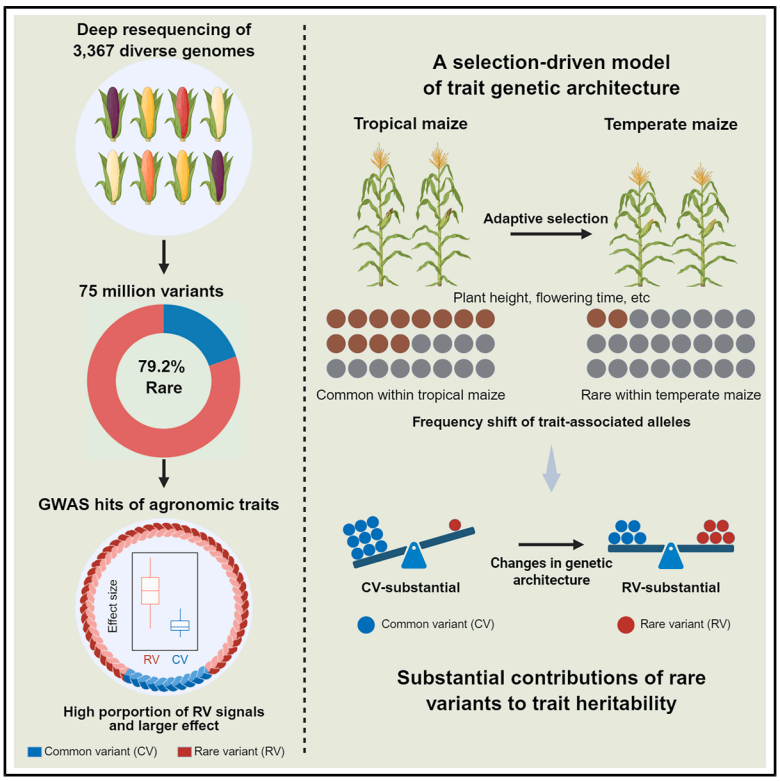
Cell

Article

Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits

Graphical abstract



Highlights

- A deep-sequenced reference panel of 3,367 diverse maize genomes
- Identification of 411 GWAS signals, with 87.3% attributable to globally rare variants
- Trait-associated rare variants show 2.4-fold larger effects than common variants
- Selection may reshape trait genetic architecture toward rare variants

Authors

Liangliang Huang, Chao Bian, Beibei Xin, ..., Haiming Zhao, Weibin Song, Jinsheng Lai

Correspondence

songwb@cau.edu.cn (W.S.), jlai@cau.edu.cn (J.L.)

In brief

Deep coverage sequencing of 3,367 maize genomes reveals substantial contributions of rare variants to trait heritability and provides insights into how selection reshapes their contribution to complex traits.

Huang et al., 2026, Cell *189*, 1–18
November 12, 2026 © 2026 Elsevier Inc. All rights reserved, including those for text and data mining, AI training, and similar technologies.
https://doi.org/10.1016/j.cell.2026.09.063

CellPress

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

Cell

CellPress

Article

Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits

Liangliang Huang,[1,22] Chao Bian,[1,22] Beibei Xin,[1,22] Xuecai Zhang,[2,3,22] Qiong Lu,[1] Yang Yue,[1] Yingshuang Li,[1] Rong Zeng,[4] Xinxiang Liu,[5] Xingming Fan,[6] Tianyu Wang,[7] Yanbo Wang,[8] Jingsheng Cao,[9] Jianguo Zhang,[9] Jianyu Wu,[10] Zhongwei Lin,[1] Jiuran Zhao,[5] Weidong Cheng,[11] Shibin Gao,[12] Yanhui Chen,[10] Yuncai Lu,[13] Junpin Yang,[14] Wenguo Liu,[15] Zehui Chen,[16] Haiyan Sun,[17] Xiaomin Lu,[18] Jiquan Xue,[19] Shuhua Yang,[4] Qiujie Liu,[1] Jian Chen,[1] Hainan Zhao,[1] Haiming Zhao,[1] Weibin Song,[1,20,21,*] and Jinsheng Lai[1,20,21,23,*]

[1]State Key Laboratory of Maize Bio-breeding, School of Interdisciplinary Life Sciences, Frontiers Science Center for Molecular Design Breeding (MOE), Key Laboratory of Maize Biology and Genetic Breeding, Key Laboratory of Genome Editing Research and Application (MARA), National Maize Improvement Center, China Agricultural University, Beijing 100193, P.R. China

[2]Ministry of Agriculture and Rural Affairs - CIMMYT Maize and Wheat Joint Laboratory, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100081, P.R. China

[3]International Maize and Wheat Improvement Center (CIMMYT), Texcoco 56237, Mexico

[4]State Key Laboratory of Plant Environmental Resilience, Frontiers Science Center for Molecular Design Breeding, College of Biological Sciences, China Agricultural University, Beijing 100193, P.R. China

[5]Beijing Key Laboratory of Maize DNA Fingerprinting and Molecular Breeding, Maize Research Institute, Beijing Academy of Agriculture and Forestry Sciences (BAAFS), Beijing 100097, P.R. China

[6]Institute of Food Crops, Yunnan Academy of Agricultural Sciences, Kunming 650205, P.R. China

[7]Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, P.R. China

[8]Institute of Maize Research, Liaoning Academy of Agricultural Sciences, Shenyang 110161, P.R. China

[9]Maize Research Institute, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, P.R. China

[10]State Key Laboratory of High-Efficiency Production of Wheat-Maize Double Cropping, Henan Agricultural University, Zhengzhou 450046, P.R. China

[11]Maize Research Institute, Guangxi Academy of Agricultural Sciences, Nanning 530007, P.R. China

[12]Maize Research Institute, Sichuan Agricultural University, Chengdu 611130, P.R. China

[13]College of Advanced Agriculture and Ecological Environment, Heilongjiang University, Harbin 150006, P.R. China

[14]Crop Research Institute, Sichuan Academy of Agricultural Sciences, Crop Germplasm Center of Sichuan Province, Chengdu 610066, P.R. China

[15]Maize Research Institute, Jilin Academy of Agricultural Sciences, Gongzhuling 136100, P.R. China

[16]Yazhouwan National Laboratory, Sanya 572024, P.R. China

[17]College of Horticulture and Landscape Architecture, Southwest University, Chongqing 400715, P.R. China

[18]Cereal Crop Research Institute, Henan Academy of Agricultural Sciences, Zhengzhou 450002, P.R. China

[19]Key Laboratory of Biology and Genetic Improvement of Maize in Arid Area of Northwest Region, College of Agronomy, Northwest A&F University, Yangling 712100, P.R. China

[20]Sanya International Maize Research Center, Sanya Institute of China Agricultural University, Sanya 572025, P.R. China

[21]Center for Crop Functional Genomics and Molecular Breeding, China Agricultural University, Beijing 100193, P.R. China

[22]These authors contributed equally

[23]Lead contact

*Correspondence: songwb@cau.edu.cn (W.S.), jlai@cau.edu.cn (J.L.)

https://doi.org/10.1016/j.cell.2026.09.063

SUMMARY

Rare variants are highly abundant, yet their genome-wide contributions to agronomic traits remain elusive in plants. Here, we performed deep whole-genome sequencing of 3,367 diverse elite maize inbred lines, uncovering 75.3 million variants, of which 79.2% were rare. Overall, we identified 411 independent association signals for 10 agronomic traits, including 359 attributable to globally rare variants, with rare variants showing effect sizes approximately 2.4 times those of common variants. Frequency-stratified genome-based restricted maximum likelihood (GREML) analyses indicated that rare variants captured a substantial proportion of phenotypic variance for traits under strong adaptive selection, consistent with the allele frequencies of trait-associated variants shifting from common in tropical to rare in temperate maize. Complementing

All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

conventional variant-level genome-wide association studies (GWAS), gene-based tests identified an additional 94 candidate gene-trait associations. Our study highlights rare variants as important components of the genetic architecture of complex agronomic traits in maize and provides much-needed resources for future functional studies and breeding.

INTRODUCTION

Over the past decades, the development of numerous segregating populations[1–4] and natural populations[5–10] has enabled significant advances in dissecting the genetic loci driven by common variants (minor allele frequency [MAF] $\geq$ 0.05) that underlie agronomic traits in plants, through linkage mapping and genome-wide association studies (GWAS). Although rare variants (MAF < 0.05[11,12]) outnumber common variants in most crop genomes,[12–17] their genome-wide impacts on agronomic traits and breeding potential remain poorly understood, with only a very limited number of rare variants linked to important agronomic traits identified to date.[18–28] This long-standing knowledge gap largely stems from the following: (1) segregating populations are typically derived from a small set of founder lines, leading to insufficient representation of rare variants; (2) small natural populations lack the sample size required for robust statistical inference, particularly when using conventional variant-level association analyses.

Research on the genome-wide contributions of rare variants to complex traits has so far been largely restricted to human populations,[29–35] with numerous rare loci associated with complex traits identified.[29–34,36–38] Except for a few traits such as coronary artery disease (CAD[39]) and age at smoking initiation (AgeSmk[40]), rare variants typically play a minor role relative to common variants in explaining phenotypic variance for well-characterized complex polygenic human traits.[37,41–43] The observed genetic effects of rare causal variants are widely attributed to negative (purifying) selection exerted on them.[41,44–46] However, because selection pressures in human populations are relatively weak,[47,48] the phenotypic effects and contribution of rare variants to traits under strong selection remain largely unknown.

As one of the world's most important crops, maize (*Zea mays* ssp. *mays*) has undergone strong natural and artificial selection, which has driven substantial directional shifts in key agronomic phenotypes across diverse environments and breeding stages.[9–11,49,50] In this study, we performed deep whole-genome sequencing (WGS) of 3,367 maize inbred lines and systematically investigated the effect of rare variants on agronomic traits. Maize agronomic traits exhibited a high proportion of phenotypic variance explained (PVE) by genetic variants, and rare variants contributed substantially to the genetic architecture of multiple key traits. Further analyses indicated that strong selection during maize adaptation from tropical to temperate environments may have reshaped the genetic architecture of key agronomic traits by shifting the frequencies of trait-associated alleles from common to rare. The large sequenced population also enabled us to demonstrate the feasibility of using gene-based tests as an efficient method for establishing gene-trait associations, complementing conventional single-variant GWAS. In addition, the 3,367 deeply sequenced diverse genomes provide a valuable resource with natural knockouts (KOs) of nearly 30,000 protein-coding genes and can serve as a high-quality reference panel.

RESULTS

Deep WGS of 3,367 elite maize inbred lines

To elucidate the genetic basis of agronomic traits, we conducted deep (median 35.7 ×, range 22–116 ×) WGS of 3,367 diverse maize inbred lines from 23 countries (Figure 1A; Figures S1A and S1B; Tables S1A and S1B), covering almost all major maize cultivation regions. Through a pipeline integrating sequence mapping, variant calling, and stringent quality control (see STAR Methods), we processed 273.6 terabases (Tb) of raw sequencing bases and identified 63.64 million single-nucleotide polymorphisms (SNPs) and 11.67 million short insertions and deletions (InDels, <50 bp), corresponding to an average density of one variant per 28 bp across the B73 reference genome.[51] Of these SNPs, 79.6% were absent from the HapMap3 database,[52] and 89.89% of them occurred at rare allele frequencies (AFs) (MAF < 0.05; Figure S1C).

Population structure analysis categorized these 3,367 accessions into 19 subgroups (Figures 1B and 1C; Figures S1D and S1E; Table S1C; see STAR Methods). The 12 temperate subgroups primarily correspond to well-known heterotic groups[8,10,53] (Figures 1B and 1C; Figure S1E), whereas the seven tropical subgroups are distinguished by their geographic associations (Figures 1D and 1E; Table S1D). Specifically, one tropical subgroup mainly originated from Southeast Asia (SEA); four subgroups derived from Mexico exhibited distinct altitude-related adaptive characteristics (Figure 1E), one of which was a lowland-adapted subgroup predominantly composed of DTMA (Drought Tolerant Maize for Africa) germplasm (Figure 1F); and the remaining two subgroups were derived from India and Yunnan (China).

A typical tropical genome contains an average of 3.97 million SNPs and 0.38 million InDels relative to the B73 reference genome, numbers significantly exceeding those of temperate genomes (3.34 million SNPs, 0.35 million InDels; $p < 2.2 \times 10^{-16}$). This observation is consistent with faster linkage disequilibrium (LD) decay (Figure S1F) and an elevated number of private SNPs (Figure S1G) in tropical maize, collectively reflecting greater nucleotide diversity in tropical than in temperate maize.[15,54] To trace recent genetic exchange among maize subgroups, we focused on variants detected exclusively in 2 of the 3,367 genomes (defined as f_2 variants[55,56]; see STAR Methods). Given the current population size, the rarity of recurrent variants[57] makes f_2 variants a robust indicator of recent genetic exchange. Analysis of f_2 variant distribution revealed ~1.5-fold higher genetic exchange among temperate subgroups than among tropical

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

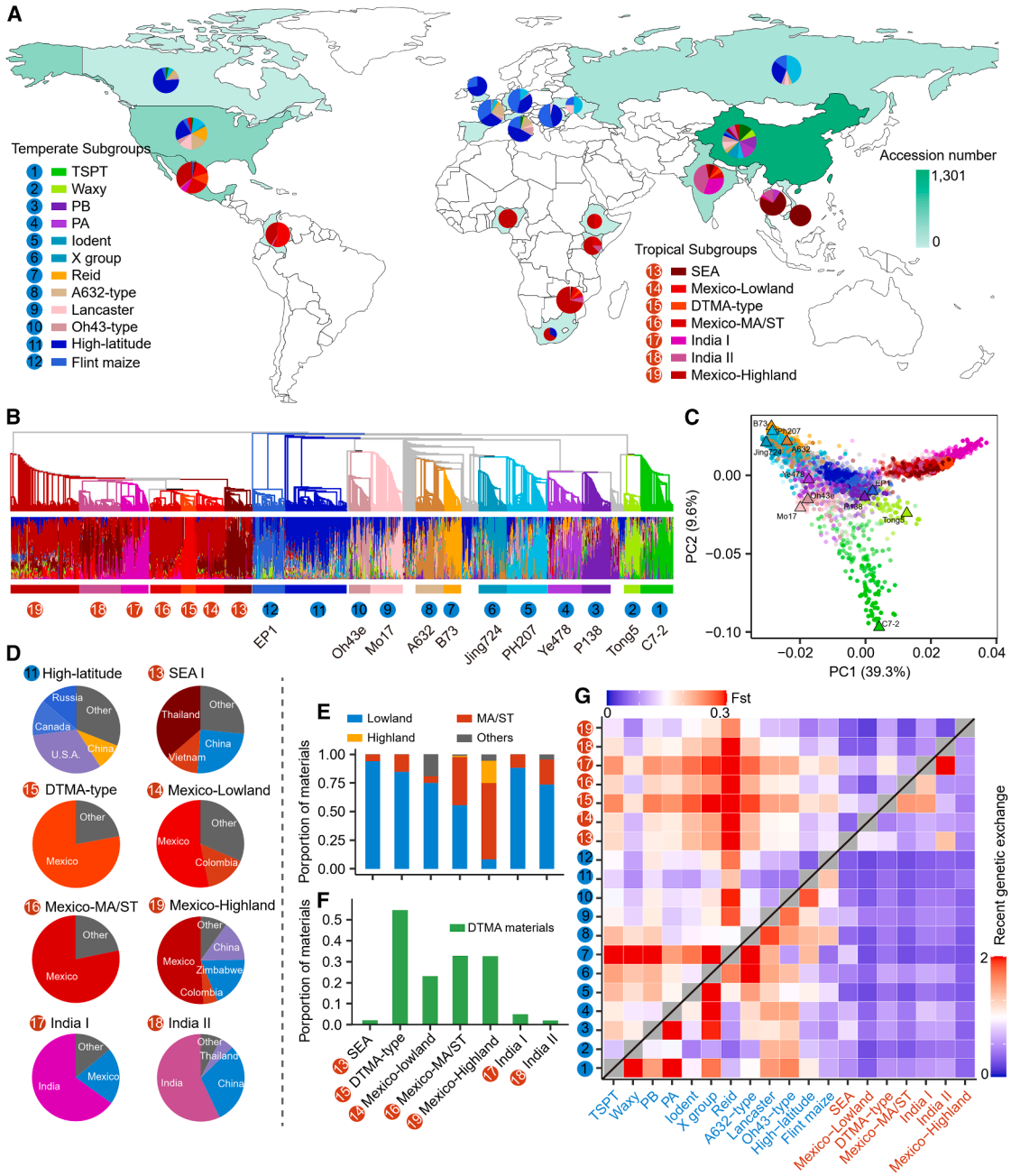


Figure 1. Geographic distribution and genetic relationships of 3,367 maize genomes

(A) Geographic distribution of the 3,367 maize inbred lines, with country-level pie charts showing subgroup compositions.

(B) Population structure of the 3,367 genomes. Top and bottom panels show the phylogenetic tree and admixture clustering results (K = 16), respectively. The phylogenetic tree further partitions two tropical ancestries into five subgroups associated with altitudinal adaptation (panel E). Black labels denote classical maize inbred lines of each heterotic group.

(C) Principal component analysis (PCA) of the 3,367 maize inbred lines.

(D) Geographic distribution of maize inbred lines from seven tropical subgroups and one temperate subgroup.

(E) Altitudinal adaptation of tropical maize inbred lines. Only CIMMYT materials with available altitudinal adaptation data are included in this analysis. MA, median altitude; ST, subtropical.

(F) The proportion of DTMA materials within each tropical subgroup.

(G) Genetic divergence (Fst) and recent genetic exchange (f_2 variants) among all 19 subgroups are presented in the left and right portions of the heatmap, respectively. Filled-circle numbers denote subgroup codes used throughout the figure.

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

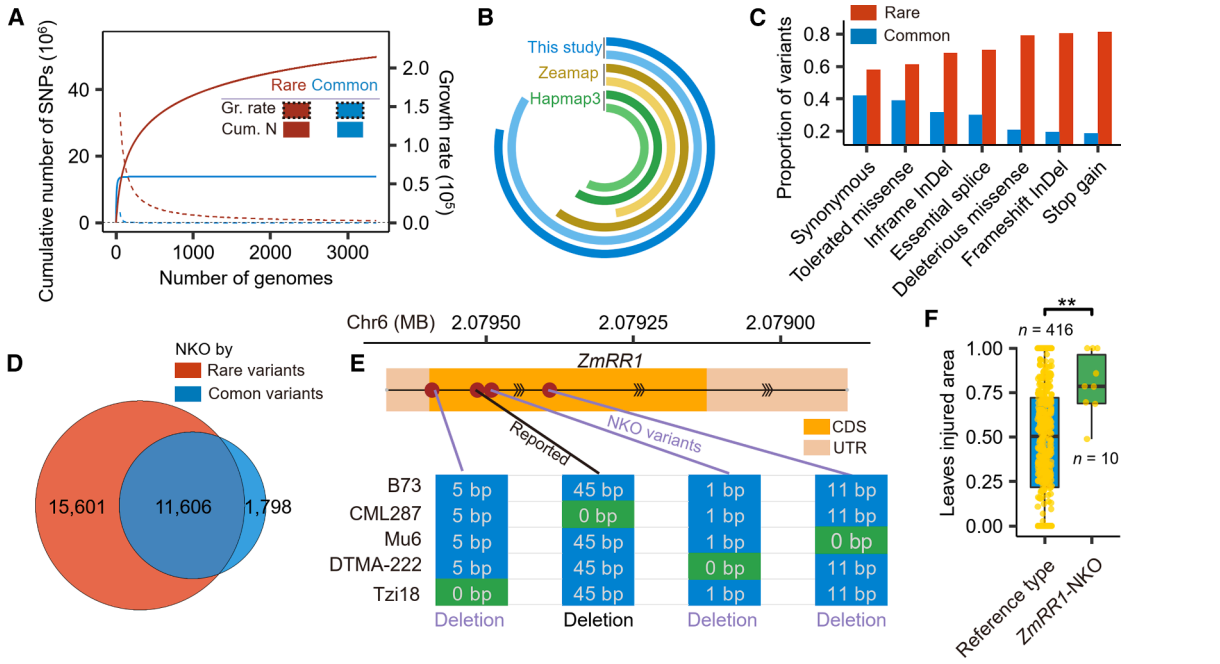


Figure 2. Genetic diversity and functional importance of rare variants

(A) Accumulation curves and growth rates with increasing sample sizes for rare and common variants. The gray horizontal dashed line indicates a growth rate of 200 variants.

(B) Proportions of rare variants among SNPs and InDels across different sequencing panels. Dark and light bars indicate SNPs and InDels, respectively.

(C) Proportions of rare and common variants across functional categories.

(D) Overlap of natural knockout (NKO) genes inactivated by common and rare variants.

(E) Gene structure of *ZmRR1* and detected NKO variants. Representative inbred lines for each haplotype are indicated on the left.

(F) Comparison of cold-induced injury between genotypes carrying or lacking the 1-bp rare frameshift InDel. Statistical significance was assessed using a two-sided *t* test of the regression coefficient for group. *n* denotes the number of inbred lines per group.

subgroups, despite an ~2.1-fold higher level of genetic divergence among temperate subgroups than among tropical subgroups (Figure 1G; Figures S1H and S1I; see STAR Methods). The coexistence of frequent recent genetic exchange and strong genetic divergence among temperate subgroups may reflect artificial breeding, in which backcrossing is widely used to maintain differences among germplasm subgroups after extensive inter-subgroup genetic exchange. This observation underscores the profound impact of modern breeding practices on maize genomes, particularly in temperate maize.

The landscape of rare variants in elite maize inbred lines

Variant discovery curves revealed distinct trajectories for common and rare variants as sample size increased. Common variants plateaued at ~430 genomes (<200 variants per additional genome), whereas rare variants maintained a steep growth trajectory, even with all 3,367 genomes included (Figure 2A). Consistent with this, 79.2% of variants detected in our panel were classified as rare (Figure 2B), a proportion much higher than that reported in prior small-sample studies ($n < 1,300$; Figure 2B).[15,52] This observation underscores the power of large-scale, high-depth WGS for rare-variant detection.

The genomic distribution of short variants varied by variant type and MAF category, with InDels enriched on chromosome arms and SNPs enriched in pericentromeric regions[58] (Figure S2A). Meanwhile, common InDels exhibited a more pronounced arm-enriched pattern than rare InDels, while rare SNPs showed stronger pericentromeric enrichment than common SNPs (Figure S2B). Recombination rates may contribute to these patterns, given their strong correlation with variant distribution (Figure S2C). A plausible mechanism is that low recombination hinders the elimination or spread of rare variants in pericentromeric regions. Given the high overall abundance of rare SNPs, combining all variant types revealed a genome-wide pericentromeric enrichment of short variants (Figure S2A), contrasting with the previously reported chromosome-arm bias observed in low-depth WGS studies.[59,60]

Driven by purifying selection, rare variants in our population exhibited a distinct functional stratification, with variants exerting more severe impacts on protein function being maintained at lower AFs (Figure 2C). Specifically, the proportion of rare variants across functional categories increased progressively from 58.0% for synonymous SNPs to 81.2% for stop-gain variants (Figure 2C). Consequently, incorporating rare variants expanded the total number of protein-altering variants by 235% compared with using only common variants (from 241,947 to 811,524; Table S2A; see STAR Methods). Meanwhile, functional annotation revealed that 18,220,804 variants overlapped with chromatin-accessible regions or peaks of active histone modifications in B73 (Table S2A). Notably, the proportion of

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

rare variants was consistently higher across all these functional regions than among synonymous SNPs (Figure S2D). Among these functional variants, protein-truncating variants have gained substantial attention in genomic analyses,[57,61–64] because they typically inactivate specific genes and function as natural knockout (NKO) variants. Notably, we identified 29,005 potential NKO candidate genes (80.9% of all protein-coding genes in the B73 reference genome; Tables S2A and S2B) in our population, with 53.8% of these NKO genes truncated exclusively by rare NKO variants (Figure 2D; see STAR Methods). Genes without NKO variants were significantly enriched in ribosomal and translational processes, consistent with their strong functional constraint. Genes carrying high-frequency NKO variants (AF > 0.5) were enriched for membrane-associated and signaling-related functions, implicating these genes in adaptation-related biological pathways (Figure S2E).

Meanwhile, these NKO alleles may be valuable breeding resources. For example, beyond the previously reported 45-bp deletion, we detected three additional rare frameshift variants in the coding region of *ZmRR1* (Figure 2E), a type-A response regulator previously implicated in enhancing maize chilling tolerance.[65] Among the 30 genomes carrying these NKO variants, 29 were from tropical subgroups (Table S2C), aligning with the relaxed selection pressure for cold adaptation in tropical regions. To validate the phenotypic impacts of these variants, we evaluated chilling tolerance in 426 tropical maize inbred lines. Ten inbred lines carrying the 1-bp deletion (one of these three NKO variants) exhibited significantly larger injured areas (73.0%) under 4°C cold treatment compared with non-carriers (47.6%; effect size: 0.35; 95% confidence interval [CI]: 0.11–0.59, $p = 0.005$; Figure 2F), indicating that the positive cold tolerance regulator may be dysfunctional in genomes harboring NKO variants. Similarly, we discovered two rare frameshift InDels and one nonsense SNP in the coding region of *ZmACO2* (Figure S2F), which encodes an ethylene biosynthesis enzyme previously shown to negatively regulate maize ear length (EL).[66] Carrier lines (5 of 1,769) displayed significantly longer ears (18.1 cm) relative to non-carriers (15.2 cm; effect size: 3.1 cm; 95% CI: 0.90–5.35; $p = 0.006$; Figure S2G), consistent with the disruption of this negative EL regulator. Although additional validation is needed, many of these NKO variants could be useful for future maize functional genomics and breeding.

Phenotypic effects of rare variants on key agronomic traits

To evaluate the phenotypic effects of rare variants on agronomic traits, we measured 10 key agronomic traits across two environments (Tables S3A and S3B; see STAR Methods). The mean inter-environment Pearson correlation (r = 0.78, range: 0.58–0.92; Tables S3A and S3B) indicated moderate-to-strong phenotypic consistency. Because increased sample sizes improved statistical power for detecting rare-variant effects,[34,36] we conducted association analyses for variants with at least 10 phenotypic observations for the minor allele (Table S3C; MAF > 0.3%; see STAR Methods) to capture associations across different allele-frequency spectra. After correcting for cryptic relatedness and population structure, we detected 5,758 significant variants ($p \leq 1 \times 10^{-7}$). The mean genomic inflation factor (λ) across GWAS analyses was 1.009, with values of 1.07, 1.04, and 0.93 for variants with MAF <1%, 1%–5%, and >5%, respectively, indicating limited overall residual inflation (Figure S3A; Table S3D). Additionally, the reconfirmation of well-known trait-associated loci, including *ZmCCT10*[67] and *ZCN8*,[68,69] provided independent validation of the GWAS results (Figure S4A). After controlling for the confounding effects of LD by conditional and joint multiple-variant analysis (COJO),[36,70] we detected 411 independent associations that were significant in the initial GWAS (Table S3E; see STAR Methods). Of these, 359 (87.3%) were globally rare variants across the entire panel (Figure 3A), with a mean AF of 1.38%. The proportion of rare signals remained robust across association models incorporating additional principal components or subgroup membership (Figure S3B), and few rare alleles were restricted to a single subgroup (Figure S3C). Although the AFs of these alleles were not uniformly distributed across ancestral groups and subgroups, ~86% were rare in both tropical and temperate germplasm and ~28% were rare in all 19 subgroups. Notably, 349 rare signals remained rare after equal weighting across subgroups, indicating that uneven subgroup representation alone cannot explain the high proportion of rare associations. Together, these results suggest that globally rare variants contribute substantially to agronomic traits.

For example, we uncovered a rare signal for days to anthesis (DTA) on chromosome 5 (Figure 3B). The lead variant was located ~5 kb upstream of *ZmAGO5c*, with no other annotated genes in the ~18-kb flanking region (Figure 3B). This signal remained robust across most subgroup configurations in leave-one-subgroup-out analyses, although significance decreased after excluding the Mexico-highland subgroup, consistent with reduced statistical power due to fewer carriers and larger standard errors (Figure S3D; Table S3F; see STAR Methods). The association remained independent after conditioning on other signals on the same chromosome (Figure S3E; see STAR Methods). *ZmAGO5c* was previously reported as the causal gene underlying the male-sterile mutant *ms28* in maize.[71] Genomes carrying this rare allele (AF = 1.1%) exhibited significantly delayed DTA compared with the non-carriers (Figure 3C), consistent with its Arabidopsis homolog, which is involved in flowering-time regulation.[72] All genomes harboring this rare variant belonged to tropical groups, aligning with the later flowering times of tropical lines (Figure 3D). To experimentally validate the function of *ZmAGO5c* in regulating DTA, we generated two KO lines using CRISPR-Cas9 (Figure 3E). Both mutant lines showed a significant ~4-day delay in DTA compared with their wild-type siblings (Figures 3F and 3G).

Beyond the proportion of rare variants contributing to agronomic traits, differences in effect sizes between common and rare variants have not been investigated in natural diversity populations in plants. We therefore compared conditional effect sizes of these detected trait-associated variants after accounting for nearby significant associations to mitigate LD-related confounding (see STAR Methods). Results indicated a strong trend of inverse correlation between the effect sizes of trait-associated variants and their MAF across all 10 traits (Figure 3H; Figure S4C). Individual common variants displayed a mean effect size of 0.21 phenotypic standard deviations (Figure 3I), consistent with previously reported QTL effect sizes in teosinte

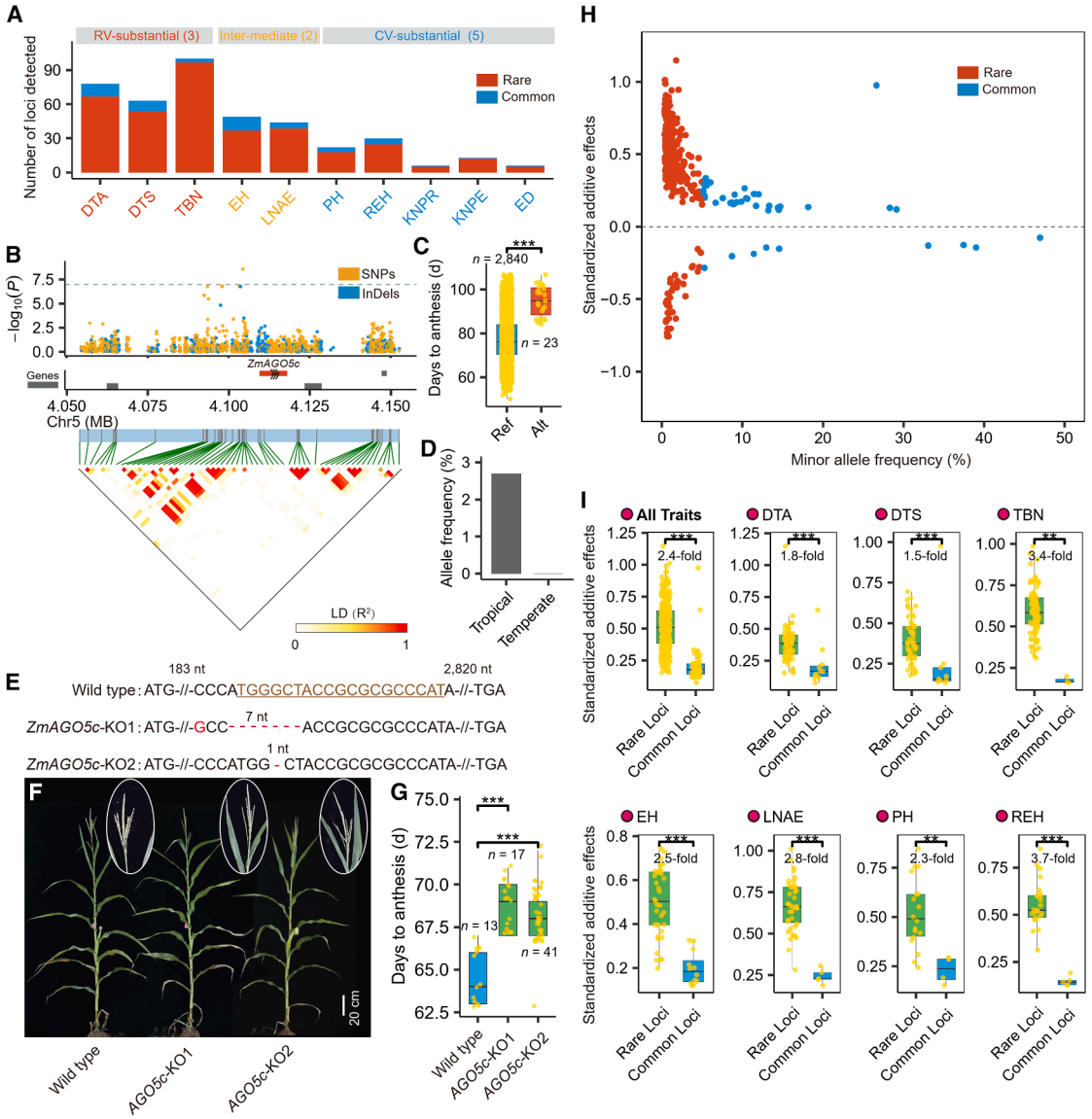


Figure 3. Phenotypic effects of rare variants on 10 key agronomic traits

(A) Number of rare and common trait-associated signals detected by GWAS for 10 traits. Trait categories are indicated above the plot and detailed in Figure 4A. RV, rare variant; CV, common variant.

(B) GWAS signals around *ZmAGO5c* for DTA. Variant significance, gene positions, and local LD structure are shown in the top, middle, and bottom panels, respectively. *ZmAGO5c* is highlighted in red.

(C) Comparison of DTAs between genotypes carrying the lead rare SNP (Alt) and non-carriers (Ref). *n* denotes the number of inbred lines per group.

(D) AFs of the lead rare SNP in temperate and tropical maize populations.

(E) CRISPR-Cas9-mediated knockout of *ZmAGO5c*, with the brown underline indicating the PAM sequence.

(F and G) Profile (F) and statistical analysis (G) of DTA for *ZmAGO5c* knockouts and their wild-type siblings. The two knockout lines were photographed separately with the same wild-type plant as a reference. Images were minimally cropped to avoid repeatedly displaying the same wild-type plant. Scale bar, 20 cm. *n* denotes the number of plants per genotype.

(H) Relationship between MAF and effect size of detected trait-associated variants.

(I) Comparison of effect sizes between detected rare and common trait-associated variants across traits. Only traits with ≥3 significant rare and ≥3 significant common loci are presented. See details in Table S3E.

Statistical significance for all pairwise comparisons was assessed using two-tailed *t* tests.

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

and maize landrace populations.[73] Within the same dataset, the average effect size of detected rare trait-associated variants was 2.4 times that of common variants across these traits (range: 1.5–3.7 times; Figure 3I; Figure S4C), highlighting the potential of these variants in genetic improvement. Sensitivity analysis using summary-statistics-based correction for winner's curse yielded consistent qualitative patterns (Figures S4D and S4E; see STAR Methods), indicating that the detected effect size difference was not solely driven by winner's curse. Although we identified few variants with effect sizes exceeding 1 phenotypic standard deviation, this may reflect limited statistical power to detect rarer loci given the current population size. Furthermore, the distribution of effect directions (trait increasing vs. trait decreasing) for minor alleles of these detected rare variants varied markedly across traits. Specifically, only 0.3% of the minor alleles of rare variants associated with DTA accelerated flowering time, whereas this proportion was markedly higher for decreased tassel branch number (TBN; 14.4%) and reduced plant height (PH; 61.1%; Figure S4C). These results were consistent with strong directional selection favoring early-flowering accessions during modern breeding, alongside relatively weaker selective pressure acting on TBN and PH.[9]

Considering modern breeding objectives[9,74] (Table S3A), 92 (25.6%) of rare trait-associated minor alleles had effect directions aligned with breeding objectives (Figure S4C), whereas only 13 associations showed potential trade-offs with the other 9 traits in our population (Table S3G), highlighting their breeding potential. For example, a rare favorable variant in the first intron of *ZmMADS1* was significantly associated with earlier flowering time (Figures S4A, S4F, S4G, S3D, and S3E; Table S3F). *ZmMADS1* promotes flowering by activating *ZCN8* and *ZmMADS69*.[75,76] Another example is a rare allele at *qPH6* associated with reduced PH and located 3.2 kb downstream of *ZmGLK31* (Figures S4B, S4I, S4J, S3D, and S3E; Table S3F), whose homologs have been reported to regulate PH in rice.[77] The favorable alleles of both variants exhibited an increased AF from tropical to temperate lines (Figures S4H and S4K), suggesting their potential active capture in breeding programs.

Contribution of rare variants to the genetic heritability of agronomic traits

To further quantify the contribution of rare variants to the genetic basis of agronomic traits, we performed frequency-stratified heritability estimation using genome-based restricted maximum likelihood (GREML) analysis,[78] a method widely applied in human genomics[31,37,79–81] (see STAR Methods). Results showed that PVE by genetic variants ranged from 0.32 (SE = 0.04) for kernel number per row (KNPR) to 0.86 (SE = 0.02) for TBN (Figures 4A and 4B), consistent with the intrinsically highly heritable properties of these traits. Seven of ten traits exhibited a significant contribution of rare variants to phenotypic variance ($h^2_{rare} > 0$, Figure 4A), indicating again that rare variants contributed broadly to agronomic traits. When stratifying the common and rare variant components of heritability, five traits (PH, relative ear height [REH], ear diameter [ED], kernel number per ear [KNPE], and KNPR) exhibited a common-variant-preferred genetic architecture, with common-variant heritability (h^2_{common}) significantly exceeding that of rare variants (Figure 4A), a pattern consistent with findings from most human complex traits.[41,43,46] By contrast, ear height (EH) and leaf number above the ear (LNAE) showed comparable estimated h^2_{rare} and h^2_{common}, while days to silking (DTS), DTA, and TBN displayed significantly higher estimated h^2_{rare} than h^2_{common} (false discovery rate [FDR] = 0.004; Figure 4A; see STAR Methods), indicating a potential rare-variant-preferred genetic architecture.

These estimates remained robust to phenotype standardization (Figure S5A; Table S4) and to correction for LD-related bias using LD- and MAF-stratified GREML (GREML-LDMS)[37,79] (Figures S5B and S5C; Table S4; see STAR Methods). Planting environments introduced greater fluctuations in heritability estimates, while the relative relationship of h^2_{rare} and h^2_{common} remained consistent (Figure S5D). Given the complex breeding history and strong population structure of maize germplasm,[82] we further controlled for structure-related confounding using additional principal components and subgroup membership. Rare-variant heritability estimates remained consistent across all analyses (Figure S5A). Nevertheless, within-subgroup phenotype permutations eliminated common-variant heritability but left residual rare-variant heritability (Figure S5E), suggesting that rare-variant heritability estimates may partly capture covariation from fine-scale structure and lineage-related genetic sharing, rather than solely reflecting the direct effects of rare causal variants. Notably, this residual signal is still significantly lower than the observed estimates (Figure S5E), although the observed differences between h^2_{rare} and h^2_{common} for DTA, DTS, and TBN did not significantly exceed the subgroup-preserving permutation null (Figure S5E).

Taken together, our results showed that rare variants contributed substantially to the genetic architecture of agronomic traits in modern maize. Although the extent of rare-variant contribution showed a trait-specific pattern, genetic architecture still differed between maize agronomic traits and most human complex traits/diseases. Specifically, maize agronomic traits exhibited a high proportion of PVE by genetic variants, with substantial rare-variant contribution observed across multiple traits. In contrast, human complex traits and diseases displayed a lower PVE, with traits typically having a common-variant-preferred genetic architecture.[37,41,43,46]

To further explore the potential role of selection in the substantial phenotypic variance captured by rare variants in maize, we focused on the dominant population structure in breeding maize. This structure reflects divergence between temperate and tropical lines (Figures 1B and 1C), coupled with marked phenotypic differentiation and intensive selection.[54,68,83] All seven traits with significant rare-variant heritability ($h^2_{rare} > 0$; Figure 4A) exhibited distinct phenotypic divergences between temperate and tropical populations (Figure S6A). The univariate Qst-Fst[84,85] analyses, comparing neutral genetic divergence (quantified by Fst) with quantitative trait divergence (quantified by Qst), rejected the neutral-drift-only hypothesis for all seven traits (FDR = 0.0178; Figure 4C; see STAR Methods), indicating that selection had driven the distinct phenotypic divergence from tropical to temperate lines for these traits rather than genetic drift alone.

Meanwhile, the distinct temperate-tropical phenotypic contrast provided a framework for examining how selection

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

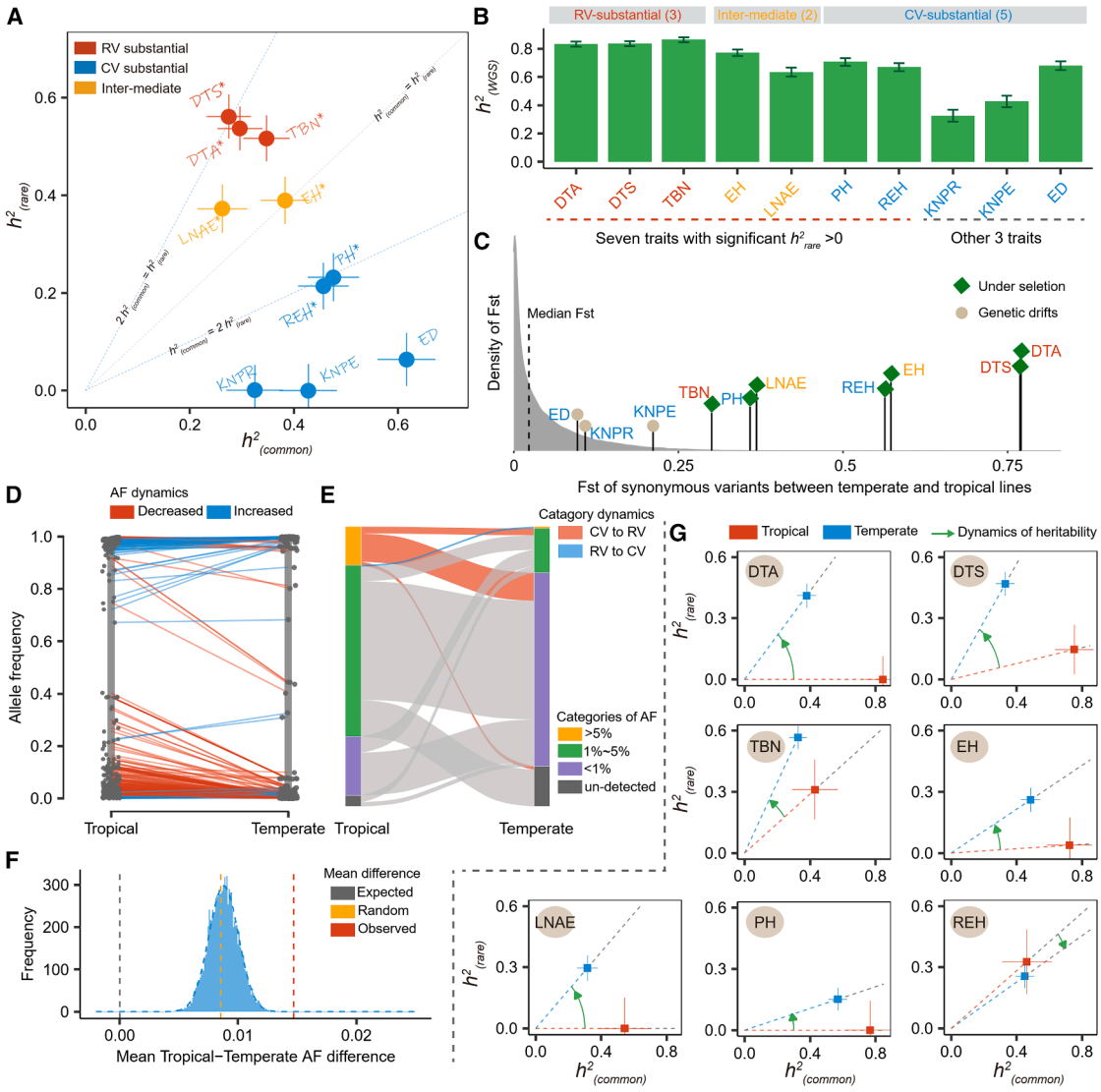


Figure 4. Dynamics in the genetic architecture of key agronomic traits in maize

(A) Comparison of the estimated common-variant component (*x* axis) and rare-variant component (*y* axis) of heritability using GREML. Asterisks denote traits with significant $h^2_{rare} > 0$, determined by normal approximation *Z* tests. Three dashed reference lines indicate different relationships between h^2_{rare} and h^2_{common}. Part of the estimated h^2_{rare} is not attributable to the direct causal effects of rare alleles; see Figure S5E. RV, rare variant; CV, common variant.

(B) Phenotypic variation explained by all WGS-derived variants across agronomic traits.

(C) Univariate Qst-Fst comparisons for all 10 traits between tropical and temperate genomes. Qst values are shown as individual lines for each trait. Traits with Qst values significantly exceeding the Fst distribution are shown in green.

(D) Frequency dynamics of tropical-preferred alleles for trait-associated variants between tropical and temperate genomes. Each line represents an individual variant.

(E) Transitions in allele-frequency categories of trait-associated rare variants between tropical and temperate genomes. Each trait-associated rare variant was classified according to its allele-frequency state in the two populations.

(F) Mean frequency differences of tropical-preferred alleles for trait-associated rare variants between tropical and temperate genomes, compared with those from 10,000 sets of randomly sampled variants. Significance was determined from the deviation of the observed value relative to the fitted normal distribution of the random sets.

(G) Comparison of heritability estimates for each trait between temperate and tropical maize lines using GREML, including only traits with $h^2_{rare} > 0$ in the 3,367 genomes. The green arcs with arrows represent shifts in the ratio of h^2_{rare} to h^2_{WGS} from tropical to temperate maize lines; see also Figures S6D and S6E.

For all panels, error bars represent SE.

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

acted on trait-associated rare variants for the seven traits. Leveraging the observed phenotypic divergence (Figure S6A), we defined tropical-preferred phenotypes as delayed flowering time, increased PH, and other corresponding agronomic traits (Table S3A). We then compared the AFs of these trait-associated rare variants and found that 67.7% of alleles linked to tropical-preferred phenotypes exhibited decreased frequencies in temperate relative to tropical lines (Figure 4D). Importantly, the magnitude of these directional AF shifts deviated significantly from that of randomly sampled variants ($p = 1.34 \times 10^{-6}$; Figure 4F; see STAR Methods), consistent with the widespread polygenic-selection-driven[86] AF changes in these trait-associated variants. Notably, 13.7% of these rare variants were present as common alleles in tropical but rare in temperate genomes, whereas only 0.6% showed the reverse pattern (Figure 4E), and these AF reductions were consistent across all 12 temperate subgroups relative to tropical germplasm (Figure S6C). Given that maize originated in tropical regions, these observations support a model in which selection may have driven a subset of trait-associated variants to shift from common (in tropical lines) to rare (in temperate lines).

To further evaluate whether these frequency shifts coincided with systematic changes in genetic architecture, we independently estimated trait heritability in tropical and temperate groups and compared the proportions captured by rare variants. Point estimates indicated that rare variants captured a greater proportion of heritability in temperate than in tropical lines for all seven traits with significant $h^2_{rare} > 0$ (Figure 4G; Figures S6D and S6E). However, estimates for the rare-variant component in tropical lines remained imprecise due to the limited sample size of lines with phenotypic data in this study (mean SE = 0.15; range: 0.11–0.21; Figure S6B). Still, the two most divergent traits (DTA and DTS) exhibited statistically significant differentiation between temperate and tropical genomes (Figure 4G). Collectively, these observations suggest that strong selection during maize tropical-temperate adaptation may have contributed to the reshaping of the genetic architecture of key agronomic traits by shifting trait-associated alleles from common to rare.

Gene-based associations for agronomic traits

To increase statistical power for detecting rare-variant associations, we also performed gene-based association testing, a method widely used in human genetics for pinpointing genetic targets underlying complex traits and diseases.[30,34,38] By integrating the effects of multiple qualifying variants within specific genes or loci, this approach enables the detection of aggregate effects of rare variants on agronomic traits. To further improve sensitivity, we implemented four variant-filtering criteria to retain qualifying variants based on functional annotations and AF thresholds (Table S5A). Synonymous variants were also included in these filtering criteria due to their potential phenotypic effects.[87–89] We subsequently tested the aggregate effects of these qualifying rare variants on 10 key agronomic traits using SKAT-O[90] (see STAR Methods). The mean inflation factors for the four filtering criteria were 1.41, 1.42, 1.37, and 1.32, respectively (Figure S6F; Table S3D), and this analysis incorporated ~1.2 million more rare alleles that were excluded by the MAF threshold in conventional single-variant GWAS per trait (Table S3C). Due to variability in qualifying variant counts across filters, the average number of gene-trait pairs tested per trait within each filter ranged from 12,200 to 34,855 (Table S5A).

Overall, we identified 151 significant candidate gene-trait associations at a genome-wide significance threshold of $p < 2 \times 10^{-6}$ after Bonferroni correction (Figure 5A; Table S5B; see STAR Methods). Even across all 10 traits and 4 filtering methods, the FDR was 1.2% after a global Benjamini-Hochberg correction. Compared with single-variant GWAS, 94 (62.3%) of these gene-based associations lacked significant variant-based signal within 1 Mb (Figure 5B), consistent with enhanced detection power. Among these, 34 (36.2%) were primarily attributed to mitigated allelic heterogeneity, with gene-based associations showing stronger significance than single-variant associations (Figure 5C). Another 42 (44.7%) associations were missed in single-variant GWAS because of the multiple-testing burden; significant variants in promoter or coding regions had low frequency (<10 samples), requiring a stringent threshold to avoid false positives (Figure 5C).

For instance, single-variant GWAS identified no significant association near the Trihelix transcription factor gene *ZmThx* for TBN ($p = 6.21 \times 10^{-7}$ for the most significant variant; Figure 5D), even when considering variants with extremely low frequency (MAF $\geq$ 0.07%). Within the coding sequence of *ZmThx*, we identified 26 rare variants (MAF < 1%; Figure 5E). Five of these variants showed nominally significant individual associations independently ($p < 0.05$; Figures 5E and 5F), and 18 variants were absent in either the top or bottom 20% of the TBN distribution (Figure 5E). Aggregating these 26 rare variants yielded a slightly stronger and significant gene-based association ($p = 2.18 \times 10^{-7}$; Figure S3D). This gene-trait association remained robust across most subgroup configurations in leave-one-subgroup-out analyses, although significance decreased after excluding the Flint subgroup, which coincided with fewer minor-allele carriers of the lead single variant (Figure S3D; Table S3F). The gene function of *ZmThx* was further validated by two *ZmThx* KO mutant lines, both of which exhibited significantly reduced TBN relative to wild-type siblings (Figures 5G and 5H). Together, these results demonstrated the feasibility of gene-based tests in detecting candidate loci underlying agronomic traits, serving as a valuable complement to the widely used single-variant GWAS.

Utility of 3,367 deep-sequenced genomes as a reference panel for imputation

High-throughput genotyping technologies, such as genotype-by-sequencing (GBS)[91] and microarrays,[92] enable cost-effective large-scale genotyping in maize but suffer from a high missing rate and low marker density. A high-quality reference panel can address these limitations by leveraging haplotype information to impute untyped genotypes. Given the negative impact of high variant missing rates on reference panel performance (Figure S7A), we constructed a reference panel of 3,367 diverse haplotypes across 30,080,804 maize SNPs, applying thresholds of minor allele count (MAC) > 5 and missing rate < 0.2. Using this panel, we evaluated genome-wide

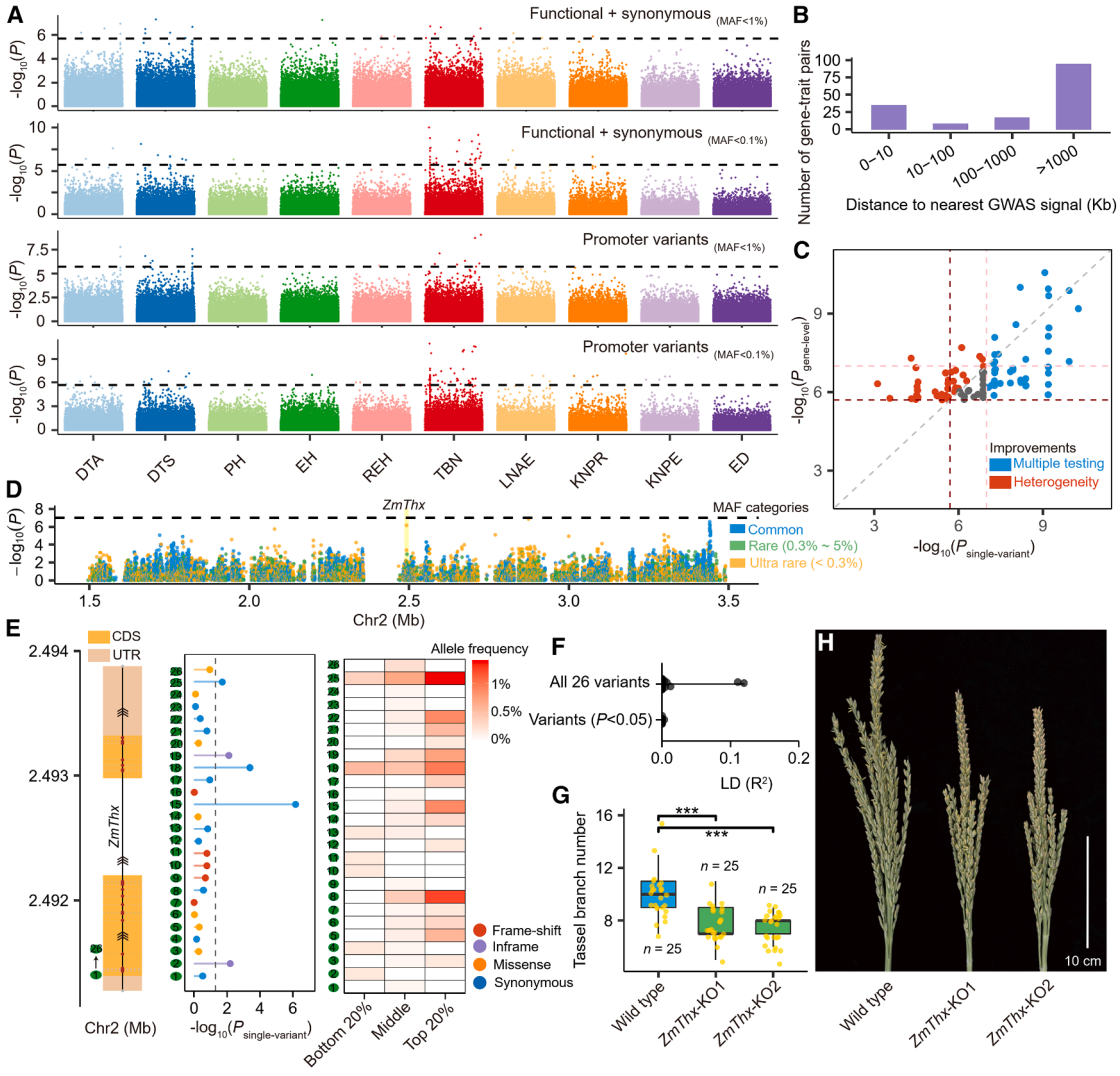


Figure 5. Boost GWAS sensitivity by gene-based association analysis

(A) Manhattan plot of gene-based association tests under four variant-filtering criteria. Each point represents a gene-trait pair, with genes ordered by chromosome and genomic position.

(B) Number of candidate gene-trait associations identified by gene-based tests, stratified by distance to single-variant GWAS signals. Variants with MAC > 10 were included in this single-variant GWAS analysis. Gene-trait pairs without significant single-variant signals within 1 Mb are classified as gene-based-specific candidate associations.

(C) Significance comparison of gene-based vs. single-variant GWAS for identified gene-based-specific candidate associations. All variants within gene bodies or promoters of corresponding genes are included in the single-variant GWAS analysis. The brown and pink dashed lines represent the significance thresholds for gene-based and single-variant association analyses, respectively.

(D) Manhattan plot of TBN from single-variant GWAS around *ZmThx*. The yellow rectangle indicates the genomic position of *ZmThx*.

(E) The three panels (left, middle, and right) respectively show the genomic positions, association significance with TBN, and frequency distributions of the 26 rare variants within the CDS region of *ZmThx*. Accessions are stratified by TBN phenotypic values.

(F) LD among these 26 rare variants.

(G and H) Phenotypic statistical analysis (G) and profile (H) of TBN in *ZmThx* knockouts and their wild-type siblings. Scale bar, 10 cm. Statistical significance was determined using a two-tailed *t* test. *n* denotes the number of plants per genotype.

imputation accuracy across the AF spectrum via downsampling our genotypes to ~680,000 polymorphic sites[82] (see STAR Methods). The mean imputation accuracy (r^2) reached 0.924 for common variants (Figure 6A). Accuracy decreased with declining MAF, yet rare variants with AFs > 0.1% could remain well imputed, achieving a mean r^2 of 0.703 for variants in the 0.1%–0.5% AF bin (Figure 6A). Increasing the original marker density before imputation could further improve accuracy (Figure 6B). Notably, imputation accuracy for rare variants was substantially higher in tropical than in temperate maize (Figure 6A). This pattern mirrors the higher imputation accuracy of rare variants observed in human populations of African

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

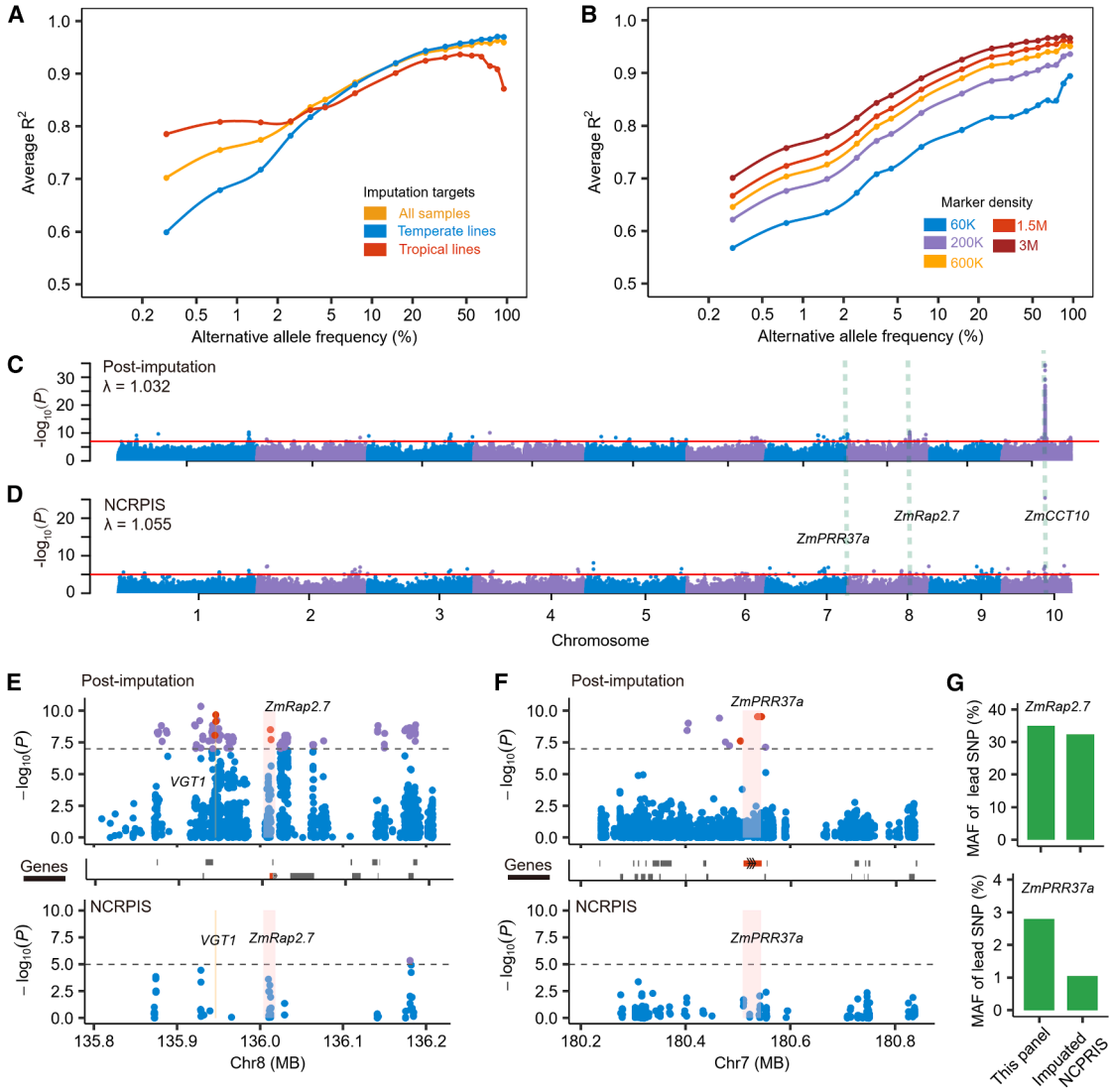


Figure 6. Imputation performance of the 3,367-genome reference panel

(A) Imputation accuracy of variants across different AFs using the 3,367-genome reference panel. AFs were calculated from the 3,367 maize genomes.

(B) Imputation accuracy under varying original genotypic marker densities. Original markers were randomly selected from the WGS data.

(C and D) GWAS signals for growing degree days to silking based on post-imputation data (C) and based on original data (D). The red horizontal line represents the significance threshold.

(E and F) GWAS signals around *ZmRap2.7* (E) and *ZmPRR37a* (F). GWAS signals from post-imputation data, gene positions, and GWAS signals from original data are shown in the top, middle, and bottom panels, respectively. Purple dots, significant variants; red dots, significant variants within 5 kb of the causal gene. *ZmRAP2.7* and *ZmPRR37a* are highlighted in red. The enhancer *VGT1* and its flanking 5-kb region are highlighted in yellow.

(G) MAF of lead variants near *ZmRap2.7* and *ZmPRR37a*. NCRPIS, the public collection of maize inbred accessions from the US Department of Agriculture North Central Region Plant Introduction Station.

ancestry, underscoring the importance of including sufficient tropical genomes in maize reference panels (Figure S7B).

To further demonstrate the utility of our reference panel, we used genotypic data from ~3,000 maize inbred lines from the US Department of Agriculture (USDA) North Central Region Plant Introduction Station (NCRPIS).[82] These data were generated via GBS and encompass approximately 680,000 polymorphic sites.[82] We re-ran GWAS for silking time with the EMMAX[93] model using both original and imputed genotypes (see STAR Methods). GWAS signals became more significant and denser after imputation, without increased genomic inflation (Figures 6C and 6D). The signal near *ZmRAP2.7*[94] on chromosome 8 was detected with both the original and imputed genotypes. However, imputation improved resolution as the second significant variant mapped directly to the well-known causal enhancer *VGT1*[94] after imputation, whereas no markers were detected in this region using the original data (Figure 6E). Importantly, imputation with the 3,367-genome reference panel also enhanced the detection of rare

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

trait-associated variants. For instance, we identified significant rare associations near *ZmPRR37a* (AF = 0.028; Figures 6F and 6G), a gene recently reported to regulate flowering time in maize.[95] After imputation, four equally significant rare variants were identified near *ZmPRR37a*, two of which were located within the gene body. In contrast, no significant association signal was detected using the original data (Figure 6F).

DISCUSSION

Elucidating the genetic basis of agronomic traits and expanding the exploitable genetic diversity are crucial for enhancing breeding efficiency.[96–99] Rare variants represent a largely untapped source of genetic diversity within crop genomes, with their phenotypic impacts and genetic contributions remaining poorly defined in plants. Our analysis showed that 79.2% of identified variants were rare, accounted for a large fraction of our GWAS signals, captured a substantial proportion of agronomic trait variance, and showed a 2.4-fold larger effect size than common variants for detected associations. These findings highlight the critical role of rare variants in determining maize phenotype and their potential value for future maize breeding.

The intensity of selection affects the genetic architecture of complex traits

In human populations where selection is generally weaker, low-frequency variants (0.5% < MAF < 5%; standard definition in human genetics, broadly consistent with our rare-variant cutoff) generally capture a smaller proportion of heritability[41,43] (~1/6 of common variants). In contrast, rare variants contributed substantially to the heritability of maize agronomic traits, with the observed estimates of h^2_{rare} being comparable to or exceeding h^2_{common} for 5 out of 10 traits investigated, although part of h^2_{rare} estimates may reflect fine-scale structure and lineage-related genetic sharing rather than direct causal effects. Traits with substantial rare-variant contributions in this study were characterized by significant selection-related phenotypic shifts and a high proportion of PVE by genetic variants, consistent with strong coupling (τ) between causal variant effects and selection pressure. This pattern aligned well with the genetic architecture simulations, where traits with smaller mutational target sizes (T) and strong selective coupling (τ) were predicted to be predominantly governed by rare variants.[43,100] Consistent with this, our results also support a model in which selection shifted a subset of trait-associated variants from common (in tropical lines) to rare (in temperate lines) for these traits. However, demographic history, founder effects, breeding bottlenecks, and sampling composition cannot be formally ruled out without demographic modeling that incorporates teosinte and maize. Overall, our results suggest that selection intensity can be an important factor in shaping the genetic architecture of complex traits and may be linked to the elevated contribution of rare variants to trait heritability in maize.

Large-scale sequencing resources across crops provide opportunities to investigate rare allele contributions,[101,102] although the predominantly low sequencing depth likely underdetects rare variants and hampers robust estimation of their contribution to heritability.[79] The outcrossing nature and heterotic breeding history of maize may favor the retention of rare variation, potentially magnifying the contribution of rare variants compared with some self-pollinating species. However, rare trait-associated alleles have been increasingly identified and functionally validated in multiple crops,[18–28] suggesting that their importance is not unique to maize. Given the pervasive action of phenotypic selection during crop domestication and improvement, and its impact on AFs,[103,104] we propose that the evolutionary dynamics of rare functional alleles are likely broadly shared across crop populations, although their quantitative contributions may differ among species.

Rare variants as genetic resources for crop breeding

Natural and artificial selection reshape the genetic architecture of complex traits by reducing the frequency of unfavorable variants.[9,68,105] However, due to the polygenic nature of complex traits,[46,106,107] genetic interactions,[108] and the finite efficiency of selection, it remains challenging to fully eliminate particular variants underlying complex traits through phenotype-based selection. Our analysis detected a systematic shift in AF from common to rare for trait-associated variants from tropical to temperate maize. This finding suggests that a substantial number of trait-related variants may persist as a rare legacy in modern maize populations, leading to an underestimation of true genetic diversity in maize when examined in small populations. Accordingly, harnessing rare variants offers a promising approach to retrieving these presumed-lost alleles, complementary to using wild relatives of maize.[109–111] This approach could be advantageous, given the reduced genetic divergence between the allele donors and recipients. Meanwhile, our results identified 92 rare variants exhibiting effect directions consistent with modern breeding objectives, while only 13 alleles showed potential trade-offs with the other 9 traits in our population. Considering their larger observed effect sizes, these rare variants may represent potentially valuable genetic resources for breeding applications. However, their broader utility remains to be evaluated, as their pleiotropic effects and stability across diverse genetic backgrounds and environments require systematic validation.

Additionally, our data may provide a valuable resource for further dissection of the genetic basis of heterosis. We identified ~11 million rare alleles private to specific heterotic groups (Table S6A), and annotated rare deleterious variants showed strong cross-group complementarity. For example, only 18% of rare deleterious variants were shared between the Reid and Lancaster subgroups, two classical heterotic groups that produce highly vigorous hybrids, consistent with the dominance-complementation model of heterosis[112] (Table S6B). Given that deleterious variants improve the accuracy of genomic selection in other crops,[89] these complementary rare deleterious alleles may represent promising markers for enhancing hybrid genomic selection accuracy.

Perspectives on rare functional alleles in crops

The low AF of rare alleles leads to reduced carrier numbers, resulting in limited power and unstable effect size estimates. At the current sample size, a substantial fraction of genuine rare-variant associations is likely to remain undetected. Increasing sample sizes improves statistical power for detecting

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

the functional effects of both common and rare variants.[34,36] The 3,367 deeply sequenced maize genomes can serve as a well-suited reference panel, enabling scalable and cost-efficient dissection of rare variants underlying key agronomic traits in larger breeding populations in the future. Meanwhile, 62.3% of the signals detected by gene-based tests were not identified by single-variant association analyses. This observation aligns with the high allelic heterogeneity of rare variants[34] and validates the power of this method for locus detection. Future approaches that integrate multiple rare variants across biologically meaningful units, beyond proximal gene regions, may further improve the sensitivity and interpretability of these frameworks. Because rare variants are abundant across diverse crops,[12–17] such gene-based testing and reference panel construction strategies are expected to be broadly applicable to other crops. More importantly, breeding panels of most crops always exhibit strong structure and genetic non-independence due to recurrent use of core elite germplasm. This fine-scale structure acts as both a statistical confounder and a reflection of the non-random distribution of trait-relevant alleles. This challenge is particularly acute for rare alleles that are carried by a small subset of individuals, so methods that disentangle their true effects from confounding signals while balancing FDR and detection power would be highly valuable.

Limitations of the study

We acknowledge several limitations. First, variant discovery was based on a single reference genome (B73). Due to the limited representation of the maize sequence space with a single reference,[59,113] the actual variant set is likely underestimated. Second, further increasing population size is expected to yield more precise heritability estimates (i.e., with reduced SE), especially for the quantification of h^2_{rare} in tropical maize germplasm. In addition, given maize's complex breeding history and strong population structure, some of the h^2_{rare} estimates may reflect covariance from fine-scale structure and lineage-related genetic sharing, rather than purely the direct effects of rare causal variants. Future work using diverse populations with well-defined pedigrees or large independent validation panels can rigorously quantify the heritability directly attributable to rare causal variants.

RESOURCE AVAILABILITY

Lead contact

Requests for resources and further information should be directed to and will be fulfilled by the lead contact, Jinsheng Lai (jlai@cau.edu.cn).

Materials availability

Constructs and transgenic seeds generated in this study are available from the lead contact upon reasonable request. Seeds of the inbred lines used in this study are available from W.S. upon reasonable request.

Data and code availability

WGS data were deposited in the NCBI SRA database under BioProject: PRJNA1417716. Raw data for the assay for transposase-accessible chromatin (ATAC) peak calling were obtained from published studies (NCBI: SRR11955290, SRR14327897, SRR10873335, SRR10873333, and SRR12321695). Scripts and pipelines to reproduce the analyses are available on GitHub: https://github.com/huangliang796-source/3367_maize_RVs_code_and_file/. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (32425041), the National Key Research and Development Program of China (2021YFD1200701-06), the Agriculture Science and Technology Major Project, the National Natural Science Foundation of China (32572347 and 32501975), the University-Local Collaborative Project (2024YLCAU005), CAU High-Talent Startup Funding (2024RC011 and 2023RC038), the 2115 Talent Development Program of China Agricultural University, and the Henan Provincial Key Research and Development Program (251111111500).

Transgenic maize seeds harboring the knockout of *ZmAGO5c* were generated by the Center for Crop Functional Genomics and Molecular Breeding, China Agricultural University. Prof. Feng Qin from China Agricultural University also provided valuable assistance in supplying the transgenic maize seeds. We thank the High-Performance Computing Platform of China Agricultural University for its support of large-scale computation. We also thank Beijing PARATERA Tech Co., Ltd. (https://paratera.com/), for providing HPC resources that have contributed to the research results reported in this study.

AUTHOR CONTRIBUTIONS

J.L. conceived, designed, and directed the project. W.S. and X.Z. collected the experimental materials and implemented the project. L.H. performed the majority of bioinformatic analyses with assistance from B.X. and C.B. Q. Lu and W.S. collected the phenotypic data of the inbred lines. Y.Y. and L.H. performed raw read mapping and variant calling. Y. Li performed maize transformation of *ZmThx* and collected phenotypic data of the mutant lines. All other authors contributed to data analysis or provided experimental materials. J.L., L.H., W.S., C.B., and B.X. participated in the interpretation of the results. L.H. and J.L. drafted the manuscript with input from all authors. L.H., C.B., B.X., and J.L. revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no conflicts of interest.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Plant materials and growth conditions
- METHOD DETAILS
  - Definition of rare alleles
  - Collection and statistical analysis of phenotypes
  - Whole genome sequencing and variant calling
  - Population structure and parameter estimation
  - Variant annotation and identification of natural knockout gene candidates
  - Single-variant GWAS analysis
  - Determination of independent GWAS signals and their phenotypic effects
  - Robustness and independence validation of GWAS signals
  - Estimation of heritability for 10 key agronomic traits
  - Sensitivity analysis of heritability estimation with different covariates, models and planting environments
  - Qst-Fst analysis between temperate and tropical genomes
  - Selection of random variants matched to trait-associated variants
  - Gene-based association test
  - Imputation accuracy of the 3,367-genome reference panel
  - Re-GWAS analysis using imputed genotypes
- QUANTIFICATION AND STATISTICAL ANALYSIS

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2026.09.063.

Received: March 13, 2026
Revised: July 14, 2026
Accepted: September 30, 2026

REFERENCES

1. Mcmullen, M.D., Kresovich, S., Villeda, H.S., Bradbury, P., Li, H., Sun, Q., Flint-Garcia, S., Thornsberry, J., Acharya, C., Bottoms, C., et al. (2009). Genetic properties of the maize nested association mapping population. Science *325*, 737–740. https://doi.org/10.1126/science.1174320.
2. Li, Y.X., Li, C., Bradbury, P.J., Liu, X., Lu, F., Romay, C.M., Glaubitz, J.C., Wu, X., Peng, B., Shi, Y., et al. (2016). Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. Plant J. *86*, 391–402. https://doi.org/10.1111/tpj.13174.
3. Liu, H.J., Wang, X., Xiao, Y., Luo, J., Qiao, F., Yang, W., Zhang, R., Meng, Y., Sun, J., Yan, S., et al. (2020). CUBIC: an atlas of genetic architecture promises directed maize improvement. Genome Biol. *21*, 20. https://doi.org/10.1186/s13059-020-1930-x.
4. Wei, X., Chen, M., Zhang, Q., Gong, J., Liu, J., Yong, K., Wang, Q., Fan, J., Chen, S., Hua, H., et al. (2024). Genomic investigation of 18,421 lines reveals the genetic architecture of rice. Science *385*, eadm8762. https://doi.org/10.1126/science.adm8762.
5. Xiao, Y., Liu, H., Wu, L., Warburton, M., and Yan, J. (2017). Genome-wide association studies in maize: praise and stargaze. Mol. Plant *10*, 359–374. https://doi.org/10.1016/j.molp.2016.12.008.
6. Zeng, R., Shi, Y., Guo, L., Fu, D., Li, M., Zhang, X., Li, Z., Zhuang, J., Yang, X., Zuo, J., et al. (2025). A natural variant of *COOL1* gene enhances cold tolerance for high-latitude adaptation in maize. Cell *188*, 1315–1329.e13. https://doi.org/10.1016/j.cell.2024.12.018.
7. Wang, X., Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., Yang, X., and Qin, F. (2016). Genetic variation in *ZmVPP1* contributes to drought tolerance in maize seedlings. Nat. Genet. *48*, 1233–1241. https://doi.org/10.1038/ng.3636.
8. 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. Mol. Breed. *28*, 511–526. https://doi.org/10.1007/s11032-010-9500-7.
9. Wang, B., Lin, Z., Li, X., Zhao, Y., Zhao, B., Wu, G., Ma, X., Wang, H., Xie, Y., Li, Q., et al. (2020). Genome-wide selection and genetic improvement during modern maize breeding. Nat. Genet. *52*, 565–571. https://doi.org/10.1038/s41588-020-0616-3.
10. Li, C., Guan, H., Jing, X., Li, Y., Wang, B., Li, Y., Liu, X., Zhang, D., Liu, C., Xie, X., et al. (2022). Genomic insights into historical improvement of heterotic groups during modern hybrid maize breeding. Nat. Plants *8*, 750–763. https://doi.org/10.1038/s41477-022-01190-2.
11. Kremling, K.A.G., Chen, S.Y., Su, M.H., Lepak, N.K., Romay, M.C., Swarts, K.L., Lu, F., Lorant, A., Bradbury, P.J., and Buckler, E.S. (2018). Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. Nature *555*, 520–523. https://doi.org/10.1038/nature25966.
12. Wang, T., He, W., Li, X., Zhang, C., He, H., Yuan, Q., Zhang, B., Zhang, H., Leng, Y., Wei, H., et al. (2023). A rice variation map derived from 10 548 rice accessions reveals the importance of rare variants. Nucleic Acids Res. *51*, 10924–10933. https://doi.org/10.1093/nar/gkad840.
13. Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W., Chai, Y., Yang, L., Liu, K., Lu, H., et al. (2013). A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). Nat. Genet. *45*, 957–961. https://doi.org/10.1038/ng.2673.
14. Dong, Y., Duan, S., Xia, Q., Liang, Z., Dong, X., Margaryan, K., Musayev, M., Goryslavets, S., Zdunić, G., Bert, P.F., et al. (2023). Dual domestications and origin of traits in grapevine evolution. Science *379*, 892–901. https://doi.org/10.1126/science.add8655.
15. 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*. Nat. Genet. *54*, 1736–1745. https://doi.org/10.1038/s41588-022-01184-y.
16. Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., Wang, B., Liu, Z., Chen, J., Li, W., et al. (2012). Genome-wide genetic changes during modern breeding of maize. Nat. Genet. *44*, 812–815. https://doi.org/10.1038/ng.2312.
17. Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., Zhan, Q., Lu, Y., Zhang, L., Huang, T., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat. Genet. *50*, 278–284. https://doi.org/10.1038/s41588-018-0041-z.
18. Wei, L., Zhang, X., Zhang, Z., Liu, H., and Lin, Z. (2018). A new allele of the *Brachytic2* gene in maize can efficiently modify plant architecture. Heredity. (Edinb) *121*, 75–86. https://doi.org/10.1038/s41437-018-0056-3.
19. Hu, J., Wang, Y., Fang, Y., Zeng, L., Xu, J., Yu, H., Shi, Z., Pan, J., Zhang, D., Kang, S., et al. (2015). A rare allele of *GS2* enhances grain size and grain yield in rice. Mol. Plant *8*, 1455–1465. https://doi.org/10.1016/j.molp.2015.07.002.
20. Lu, P., Guo, L., Wang, Z., Li, B., Li, J., Li, Y., Qiu, D., Shi, W., Yang, L., Wang, N., et al. (2020). A rare gain of function mutation in a wheat tandem kinase confers resistance to powdery mildew. Nat. Commun. *11*, 680. https://doi.org/10.1038/s41467-020-14294-0.
21. Xie, J., Guo, G., Wang, Y., Hu, T., Wang, L., Li, J., Qiu, D., Li, Y., Wu, Q., Lu, P., et al. (2020). A rare single nucleotide variant in *Pm5e* confers powdery mildew resistance in common wheat. New Phytol. *228*, 1011–1026. https://doi.org/10.1111/nph.16762.
22. Xu, Z., Zhou, Z., Cheng, Z., Zhou, Y., Wang, F., Li, M., Li, G., Li, W., Du, Q., Wang, K., et al. (2023). A transcription factor *ZmGLK36* confers broad resistance to maize rough dwarf disease in cereal crops. Nat. Plants *9*, 1720–1733. https://doi.org/10.1038/s41477-023-01514-w.
23. Zhang, Q., Wu, R., Hong, T., Wang, D., Li, Q., Wu, J., Zhang, H., Zhou, K., Yang, H., Zhang, T., et al. (2024). Natural variation in the promoter of *qRBG1/OsBZR5* underlies enhanced rice yield. Nat. Commun. *15*, 8565. https://doi.org/10.1038/s41467-024-52928-9.
24. Song, X.J., Kuroha, T., Ayano, M., Furuta, T., Nagai, K., Komeda, N., Segami, S., Miura, K., Ogawa, D., Kamura, T., et al. (2015). Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. Proc. Natl. Acad. Sci. USA *112*, 76–81. https://doi.org/10.1073/pnas.1421127112.
25. Yan, J., Kandianis, C.B., Harjes, C.E., Bai, L., Kim, E.H., Yang, X., Skinner, D.J., Fu, Z., Mitchell, S., Li, Q., et al. (2010). Rare genetic variation at *Zea mays crtRB1* increases β—carotene in maize grain. Nat. Genet. *42*, 322–327. https://doi.org/10.1038/ng.551.
26. Huang, R., Jin, Z., Zhang, D., Li, L., Zhou, J., Xiao, L., Li, P., Zhang, M., Tian, C., Zhang, W., et al. (2024). Rare variations within the serine/arginine-rich splicing factor *PtoRSZ21* modulate stomatal size to determine drought tolerance in *Populus*. New Phytol. *243*, 1776–1794. https://doi.org/10.1111/nph.19934.
27. Song, L., Liu, J., Cao, B., Liu, B., Zhang, X., Chen, Z., Dong, C., Liu, X., Zhang, Z., Wang, W., et al. (2023). Reducing brassinosteroid signalling enhances grain yield in semi-dwarf wheat. Nature *617*, 118–124. https://doi.org/10.1038/s41586-023-06023-6.
28. Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D.M., Burzynski-Chang, E.A., Fish, T.L., Stromberg, K.A., Sacks, G.L., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

fruit flavor. Nat. Genet. *51*, 1044–1051. https://doi.org/10.1038/s41588-019-0410-2.

29. Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A., et al. (2016). The allelic landscape of human blood cell trait variation and links to common complex disease. Cell *167*, 1415–1429.e19. https://doi.org/10.1016/j.cell.2016.10.042.
30. Marouli, E., Graff, M., Medina-Gomez, C., Lo, K.S., Wood, A.R., Kjaer, T.R., Fine, R.S., Lu, Y., Schurmann, C., Highland, H.M., et al. (2017). Rare and low-frequency coding variants alter human adult height. Nature *542*, 186–190. https://doi.org/10.1038/nature21039.
31. International Multiple Sclerosis Genetics Consortium; International Multiple Sclerosis Genetics Consortium (2018). Low frequency and rare-coding variation contributes to multiple sclerosis risk. Cell *175*, 1679–1687.e7. https://doi.org/10.1016/j.cell.2018.09.049.
32. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell *180*, 568–584.e23. https://doi.org/10.1016/j.cell.2019.12.036.
33. Akbari, P., Gilani, A., Sosina, O., Kosmicki, J.A., Khrimian, L., Fang, Y.Y., Persaud, T., Garcia, V., Sun, D., Li, A., et al. (2021). Sequencing of 640,000 exomes identifies *GPR75* variants associated with protection from obesity. Science *373*, eabf8683. https://doi.org/10.1126/science.abf8683.
34. Wang, Q., Dhindsa, R.S., Carss, K., Harper, A.R., Nag, A., Tachmazidou, I., Vitsios, D., Deevi, S.V.V., Mackay, A., Muthas, D., et al. (2021). Rare variant contribution to human disease in 281,104 UK biobank exomes. Nature *597*, 527–532. https://doi.org/10.1038/s41586-021-03855-y.
35. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., Mccarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753. https://doi.org/10.1038/nature08494.
36. Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A.U., Jiang, Y., Raghavan, S., et al. (2022). A saturated map of common genetic variants associated with human height. Nature *610*, 704–712. https://doi.org/10.1038/s41586-022-05275-y.
37. Wainschtein, P., Zhang, Y., Schwartzentruber, J., Kassam, I., Sidorenko, J., Fiziev, P.P., Wang, H., Mcrae, J., Border, R., Zaitlen, N., et al. (2026). Estimation and mapping of the missing heritability of human phenotypes. Nature *649*, 1219–1227. https://doi.org/10.1038/s41586-025-09720-6.
38. Dhindsa, R.S., Burren, O.S., Sun, B.B., Prins, B.P., Matelska, D., Wheeler, E., Mitchell, J., Oerton, E., Hristova, V.A., Smith, K.R., et al. (2023). Rare variant associations with plasma protein levels in the UK biobank. Nature *622*, 339–347. https://doi.org/10.1038/s41586-023-06547-x.
39. Rocheleau, G., Clarke, S.L., Auguste, G., Hasbani, N.R., Morrison, A.C., Heath, A.S., Bielak, L.F., Iyer, K.R., Young, E.P., Stitziel, N.O., et al. (2024). Rare variant contribution to the heritability of coronary artery disease. Nat. Commun. *15*, 8741. https://doi.org/10.1038/s41467-024-52939-6.
40. Jang, S.K., Evans, L., Fialkowski, A., Arnett, D.K., Ashley-Koch, A.E., Barnes, K.C., Becker, D.M., Bis, J.C., Blangero, J., Bleecker, E.R., et al. (2022). Rare genetic variants explain missing heritability in smoking. Nat. Hum. Behav. *6*, 1577–1586. https://doi.org/10.1038/s41562-022-01408-5.
41. Gazal, S., Loh, P.R., Finucane, H.K., Ganna, A., Schoech, A., Sunyaev, S., and Price, A.L. (2018). Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. Nat. Genet. *50*, 1600–1607. https://doi.org/10.1038/s41588-018-0231-8.
42. Pathan, N., Deng, W.Q., Di Scipio, M., Khan, M., Mao, S., Morton, R.W., Lali, R., Pigeyre, M., Chong, M.R., and Paré, G. (2024). A method to estimate the contribution of rare coding variants to complex trait heritability. Nat. Commun. *15*, 1245. https://doi.org/10.1038/s41467-024-45407-8.
43. Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., Ma, C., Fontanillas, P., Moutsianas, L., Mccarthy, D.J., et al. (2016). The genetic architecture of type 2 diabetes. Nature *536*, 41–47. https://doi.org/10.1038/nature18642.
44. Zeng, J., de Vlaming, R., Wu, Y., Robinson, M.R., Lloyd-Jones, L.R., Yengo, L., Yap, C.X., Xue, A., Sidorenko, J., Mcrae, A.F., et al. (2018). Signatures of negative selection in the genetic architecture of human complex traits. Nat. Genet. *50*, 746–753. https://doi.org/10.1038/s41588-018-0101-4.
45. Pritchard, J.K. (2001). Are rare variants responsible for susceptibility to complex diseases? Am. J. Hum. Genet. *69*, 124–137. https://doi.org/10.1086/321272.
46. Lappalainen, T., Li, Y.I., Ramachandran, S., and Gusev, A. (2024). Genetic and molecular architecture of complex traits. Cell *187*, 1059–1075. https://doi.org/10.1016/j.cell.2024.01.023.
47. Lu, J., and Wu, C.I. (2005). Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee. Proc. Natl. Acad. Sci. USA *102*, 4063–4067. https://doi.org/10.1073/pnas.0500436102.
48. Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., et al. (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature *493*, 216–220. https://doi.org/10.1038/nature11690.
49. Lozano, R., Gazave, E., Dos Santos, J.P.R., Stetter, M.G., Valluru, R., Bandillo, N., Fernandes, S.B., Brown, P.J., Shakoor, N., Mockler, T.C., et al. (2021). Comparative evolutionary genetics of deleterious load in sorghum and maize. Nat. Plants *7*, 17–24. https://doi.org/10.1038/s41477-020-00834-5.
50. Sun, S., Wang, B., Li, C., Xu, G., Yang, J., Hufford, M.B., Ross-Ibarra, J., Wang, H., and Wang, L. (2023). Unraveling prevalence and effects of deleterious mutations in maize elite lines across decades of modern breeding. Mol. Biol. Evol. *40*, msad170. https://doi.org/10.1093/molbev/msad170.
51. Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M.C., Wang, B., Campbell, M.S., Stein, J.C., Wei, X., Chin, C.S., et al. (2017). Improved maize reference genome with single-molecule technologies. Nature *546*, 524–527. https://doi.org/10.1038/nature22971.
52. Bukowski, R., Guo, X., Lu, Y., Zou, C., He, B., Rong, Z., Wang, B., Xu, D., Yang, B., Xie, C., et al. (2018). Construction of the third-generation *Zea mays* haplotype map. GigaScience *7*, 1–12. https://doi.org/10.1093/gigascience/gix134.
53. Haberer, G., Kamal, N., Bauer, E., Gundlach, H., Fischer, I., Seidel, M.A., Spannagl, M., Marcon, C., Ruban, A., Urbany, C., et al. (2020). European maize genomes highlight intraspecies variation in repeat and gene content. Nat. Genet. *52*, 950–957. https://doi.org/10.1038/s41588-020-0671-9.
54. Liu, H., Wang, X., Warburton, M.L., Wen, W., Jin, M., Deng, M., Liu, J., Tong, H., Pan, Q., Yang, X., et al. (2015). Genomic, transcriptomic, and phenomic variation reveals the complex adaptation of modern maize breeding. Mol. Plant *8*, 871–884. https://doi.org/10.1016/j.molp.2015.01.016.
55. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.
56. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65. https://doi.org/10.1038/nature11632.
57. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291. https://doi.org/10.1038/nature19057.

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

58. Grzybowski, M.W., Mural, R.V., Xu, G., Turkus, J., Yang, J., and Schnable, J.C. (2023). A common resequencing-based genetic marker data set for global maize diversity. Plant J. *113*, 1109–1121. https://doi.org/10.1111/tpj.16123.

59. Hufford, M.B., Seetharam, A.S., Woodhouse, M.R., Chougule, K.M., Ou, S., Liu, J., Ricci, W.A., Guo, T., Olson, A., Qiu, Y., et al. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science *373*, 655–662. https://doi.org/10.1126/science.abg5289.

60. Chia, J.M., Song, C., Bradbury, P.J., Costich, D., de Leon, N., Doebley, J., Elshire, R.J., Gaut, B., Geller, L., Glaubitz, J.C., et al. (2012). Maize hapmap2 identifies extant variation from a genome in flux. Nat. Genet. *44*, 803–807. https://doi.org/10.1038/ng.2313.

61. Saleheen, D., Natarajan, P., Armean, I.M., Zhao, W., Rasheed, A., Khetarpal, S.A., Won, H.H., Karczewski, K.J., O'Donnell-Luria, A.H., Samocha, K.E., et al. (2017). Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. Nature *544*, 235–239. https://doi.org/10.1038/nature22034.

62. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. Nature *590*, 290–299. https://doi.org/10.1038/s41586-021-03205-y.

63. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.

64. Sulem, P., Helgason, H., Oddson, A., Stefansson, H., Gudjonsson, S.A., Zink, F., Hjartarson, E., Sigurdsson, G.T., Jonasdottir, A., Jonasdottir, A., et al. (2015). Identification of a large set of rare complete human knockouts. Nat. Genet. *47*, 448–452. https://doi.org/10.1038/ng.3243.

65. Zeng, R., Li, Z., Shi, Y., Fu, D., Yin, P., Cheng, J., Jiang, C., and Yang, S. (2021). Natural variation in a type-A response regulator confers maize chilling tolerance. Nat. Commun. *12*, 4713. https://doi.org/10.1038/s41467-021-25001-y.

66. Ning, Q., Jian, Y., Du, Y., Li, Y., Shen, X., Jia, H., Zhao, R., Zhan, J., Yang, F., Jackson, D., et al. (2021). An ethylene biosynthesis enzyme controls quantitative variation in maize ear length and kernel yield. Nat. Commun. *12*, 5832. https://doi.org/10.1038/s41467-021-26123-z.

67. Yang, Q., Li, Z., Li, W., Ku, L., Wang, C., Ye, J., Li, K., Yang, N., Li, Y., Zhong, T., et al. (2013). CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc. Natl. Acad. Sci. USA *110*, 16969–16974. https://doi.org/10.1073/pnas.1310949110.

68. Guo, L., Wang, X., Zhao, M., Huang, C., Li, C., Li, D., Yang, C.J., York, A.M., Xue, W., Xu, G., et al. (2018). Stepwise cis-regulatory changes in *ZCN8* contribute to maize flowering-time adaptation. Curr. Biol. *28*, 3005–3015.e4. https://doi.org/10.1016/j.cub.2018.07.029.

69. Meng, X., Muszynski, M.G., and Danilevskaya, O.N. (2011). The FT – like *ZCN8* gene functions as a floral activator and is involved in photoperiod sensitivity in maize. Plant Cell *23*, 942–960. https://doi.org/10.1105/tpc.110.081406.

70. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., , et al.; Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Madden, P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. *44*, 369–375. https://doi.org/10.1038/ng.2213.

71. Li, Y., Huang, Y., Pan, L., Zhao, Y., Huang, W., and Jin, W. (2021). Male sterile 28 encodes an argonaute family protein essential for male fertility in maize. Chromosome Res. *29*, 189–201. https://doi.org/10.1007/s10577-021-09653-6.

72. Roussin-Léveillée, C., Silva-Martins, G., and Moffett, P. (2020). *Argonaute5* represses age-dependent induction of flowering through physical and functional interaction with miR156 in *Arabidopsis*. Plant Cell Physiol. *61*, 957–966. https://doi.org/10.1093/pcp/pcaa022.

73. Chen, Q., Samayoa, L.F., Yang, C.J., Bradbury, P.J., Olukolu, B.A., Neumeyer, M.A., Romay, M.C., Sun, Q., Lorant, A., Buckler, E.S., et al. (2020). The genetic architecture of the maize progenitor, teosinte, and how it was altered during maize domestication. PLoS Genet. *16*, e1008791. https://doi.org/10.1371/journal.pgen.1008791.

74. Jafari, F., Wang, B., Wang, H., and Zou, J. (2024). Breeding maize of ideal plant architecture for high-density planting tolerance through modulating shade avoidance response and beyond. J. Integr. Plant Biol. *66*, 849–864. https://doi.org/10.1111/jipb.13603.

75. Han, X., Li, J., Li, D., Guo, L., Wu, L., Liang, Y., Jia, H., Xia, J., Qin, C., Qin, W., et al. (2025). A copy number variation in the *ZmMADS1* promoter enhances maize adaptation to high altitudes. J. Integr. Plant Biol. *67*, 1994–1996. https://doi.org/10.1111/jipb.13924.

76. Liang, Y., Liu, Q., Wang, X., Huang, C., Xu, G., Hey, S., Lin, H.Y., Li, C., Xu, D., Wu, L., et al. (2019). *ZmMADS69* functions as a flowering activator through the *ZmRap2.7-ZCN8* regulatory module and contributes to maize flowering time adaptation. New Phytol. *221*, 2335–2347. https://doi.org/10.1111/nph.15512.

77. Zhang, Y., Yu, C., Lin, J., Liu, J., Liu, B., Wang, J., Huang, A., Li, H., and Zhao, T. (2017). *OsMPH1* regulates plant height and improves grain yield in rice. PLoS One *12*, e0180825. https://doi.org/10.1371/journal.pone.0180825.

78. Yang, J., Benyamin, B., Mcevoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569. https://doi.org/10.1038/ng.608.

79. Wainschtein, P., Jain, D., Zheng, Z., , et al.; TOPMed Anthropometry Working Group, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Cupples, L.A., Shadyab, A.H., McKnight, B., Shoemaker, B.M., Mitchell, B.D. (2022). Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nat. Genet. *54*, 263–273. https://doi.org/10.1038/s41588-021-00997-7.

80. Mancuso, N., Rohland, N., Rand, K.A., Tandon, A., Allen, A., Quinque, D., Mallick, S., Li, H., Stram, A., Sheng, X., et al. (2016). The contribution of rare variation to prostate cancer heritability. Nat. Genet. *48*, 30–35. https://doi.org/10.1038/ng.3446.

81. Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A.E., Lee, S.H., Robinson, M.R., Perry, J.R.B., Nolte, I.M., van Vliet-Ostaptchouk, J.V., et al. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. *47*, 1114–1120. https://doi.org/10.1038/ng.3390.

82. Romay, M.C., Millard, M.J., Glaubitz, J.C., Peiffer, J.A., Swarts, K.L., Casstevens, T.M., Elshire, R.J., Acharya, C.B., Mitchell, S.E., Flint-Garcia, S.A., et al. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. *14*, R55. https://doi.org/10.1186/gb-2013-14-6-r55.

83. Sun, G., Yu, H., Wang, P., Lopez-Guerrero, M., Mural, R.V., Mizero, O.N., Grzybowski, M., Song, B., van Dijk, K., Schachtman, D.P., et al. (2023). A role for heritable transcriptomic variation in maize adaptation to temperate environments. Genome Biol. *24*, 55. https://doi.org/10.1186/s13059-023-02891-3.

84. Leinonen, T., Mccairns, R.J.S., O'Hara, R.B., and Merilä, J. (2013). Qst – fst comparisons: evolutionary and ecological insights from genomic heterogeneity. Nat. Rev. Genet. *14*, 179–190. https://doi.org/10.1038/nrg3395.

85. Whitlock, M.C., and Guillaume, F. (2009). Testing for spatially divergent selection: comparing qst to fst. Genetics *183*, 1055–1063. https://doi.org/10.1534/genetics.108.099812.

86. Turchin, M.C., Chiang, C.W.K., Palmer, C.D., Sankararaman, S., Reich, D., and Genetic Investigation of ANthropometric Traits (GIANT)

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

Consortium, and Hirschhorn, J.N. (2012). Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat. Genet. *44*, 1015–1019. https://doi.org/10.1038/ng.2368.

87. Xin, T., Zhang, Z., Zhang, Y., Li, X., Wang, S., Wang, G., Li, H., Wang, B., Zhang, M., Li, W., et al. (2025). Recessive epistasis of a synonymous mutation confers cucumber domestication through epitranscriptomic regulation. Cell *188*, 4517–4529.e15. https://doi.org/10.1016/j.cell.2025.06.007.

88. Lan, Y., Xia, Z., Shao, Q., Lin, P., Lu, J., Xiao, X., Zheng, M., Chen, D., Dou, Y., and Xie, Q. (2025). Synonymous mutations promote tumorigenesis by disrupting m6A-dependent mRNA metabolism. Cell *188*, 1828–1841.e15. https://doi.org/10.1016/j.cell.2025.01.026.

89. Wu, Y., Li, D., Hu, Y., Li, H., Ramstein, G.P., Zhou, S., Zhang, X., Bao, Z., Zhang, Y., Song, B., et al. (2023). Phylogenomic discovery of deleterious mutations facilitates hybrid potato breeding. Cell *186*, 2313–2328.e15. https://doi.org/10.1016/j.cell.2023.04.008.

90. Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., NHLBI GO Exome Sequencing Project—ESP Lung Project Team, Christiani, D.C., Wurfel, M.M., and Lin, X. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am. J. Hum. Genet. *91*, 224–237. https://doi.org/10.1016/j.ajhg.2012.06.007.

91. Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One *6*, e19379. https://doi.org/10.1371/journal.pone.0019379.

92. Tian, H., Yang, Y., Yi, H., Xu, L., He, H., Fan, Y., Wang, L., Ge, J., Liu, Y., Wang, F., et al. (2021). New resources for genetic studies in maize (*Zea mays L.*): A genome-wide maize6H-60K single nucleotide polymorphism array and its application. Plant J. *105*, 1113–1122. https://doi.org/10.1111/tpj.15089.

93. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. *42*, 348–354. https://doi.org/10.1038/ng.548.

94. Salvi, S., Sponza, G., Morgante, M., Tomes, D., Niu, X., Fengler, K.A., Meeley, R., Ananiev, E.V., Svitashev, S., Bruggemann, E., et al. (2007). Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc. Natl. Acad. Sci. USA *104*, 11376–11381. https://doi.org/10.1073/pnas.0704145104.

95. Yang, N., Wang, Y., Liu, X., Jin, M., Vallebueno-Estrada, M., Calfee, E., Chen, L., Dilkes, B.P., Gui, S., Fan, X., et al. (2023). Two teosintes made modern maize. Science *382*, eadg8940. https://doi.org/10.1126/science.adg8940.

96. Cheng, S., Feng, C., Wingen, L.U., Cheng, H., Riche, A.B., Jiang, M., Leverington-Waite, M., Huang, Z., Collier, S., Orford, S., et al. (2024). Harnessing landrace diversity empowers wheat breeding. Nature *632*, 823–831. https://doi.org/10.1038/s41586-024-07682-9.

97. Guo, D., Li, Y., Lu, H., Zhao, Y., Kurata, N., Wei, X., Wang, A., Wang, Y., Zhan, Q., Fan, D., et al. (2025). A pangenome reference of wild and cultivated rice. Nature *642*, 662–671. https://doi.org/10.1038/s41586-025-08883-6.

98. Li, N., He, Q., Wang, J., Wang, B., Zhao, J., Huang, S., Yang, T., Tang, Y., Yang, S., Aisimutuola, P., et al. (2023). Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. Nat. Genet. *55*, 852–860. https://doi.org/10.1038/s41588-023-01340-y.

99. Schulthess, A.W., Kale, S.M., Liu, F., Zhao, Y., Philipp, N., Rembe, M., Jiang, Y., Beukert, U., Serfling, A., Himmelbach, A., et al. (2022). Genomics-informed prebreeding unlocks the diversity in genebanks for wheat improvement. Nat. Genet. *54*, 1544–1552. https://doi.org/10.1038/s41588-022-01189-7.

100. Agarwala, V., Flannick, J., Sunyaev, S., and GoT2D Consortium, and Altshuler, D. (2013). Evaluating empirical bounds on complex disease genetic architecture. Nat. Genet. *45*, 1418–1427. https://doi.org/10.1038/ng.2804.

101. Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R., Zhang, F., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature *557*, 43–49. https://doi.org/10.1038/s41586-018-0063-9.

102. Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.A., Zhang, H., Liu, Z., Shi, M., et al. (2020). Pan-genome of wild and cultivated soybeans. Cell *182*, 162–176.e13. https://doi.org/10.1016/j.cell.2020.05.023.

103. Niu, J., Ma, S., Zheng, S., Zhang, C., Lu, Y., Si, Y., Tian, S., Shi, X., Liu, X., Naeem, M.K., et al. (2023). Whole-genome sequencing of diverse wheat accessions uncovers genetic changes during modern breeding in china and the united states. Plant Cell *35*, 4199–4216. https://doi.org/10.1093/plcell/koad229.

104. Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., Zhou, H., Yu, S., Xu, C., Li, X., et al. (2008). Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. Nat. Genet. *40*, 761–767. https://doi.org/10.1038/ng.143.

105. Jiao, C., Xie, X., Hao, C., Chen, L., Xie, Y., Garg, V., Zhao, L., Wang, Z., Zhang, Y., Li, T., et al. (2025). Pan-genome bridges wheat structural variations with habitat and breeding. Nature *637*, 384–393. https://doi.org/10.1038/s41586-024-08277-0.

106. Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B., Brown, P.J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J.C., et al. (2009). The genetic architecture of maize flowering time. Science *325*, 714–718. https://doi.org/10.1126/science.1174276.

107. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An expanded view of complex traits: from polygenic to omnigenic. Cell *169*, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038.

108. Liao, H., Zhao, X., Ren, K., Guo, L., Li, Z., Liu, Z., Zhang, X., Su, T., Fu, D., Zhang, Z., et al. (2026). Rewiring an E3 ligase enhances cold resilience and phosphate use in maize. Nature *653*, 831–839. https://doi.org/10.1038/s41586-026-10142-1.

109. Tian, J., Wang, C., Xia, J., Wu, L., Xu, G., Wu, W., Li, D., Qin, W., Han, X., Chen, Q., et al. (2019). Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. Science *365*, 658–664. https://doi.org/10.1126/science.aax5482.

110. Huang, Y., Wang, H., Zhu, Y., Huang, X., Li, S., Wu, X., Zhao, Y., Bao, Z., Qin, L., Jin, Y., et al. (2022). *THP9* enhances seed protein content and nitrogen-use efficiency in maize. Nature *612*, 292–300. https://doi.org/10.1038/s41586-022-05441-2.

111. Romero Navarro, J.A., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S., Preciado, E., Terron, A., Delgado, H.V., Vidal, V., et al. (2017). A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat. Genet. *49*, 476–480. https://doi.org/10.1038/ng.3784.

112. Charlesworth, D., and Willis, J.H. (2009). The genetics of inbreeding depression. Nat. Rev. Genet. *10*, 783–796. https://doi.org/10.1038/nrg2664.

113. Chen, J., Wang, Z., Tan, K., Huang, W., Shi, J., Li, T., Hu, J., Wang, K., Wang, C., Xin, B., et al. (2023). A complete telomere-to-telomere assembly of the maize genome. Nat. Genet. *55*, 1221–1231. https://doi.org/10.1038/s41588-023-01419-6.

114. Crisp, P.A., Marand, A.P., Noshay, J.M., Zhou, P., Lu, Z., Schmitz, R.J., and Springer, N.M. (2020). Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. Proc. Natl. Acad. Sci. USA *117*, 23991–24000. https://doi.org/10.1073/pnas.2010250117.

115. Sun, Y., Dong, L., Zhang, Y., Lin, D., Xu, W., Ke, C., Han, L., Deng, L., Li, G., Jackson, D., et al. (2020). 3D genome architecture coordinates trans and cis regulation of differentially expressed ear and tassel genes in maize. Genome Biol. *21*, 143. https://doi.org/10.1186/s13059-020-02063-7.

116. Tu, X., Mejía-Guerra, M.K., Valdes Franco, J.A., Tzeng, D., Chu, P.Y., Shen, W., Wei, Y., Dai, X., Li, P., Buckler, E.S., et al. (2020). Reconstructing the maize leaf regulatory network using Chip-seq data of 104 transcription factors. Nat. Commun. *11*, 5089. https://doi.org/10.1038/s41467-020-18832-8.

117. Bates, D., Mächler, M., Bolker, B.M., and Walker, S.C. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67*, 1–48. https://doi.org/10.18637/jss.v067.i01.

118. Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit: a map-reduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

119. Chen, S. (2025). Fastp 1.0: an ultra-fast all-round tool for fastq data quality control and preprocessing. IMETA *4*, e70078. https://doi.org/10.1002/imt2.70078.

120. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

121. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., Mccarthy, S.A., Davies, R.M., et al. (2021). Twelve years of samtools and bcftools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

122. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.

123. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109.

124. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.

125. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., Depristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and vcftools. Bioinformatics *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

126. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X.Y., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SNPeff: SNPs in the genome of drosophila melanogaster strain *w1118; iso-2; iso-3*. Fly *6*, 80–92. https://doi.org/10.4161/fly.19695.

127. Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M., and Ng, P.C. (2016). SIFT missense predictions for genomes. Nat. Protoc. *11*, 1–9. https://doi.org/10.1038/nprot.2015.123.

128. Loh, P.R., Palamara, P.F., and Price, A.L. (2016). Fast and accurate long-range phasing in a UK biobank cohort. Nat. Genet. *48*, 811–816. https://doi.org/10.1038/ng.3571.

129. Browning, B.L., Tian, X., Zhou, Y., and Browning, S.R. (2021). Fast two-stage phasing of large-scale sequence data. Am. J. Hum. Genet. *108*, 1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005.

130. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of Chip-seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

131. Forde, A., Hemani, G., and Ferguson, J. (2023). Review and further developments in statistical corrections for winner's curse in genetic association studies. PLoS Genet. *19*, e1010546. https://doi.org/10.1371/journal.pgen.1010546.

132. Stephens, M. (2017). False discovery rates: a new deal. Biostatistics *18*, 275–294. https://doi.org/10.1093/biostatistics/kxw041.

133. Zhang, C., Dong, S.S., Xu, J.Y., He, W.M., and Yang, T.L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics *35*, 1786–1788. https://doi.org/10.1093/bioinformatics/bty875.

134. Porebski, S., Bailey, L.G., and Baum, B.R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol. Biol. Rep. *15*, 8–15. https://doi.org/10.1007/bf02772108.

135. Oehlert, G.W. (1992). A note on the delta method. Am. Stat. *46*, 27–29. https://doi.org/10.1080/00031305.1992.10475842.

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Maize association panel of 3,367 inbred lines	This study	N/A
NCRPIS association panel	Romay et al.[82]	https://cbsusrv04.tc.cornell.edu/users/panzea/download.aspx?filegroupid=6
Deposited data		
Raw data of whole genome sequence	This paper	PRJNA1417716
ATAC data of B73 Root	Crisp et al.[114]	SRR11955290
ATAC data of B73 Seedling	Hufford et al.[59]	SRR14327897
ATAC data of B73 Ear	Sun et al.[115]	SRR10873335
ATAC data of B73 Tassel	Sun et al.[115]	SRR10873333
ATAC data of B73 Leaves	Tu et al.[116]	SRR12321695
Experimental models: Organisms/strains		
Zea mays: ZmAGO5c-1	This paper	N/A
Zea mays: ZmAGO5c-2	This paper	N/A
Zea mays: ZmThx-1	This paper	N/A
Zea mays: ZmThx-2	This paper	N/A
Software and algorithms		
lme4	Bates et al.[117]	https://cran.r-project.org/web/packages/lme4/index.html
GATK4	Mckenna et al.[118]	https://github.com/broadinstitute/gatk
Fastp	Chen et al.[119]	https://github.com/OpenGene/fastp
BWA	Li and Durbin[120]	https://github.com/lh3/bwa
Samtools	Danecek et al.[121]	https://github.com/samtools/samtools
PLINK	Purcell et al.[122]	https://github.com/chrchang/plink-ng
ADMIXTURE	Alexander et al.[123]	https://dalexander.github.io/admixture/
GCTA	Yang et al.[124]	https://yanglab.westlake.edu.cn/software/gcta
VCFtools	Danecek et al.[125]	https://vcftools.github.io/
SNPeff	Cingolani et al.[126]	https://pcingola.github.io/SnpEff/
SIFT4G	Vaser et al.[127]	https://github.com/rvaser/sift4g
COJO	Yang et al.[70]	https://yanglab.westlake.edu.cn/software/gcta/#COJO
SKAT-O	Lee et al.[90]	https://cran.r-project.org/web/packages/SKAT/index.html
Eagle	Loh et al.[128]	https://alkesgroup.broadinstitute.org/Eagle/
Beagle	Browning et al.[129]	https://faculty.washington.edu/browning/beagle/beagle.html
MACS2	Zhang et al.[130]	https://github.com/macs3-project/MACS
Winnerscurse	Forde et al.[131]	https://github.com/zrmacc/WinCurse/
Ashr	Stephens et al.[132]	https://github.com/stephens999/ashr
PopLDdecay	Zhang et al.[133]	https://github.com/BGI-shenzhen/PopLDdecay
R	R foundation	https://www.r-project.org

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Plant materials and growth conditions

CRISPR-Cas9-generated loss-of-function mutant lines of *ZmAGO5c* and *ZmThx* in maize were used in this study. The *ZmAGO5c* loss-of-function mutants were generated by the Center for Crop Functional Genomics and Molecular Breeding at China Agricultural University, and the *ZmThx* loss-of-function mutants were generated by the Maize Genetic Transformation Platform at China Agricultural University. Mutant lines for both genes were derived in the LH244 genetic background.

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

Each genotype was grown in a single plot under field conditions in Zhuozhou, Hebei Province, China (115.9°E, 39.4°N), with wild-type LH244 plants interspersed throughout the experimental field. Row spacing was 0.600 m, and plant spacing was 0.278 m. Five border rows were planted around the experimental field to minimize edge effects. Phenotypic data were collected from individual plants at the VT-R1 developmental stage.

METHOD DETAILS

Definition of rare alleles

We defined rare alleles as variants with a minor allele frequency (MAF) below 5%. This threshold was selected considering the current sample size at the thousand-scale and the widely accepted standard within plant genetics.[11,12] Unless otherwise stated, the term "rare variants" refers to the broad set with MAF < 0.05.

Collection and statistical analysis of phenotypes

Phenotypic data for 10 key agronomic traits were collected from two growing seasons of field trials conducted in Zhuozhou, Hebei Province (115.9°E, 39.4°N), China, in 2020.

In each growing season, all 3,367 maize inbred lines were evaluated under field conditions and randomly assigned to field positions. Each genotype was represented by a single plot in each environment. The row and column spacing were 0.600 m and 0.278 m, respectively, and five border rows were planted around the experimental field to minimize edge effects. Details on the measurement methods of each trait can be found in Table S3A. To improve measurement accuracy, three individual plants in the middle of each plot were selected and recorded. Phenotypic values for these 10 traits across environments were fitted with a linear mixed model to calculate best linear unbiased predictions (BLUPs) using the lme4 package in R. The model was specified as follows:

$$Y_{ij} = u + E_i + G_j + (GE)_{ij} + \mathcal{E}_{ij}$$

where u is the overall mean, E_i is the random main effect of the i-th environment, the G_j is the random genetic main effect of the j-th inbred line, $(GE)_{ij}$ is the random genotype-by-environment (G×E) interaction effect, and $\mathcal{E}_{ij}$ is the random residual error. All random effects were assumed to follow independent normal distributions with mean zero and their respective variance components.

Phenotypic data for ear length were collected in winter 2019 in Sanya, Hainan Province, China. These accessions were planted under open-pollinated field conditions using the same randomized complete block design (RCBD) as described above. Three individual plants in the middle of each plot were sampled and recorded. Meanwhile, the chilling resistance phenotype of maize was measured following the protocol described previously.[65] In brief, maize seedlings at the V2 stage were subjected to a chilling stress treatment at 4°C for 2–4 days, and the injured area was quantified. Due to the lack of cross-environmental replication and a limited number of phenotyped samples, both traits were excluded from further association analysis and heritability estimation.

Whole genome sequencing and variant calling

Three individual plants were chosen from each accession, and their young leaves were collected for sequencing library construction. Genomic DNA was extracted following the cetrimonium bromide (CTAB) method.[134] DNA fragments with an average insert size of 350 bp were selected, and paired-end sequencing was conducted on these fragments using the DNBSEQ-T7 sequencing platform.

Raw reads were initially subjected to quality assessment and filtering using fastp software. After quality control, the resulting clean reads were mapped to the B73 reference genome (AGPv4)[51] via the Burrows-Wheeler backtrack alignment algorithm (BWA-aln).[120] Reads with a mapping quality score > 20 were retained, and PCR duplicates were marked using the Genome Analysis Toolkit[118] (GATK). Subsequently, variants for each accession were identified using the HaplotypeCaller module in GATK, based on the duplicate-marked BAM files. Finally, population-level variant information was integrated and generated using the GenomicsDBImport and GenotypeGVCFs modules in GATK.

High-quality SNPs and Indels were obtained via the Hard-filter method with different parameters using the SelectVariants module of GATK. Specifically, SNP variants meeting the following criteria were retained: QD > 2.0, QUAL > 30.0, SOR < 3.0, FS < 60.0, MQ > 40.0, MQRankSum > -12.5, and ReadPosRankSum > -8.0. For Indel variants, the retention criteria were set as: QD > 2.0, QUAL > 30.0, SOR < 10.0, FS < 150.0, MQ > 20.0, MQRankSum > -15.0, and ReadPosRankSum > -10.0. Since all accessions used for variant calling were maize inbred lines, variants with the observed number of heterozygous genotypes exceeding 0.51 times the expected number of heterozygous genotypes under Hardy-Weinberg equilibrium (HWE) were removed. Here, 0.51 corresponds to five times (1-F), where F represents the estimated inbreeding coefficient calculated following the pipeline employed in the 3,010 Rice Genomes Project.[101]

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

Population structure and parameter estimation

Phylogenetic trees were constructed using PHYLIP software based on the genetic distance matrix. The genetic distance was calculated using PLINK[122] based on synonymous SNPs with minor allele frequency (MAF) > 5% and missing rate < 50%. Population genetic structure was inferred using ADMIXTURE with a SNP set matching the phylogenetic analysis, with a MAF > 1% filtering threshold. The putative number of ancestral populations (*K*) was set from 2 to 24. Principal component analysis (PCA) was performed using GCTA[124] software with SNPs (MAF > 5%, missing rate < 50%) after LD pruning.

LD decay was calculated using PopLDdecay software for each subgroup. Private variants were defined as SNPs with an allele frequency (AF) > 1% that were absent in all other subgroups. Genetic divergence among these subgroups was quantified using the Weir and Cockerham weighted Fst metric, which was calculated with VCFtools via pairwise comparisons between subgroups, with a 100-kb analysis window and a 100-kb step size.

Recent genetic exchange was quantified using f_2 variants. Given the inbred nature of the materials in this study, f_2 variants were defined as SNPs detected exclusively in two genomes, rather than the standard definition of minor allele count (MAC) = 2. After identifying f_2 variants between each pair of groups, the total number of f_2 variants in the corresponding two subgroups was used to normalize the inter-group f_2 variant counts.

Variant annotation and identification of natural knockout gene candidates

Functional annotation was performed using SNPeff[126] software with the B73 gene annotation file downloaded from Ensembl Plants (release-49/gff3/zea_mays/). Both SNPs and Indels annotated with the term "splice_acceptor_variant" or "splice_donor_variant" were classified as essential splice site variants. SNPs annotated with "stop_gained", "missense_variant", and "synonymous_variant" were categorized as stop-gained, missense, and synonymous variants, respectively. InDels annotated with "frameshift_variant" were defined as frameshift variants, while those annotated with ″inframe_deletion" or ″inframe_insertion" were designated as inframe variants. For variants that were assigned to multiple functional categories, higher priority was assigned to categories with higher impact. We additionally performed functional annotation of all missense variants using SIFT4G. The UniRef90 protein database was used as the reference to predict the functional impact of each variant on B73 protein sequences. Among all annotated missense variants, 264,120 were classified as deleterious and 363,094 as tolerated.

In addition, we also annotated intergenic variants via overlap with ATAC-seq and epigenetic mark peaks to improve functional interpretation. All peaks were called with MACS2 from high-quality reads aligned to the B73 AGPv4 reference genome.

Variants annotated with the "High"-effect term by SNPeff were classified as potential protein-truncating variants (PTVs). These variants primarily include stop-gained, stop-lost, essential splice site variants, and frameshift variants. Genes containing any PTV within the gene body were designated as potential natural knockout genes. Genes carrying NKO variants were further grouped based on the highest-frequency NKO variant per gene. GO term enrichment analysis was performed using the singular enrichment analysis (SEA) module on the AgriGO web server, with the B73v4 Ensembl gene set as the reference background. *P* values were calculated using Fisher's exact test.

Single-variant GWAS analysis

For each trait, single-variant GWAS analysis was performed using EMMAX[93] for variants with a minimum of 10 phenotypic observations for the minor allele. This model can incorporate cryptic relatedness, other fixed effects, and population stratification via the kinship matrix and the first four principal components, respectively. The PCA performed for population structure analysis was also included as covariates in GWAS, and the kinship matrix was estimated using the emmax-kin subprogram based on the same variant set under a more stringent missing rate threshold of < 10%.

The number of effective markers was estimated via LD pruning analysis in PLINK, using a sliding window encompassing 50 variants and a step size of 20 variants; variants with an $r^2 > 0.1$ were pruned out. The effective variants number ranged from 4,507,001 to 8,159,960. The GWAS threshold was determined as 1 divided by the number of effective markers, and the final threshold was set to 10^{-7}.

Determination of independent GWAS signals and their phenotypic effects

For each trait, independent GWAS signals were detected using conditional and joint multiple-SNP analysis[70] (COJO). The core principle of this analysis is to test whether nearby variants can still reach the significance threshold after accounting for the most significant marker as a control. If so, this indicates that there are multiple independent GWAS signals at this locus. This analysis was performed in GCTA, using input summary statistics derived from EMMAX-based GWAS results following a format conversion. The genotypes of the 3,367 maize inbred lines were used as the LD reference panel for this analysis. The significance threshold for the conditional analysis was set to match the genome-wide significance threshold used in the single-variant GWAS. Finally, only variants that were significant in both the original GWAS and the conditional analysis were retained as independent GWAS signals for subsequent analyses. The beta values from joint-analysis of COJO results were used to determine the effect size and effect direction of these independent GWAS signals directly, which has considered the influence of other significant trait-related signals.

To further assess the impact of winner's curse on effect size estimation, we applied five summary statistic-based correction methods to adjust the observed effect sizes from the winnerscurse and ashr R packages, with raw effect sizes and standard errors

as input. Conditional likelihood correction was restricted to genome-wide significant variants, while the other four methods were run across all GWAS markers.

Robustness and independence validation of GWAS signals

We first performed subgroup-aware robustness testing of representative association signals via leave-one-subgroup-out analysis. Using subgroup membership from Figure 1, we sequentially excluded each subgroup and re-ran GWAS with the original kinship matrix and top four principal components (PCs) as covariates.

To assess the stability of the rare variant proportion among all detected association signals, we conducted two complementary replication GWAS: one with varying PC counts as covariates, and another adding subgroup membership as a fixed effect alongside the top four PCs. Independent signals were identified via COJO correction following the primary analysis pipeline.

Finally, we evaluated the conditional stability of three representative signals (*ZmAGO5c, ZmMADS1* and *qPH6*). For each focal locus, genotypes of the lead variant or its co-chromosome independent significant variants were included as covariates alongside PCs to control for their effects; self-imputed genotypes were used to handle missing data.

Estimation of heritability for 10 key agronomic traits

Phenotypic variance explained by different genomic components was estimated using the genome-based restricted maximum likelihood (REML) approach implemented in GCTA. Specifically, genomic relationship matrices (GRMs) were constructed separately for rare and common variants using the make-grm-inbred command. Only variants with minor allele count (MAC) > 10 and missing rate < 0.5 were retained for this analysis. To partition rare- and common-variant genetic variance conditional on each other, these two GRMs were then fitted jointly in a single GREML model to estimate the heritability proportions attributable to rare and common variants for the 10 key agronomic traits, with the top 10 principal components (derived from LD-pruned common SNPs) included as covariates. Whole-genome heritability (h^2_{WGS}) was calculated as the sum of the rare- and common-variant variance components estimated from the joint GREML model. Standard errors for the directly estimated h^2_{rare} and h^2_{common} were obtained from the REML procedure implemented in GCTA.

The dynamic changes in heritability between temperate and tropical genomes were estimated with the same pipeline. In this analysis, variants were processed separately within the temperate and tropical subgroups, with allele frequencies and missing rates computed exclusively within each corresponding group. The standard errors for the derived parameter, the h^2_{rare} /h^2_{WGS} ratio, were approximated with the delta method.[135] *P*-values for all comparisons between different point estimates were determined using normal-approximation Z-tests. A threshold of $P < 0.05$ was considered statistically significant.

Sensitivity analysis of heritability estimation with different covariates, models and planting environments

To assess the impact of covariate settings on heritability estimation, we performed PCA using different variant datasets and estimated heritability with varying numbers of principal components (PCs). Specifically, we also used the top 2, 3, 5, 20, and 50 PCs derived from LD-pruned common SNPs, as well as the top 10, 20, and 50 PCs calculated from LD-pruned rare SNPs. Additionally, we incorporated membership of the 19 subgroups as a covariate in the heritability model. We further estimated heritability using differently processed phenotypic values, including Z-score normalization and the exclusion of samples with significantly deviated phenotypic values identified based on the Z-score normalization.

To mitigate LD-related and MAF-related biases of phenotypic effects of causal variants (i.e., to relax the assumption that variants with different global LD levels and MAFs exhibit similar phenotypic effects), we further stratified variants into multiple bins according to LD score and MAF, constructed genomic relationship matrices (GRMs) independently for each bin, and jointly estimated their contributions to phenotypic variance. LD scores were calculated within a 5-Mb window with GCTA-LDS, and variants were stratified in two different groups based on the LD scores of each variant. The estimation for h^2_{rare} and h^2_{common} were obtained by summing variance components from bins corresponding to their respective MAF categories. The standard errors for the estimation of h^2_{rare} and h^2_{common} were also approximated with the delta method. Given the substantially smaller sample size relative to near-million-scale human genetic studies,[37,79] we applied a maximum of six bins (two LD scores based bins × three MAF-based bins) in this analysis.

We also assessed the influence of planting environments on heritability estimates by comparing results derived from phenotypic values measured in individual environments with those based on BLUP-derived phenotypic values. For all the analyses in this section, only one parameter was varied at a time.

Furthermore, we performed within-subgroup phenotype permutation for all seven traits with $h^2_{rare} > 0$. Phenotypes were randomly shuffled within each of the 19 subgroups, and GREML analysis was subsequently conducted with subgroup labels and the top 10 PCs retained as covariates. The entire permutation procedure was repeated 3,000 times.

Qst-Fst analysis between temperate and tropical genomes

The null FST model was generated using synonymous SNPs annotated by SNPEff. Only SNPs with MAF > 0.05 in both temperate and tropical populations were retained for this analysis. QST-FST tests were additionally performed for each trait using the R scripts provided by Whitlock and Guillaume.[85]

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

Selection of random variants matched to trait-associated variants

To measure the random frequency dynamics between temperate and tropical genomes, random variants were selected that matched the analyzed trait-associated variants on a per-variant basis.[86] Prior to this selection, genome-wide data had been pruned to eliminate the impact of highly correlated variants: a sliding window encompassing 500 variants was used, with a step size of 50 variants and an r^2 threshold of 0.8. For each trait-associated variant, control variants were matched for MAF (±0.2% tolerance) and LDscores stratum (five equal bins), where LDscores were calculated with all 3,367 genomes. Finally, 10,000 sets of random variants were drawn.

Gene-based association test

After annotation with SNPeff, variants with high or moderate effects on specific genes were regarded as functional variants for the corresponding genes. Variants located within 1 kb upstream of a gene were classified as promoter variants for that gene. We defined four distinct variant sets based on MAF and functional categories: (i) functional + synonymous variants with MAF < 1%; (ii) functional + synonymous variants with MAF < 0.1%; (iii) promoter variants with MAF < 1%; and (iv) promoter variants with MAF < 0.1%. After rare variant aggregation and quantification, we retained only genes satisfying the following criteria for subsequent association analysis: (i) containing no fewer than two qualifying variants; (ii) having a minimum of 10 phenotyped samples that carry any of these quantified rare variants.

Association analyses between aggregated rare variants in genes or loci and traits were performed using the SKAT[90] software, with the SKAT-O model applied for these tests. Confounding effects were controlled using the *SKAT_NULL_emmaX()* function, which accounts for the effects of the kinship matrix and the first four principal components. The calculation of kinship and principal components was consistent with that used in the single-variant GWAS analysis.

On average, 23,995 (range: 12,200–34,855) gene-trait pairs were tested for each variant set. The significance threshold was determined using Bonferroni correction as 0.05 / 23,995 (a value slightly greater than 2×10^{-6}), with the final threshold set at 2×10^{-6}. Additionally, to avoid the influence of LD among rare variants, for significant genes associated with the same trait (from the same variant set) and located within 100-kb region to each other, only the most significant gene was retained.

Imputation accuracy of the 3,367-genome reference panel

Due to the impact of missing rate on imputation accuracy, only SNP with missing rate less than 0.2 were included in reference panel. The Eagle[128] software was used to phase the genotypes of the 3,367 maize inbred lines for each chromosome. The genetic map was set at a density of 1 centimorgan (cM) per 1 megabase (Mb) on the physical map. Only variants with a minor allele count > 5 were included.

To evaluate the imputation performance of this reference panel, 200 samples were randomly selected from the 3,367 genomes, with even distribution across 19 subgroups. All samples with an identity-by-state (IBS) > 0.95 with these 200 samples were excluded from the reference panel to avoid overestimating imputation accuracy due to the inclusion of highly similar accessions. After downsampling the identified SNPs of these randomly selected samples to the previously used 680K polymorphic sites that revealed by GBS,[82] Beagle was used to impute these samples with the modified reference panel. The imputed genotypes were then compared with the true genotypes (from GATK) to assess imputation accuracy. The average coefficient of determination (r^2) of polymorphic variants across the 200 samples was used to quantify imputation accuracy. When estimating the imputation accuracy of this reference panel for temperate and tropical genomes, the same 200 samples were re-analyzed, with only the corresponding samples (temperate or tropical) included in each respective analysis.

The same pipeline was used for other specific accuracy evaluations of the reference panel, but different modifications were applied to the reference panel. To evaluate the imputation performance of this reference panel without tropical genomes, all tropical genomes and IBS-related samples were removed from the reference panel. To evaluate the influence of variant missing rates on the imputation performance of the reference panel, all variants were included without applying a missing rate filter.

Re-GWAS analysis using imputed genotypes

The original genotypic data included approximately 3,000 samples from the U.S. National Maize Inbred Seed Bank.[82] These raw genotypes were converted to the B73v4 coordinates, and variants with discordant genomic base were removed. The genotypes were then imputed using the 3,367-genome reference panel with Beagle. Only variants with a minor allele frequency (MAF) > 1% were included for further analysis, as imputation accuracy declines with decreasing MAF. Phenotypic data for growing degree days to silking were used for GWAS analysis. GWAS was re-conducted using EMMAX, following the same pipeline as the single-variant GWAS analysis in this study. We noted here that it should be preferable to treat imputed genotypes as dosages rather than discrete hard genotype calls for reference or alternative alleles, as this treatment would enhance the statistical power when imputation accuracy is not perfectly high. Here the discrete genotype calls were used for GWAS analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using R and the specialized software described in the Method Details. Four classes of significance tests were applied without formally testing whether the data met their underlying assumptions: (1) Two-group

Please cite this article in press as: Huang et al., Deep sequencing of 3,367 maize genomes uncovers substantial contribution of rare variants to agronomic traits, Cell (2026), https://doi.org/10.1016/j.cell.2026.09.063

comparisons used two-tailed t-tests. (2) Two-tailed t-tests from linear models were used for two-group comparisons, adjusted for population structure and membership. (3) Observed values were compared with random permutation replicates; significance was determined by the deviation of observed values relative to the fitted normal distribution of the permutation outputs. (4) Normal-approximation Z-tests were used either to compare pairs of estimated parameters or to test whether individual estimates differed from zero; One-sided standard-normal *P* values assessed directional differences. Details are provided in the figures and figure legends. Asterisks above bars indicate significant differences at $P < 0.05$. For all box plots presented in the figures, the horizontal line inside the box represents the median; the lower and upper edges of the box represent the 25th and 75th percentiles.