

Genome assembly and genetic dissection of a prominent drought-resistant maize germplasm

Received: 24 November 2021

Accepted: 11 January 2023

Published online: 20 February 2023

Tian Tian^{1,5}, Shuhui Wang^{1,5}, Shiping Yang¹, Zhirui Yang¹, Shengxue Liu¹, Yijie Wang¹, Huajian Gao^{1,4}, Shuaisong Zhang², Xiaohong Yang^{1,2,3}, Caifu Jiang^{1,2} & Feng Qin^{1,2}✉

 Check for updates

In the context of climate change, drought is one of the most limiting factors that influence crop production. Maize, as a major crop, is highly vulnerable to water deficit, which causes significant yield loss. Thus, identification and utilization of drought-resistant germplasm are crucial for the genetic improvement of the trait. Here we report on a high-quality genome assembly of a prominent drought-resistant genotype, CIMBL55. Genomic and genetic variation analyses revealed that 65 favorable alleles of 108 previously identified drought-resistant candidate genes were found in CIMBL55, which may constitute the genetic basis for its excellent drought resistance. Notably, *ZmRtn16*, encoding a reticulon-like protein, was found to contribute to drought resistance by facilitating the vacuole H⁺-ATPase activity, which highlights the role of vacuole proton pumps in maize drought resistance. The assembled CIMBL55 genome provided a basis for genetic dissection and improvement of plant drought resistance, in support of global food security.

Drought is a great environmental constraint on crop production and is responsible for causing a higher loss in annual crop yield than all pathogens combined¹. Due to global climate change and the resulting unpredictable precipitation patterns, the need for breeding high-yielding and stress-resilient varieties is more urgent than ever^{2,3}. Maize (*Zea mays*) is a crop that is cultivated worldwide and serves as a resource for food, feed and industry materials⁴. Based on the reports published between 1980 and 2015, drought was responsible for a 39.3% reduction in the yield of maize, which was even greater than the yield reduction (20%) in wheat (*Triticum aestivum*), indicating that maize is especially vulnerable to water deficit⁵. It has been estimated that less than 5% of the world's maize germplasm has been utilized in breeding programs in the United States⁶. This suggests that a diverse amount of genetic resources exist in maize that may represent excellent sources

for breeding drought-resistant varieties. Given that more than 85% of genetic content consists of complex repetitive sequences⁷, accurate and complete SV identification in maize genome is challenging as it relies heavily on the existence of high-quality genome sequences and identification methods⁸. The high-quality genome assemblies of B73 (ref. ⁹) and Mo17 (ref. ¹⁰), representatives of two heterotic groups of maize, have been accomplished, and they serve as excellent references for exploring the maize genome. Recently, the germplasms of 25 founder lines of a nested association mapping (NAM) population¹¹, SK¹² (small kernel), A188¹³, K0326Y¹⁴ and four European flint maize lines¹⁵ were de novo assembled, thus providing a pan-genome view for maize. None of these genotypes, however, were recognized as an elite drought-tolerant germplasm. Previously, we identified a tropical maize germplasm, CIMBL55, with prominent drought resistance and

¹State Key Laboratory of Plant Environmental Resilience, College of Biological Science, China Agricultural University, Beijing, China. ²Center for Crop Functional Genomics and Molecular Breeding, China Agricultural University, Beijing, China. ³National Maize Improvement Center of China, China Agricultural University, Beijing, China. ⁴Present address: School of Life Sciences and Medicine, Shandong University of Technology, Zibo, China.

⁵These authors contributed equally: Tian Tian, Shuhui Wang. ✉e-mail: qinfeng@cau.edu.cn

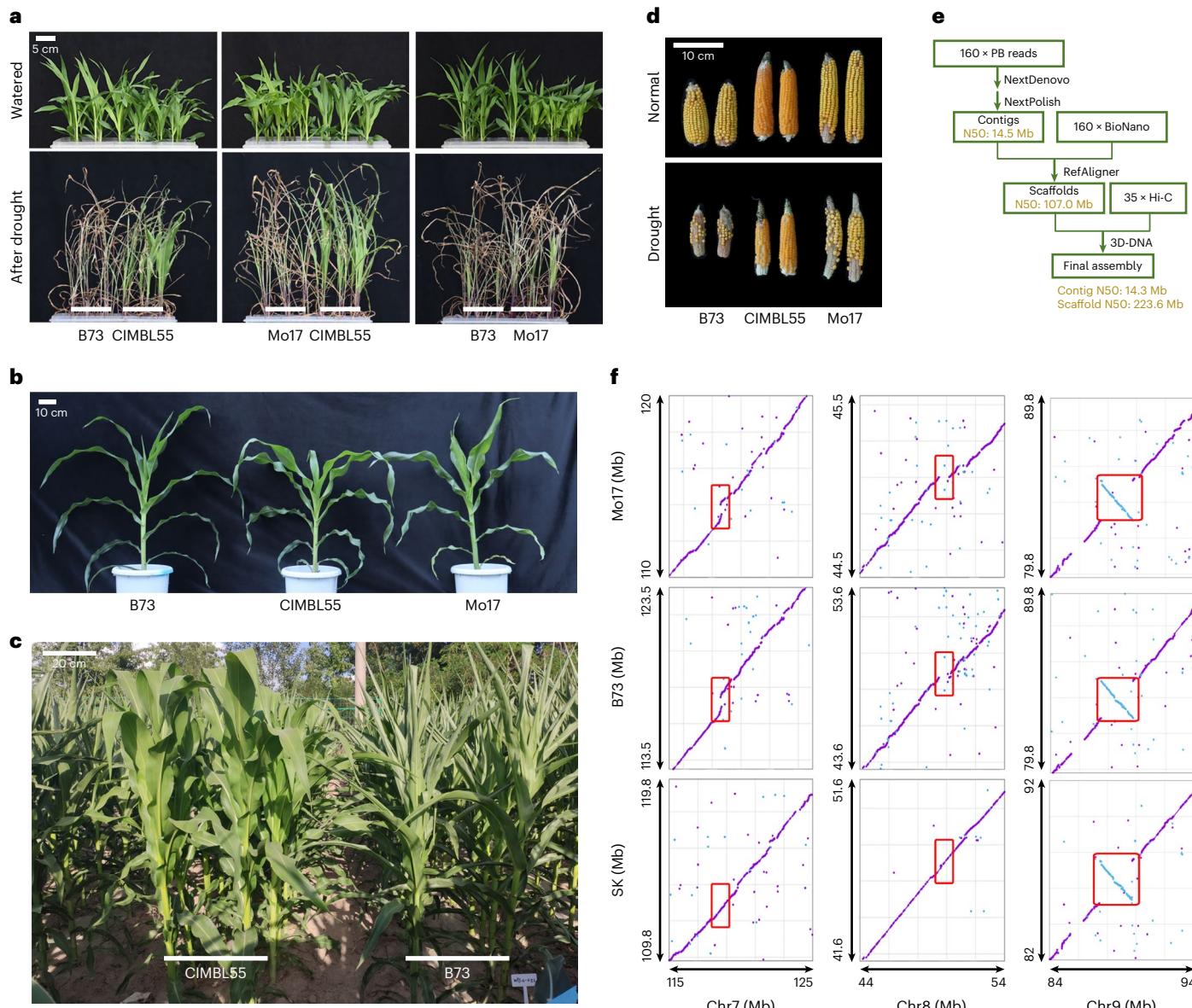


Fig. 1 | The drought-resistant phenotypes and genome assembly of CIMBL55. **a**, Seedling survival of B73, CIMBL55 and Mo17. Photographs were taken of the seedlings before or after the drought treatment when the differences were most evident. Scale bar, 5 cm. **b**, The morphology of representative, 1-month-old plants grown under normal, well-watered conditions in a greenhouse. Scale bar, 10 cm. **c**, Two-month-old plants of CIMBL55 and B73 subjected to water

deficit conditions in the field. Scale bar, 20 cm. **d**, Ears of B73, CIMBL55 and Mo17 collected under well-watered and drought conditions (see Methods). Scale bar, 10 cm. **e**, The pipeline used for the CIMBL55 genome assembly. **f**, Large SVs in CIMBL55, relative to B73, Mo17 and SK genomes. The x axis in each graph represents the CIMBL55 genome sequence of the indicated regions.

identified several favorable alleles contributing to this trait, such as *ZmVPP1* (ref. ¹⁶), *ZmNAC111* (ref. ¹⁷) and *ZmABH2* (ref. ¹⁸). In the present study, we provide a high-quality, de novo assembly of the CIMBL55 genome. The genome assembly enabled an intensive gene synteny survey, SV identification, epigenome analysis and drought-resistant gene cloning, which provided a better understanding of the genetic basis for maize drought resistance.

Results

Genome assembly and annotation of CIMBL55

CIMBL55 was derived from the tropical/subtropical (TST) maize germplasms that were included in a global germplasm collection for genetic association studies of complex quantitative traits¹⁹. CIMBL55 was the genotype with the most notable drought resistance within this

collection. For example, >70% of CIMBL55 seedlings survived a drought treatment while only ~20% of B73 and Mo17 seedlings survived the same treatment (Fig. 1a, Supplementary Fig. 1a–c and Supplementary Data). CIMBL55 plants grew and developed normally in a manner similar to most elite maize germplasms (Fig. 1b). Leaves of B73 began to exhibit rolling and wilt under a water deficit in the field, while CIMBL55 plants appeared normal and healthy (Fig. 1c). Notably, the yield loss in CIMBL55 (~40%) was substantially less than B73 (~98%) and Mo17 (~70%) under the same drought stress in field (Fig. 1d, Supplementary Fig. 1d and Supplementary Data).

To explore the genetic elements contributing to the drought resistance trait, the CIMBL55 genome was sequenced and de novo assembled using a combination of technologies, including 60× RSII and 100× Sequel 2 of PacBio single-molecule real-time sequencing,

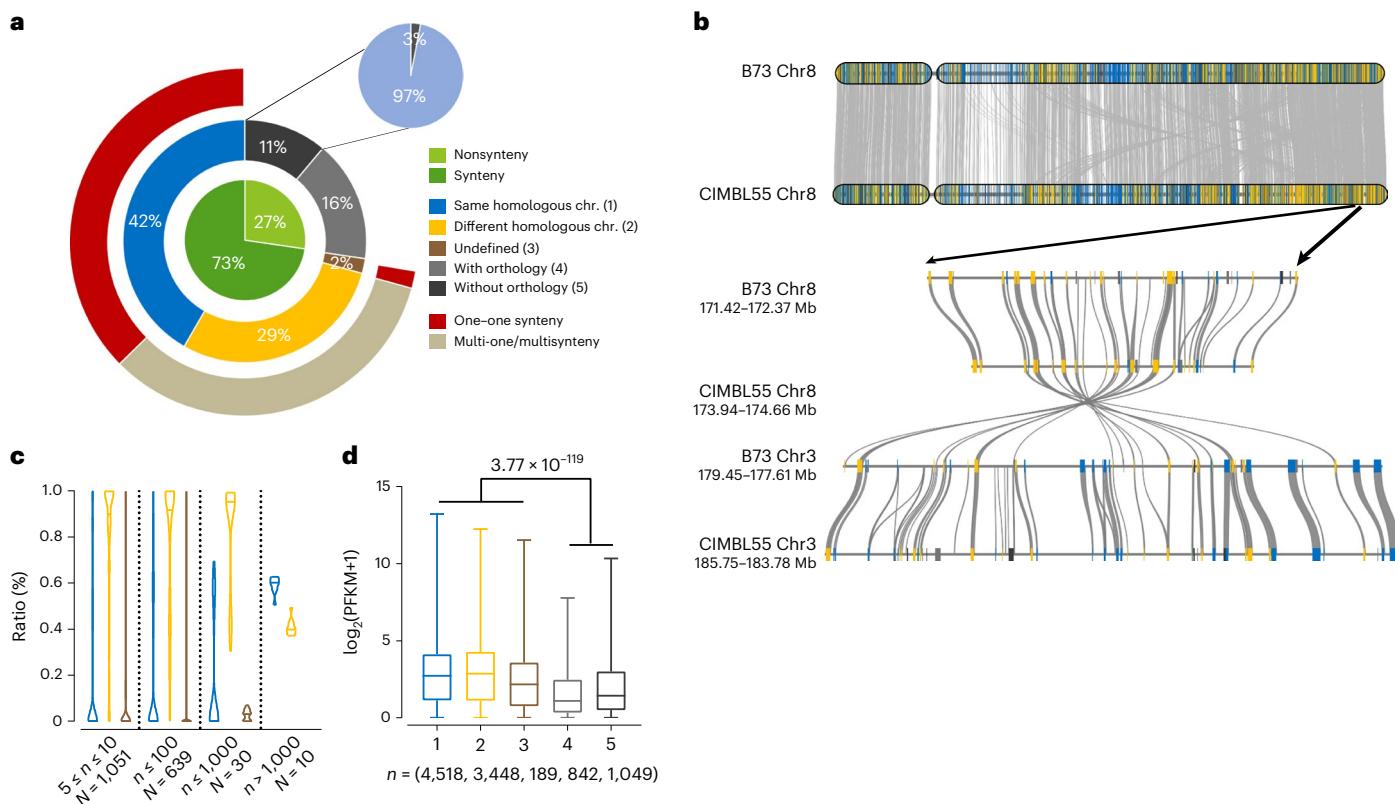


Fig. 2 | Gene synteny analysis in CIMBL55 and B73. **a**, Circular graph indicating the percentage of the five classes of genes in regard to their synteny status. The center circle represents the percentage of syntenic and nonsyntenic genes in CIMBL55, relative to B73. The middle circle indicates that the syntenic genes comprised of Classes 1, 2 and 3 genes, while the nonsyntenic genes comprised Classes 4 and 5 genes. The small pie chart indicates that Class 5 genes were further divided into two categories based on whether a homologous DNA sequence could be identified in B73 (see Methods). The outer ring indicates if the syntenic genes were present as a one-one synteny or multi-one/multi synteny arrangement. **b**, Synteny map between the B73 and CIMBL55 genomes. Upper panel, entire chromosome 8 of B73 and CIMBL55. Lower panel, zoomed view of the synteny blocks on B73 Chr8:171.42–172.37 Mb and CIMBL55 Chr8:

173.94–174.66 Mb, which also share homology with the synteny block on B73 Chr3: 179.45–177.61 Mb and CIMBL55 Chr3: 185.75–183.78 Mb. The syntenic genes were connected by gray lines. **c**, The ratio of Class 1–3 genes within the different-sized synteny blocks. N , the number of the synteny blocks of the indicated size; n , the number of syntenic genes in the synteny block. **d**, The expression level of Class 1–5 genes in seedlings, without the inclusion of nonexpressed genes. n : the number of expressed genes ($\text{FPKM} > 1$) in each class. For the box plots, boxes span the first and third quartiles, center lines represent medians and whiskers represent the smallest and largest sample values. Statistical significance was determined by a two-sided t -test. In all of the panels, the five classes of genes are presented in different colors as indicated in **a**.

160 \times BioNano Single molecule optical mapping and 35 \times Hi-C (high-throughput chromosome conformation capture) sequencing. The PacBio reads were initially assembled into 334 contigs with an N50 of 14.5 Mb. The initial contigs were then arrayed based on the BioNano optical map. A final total of 338 contigs were obtained after the removal of 96 conflicts and arrayed into 43 scaffolds with an N50 of 107.0 Mb. Then, the scaffolds were then placed on the ten chromosomes of maize, based on the Hi-C detection. The final assembly resulted in 116 scaffolds covering ~99.0% of the genome, having a final assembled genome size of 2,159.5 Mb and a scaffold N50 of 223.6 Mb (Fig. 1e). Genome integrity was calculated to be ~98.0% by bench-marking universal single-copy orthologs²⁰. Nine CIMBL55 bacterial artificial chromosome clones were also randomly selected and sequenced and could be perfectly aligned to the de novo assembled genome, supporting the quality of the assembly (Supplementary Table 1). Large translocation, duplication and inversion events between B73 and CIMBL55 were identified (Supplementary Fig. 2). A comparison with the B73 and Mo17 genomes identified an ~440 kb deletion on Chr7 and an ~560 kb insertion on Chr8 in both the CIMBL55 and SK¹² (another TST germplasm) genomes, indicating that these differences were likely common within the TST germplasms (Fig. 1f). An ~2.2 Mb inversion on Chr9 was specific to CIMBL55 among the four compared germplasms (CIMBL55, SK,

B73 and Mo17); however, it was also identified in 15 of 25 of the NAM founders¹¹, indicating that it was a common genetic variation within maize germplasms (Fig. 1f).

The annotation of the CIMBL55 genome revealed that ~83.95% of the genomic content was transposon elements (Supplementary Table 2). A total of 38,439 high-confidence genes were annotated in CIMBL55, having an average gene length of 4,289 bp. The circos plot revealed a similar distribution pattern for DNA-TEs and protein-coding genes that spanned the chromosome arms, and that the distribution of RNA-TEs was associated with the high level of CG and CHG methylation (revealed by the bisulfite sequencing conducted in this study) that was prominent near centromeres (Supplementary Fig. 3).

Gene synteny between CIMBL55 and B73 maize genomes
Gene synteny reflects the gene content and their arrangement on chromosomes between different genomes^{21,22}. The synteny analysis between the CIMBL55 and B73 genomes indicated that 73% of CIMBL55 genes were located in synteny blocks, and 27% of the genes were considered nonsyntenic genes (Fig. 2a). The syntenic genes were further divided into three classes. Class 1 genes (42%) were located on the homologous chromosomes of B73 and CIMBL55, with majority of them being single-copy genes (one-one synteny). Class 2 genes (29%) are

duplicated genes residing on two different homologous chromosomes in their respective genome (multi–multi synteny; Fig. 2a). For example, 15 pairs of Class 2 genes were located in the synteny block on Chr8 (B73, 171.42–172.37 Mb; CIMBL55, 173.94–174.66 Mb) and in the other synteny block on Chr3 (B73, 179.45–177.61 Mb; CIMBL55, 185.75–183.78 Mb) (Fig. 2b). Class 3 genes (2%) were in the contigs unplaced onto chromosomes in the current assembly of CIMBL55 but were in synteny blocks with their counterparts in B73. The nonsyntenic genes comprised Class 4 (16%) and Class 5 (11%) genes (Supplementary Table 3). Class 4 genes had orthologs but were not positioned in synteny as were their counterparts in B73. No orthologous protein sequences for the Class 5 genes (*e* value < 10⁻²⁰) could be identified, although 97% of them had DNA orthologs in B73, indicating that this class of genes may have undergone significant gene structural diversification or functional loss. The identified ten greatest synteny blocks correspond to the ten maize chromosomes. Class 2 genes tended to be present in smaller synteny blocks (containing <1,000 genes), relative to Class 1 genes (Fig. 2c). Syntenic genes had a higher transcript abundance than nonsyntenic genes based on the RNA-seq data obtained from leaf samples, indicating that genomic rearrangement likely impaired gene expression¹⁸ (Fig. 2d). The maize genome underwent a whole genome duplication event from its ancient ancestor, followed by biased fractionation and diploidization²³. Referring to the sorghum genome, the subgenome organization of CIMBL55 was constructed and the fractionation bias was found to be consistent with B73, supporting the evolution conservation within *Zea mays* (Supplementary Fig. 4a,b). Then, we found that 61.5% Class 2 genes were pair-retained genes in the two subgenomes of CIMBL55, indicating that Class 2 genes remained relatively conserved during maize evolution (Supplementary Fig. 4c). Interestingly, they were found to be enriched in the Gene Ontology (GO) term²⁴ response to various stimuli (Supplementary Table 4). To gain insights into it, we further investigated the gene synteny of several gene families involved in ABA signaling and stress response, including genes encoding 13 ABA receptor (PYLs)²⁵, 13 SnRK2s protein kinase²⁶, 103 protein phosphatases (PP2Cs)²⁷ and 55 dehydration responsive element binding proteins/C-repeat binding factor²⁸, 130 NAM, ATAF and CUC (NACs)²⁹ and 128 bZIPs²⁹ transcription factors (Supplementary Table 5). Among them, 72% belonged to Class 2 genes, which was remarkably higher than 29% Class 2 genes on the whole genome level (Supplementary Fig. 5a,b and Fig. 2a). These data indicate that genes involved in environmental response and adaptation tended to be retained in duplication during maize evolution.

Genetic variants contributing to drought resistance

To identify the genetic elements contributing to the prominent drought resistance of CIMBL55, we explored the genetic variants on a pan-genome level, including 25 NAM founder lines¹¹ and another four maize accessions, SK¹², Mo17 (ref. ¹⁰), K0326Y¹⁴, A188¹³ (all of them assembled with contig N50 > 5.0 Mb). To identify a maximum number of genetic variants, 30 assembled maize genome sequences were aligned and compared with B73_v5 (ref. ²¹) through the following four programs: SyRi³⁰, Smartie-sv³¹, CuteSV³² and Sniffles³³. As a result of this analysis, 17,581,014 structure variations (SVs), including 16,311,409 insertions/deletions (InDels), 239,611 duplications, 496,685 inversions, 499,973 translocations and 33,336 copy number variations, were obtained non-redundantly (Fig. 3a and Supplementary Table 6). SyRi (based on the assembled genome sequence comparisons) showed effectiveness in discerning various types of variations, while Smartie-sv (based on the assembled contig alignments) and CuteSV (based on the long-read alignments) performed better in identifying InDels (>50 bp) than other programs (Fig. 3a). To gain more insight into the genomic variations in CIMBL55, we also called the genetic variants identified from B73 and Mo17, using CIMBL55 as a reference. As a result, 841,911 DNA variants (>20 bp) were subsequently identified (Supplementary Table 7). Then, a total of 544,853 SVs were successfully genotyped in the association

panel consisting of 368 maize germplasms, of which the drought tolerance was previously analyzed regarding the seedling survival rate after drought^{16,34}.

Previously, genome-wide association study (GWAS)¹⁶ and Mendelian randomization analysis¹⁸ of this association panel using SNPs predicted 42 and 97 genes as candidate drought resistance genes, respectively. To understand the prominent drought resistance of CIMBL55, we investigated the genotype of these loci in CIMBL55. Based on the previous SNPs information, the genotypes of 108 genes were available in CIMBL55. The superior alleles of 79 genes were contained in CIMBL55, which were defined as if the germplasm carrying the allele averaged had a markedly higher survival rate after drought than those harboring the counterpart allele (Fig. 3b and Supplementary Table 8). Because many of these candidate genes were predicted to be associated with drought resistance based on their variations in gene expression¹⁸, we looked for the potential SVs substantially associated with drought resistance that may have potential roles in affecting the gene expressions. Resultantly, significant SVs associated with drought tolerance were identified for 69 genes, which were in strong linkage disequilibrium (LD, $r^2 > 0.5$) with the previous lead SNPs of 79 genes (Fig. 3c). Notably, the superior haplotype (both the SNP and SV were the drought-resistant alleles) of 65 genes were found in CIMBL55, explaining the predominant drought resistance of CIMBL55 among the population (Supplementary Table 8). Moreover, among 108 genes, the superior haplotype of 11 candidate genes was identified in germplasms other than CIMBL55 (non-CIMBL55 allele), suggesting that other germplasms may contain complementary genetic resources to CIMBL55, concerning maize drought resistance (Fig. 3b and Supplementary Table 8).

The SNP (S1474) and SV (S3205, 42 bp) located in the second intron of *ZmABF4* (*Zm00001d031790*, the B73_v4 gene code in maizeGDB, <https://www.maizegdb.org/>) were found to be substantially associated with the gene expression and plant drought resistance (Supplementary Fig. 6a). *ZmABF4* encodes a bZIP transcription factor, an ortholog of *Arabidopsis* ABA-responsive factor 4, which has been characterized as a major regulator of ABA- and drought-inducible gene expression^{35,36}. The germplasms harboring the allele of *ZmABF4*^{CIMBL55} genotype had a substantially higher level of gene expression and drought resistance, compared with those carrying the *ZmABF4*^{B73} allele (Fig. 3d,e). Transgenic maize overexpressing the coding sequence of *ZmABF4*^{B73} was generated to determine if increased gene expression conferred a drought-resistant phenotype (Supplementary Fig. 6b and Supplementary Data). Results showed that the *ZmABF4* transgenic lines had a remarkably enhanced survival rate after drought exposure compared to nontransgenic wild-type (WT) plants and reduced rates of leaf water loss in response to dehydration treatments (Fig. 3f–h). It indicated that *ZmABF4* positively regulated maize drought resistant and *ZmABF4*^{CIMBL55}, with a higher level of gene expression, was probably the favorable allele.

Epigenomic variations that affect drought resistance

The high-quality assembled genome sequence in our study provided the ability to accurately evaluate and compare the DNA methylation status in the B73, Mo17 and CIMBL55 genomes by conducting whole genome bisulfite sequencing (BS-seq) of the three germplasms (Supplementary Table 9). The overall level of mCG and mCHG was considerably high throughout the genome but clearly decreased at gene boundaries (Supplementary Fig. 7a–c). In contrast, mCHH content was generally low level across the genome, including RNA-TE regions, whereas it was relatively high in DNA-TE elements, suggesting that different mechanisms regulate CHH methylation in these two types of TEs (Supplementary Fig. 7a–c). Interestingly, 24-nt small RNA (sRNA) preferentially mapped to gene boundaries and within DNA-TE regions, resembling the pattern of mCHH, indicating a possible involvement of RNA-directed DNA CHH methylations in these regions (Supplementary Fig. 7d).

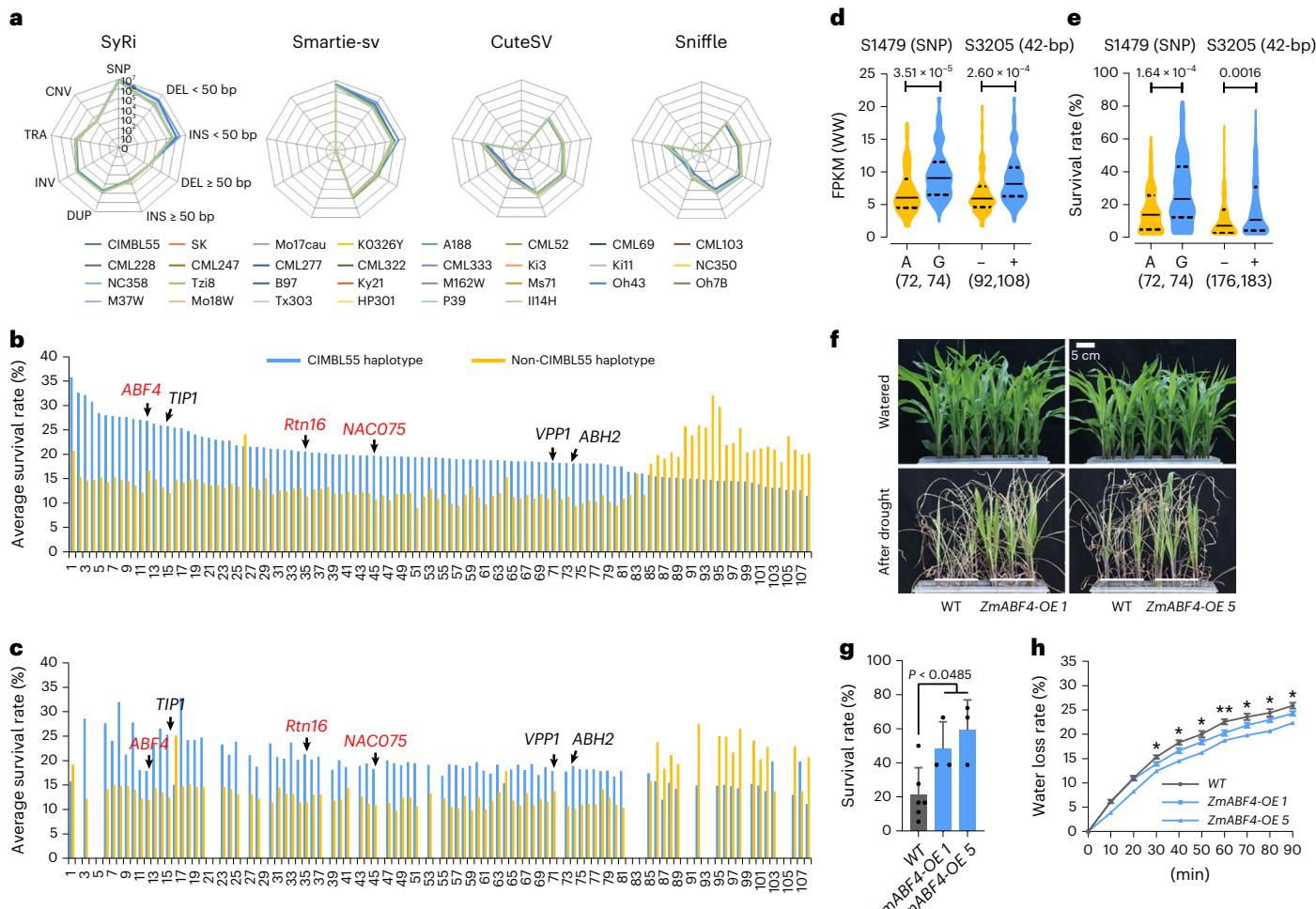


Fig. 3 | Identification of genetic variants and their associations with drought resistance. **a**, Statistics for SNP, deletion (DEL), insertion (INS), duplication (DUP), inversion (INV), translocation (TRA) and copy number variation (CNV) computed by four kinds of programs between B73 and other 30 high-quality maize assemblies. **b**, Average drought-survival rate of the germplasms harboring either the CIMBL55 or non-CIMBL55 genotype of 108 candidate drought-resistant genes based on the previously identified lead SNPs. The loci are arrayed in descending order of the CIMBL55 allelic effect. **c**, Average drought-survival rate of the germplasms harboring either the CIMBL55 or non-CIMBL55 haplotype of significant SVs. Three genes analyzed in the present research are indicated in red, and three previously reported genes^{16,18,50} are indicated in black. The blue and yellow colors represent the CIMBL55 and non-CIMBL55 genotypes. In **b** and **c**, 'Zm' for *Zeae mays* was omitted in the gene names. **d**, Violin plots of *ZmABF4* expression levels and survival rate (**e**) of the germplasms with different genotypes at the S1479 and S3205 loci of *ZmABF4*. For the violin plots (in this and

all following figures), dashed lines represent the first and third quartiles, center lines represent medians, and whiskers represent the smallest and largest sample values. The yellow and blue colors represent the B73 and CIMBL55 genotypes. 't', with the insertion; '−', without the insertion. The numbers in the parentheses denote the sample size used in the two-sided *t*-test. **f**, Drought survival assay of two *ZmABF4*-overexpression (*ZmABF4-OE*) lines. Scale bar, 5 cm. **g**, Percentage survival rate. Eighteen plants of each genotype were tested in each experiment, and at least three replicated experiments were conducted. Data represent the mean ± s.d. **h**, Rate of leaf water loss of WT and *ZmABF4-OE* lines. Four detached leaves of each genotype were left to dehydrate on a clean bench and weighed over a period of 90 min. The experiment was repeated three times for WT and OE 1 (mean ± s.d.) and twice for OE 5 (mean value). Statistical significance between WT and all the OE plants was determined, by a two-sided *t*-test: **P* < 0.05, ***P* < 0.01 (in **g** and **h**).

We identified differential DNA methylations in SV-related genomic regions based on the assembled CIMBL55 and B73 genome sequences. A total of 5,346 insertional fragments identified in CIMBL55, which could be bidirectionally confirmed (see Methods), were used for SV-related DNA methylation analysis. The DNA methylation patterns of the insertional sequence (including its 0.5-kb-flanking regions) were clustered into five groups as determined by hierarchical clustering based on the levels of mCG, mCHG and mCHH (Fig. 4a–c and Supplementary Table 10). Insertions belonging to clusters 1–3 exhibited considerably high levels of mCG, mCHG and mCHH, whereas those belonging to cluster five displayed generally low levels of DNA methylation (Fig. 4b). Notably, we found that insertions of cluster four had remarkably higher levels of mCG, mCHG and mCHH within the insertional fragments

compared to those in the flanking sequences (Fig. 4b,c). DNA-TEs, especially DTH (DNA transposon terminal inverted repeat Harbinger), were enriched in this cluster of insertions (Fig. 4b). The insertion length was concentrated in the range of 100–1,000 bp and was relatively close to the transcription start site of the neighboring gene, indicating that DNA-TEs, especially DTH, proximal to genes tended to be methylated CHH (Fig. 4b, Supplementary Fig. 8a,b and Supplementary Data).

Differential DNA methylation was identified in the promoter region of *ZmNAC075* (*Zm00001d008817*), which encodes a NAC-domain-containing transcription factor. Several members of this gene family have been reported to positively function in water deficit tolerance in several plant species^{17,37–39}. A comparison of the genomic sequence of CIMBL55 and B73 revealed two insertions in the upstream

region of *ZmNAC075^{B73}*, S-9041 (182 bp) and S-1425 (5,123 bp), which led to the observed hypermethylation of CG, CHG and CHH, relative to *ZmNAC075^{CIMBL55}*. The two insertions showed a strong LD with a previously identified SNP in the 3'-UTR, which was predicted to be a local eQTL for gene expression¹⁸ (Fig. 4d). Moreover, germplasm carrying the *ZmNAC075^{B73}* allele had a substantially lower level of gene expression and a reduced survival under drought stress than germplasm containing the *ZmNAC075^{CIMBL55}* allele (Fig. 4e,f). To unravel the effect of the hypermethylation of two SVs on the *ZmNAC075^{B73}* expression, we analyzed two CRISPR-targeted gene knock-out (KO) lines of a maize *DRD1* gene (*zmdrd1-KO1* and *zmdrd1-KO2*) in an LH244 genetic background which carries the *ZmNAC075^{B73}* allele (Supplementary Fig. 9). Arabidopsis *DRD1* (defective in RNA-directed DNA methylation 1) was reported to facilitate RNA-directed de novo DNA methylation which might function to epigenetically regulate the expression of adjacent genes^{40,41}. The whole genome BS-seq analysis revealed that the mCHH levels of both S-9041 and S-1425 in the *zmdrd1* mutants were substantially diminished compared with the WT (LH244), while the mCG and mCHG levels were not changed (Fig. 4g–k). Concomitantly, the expression of *ZmNAC075^{B73}* was revealed to be substantially increased, indicating that the CHH hypermethylation of the two newly identified SVs upstream *ZmNAC075^{B73}* may cause the suppression of gene expression (Fig. 4l).

ZmRtn16 has a positive role in drought resistance

ZmRtn16 (*Zm00001d047517*) which encodes a reticulon-like protein was among the 65 superior alleles identified in CIMBL55. Members of this protein family were reported to function in endomembrane autophagy in aleurone cells during seed germination⁴² and protein trafficking from the endoplasmic reticulum (ER) to the Golgi apparatus⁴³. Several genetic variants within or near to *ZmRtn16* were determined to be substantially associated with the gene expression level under different conditions, indicating a strong local regulation of the gene expression (Fig. 5a–c). We found that a newly identified insertion, S2290 (28-bp), in the 3'-UTR was substantially associated with the plant survival rate after drought exposure and it was in strong LD with other significant nonsynonymous SNPs located in the intron and last exon (Fig. 5d). The germplasms harboring the *ZmRtn16^{CIMBL55}* allele (without 28 bp) had a substantially higher expression level and greater drought stress resistance than those carrying the *ZmRtn16^{B73}* allele (with 28 bp), indicating that this variant may affect the gene expression and drought resistance (Fig. 5e,f). To determine if this variant really affected *ZmRtn16* mRNA abundance, we cloned the 3'-UTR sequence from B73 and CIMBL55, which contained a previously identified significant SNP (S2181) and the 28-bp SV (S2290). The two fragments were fused downstream to the coding sequence of *ZmRtn16^{B73}*, and the constructed plasmids were transfected into maize leaf protoplasts. Results indicated that the construct containing the sequence cloned from CIMBL55 produced a dramatically greater abundance of *ZmRtn16* transcripts than the B73 construct. Moreover, changing the nucleotide of S2181

from the B73 genotype to CIMBL55 genotype failed to make any clear difference in *ZmRtn16* expression, while deletion of the 28 bp from the B73 sequence dramatically increased transcript abundance to a similar level to the CIMBL55 sequence (Fig. 5g). These data indicated that the 28 bp insertion in the *ZmRtn16^{B73}* 3'-UTR, but not the SNP, was the causal variant affecting *ZmRtn16* expression. We explored the *ZmRtn16* mRNA stability of the two constructs in maize leaf protoplasts. The mRNA decay of *ZmRtn16^{CIMBL55}* was found to be slower than *ZmRtn16^{B73}* in the presence of Triptolide, an inhibitor of transcription initiation, indicating that the 3'-UTR of *ZmRtn16^{CIMBL55}* might be more favorable for the mRNA stability than that of *ZmRtn16^{B73}* (Supplementary Fig. 10a and Supplementary Data). Interestingly, an RNA motif (G/U/C/GTGA) potentially recognized by two predicted maize RNA-binding proteins was annotated (<http://cisbp-rna.ccbr.utoronto.ca>) within the 28-bp insertion of the 3'-UTR of *ZmRtn16^{B73}*, which was likely the cause for the observed difference (Supplementary Fig. 10b).

We then generated *ZmRtn16* overexpression lines using a *ZmUbi1* constitutive promoter (Supplementary Fig. 11a and Supplementary Data). Greater drought resistance was observed in *ZmRtn16*-overexpressing plants, relative to WT plants (Fig. 5h). In contrast, the interruption of *ZmRtn16* function in two CRISPR-targeted gene knock-out lines resulted in compromised drought stress resistance (Fig. 5i and Supplementary Fig. 11b). Notably, the leaf-water-loss rate was consistently, substantially slower in *ZmRtn16*-overexpressing plants, while it was quicker in the KO lines, relative to WT plants (Fig. 5h–i). The stomatal apertures in the KO plants were more open than in the WT and the opposite phenotype was observed in the OE plants, suggesting that the enhanced stress resistance in the *ZmRtn16*-overexpressing plants was due to the size of stomatal aperture (Supplementary Fig. 11c and Supplementary Data). We compared the growth and development of *ZmRtn16-OE*, *zmrt16-crispr* and WT plants under well-watered and drought conditions in fields within a rain-out shelter. No significant differences were observed between the three groups of plants under well-watered conditions, except slightly shorter stature of the *zmrt16-crispr* plants. Under drought conditions, however, reduced plant height and prolonged anthesis and silking interval (ASI, which is an important symptom of drought stress on maize development) were exhibited in the *zmrt16-crispr* plants, relative to WT plants. In contrast, *ZmRtn16-OE* plants exhibited the reversed phenotypes (Supplementary Fig. 11d,e and Supplementary Data). The grain yields of *ZmRtn16-OE* plants were uncovered to be substantially higher than WT under drought conditions, but they were comparable under conditions of adequate water availability (Fig. 5j). Collectively, the data indicated that *ZmRtn16* played positive roles in plant drought resistance.

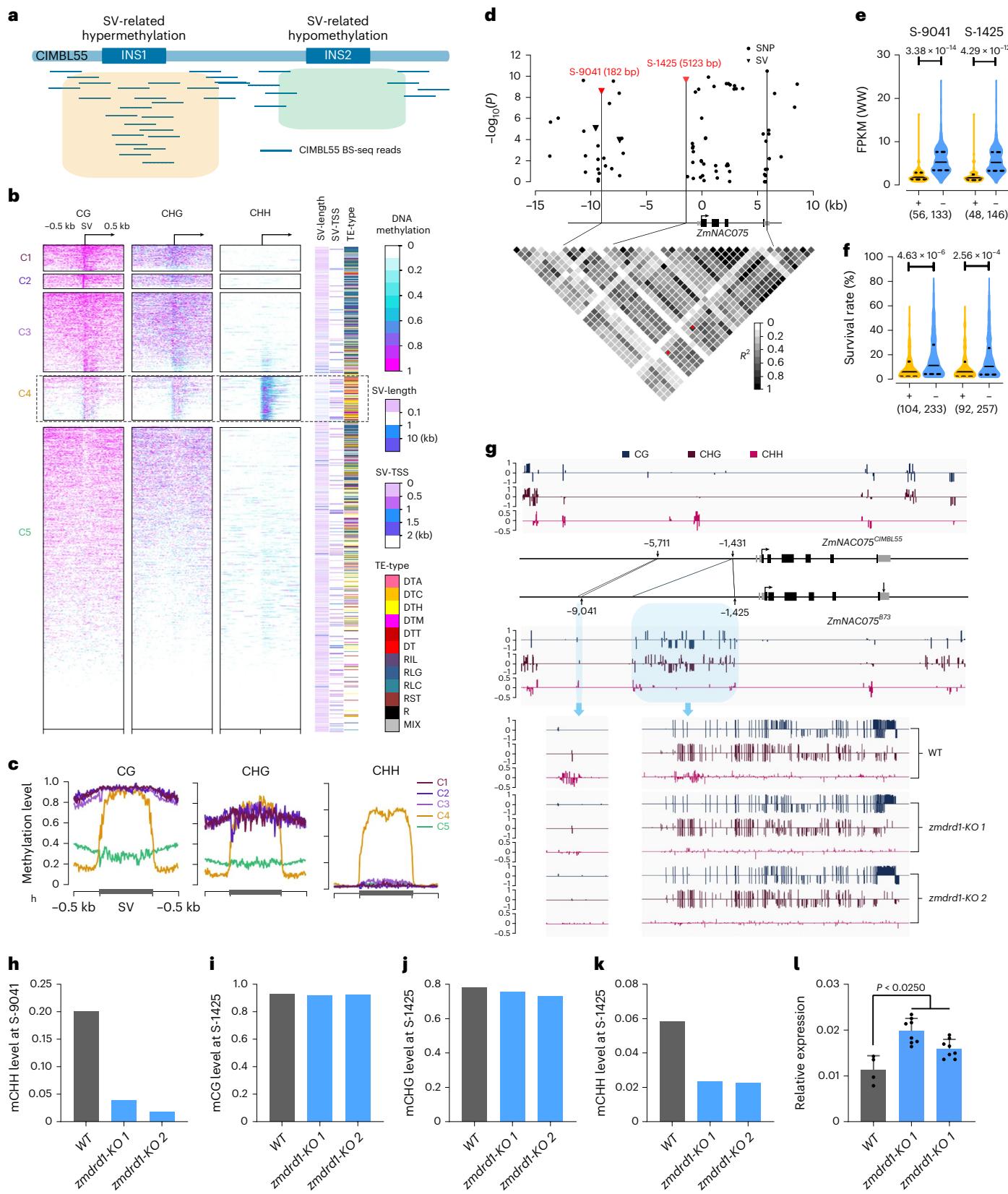
To understand the biological function of *ZmRtn16*, subcellular localization analysis of the *ZmRtn16-GFP* protein revealed that it colocalized with ER fractions (Supplementary Fig. 11f). Then, co-immunoprecipitation (co-IP) purification followed by mass spectrometry analyses identified that the vacuole H⁺-ATPase subunit A

Fig. 4 | SV-related alterations in DNA methylation. **a**, Diagram illustrating SV-related hyper- or hypo-DNA methylation. The insertional sequences, identified in CIMBL55 referenced to B73 and Mo17. The small lines represent the bisulfite sequencing (BS-seq) reads mapped to CIMBL55. The orange box denotes the DNA hypermethylation, and the blue-green box denotes the DNA hypo-methylation in the two insertional regions in CIMBL55. **b**, Heatmap of the level of mCG, mCHG and mCHH within the insertion start sites and their 0.5-kb-flanking regions. SV-length, the length of the insertion in CIMBL55. SV-TSS, distance of the insertional site to the transcriptional start site of the closest gene. TE-type, insertion sequences are characterized as TE. The abbreviations for the types of TE are listed in Supplementary Table 2. MIX, multiple types of DNA-TE and RNA-TE. **c**, Profiles of DNA methylation level in the 0.5-kb-flanking region of the insertional fragments of five clusters. **d**, *ZmNAC075*-based association mapping of gene expression under adequate water conditions. The association analysis is based on a mix linear model²⁸, in this and all the following figures. The position

of the lead SNP (8_21515682) and the two insertions are indicated by lines, and their LD levels are indicated by red asterisks in the pairwise LD heatmap of all the variants ($P < 10^{-3}$). **e,f**, Violin plots of *ZmNAC075* expression levels (e) and survival rate (f) of the germplasms with different alleles at the S-9014 and S-1425 sites. The numbers in the parentheses denote the sample size used in the two-sided t-test. **g**, Upper panel, sequence comparison and DNA methylation status of the *ZmNAC075* locus in B73 and CIMBL55. Hypermethylation of two insertions, S-9014 (182 bp) and S-1425 (5123 bp), upstream of *ZmNAC075^{B73}* indicated by the light-blue boxes. Lower panel, The DNA methylation status of the two insertional fragments in WT and two *zmdrd1-KO* lines. **h–k**, DNA methylation levels at S-9041 and S-1425 between WT and two *zmdrd1-KO* lines. Mean value of two biological replicates was shown. **l**, Relative expression of *ZmNAC075^{B73}* in WT and two *zmdrd1-KO* lines. The statistical significance is determined by a two-sided t-test from four biological replicates.

(ZmVHA-A) and E3 (ZmVHA-E3) were the candidate proteins interacting with ZmRtn16. The split-luciferase complementation and co-IP assays confirmed their interactions with ZmRtn16-GFP (Fig. 6a,b). Subsequently, we compared the subcellular localization of ZmVHA-A and ZmVHA-E3 in WT and *zmrtn16-crispr* plants. A clear defect in the

tonoplast localization of ZmVHA-A and ZmVHA-E3 was observed in *zmrtn16-crispr* plants, suggesting that ZmRtn16 facilitates the tonoplast localization of ZmVHA-A and ZmVHA-E3 (Fig. 6c). Furthermore, the V-H⁺-ATPase activity was determined to be greatly reduced in *zmrtn16-crispr* plants and enhanced in plants of *ZmRtn16-OE* lines,



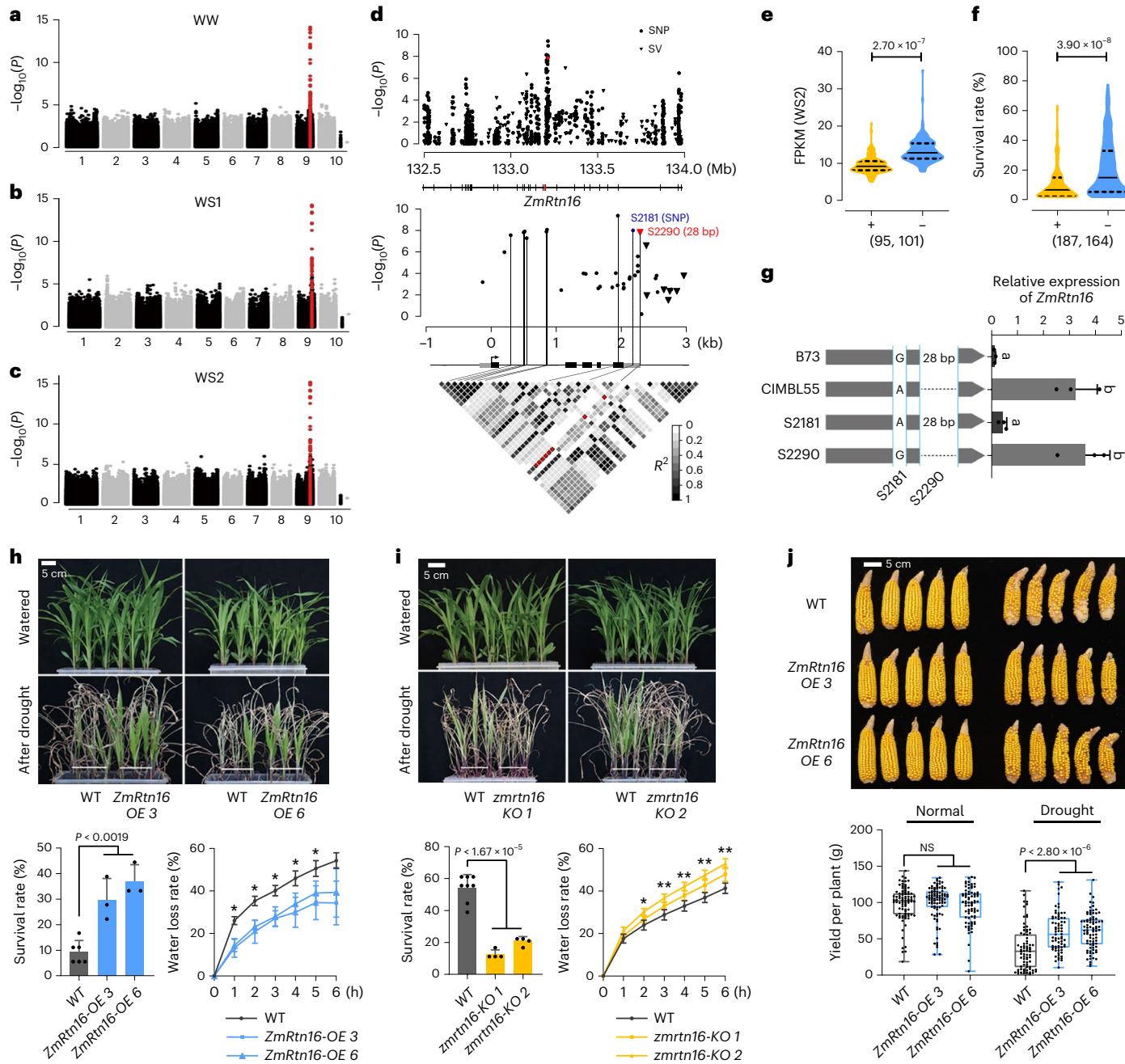


Fig. 5 | *ZmRtn16* positively contributes to drought resistance in maize.

a-c. Manhattan plot of the GWAS identifying the genetic loci associated with *ZmRtn16* expression levels under well-watered (WW), moderate drought (WS1) and severe drought (WS2) conditions. SNPs located within the *ZmRtn16* and its flanking 10-kb region are indicated by red dots. **d.** Upper panel, *ZmRtn16*-based association mapping for survival rate after exposure to drought treatment. The 132.5–134.0 Mb region of Chr9 is shown. Lower panel, Zoom view of the *ZmRtn16* gene region and pairwise LD heatmap for the variants. Significant variants ($P < 10^{-7}$) are connected to the pairwise LD heatmap by lines. Red asterisks indicate a strong LD between the variants. **e, f.** Violin plots of the expression level of *ZmRtn16* under WS2 condition and the survival rate of germplasms carrying the different alleles of *ZmRtn16* at S2290, with (+) or without (-) the 28-bp insertion. Yellow and blue colors represented the germplasms carrying the alleles of *ZmRtn16*^{B73} and *ZmRtn16*^{CIMBL55}, respectively. The numbers in the parentheses denote the sample size in a two-sided *t*-test. **g.** Left panel, diagram indicating the genotypes of the downstream sequence of *ZmRtn16*^{B73} and *ZmRtn16*^{CIMBL55}. Right panel, mRNA abundance of *ZmRtn16* when it was fused with different 3'-UTR sequences. Different letters indicate a significant difference ($P < 0.05$) based on the one-way ANOVA (Tukey's test). **h, i.** Drought resistance assays and the rate of leaf water loss. Eighteen plants of each genotype were tested in each experiment, and at least three replicates were conducted. Four detached leaves of each genotype were dehydrated on a clean bench and weights were measured over the course of 6 h. Data represent the mean \pm s.d., based on three replicated experiments. Statistical significance was determined by a two-sided *t*-test: * $P < 0.05$, ** $P < 0.01$. **j.** Grain yield of *ZmRtn16*-OE and WT plants under well-watered and drought conditions in field. Representative photographs were taken of the ears of WT and two *ZmRtn16*-OE plants. The data were obtained from at least 72 plants for each genotype. Statistical significance was determined by a two-sided *t*-test in **h-j**. Scale bar, 5 cm.

the genotypes of the downstream sequence of *ZmRtn16*^{B73} and *ZmRtn16*^{CIMBL55}. Right panel, mRNA abundance of *ZmRtn16* when it was fused with different 3'-UTR sequences. Different letters indicate a significant difference ($P < 0.05$) based on the one-way ANOVA (Tukey's test). **h, i.** Drought resistance assays and the rate of leaf water loss. Eighteen plants of each genotype were tested in each experiment, and at least three replicates were conducted. Four detached leaves of each genotype were dehydrated on a clean bench and weights were measured over the course of 6 h. Data represent the mean \pm s.d., based on three replicated experiments. Statistical significance was determined by a two-sided *t*-test: * $P < 0.05$, ** $P < 0.01$. **j.** Grain yield of *ZmRtn16*-OE and WT plants under well-watered and drought conditions in field. Representative photographs were taken of the ears of WT and two *ZmRtn16*-OE plants. The data were obtained from at least 72 plants for each genotype. Statistical significance was determined by a two-sided *t*-test in **h-j**. Scale bar, 5 cm.

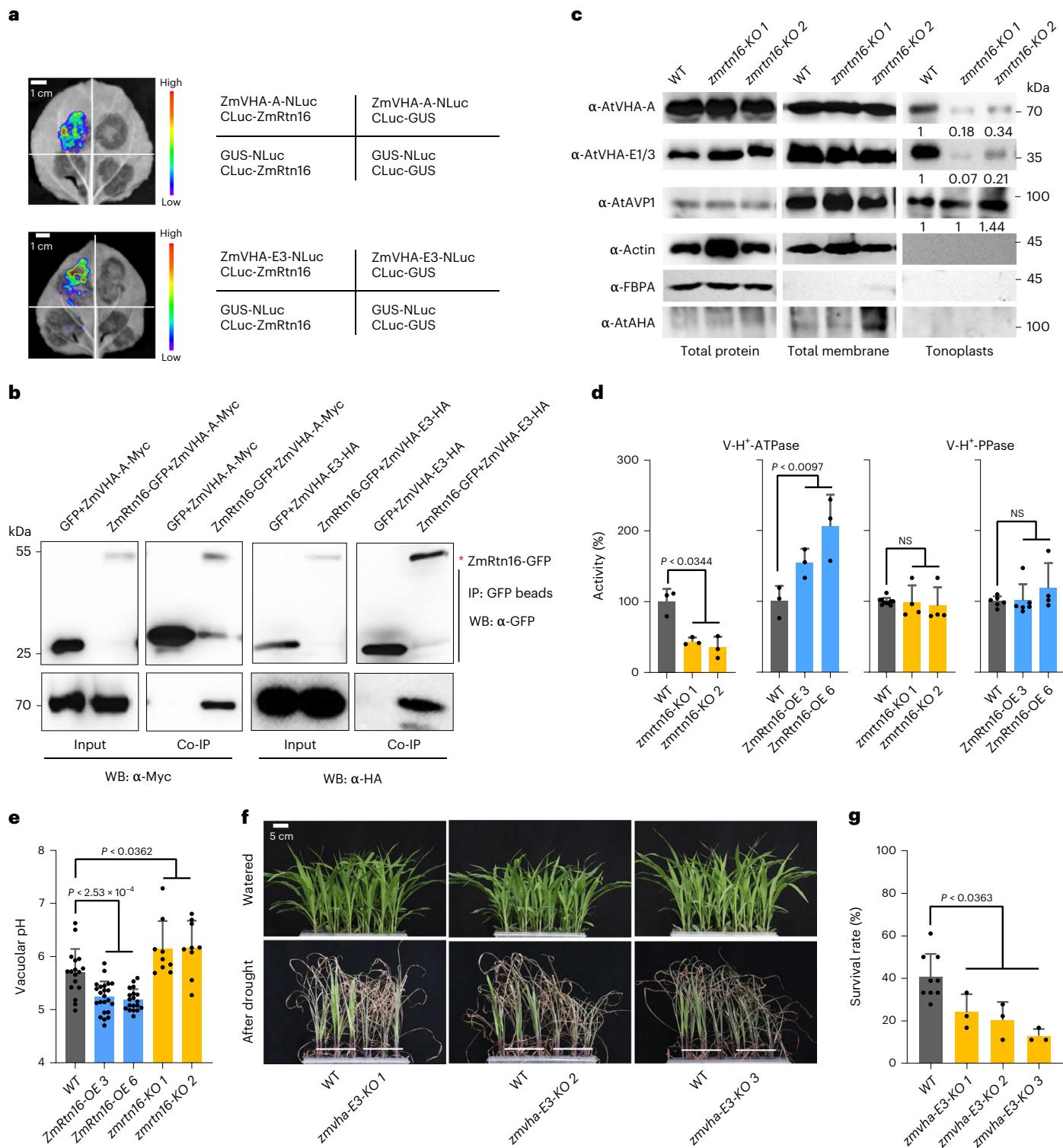


Fig. 6 | ZmRtn16 interacts with ZmVHA-A and E3 and facilitates ZmVHA functions. **a**, Luciferase complementation assay detecting the interaction of ZmRtn16 with ZmVHA-A and ZmVHA-E3 in tobacco leaf cells. Scale bar, 1 cm. **b**, Co-immunoprecipitation assay of ZmRtn16, ZmVHA-A and ZmVHA-E3 using maize leaf protoplasts. **c**, Immunoblot analysis of ZmVHA-A and ZmVHA-E1/E3 abundance present on tonoplasts in WT and KO plants. The immunoblots of anti-Actin, anti-AVP1 (*Arabidopsis* Vacuole H⁺-pyrophosphatase, V-H⁺-PPase), anti-FBPA (fructose-bisphosphate aldolase, a cytoplasmic protein) and anti-AHA (*Arabidopsis* plasma membrane H⁺-ATPase) indicate equal amounts of total protein from leaf tissue that were analyzed and successfully fractionated into the total membrane, soluble cytoplasmic and tonoplast fractions. The full scans of all the blots are included in Source Data. The immunoblots shown in **b** and **c** are the representative

results from at least twice independent replicates. **d**, V-H⁺-ATPase and V-H⁺-PPase activity assay. A total of 20 g of seedling leaves were used for each genotype in the tonoplast preparation and the V-H⁺-ATPase and V-H⁺-PPase activity assays. Three replicated measurements were performed. **e**, Root vacuolar pH quantification. The statistical data were obtained from at least nine measurements from six seedlings of each plant type. **f**, Comparison of the plant drought resistance. Photographs were taken under normal growing conditions and after drought treatment when the difference was most apparent. Eighteen plants of each genotype were tested in each experiment, and at least three replicated experiments were conducted. Scale bar, 5 cm. **g**, Statistical data of the survival rate of zmva-E3-KO lines in relation to WT plants. Data represent the mean ± s.d. Statistical significance was determined by a two-sided *t*-test (in **d**, **e** and **g**).

relative to WT plants. Notably, neither the tonoplast localization nor protein activity of vacuole-type H⁺-pyrophosphatase (V-H⁺-PPase) was substantially altered in the mutants, suggesting that ZmRtn16 was not required for the biological function of V-H⁺-PPase (Fig. 6c,d). In relative comparison to WT plants, the vacuolar proton levels were determined to be higher (with lower pH values) in *ZmRtn16-OE* plants, whereas they were lower in *ZmRtn16-KO* plants (Fig. 6e and Supplementary Fig. 11g,h). Then, we obtained the CRISPR-targeted *zmwha-E3-KO* plants (Supplementary Fig. 12), probably owing to the potential functional redundancy with *ZmVHA-E1*. Similar drought sensitivity phenotype was observed on the *zmwha-E3-KO* plants, providing direct evidence that defects in V-H⁺-ATPase due to the absence of the ZmVHA-E3 subunit on tonoplasts compromised plant drought resistance (Fig. 6f,g). Taken together, we found that a 28-bp deletion in the 3'-UTR of *ZmRtn16*^{CIMBL55} increased the expression of *ZmRtn16*, which had a positive effect on drought resistance by facilitating the localization of ZmVHA-A and ZmVHA-E3 to tonoplasts and thus, enhancing V-H⁺-ATPase activity.

Discussion

We constructed a high-quality, de novo genome assembly of a drought-resistant maize genotype CIMBL55. In total, 544,853 SVs were identified on a pan-genome level and genotyped on a maize association panel, which facilitated the SV identification linked with drought-resistant candidate genes and the analysis of SV-related DNA methylation. The newly identified SVs that were substantially associated with drought resistance may not be the actual causal variants; however, they can serve as potential targets for further validation and allelic mining.

There are two types of H⁺-pumps present on tonoplasts, V-H⁺-ATPase and V-H⁺-PPase. V-H⁺-ATPase is a protein complex consisting of 13 subunits that are part of the peripheral V1 (including A and E3 subunits) and membrane-integral V0 complexes. V-H⁺-ATPase hydrolyzes one ATP and transports two protons from the cytosol into vacuole⁴⁴. V-H⁺-PPase is a homodimer, and it utilizes the energy of inorganic pyrophosphate (PPi) to drive one proton translocation across tonoplast membranes⁴⁵. These two types of proton pumps are assumed to work synergistically using different energy resources, ensuring that plants are able to maintain the appropriate proton gradient for vacuole transport, even when cells are subject to environmental stress conditions⁴⁶. We previously reported that genetic variation in *ZmVPP1*, which encodes a V-H⁺-PPase, contributes to drought resistance in maize¹⁶. In the present study, we demonstrated that ZmRtn16 facilitates the tonoplast localization of the A and E3 subunits of V-H⁺-ATPase, which ensured V-H⁺-ATPase activity and had a positive role in drought stress resistance. Taken together, it highlighted the role of vacuole proton pumps in maize drought resistance. Vacuolar acidification was found to be necessary for rapid stomatal closure in response to ABA because an *Arabidopsis* double mutant of *vha-a2 vha-a3*, lacking two subunits of V-H⁺-ATPase V0 complex, showed delayed stomatal closure, and a similar phenotype was also observed in the V-H⁺-PPase mutant (*vhp1*)⁴⁷. Mutation in VHA-C, the C subunit of V-H⁺-ATPase V1 complex, rendered severe salt sensitivity in *Arabidopsis*⁴⁸. The OsVHA-A RNA-interference rice plants showed expanded stomatal aperture and compromised salt tolerance⁴⁹. Utilizing the superior allele of *ZmRtn16* through selection of the variation in the 3'-UTR or by targeted gene editing to enhance *ZmRtn16* gene expression represents a promising strategy for improving the trait of drought resistance. The high-quality assembly of CIMBL55, a prominent drought-resistant germplasm, provides an important resource for both gene function discovery and the genetic improvement of drought resistance.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-023-01297-y>.

References

- Gupta, A., Rico-Medina, A. & Cano-Delgado, A. I. The physiology of plant responses to drought. *Science* **368**, 266–269 (2020).
- Hu, H. & Xiong, L. Genetic engineering and breeding of drought-resistant crops. *Annu Rev Plant Biol.* **65**, 715–741 (2014).
- Ashraf, M. Inducing drought tolerance in plants: recent advances. *Biotechnol. Adv.* **28**, 169–183 (2010).
- Godfray, H. C. et al. Food security: the challenge of feeding 9 billion people. *Science* **327**, 812–818 (2010).
- Daryanto, S., Wang, L. & Jacinthe, P. A. Global synthesis of drought effects on maize and wheat production. *PLoS ONE* **11**, e0156362 (2016).
- Smith, C. W., Betran J., Runge E. C. A. (eds.). *Corn: Origin, Technology, and Production*, 99–132 (John Wiley & Sons, 2004).
- Schnable, P. S. et al. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115 (2009).
- Mahmoud, M. et al. Structural variant calling: the long and the short of it. *Genome Biol.* **20**, 246 (2019).
- Jiao, Y. et al. Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017).
- Sun, S. et al. Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet.* **50**, 1289–1295 (2018).
- Hufford, M. B. et al. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).
- Yang, N. et al. Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat. Genet.* **51**, 1052–1059 (2019).
- Lin, G. et al. Chromosome-level genome assembly of a regenerable maize inbred line A188. *Genome Biol.* **22**, 175 (2021).
- Li, C. et al. Long-read sequencing reveals genomic structural variations that underlie creation of quality protein maize. *Nat. Commun.* **11**, 17 (2020).
- Haberer, G. et al. European maize genomes highlight intraspecies variation in repeat and gene content. *Nat. Genet.* **52**, 950–957 (2020).
- Wang, X. et al. Genetic variation in *ZmVPP1* contributes to drought tolerance in maize seedlings. *Nat. Genet.* **48**, 1233–1241 (2016).
- Mao, H. et al. A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nat. Commun.* **6**, 8326 (2015).
- Liu, S. et al. Mapping regulatory variants controlling gene expression in drought response and tolerance in maize. *Genome Biol.* **21**, 163 (2020).
- Yang, X. et al. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* **28**, 511–526 (2010).
- Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019).
- Tang, H. et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
- Liu, D., Hunt, M. & Tsai, I. J. Inferring synteny between genome assemblies: a systematic evaluation. *BMC Bioinform.* **19**, 26 (2018).
- Woodhouse, M. R. et al. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS Biol.* **8**, e1000409 (2010).
- Tian, T. et al. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **45**, W122–W129 (2017).

25. He, Z. et al. The maize ABA receptors ZmPYL8, 9 and 12 facilitate plant drought resistance. *Front Plant Sci.* **9**, 422 (2018).
26. Huai, J. et al. Cloning and characterization of the SnRK2 gene family from Zea mays. *Plant Cell Rep.* **27**, 1861–1868 (2008).
27. Wei, K. & Pan, S. Maize protein phosphatase gene family: identification and molecular characterization. *BMC Genom.* **15**, 773 (2014).
28. Liu, S. et al. Genome-wide analysis of ZmDREB genes and their association with natural variation in drought tolerance at seedling stage of Zea mays L. *PLoS Genet.* **9**, e1003790 (2013).
29. Jin, J. et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **45**, D1040–D1045 (2017).
30. Goel, M., Sun, H., Jiao, W. B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277 (2019).
31. Kronenberg, Z. N. et al. High-resolution comparative analysis of great ape genomes. *Science* **360**, eaar6343 (2018).
32. Jiang, T. et al. Long-read-based human genomic structural variation detection with cuteSV. *Genome Biol.* **21**, 189 (2020).
33. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).
34. Cao, Y. et al. Natural variation of an EF-hand Ca⁽²⁺⁾-binding-protein coding gene confers saline-alkaline tolerance in maize. *Nat. Commun.* **11**, 186 (2020).
35. Kang, J. Y., Choi, H. I., Im, M. Y. & Kim, S. Y. Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. *Plant Cell* **14**, 343–357 (2002).
36. Fujita, Y. et al. AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. *Plant Cell* **17**, 3470–3488 (2005).
37. Hu, H. et al. Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proc. Natl Acad. Sci. USA* **103**, 12987–12992 (2006).
38. Fujita, M. et al. A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *Plant J.* **39**, 863–876 (2004).
39. Tran, L. S. et al. Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *Plant Cell* **16**, 2481–2498 (2004).
40. Kanno, T. et al. Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr. Biol.* **14**, 801–805 (2004).
41. Huettel, B. et al. Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *EMBO J.* **25**, 2828–2836 (2006).
42. Zhang, X. et al. Reticulon proteins modulate autophagy of the endoplasmic reticulum in maize endosperm. *eLife* **9**, e51918 (2020).
43. Yang, Y. S. & Strittmatter, S. M. The reticulons: a family of proteins with diverse functions. *Genome Biol.* **8**, 234 (2007).
44. Cipriano, D. J. et al. Structure and regulation of the vacuolar ATPases. *Biochim. Biophys. Acta* **1777**, 599–604 (2008).
45. Gaxiola, R. A., Palmgren, M. G. & Schumacher, K. Plant proton pumps. *FEBS Lett.* **581**, 2204–2214 (2007).
46. Li, Y., Zeng, H., Xu, F., Yan, F. & Xu, W. H(+)-ATPases in plant growth and stress responses. *Annu. Rev. Plant Biol.* **73**, 495–521 (2022).
47. Bak, G. et al. Rapid structural changes and acidification of guard cell vacuoles during stomatal closure require phosphatidylinositol 3,5-bisphosphate. *Plant Cell* **25**, 2202–2216 (2013).
48. Batelli, G. et al. SOS2 promotes salt tolerance in part by interacting with the vacuolar H+-ATPase and upregulating its transport activity. *Mol. Cell. Biol.* **27**, 7781–7790 (2007).
49. Zhang, H. et al. RNAi-directed downregulation of vacuolar H(+) -ATPase subunit a results in enhanced stomatal aperture and density in rice. *PLoS ONE* **8**, e69046 (2013).
50. Zhang, X. et al. Genetic variation in ZmTIP1 contributes to root hair elongation and drought tolerance in maize. *Plant Biotechnol. J.* **18**, 1271–1283 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

Methods

CIMBL55 sequencing and assembly

The maize germplasm CIMBL55 was originated from the International Maize and Wheat Improvement Center (CIMMYT). The genomic DNAs of CIMBL55 were prepared from the 2-week-old plants grown in greenhouse. The sequencing library for Sequel 2 platform (Pacific Biosciences), HiSeq X Ten (Illumina) and Saphyr Instruments (BioNano Genomics) was constructed, based on the manufacturer protocols by Biomarker Technologies and BerryGenomics. For Hi-C data generation, the genomic DNA was fixed with formaldehyde, fragmented with Hind 3 as restriction enzyme, blunt-end ligation and purified to obtain qualified DNA products for the library construction. Further, 74.78 Gb clean data for 249,682,204 pairs of reads were obtained by Biomarker Technologies.

For the genome assemble, the PacBio raw data of 60×RSII and 100×Sequel 2 were corrected and assembled by NextDenovo v2.3.1 (<https://github.com/Nextomics/NextDenovo>) and set the correct_option with (minimap2_options_raw = -x ava-pb -t 30) and assemble_option with (minimap2_options_cns = -x ava-pb -t 30 -k17 -w17). Combing 60× coverage Illumina sequencing, NextPolish v1.3.0 was used to polish the contigs. Bionano Solve v3.6 was used to construct the optical maps and assemble the contigs to produce scaffolds with default parameters. Finally, Hi-C data helped mounting scaffolds to chromosomes by 3D-DNA pipeline⁵¹ with default parameters and manually corrected based on the Hi-C interaction map at juicebox website (<https://aidenlab.org/juicebox/>).

Genome annotation

Repeat regions in the CIMBL55 genome were identified by de novo predictions and homology-based annotation through RepeatModeler v2.0.1 (ref. ⁵²), RepeatMasker v4.1.1 (ref. ⁵³) and HelitronScanner⁵⁴. Gene annotation was combined with ab initio training and evidence-based predictions. For the homolog-evidence-based prediction, protein sequence of B73, Mo17, SK, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Oryza sativa*, *Setaria italica* and *Sorghum bicolor* was used to align to CIMBL55 genome. Transcript evidence includes 69,163 publicly available full-length cDNAs from B73 deposited in GenBank, 302,909 Trinity (v2.11.0)-assembled transcripts from 58 B73 RNA-seq experiments (SRP010680) and 36 B73 iso-seq (SRP067440), 644,700 Trinity-assembled transcripts from 21 CIMBL55 RNA-seq experiment including four inhouse data and 17 public samples (SRP132192). The detailed information is listed in Supplementary Table 11. These transcripts were further validated by PASA v2.4.1 (ref. ⁵⁵). For de novo prediction, we use Augustus v3.2.3 and FGENESH v8.0.0a to de novo training on the repeat masked CIMBL55 genome. All the evidence was submitted to MAKER v3.01.03 (ref. ⁵⁶) result in 38,439 genes and 59,509 transcripts. For function annotation of the protein-coding genes, protein domain was achieved by using PfamScan⁵⁷ and GO terms annotation was calculated by Blast2GO⁵⁸ with default parameters.

Synteny analysis

The longest proteins of each gene of four maize inbred lines (CIMBL55, B73, Mo17, SK) were put together and aligned with blast toolkit (-p blastp -e 1e-20). A synteny block was defined based on MCscanX⁵⁹ with the default parameter (MCscanX_h-b2-c1-m25). It is interpreted as the genomic regions (of the two compared genomes) contain at least five orthologous gene pairs, with no more than 25 interrupting genes between any neighboring pairs. We tested different gap parameters (-m5/10/15/20/25/30) and obtained consistent proportion (from 69.15% to 73.01%) of syntenic genes. We chose 25 interrupting genes as the parameter because, under this condition, the obtained ten biggest synteny blocks coincided with the ten chromosomes. For the Class 5 genes, we aligned their DNA sequences to the B73 genome. With the parameters of identity >90 and e value <1×10⁻²⁰ and score > 200, the successfully aligned genes were further grouped into homologous

chromosomes and nonhomologous chromosomes groups, based on their chromosomal location in the two genomes. The subgenome and fractionation bias analyses of CIMBL55 were performed by SynMap at CoGe website (<https://genomevolution.org/CoGe/SynMap.pl>) with default parameters.

Genetic variation identification and genotyping in an association panel

As for the DNA variants identification for the 30 maize genomes, the genome sequence of B73_v5 (ref. ¹¹) was used as a reference. SyR³⁰ was used to compare the two assembled genomes as previously described in ref. ¹³ based on the MUMmer⁶⁰ alignment output (nuclmer--maxmatch-c 500-b 500-l 1200-g 1000; delta-filter -m -190-l 1000). Smartie-sv³¹ identified variants by aligning the contig sequence of the 30 maize germplasms to B37_v5 based on blasr with default parameters. CuteSV³² and Sniffles³³ were used for the prediction of variants after the alignment of the long reads of the 30 maize germplasms to B37_v5 through its supporting toolkit ngmlr. The variants were merged by BEDtools⁶¹, parameter settings intersect -r -f 0.8 -v for insertions and window -l100 -r100 -v for deletions. As for the identification of the variants between B73/Mo17 and CIMBL55, CIMBL55 was used as the reference genome. Four kinds of software were employed. MUMmer v4.0 (ref. ⁶⁰) was used to compare the two assembled genomes as previously described in ref. ¹⁰. The show-snps (-ChrT -x 1-T) option was used to obtain the initial variations, and the mummerplot (-x -y) option was used to view large variation steps by a 10-Mb window along each pair of corresponding chromosomes. Smartie-sv³¹ identified variants through aligning the B73/Mo17 contigs to CIMBL55, based on the blasr with default parameters. Pbsv (<https://github.com/PacificBiosciences/pbsv>) and Sniffles³³ were used for the prediction of variants after alignment of the PacBio reads to CIMBL55 through its supporting toolkits pbmm2 and ngmlr, respectively. The variants were classified into SNPs, InDels (≤ 20 bp) and variants (>20 bp). The identified SVs ≥ 50 bp obtained from the 30 maize genomes and DNA variants ≥ 20 bp were genotyped by Paragraph⁶² based on the 368 DNA deep re-sequencing data¹². Finally, 544,853 variants were genotyped with a minor allele frequency > 0.05 and a missing rate < 0.6 . As for bidirectional SV identification, at first, the DNA variants were identified through four programs, MUMmer, Smartie-sv, Pbsv and Sniffles. The two output files of a program (for example, MUMmer) were obtained based on the bidirectional analysis of either CIMBL55 as reference and B73/Mo17 as query or CIMBL55 as query and B73/Mo17 as reference. Then, they were processed by a customized perl script (https://github.com/tian627/CIMBL55_genome_assembly) to filter out the SNPs present only in one file. Secondly, the identified bidirectional SNPs were used as anchors. If an insertion was identified in CIMBL55 referred to B73, a deletion should be called in B73 in a comparison with CIMBL55, with the difference of their distances to the anchor SNP being less than 500 bp. Finally, 5,346 bidirectionally confirmed insertional fragments (≥ 20 bp) identified in CIMBL55 in comparison with B73 (also Mo17) were used for SV-related differential DNA methylation analysis.

Whole genome bisulfite sequencing and differential methylation analysis

The DNA libraries were constructed according to the manufacturer's instructions. Three replicates of each genome were sequenced, and >30 -fold coverage of reads was obtained for each replicate. In brief, the ligated DNA was bisulfite converted using the EZ DNA Methylation-Gold kit (Zymo Research) which was then purified and amplified by PCR. Sequencing was performed on HiSeq X Ten conducted by Beijing Genomics institution. After removing adapter sequences, contamination and low-quality reads from the raw data, the clean reads of CIMBL55, B73 and Mo17 were then mapped to the CIMBL55 reference genome by BSMAP v2.9 (ref. ⁶³) and the methylation ratio was calculated. Each methylation ratio file was split into CG, CHG and CHH

containing files. To calculate the DNA methylation level, the sites with more than 4-read-coverage in all three replicates were screened and converted to bw format files. The methylation levels of the insertions in CIMBL55, compared with B73/Mo17, and their flanking regions were clustered by DeepTools 3.4.3 (ref. ⁶⁴). *Zmdrd1-related* bisulfite sequencing was performed on DNBSEQ with two repeats for each variant line and WT conducted by Beijing Genomics institution.

sRNA analysis

The raw data were downloaded from NCBI (Supplementary Table 11), and analysis was performed according to ref. ⁶⁵ with some modifications. After removing adapters with Trimmomatic v0.35 (ref. ⁶⁶), the trimmed reads aligned to B73_v4 genome with Bowtie⁶⁷ (-v 0), and the output SAM file was converted to bed format by BEDTools v2.29.2 (ref. ⁶¹). After removing the tRNA, rRNA, snoRNA and miRNA, the length of 24-nt alignments was obtained and converted to bw format as input of DeepTools 3.4.3 (ref. ⁶⁴).

Candidate gene association analysis

The normalized gene expression values for well water, moderate drought (WS1) and severe drought (WS2) in seedling leaves and survival rate after drought were obtained from the previous work^{16,18}. We computed GWAS analysis using a linear mixed model (-mlm -mlmVarCompEst EachMarker -mlmCompressionLevel None) or general linear model (-glm) by TASSEL (version 5.0).

Transgenic plants generation

The maize transgenic plants were from the Center for Crop Functional Genomics and Molecular Breeding, CAU. The construct was transformed into the inbred line LH244 (PI 612589), which is available at GRIN-Global (<https://npgsweb.ars-grin.gov/gringlobal/search>). The expression of *ZmABF4* and *ZmRtn16* in the transgenic plants was determined by reverse-transcription quantitative-PCR (RT-qPCR), in which *ZmUbi1* (ref. ¹⁷) was used as an internal control. For CRISPR-targeted knock-out plants of *zmdrd1*, *zmrttn16*, *zmwha-E3*, the targeted editing was identified by genomic DNA PCR and sequence analysis. The edited T0 plants were self-pollinated for T1 seeds. The gene knock-out and cas9-free plants were identified in T1 generation which was propagated for T2 seeds. They were used for further research.

RNA extraction and RT-qPCR

Total RNAs were extracted from 14-day-old seedlings grown in soil or maize protoplasts with TRIzol (Invitrogen), followed by reverse transcription (2.0 µg of RNAs) with M-MLV reverse transcriptase (Promega). Then, RT-qPCRs were performed with SYBR Green reagent (Takara) on a 7,500 real-time PCR system to detect specific genes. The relative expression of the gene was calculated by the $2^{-\Delta\Delta Ct}$ method.

Protoplasts transfection and gene transient expression analysis

The 646 bp and 1016 bp downstream sequences of *ZmRtn16* were cloned from B73 and CIMBL55, respectively, and they were ligated after the *ZmRtn16^{B73}* coding sequence (driven by the *Zmubi1* promoter) in the pGreenII vector. Protoplast transfection assays were carried out as described in ref. ⁶⁸. Briefly, 100 µl of a protoplast suspension (10⁶/ml) was transfected with 5 µg of DNA and incubated at 22 °C for 12 h. The cells were collected for the expression quantification of *ZmRtn16*.

Affinity purification and co-immunoprecipitation assays

The *ZmRtn16* CDS was inserted into pGreenII-GFP vector using *Sma I* sites⁶⁹. Ubi:ZmRtn16-GFP and GFP vectors were transformed into maize protoplasts and incubated at 22 °C for 12 h. Affinity purification was performed according to ref. ⁶⁹. Candidate proteins specifically bound to ZmRtn16-GFP were identified by anti-GFP (Abclonal,

AE012) immunoprecipitation in two biological repeats with the subtraction of those present in GFP vector transfected samples. The full length of *ZmRtn16* CDS was cloned into pGreenII-Ubi-GFP vector using *Sma I*. The CDS of *ZmVHA-E3-HA* and *ZmVHA-A-Myc* with a stop codon was cloned into the same vector using *Sma I* and *Not I* sites. The co-immunoprecipitation assays were performed as described by ref. ⁵⁰. Briefly, 1.2 ml of protoplasts transformed with different combinations of plasmids were used for total protein extraction and immunoprecipitation with anti-GFP Magnetic Beads (MBL, D153-11). The immunoprecipitates were detected with anti-GFP (Abclonal, AE012), anti-HA antibody (Abclonal, AE008) and anti-Myc antibody (Sigma-Aldrich, M4439).

Subcellular localization

To detect the subcellular localization of ZmRtn16, Ubi:ZmRtn16-GFP and HDEL-mCherry were transformed into maize protoplasts incubated at 22 °C for 12 h. HDEL-mCherry was the ER marker. The fluorescence was observed by confocal microscopy (ZEISS880, Carl Zeiss) with exciting light of 488 nm for GFP and 561 nm for mCherry.

Leaf water loss measurement and stomatal aperture observation

For the measurement of leaf water loss, for each sample, four fully expanded leaves were detached from 21-day-old plants of each genotype. All the samples were placed on a clean bench at room temperature. The fresh weights of the samples were measured at the indicated time points. Three repeated experiments were performed to obtain the statistical data. For stomatal aperture analyses, seedlings were cultured in soil in a growth chamber for 6 d and then transplanted to soil in a pot under natural growth conditions. TM3000 was used to photograph the lower epidermis of the second fully expanded maize leaf before or after 5 h dehydration treatment. Stomata were divided into three states: fully open, partially closed and fully closed. The percentage of each type was calculated.

Firefly luciferase complementation imaging

The full-length CDS of *ZmRtn16*, *ZmVHA-E3* and *ZmVHA-A* was cloned into pCAMBIA1300-NLuc and pCAMBIA1300-CLuc⁷⁰ to produce recombinant of the *ZmVHA-A-NLuc*, *ZmVHA-E3-NLuc*, CLuc-ZmRtn16 vectors, respectively. Fully expanded *Nicotiana benthamiana* leaves (8–10 weeks old) were injected with *Agrobacterium tumefaciens* GV3101 that had been transfected with the above constructs. At four different areas on a leaf, one sample and three negative control combinations were injected. After 54 h of growth under a 14 h light/10 h dark cycle, 1 mM luciferin (Promega) was sprayed onto the inoculated leaves. Sprayed leaves were left in dark for 5 min. The luciferase signals were captured using the LUCK2019 (LB985) imaging system. The experiments were repeated three times, and consistent results were obtained.

Vacuolar membrane protein preparation and western blot analysis

Vacuolar membrane protein was extracted from maize as described in ref. ¹⁶. Briefly, 20 g leaf samples were collected from 3-week-old plants grown under normal conditions and quickly frozen in liquid nitrogen. Cold extraction buffer (80 ml) (50 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid/tris(hydroxymethyl) aminomethane (HEPES/Tris) (pH 7.4), 0.25 M sucrose, 10% glycerinum, 2 mM ethyleneglycol-bis(β-aminoethyl ether)-N,N,N',N'-ethy-tetraacetic acid (EGTA), 0.5% bovine serum albumin (wt/vol), 5 mM dithiothreitol (DTT), 2 mM MgSO₄, 1% polyvinylpolypyrrolidone (wt/vol) and 1 mM phenylmethylsulphonyl fluoride (PMSF)) was added. After fully mixed, the samples were centrifuged at 4,200g for 10 min at 4 °C. The supernatant was filtered with four layers of Miracloth (Calbiochem, 475855-IR) and then centrifuged at 100,000g for 1 h at 4 °C, and the resultant pellet was the crude membranes protein. The pellet in the storage buffer containing

5 mM HEPES/Tris (pH 7.4), 2 mM MgSO₄, 1 mM PMSF and 2 mM DTT was resolved. Then, the crude membranes protein was slowly added to the upper layer of a 10%/30% discontinuous sucrose density gradient, and centrifuged at 100,000g for 2 h at 4 °C. Vacuolar membrane proteins were concentrated at the 10%/30% sucrose gradient boundary. Then, the fraction was resuspended and cleaned with ten volumes of cleaning buffer (5 mM HEPES/Tris (pH 7.4), 2 mM MgSO₄, 1 mM PMSF, 2 mM DTT). Centrifugation was performed at 146,000g at 4 °C for 1 h. Discarding the supernatant, the precipitate was suspended again with the storage buffer, namely, vacuolar membrane protein, which was stored at -80 °C for further experiments. The samples were analyzed by western blotting using the anti-AtVHA-A (PHYTOAB, PHY1480S) and anti-AtVHA-E1/3 (PHYTOAB, PHY1035S), the protein bands were quantified by ImageJ. The western blot of anti-OsFBPA (Beijing Protein Innovation, Abp8247-A-SE), anti-AtAHA (PHYTOAB, PHY2285A) and anti-AtAVP1 (PHYTOAB, PHY2176S) indicated the cytosol, whole membrane, tonoplast fractionation, respectively. Anti-Actin (Abclonal, AC009) showed that the equal amount of initial materials was analyzed.

The vacuole H⁺-ATPase and H⁺-PPase activity assay

The vacuole H⁺-ATPase and H⁺-PPase activity assays were carried out as described in refs. ^{16,71}. Briefly, the hydrolytic activity of H⁺-ATPase was determined by measuring the release of Pi from hydrolysis of ATP. The tonoplasts fraction was prepared from 20 g leaves of 3-week-old maize plants cultivated under normal conditions. Fractionated tonoplast protein of 30 µg was used for the assay in a reaction volume of 0.5 ml.

Drought resistance assay in lab and field conditions

As for the drought resistance assay in lab conditions, different genotype plants were planted side by side in a cultivation box (35 × 20 × 10 cm, length × width × depth) that was filled with 1.5 kg of enriched soil (turf to vermiculite in a ratio of 1:1). A drought treatment was applied to plants at the three-leaf seedling stage by withholding water. Approximately, 17 d water-withholding was applied to the plants for the comparison between the WT and *ZmABF4-OE* (or *ZmRtn16-OE*) plants and 14 d for comparison of the WT and *zmrtn16-KO* (or *zmwha-E3-KO*) plants. Water was resumed to allow plants to recover, when the difference of leaf-wilt became most obvious. The number of surviving plants was recorded around 3–5 d later. As for the field test, the *ZmRtn16* over-expressing, CRISPR-targeted KO lines and WT were planted in a rain-out shelter in Beijing (39°980 N, 116°210 E; 45 m above mean sea level), where the mean day length was ~14 h and the average temperature was -23.1 °C during the growing season (from May 2021 to September 2021). The irrigation was cut off 30 d after sowing and resumed after all plants finished silking. In this period of time, the soil water potential gradually dropped to about -150 kPa. The phenotypes of plant height, days to anthesis, day to silking and anthesis and silking interval and grain yield per plant were collected. The *ZmRtn16* overexpressing plant and WT were planted in fields in Zhangye (38°56' N, 100°27' E; 1551 m above mean sea level), where the mean day length was ~14 h and the average temperature was -20.5 °C, during the growing season. Four replicates for normal and drought treatments were carried out for each genotype plant. For each replicate, 13 seeds per row were sown at a density of 58,000 plants per hectare, and two rows were designed. The drought treatment was applied to the plants from the V5 (the fifth leaf with visible ligule) stage to the completion of silking. The soil water potential dropped gradually from -20 to -220 kPa. Typically, irrigation was supplied approximately every 10 d. The total irrigation for the drought plot was approximately 40% of that for the normal growing plot. At both sites, the drop irrigation system was set up for water supply. The soil water potential was automatically monitored by Watermark 200SS Sensor buried underground. For normal watering treatment, the soil water potential was maintained at 0 to -20 kPa during the whole growing season.

RNA motif scan

CISBP-RNA database (<http://cisbp-rna.ccb.utoronto.ca/>) was used to predict the RNA motif within the 28-bp insertion sequences in the 3'-UTR of *ZmRtn16*⁸⁷³.

mRNA decay assay

The plasmids of the *ZmRtn16*⁸⁷³ CDS (in a Myc-tag fusion) followed by either the 3'-UTR of *ZmRtn16*⁸⁷³ or that of *ZmRtn16*^{CIMBL55} were transfected into maize leaf mesophyll protoplasts. After 12 h incubation under room temperature, 1 µM Triptolide was supplemented into the cultivation. Samples were obtained at 0, 15 and 30 min, subsequently. The exogenous transcripts levels of the two alleles were determined by RT-qPCR analysis. All the primers used in this research and their sequences are listed in Supplementary Table 12.

Vacuole pH measurements

The vacuole pH measurements were carried out as described in ref. ⁷¹ with minor modifications. Briefly, the primary root samples from 5-day-old plants were soaked in the buffer, containing 0.5% sucrose, 10 mM Mes-KOH (pH 5.8), 10 µM 2',7'-bis-(2-carboxyethyl)-5-(and-6)-carboxyfluorescein (BCECF, Beyotime Biotechnology) and 0.02% Pluronic F-127, in darkness at 22 °C and then washed in sterile dH₂O for 10 min. BCECF fluorescence photographs were taken by a confocal laser scanning microscope (Zeiss880). All images were obtained of cells at the elongation zones of primary roots of 5-day-old plants. Fluorescence was excited at 488 and 458 nm, and the emissions of the wavelength of 530–550 nm were collected. Both fluorescence intensity and ratio were obtained by ImageJ.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All genomic data mentioned in this paper are available at NCBI under the project of **PRJNA765111**, including the CIMBL55 genome sequence (**JAJHUUH000000000**) and the raw sequencing data used for the assembly (PacBio, illumina, Bionano, Hi-C) and DNA methylation analysis (BS-seq) of CIMBL55, B73, Mo17, wild type (LH244) and *zmdrd1* mutants (**SRP338635**). Source data are provided with this paper. Supplementary figures and source data are available at Figshare (<https://doi.org/10.6084/m9.figshare.21709943>)⁷².

Code availability

We deposited customized scripts in the following GitHub repository (https://github.com/ttian627/CIMBL55_genome_assembly) and Zenodo (<https://doi.org/10.5281/zenodo.7523457>)⁷³.

References

51. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
52. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
53. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.*, <https://doi.org/10.1002/0471250953.bi0410s25> (2009).
54. Xiong, W., He, L., Lai, J., Dooner, H. K. & Du, C. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proc. Natl Acad. Sci. USA* **111**, 10263–10268 (2014).
55. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).

56. Campbell, M. S. et al. MAKER-P: a tool kit for the rapid creation, management and quality control of plant genome annotations. *Plant Physiol.* **164**, 513–524 (2014).
57. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
58. Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
59. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
60. Marcais, G. et al. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 (2018).
61. Quinlan, A. R. BEDTools: the swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinforma.* **47**, 1–34 (2014).
62. Chen, S. et al. Paragraph: a graph-based structural variant genotyper for short-read sequence data. *Genome Biol.* **20**, 291 (2019).
63. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinform.* **10**, 232 (2009).
64. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
65. Crisp, P. A. et al. Variation and inheritance of small RNAs in maize inbreds and F1 hybrids. *Plant Physiol.* **182**, 318–331 (2020).
66. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
67. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
68. Yoo, S. D., Cho, Y. H. & Sheen, J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).
69. Ding, S., Zhang, B. & Qin, F. Arabidopsis RZFP34/CHYR1, a ubiquitin E3 ligase, regulates stomatal movement and drought tolerance via SnRK2.6-mediated phosphorylation. *Plant Cell* **27**, 3228–3244 (2015).
70. Chen, H. et al. Firefly luciferase complementation imaging assay for protein–protein interactions in plants. *Plant Physiol.* **146**, 368–376 (2008).
71. Krebs, M. et al. Arabidopsis V-ATPase activity at the tonoplast is required for efficient nutrient storage but not for sodium accumulation. *Proc. Natl Acad. Sci. USA* **107**, 3251–3256 (2010).
72. Qin, F. et al. Supplementary information for 'Genome assembly and genetic dissection of a prominent drought-resistant maize germplasm', <https://doi.org/10.6084/m9.figshare.21709943> (2023).
73. Tian, T. et al. Perl scripts used in 'genome assembly and genetic dissection of a prominent drought-resistant maize germplasm'. Zenodo. <https://doi.org/10.5281/zenodo.7523457> (2023).

Acknowledgements

We thank the great technical support of generating and propagating the transgenic maize provided by the staff in the Center of Crop Functional Genomics and Molecular Breeding at CAU. This research was supported by the Beijing Outstanding Young Scientist Program (BJJWZYJH01201910019026), the National Key Research and Development Program of China (2021YFD1200703), the National Natural Science Foundation of China (31625022, 32171940) and Chinese Postdoctoral Science Foundation (2019M660874, 2021T140714).

Author contributions

F.Q. designed and supervised the study and revised the manuscript. T.T. and S.W. analyzed the data, performed the experiments and drafted the manuscript. S.Y. identified the genetic variants in 30 maize accessions. Z.Y. generated the *zmdrd1-ko* plants. Y.W. and H.G. provided phenotypes for CIMBL55, B73 and Mo17 in fields. S.L. assisted in the candidate gene association analysis. S.Z. assisted in all the greenhouse experiments. X.Y. and C.J. provided maize materials and important suggestions for the work. All the author(s) read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-023-01297-y>.

Correspondence and requests for materials should be addressed to Feng Qin.

Peer review information *Nature Genetics* thanks Klaus Mayer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Pacbio RSII and sequel II platform produced the SMRT long reads; Bionano irys platform produced the labeled DNA molecules; Illumina Hiseq X10 platform produced whole genome resequencing reads, RNA-seq reads, Hi-C reads and bisulfite-seq reads for B73, Mo17, CIMBL55; BGI DNBSEQ-500 produced bisulfite-seq reads for two zmdrd1-KO lines and WT (LH244).

Data analysis

Genome assembly: NextDenovo v2.3.1, NextPolish v1.3.0, Bionano Solve v3.6, SALSA; Genome annotation: RepeatModeler v2.0.1, RepeatMasker v4.1.1, HelitronScanner_V1, Trinity v2.11.0, Augustus v 3.2.3, PASA v 2.4.1, MAKER v 3.01.03; Synteny analysis: MCscanX (<https://github.com/wyp1125/MCScanX>); Structure variation analysis: MUMmer v4.0, Pbsv v2.2.1, pbmm2 v1.0.0, Sniffles v1.0.11, ngmlr v0.2.7, Smartie-sv (<https://github.com/zeeev/smartie-sv>), cuteSV 1.0.12, SyRi v1.4, Paragraph v2.4a; BS-seq data analysis: BSMAP v2.9, Deeptools 3.4.3; sRNA-seq analysis: Trimmomatic v0.35, Bowtie v1.2.3, BEDTools v2.29.2, Deeptools 3.4.3; GWAS analysis:TASSEL v5.0. We deposited customized scripts in the following GitHub repository: https://github.com/ttian627/CIMBL55_genome_assembly.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All genomic data in this paper is available at NCBI under the project of PRJNA765111, including the CIMBL5 genome sequence and gene annotations, the raw data used for the assembly (PacBio, illumina, Bionano, Hi-C), as well as BS-seq data for CIMBL55, B73 and Mo17, wild type (LH244) and zmdrd1 mutants. We deposited customized scripts for bi-directional SV identification in the GitHub repository: https://github.com/ttian627/CIMBL55_genome_assembly.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

N/A

Population characteristics

N/A

Recruitment

N/A

Ethics oversight

N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The sample size for GWAS was the number of inbred lines, 224 (Liu et al., 2021) and 368 (Wang et al., 2016), in an association mapping panel from previous studies.
The sample size for the statistic analysis of each experiment was clearly mentioned in each figure legend or Methods.

Data exclusions

Genotyped SVs were filtered with a minor allele frequency (MAF) > 0.05 and a missing rate (MR) < 0.6. This is commonly applied since the current method is not suitable to discover the association of this kind of rare alleles with the trait.

Replication

Replications for each experiment were clearly stated in figure legends or Methods section.
Replications of repeated experiments were evaluated by proper statistic analyses and confirmed to be successful.

Randomization

For phenotype analysis of the gene overexpression or mutant, they were planted with wild-type plants side by side in a random arrangement.

Blinding

Blinding was not considered as the authors were aware of genotypes when performing analyses. No blinding should not affect interpretation as all our experiment measures were objective.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Mouse-anti GFP, Abclonal, Cat#AE012, immune incubation: 1/3000;
 Mouse-anti HA, Abclonal, Cat#AE008, immune incubation: 1/3000;
 Mouse-anti Myc, Sigma-Aldrich, Cat#M4439, immune incubation: 1/3000;
 Mouse-anti Actin, Abclonal, Cat#AC009, immune incubation: 1/10000;
 Rabbit-anti AtVHA-A, PHYTOAB, cat#PHY1480S, immune incubation: 1/1000;
 Rabbit-anti AtVHA-E1/E3, PHYTOAB, cat#PHY1035S, immune incubation: 1/1000;
 Rabbit-anti OsFBPA, Beijing Protein Innovation, cat#Abp8247-A-SE, immune incubation: 1/2000;
 Rabbit-anti AtAHA, PHYTOAB, cat#PHY2285A, immune incubation: 1/1000;
 Rabbit-anti AtAVP1, PHYTOAB, cat#PHY2176S, immune incubation: 1/1000;

Validation

Mouse-anti GFP (<https://abclonal.com.cn/catalog/AE012>), mouse-anti Actin (<https://abclonal.com.cn/catalog/AC009>) and mouse-anti HA (<https://abclonal.com.cn/catalog/AE008>), Mouse-anti Myc (<https://www.sigmaldrich.cn/CN/zh/product/sigma/m4439>), have been widely used, and their validations were provided by the manufacturer's website.
 The validation of Rabbit-anti AtVHA-A (<https://www.phytoab.com/vha-a%20antibody>), rabbit- anti AtVHA-E1/E3 (<https://www.phytoab.com/vha-e3%20antibody>), rabbit- anti AtAHA (<https://www.phytoab.com/aha1%20antibody>) and rabbit- anti-AtAVP1 (<https://www.phytoab.com/avp1%20antibody>) can be found in the manufacturer's website. In addition, we validated rabbit- anti-AtAVP1 and rabbit-anti AtVHA-E1/E3 antibody using maize transgenic plants.