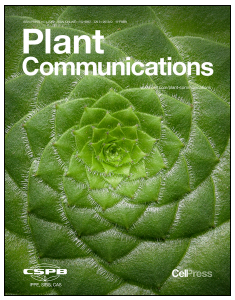
Journal Pre-proof



GWAS-based population genetic analysis identifies *bZIP29* as a heterotic gene in maize

Jie Zhang, Riliang Gu, Xinxin Miao, Renate H. Schmidt, Zhenxiang Xu, Jiawen Lu, Yuting Ma, Tao Yang, Pingxi Wang, Yangyang Liu, Xiaoli Wang, Xuemei Du, Nannan Zheng, Sihan Zhen, Chengyong Liang, Yuxin Xie, Yongrui Wu, Lin Li, Jochen C. Reif, Yong Jiang, Jianhua Wang, Junjie Fu, Hongwei Zhang

PII: S2590-3462(25)00051-3

DOI: https://doi.org/10.1016/j.xplc.2025.101289

Reference: XPLC 101289

To appear in: *PLANT COMMUNICATIONS*

Received Date: 27 August 2024

Revised Date: 10 November 2024

Accepted Date: 19 February 2025

Please cite this article as: Zhang, J., Gu, R., Miao, X., Schmidt, R.H, Xu, Z., Lu, J., Ma, Y., Yang, T., Wang, P., Liu, Y., Wang, X., Du, X., Zheng, N., Zhen, S., Liang, C., Xie, Y., Wu, Y., Li, L., Reif, J.C, Jiang, Y., Wang, J., Fu, J., Zhang, H., GWAS-based population genetic analysis identifies *bZIP29* as a heterotic gene in maize, *PLANT COMMUNICATIONS* (2025), doi: https://doi.org/10.1016/j.xplc.2025.101289.

GWAS-based population genetic analysis identifies *bZIP29* as a heterotic gene in maize

Jie Zhang[1,2,§], Riliang Gu[2,§], Xinxin Miao[3,§], Renate H Schmidt[4,§], Zhenxiang Xu[2], Jiawen Lu[1], Yuting Ma[1], Tao Yang[5,a], Pingxi Wang[1,b], Yangyang Liu[1,c], Xiaoli Wang[1], Xuemei Du[2], Nannan Zheng[2], Sihan Zhen[1], Chengyong Liang[3], Yuxin Xie[1], Yongrui Wu[5], Lin Li[3], Jochen C Reif[4], Yong Jiang[4,*], Jianhua Wang[2,*], Junjie Fu[1,*], and Hongwei Zhang[1,*]

1, State Key Laboratory of Crop Gene Resources and Breeding, National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

2 State Key Laboratory of Maize Bio-breeding, Beijing Innovation Center for Crop Seed Technology of Ministry of Agriculture and Rural Affairs, College of Agronomy and Biotechnology, China Agricultural University, Beijing 100193, China

3 National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

4 Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland 06466, Germany

5 National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China

a, present address: State Key laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu, 611130, China

b, present address: School of Agriculture, Henan Institute of Science and Technology, Xinxiang 453003, China

c, present address: Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

§ These authors contribute equally to this work

*Correspondence: Hongwei Zhang: zhanghongwei@caas.cn; Junjie Fu: fujunjie@caas.cn; Jianhua Wang: wangjh63@cau.edu.cn; Yong Jiang: jiang@ipk-gatersleben.de

Short Summary

Six candidate heterotic genes, including *bZIP29*, were identified in a maize hybrid population, with *bZIP29* demonstrating significant heterotic effects in both the hybrid population and transgenic-derived hybrid lines. Functional characterization using tsCUT&Tag assays revealed that bZIP29 protein directly binds to a target gene regulated by its associated eQTL, as well as to six additional genes within co-expression modules. These findings offer new insights into the critical roles of key genes and their regulatory interactions in enhancing hybrid performance.

Abstract

Understanding the role of heterotic genes in contributing to heterosis is essential for advancing hybrid breeding. We analyzed plant height (PH), ear height (EH), and transcriptomic data from a maize hybrid population. GWAS identified that dominance effects of QTL play a significant role in hybrid traits and mid-parent heterosis. By integrating GWAS, expression GWAS (eGWAS) analysis, and module eGWAS, six candidate heterotic genes underlying six QTL were prioritized, including one QTL spanned the *bZIP29* gene. In the hybrid population, *bZIP29* exhibited additive expression and dominance effects for both traits and mid-parent heterosis, with its favorable allele correlating positively with PH and EH. *bZIP29* demonstrated dominance or over-dominance patterns in the hybrids derived from crosses between transgenic and wildtype lines, contingent upon its expression. tsCUT&Tag assay revealed that bZIP29 protein bound directly to a gene regulated by its associated eQTL and six genes within expression modules governed by their meQTL. Regulatory networks involving *bZIP29* were more extensive in hybrid sub-populations compared to the parental population. This study offers insights into key heterotic genes and networks underpinning the robust growth of hybrid maize.

Keywords: hybrid maize, mid-parent heterosis, dominance, heterotic gene, *bZIP29*

INTRODUCTION

Hybrids generated by crossing two homozygous lines frequently exhibit superior agronomic traits compared to their parental lines. Enhancing hybrid performance is a crucial strategy for increasing crop production (Labroo et al., 2021). Future breeding efforts require a comprehensive understanding of the molecular regulation of diverse crop traits, particularly the quantitative trait locus/loci (QTL) and genes linked to agronomic characteristics (Wallace et al., 2018). Most studies elucidating gene functions used inbred lines as genetic background, which provide limited reference value for the understanding of phenotypic variation in hybrid population, since the QTL and genetic effects underlying trait performance are often different between inbred lines and hybrid lines (Guo et al., 2014; Ma et al., 2022). Therefore, studying the molecular basis and regulation of hybrid performance is essential for optimizing the use of heterosis.

Conventional linkage-based studies using bi-parental populations, such as immortalized F_2 populations, had identified many QTL associated with hybrid performance across various crop species, including rice (Hua et al., 2002; Zhou et al., 2012), maize (Guo *et al.*, 2014; Tang et al., 2010), and cotton (Liu et al., 2011). Notably, different QTL associated with hybrid performance and mid-parent heterosis (MPH) have been identified when distinct traits or populations were analyzed (Fiévet et al., 2018; Xiao et al., 2021). Bi-parental populations inherently offer low mapping resolution, which complicates the identification of candidate genes underlying hybrid performance (Liu et al., 2020). As an alternative, hybrid populations generated by crossing diverse inbred lines have been used to dissect hybrid performance (Seymour et al., 2016). Such populations have significantly advanced our understanding of the genetic regulation of hybrid performance due to their high genetic diversity and mapping resolution (Huang et al., 2015; Seymour *et al.*, 2016; Xiao *et al.*, 2021). These analysis have even contributed to the expansion theories of heterosis (Bonnafous et al., 2018; Jiang et al., 2017; Xie et al., 2022) and revealed that hybrid performance and MPH possess complicated genetic architectures shaped by numerous genes and their interactions (Jiang *et al.*, 2017). While extensive genetic analysis can identify QTL associated with heterosis or hybrid performance, such investigations offer only limited insights into the regulatory patterns of heterotic QTL.

Systems genetics approaches are powerful tools for investigating the genes,

pathways, and networks underlying traits in animals and plants (Civelek and Lusis, 2014). Among these approaches, expression QTL (eQTL) mapping is widely used to establish genome-wide regulatory relationships between genomic variants and gene expressions (Fu et al., 2013; Li et al., 2013b). Genomic loci that influence the expression of multiple genes are typically defined as *trans*-acting hotspots, which are likely to play central roles in the regulatory networks governing target traits (Li et al., 2020; Tan et al., 2022). However, identifying *trans*-acting associations can be challenging, as they are often sensitive to confounding factors stemming from biological and technical variations, as well as differences in tissues and developmental stages (Kolberg et al., 2020). Transcriptome profiles can be organized into dozens of modules, derived through dimensionality reduction analysis (Cote et al., 2022; Rotival et al., 2011). Recent studies indicated that module-eQTL (meQTL) mapping, which treats the eigenvalues of co-expression modules as composite traits, can improve the statistical power for detecting *trans*-regulatory hotspots (Kolberg *et al.*, 2020; Rotival *et al.*, 2011; Sun et al., 2023) and prioritize candidate genes associated with trait performance (Momozawa et al., 2018; Tang et al., 2021). Systems genetics studies in plants have primarily focused on establishing the relationships among genotypes, gene expressions, and phenotypes in inbred populations (Fu *et al.*, 2013; Tang *et al.*, 2021), with limited application hybrid populations.

Several hypotheses, including dominance and over-dominance, have been proposed to explain heterosis (Hochholdinger and Baldauf, 2018). According to the dominance hypothesis, favorable and unfavorable alleles exhibit dominance and recessive effects on plant growth in hybrid lines (Jones, 1917). East proposed that the interaction of heterogeneous alleles caused over-dominance (East, 1936). However, evidence linking allele expression to heterotic effects remains limited. Springer and Stupar (Springer and Stupar, 2007) proposed that gene expression levels in inbred lines may be fixed above or below the optimal range, and mid-parent expression levels in hybrid lines may adjust gene expression level to within the optimal range. Consequently, hybrids may show dominance or over-dominance advantages due to mid-parent expression levels. Despite this, limited experimental evidence supports this model.

To identify heterotic genes and regulatory networks that are associated with hybrid performance and heterosis, we constructed a hybrid population based on a North Carolina II (NCII) design. As PH and ear height (EH) showed strong heterosis and high heritability in maize (Hu et al., 2017; Li et al., 2021b), these traits were particularly

well suited for studying hybrid performance and heterosis. Additionally, the RNA expression profiles from this population provided valuable insights into the regulation of PH and EH. Integrating RNA expression data with genetic mapping facilitates deciphering the molecular regulation of heterosis. Specifically, this study included the following objectives: (1) conducting GWAS to uncover the genetic basis of hybrid performance for PH and EH; (2) identifying candidate genes and their regulatory networks underlying QTL through eGWAS and meGWAS analysis; and (3) demonstrate how the expression of the heterotic gene *bZIP29* in hybrid lines influences heterosis. The identified heterotic genes and their associated networks contribute to our understanding of the mechanism of heterosis.

RESULTS

Identification of QTL for PH and EH in the hybrid population

We developed an NCII population by crossing 78 maternal lines to four paternal lines, generating 314 hybrid lines (including two hybrid lines obtained by crossing parental lines). Whole genome re-sequencing data of the 82 parental lines identified 4,628,240 high-quality SNPs (Fig. S1A). The heterozygous rates of parental lines were below 0.002 (Fig. S1B), while those of hybrid lines ranged from 0.19 to 0.42 (Fig. S1C). Phylogenetic analysis showed a relatively balanced distribution among the parental lines, with no highly divergent clades, and placed the four male parents in separate branches (Fig. S2A). The hybrid lines grouped into four divergent clusters, each corresponding to one of the four parental inbred lines (Fig. S2B).

Phenotypic analysis revealed strong correlations between replicates within and among environments (Fig. S3A-B), as well as between PH and EH (Fig. S4), indicating the reliability of the phenotypic data. Comparisons between hybrid performance and mid-parent phenotypes revealed strong heterosis for both traits in all hybrids (Fig. 1A), with MPH ranging from 45 cm to 116 cm for PH (Fig. S5A), and from 19 cm to 68 cm for EH (Fig. S5B). The broad-sense heritability (H^2) of PH and EH is 0.96, while that of MPH is 0.70 for PH and 0.77 for EH (Fig. 1B), indicating that hybrid lines showed stable performance across environments.

We focused on the genetic dissection of hybrid performance, but not MPH, because hybrid performance had higher H^2 than its derivative MPH (Li *et al.*, 2021b) (Fig. 1B), and both traits and their derivative MPH values showed strong correlation (Fig. S5C-D). A mixed linear model incorporating both the additive and dominance effects of each

SNP, totaling 2,324,652 SNPs, was applied for GWAS and referred to as Additive-dominance (AD) model (see Materials and Methods). The AD model effectively reduces potential type I error caused by population structure and cryptic relatedness (Fig. S6A-D). A total of 11 additive-QTL and 41 dominance-QTL were identified for PH, while 21 additive-QTL and six dominance-QTL were identified for EH (Fig. S7A-D, Table S1). Comparison of the physical positions of QTL for PH and EH revealed that one additive-QTL and three dominance-QTL were associated with both traits (Fig. 1C; Table S1). The phenotypic variance explained (PVE) of dominance-QTL were significantly higher than those of additive-QTL for both traits (Fig. 1D). We then screened the QTL regions for annotated genes or genes with reported homologs involved in plant growth or development as candidate genes. Six additive- and 15 dominance-QTL for PH, as well as eight additive- and four dominance-QTL for EH, encompassed such candidate genes related to plant growth or development (Fig. S7A-D; Table S1).

We found that the favor scores, representing the cumulative effects of favorable QTL genotypes, were significantly correlated with hybrid performance for both traits (Fig. 1E-F), supporting the reliability of these QTL. The d/a (degree of dominance) for most dominance-QTL exceeded 1.0, whereas the d/a values for all additive-QTL ranged from -1.0 to 1.0 (Fig. S8A-B), consistent with the genetic effects of these QTL (Table S1), and indicating that the AD model can reliably differentiate between additive- and dominance-QTL. Further, QTL with higher d/a tended to contribute more significantly to phenotypic variances (Fig. S8A-B). The genetic effects of these QTL were further supported by the phenotypic comparisons among different genotype groups. Specifically, significant differences in hybrid performances were observed among all three genotype groups for additive-QTL (Fig. S9A-B). In contrast, for dominance-QTL, hybrids with heterozygous genotypes generally exhibited higher trait values than those with homozygous unfavorable alleles, with no significant differences compared to hybrids with homozygous favorable alleles (Fig. S9C-D). The MPH values of hybrids carrying heterozygous genotypes at dominance-QTL were generally higher than those of hybrids with homozygous alleles, a pattern not observed for additive-QTL (Fig. S10).

Genome-wide identification of eQTL and eQTL hotspots

To explore the regulation of hybrid performance, we conducted a genome-wide scan for eQTL by performig eGWAS on 28,706 expressed genes. A modified AD model

testing the combined effect (additive and dominance effects) of each marker was used for eGWAS so as to reduce the computation load of eGWAS. In total, 36,946 eQTL associated with 18,517 eGenes (eQTL-associated genes), accounting for 64.51% of the total analyzed genes, were identified across the genome, including 10,696 (28.95%) local-eQTL and 26,250 (71.05%) distant-eQTL (Fig. 2A). Notably, local-eQTL were distributed along the diagonal lines, and distant-eQTL were dispersed throughout the genome (Fig. S11). The number of eQTL regulating housekeeping genes (Lin et al., 2014) were significantly less than those regulating non-housekeeping genes (Fig. S12, *P*-value of Wilcoxon rank sum test = 8.29E-6). Among the eGenes, 10,569 (57.08%) were regulated by distant-eQTL, 5,936 (32.06%) by local-eQTL, and 2,012 (10.87%) by both local- and distant-eQTL (Fig. 2B). Out of the local-eQTL, 42.84% were located within 10 kb of transcription start sites (TSS) of their target eGenes (Fig. 2C). Most eGenes were associated with a single eQTL (Fig. 2D). Comparison of the PVEs of lead SNPs in eQTL revealed that local-eQTL had larger contributions to the variances of eGene expressions than distant-eQTL (Fig. 2E).

To identify potential regulatory nodes and candidate genes associated with hybrid performance, we analyzed eQTL hotspots co-localizing with QTL. A genome-wide scan of global eQTL identified 552 eQTL hotspots, of which, 42 were located within 200 kb of 35 QTL (Fig.2F, Table S2). For example, hotspot C10_Hot26 co-located with PH_Dom3 (a dominance-QTL for PH) encompassing the putative candidate gene *GRF5* (Fig. S13A). C10_Hot26 regulated the expression of 69 eGenes (Table S3; Fig. 2G). Correlation analysis revealed that eGenes involved in stress response and reproductive development, such as *GSNAP*, *MNS1*, *WRKY11*, *CBR1*, *PPT1*, and *ALY3*, were positively correlated with *GRF5* expression but negatively corelated or uncorrelated with PH. In contrast, eGenes involved in nutrient metabolism and signal transduction, such as *SHR5*, *ATG8C*, and *CML*, were negatively correlated with *GRF5* expression but positively corelated with PH (Fig. S13B). Comparisons of gene expression between hybrids with different genotypes at PH_Dom3 showed that the hybrids carrying the homozygous favorable genotype (TT) had reduced expression of *GRF5*, *CBR1*, and *ALY3*, and elevated expression of *ATG8C* and *CML* (Fig. S13C). These results align with the hypothesis that the favorable allele of PH_Dom3 may increase PH by modulating the expression of *GRF5* and associated eGenes. Similar results were obtained for other hotspots. Hotspot C4_Hot21 co-located with PH_Dom18, encompassing the putative candidate gene *SBP7* and associated with 49 downstream

eGenes, including key genes involved in nutrient metabolism and plant development (Table S3; Fig. 2G). Hotspot C7_Hot08, co-located with PH_Dom34, spanned the putative candidate gene *CRF4*. Hotspot C7_Hot08 was associated with 36 eGenes, including essential genes involved in photosynthesis, nutrient metabolism, and the cell cycle (Table S3; Fig. 2G). These findings suggest that the candidate genes within the QTL hotspots may regulate PH by influencing the expression of downstream eGenes.

Identification of co-expression modules and module-eQTL

To further elucidate the regulatory networks underlying hybrid performance, co-expression modules in the hybrid population were identified by independent component (IC) analysis. As a result, 205 ICs were identified as co-expression modules (Table S4). The number of genes assigned to IC modules ranged from 10 to 841, with a median of 250 (Fig. S14). Correlations between the latent features of ICs and the tested traits revealed that 25 ICs were significantly correlated with PH, 47 ICs with EH, and 20 ICs with both traits (Pearson correlation, $P < 0.01$) (Fig. 3A, Table S4). Additionally, more ICs showed positive correlations with the tested traits than negative correlations (Fig. 3A). GO enrichment analysis of the top ten positively correlated and top five negatively correlated modules indicated that genes in positively correlated modules were significantly enriched in biological processes related to plant vegetative growth, while genes in negatively correlated modules were enriched in processes associated with plant reproductive growth and stress response (Table S4; Fig. 3B-C).

To identify additional key regulators not detected in eGWAS, we conducted module eGWAS (meGWAS) using the AD model for the latent features of the 205 ICs. A total of 234 meQTL associated with 84 ICs were identified, including 142 meQTL associated with 46 trait-correlated ICs (Fig. 3D, Table S5). We employed two strategies to confirm the reliability of these meQTL: Firstly, we compared the physical positions of these meQTL with those of eQTL hotspots, finding that 125 (53.42%) meQTL were located within 200 kb of eQTL hotspots (Table S5). Secondly, we tested whether eQTL regulating genes within a particular IC were significantly enriched in the 200 kb flanking regions of a corresponding meQTL by carrying out Fisher's exact tests (see Materials and Methods), and found significant enrichments in the flanking regions of 66 (28.21%) meQTL (Table S5). These comparisons indicated a high degree of concordance between the eGWAS and meGWAS findings; however, meGWAS also detected meQTL that did not co-locate with any eQTL hotspot.

We then focused on the meQTL associated with trait-correlated ICs. Among the 104 meQTL associated with PH-correlated ICs, 28 (26.92%) co-located with QTL for PH, 40 (38.46%) overlapped with eQTL hotspots, and eQTL underlying genes assigned to the relevant IC modules showed enrichment in the flanking regions of 33 (31.73%) meQTL (Fig. 3E). For the 101 meQTL associated with EH-correlated ICs, nine (8.91%) co-located with QTL for EH, 45 (44.55%) overlapped with eQTL hotspots, and eQTL regulating genes assigned to the relevant IC modules showed enrichment in the flanking regions of 37 (36.63%) meQTL (Fig. 3F). The regulatory networks constructed from genes assigned to IC modules may illustrate the regulatory relationships of QTL that co-locate with meQTL. For instance, PH_Add1 (an additive-QTL for PH) co-located with an meQTL (IC187_C10M9366861) that regulates the latent feature of IC187, and six genes of IC187 were regulated by an eQTL hotspot (C10_Hot10) in the same region (Fig. S15A). Additionally, 29 genes of IC187 showed co-expression with *GATA7*, a candidate gene within PH_Add1 (Fig. S15B). Genes assigned to IC187 were significantly enriched in biological processes such as carbohydrate metabolism, microtubule assembly, and auxin biosynthesis (Fig. 3B). Given that IC187 was positively correlated with PH, it is plausible that PH_Add1 and its candidate gene *GATA7* may function as a regulatory hub to control PH by up-regulating genes in module IC187.

Similarly, an meQTL (IC176_C8M170930342) associated with the latent feature of IC176 co-located with PH_Dom38 (a dominance-QTL for PH), EH_Dom6 (a dominance-QTL for EH), and an eQTL hotspot (C8_Hot40) regulating eight eGenes. The QTL encompassed the putative candidate gene *MYB163* (Fig. S16A). Among the 126 genes in IC176, four were regulated by C8_Hot40 (Fig. S16B). GO enrichment analysis revealed that the genes in this module were enriched in biological processes related to plant reproductive development and stress response (Fig. 3C). Since IC176 was negatively correlated with EH, it is plausible that PH_Dom38 and EH_Dom6 may regulate PH and EH by repressing genes in module IC176.

bZIP29 **represents a candidate gene for hybrid growth**

To select a candidate gene for functional analysis, we focused on the three dominance-QTL shared by both PH and EH: PH_Dom10 (EH_Dom3), PH_Dom38 (EH_Dom6), and PH_Dom29 (EH_Dom5) (Table S1), as dominance-QTL contributed more to variations of hybrid performance than additive-QTL (Fig. S8-10). We found

292 that PH_Dom10 encompassed *bZIP29*, and PH_Dom38 encompassed *MYB163* (Table
293 S1). Notably, PH_Dom10 co-located with two closely linked eQTL hotspots
294 (C1_Hot60 and C1_Hot61; Table S2) that merged into a single hotspot (C1_Hot60*)
295 when the hot_scan distance cutoff was increased to > 50 kb (Fig. 4A). Four meQTL
296 associated with IC181 and IC176 were also located in this region (Fig. 3C; 4A; Table
297 S5). Thus, *bZIP29* was selected as a candidate gene for hybrid growth.

298 *bZIP29* encoded a basic leucine zipper (bZIP) transcription factor and showed
299 strong linkage disequilibrium (LD) with the peak SNPs of the QTL, eQTL, and meQTL
300 (Fig. 4B). Notably, *bZIP29* was significantly and positively correlated with both traits,
301 and most target eGenes regulated by C1_Hot60* were also significantly correlated with
302 both traits and *bZIP29* expression (Fig. 4C). Among these eGenes, *AGD1* (involved in
303 cytoskeleton organization), *SLC35F5* (encoding solute carrier), *bHLH48* (participating
304 in developmental regulation), *RPS17* (30S ribosomal protein), and *PAP1* (associated
305 with chloroplast development) were positively correlated with both traits and *bZIP29*
306 expression, whereas genes involved in disease resistance or stress responses (such as
307 *GSTF8*, *UBP12*) were negatively correlated with *bZIP29* expression (Fig. 4C). A local-
308 eQTL corresponding to *bZIP29* may drive differential regulation of this gene in the
309 hybrid population and influence the expression of downstream eGenes.

310 A common peak SNP (C1M296443013) with dominance effects was significantly
311 associated with both traits (Fig. 4A). Comparisons of hybrid performance and MPH
312 values among hybrids with different genotypes at C1M296443013 revealed that those
313 with the homozygous favorable genotype (CC) or the heterozygous genotype (CT)
314 exhibited higher phenotypic and MPH values than those with the homozygous
315 unfavorable genotype (TT) (Fig. 4D-E; Table S6), indicating dominance effects at this
316 locus for both traits and their MPH values. In contrast, *bZIP29* expressions in hybrids
317 with the heterozygous genotype were significantly lower than in those with the
318 homozygous favorable genotype, yet significantly higher than those with the
319 homozygous unfavorable genotype (Fig. 4F), suggesting an additive effect on *bZIP29*
320 expression. Notably, *bZIP29* expression was significantly associated with MPH for
321 each trait (Fig.S17), supporting that *bZIP29* exerts a heterotic effect on both PH and
322 EH. Further comparisons of eGene expressions among the three genotype groups
323 showed that hybrids with the homozygous favorable or heterozygous genotype had
324 lower expression levels of negatively correlated eGenes (*GSTF8* and *UBP12*) (Fig.
325 S18A) and higher expression levels of positively correlated eGenes (*SLC35F5*,

bHLH48, and *RPS17*) (Fig. S18B) compared to those with the homozygous unfavorable genotype. The above results suggest that the favorable allele of this QTL may enhance PH and EH by influencing the expression of *bZIP29* and its downstream eGenes.

To further investigate the functional variations associated with *bZIP29*, LD analysis was performed between the peak SNP, C1M296443013, and all potential variations within the gene body and upstream region of *bZIP29*. Two variations, C1M296588268 (a SNP marker) and C1M296596845 (an INDEL marker), showed strong LD with the peak SNP (Fig. S19A). Notably, both variations are located within the upstream region of *bZIP29*, suggesting their potential role as cis-regulatory elements influencing *bZIP29* expression. Among the haplotypes clustered around these two variations, four haplotype-based genotypes (HapGs) had frequencies exceeding 0.05 in the hybrid population. Subsequent phenotypic and expression-level comparisons of *bZIP29* across these HapGs revealed that hybrids carrying the heterozygous haplotype (HapG2) exhibited significantly higher plant height (PH), ear height (EH), and *bZIP29* expression levels compared to those with the homozygous haplotype (HapG4) (Fig. S19B). These findings highlight C1M296588268 and C1M296596845 as strong candidates for validating the functional variations of *bZIP29*.

Validation of the heterotic effects of *bZIP29* on PH and EH

To validate the dominance effects of *bZIP29*, we obtained two independent loss-of-function mutants (DEL lines) generated with the CRISPR-Cas9 system, along with two independent over-expression lines (OE lines) (Yang et al., 2022). The expression levels of *bZIP29* in WT lines were significantly higher than those in both DEL lines and DEL-derived F_1 lines, and significantly lower than those in both OE lines and OE-derived F_1 lines (Fig. 5A-B), reflecting the additive expression pattern observed in the hybrid population (Fig. 4E). DEL-derived F1 lines were taller than DEL lines, but smaller than the WT line (Fig. 5C-D), indicating an additive effect on both PH and EH. However, over-expression of *bZIP29* resulted in increased PH and EH, displaying an over-dominance and dominance pattern in OE1F1 and OE2F1, respectively (Fig. 5E-F). The differences in effects between the two OE-derived F1 lines may be related to the expression level of *bZIP29*.

A tsCUT&Tag assay with two high-quality replicates (Fig. 5G) was performed to identify downstream genes targeted by the bZIP29 protein using maize inbred line B73. A total of 3,363 genes were targeted by the bZIP29 protein in both biological replicates.

These target genes were significantly enriched in various biological processes, including metabolism, stress response, development, and plant hormones (Fig. S20). Notably, 28 of these genes were encompassed by 21 QTL (Table S7). Among the 18 eGenes regulated by C1_Hot60*, the promoter of one target eGene (*PTR3*) was bound by the bZIP29 protein. Additionally, the promoters of six of the 186 genes in IC176 and IC181 were bound by the bZIP29 protein (Fig. 5H-I, Table S7), indicating significant enrichment of eGenes (*P*-value of Chi-square test = 1.24E-3) and module genes (Chi-square test *P* = 4.05E-27) among the genes bound by the bZIP29 protein. Furthermore, other eGenes and module genes may be indirectly regulated by *bZIP29*. Since genes in IC181 were enriched for meristem development, while genes in IC176 were enriched for stress response and meiosis (Fig. 3C), *bZIP29* may positively regulate PH and EH by activating the genes involved in vegetative growth and repressing those related to stress response and reproductive development. A complex regulatory network, constructed by integrating eGenes and genes in the two ICs, along with genes targeted by the bZIP29 protein (Fig. 5J), suggests that *bZIP29* participate in multiple biological processes.

bZIP29-centered regulatory networks in hybrids are more extensive and complex than that in the parental lines

To further investigate the role of *bZIP29* in heterosis, we divided the entire hybrid population into four sub-populations based on the four male testers and compared the *bZIP29*-centered co-expression networks of these sub-populations with that of the parental population. Co-expression analysis using Pearson correlation showed that the numbers of genes co-expressed with *bZIP29* in the hybrid sub-populations are 9.67 to 25.42 times greater than that in the parental population (Fig. 6A). Additionally, we found that the co-expression networks of *bZIP29* in the hybrid and parental populations shared both common and unique genes (Fig. 6A). GO enrichment analysis revealed that the co-expressed genes in the hybrid sub-populations were significantly enriched for common GO terms related to metabolic processes, stress responses, developmental processes, and plant hormones. In contrast, co-expressed genes in the parental population were enriched for only four GO terms (Fig. 6B). Furthermore, the co-expression networks of the hybrid sub-populations exhibited similar patterns when compared to each other, and were more extensive and complex than that of the parental population (Fig. 6C). These findings suggest that *bZIP29* contributes to heterosis and hybrid performance in PH and EH by participating in complex regulatory networks that

display partial conservation among hybrid sub-populations.

DISCUSSION

Systems genetics approaches have been used to identify genes associated with kernel size and benzoxazinoid biosynthesis in maize (Li et al., 2022; Wang et al., 2018), seed size in cotton (Zhao et al., 2023), and flowering time and growth related traits in *Brassica napus* (Li et al., 2018). In this study, we applied systems genetics for the first time to a hybrid population to dissect the molecular basis of hybrid performance in maize. We conducted a GWAS of PH and EH using a diverse hybrid population, and integrated these results with eGWAS and meGWAS. Candidate genes for 41% of the detected QTL (33 out of 79 QTL) were identified based on literature information on annotated genes within QTL regions. Six genes in six QTL regions were prioritized by assessing their co-localization with eQTL and meQTL. Furthermore, we evaluated the correlations between the expression levels of candidate genes with the tested traits. Additionally, we analyzed their co-expression patterns with genes located in eQTL hotspots and/or in ICs that are regulated by meQTL co-localized within the identified QTL regions. The systems genetics approach was suitable to uncover candidate genes for hybrid performance and heterosis and also provided insights into the molecular networks of candidate genes. Notably, one candidate gene, *bZIP29*, underlying a dominance-QTL for both PH and EH and overlapping with an eQTL hotspot and four meQTL, was validated in transgenic experiments to show heterotic effects for the tested traits. Co-expression analysis revealed that *bZIP29* was involved in more extensive and complex networks in hybrid sub-populations than in the parental population.

The advantages of using PH and EH to study hybrid performance and heterosis in maize

To study the genetic basis of hybrid performance and heterosis in maize, we chose two traits showing high H^2 to mitigate effects by environmental factors. The identification of 79 QTL (Table S1) backed this selection. PH and EH are strongly correlated traits (Li et al., 2017; Yin et al., 2022) (Fig. S4), and it was therefore reasoned that QTL in common to both traits would be particularly strong candidates for the identification of genes associated with hybrid performance. The identification of candidate genes *MYB163* and *bZIP29* for two out of four common QTL confirmed this notion. Although MPH is used frequently to study heterosis, it is a derived composite

trait calculated based on trait performance, and is therefore even more likely to be influenced by unexpected factors than hybrid performance (Fig. 1B) (Li *et al.*, 2021b). Hence, we focused on PH and EH in the hybrid population rather than on MPH. Nonetheless, given the high correlation between hybrid performance with MPH (Fig. S5B-C) (Li *et al.*, 2021b), the obtained insights may also be relevant for heterosis as we exemplified for the QTL spanning *bZIP29*. Gene expression in elongating internodes was informative to reveal the regulation of PH and EH (Sun et al., 2024; Wang et al., 2022). In this study, genome-wide expression profiles of internodes for the diverse hybrid population revealed several gene regulatory networks associated with PH and EH, and candidate genes associated with the tested traits could be prioritized based on co-localization of QTL for both traits with eQTL and meQTL.

A diverse NCII population is particularly well-suited for investigating the genetic basis of hybrid performance and heterosis

The parental lines of the NCII hybrid population showed great genetic diversity, and had negligible population structure, enabling high-resolution mapping of candidate genes. We used the AD model that integrates both additive and dominance kinship matrices, and respectively maps additive- and dominance-QTL by testing the additive and dominance effects of markers separately. The detected QTL did not have consistent effects across the four hybrid sub-populations, but overall positive contributions were noted in each sub-population (Fig. 1E-F). Although false positive can't be completely excluded, the reliability of the QTL was supported by the results that the dominance degrees of the dominance-QTL were larger than those of the additive-QTL, and that the cumulative effects of all detected QTL showed strong correlation with hybrid performance (Fig. 1E-F), and that 33 QTL contained genes related to plants growth and development (Table S1). The effects of dominance-QTL were larger than those of the additive-QTL for PH and EH (Fig. S8A-B; Table S1). We therefore focused our study largely on dominance-QTL, and detailed analysis prioritized candidate genes for five of these dominance-QTL. Thus, in accordance with previous reports (Hashimoto et al., 2021; Li *et al.*, 2021b; Xiao et al., 1995), our results supported the substantial contribution of dominance-QTL to hybrid performance and heterosis of hybrid maize.

Systems genetics approaches reveal candidate genes for hybrid performance and heterosis in maize

Genome-wide expression profiles can be exploited to map individual genes or groups of co-expression network modules (Feltus, 2014). Combining GWAS, eGWAS and eQTL hotspot analysis, we found that 42 eQTL hotspots were located within 200 kb of 35 QTL (Table S2). Three eQTL hotspots were studied in detail and the candidate genes *GRF5*, *SBP7*, and *CRF4* were prioritized. The three genes were selected because they were in strong LD with the lead SNP in the QTL and eQTL hotspot regions. Furthermore, all three genes showed co-expression with genes related to plant development or growth (Fig. 2G). Genes that were not associated with a local eQTL were also considered as candidate genes due to the possibility that a mutation of the coding sequence could affect the gene function without changing the transcript level. Likewise, it is possible that genome-wide thresholds may not permit the detection of some local eQTL. The *GRF5* gene showed for example significant differences among different genotypic classes of the QTL (Fig. S13) but no local eQTL.

The reported functions for *GRF5*, *SBP7*, and *CRF4* are consistent with a putative role in controlling PH and EH. For example, SBP domain-containing genes participate in multiple growth and development processes. *IPA1* regulates plant architecture in rice (Song et al., 2017) and *TaSPL14* controls plant height in wheat (Cao et al., 2021). There are 30 *SBP* genes in maize, and these genes are regarded as promising candidate genes for improving plant architecture for high-density planting in maize (Wei et al., 2018). *GRF5* was shown to regulate cell size, leaf size and root length in *Arabidopsis* (Lantzouni et al., 2020), leaf size and the content of zeatin and isopentenyladenine in poplar (Wu et al., 2021a), and transformation efficiency in some dicot and monocot species (Kong et al., 2020). *CRFs* represent a small subset of genes belonging to the AP2/ERF gene family of transcription factors, and *CRF4* regulates biomass, root development, and $^{15}NO_3^-$ uptake in *Arabidopsis* (Varala et al., 2018). However, it should be pointed out that genes in LD with the candidate genes may also contribute to the eQTL hotspots. Annotation of the downstream eGenes and correlation analysis between the candidate genes with downstream eGenes or tested traits (Fig. S13; S15-16; 4C) are therefore important to provide auxiliary evidence for the functional role of candidate genes.

The extraction of key features representative for groups of co-expressed genes reduces dimensions, and facilitates genetic mapping of *trans*-regulatory eQTL that regulate a large number of genes. It was reasoned that meGWAS may complement but also support the results of eGWAS. Indeed, 13 out of 37 meQTL (35.14%), which were

associated with trait-correlated modules and were mapped within of 10 kb of QTL, were located within the flanking 10 kb of eQTL hotspots (Table S5). Integrating the meGWAS results and the genes of the co-expressed modules into the analysis provided detailed information on QTL regions, even if the associated hotspots did not encompass many genes as was shown for the three QTL spanning the candidate genes *GATA7*, *MYB163*, and *bZIP29*. Studies of candidate genes *GATA7* and *MYB163* in other plant species support a putative function in controlling PH and EH. Rice *OsGATA7* modulates brassinosteroids-mediated regulation of plant growth and architecture (Zhang et al., 2018) and *MYB163* belongs to the R2R3-type MYB transcription factor group, which regulates specialized metabolism, development, and responses to stresses (Dubos et al., 2010).

bZIP29 represents a novel heterotic gene in plants

bZIP29 was prioritized as a candidate gene for a common dominance-QTL for both PH and EH based on its correlations with both traits and with downstream genes belonging to the relevant hotspot and ICs (Fig. 4A-C). Two natural variations in the upstream region of *bZIP29* were significantly associated with *bZIP29* expression and may contribute to phenotypic differences between the different haplotype groups (Fig. S19). *bZIP29* expression showed great variation, and had strong positive correlation with each trait (Fig 4C; 4E; S17), and such correlation was supported by OE- and DEL lines. Importantly, *bZIP29* was associated with dominance effects for hybrid performance and heterosis in the hybrid population, and transgenic approaches confirmed its heterotic effects for PH and EH.

bZIP29 regulated plant growth and the development of reproductive organs in *Arabidopsis* (Lozano-Sotomayor et al., 2016), and its maize ortholog regulated seed development by interacting with *ZmABI19* (Yang *et al.*, 2022). Thus, *bZIP29* may function in different organs or at different stages of plant development. The organ- and stage-specific expression of *bZIP29* supported this notion (www.maizegdb.org; http://minteractome.ncpgr.cn/searchelement.php). The result that the 3,363 target genes of bZIP29 protein were enriched on a variety of GO terms (Fig. S20) indicated that *bZIP29* might have different functional roles depending on its organ- and stage-specific expression. Given that the roles of *bZIP29* on plant growth have been validated in our study, as well as on kernel weight in a previous work (Yang *et al.*, 2022), it would be meaningful to study the contribution of *bZIP29* to hybrid performance of other traits in

524 the future.

525 In addition to the large-effect Mendelian loci affecting heterosis, evidence for a
526 substantial role of epistatic interactions in heterosis has also been documented (Jiang *et*
527 *al.*, 2017; Torgeman and Zamir, 2023). To study epistatic effects in a diverse hybrid
528 panel was computationally challenging for a high-density SNP data set as the one used
529 in this study. Therefore, we used an AD model which cannot discriminate dominance
530 and epistatic effects, and it was possible that some detected dominance-QTL may act
531 as epistatic QTL to affect the variations of the tested traits and gene expressions.
532 Considering that *bZIP29* was spanned by a dominance-QTL, and was involved in
533 complex interactions (Fig. 5J; Table S7), it might be risky to verify its heterotic effect
534 using one or a few hybrid lines derived from two distinct inbreds. Especially. bZIP29
535 protein interacted with 3.3363 genes, including 28 QTL-spanned genes (Table S7), and
536 acted as a core transcriptional regulator to regulate 18 eGenes (Table S2), some of
537 which were related to plant development or stress tolerance (Fig. 4C). Above all, it is
538 tempting to speculate that *bZIP29* may regulate hybrid performance and heterosis by
539 interacting with other genes as was shown for seed development in maize (Yang *et al.*,
540 2022).

541 Both additive and non-additive expression is observed in hybrid lines, and these
542 expression patterns may be related to heterosis (Pea et al., 2008) (Zhang et al., 2023)
543 (Zhou et al., 2022). Springer and Stupar (Springer and Stupar, 2007) proposed that the
544 expression levels may be too low or two high in inbred lines, causing adverse effects
545 on phenotypes. The between-parental expression in hybrid lines may dilute the adverse
546 effects, thereby causing dominance or over-dominance effects. OE-derived F_1 lines are
547 ideal materials for verifying their proposal. In this study, we found that the expression
548 of *bZIP29* in OE1 lines was higher than that in OE2 lines, and the adverse effects on
549 OE1 lines were more severe than those on OE2 lines, indicating that the levels of
550 *bZIP29* expression were associated with their adverse effects. *bZIP29* expressions were
551 reduced in both OE-derived F_1 lines, thus alleviating the adverse effects of *bZIP29* on
552 PH in the OE-derived F1 lines, resulting in over-dominance and dominance patterns for
553 these traits in OE-derived hybrid lines. These results are consistent with Springer and
554 Stupar's proposal.

555 The genes in the regulatory networks of *bZIP29* in the four hybrid sub-populations
556 showed enrichment in common GO terms (Fig. 6B-C), indicating that genes in these
557 GO terms may be related to the vigorous growth of hybrid maize. Strikingly, the gene

regulatory network of *bZIP29* in the parental population was much smaller than those of the hybrid sub-populations. Different network properties had been observed previously when metabolite correlation networks were analyzed, with most maize hybrids showing a higher network density than the corresponding parental lines (Lisec et al., 2011). Likewise, a higher connectivity was observed for the two reciprocal Arabidopsis hybrids than for the corresponding parents Col-0 and C24 (Andorf et al., 2009). Based on these results, it was advocated that a systems biological approach focusing on interactions should be regarded complementary to conventional quantitative genetics approaches focusing on elucidating the roles of epistatic interactions of individual loci with the genetic background (Andorf et al., 2010). Future studies focusing on interactions should aim to decipher the underlying molecular basis for the complex regulatory networks in hybrids and their phenotypic consequences.

Identification of genes associated with hybrid performance and heterosis has been proven to be a formidable task, nonetheless few successful efforts provided invaluable information on the molecular basis of heterosis in plants. In rice, *OsMADS1*, encoding a MIKC-type MADS-box transcription factor, showed incomplete dominance in an F_2 population. Introduction of a 15-bp genomic fragment spanning the intron-exon junction of *OsMADS1* changed the sequence of the mature mRNA and the translated protein, leading to the increase of grain length and weight (Wang et al., 2019; Wang et al., 2024). Through integrated genetic and transcriptome analysis, it was postulated that the allelic heterozygosity of *RH8*, also known as *Ghd8*, could be the causal gene for yield heterosis in many rice hybrid varieties (Li et al., 2016). In tomato, the flowering gene *SINGLE FLOWER TRUSS* (*SFT*) showed an over-dominance effect for yield in distinct genetic backgrounds, and the genetic effect was associated with the suppression of growth termination mediated by *SELF PRUNING* (*SP*) (Krieger et al., 2010). *BnaA9.CYP78A9* represents a heterotic gene for yield heterosis in oilseed rape. Heterozygosity in its upstream regulatory region caused partial dominance at expression and auxin levels, and resulted in non-additive expression of downstream genes (Ye et al., 2023). GWAS analysis based on an *Arabidopsis* half diallel population found that *AGL50* (*AGAMOUS-LIKE 50*) dominantly controlled flowering time, and the function of *AGL50* was verified in four genetic backgrounds (Seymour *et al.*, 2016). In maize, population analysis revealed that *ZAR1* and *ZmACO2* showed over-dominance effects for yield traits (Wang et al., 2023). Our results, together with previous results, showed that heterosis in plants is associated with many different

pathways, and despite these advancements information on how heterosis is regulated in crop plants is still fragmentary.

CONCLUSION

This study described the regulation relationships of QTL in a maize hybrid population. Combining GWAS, eGWAS, meGWAS and Fisher's exact test prioritized six potential heterotic genes, among which, *bZIP29* was spanned by a dominance-QTL in common to both PH and EH, two eQTL hotspots, and four meQTL underlying two ICs. The QTL spanning *bZIP29* showed dominance effects for both hybrid performance and heterosis. Transgenic experiment confirmed that *bZIP29* showed heterotic effects for plant growth. The tsCUT&Tag experiment and regulatory network analysis revealed the regulatory mechanism of heterosis involving *bZIP29*. This study revealed key genes and regulatory networks of hybrid performance and heterosis in maize.

MATERIALS AND METHODS

Population construction, phenotype evaluation, and phenotypic data analysis

In this study, 78 maternal inbred lines (Table S8) were selected from a large maize association mapping panel which included 527 inbred lines with tropical, subtropical and temperate backgrounds (Gui et al., 2022; Li et al., 2013a; Yang et al., 2011). We selected the 78 maternal lines from the temperate sub-population of the association panel, so that the obtained hybrid lines can grow and flower normally in Northern China. Notably, the selected maternal lines exhibited a relatively balanced phylogeny distribution with no extremely divergent clades (Fig.S2A), which must benefit the genetic mapping of hybrid traits of the diverse hybrid population.

The four male parents, including M01, M02, M03, and M04, were used in our previous study (Li et al., 2021a). Notably, the four male parents are in divergent clades (Fig.S2A). A hybrid population was constructed following NCII design, producing four hybrid sub-populations with each containing 78 hybrid lines. Two hybrid lines obtained by crossing M01 to M02, and M03 to M04. In total, 314 hybrid lines were used for phenotypic investigation and transcriptome sequencing.

In the summer of 2018, the hybrid population was sown in Zhuozhou (ZZ, Hebei province; 39°29′N and 115°58′E), Xinxiang (XX, Henan province; 35°22′N and 113°54′E) and Gongzhuling (GZL, Jilin province; 43°30′N and 124°49′E). Each hybrid was sown in a two-row plot with row space being 60 cm and plant space being 25 cm.

The population was arranged in an incomplete block design with each block containing a hybrid sub-population. Ten plants in each plot were evaluated for PH and EH after the flowering stage, and the outliers beyond the 1.5 interquartile range were excluded.

The best linear unbiased estimates (BLUEs) were estimated in two steps as described in a previous study (Jiang *et al.*, 2017). Briefly, in the first step, BLUEs of the population in each environment were obtained with a linear mixed model, which included the effects of genotypes, replicates and blocks nested within replicates. Genotype effect was used as the fixed effect and others as random effects. In the second step, BLUEs across environments were obtained with a linear mixed model, in which the genotype and environment effects were treated as fixed and random effects, respectively. BLUEs of parents and hybrids were used for subsequent genetic analysis. MPH was calculated as:

$$MPH = BLUE_{F1} - (BLUE_M + BLUE_F)/2 \quad (1)$$

To estimate H^2, best linear unbiased predictions (BLUPs) were calculated by fitting a linear mixed model, in which all effects were treated as random effects. H^2 was calculated as (Cullis et al., 2006):

$$H^2 = 1 - \frac{\overline{v}_{BLUP}}{2\sigma_G^2} \quad (2)$$

where σ_G^2 is the genotypic variance and $\bar{v}_{BLUP}$ is the mean variance of the differences between each pair of BLUPs.

Transcriptome sequencing and analysis

All 314 hybrids were sown in Haidian (Beijing) in June of 2018. When plants reached V7 stage, the decapitated juvenile internodes of three plants were bulked as one sample for each line. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Germany), and RNA purity was checked using the kaiaoK5500 Spectrophotometer (Kaiao, Beijing, China). RNA integrity and concentration were assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). mRNA from total RNA was purified using poly-T oligo-attached magnetic beads. Then, sequencing libraries were constructed using NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina (NEB, Ispawich, USA), and sequenced using the Illumina Novaseq system, producing 150 bp (pair-end) sequencing reads. More than 10Gb of data were obtained for each library.

To quantify gene expression, the raw reads of RNA-seq were firstly filtered by

removing adapter sequences, low-quality bases (quality score < 20), and short reads (length >= 30bp) with Trimmomatic-0.36 (Bolger et al., 2014). Secondly, the filtered reads were mapped to the maize reference genome (ARGv4) by STAR (v2.6) in two-pass mode (Dobin et al., 2013), and the uniquely mapped reads were extracted for gene expression quantification. Then, the read counts of genes annotated in maize B73 ARGv4 genome were calculated by HTseq-count with parameters of -s reverse --mode union (Anders et al., 2015). Next, DEseq2 was used to correct the library size and calculate normalized FPKM values of genes (Love et al., 2014). Finally, after filtering with the median of FPKM > 0 among the hybrid population, 28,706 filtered genes were obtained and their expressions were normalized with R *qqnorm*() function. PEER (Stegle et al., 2010) was used to estimate technical variation in expression profile and 16 hidden confounding factors were detected.

SNP calling with genomic re-sequencing data

The public genome re-sequencing data for the parental lines (Gui *et al.*, 2022) were exploited for SNP calling. The raw reads were filtered by removing adapter sequences, low-quality bases (quality score < 20), and short reads (length >= 30bp) with Trimmomatic-0.36 (Bolger *et al.*, 2014). The processed reads were mapped to the maize reference genome (ARGv4) using BWA mem (Li, 2013) with default parameters. Only uniquely mapped reads were extracted for SNP calling, which was carried out following the GATK (v3.8.1) best-practice guidelines for variant calling (DePristo et al., 2011). In brief, the alignment files (BAM format) were sorted, duplicate reads were marked, and group information was added and indexed with Picard Tools (http://broadinstitute.github.io/picard). Next, IndelRealigner of GATK was used to realign reads containing indels, and BaseRecalibrator of GATK was performed for base recalibration with the SNPs of maize Hapmap v3.2.1 as knownSites (Bukowski et al., 2018). Genomic variants were called using HaplotypeCaller with the parameters -dontUseSoftClippedBases -stand_call_conf 20.0 -ERC GVCF -variant_index_type LINEAR. The joint genotyping of all inbred lines was performed with GenotypeGVCFs. Finally, raw SNPs were filtered with quality parameters set as QD >= 2.0, FS <= 60.0, and MQ >= 20.0.

To prepare SNP markers for genetic analysis, SNPs that passed quality control were further filtered with the heterozygous rate <= 20% and missing rate <= 20% among the parent population. Then, Beagle was used to impute the filtered SNPs with

the major parameters set as window=50000 and overlap=10000 (Browning et al., 2018). The additive genotypes of the parent population were coded as 0,1,2, where 0 represents the major homozygous genotype, 1 for the heterozygous genotype, and 2 for the minor homozygous genotype. Owing to the limited sample size of the parental population, there was extensive long-distance LD among SNP loci. To remove such SNPs, we used the method proposed by (Seymour *et al.*, 2016). In brief, in the additive genotype matrix, we regarded SNPs having the same genotypes across the whole population as one genotype combination, and found that there were 1,823,526 genotype combinations. The number of SNPs in these combinations ranged from 1 to 33,586. We retained the combinations with less than 50 SNPs, and their corresponding SNPs. In addition, SNPs showing perfect long-range LD across chromosomes were removed. Finally, 4,628,240 SNPs were obtained for the parental population. The genotypes of hybrids were inferred based on the genotypes of their parents.

The neighbour-joining (NJ) tree was constructed based on the distance matrix computed using the genotypic data. NJ tree construction was implemented using R package APE (Paradis et al., 2004).

GWAS of hybrid performance and identification of QTL

GWAS was implemented following a linear mixed model incorporating both additive and dominance effects (AD model). The AD model (Bonnafous *et al.*, 2018; Ramstein et al., 2020) is:

$$y = \mu + x_i\theta_{ai} + z_i\theta_{di} + A + D + e \qquad (3)$$

where y is the BLUEs; μ is the intercept; x_i and z_i are the additive and dominance genotypes of the ith SNP; θ_{ai} and θ_{di} are the additive and dominance effects of the ith SNP, respectively; A is the random additive effect; D is the random dominance effect. Let $A \sim N(0, \sigma_a^2 K_a)$, $D \sim N(0, \sigma_d^2 K_d)$, where K_a is the additive kinship matrix (VanRaden, 2008), and K_d the dominance kinship matrix (Vitezica et al., 2013). The additive genotypes were coded as stated above. For the dominance genotype matrix, the homozygous and heterozygous genotypes of each SNP were coded as 0 and 1, respectively. To reduce the computational load, the mixed model was approximately transformed into a simple linear regression model based on the population parameters previously determined (P3D) algorithm (Zhang et al., 2010). The significance of additive and dominance effects for each marker were determined with *t*-tests. A total of

2,324,652 SNPs, whose frequencies of each SNP genotype (two homozygous genotypes and a heterozygous genotype) were larger than 0.05, were used for association mapping

The genome-wide significance threshold for the association analysis was determined based on the number of effective markers (Me). A total of 74,316 effective SNPs were identified using GEC (v0.2) (Li et al., 2012). To establish an optimal threshold, four approaches were compared: (1) a modified Bonferroni correction threshold of 1.34×10^{-5}, determined as 1/Me, based on previous recommendations (Li *et al.*, 2013a; Wen et al., 2014); (2) a Bonferroni correction threshold of 6.73×10^{-7}, calculated as 0.05/Me; (3) a false discovery rate (FDR) threshold implemented using the p.adjust() function in R with FDR < 0.05; and (4) permutation tests (PT). For the PT threshold, the BLUE values for each trait were randomly shuffled 1,000 times and subjected to GWAS using the AD model. The *P*-value corresponding to a one-tailed probability of 0.01 was set as the empirical threshold. We compared the four methods and observed the following (Table S9): the second approach yielded zero dominance-QTLs for EH and only a few additive-QTLs for both traits; the third approach resulted in zero additive-QTLs for PH and zero dominance-QTLs for EH; and the fourth approach identified more than 200 QTLs. Comparatively, the first approach better reflected the quantitative inheritance patterns of PH and EH. Therefore, we selected the first approach to determine the QTL significance threshold.

The QTL for hybrid performance were identified with the following steps. Firstly, all significant SNPs with additive and dominance effects were extracted and clustered into intervals with a distance cut-off of 200 kb. Then, only intervals that spanned at least 3 significant SNPs were kept and treated as QTL. The lead SNPs in the interval/QTL were used to represent the QTL. The likelihood-ratio-based R^2 was adopted to estimate the PVEs by the lead SNPs following a mixed linear model, which contained both the additive and dominant kinship matrices (Sun et al., 2010). The model was fitted by GridLMM package (Runcie and Crawford, 2019). A ratio of dominance effect / additive effect (d/a) was used to indicate the degree of dominance for each QTL. The additive effect (a) is half of the difference of the average phenotypic values between the two homozygous genotypes. The dominant effect (d) is the phenotypic difference between the heterozygous genotypes and mid-parent values. The favorable and unfavorable alleles of the lead SNP for each QTL were defined as follows: the hybrids carrying homozygous favorable alleles should display higher average phenotypic

values than those carrying homozygous unfavorable alleles among the hybrid population.

The favor score of each hybrid was calculated as the sum of numeric genotypes at all detected QTL. In brief, the genotypes of the lead SNPs in each QTL were firstly coded as numeric format based on their genetic effects. For additive-QTL, the genotypes of lead SNPs were coded as 1, 0, -1, where 1 represents the homozygous favorable genotype, 0 the heterozygous genotype, and -1 the homozygous unfavorable genotype. For dominance-QTL, the genotypes of lead SNPs were coded as 1, 0, where 1 represents the heterozygous genotype, and 0 the homozygous genotype. Thus, the favor score indicated the cumulative effects of favorable genotypes of all detected QTL in each hybrid.

eQTL and hotspot analysis

eGWAS was performed with an AD model modified from formula (3). In this model, the 16 PEER factors were incorporated as fixed effects to correct the technical variation, and the significance of combined effect (additive and dominance effects) of each marker was tested with Wald test. To reduce the computational load of eQTL mapping, the redundant SNPs were removed through a genome-wide LD-based pruning ($r^2 \geqslant 0.99$) by PLINK with key parameters set as --indep-pairwise 100 5 0.99, producing 889,673 informative SNPs. The significance threshold (1.34E-5) corresponded to that of the GWAS analysis.

For a specific gene, the eQTL were identified with the following steps: (1) the significant SNPs (eSNPs) within a distance of 10 kb were grouped into a candidate eQTL interval, and the candidate eQTL containing less than three eSNPs were removed. The lead SNPs were chosen to represent the candidate eQTL. (2) If multiple candidate eQTL showed strong LD (r2 > 0.1) with each other, then the less significant eQTL was dropped. The target genes of eQTL were referred as eGenes. The local- and distant-eQTL were determined based on the relative distance between eQTL and the target genes with a distance cut-off of 200 kb. eQTL hotspots were identified with hot_scan software (Silva et al., 2014) by setting the window size as 10 kb, and the threshold of adjusted *P*-value < 0.01.

IC modules and module-eQTL identification

Independent component (IC) analysis was performed to obtain co-expression

modules using the package picaplot v0.99.7 (https://github.com/jinhyunju/picaplot) with fastICA algorithm by setting the key parameters as n_runs = 20, n_cores = 40, max_iter = 500. The optimal number of ICs was determined using a cut-off value of 95% cumulative variance. Gene assignment to each module was determined using the R fdrtool package. Genes with adjusted *P*-values < 0.001 for false discovery rate (FDR) estimation of the source signal weights were added to the module. Finally, only ICs with assigned genes ≥ 10 were kept.

meGWAS for the latent features of IC modules was performed with the same AD model as mentioned in eGWAS analysis. meQTL were identified with the same pipeline as adopted for the identification of eQTL. To check whether the eQTL were significantly enriched in 200 kb flanking regions of meQTL, a fisher exact test (Sun *et al.*, 2023) was performed with four categories: (1) eGenes assigned to ICs are regulated by eQTL in the flanking regions of meQTL regulating the specific ICs, (2) eGenes not assigned to ICs are regulated by eQTL in the flanking regions of meQTL regulating the specific ICs, (3) genes assigned to ICs but not regulated by eQTL in the flanking regions of meQTL regulating the specific ICs, (4) and eGenes neither assigned to ICs nor regulated by eQTL in the flanking regions of meQTL regulating the specific ICs. Significant *P*-values (≤ 0.05) indicate that eQTL of more genes assigned to ICs then expected by chance are associated with the meQTL.

GO functional enrichment analysis for the genes allocated to each IC was performed based on maize-GAMER annotation (Wimalanathan et al., 2018), and was implemented using Goatools (Klopfenstein et al., 2018) with a hypergeometric test (Bonferroni-correction $P \leq 0.05$).

Haplotype analysis for *bZIP29*

Re-sequencing-based variations (SNPs and INDELs identified using GATK) located within the gene body and the 10 kb upstream region of bZIP29 were filtered based on quality and genotype frequency (frequencies of the three genotypes in the hybrid population > 0.05). These variations were then used in association analysis for PH and EH with the AD model. Significant trait-associated variations ($P < 0.01$) were clustered into genotype groups, which were analyzed using a one-way ANOVA model (trait ~ group, data) in R. The most significant genotype group was treated as the associated haplotype. The phenotypic and expression-level effects of the associated haplotype were further evaluated through pairwise comparison of means using Tukey's

HSD test (adjusted $P < 0.05$), with only HapGs with frequencies exceeding 0.05 included in the pairwise comparison.

Construction and phenotypic investigation of transgenic lines, and expression quantification of *bZIP29* with qRT-PCR

The generation of *bZIP29* CRISPR-Cas9 transgenic lines (DEL) and the *bZIP29* over-expression transgenic (OE) lines were described in a previous work (Yang *et al.*, 2022). Both DEL and OE lines were reliably used to compare the differences of seed development between transgenic lines and wildtype. DEL lines were generated through Agrobacterium tumefaciens-mediated transformation in the B104 inbred background. The obtained DEL plants were crossed and backcrossed to B104 to select the BC1 seeds that didn't contain Cas9 construct. The selected BC1 plants were self-pollinated to obtain the wild-type and homozygous DEL lines. The full-length CDS of *bZIP29* driven by the 27-kD γ-zein promoter was cloned into pTF102 plasmids, which were transformed to the Hi-II hybrid, and then backcrossed to B104 for more than four generations. The null segregants from the self-pollinated OE lines were used as wild type control.

Phenotypes of transgenic lines were investigated in the summer of 2024 in Beijing, plant height and ear height of DEL, OE, WT, and their F1 lines were measured. The juvenile internodes of DEL, OE, WT, and their F1 lines at V7 stage were collected for expression quantification of *bZIP29*. Total RNA samples were extracted with FastPure Universal Plant Total RNA Isolation Kit (catalogue No. RC411, Vazyme, Nanjing, China). For each sample, 1 ng RNA was used for synthesis of the first-strand complementary DNAs (cDNAs) using the HiScript II Q RT SuperMix for qPCR (+gDNA wiper) (Vazyme, Nanjing, China). qRT-PCR was conducted with the Taq Pro Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China) on a CFX-96 Real-time PCR platform (BIO-RAD, Hercules, USA). The *ZmActin* gene was used as internal reference (Zhang et al., 2020). The primer sequences used for qRT-PCR were listed in Table S10.

tsCUT&Tag experiment and data analysis

A tsCUT&Tag technology was employed to analyze the regulatory network of *bZIP29* (Wu et al., 2022). The Hyperactive In-Situ ChIP Library Prep Kit for Illumina (pG-Tn5) kit (Vazyme TD901) was used for experimental operations. The transformed

protoplasts of B73 were observed by fluorescence microscope to detect the transformation efficiency. Selected samples with conversion efficiency >60% were used for subsequent tsCUT&Tag experiments. Two biological replicates were set for the constructed vector containing CDS of *bZIP29*. Cells were collected by low-speed centrifugation at 100 r/min for 2 min, and resuspended using 100 ml resuspension solution. After treating the resultant with ConA beads, incubation was performed with GFP antibody and the corresponding secondary antibody. pG-Tn5 Transposon was used to fragment the DNAs and insert adaptors. Finally, the fragmentated DNA was extracted for library construction. After quantifying by Qubit, the libraries were sequenced with pair-end 150 bp in Illumina Hiseq X-Ten platform. Protoplasts transformed with the pM999-GFP vector were used as control. The cleaned reads were mapped to the B73 reference genome (AGPv4) using bowtie2 (Langmead and Salzberg, 2012). The high confidence peaks ($P_{peak} < 0.0001$) were scanned across the whole genome with Macs (Feng et al., 2012). The distribution of peaks in the whole genome was analyzed with ChIPseeker in R (Yu et al., 2015). If a peak was located within 3 KB from TSS of the gene, the gene was considered to be a target.

Funding

This study was supported by the National Key R&D Program of China (2023YFF1000400), Biological Breeding-National Science and Technology Major Project (2023ZD04076), the China Agriculture Research System of Maize (CARS-02-13), and the Innovation Program of the Chinese Academy of Agricultural Sciences

Author contributions

H.Z., J.F., and J.W. conceived and designed the experiments. J.Z., R.G., X.M., Z.X., Y.M., P.W., Y.L., X.W., X.D., N.Z., S.Z., T.Y., Y.W., and C.L. performed the experiments. J.Z., R.S., J.L., X.M., J.R. and Y.J. analyzed the data. J.Z., H.Z., J.F., R.S., Y.J., Y.X., L.L., and J.R. wrote the paper.

Acknowledgments

We thank the high-performance Bioinformatics Computing Platform of Institute of Crop Sciences, Chinese Academy of Agricultural Sciences for providing computational support.

Data availability

RNA-seq data of maize hybrids have been deposited in the NCBI SRA database, accession number PRJNA1000225. Genomic Re-seq data of maize parental lines were downloaded from NCBI SRA database under accession number PRJNA531553.

Competing interests

All authors declare that they have no competing interests.

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics (Oxford, England) **31**:166-169. 10.1093/bioinformatics/btu638.

Andorf, S., Gärtner, T., Steinfath, M., Witucka-Wall, H., Altmann, T., and Repsilber, D. (2009). Towards systems biology of heterosis: a hypothesis about molecular network structure applied for the Arabidopsis metabolome. EURASIP journal on bioinformatics & systems biology **2009**:147157. 10.1155/2009/147157.

Andorf, S., Selbig, J., Altmann, T., Poos, K., Witucka-Wall, H., and Repsilber, D. (2010). Enriched partial correlations in genome-wide gene expression profiles of hybrids (A. thaliana): a systems biological approach towards the molecular basis of heterosis. Theor Appl Genet **120**:249-259. 10.1007/s00122-009-1214-z.

Austin, D.F., Lee, M., and Veldboom, L.R. (2001). Genetic mapping in maize with hybrid progeny across testers and generations: plant height and flowering. Theoretical and Applied Genetics **102**:163-176. 10.1007/s001220051632.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) **30**:2114-2120. 10.1093/bioinformatics/btu170.

Bonnafous, F., Fievet, G., Blanchet, N., Boniface, M.-C., Carrère, S., Gouzy, J., Legrand, L., Marage, G., Bret-Mestries, E., Munos, S., et al. (2018). Comparison of GWAS models to identify non-additive genetic control of flowering time in sunflower hybrids. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **131**:319-332. 10.1007/s00122-017-3003-4.

Browning, B.L., Zhou, Y., and Browning, S.R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. American Journal of Human Genetics **103**:338-348. 10.1016/j.ajhg.2018.07.015.

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. 10.1093/gigascience/gix134.

Cao, J., Liu, K., Song, W., Zhang, J., Yao, Y., Xin, M., Hu, Z., Peng, H., Ni, Z., Sun, Q., et al. (2021). Pleiotropic function of the SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE gene TaSPL14 in wheat plant architecture. Planta **253**:44. 10.1007/s00425-020-03531-x.

Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. Nature Reviews. Genetics **15**:34-48. 10.1038/nrg3575.

Cote, A.C., Young, H.E., and Huckins, L.M. (2022). Comparison of confound adjustment methods in the construction of gene co-expression networks. Genome Biology **23**:44. 10.1186/s13059-022-02606-0.

Cui, J., Cui, Z., Lu, Y., Lv, X., Cao, Q., Hou, Y., Yang, X., and Gu, Y. (2022). Maize grain yield enhancement in modern hybrids associated with greater stalk lodging resistance at a high planting density: a case study in northeast China. Scientific Reports **12**:14647. 10.1038/s41598-022-18908-z.

Cullis, B.R., Smith, A.B., and Coombes, N.E. (2006). On the design of early generation variety trials with correlated data. Journal of Agricultural, Biological, and Environmental Statistics **11**:381-393. 10.1198/108571106X154443.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics **43**:491-498. 10.1038/ng.806.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England) **29**:15-21. 10.1093/bioinformatics/bts635.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. Trends in Plant Science **15**:573-581. 10.1016/j.tplants.2010.06.005.

East, E.M. (1936). Heterosis. Genetics **21**:375-397. 10.1093/genetics/21.4.375.

Feltus, F. (2014). Systems genetics: a paradigm to improve discovery of candidate genes and mechanisms underlying complex traits. Plant science : an international journal of experimental plant biology **223**10.1016/j.plantsci.2014.03.003.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nature Protocols **7**:1728-1740. 10.1038/nprot.2012.101.

Fiévet, J.B., Nidelet, T., Dillmann, C., and de Vienne, D. (2018). Heterosis Is a Systemic Property Emerging From Non-linear Genotype-Phenotype Relationships: Evidence From in Vitro Genetics and Computer Simulations. Frontiers in Genetics **9**:159. 10.3389/fgene.2018.00159.

Fu, J., Cheng, Y., Linghu, J., Yang, X., Kang, L., Zhang, Z., Zhang, J., He, C., Du, X., Peng, Z., et al. (2013). RNA sequencing reveals the complex regulatory network in the maize kernel. Nature Communications **4**:2832. 10.1038/ncomms3832.

Gui, S., Wei, W., Jiang, C., Luo, J., Chen, L., Wu, S., Li, W., Wang, Y., Li, S., Yang, N., et al. (2022). A pan-Zea genome map for enhancing maize improvement. Genome Biol **23**:178. 10.1186/s13059-022-02742-7.

Guo, T., Yang, N., Tong, H., Pan, Q., Yang, X., Tang, J., Wang, J., Li, J., and Yan, J. (2014). Genetic basis of grain yield heterosis in an "immortalized F_2" maize population. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **127**:2149-2158. 10.1007/s00122-014-2368-x.

Hashimoto, S., Wake, T., Nakamura, H., Minamiyama, M., Araki-Nakamura, S., Ohmae-Shinohara, K., Koketsu, E., Okamura, S., Miura, K., Kawaguchi, H., et al. (2021). The dominance model for heterosis explains culm length genetics in a hybrid sorghum variety. Scientific Reports **11**:4532. 10.1038/s41598-021-84020-3.

Hochholdinger, F., and Baldauf, J.A. (2018). Heterosis in plants. Current biology: CB **28**:R1089-R1092. 10.1016/j.cub.2018.06.041.

Hu, S., Wang, C., Sanchez, D.L., Lipka, A.E., Liu, P., Yin, Y., Blanco, M., and Lübberstedt, T. (2017). Gibberellins Promote Brassinosteroids Action and Both Increase Heterosis for Plant Height in Maize (Zea mays L.). Frontiers in Plant Science **8**:1039. 10.3389/fpls.2017.01039.

Hua, J.P., Xing, Y.Z., Xu, C.G., Sun, X.L., Yu, S.B., and Zhang, Q. (2002). Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics **162**:1885-1895. 10.1093/genetics/162.4.1885.

Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., Li, W., Zhan, Q., Cheng, B., Xia, J., et al. (2015). Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. Nature Communications **6**:6258. 10.1038/ncomms7258.

Jiang, Y., Schmidt, R.H., Zhao, Y., and Reif, J.C. (2017). A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. Nature Genetics **49**:1741-1746. 10.1038/ng.3974.

Jones, D.F. (1917). Dominance of Linked Factors as a Means of Accounting for Heterosis. Genetics **2**:466-479. 10.1093/genetics/2.5.466.

Klopfenstein, D.V., Zhang, L., Pedersen, B.S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C.J., Yunes, J.M., Botvinnik, O., Weigel, M., et al. (2018). GOATOOLS: A Python library for Gene Ontology analyses. Scientific Reports **8**:10872. 10.1038/s41598-018-28948-z.

Kolberg, L., Kerimov, N., Peterson, H., and Alasoo, K. (2020). Co-expression analysis reveals interpretable gene modules controlled by trans-acting genetic variants. eLife **9**:e58705. 10.7554/eLife.58705.

Kong, J., Martin-Ortigosa, S., Finer, J., Orchard, N., Gunadi, A., Batts, L.A., Thakare, D., Rush, B., Schmitz, O., Stuiver, M., et al. (2020). Overexpression of the Transcription Factor GROWTH-REGULATING FACTOR5 Improves Transformation of Dicot and Monocot Species. Frontiers in Plant Science **11**:572319. 10.3389/fpls.2020.572319.

Krieger, U., Lippman, Z.B., and Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nature Genetics **42**:459-463. 10.1038/ng.550.

Labroo, M.R., Studer, A.J., and Rutkoski, J.E. (2021). Heterosis and Hybrid Crop Breeding: A Multidisciplinary Review. Frontiers in Genetics **12**:643761. 10.3389/fgene.2021.643761.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods **9**:357-359. 10.1038/nmeth.1923.

Lantzouni, O., Alkofer, A., Falter-Braun, P., and Schwechheimer, C. (2020). GROWTH-REGULATING FACTORS Interact with DELLAs and Regulate Growth in Cold Stress. The Plant Cell **32**:1018-1034. 10.1105/tpc.19.00784.

Li, D., Xu, Z., Gu, R., Wang, P., Xu, J., Du, D., Fu, J., Wang, J., Zhang, H., and Wang, G. (2021a). Genomic Prediction across Structured Hybrid Populations and Environments in Maize. Plants **10**:1174. 10.3390/plants10061174.

Li, D., Zhou, Z., Lu, X., Jiang, Y., Li, G., Li, J., Wang, H., Chen, S., Li, X., Würschum, T., et al. (2021b). Genetic Dissection of Hybrid Performance and Heterosis for Yield-Related Traits in Maize. Frontiers in Plant Science **12**:774478. 10.3389/fpls.2021.774478.

Li, D., Huang, Z., Song, S., Xin, Y., Mao, D., Lv, Q., Zhou, M., Tian, D., Tang, M., Wu, Q., et al. (2016). Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. Proceedings of the National Academy of Sciences of the United

States of America **113**:E6026-E6035. 10.1073/pnas.1610115113.
Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics.
Li, H., Yang, Q., Fan, N., Zhang, M., Zhai, H., Ni, Z., and Zhang, Y. (2017). Quantitative trait locus analysis of heterosis for plant height and ear height in an elite maize hybrid zhengdan 958 by design III. BMC genetics **18**:36. 10.1186/s12863-017-0503-9.
Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., et al. (2013a). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nature Genetics **45**:43-50. 10.1038/ng.2484.
Li, L., Petsch, K., Shimizu, R., Liu, S., Xu, W.W., Ying, K., Yu, J., Scanlon, M.J., Schnable, P.S., Timmermans, M.C.P., et al. (2013b). Mendelian and non-Mendelian regulation of gene expression in maize. PLoS genetics **9**:e1003202. 10.1371/journal.pgen.1003202.
Li, M.-X., Yeung, J.M.Y., Cherny, S.S., and Sham, P.C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Human Genetics **131**:747-756. 10.1007/s00439-011-1118-2.
Li, R., Jeong, K., Davis, J.T., Kim, S., Lee, S., Michelmore, R.W., Kim, S., and Maloof, J.N. (2018). Integrated QTL and eQTL Mapping Provides Insights and Candidate Genes for Fatty Acid Composition, Flowering Time, and Growth Traits in a F2 Population of a Novel Synthetic Allopolyploid Brassica napus. Frontiers in Plant Science **9**:1632. 10.3389/fpls.2018.01632.
Li, Y.-X., Lu, J., He, C., Wu, X., Cui, Y., Chen, L., Zhang, J., Xie, Y., An, Y., Liu, X., et al. (2022). cis-Regulatory variation affecting gene expression contributes to the improvement of maize kernel size. The Plant Journal: For Cell and Molecular Biology **111**:1595-1608. 10.1111/tpj.15910.
Li, Z., Wang, P., You, C., Yu, J., Zhang, X., Yan, F., Ye, Z., Shen, C., Li, B., Guo, K., et al. (2020). Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton. The New Phytologist **226**:1738-1752. 10.1111/nph.16468.
Lin, F., Jiang, L., Liu, Y., Lv, Y., Dai, H., and Zhao, H. (2014). Genome-wide identification of housekeeping genes in maize. Plant Molecular Biology **86**:543-554. 10.1007/s11103-014-0246-1.
Lisec, J., Römisch-Margl, L., Nikoloski, Z., Piepho, H.-P., Giavalisco, P., Selbig, J., Gierl, A., and Willmitzer, L. (2011). Corn hybrids display lower metabolite variability and complex metabolite inheritance patterns. The Plant Journal: For Cell and Molecular Biology **68**:326-336. 10.1111/j.1365-313X.2011.04689.x.
Liu, J., Li, M., Zhang, Q., Wei, X., and Huang, X. (2020). Exploring the molecular basis of heterosis for plant breeding. Journal of Integrative Plant Biology **62**:287-298. 10.1111/jipb.12804.
Liu, R., Wang, B., Guo, W., Wang, L., and Zhang, T. (2011). Differential gene expression and associated QTL mapping for cotton yield based on a cDNA-AFLP transcriptome map in an immortalized F2. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **123**:439-454. 10.1007/s00122-011-1597-5.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology **15**:550. 10.1186/s13059-014-0550-8.
Lozano-Sotomayor, P., Chávez Montes, R.A., Silvestre-Vañó, M., Herrera-Ubaldo, H., Greco, R., Pablo-Villa, J., Galliani, B.M., Diaz-Ramirez, D., Weemen, M., Boutilier, K., et al. (2016). Altered expression of the bZIP transcription factor DRINK ME affects growth and reproductive development in Arabidopsis thaliana. The Plant Journal: For Cell and Molecular Biology **88**:437-451. 10.1111/tpj.13264.
Ma, Y., Li, D., Xu, Z., Gu, R., Wang, P., Fu, J., Wang, J., Du, W., and Zhang, H. (2022). Dissection of the Genetic Basis of Yield Traits in Line per se and Testcross Populations and Identification of Candidate Genes for Hybrid Performance in Maize. International Journal of Molecular Sciences **23**:5074. 10.3390/ijms23095074.
Momozawa, Y., Dmitrieva, J., Théâtre, E., Deffontaine, V., Rahmouni, S., Charloteaux, B., Crins, F., Docampo, E., Elansary, M., Gori, A.-S., et al. (2018). IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. Nature Communications **9**:2427. 10.1038/s41467-018-04365-8.
Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics (Oxford, England) **20**:289-290. 10.1093/bioinformatics/btg412.
Pea, G., Ferron, S., Gianfranceschi, L., Krajewski, P., and Enrico Pè, M. (2008). Gene expression non-additivity in immature ears of a heterotic F1 maize hybrid. Plant Science **174**:17-24. 10.1016/j.plantsci.2007.09.005.
Ramstein, G.P., Larsson, S.J., Cook, J.P., Edwards, J.W., Ersoz, E.S., Flint-Garcia, S., Gardner, C.A., Holland, J.B., Lorenz, A.J., McMullen, M.D., et al. (2020). Dominance Effects and Functional Enrichments Improve Prediction of Agronomic Traits in Hybrid Maize. Genetics **215**:215-230. 10.1534/genetics.120.303025.

Rotival, M., Zeller, T., Wild, P.S., Maouche, S., Szymczak, S., Schillert, A., Castagné, R., Deiseroth, A., Proust, C., Brocheton, J., et al. (2011). Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. PLoS genetics **7**:e1002367. 10.1371/journal.pgen.1002367.

Runcie, D.E., and Crawford, L. (2019). Fast and flexible linear mixed models for genome-wide genetics. PLoS genetics **15**:e1007978. 10.1371/journal.pgen.1007978.

Schnable, P.S., and Springer, N.M. (2013). Progress toward understanding heterosis in crop plants. Annual Review of Plant Biology **64**:71-88. 10.1146/annurev-arplant-042110-103827.

Seymour, D.K., Chae, E., Grimm, D.G., Martín Pizarro, C., Habring-Müller, A., Vasseur, F., Rakitsch, B., Borgwardt, K.M., Koenig, D., and Weigel, D. (2016). Genetic architecture of nonadditive inheritance in Arabidopsis thaliana hybrids. Proceedings of the National Academy of Sciences of the United States of America **113**:E7317-E7326. 10.1073/pnas.1615268113.

Silva, I.T., Rosales, R.A., Holanda, A.J., Nussenzweig, M.C., and Jankovic, M. (2014). Identification of chromosomal translocation hotspots via scan statistics. Bioinformatics (Oxford, England) **30**:2551-2558. 10.1093/bioinformatics/btu351.

Song, X., Lu, Z., Yu, H., Shao, G., Xiong, J., Meng, X., Jing, Y., Liu, G., Xiong, G., Duan, J., et al. (2017). IPA1 functions as a downstream transcription factor repressed by D53 in strigolactone signaling in rice. Cell Research **27**:1128-1141. 10.1038/cr.2017.102.

Springer, N.M., and Stupar, R.M. (2007). Allelic variation and heterosis in maize: how do two halves make more than a whole? Genome Research **17**:264-275. 10.1101/gr.5347007.

Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS computational biology **6**:e1000770. 10.1371/journal.pcbi.1000770.

Sun, C., Liu, Y., Li, G., Chen, Y., Li, M., Yang, R., Qin, Y., Chen, Y., Cheng, J., Tang, J., et al. (2024). *ZmCYP90D1* regulates maize internode development by modulating brassinosteroid-mediated cell division and growth. The Crop Journal **12**:58-67. 10.1016/j.cj.2023.11.002.

Sun, G., Zhu, C., Kramer, M.H., Yang, S.-S., Song, W., Piepho, H.-P., and Yu, J. (2010). Variation explained in mixed-model association mapping. Heredity **105**:333-340. 10.1038/hdy.2010.11.

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 Biology **24**:55. 10.1186/s13059-023-02891-3.

Tan, Z., Peng, Y., Xiong, Y., Xiong, F., Zhang, Y., Guo, N., Tu, Z., Zong, Z., Wu, X., Ye, J., et al. (2022). Comprehensive transcriptional variability analysis reveals gene networks regulating seed oil content of Brassica napus. Genome Biology **23**:233. 10.1186/s13059-022-02801-z.

Tang, J., Yan, J., Ma, X., Teng, W., Wu, W., Dai, J., Dhillon, B.S., Melchinger, A.E., and Li, J. (2010). Dissection of the genetic basis of heterosis in an elite maize hybrid by QTL mapping in an immortalized F2 population. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **120**:333-340. 10.1007/s00122-009-1213-0.

Tang, S., Zhao, H., Lu, S., Yu, L., Zhang, G., Zhang, Y., Yang, Q.-Y., Zhou, Y., Wang, X., Ma, W., et al. (2021). Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in Brassica napus. Molecular Plant **14**:470-487. 10.1016/j.molp.2020.12.003.

Torgeman, S., and Zamir, D. (2023). Epistatic QTLs for yield heterosis in tomato. Proceedings of the National Academy of Sciences of the United States of America **120**:e2205787119. 10.1073/pnas.2205787119.

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. Journal of Dairy Science **91**:4414-4423. 10.3168/jds.2007-0980.

Varala, K., Marshall-Colón, A., Cirrone, J., Brooks, M.D., Pasquino, A.V., Léran, S., Mittal, S., Rock, T.M., Edwards, M.B., Kim, G.J., et al. (2018). Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants. Proceedings of the National Academy of Sciences of the United States of America **115**:6494-6499. 10.1073/pnas.1721487115.

Vitezica, Z.G., Varona, L., and Legarra, A. (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics **195**:1223-1230. 10.1534/genetics.113.155176.

Wallace, J.G., Rodgers-Melnick, E., and Buckler, E.S. (2018). On the Road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of Crop Quantitative Genomics. Annual Review of Genetics **52**:421-444. 10.1146/annurev-genet-120116-024846.

Wang, B., Hou, M., Shi, J., Ku, L., Song, W., Li, C., Ning, Q., Li, X., Li, C., Zhao, B., et al. (2023). De novo genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. Nature Genetics **55**:312-323. 10.1038/s41588-022-01283-w.

Wang, C., Wang, Z., Cai, Y., Zhu, Z., Yu, D., Hong, L., Wang, Y., Lv, W., Zhao, Q., Si, L., et al. (2024). A higher-yield hybrid rice is achieved by assimilating a dominant heterotic gene in inbred parental lines. Plant Biotechnology Journal 10.1111/pbi.14295.

Wang, C., Tang, S., Zhan, Q., Hou, Q., Zhao, Y., Zhao, Q., Feng, Q., Zhou, C., Lyu, D., Cui, L., et al. (2019). Dissecting a heterotic gene through GradedPool-Seq mapping informs a rice-improvement strategy. Nature Communications **10**:2982. 10.1038/s41467-019-11017-y.

Wang, F., Yu, Z., Zhang, M., Wang, M., Lu, X., Liu, X., Li, Y., Zhang, X., Tan, B.-C., Li, C., et al. (2022). ZmTE1 promotes plant height by regulating intercalary meristem formation and internode cell elongation in maize. Plant Biotechnology Journal **20**:526-537. 10.1111/pbi.13734.

Wang, X., Chen, Q., Wu, Y., Lemmon, Z.H., Xu, G., Huang, C., Liang, Y., Xu, D., Li, D., Doebley, J.F., et al. (2018). Genome-wide Analysis of Transcriptional Variability in a Large Maize-Teosinte Population. Molecular Plant **11**:443-459. 10.1016/j.molp.2017.12.011.

Wei, H., Zhao, Y., Xie, Y., and Wang, H. (2018). Exploiting SPL genes to improve maize plant architecture tailored for high-density planting. Journal of Experimental Botany **69**:4675-4688. 10.1093/jxb/ery258.

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J., et al. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nature Communications **5**:3438. 10.1038/ncomms4438.

Wimalanathan, K., Friedberg, I., Andorf, C.M., and Lawrence-Dill, C.J. (2018). Maize GO Annotation-Methods, Evaluation, and Review (maize-GAMER). Plant Direct **2**:e00052. 10.1002/pld3.52.

Wu, L., Luo, Z., Shi, Y., Jiang, Y., Li, R., Miao, X., Yang, F., Li, Q., Zhao, H., Xue, J., et al. (2022). A cost-effective tsCUT&Tag method for profiling transcription factor binding landscape. Journal of Integrative Plant Biology **64**:2033-2038. 10.1111/jipb.13354.

Wu, W., Li, J., Wang, Q., Lv, K., Du, K., Zhang, W., Li, Q., Kang, X., and Wei, H. (2021a). Growth-regulating factor 5 (GRF5)-mediated gene regulatory network promotes leaf growth and expansion in poplar. The New Phytologist **230**:612-628. 10.1111/nph.17179.

Wu, X., Liu, Y., Zhang, Y., and Gu, R. (2021b). Advances in Research on the Mechanism of Heterosis in Plants. Frontiers in Plant Science **12**:745726. 10.3389/fpls.2021.745726.

Xiao, J., Li, J., Yuan, L., and Tanksley, S.D. (1995). Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics **140**:745-754. 10.1093/genetics/140.2.745.

Xiao, Y., Jiang, S., Cheng, Q., Wang, X., Yan, J., Zhang, R., Qiao, F., Ma, C., Luo, J., Li, W., et al. (2021). The genetic mechanism of heterosis utilization in maize improvement. Genome Biology **22**:148. 10.1186/s13059-021-02370-7.

Xie, J., Wang, W., Yang, T., Zhang, Q., Zhang, Z., Zhu, X., Li, N., Zhi, L., Ma, X., Zhang, S., et al. (2022). Large-scale genomic and transcriptomic profiles of rice hybrids reveal a core mechanism underlying heterosis. Genome Biology **23**:264. 10.1186/s13059-022-02822-8.

Yang, T., Wang, H., Guo, L., Wu, X., Xiao, Q., Wang, J., Wang, Q., Ma, G., Wang, W., and Wu, Y. (2022). ABA-induced phosphorylation of basic leucine zipper 29, ABSCISIC ACID INSENSITIVE 19, and Opaque2 by SnRK2.2 enhances gene transactivation for endosperm filling in maize. The Plant Cell **34**:1933-1956. 10.1093/plcell/koac044.

Yang, X., Gao, S., Xu, S., Zhang, Z., Prasanna, B.M., Li, L., Li, J., and Yan, J. (2011). Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. Molecular Breeding **28**:511-526. 10.1007/s11032-010-9500-7.

Ye, J., Liang, H., Zhao, X., Li, N., Song, D., Zhan, J., Liu, J., Wang, X., Tu, J., Varshney, R.K., et al. (2023). A systematic dissection in oilseed rape provides insights into the genetic architecture and molecular mechanism of yield heterosis. Plant Biotechnology Journal **21**:1479-1495. 10.1111/pbi.14054.

Yin, X., Bi, Y., Jiang, F., Guo, R., Zhang, Y., Fan, J., Kang, M.S., and Fan, X. (2022). Fine mapping of candidate quantitative trait loci for plant and ear height in a maize nested-association mapping population. Frontiers in Plant Science **13**:963985. 10.3389/fpls.2022.963985.

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics (Oxford, England) **31**:2382-2383. 10.1093/bioinformatics/btv145.

Zhang, H., Xiang, Y., He, N., Liu, X., Liu, H., Fang, L., Zhang, F., Sun, X., Zhang, D., Li, X., et al. (2020). Enhanced Vitamin C Production Mediated by an ABA-Induced PTP-like Nucleotidase Improves Plant Drought Tolerance in Arabidopsis and Maize. Mol Plant **13**:760-776. 10.1016/j.molp.2020.02.005.

Zhang, J., Xie, Y., Zhang, H., He, C., Wang, X., Cui, Y., Heng, Y., Lin, Y., Gu, R., Wang, J., et al. (2023). Integrated Multi-Omics Reveals Significant Roles of Non-Additively Expressed Small RNAs in Heterosis for Maize Plant Height. International Journal of Molecular Sciences **24**:9150. 10.3390/ijms24119150.

Zhang, Y.-J., Zhang, Y., Zhang, L.-L., Huang, H.-Y., Yang, B.-J., Luan, S., Xue, H.-W., and Lin, W.-H. (2018). OsGATA7 modulates brassinosteroids-mediated growth regulation and influences architecture and grain shape. Plant Biotechnology Journal **16**:1261-1264. 10.1111/pbi.12887.
Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. Nature Genetics **42**:355-360. 10.1038/ng.546.
Zhao, T., Wu, H., Wang, X., Zhao, Y., Wang, L., Pan, J., Mei, H., Han, J., Wang, S., Lu, K., et al. (2023). Integration of eQTL and machine learning to dissect causal genes with pleiotropic effects in genetic regulation networks of seed cotton yield. Cell Reports **42**:113111. 10.1016/j.celrep.2023.113111.
Zhou, G., Chen, Y., Yao, W., Zhang, C., Xie, W., Hua, J., Xing, Y., Xiao, J., and Zhang, Q. (2012). Genetic composition of yield heterosis in an elite rice hybrid. Proceedings of the National Academy of Sciences of the United States of America **109**:15847-15852. 10.1073/pnas.1214141109.
Zhou, T., Afzal, R., Haroon, M., Ma, Y., Zhang, H., and Li, L. (2022). Dominant complementation of biological pathways in maize hybrid lines is associated with heterosis. Planta **256**:111. 10.1007/s00425-022-04028-5.

Figure legends:

Fig. 1 Phenotypic and genetic analysis of PH and EH in the hybrid population. (A) The distribution of hybrid performance against the mid-parent values for PH and EH. The colors of points indicated the tested traits. The red dashed line indicated the diagonal line $y = x$. (B) H^2 estimates for hybrid performance and MPH, respectively. (C) Overlaps of different types of QTL. PH_Add and PH_Dom indicated additive-QTL and dominance-QTL for PH, respectively. EH_Add and EH_Dom indicated additive-QTL and dominance-QTL for EH, respectively. (D) PVEs of the lead SNPs of different types of QTL. Statistical comparisons were performed using two-sided Wilcoxon rank-sum tests. (E) and (F) Correlation between favor scores and hybrid traits.

Fig. 2 Identification of eQTL and their genetic characteristics. (A) Proportions of distant- and local-eQTL. (B) Proportions of eGenes regulated by distant-eQTL (Distant), local-eQTL (Local), and both (Distant + Local). (C) Frequencies of distances between eQTL and their regulating eGenes. (D) Frequencies of the number of eQTL that regulate eGenes. (E) Comparison of expression variances explained (R^2_{LR}) by lead SNPs of distant- and local-eQTL. Statistical significance was assessed using a two-sided Wilcoxon rank-sum test. *, $P < 0.05$; **, $P < 0.01$. (F) Distribution of QTL, eQTL and eQTL hotspots in the maize genome, shown as five circles: [I] The ten chromosomes; [II] Heatmap showing the density of eQTL along the genome in 200-kb bins; [III] Barplot showing the distribution of eQTL hotspots with the height of bars representing the number of eGenes regulated by eQTL hotspots; [IV] and [V] Distribution of QTL for PH and EH, respectively. The blue and red bars indicated additive- and dominance-QTL, respectively. The red, blue and purple linked lines within the circle indicated the associations of three representative eQTL hotspots (C4_Hot21, C7_Hot08, and C10_Hot26, respectively) with their downstream eGenes. (G) Regulatory relationship of the three representative eQTL hotspots and their downstream eGenes. The genes potentially related to plant growth and development were labeled with their abbreviated names.

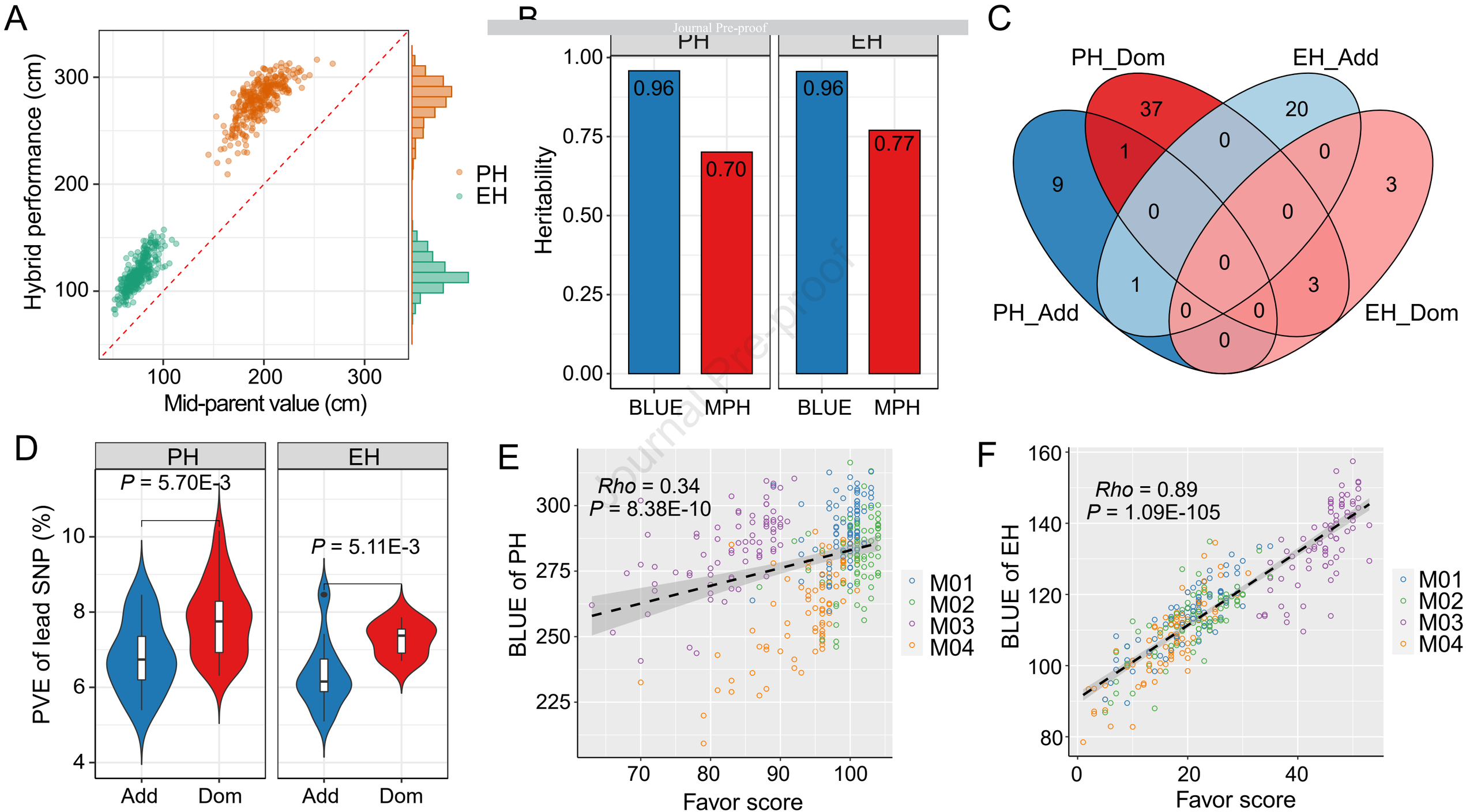
Fig. 3 Co-expression modules and meGWAS analysis. (A) x-axis and y-axis indicated the correlation coefficients (r) between latent features of ICs and PH, and between latent features of ICs and EH, respectively. The colors of points indicated the correlations between latent features of ICs and the tested traits. NoCor indicated non-significant correlations between latent features of ICs and either trait. The sizes of points indicate the number of genes assigned to the ICs. (B-C) GO terms significantly enriched for the top ten positively correlated modules and top five negatively correlated modules for PH (B) and EH (C), respectively. Line color indicated correlations between latent features and the tested traits. The colored GO IDs indicated functional categories of GO terms. (D) Distribution of meQTL for trait-correlated ICs. The colors of points indicated the $-\log_{10}(P\text{-values})$ of the lead SNPs in meQTL. The colors of labels on y-axis indicated that there were significant correlations between latent features of ICs and the tested traits. The colors of bars in the two top rows indicated the additive- (blue) and dominance-QTL (red) for PH and EH, respectively. The shapes of points indicated the levels of P-values for the enrichment analysis with Fisher's exact tests (FET, see "Materials and methods"). (E) and (F) Venn diagrams showing the overlaps of QTL, eQTL hotspots, and meQTL whose flanking regions were significantly enriched for eQTL regulating

eGenes assigned to relevant ICs (as detected by FET analysis). A distance cut-off of 10 kb was adopted for identifying co-localization of meQTL with QTL and eQTL. 104 and 101 meQTL were associated with PH-correlated ICs (E) and EH-correlated ICs (F), respectively.

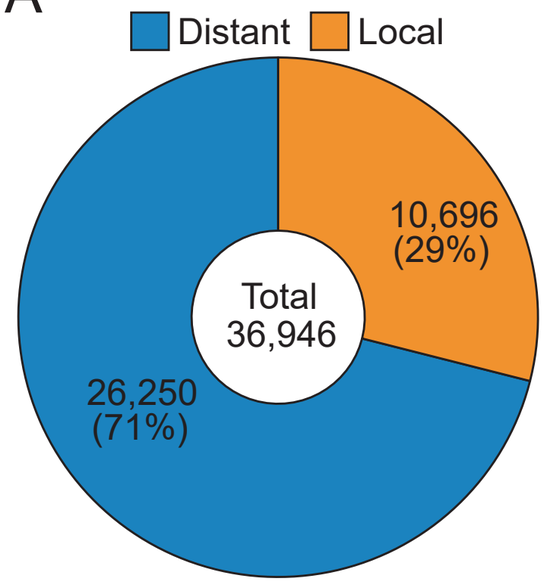
Fig. 4 Association mapping and expression analysis revealed *bZIP29* as candidate gene. (A) Local Manhattan plots showing GWAS signals (upper panel), meGWAS signals (middle panels), and eGenes regulated by the hotspot (C1_Hot60*, lower panel). (B) LD structure of the QTL. (C) Heatmap showing correlations between the tested traits or the expression of *bZIP29* and the expressions of 15 eGenes plus *bZIP29*. (D-E) Hybrid performance, MPH (D), and *bZIP29* expression (E) in hybrids carrying different genotypes at the peak SNP (C1M296443013) of the QTL. We used Student's *t*-test to compare the difference among genotype groups. *, $P < 0.05$; **, $P < 0.01$. The code for performing Student's *t*-test was available at https://github.com/Comp-Bio-ZhangJ/Population-analysis-of-hybrid-maize. The left, middle, and right genotypes on the x-axis of each panels indicated the homozygous favorable genotype, the heterozygous genotype, and homozygous unfavorable genotype, respectively.

Fig. 5 Functional characterization of the dominance effects of *bZIP29* on PH and EH. (A) Comparison of expression levels of *bZIP29* in WT lines, DEL lines (DEL1 and DEL2), and DEL-derived F_1 lines (DEL1F_1 and DEL2F_1). (B) Comparison of expression levels of *bZIP29* in WT lines, OE lines (OE1 and OE2), and OE-derived F_1 lines (OE1F_1 and OE2F_1). The expression levels of *bZIP29* in juvenile internodes of the indicated lines were measured by qPCR. (C) Comparison of PH and EH between WT lines, DEL lines, and DEL-derived F_1 lines. The number of plants for DEL1, DEL1F1, WT, DEL2, and DEL2F1 were 23, 36, 61, 31, and 31, respectively. (D) Phenotypes of WT, DEL, and DEL-derived F_1 lines several days after flowering. (E) Comparison of PH and EH in WT lines, OE lines, and OE-derived F_1 lines. The number of OE1, OE1F1, WT, OE2, and OE2F1 were 15, 37, 15, 39, and 31, respectively. (F) Phenotypes of OE, WT and OE-derived F_1 lines several days after flowering. Statistical significance in panels A–C and E was assessed using Student's *t*-test. *, $P < 0.05$; **, $P < 0.01$. (G) Heatmap showing the quality of the two replicates of tsCUT&Tag libraries. (H) and (I) tsCUT&Tag read counts across the genome sequences of RPS17 and FLS1, respectively. (J) Regulatory network featuring *bZIP29*. Top node indicated *bZIP29*. Orange nodes indicated the eGenes regulated by the eQTL hotspots (C1_Hot60*). Purple and green nodes indicated the co-expressed genes assigned in modules IC176 and IC181, respectively. Red linked lines indicated the genes regulated by *bZIP29* were supported by both evidences of genetic analysis (eQTL or meQTL analysis) and tsCUT&Tag assay. Blue nodes indicated the other target genes identified by tsCUT&Tag assay.

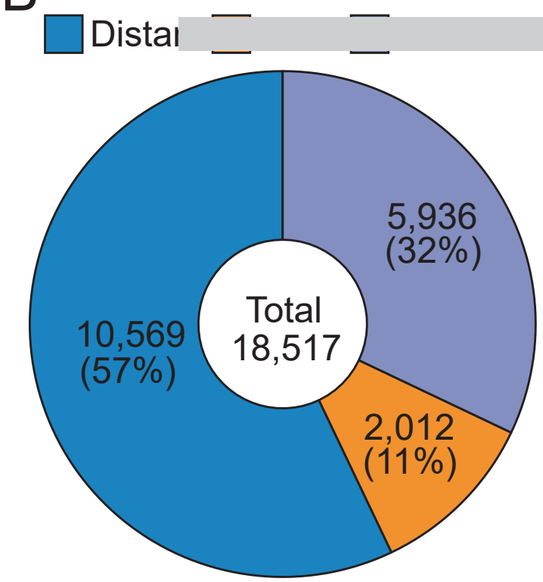
Fig. 6 Comparison of *bZIP29*-centered co-expression networks between the four hybrid sub-populations and the parental population. (A) The Venn diagram displays the genes co-expressed with *bZIP29* in the four hybrid sub-populations and the parental population. The co-expressed genes of *bZIP29* were identified based on Pearson correlation ($P < 0.001$) in the four hybrid sub-populations and parental population. (B) GO enrichment of genes co-expressed with *bZIP29* in the five populations. The colors of labels on y-axis indicated the significantly-enriched GO terms for genes co-expressed with *bZIP29* in the five populations. The colors and sizes of the points indicate the significance levels and the number of co-expressed genes assigned to the enriched GO terms. (C) Co-expression networks of *bZIP29* in the four hybrid sub-populations (M01, M02, M03 and M04), as well as in the parental population. The colors of nodes represent the functional categories of the co-expressed genes based on their GO annotation. The co-expression networks were displayed with Gephi. The numbers under the labels of networks indicated the numbers of hybrids or parents carrying different genotypes at the peak SNP (C1M296443013) of the QTL colocalized with *bZIP29*. The left, middle, and right numbers represent the counts of lines carrying the homozygous favorable alleles, the heterozygous genotype, and homozygous unfavorable alleles, respectively.



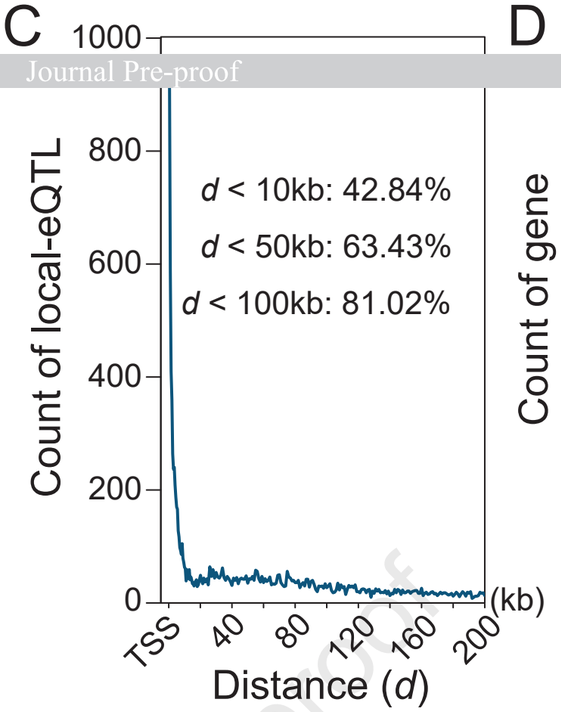
A



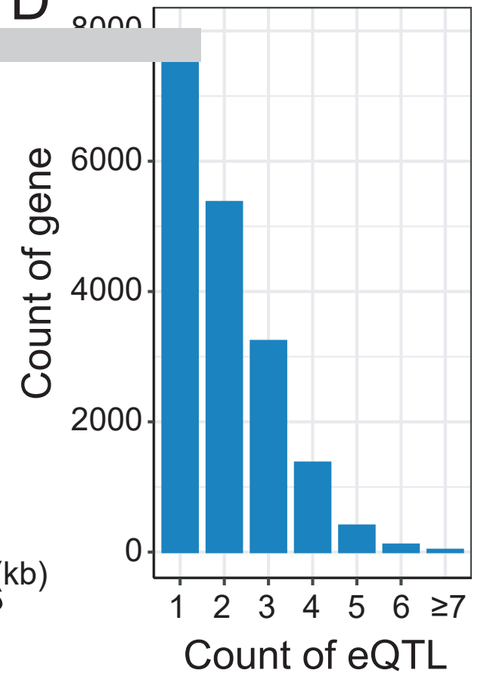
B



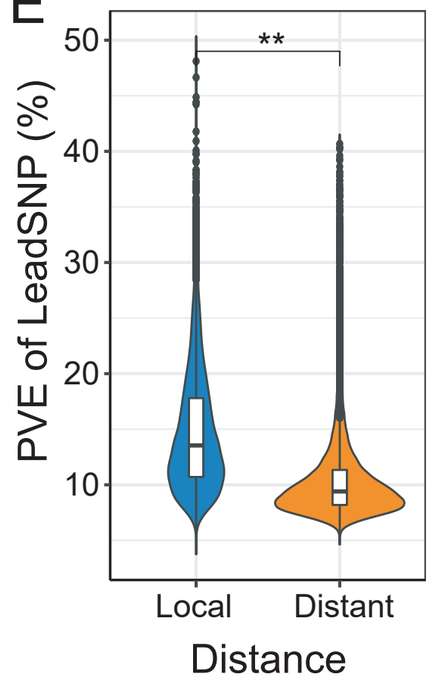
C



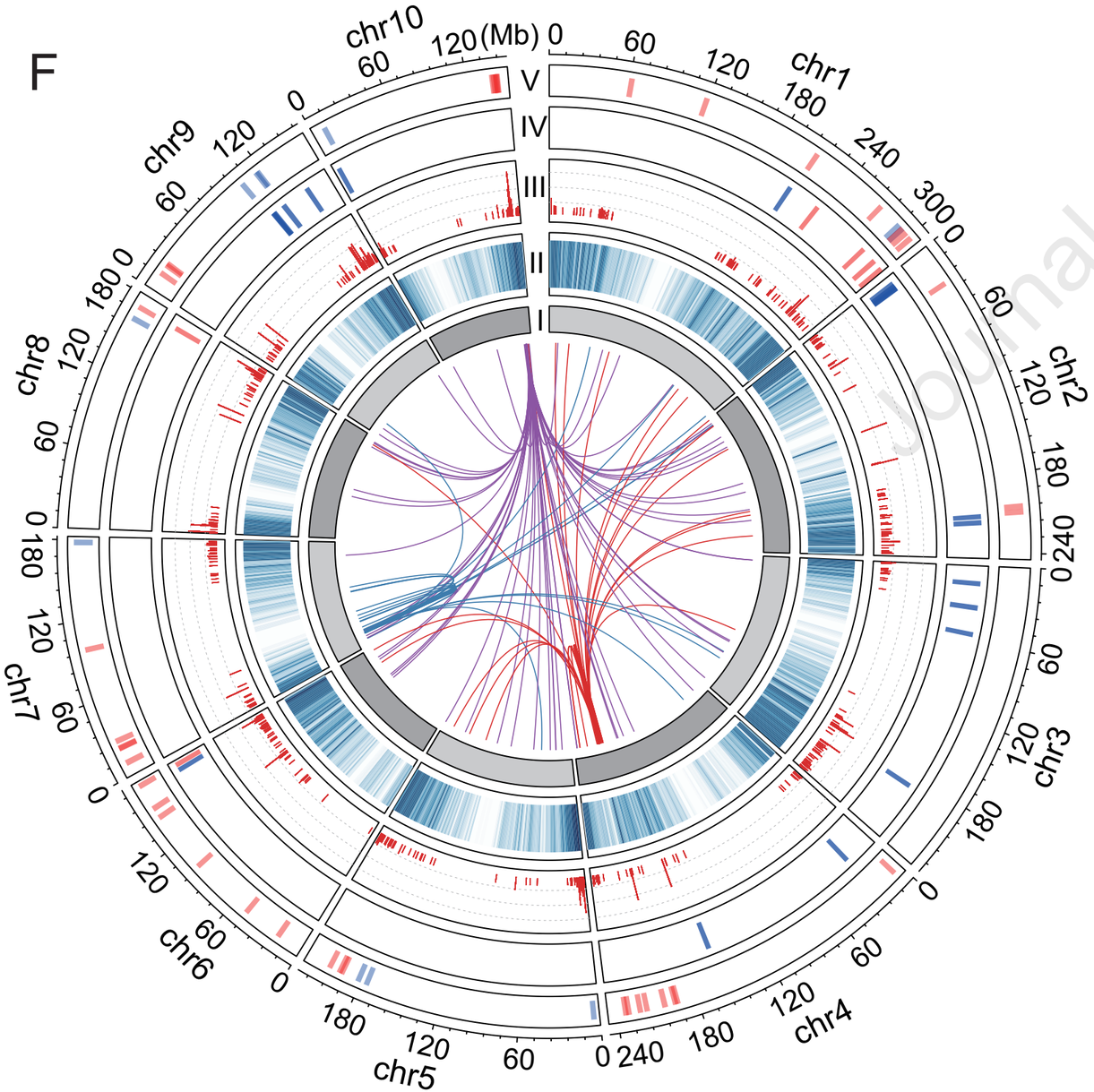
D



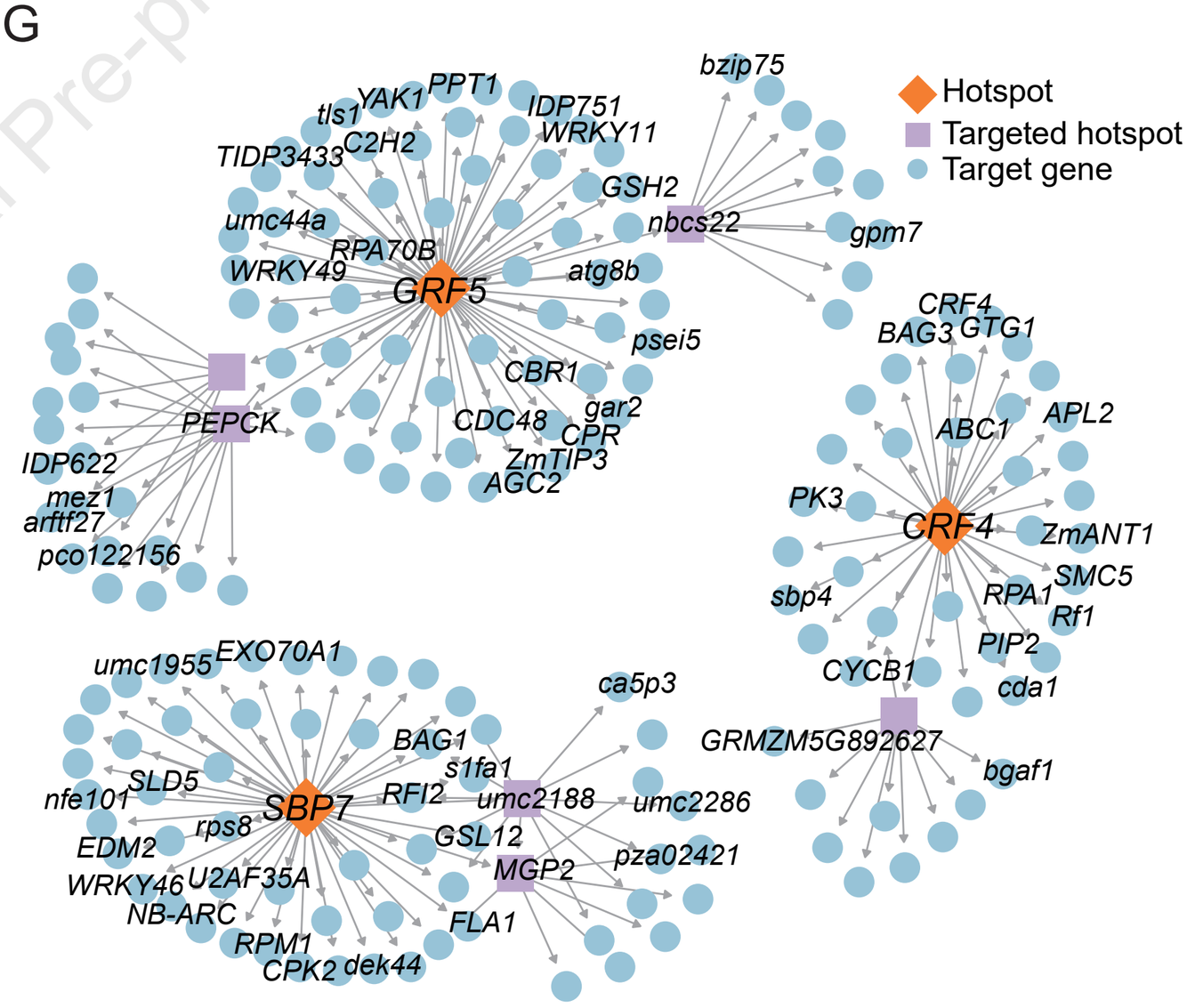
E



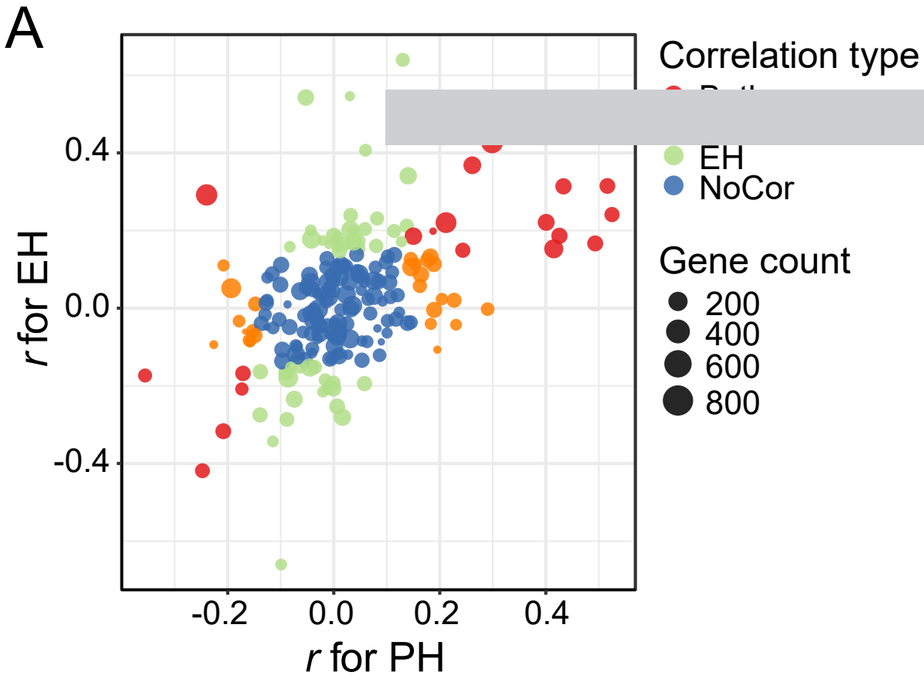
F



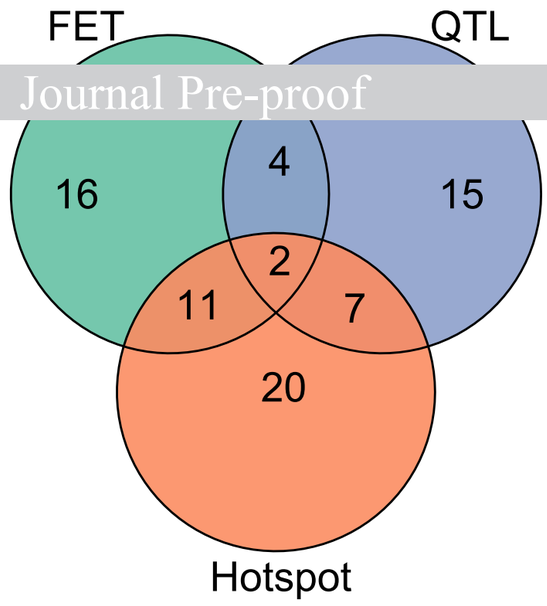
G



A

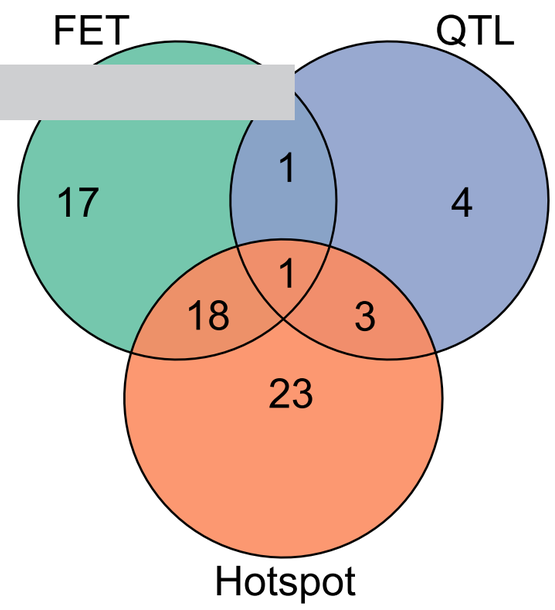


E



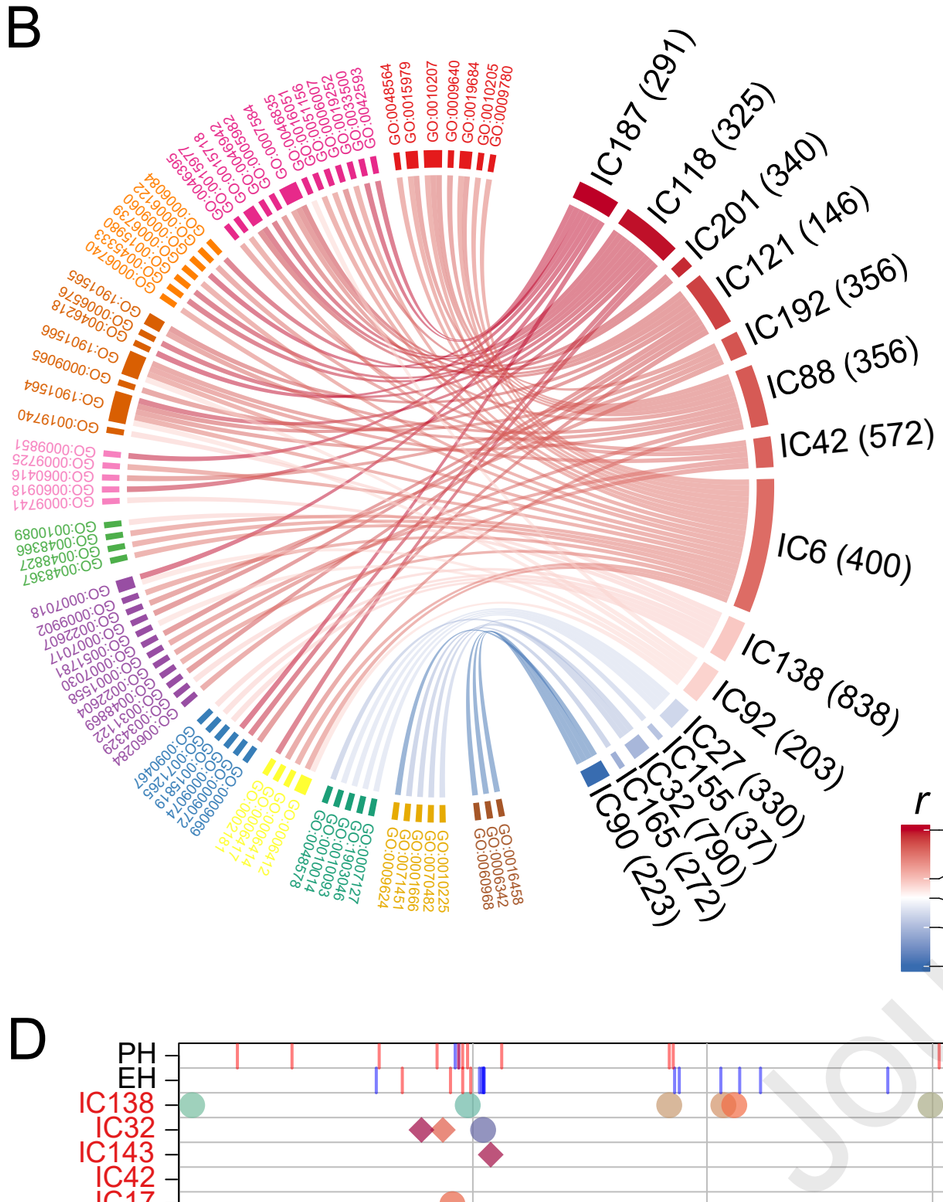
Number of meQTL for PH ICs = 104

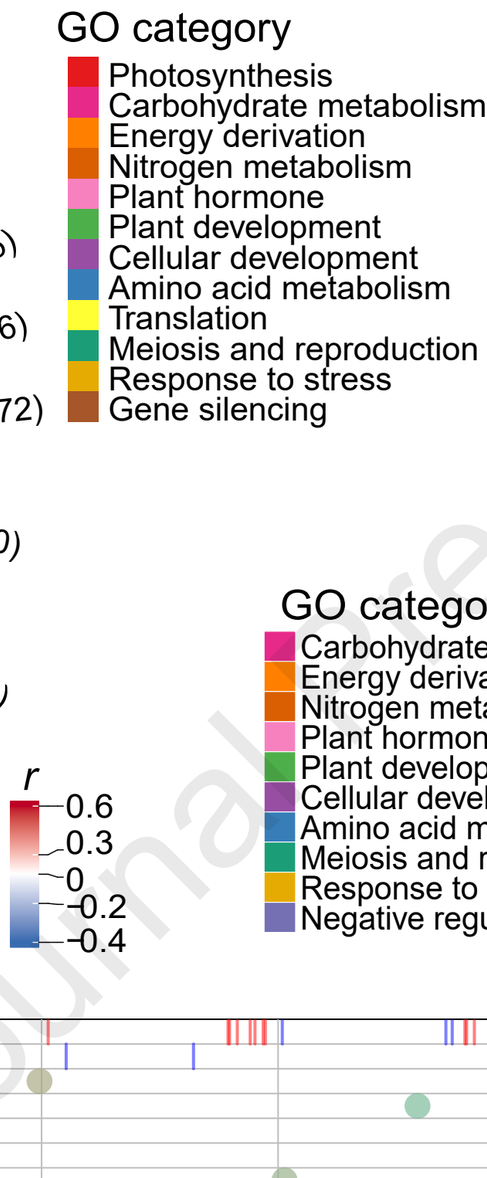
F



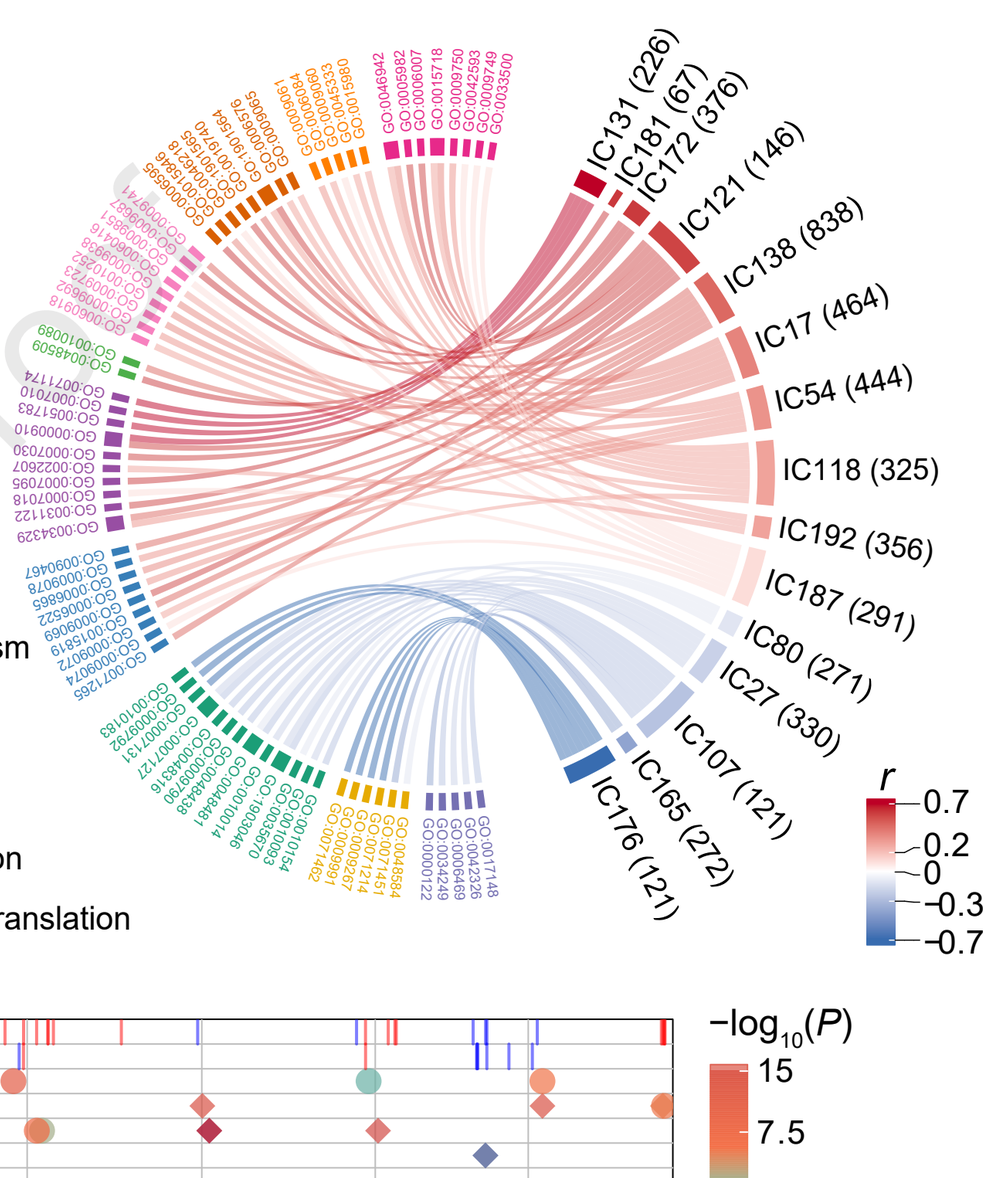
Number of meQTL for EH ICs = 101

B

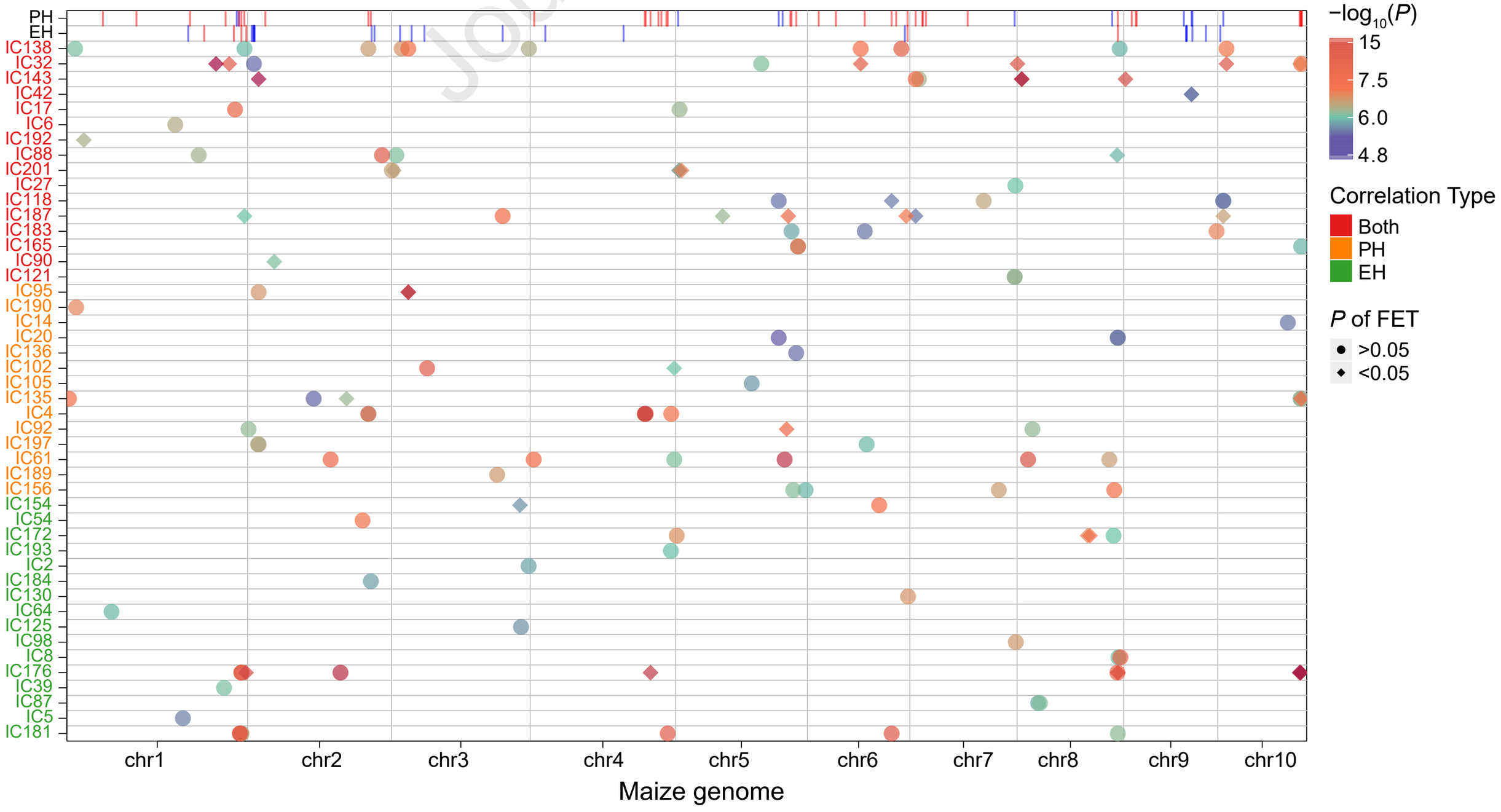




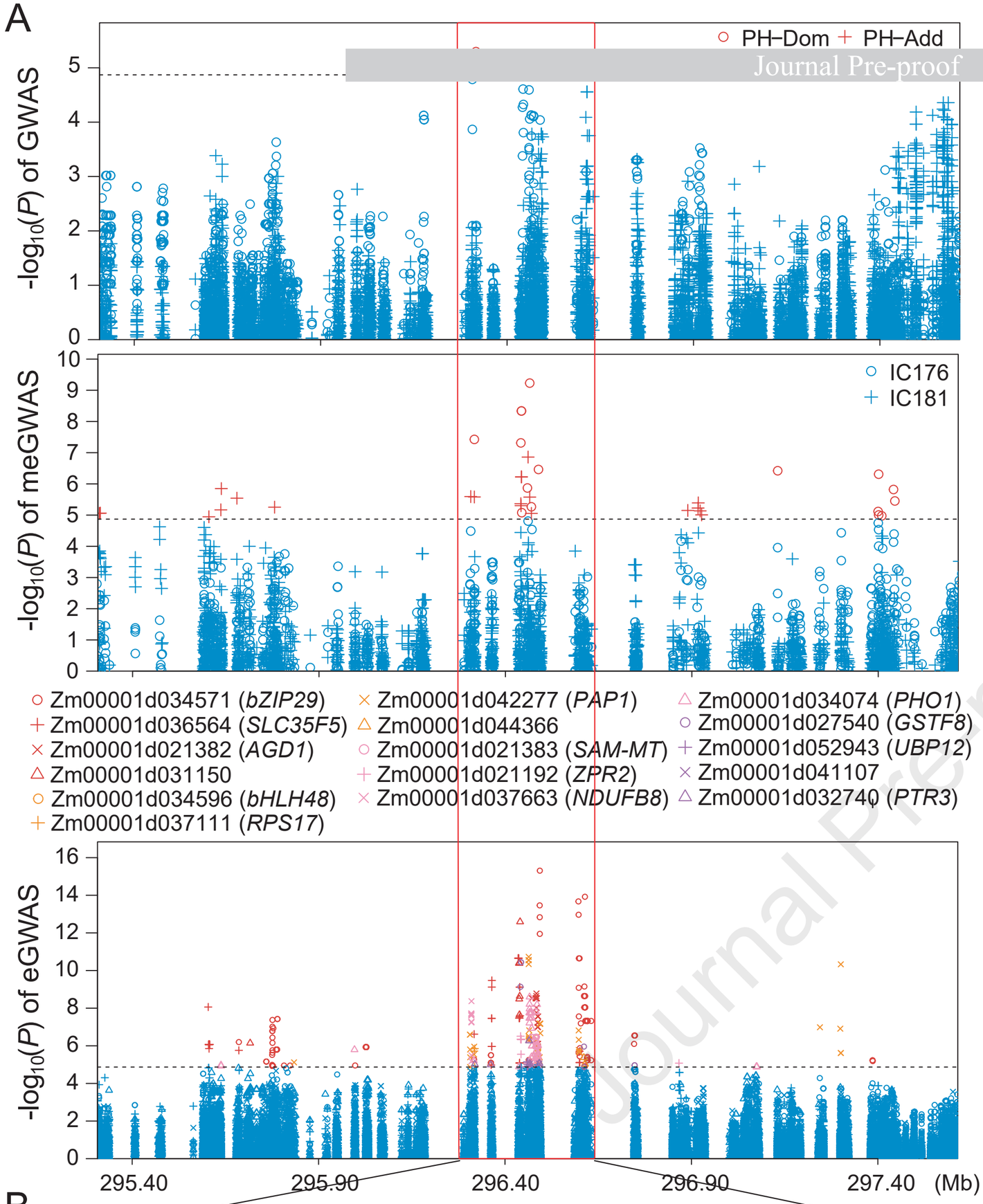
C



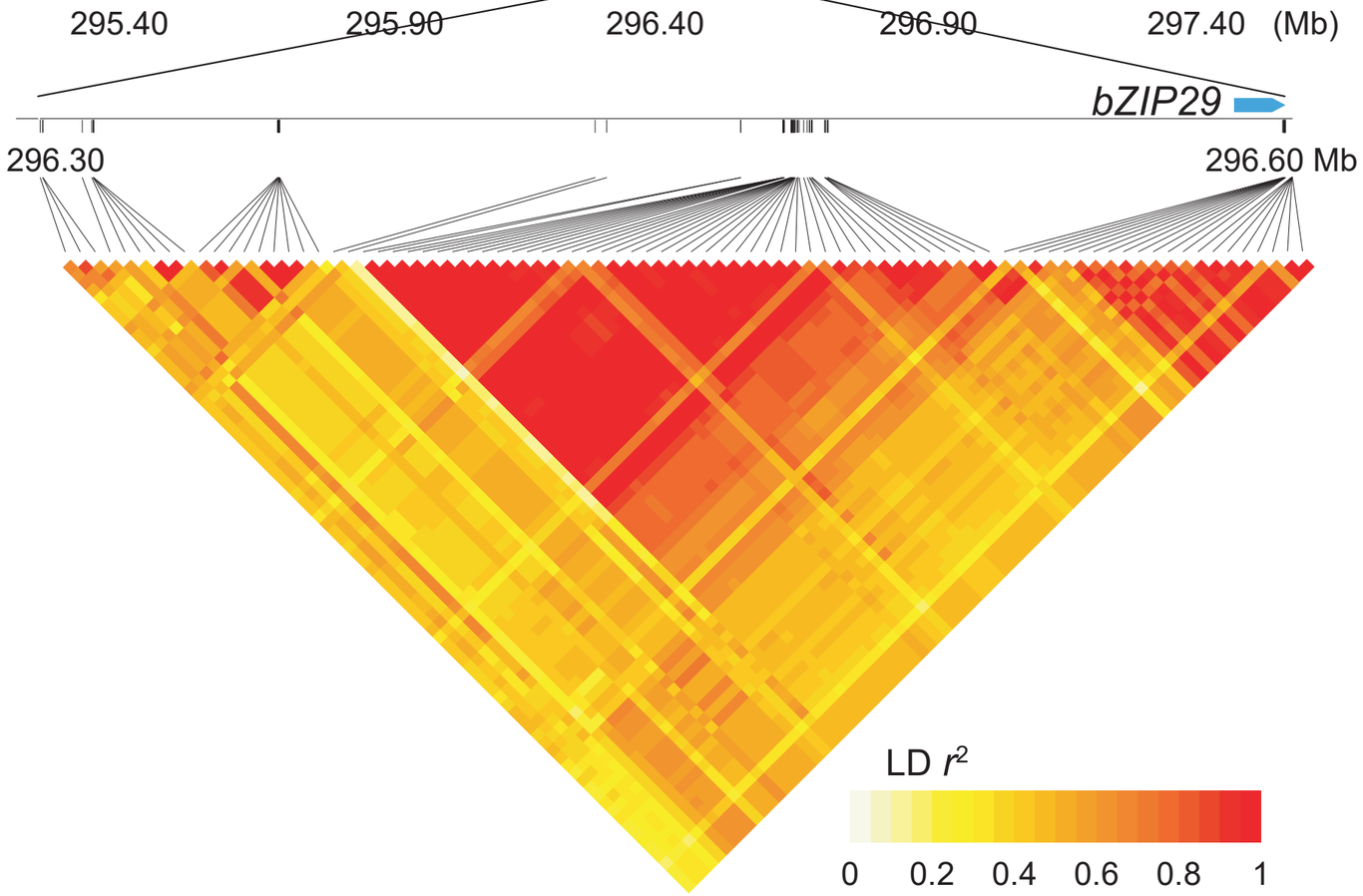
D



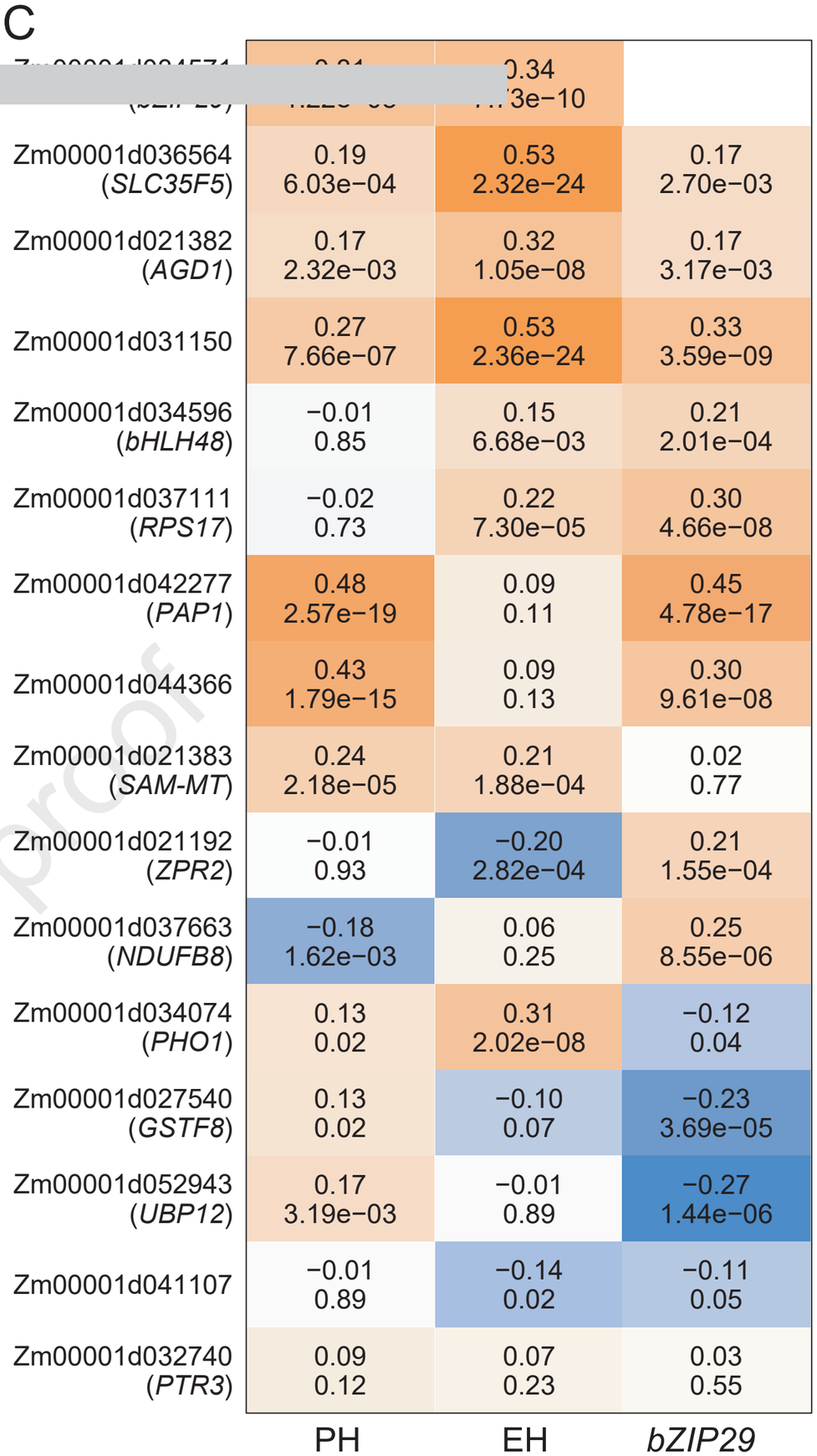
A



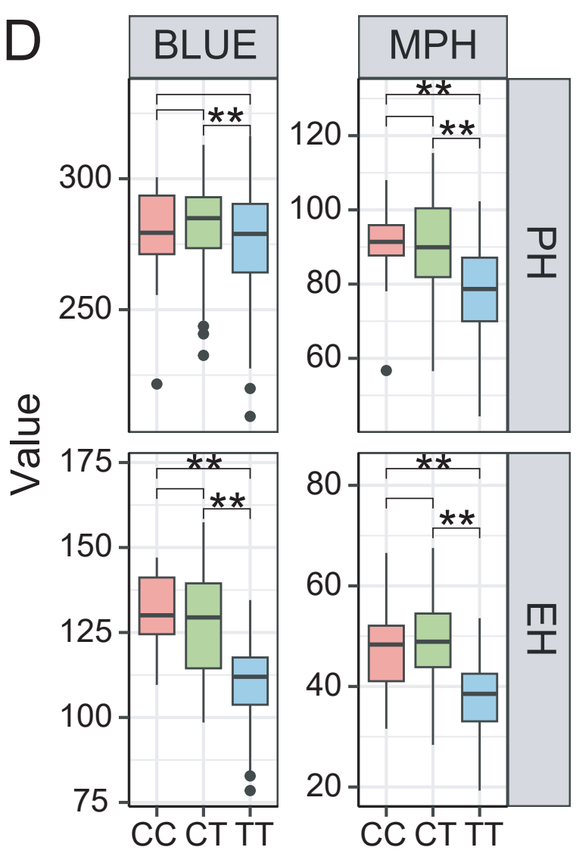
B



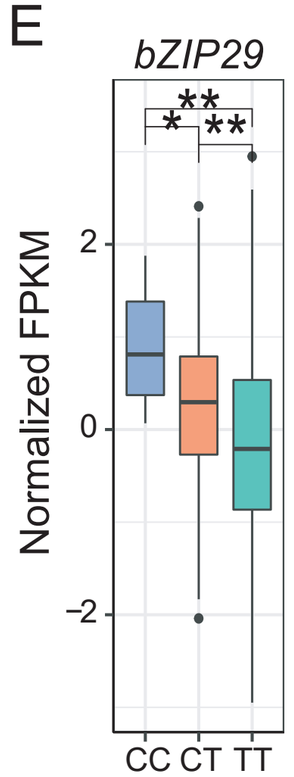
C



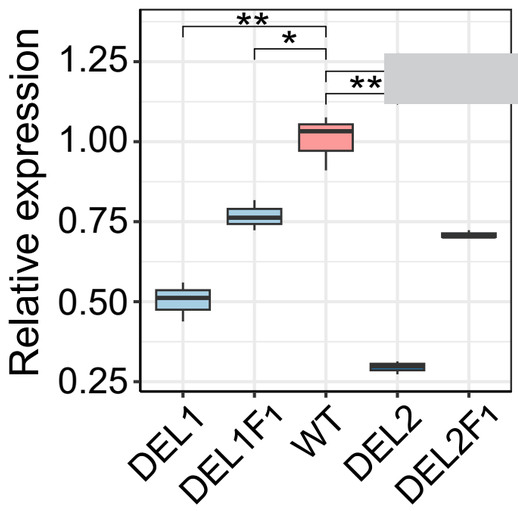
D



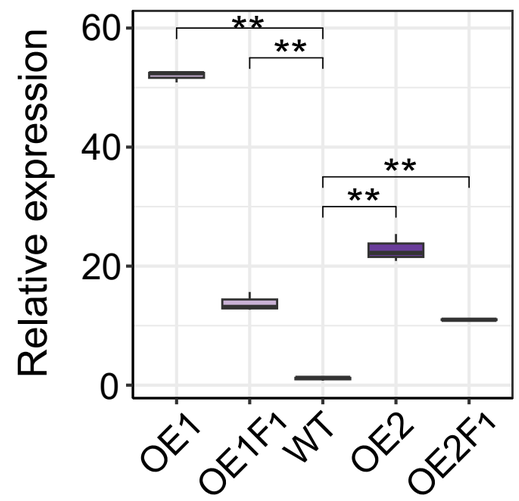
E



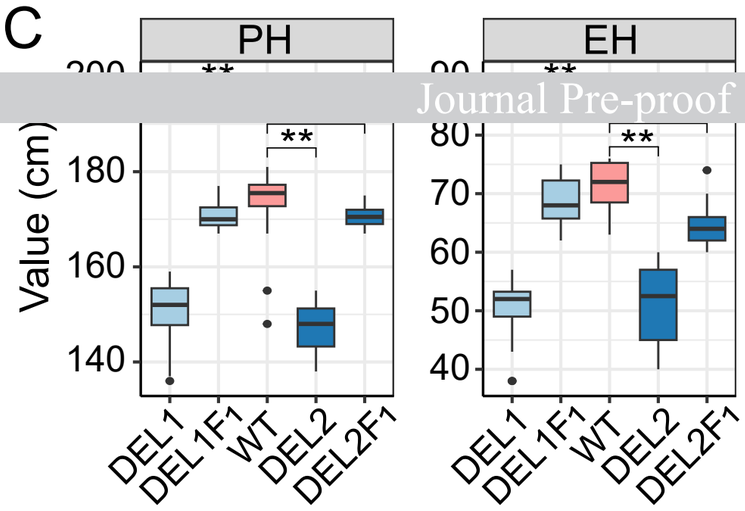
A



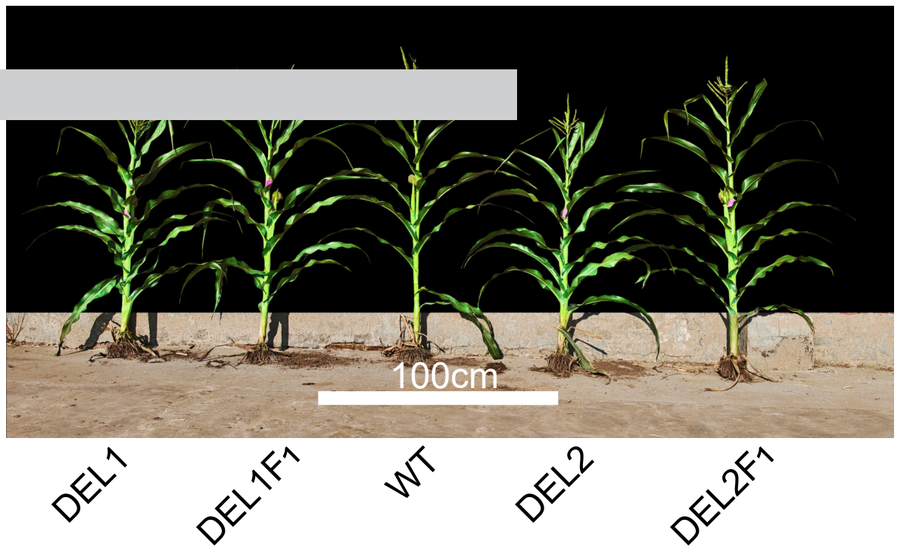
B



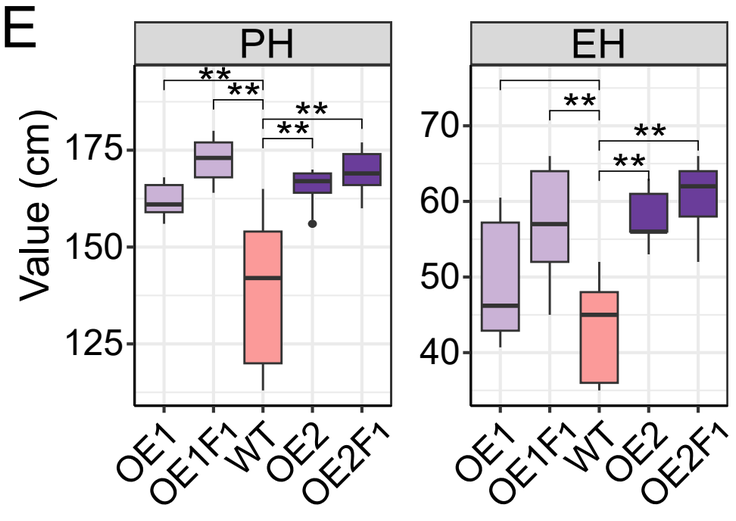
C



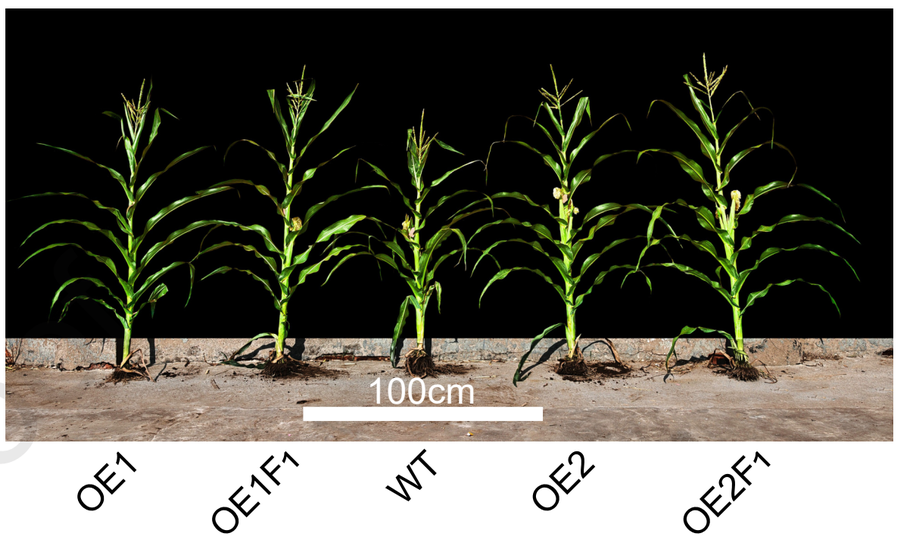
D



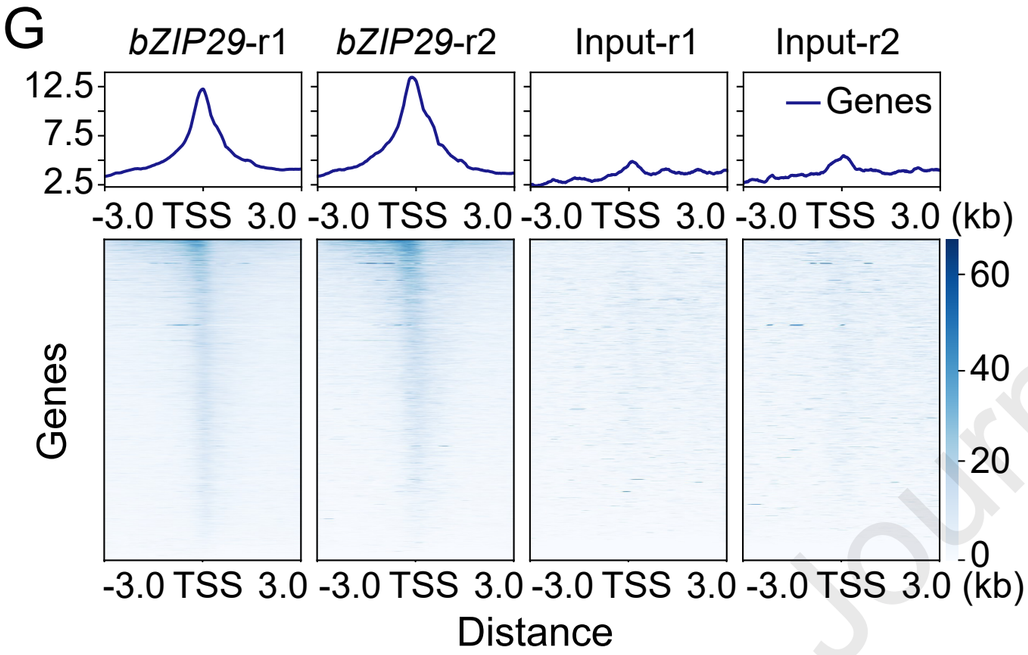
E



F



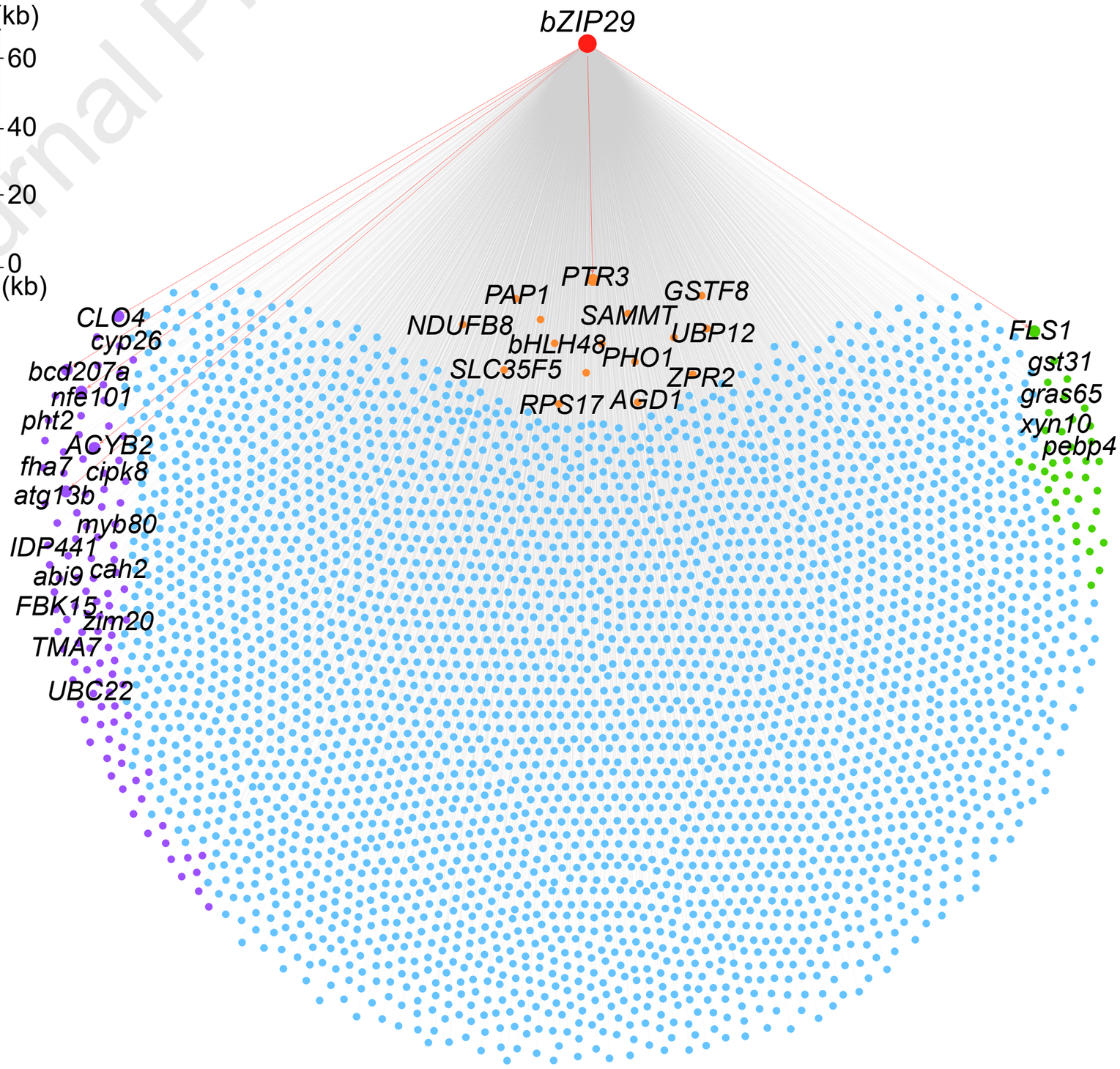
G



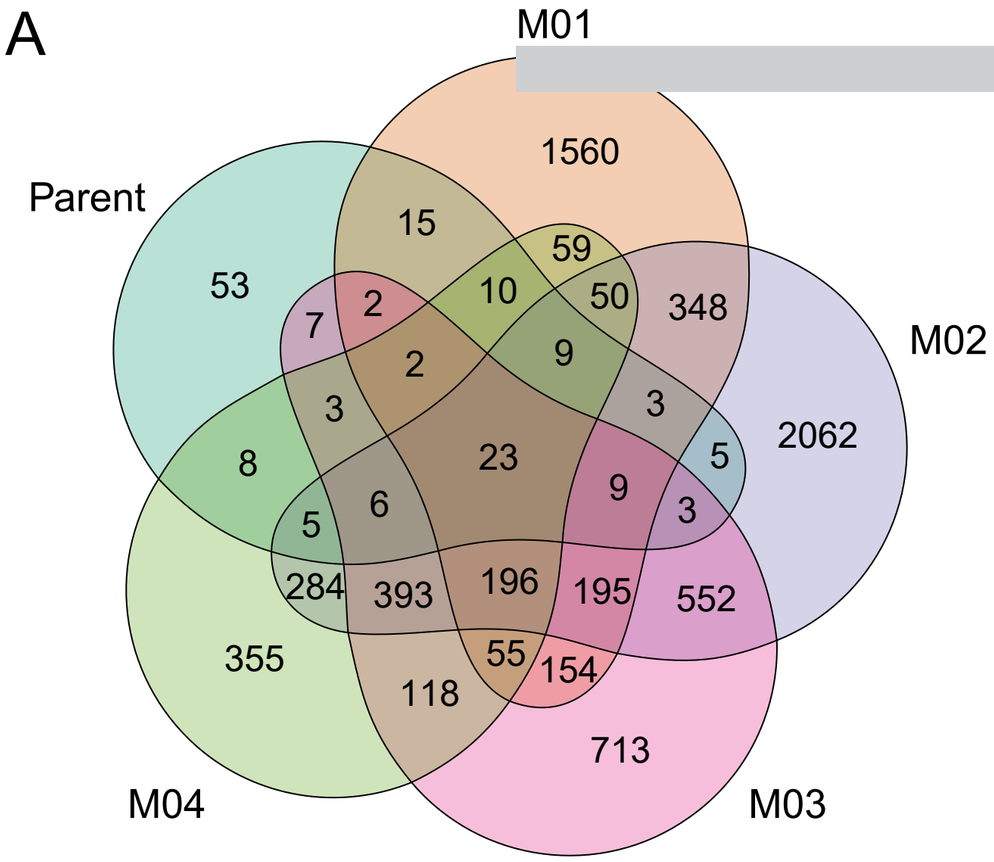
H

I

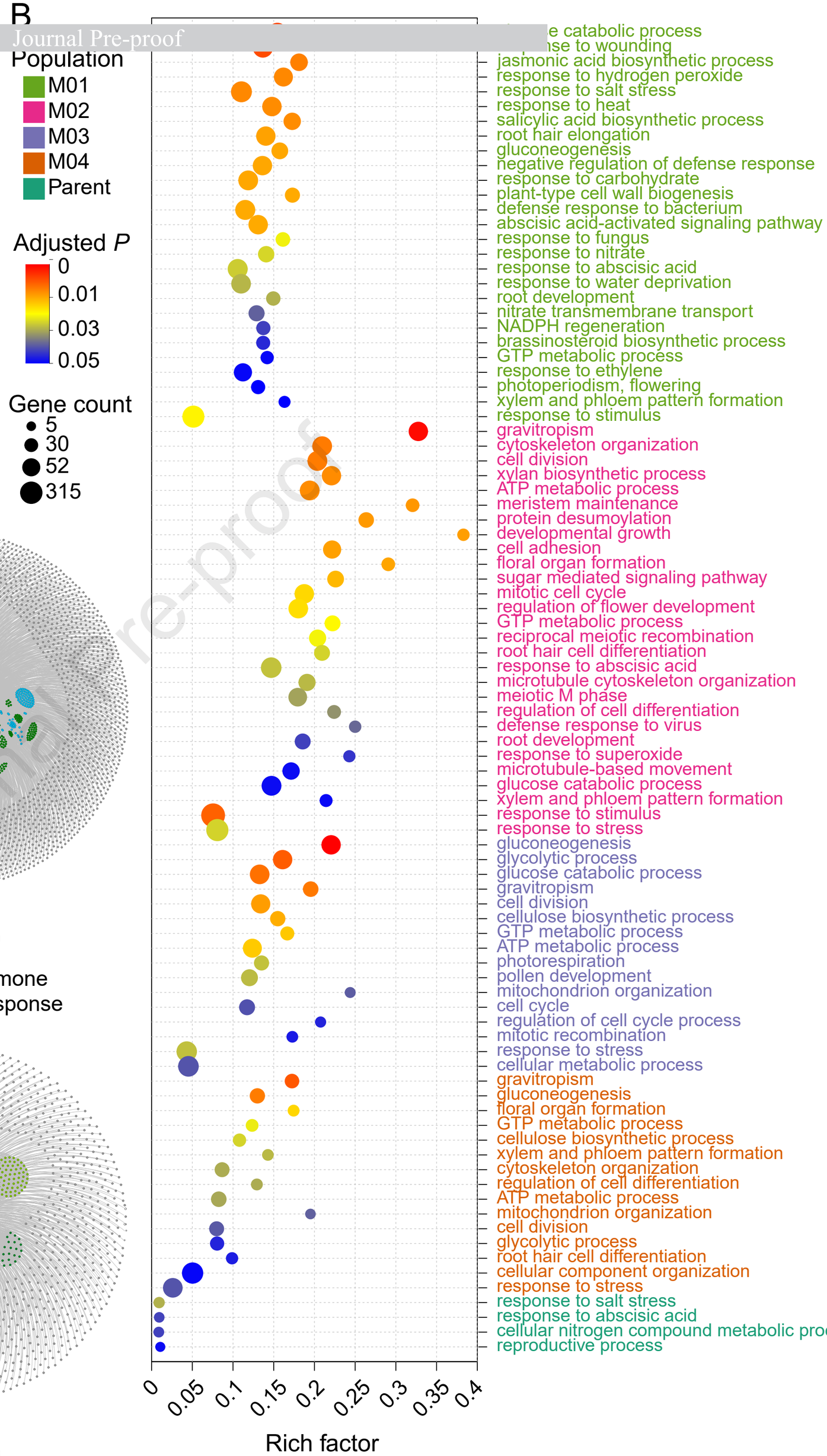
J



A



B



C