wickham H, Vaughan D, Girlich M. tidyr: Tidy Messy Data (Version 1.3.1). 2024. https://CRAN.R-project.org/package=tidyr.
19. Wickham H, Hester J, Bryan J, readr. Read Rectangular Text Data (Version 2.1.5). 2024. https://CRAN.R-project.org/package=readr.
20. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4 (Version 1.1–35). J Stat Softw. 2015;67(1):1–48. https://doi.org/10.18637/jss.v067.i01.
21. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models (Version 3.1-3). J Stat Softw. 2017;82(13):1–26. https://doi.org/10.18637/jss.v082.i13.
22. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 1980;8:4321–5. https://doi.org/10.1093/nar/8.19.4321.
23. Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U.S.A.*. 1984; 81: 8014–8018. https://doi.org/10.1073/pnas.81.24.8014.
24. Joshi NA, Fass JN, Sickle. A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). 2011. https://github.com/najoshi/sickle.
25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324.
26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.
27. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303. https://doi.org/10.1101/gr.107524.110.
28. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75. https://doi.org/10.1086/519795.
29. Zhang SS, Chen X, Jin EH, Wang AK, Chen TT, Zhang XL, Zhu JW, Dong LL, Sun YL, Yu CX, Zhou YB, Fan ZJ, Chen HX, Zhai S, Sun YB, Chen QC, Xiao JF, Song SH, Zhang Z, Bao YM, Wang YQ, Zhao WM. The GSA Family in 2025: A Broadened Sharing Platform for Multi-Omics and Multimodal Data. Genom Proteom Bioinform. 2025;23(4):qzaf072. https://doi.org/10.1093/gpbjnl/qzaf072.
30. CNCB-NGDC Members and Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2025. Nucleic Acids Res. 2025;53(D1): D30-D44. https://doi.org/10.1093/nar/gkae978.
31. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64. https://doi.org/10.1101/gr.094052.109.
32. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. Array programming with NumPy. Nature. 2020;585:357–62. https://doi.org/10.1038/s41586-020-2649-2.
33. Hunter JD, Matplotlib. A 2D graphics environment. Comput Sci Eng. 2007;9:90–5. https://doi.org/10.1109/MCSE.2007.55.
34. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17:261–72. https://doi.org/10.1038/s41592-019-0686-2.
35. Waskom ML. seaborn: statistical data visualization. J Open Source Softw. 2021;6:3021. https://doi.org/10.21105/joss.03021.
36. Slowikowski K, ggrepel. Automatically Position Non-Overlapping Text Labels with 'ggplot2' (Version 0.9.6). 2024. https://CRAN.R-project.org/package=ggrepel.
37. Neuwirth E, RColorBrewer. ColorBrewer Palettes (Version 1.1-3). 2022. https://CRAN.R-project.org/package=RColorBrewer.
38. Kolde R, pheatmap. Pretty Heatmaps (Version 1.0.13). 2025. https://CRAN.R-project.org/package=pheatmap.
39. Wickham H. Reshaping data with the reshape package. J Stat Softw. 2007;21:1–20. https://doi.org/10.18637/jss.v021.i12.
40. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8. https://doi.org/10.1093/bioinformatics/btr330.
41. Pedersen T. patchwork: The Composer of Plots (Version 1.3.0). 2024. https://CRAN.R-project.org/package=patchwork.
42. Wickham H, Pedersen T, Seidel D. scales: Scale Functions for Visualization (Version 1.4.0). 2025. https://CRAN.R-project.org/package=scales.
43. Wright S. Variability Within and Among Natural Populations. Evolution and the Genetics of Populations. Volume 4. Chicago: University of Chicago Press; 1978.
44. Hartl DL, Clark AG. Principles of Population Genetics. 3rd ed. Sunderland, MA: Sinauer Associates; 1997.
45. Suarez-Kurtz G, de Araújo GS. Pharmacogenetic differentiation across Latin America. Pharmacogenomics. 2022;23(4):225–33. https://doi.org/10.2217/pgs-2021-0152.
46. Li YX, Wang TY, Li Y, Formation. Research and Utilization of Founder Parents in Major Crops. J Plant Genetic Resour. 2019;20:1093–102. https://doi.org/10.13430/j.cnki.jpgr.20190505003. (in Chinese with English abstract).
47. Wang SQ, Zhang LW, Zhang ZY, Wang SY. High-yield Maize New Variety—Xianyu 335. Sci Plant Breed. 2010;0749–50. https://doi.org/10.13270/j.cnki.kxzh.2010.07.030. (in Chinese).
48. Li CH, Guan HH, Jing X, Li YY, Wang BB, Li YX, Liu XY, Zhang DF, Liu C, Xie XQ, Zhao HY, Wang YB, Liu JB, Zhang PP, Hu GH, Li GL, Li SY, Sun DQ, Wang XM, Shi YS, Song YC, Jiao CZ, Ross-Ibarra J, Li Y, Wang TY, Wang HY. Genomic insights into historical improvement of heterotic groups during modern hybrid maize breeding. Nat Plants. 2022;8:750–63. https://doi.org/10.1038/s41477-022-01190-2.
49. Li ZY, Li CH, Zhang RY, Duan MX, Tian HL, Yi HM, Xu LW, Wang FG, Shi Z, Wang XQ, Wang JD, Su AG, Wang S, Sun X, Zhao YX, Wang SS, Zhang YX, Wang YD, Song W, Zhao JR. Genomic analysis of a new heterotic maize group reveals key loci for pedigree breeding. Front Plant Sci. 2023;14:1213675. https://doi.org/10.3389/fpls.2023.1213675.
50. Lu YL, Yan JB, Guimarães CT, Taba S, Hao ZF, Gao SB, Chen SJ, Li JS, Zhang SH, Vivek BS, Magorokosho C, Mugo S, Makumbi D, Parentoni SN, Shah T, Rong TZ, Crouch JH, Xu YB. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet. 2009;120:93–115. https://doi.org/10.1007/s00122-009-1162-7.
51. Wang RH, Yu YT, Zhao JR, Shi YS, Song YC, Wang TY, Li Y. Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. Theor Appl Genet. 2008;117:1141–53. https://doi.org/10.1007/s00122-008-0852-x.
52. Xue YJ, Zhao YK, Zhang YL, Wang R, Li XH, Liu ZH, Wang WW, Zhu SX, Fan YM, Xu LW, Zhao W, Zhao JR, Wang FG. Insights into the genomic divergence of maize heterotic groups in China. J Integr Plant Biol. 2025;67:1467–86. https://doi.org/10.1111/jipb.13884.
53. Wang XJ, Gao S, Yang HY, Li E, Xie RZ, Wang KR, Zhang GQ, Yu HB, Ming B, Li SK. Spatiotemporal dynamics of maize planting density in China: Yield responses and region-specific intensification thresholds. Crop Environ. 2025;100116. https://doi.org/10.1016/j.crope.2025.100116.
54. Liu HJ, Liu J, Zhai ZW, Dai MQ, Tian F, Wu YR, Tang JH, Lu YL, Wang HY, Jackson D, Yang XH, Qin F, Xu ML, Fernie AR, Zhang ZX, Yan JB. Maize2035: A decadal vision for intelligent maize breeding. Mol Plant. 2025;18:313–32. https://doi.org/10.1016/j.molp.2025.01.012.

Publisher's note