nature plants

LETTERS

https://doi.org/10.1038/s41477-019-0443-7

Mutation of *ZmDMP* enhances haploid induction in maize

Yu Zhong[1,2], Chenxu Liu[1,2], Xiaolong Qi[1,2], Yanyan Jiao[1,2], Dong Wang[1], Yuwen Wang[1], Zongkai Liu[1], Chen Chen[1], Baojian Chen[1], Xiaolong Tian[1], Jinlong Li[1], Ming Chen[1], Xin Dong[1], Xiaowei Xu[1], Liang Li[1], Wei Li[1], Wenxin Liu[1], Weiwei Jin[1], Jinsheng Lai[1] and Shaojiang Chen[1*]

Doubled haploid (DH) breeding based on in vivo haploid induction has led to a new approach for maize breeding[1]. All modern haploid inducers used in DH breeding are derived from the haploid inducer line *Stock6*. Two key quantitative trait loci, *qhir1* and *qhir8*, lead to high-frequency haploid induction[2]. Mutation of the gene *MTL/ZmPLA1/NLD* in *qhir1* could generate a ~2% haploid induction rate (HIR)[3–5]; nevertheless, this mutation is insufficient for modern haploid inducers whose average HIR is ~10%[6]. Therefore, cloning of the gene underlying *qhir8* is important for illuminating the genetic basis of haploid induction. Here, we present the discovery that mutation of a non-*Stock6*-originating gene in *qhir8*, namely, *ZmDMP*, enhances and triggers haploid induction. *ZmDMP* was identified by map-based cloning and further verified by CRISPR–Cas9-mediated knockout experiments. A single-nucleotide change in *ZmDMP* leads to a 2–3-fold increase in the HIR. *ZmDMP* knockout triggered haploid induction with a HIR of 0.1–0.3% and exhibited a greater ability to increase the HIR by 5–6-fold in the presence of *mtl/zmpla1/nld*. *ZmDMP* was highly expressed during the late stage of pollen development and localized to the plasma membrane. These findings provide important approaches for studying the molecular mechanism of haploid induction and improving DH breeding efficiency in maize.

Doubled haploid (DH) breeding via in vivo haploid induction has been extensively used in modern maize breeding[1]. The production of haploids depends on haploid inducers, and the ancestral haploid inducer line *Stock6* can induce ~3% of maternal haploids[7]. The haploid induction rate (HIR) is heritable and can be improved by phenotypic selection, allowing the breeding of modern haploid inducers with increased HIR values[6,8]. The HIR of modern haploid inducers has been improved to ~10%, which makes it possible to produce DH lines efficiently in large-scale commercial breeding programmes[1]. With respect to the genetic basis of haploid induction, a comprehensive quantitative trait loci mapping study of the HIR demonstrated that two loci function significantly[2], that is, *qhir1* in bin 1.04 and *qhir8* in bin 9.01 explained ~66% and ~20%, respectively, of the genetic variance. Based on the fine mapping of *qhir1* (ref. [9]), the *Stock6*-derived haploid-induction gene *MTL/ZmPLA1/NLD* was cloned[3–5,10]. *qhir8* exists in high-HIR inducer lines such as CAU5, whereas *qhir8* is absent in low-HIR inducer lines such as CAUHOI. *qhir8* increases the HIR by 2–3-fold in the presence of *qhir1* (ref. [11]). Moreover, mutation of the orthologues of *ZmPLA1/MTL/NLD* could trigger haploid induction in other crop species such as rice, which makes cloning of *ZmDMP* extremely promising with respect to crop breeding[12].

In 2016, based on the 789-kb mapping region of *qhir8*, 16 F_3 families were used to narrow the mapping region to a 138-kb region flanked by markers ZS4307 and ZS4446 (Fig. 1a and Supplementary Fig. 1). To analyse the sequence within the mapping region in CAU5, two positive clones covering the region were screened from a CAU5 bacterial artificial chromosome (BAC) library[4]; these clones provided sequence information and polymorphic markers (Supplementary Fig. 2 and Supplementary Table 1). Using these markers, we ultimately narrowed the mapping region to a 318-bp region via the use of 21 new F_3 families (Fig. 1a and Supplementary Fig. 1). The final mapping region was located within the protein-coding sequence of the gene GRMZM2G465053, which encodes a DUF679 domain membrane protein and was named *ZmDMP* (Fig. 1b). One single-nucleotide substitution was found in the mapping region between CAU5 and CAUHOI, that is, from thymine (T) (CAUHOI) to cytosine (C) (CAU5), at 131 bp from the initiating codon ATG, which led to an amino acid substitution from methionine to threonine (Fig. 1b). Therefore, *ZmDMP* was considered the candidate gene, and the single-base substitution was considered to contribute to the high HIR observed for CAU5.

To verify the candidate, we knocked out *ZmDMP* (*zmdmp*-ko) with the clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9) system[13]. Two transgenic events, T0–15 and T0–17, which harboured frameshift mutations within *ZmDMP*, were chosen for subsequent experiments (Fig. 1c). Individual plants with the genotype combinations *zmpla1*-(*zmdmp*-ko), *zmpla1-ZmDMP* and *zmpla1-ZmDMP/zmdmp*-ko (*zmpla1*-heterozygous) were screened in the F_2 population derived from T0–15 × CAUHOI and T0–17 × CAUHOI with molecular markers (Supplementary Table 1). These individuals were planted and pollinated to hybrid ZhengDan958 (ZD958) to evaluate their HIR. For the F_2 individuals derived from either T0–15 × CAUHOI or T0–17 × CAUHOI, the average HIR was ~7% for the genotype combination *zmpla1*-(*zmdmp*-ko), which was significantly higher than the ~1% value for the genotype combination *zmpla1-ZmDMP*. The average HIR for the genotype combination *zmpla1*-heterozygous was in between that for *zmpla1*-(*zmdmp*-ko) and *zmpla1-ZmDMP* (Wilcoxon rank-sum test, $P < 0.01$) (Fig. 2a). When grown in the field, all putative haploid plants exhibited typical haploid characteristics, for example, plants were short and sterile (Fig. 3a–c). Haploids were further verified using ten polymorphic molecular markers (Fig. 3d, Supplementary Fig. 3 and Supplementary Table 1) and flow cytometry (Fig. 3e), which revealed that all were true haploids with only the female genotype, that is, demonstrated their maternal haploid properties. In accordance with the functional

[1]National Maize Improvement Center of China, Key Laboratory of Crop Heterosis and Utilization (MOE), China Agricultural University, Beijing, China. [2]These authors contributed equally: Yu Zhong, Chenxu Liu, Xiaolong Qi, Yanyan Jiao. *e-mail: chen368@126.com

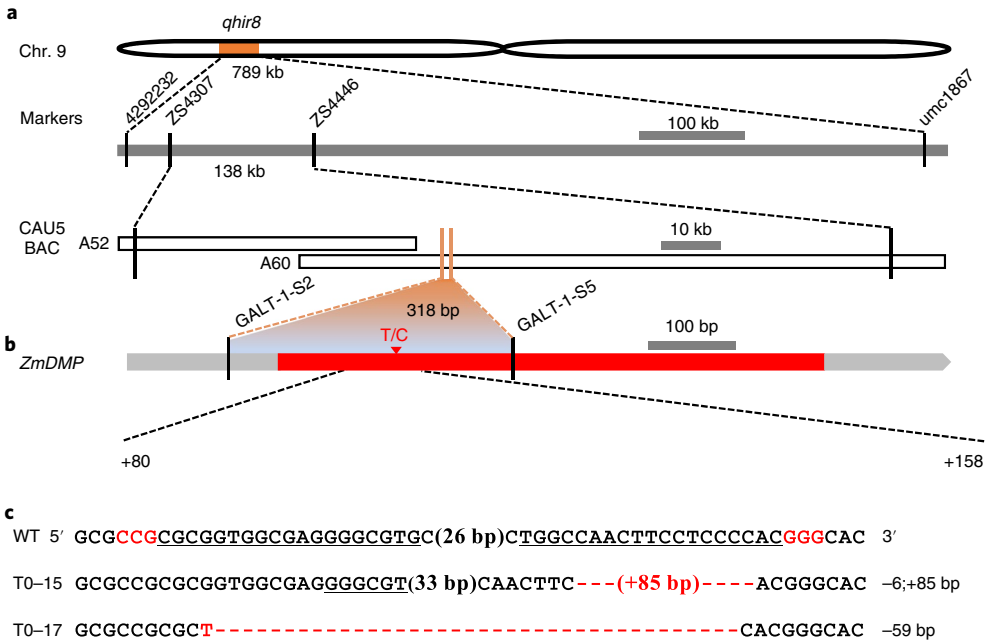

Fig. 1 | Map-based cloning of *ZmDMP* and knockout experiments. a, Sequential fine mapping of the quantitative trait locus *qhir8*. The vertical lines represent the sites of key molecular markers, and the two orange vertical lines indicate the flanking markers GALT-1-S2 and GALT-1-S5 in the final mapping step. CAU5 BAC contigs are represented by overlapping, unfilled rectangles. Chr., chromosome. **b**, Structure of *ZmDMP* and the causal mutation leading to an amino acid substitution. The red region is the protein-coding region, and the grey ends represent the 5′ untranslated region and 3′ untranslated region. **c**, The CRISPR–Cas9 target sequences are underlined, and the protospacer adjacent motif sequences are marked in red on the candidate gene. Also shown is the sequence verification of the knockout lines. WT, wild type.

attributes of *qhir8*, which can lead to a high frequency of endosperm aborted kernels (EnAs)[11], we also observed a high frequency of EnAs in ZD958 ears pollinated by *zmpla1*-(*zmdmp*-ko) (Fig. 2c). Statistical analysis revealed that, in F_2 individuals derived from either T0–15 or T0–17, the presence of homozygous *zmdmp*-ko significantly increased the EnA rate from ~10% to ~40% (Fig. 2b). The fact that the *zmdmp*-ko allele resulted in a significant increase in both the HIR and the EnA rate suggested that *zmdmp* is responsible for haploid induction. Moreover, we found that the *zmdmp*-ko allele contributed significantly more (approximately sixfold increase) to the HIR than did the natural allele of CAU5 (Fig. 2d). This effect of the *zmdmp*-ko allele on the HIR may represent a new way to improve the HIR of modern inducer lines using CRISPR–Cas9.

Considering the independent role of *zmpla1* in haploid induction, we also tested whether *zmdmp*-ko could act independently in haploid induction when pollinated to hybrids or self-pollinated. Six haploids were found among the 5,080 progenies obtained from the ZD958 × *zmdmp*-ko cross, whereas no haploid was detected among 4,039 control individuals from ZD958 × *ZmDMP* (Supplementary Table 2). Three haploids were identified among the 949 self-pollinated progenies of *zmdmp*-ko, whereas no haploid was detected among the 844 self-pollinated progenies of *ZmDMP*. Therefore, these results imply that knockout of *ZmDMP* can lead to independent haploid induction.

Although the ability of *qhir8* to promote haploid induction was well established in our previous studies[2,11], the implied interaction between *qhir1* and *qhir8* has not been assessed. Therefore, by constructing near-isogenic lines (NILs) in a B73 background, including CAU6$^{(ZmPLA1\text{-}ZmDMP)}$, CAU6$^{(ZmPLA1\text{-}zmdmp)}$, CAU6 $^{(zmpla1\text{-}ZmDMP)}$ and CAU6$^{(zmpla1\text{-}zmdmp)}$, we studied the genetic interaction between *ZmPLA1* and *ZmDMP* via pollinating the hybrid ZD958. We first determined the percentage of undeveloped zygotes by calculating the ratio of visible seeds to the number of embryo sacs, that is, the silk number. The wild-type NIL CAU6$^{(ZmPLA1\text{-}ZmDMP)}$ had ~15% undeveloped kernels and 1.36% kernel abortions, and the HIR was as low as 0.13% (Supplementary Fig. 4a). The presence of *zmdmp* in CAU6$^{(ZmPLA1\text{-}zmdmp)}$ led to a slight increase in the percentage of undeveloped kernels, that is, from 15% to 21%, and the percentage of diploid kernels decreased to 76.9%; the HIR was 0.15%. For CAU6$^{(zmpla1\text{-}ZmDMP)}$, we observed a sharp decrease in the percentage of diploid kernels, that is, from 82.8% to 39.6%, whereas the percentage of undeveloped kernels increased to 45.9%. Moreover, the percentages of haploids, EnA and embryo aborted kernels (EmA) increased significantly. In the double-mutant line CAU6$^{(zmpla1\text{-}zmdmp)}$, the percentage of undeveloped kernels increased from 45.9% to 56.8%, and the proportion of both EnA and EmA nearly doubled relative to CAU6$^{(zmpla11\text{-}ZmDMP)}$. The HIRs for both CAU6$^{(ZmPLA1\text{-}ZmDMP)}$ and CAU6$^{(ZmPLA1\text{-}zmdmp)}$ were very low but were significantly higher in CAU6$^{(zmpla1\text{-}ZmDMP)}$. The HIR of CAU6$^{(zmpla1\text{-}zmdmp)}$ was nearly triple that of CAU6$^{(zmpla1\text{-}ZmDMP)}$ (Supplementary Fig. 4a). Therefore, *zmdmp* had a greater contribution to the HIR in the presence of *zmpla1*.

Staining of pollen with fluorescein diacetate (FDA) among the four genotype combinations revealed that the percentage of high-viability pollen ranged from 42.4% to 44.5%, the percentage of low-viability pollen ranged from 53.6% to 55.3% and that of dead pollen ranged from 1.8% to 2.3% (Supplementary Fig. 4b,c). For each pollen viability class, the *t*-test revealed no significant difference between any two of the four genotype combinations. This result was further verified by a pollen germination experiment, which also did not show any difference among the four genotype combinations (Supplementary Fig. 4b,d). These results indicated that pollen viability was not significantly affected by *zmpla1* or *zmdmp* or their combination.

As previously reported, wild-type *ZmPLA1* is a plasma membrane-anchored protein that is highly expressed in developing and mature pollen[5]. Expression pattern analysis of *ZmDMP* and *zmdmp* via quantitative reverse transcription PCR (qRT–PCR) revealed that each of *ZmDMP* and *zmdmp* is highly expressed in mature pollen

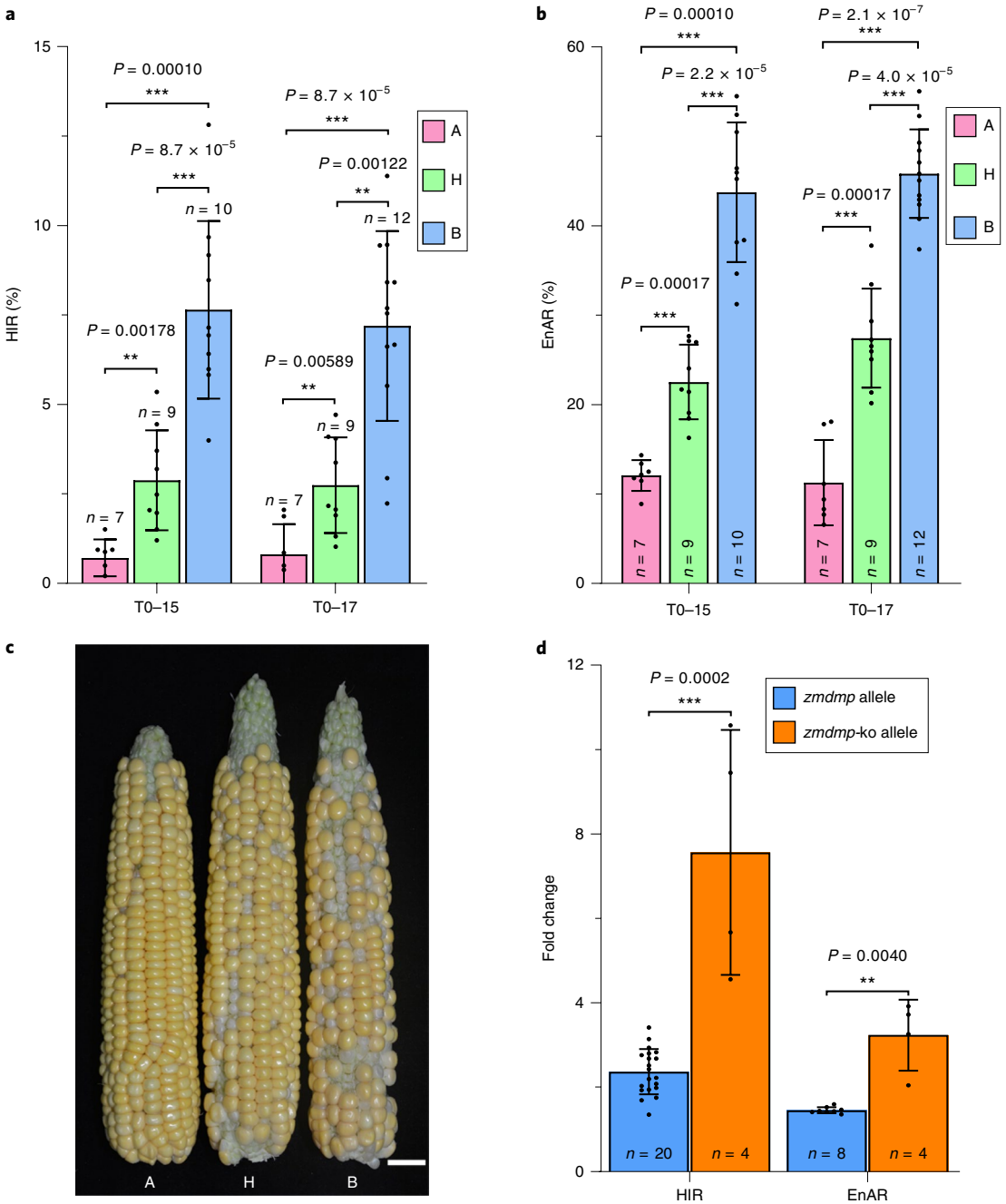

Fig. 2 | Phenotypic evidence of transgenic events that enhanced HIRs. a, Bar plot of the HIR of ZD958 ears pollinated by genotype classes A, H and B derived from transgenic events T0-15 and T0-17. **b**, Bar plot of the EnA rate (EnAR) of ZD958 ears pollinated by genotype classes A, H and B derived from transgenic events T0-15 and T0-17. *n* indicates the number of ears used for calculating the HIR and the EnAR of each genotype. **c**, Performance of ZD958 ears pollinated by genotype classes A ($n = 30$), H ($n = 46$) and B ($n = 63$). Scale bar, 2 cm. Genotype classes A, H and B represent the genotype combinations *zmpla1-ZmDMP*, *zmpla1*-heterozygous and *zmpla1*-(*zmdmp*-ko), respectively (**a**–**c**). **d**, The blue bar plot shows the fold change of the HIR and the EnAR between genotype classes *zmpla1-zmdmp* (natural allele) and *zmpla1-ZmDMP* in the CAU5–CAUHOI F_2 population; the orange bar plot shows the fold change of the HIR and the EnAR between *zmpla1*-(*zmdmp*-ko) (knockout allele) and *zmpla1-ZmDMP* in the F_2 population. *n* indicates the number of F_2 population. Data represent the mean ± s.d.; **, $P < 0.01$, ***$P < 0.001$ (two-sided Wilcoxon rank-sum test) (**a**,**b**,**d**).

but at much lower levels in immature anthers and kernels at different stages (Fig. 4a), implying that *ZmDMP* may function during the late stage of gametophyte development. In mature pollen, *zmdmp* is expressed at levels significantly higher than *ZmDMP*, presumably as a consequence of feedback regulation. To confirm the developmental expression pattern of *ZmDMP*, we generated transgenic plants harbouring either pZmDMP::GUS (β-glucuronidase) or pzmdmp::GUS. Staining for GUS revealed strong expression of both pZmDMP::GUS and pzmdmp::GUS in pollens, silks and seeds before 6 days after pollination (DAP) (Fig. 4b–d), which was consistent with the qRT–PCR results. The proteins encoded by either the wild-type Ubi::ZmDMP-eGFP (Ubi, ubiquitin; eGFP, enhanced green fluorescent protein) (CAUHOI) or mutant Ubi::zmdmp-eGFP (CAU5) colocalized with the plasma membrane marker AtPIP2a-mCherry in maize protoplasts (Fig. 4e,f), whereas the GFP signal for the control group was

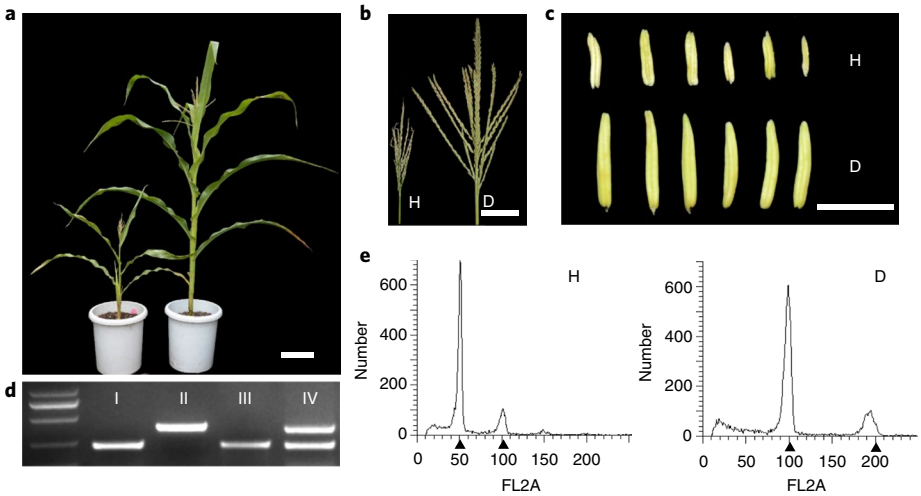

Fig. 3 | Verification of the haploid phenotypes. a–c, Phenotypes of plants (**a**), tassels (**b**) and anthers (**c**) of haploids (H) and diploids (D). Scale bars, 0.2 m (**a**), 10 cm (**b**) and 6 mm (**c**). **d**, Identification of haploids and diploids with polymorphic markers between the transgenic receptor line and tester. The left lane shows DNA markers, and I, II, III and IV represent the bands of the tester hybrid ZD958, receptor, haploid and diploid, respectively. **e**, Verification of the ploidy of haploids and diploids with flow cytometry. The *x* axis represents the signal peak for the nucleus, and the *y* axis represents the number of nuclei. In **a**–**e**, experiments were repeated 514 times and similar results were obtained.

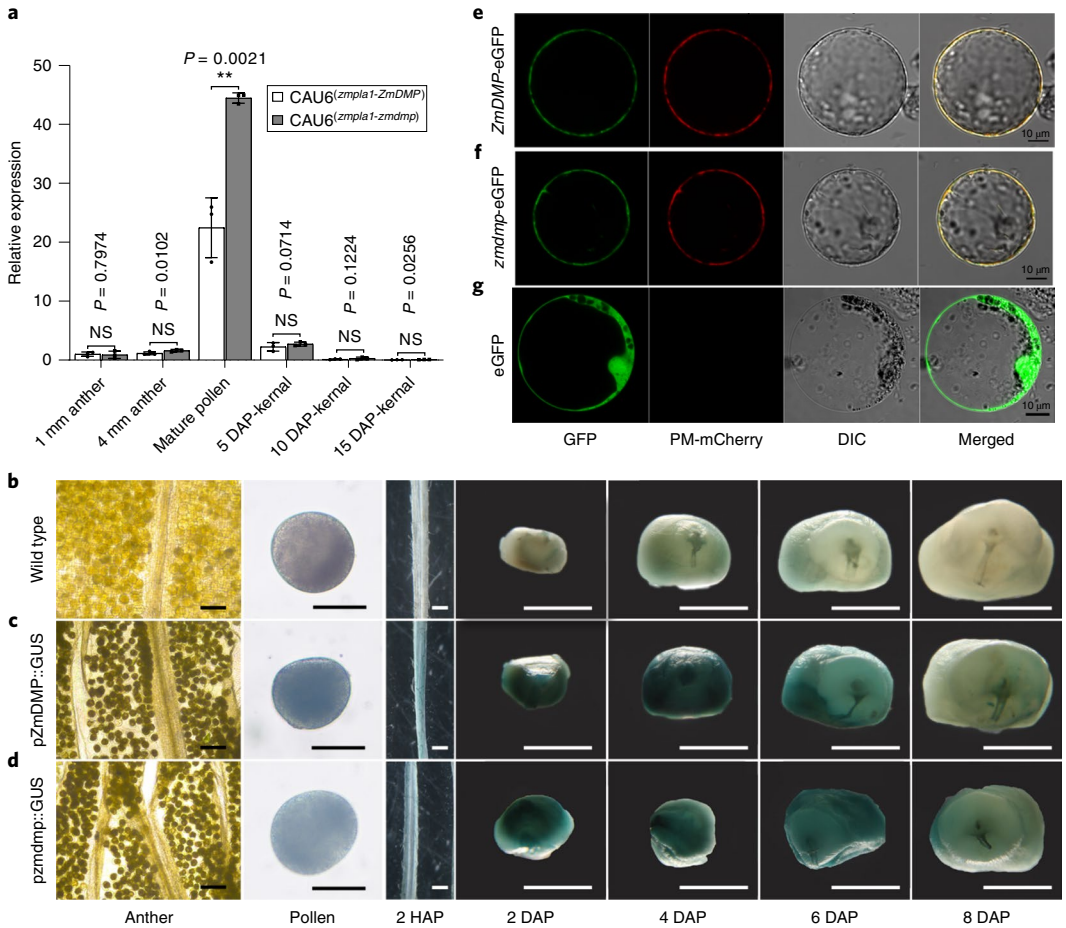

Fig. 4 | Expression and subcellular localization profiling of *ZmDMP* and *zmdmp*. a, Expression analysis of *ZmDMP* and *zmdmp* in NILs between the genotypes CAU6$^{(zmpla1\text{-}ZmDMP)}$ and CAU6$^{(zmpla1\text{-}zmdmp)}$ among different tissues. Data are the means ± s.d. of three biologically independent samples (each sample with three technical repetitions). A two-sided Student's *t*-test was used to determine significant differences. $^{**}P < 0.01$; NS, not significant. **b**–**d**, GUS staining pattern of anthers, pollen, silks (2 h after pollination) and seeds from 2 DAP to 8 DAP in the wild-type plant (**b**), the pZmDMP::GUS transgenic plant (**c**) and the pzmdmp::GUS transgenic plant (**d**). Scale bars, 200 μm (anthers), 20 μm (pollen), 500 μm (silks) and 2.5 mm (seeds). Experiments were repeated three times and similar results were obtained. **e**–**g**, Confocal scanning (GFP), plasma membrane (PM)-mCherry signal, differential interference contrast (DIC) and merged micrographs of maize protoplast cells transformed with a Ubi::ZmDMP-eGFP construct (**e**), a Ubi::zmdmp-eGFP construct (**f**) or a Ubi::eGFP vector control (**g**). Experiments were repeated two times and similar results were obtained.

observed at both the plasma membrane and the nucleus (Fig. 4g). These results demonstrated that *ZmDMP* and *ZmPLA1* have similar expression patterns and subcellular localization.

Although *zmdmp* has already been fixed in modern haploid inducers, knowledge concerning its origin is scarce. An analysis of 50,000 single-nucleotide polymorphisms (SNPs) of 53 haploid inducers[6] revealed that the haplotype of *zmdmp* is conserved in 32 recently developed haploid inducers but diverse in 21 haploid inducers developed earlier (Supplementary Fig. 5a); the average HIR of the 32 recent haploid inducer lines with *zmdmp* was significantly higher than the 21 inducer lines without *zmdmp* (Supplementary Fig. 5b). In combination with the pedigree of the haploid inducers[6], this result suggests that *qhir8* was strongly selected and became rapidly fixed in the newly developed inducer lines under phenotype selection. Analysis of the *ZmDMP* sequence in 138 non-inducer lines revealed that 19 lines carry the same haplotype as CAU5 and 44 lines carry the same haplotype as CAUHOI (Supplementary Table 3), demonstrating that the causal SNP that we identified is not rare but exists widely in different germplasms. A recent study reported that the in vivo haploid induction system in maize can be extended to other crop species, such as rice[12], which undoubtedly improves the significance of the system. Phylogenetic analysis of *ZmDMP* in plants revealed that this gene is quite conserved among maize, sorghum and foxtail millet, with amino acid sequence identities of ~91% (maize-sorghum) and ~73% (maize-millet) (Supplementary Fig. 6 and Supplementary Table 4). These results highlight the potential application of these orthologues in DH breeding of different crop species.

These findings reveal that the SNP in *ZmDMP* significantly improves the HIR in the presence of *mtl/zmpla1/nld*. The novel functions of the knockout allele, including independent haploid induction and greater contribution to the HIR, highlight the huge potential for further improvement of maize DH breeding efficiency. The characterization of *ZmDMP* revealed its similarity with *ZmPLA1*, which provides an important clue for future studies on the molecular mechanism of haploid induction. Considering the high sequence conservation among orthologues of *ZmDMP*, our findings have the potential to be applied rapidly along with the help of gene editing to accelerate crop breeding.

Methods

Materials. During the ultrafine mapping of *qhir8*, we used two haploid inducer lines, namely, CAUHOI (HIR: ~2%) and CAU5 (HIR: ~10%, the donor parent of *qhir8*), to generate a $F_{2,3}$ mapping population[11]; these lines differed for the analysis of *qhir8*. Recombinants were screened from the F_2 population and self-pollinated to obtain F_3 families. A receptor line that lacks haploid induction ability was used for transformation. The transgenic events were crossed with CAUHOI to obtain mutant combinations between *zmpla1* and *zmdmp*-ko. To verify the effects of the candidate gene, we pollinated a single hybrid tester, ZD958, with pollen from the three combinations: *zmpla1*-(*zmdmp*-ko), *zmpla1-ZmDMP* and *zmpla1*-heterozygous. To determine sequence variations in *ZmDMP*, we sequenced the gene among 138 inbred lines that were chosen from different germplasms (Supplementary Table 3). CAU[6] is a haploid inducer line with both *qhir1* and *qhir8* in the B73 background, and B73_*R1-nj* is a B73 inbred line with the marker *R1-nj*. The NILs of $CAU6^{(ZmPLA1\text{-}ZmDMP)}$, $CAU6^{(ZmPLA1\text{-}zmdmp)}$, $CAU6^{(zmpla1\text{-}ZmDMP)}$ and $CAU6^{(zmpla1\text{-}zmdmp)}$ in the B73 background were constructed by crossing CAU6 with B73_*R1-nj*, followed by two generations of self-pollination. The four NILs share a ~95% identical background as assessed with an Illumina maize SNP 6K genotyping array. *ZmPLA1* and *ZmDMP* represent the wild-type alleles, *zmpla1* represents the CAUHOI/CAU5-mutant allele, *zmdmp* represents the CAU5-mutant allele and *zmdmp*-ko represents the knockout allele.

Ultrafine mapping of *qhir8*. Based on the 789-kb mapping region of *qhir8* that we identified[11], we further narrowed the mapping region using a progeny strategy[9,11,14]; that is, recombinants in a large F_2 population were screened with newly developed molecular markers, after which the recombinants were self-pollinated to produce F_3 progeny. Genotyping was performed to classify F_3 individuals as the homozygous CAU5, homozygous CAUHOI or the heterozygous genotype. Moreover, phenotyping was performed for each genotype class by pollinating at least six ears of hybrid ZD958. The Wilcoxon rank-sum test was used to calculate phenotypic differences among the different genotype classes[11]. For each F_3 family, significant differences among the three genotype classes were inferred to reflect the presence of *qhir8*; otherwise, *qhir8* was inferred to be absent.

Knockout of *ZmDMP* with the CRISPR–Cas9 system. Two guide RNAs were designed to target two sites within the exon of the candidate gene *ZmDMP*, and the sequences for the guide RNAs were then inserted into the pBUE411 vector[13]. At 12 DAP, embryos from the receptor line were used for *Agrobacterium* (EHA105)-mediated transformation. Positive transformation events were screened with bialaphos medium (1.5 mg l^{-1}) and verified by DNA sequencing at the seedling stage. Knockout lines with frameshift mutations were transplanted to a greenhouse and then self-pollinated to acquire homozygous knockout mutants.

Phenotyping of knockout lines. To determine whether *zmdmp*-ko could act independently to trigger haploid induction, the knockout plants with a homozygous mutant genotype were self-pollinated and then pollinated to the single hybrid ZD958. The resultant putative haploids were screened among the crossed and self-pollinated progenies, and the wild-type lines were used as a control. To verify the effect of *zmdmp*-ko in the presence of *zmpla1*, an F_2 population was obtained by crossing *zmdmp*-ko plants with CAUHOI. The three genotype combinations *zmpla1*-(*zmdmp*-ko), *zmpla1-ZmDMP* and *zmpla1*-heterozygous in the F_2 population were screened with molecular markers. The HIR values for the three genotype classes were determined by crossing with ZD958. Differences in phenotype among the three genotype classes were assessed via the Wilcoxon rank-sum test, and a significant difference was taken to mean that the knockout allele had an effect on haploid induction.

Identifying haploid plants and classifying aborted kernels. During the ultrafine mapping of *qhir8*, because all individuals carried homozygous *R1-nj* markers, haploids could be recognized accurately with *R1-nj*. However, no kernel marker was available for transgenic plants and their combinations with CAUHOI, so we applied two molecular markers (Supplementary Fig. 3 and Supplementary Table 1; Chr3–26.4 and Chr5–94.2) with polymorphism between the receptor line and ZD958 to screen putative haploids in a large population at the seedling stage. Then, we determined whether these haploids had a maternal origin using ten additional molecular markers (Supplementary Fig. 3 and Supplementary Table 1). The 12 markers distributed on 7 different chromosomes. Flow cytometry was then used to verify the ploidy of the haploids, and those haploids whose peaks were analogous to those of standard haploids were deemed true haploids, and those haploids whose peaks were analogous to those of standard diploids were deemed diploids[4]. The field performance of these putative haploids, which had short and male-sterile phenotypes, was evaluated again with flow cytometry. Except for haploids, the abnormally developed kernels appeared on the ears of ZD958, including EnA and EmA; therefore, they were classified according to previous standards[15]. The HIR, EmA rate and EnA rate were calculated according to Xu et al.[15].

Subcellular localization. Total RNA was isolated from mature pollen of both *zmdmp* (CAUHOI) and *ZmDMP* (CAU5) via TRIzol reagent. Reverse transcription was performed with oligo(dT) primers to obtain full-length complementary DNAs. The coding sequence of the gene without a termination codon (TAA) was cloned and inserted into the vector PCUN-eGFP, which is driven by the ubiquitin promoter. The construct AtPIP2a-mCherry, the product of which localizes to the cytoplasmic side of the maize protoplast membrane, was used as a marker. Sequencing-validated constructs of Ubi::ZmDMP-eGFP, Ubi::zmdmp-eGFP and AtPIP2a-mCherry were used for polyethylene glycol-mediated transformation into maize protoplasts. After culturing at 25 °C for 16 h, the colocalization signal of each of Ubi::ZmDMP-eGFP and AtPIP2a-mCherry and of Ubi::zmdmp-eGFP and AtPIP2a-mCherry was observed and imaged with a confocal microscope (Zeiss 710). As a control, protoplasts were also transformed with the unmodified vector PCUN-eGFP.

Expression analysis. To analyse the expression characteristics of *ZmDMP*, 1-mm immature anthers, 4-mm immature anthers, mature pollen and young kernels at 5, 10 or 15 DAP of the NILs $CAU6^{(zmpla1\text{-}ZmDMP)}$ and $CAU6^{(zmpla1\text{-}zmdmp)}$ were collected with three biological replicates and then frozen immediately in liquid nitrogen. Total RNA was extracted with TRIzol and then reverse transcribed into cDNA. The primer RTDMP with a single unique PCR product of 181 bp was used for qRT–PCR, which was performed with an ABI 7500 system. To generate GUS reporter lines, the promoters (upstream of the ATG start codon) of *ZmDMP* (CAUHOI, 2,743 bp) and *zmdmp* (CAU5, 2,707 bp) were amplified with primers (Supplementary Table 1), and the resultant PCR products were ligated into *HindIII*/*BamHI*-linearized pCM3300M-GUS using the Seamless Assembly Cloning Kit (C5891–25, Clone Smarter). For GUS histochemical staining[16], plant tissues were first subjected to vacuum infiltration at 37 °C for 24 h in the following solution: 50 mM sodium phosphate (pH 7.0), 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 0.1% v/v Triton X-100 and 1 mM X-Gluc. Tissues were then washed in 70% ethanol. Pollen and mature anthers were observed from both pZmDMP::GUS and pzmdmp::GUS plants and receptor tissues including silks (2 h after pollination) and seeds at 2, 4, 6 or 8 DAP crossed with both pZmDMP::GUS and pzmdmp::GUS. Corresponding tissues from the self-pollinated receptor line were used as controls. Tissues were observed and photographed using a Nikon Eclipse Ti2 microscope and a Leica M60 stereo microscope.

Pollen viability evaluation. The FDA working solution was prepared by mixing 100 µl FDA stock solution (0.5% FDA in acetone) with 4.9 ml of 0.5 M sucrose solution[17]. Fresh pollen (20 mg) sampled from three biological replicates was immediately mixed with the working solution following pollen collection in the field. Each mixture was left in the dark for 1 h without cover to guarantee an abundant supply of oxygen for the esterase reaction. Fluorescence was observed with a fluorescence microscope (BX 53, Olympus) with an excitation wavelength of 485 nm. Pollen that exhibited strong green fluorescence emanating from the cytoplasm was deemed as high-viability pollen, pollen that exhibited weak fluorescence was deemed as low-viability pollen, and pollen that could not be stained was deemed dead. In the pollen germination experiment, fresh mature pollen was sampled from plants in the field and sprinkled into a medium with 18% sucrose, 0.03% boric acid, 0.03% calcium nitrate, 0.03% potassium nitrate, 0.02% magnesium sulphate and 9% polyethylene glycol 4000 (ref. [18]). Pollen was germinated at room temperature for 1 h. Images were captured using a Nikon Eclipse Ti2 microscope. Pollen with elongated tubes was deemed as viable pollen, and pollen without such tubes was deemed dead. Pollen viability was evaluated by calculating the germination rate.

Haplotype origin analysis. The 50,000 SNP chip-based genotypes of 53 haploid inducers and 1,169 inbred lines were downloaded from http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.184234/-/DC1/FileS4.zip (ref. [7]) and used to analyse the haplotypes of *ZmDMP* in the target region, which included 10 upstream SNPs and 10 downstream SNPs. Haploid inducers were grouped according to their haplotype, and the distribution of the HIR for each group was analysed via a box plot.

Phylogenetic analysis of the *ZmDMP*-encoded protein. The amino acid sequence encoded by *ZmDMP* and sequences for its top 17 most identical orthologues were downloaded from http://www.gramene.org/, and the sequence identities were calculated with DNAMAN 8. Sequence alignment was performed with the software MEGA-X, and the phylogenetic tree was constructed with the neighbour-joining algorithm.

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

Data availability

The data sets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 19 November 2018; Accepted: 9 May 2019;
Published online: 10 June 2019

References

1. Ren, J. et al. Novel technologies in doubled haploid line development. *Plant Biotechnol. J.* **15**, 1361–1370 (2017).
2. Prigge, V. et al. New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize. *Genetics* **190**, 781–793 (2012).
3. Kelliher, T. et al. MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* **542**, 105–109 (2017).
4. Liu, C. et al. A 4-bp insertion at ZmPLA1 encoding a putative phospholipase A generates haploid induction in maize. *Mol. Plant* **10**, 520–522 (2017).
5. Gilles, L. M. et al. Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. *EMBO J.* **36**, e201796603 (2017).
6. Hu, H. et al. The genetic basis of haploid induction in maize identified with a novel genome-wide association method. *Genetics* **202**, 1267–1276 (2016).
7. Coe, E. H. A line of maize with high haploid frequency. *Am. Nat.* **93**, 381–382 (1959).
8. Lashermes, P., Gaillard, A. & Beckert, M. Gynogenetic haploid plants analysis for agronomic and enzymatic markers in maize (*Zea mays* L.). *Theor. Appl. Genet.* **76**, 570–572 (1988).
9. Dong, X. et al. Fine mapping of qhir1 influencing in vivo haploid induction in maize. *Theor. Appl. Genet.* **126**, 1713–1720 (2013).
10. Jackson, D. No sex please, we're (in) breeding. *EMBO J.* **36**, 703–704 (2017).
11. Liu, C. et al. Fine mapping of qhir8 affecting in vivo haploid induction in maize. *Theor. Appl. Genet.* **128**, 2507–2515 (2015).
12. Yao, L. et al. OsMATL mutation induces haploid seed formation in indica rice. *Nat. Plants* **4**, 530–533 (2018).
13. Xing, H. et al. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* **14**, 327 (2014).
14. Yang, Q., Zhang, D. & Xu, M. A sequential quantitative trait locus fine mapping strategy using recombinant-derived progeny. *J. Integr. Plant Biol.* **54**, 228–237 (2012).
15. Xu, X. et al. Gametophytic and zygotic selection leads to segregation distortion through in vivo induction of a maternal haploid in maize. *J. Exp. Bot.* **64**, 1083–1096 (2013).
16. Jefferson, R. A., Kavanagh, T. A. & Bevan, M. W. GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907 (1987).
17. Widholm, J. M. The use of fluorescein diacetate and phenosafranine for determining viability of cultured plant cells. *Stain Technol.* **47**, 189–194 (1972).
18. Herrero, M. P. & Johnson, R. R. High temperature stress and pollen viability of maize. *Crop Sci.* **20**, 796–800 (1980).

Acknowledgements

We acknowledge support from the National Key Research and Development Program of China—Maize heterosis utilization technology and strong heterosis hybrids breeding (2016YFD0101200, 2016YFD0101003 and 2018YFD0100201)—and the Modern Maize Industry Technology System (CARS–02–04).

Author contribution

S.C., Y.Z. and C.L. conceived and designed the experiments. Y.Z., C.L., X.Q. and Y.J. performed the experiments. Y.Z., S.C., C.L. and X.Q. analysed the data. C.L., S.C., Y.Z., D.W., Y.W., Z.L., C.C., B.C. X.T., J.Li, M.C., X.D., X.X., L.L., W.Li., W.Liu., W.J. and J.Lai. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41477-019-0443-7.

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

Correspondence and requests for materials should be addressed to S.C.

Peer review information: *Nature Plants* thanks Jose Seguí-Simarro and other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

nature research

Corresponding author(s): Shaojiang Chen

Last updated by author(s): May 9, 2019

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	Becton Dickinson FACSCalibur system
Data analysis	R 3.1.4, Excel 2016, GraphPad Prism 8, Becton Dickinson CellQuest Pro

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that all data supporting the findings of this study are available within the paper and any raw data can be obtained upon request to the corresponding author.

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

Life sciences study design

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

Sample size	no sample size calculation was performed. On average, the improvement of HIR is ~2-4%, we need enough sample size to calculate significance level between wildtype and knockout lines, so that is the standard for sample size.
Data exclusions	No data was excluded
Replication	all attempts at replication were successful
Randomization	Plants of different genetic backgrounds were randomly positioned to mitigate potential variables, which include the availability of water, light and air flow.
Blinding	Data collection, such as the measurements of haploid induction rate were conducted by a third person who was blinded to sample identities.

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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	young leaves were sampled in field without frozen. Then chopped with razor. lysate was used to release nuclears. subsquently, centrifuge with 1000rpm was conducted to collect nuclears. all nuclears were dyed for 30 min in 4℃ environment, and used for determination with flow cytrometry
Instrument	Becton Dickinson FACSCalibur system
Software	Becton Dickinson CellQuest Pro
Cell population abundance	by centrifuge
Gating strategy	bulk cell population was used in flow cytometry experimentbulk cell population was used in flow cytometry experiment

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.