



OPEN ACCESS

EDITED BY

Hakan Ozkan,
Çukurova University, Türkiye

REVIEWED BY

Zhang Hongwei,
Chinese Academy of Agricultural Sciences,
China
Chen Chen,
Beijing Academy of Agriculture and
Forestry Sciences, China
Manje S. Gowda,
The International Maize and Wheat
Improvement Center (CIMMYT), Kenya
Haochuan Li,
Henan Agricultural University, China

*CORRESPONDENCE

Thomas Lübberstedt
✉ thomasl@iastate.edu

†These authors have contributed equally to
this work

RECEIVED 06 May 2023

ACCEPTED 05 September 2023

PUBLISHED 04 October 2023

CITATION

Trentin HU, Krause MD, Zunjare RU,
Almeida VC, Peterlini E, Rotarencu V,
Frei UK, Beavis WD and Lübberstedt T
(2023) Genetic basis of maize maternal
haploid induction
beyond *MATRILINEAL* and *ZmDMP*.
Front. Plant Sci. 14:1218042.
doi: 10.3389/fpls.2023.1218042

COPYRIGHT

© 2023 Trentin, Krause, Zunjare, Almeida,
Peterlini, Rotarencu, Frei, Beavis and
Lübberstedt. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Genetic basis of maize maternal haploid induction beyond *MATRILINEAL* and *ZmDMP*

Henrique Uliana Trentin^{1,2†}, Matheus Dalsente Krause^{2†},
Rajkumar Uttamrao Zunjare^{2,3}, Vinícius Costa Almeida^{2,4},
Edicarlo Peterlini^{2,5}, Valeriy Rotarencu⁶, Ursula Karoline Frei²,
William Dale Beavis² and Thomas Lübberstedt^{2*}

¹Bayer Crop Science, Coxilha, RS, Brazil, ²Department of Agronomy, Iowa State University, Ames, IA, United States, ³Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India, ⁴Federal University of Viçosa, Viçosa, MG, Brazil, ⁵Department of Agronomy, State University of Maringá, Maringá, PR, Brazil, ⁶VAR BREEDING Ltd., Chisinau, Moldova

In maize, doubled haploid (DH) lines are created *in vivo* through crosses with maternal haploid inducers. Their induction ability, usually expressed as haploid induction rate (HIR), is known to be under polygenic control. Although two major genes (*MTL* and *ZmDMP*) affecting this trait were recently described, many others remain unknown. To identify them, we designed and performed a SNP based (~9007) genome-wide association study using a large and diverse panel of 159 maternal haploid inducers. Our analyses identified a major gene near *MTL*, which is present in all inducers and necessary to disrupt haploid induction. We also found a significant quantitative trait loci (QTL) on chromosome 10 using a case-control mapping approach, in which 793 noninducers were used as controls. This QTL harbors a *kokopelli* ortholog, whose role in maternal haploid induction was recently described in *Arabidopsis*. QTL with smaller effects were identified on six of the ten maize chromosomes, confirming the polygenic nature of this trait. These QTL could be incorporated into inducer breeding programs through marker-assisted selection approaches. Further improving HIR is important to reduce the cost of DH line production.

KEYWORDS

maize maternal haploid induction rate, maize haploid inducers, genome-wide association study, maize doubled haploid technique, single fertilization

Introduction

Today's maize maternal haploid inducers have substantially higher induction rates than paternal haploid inducers, and are, therefore, the preferred tool for DH line production (Gilles et al., 2017a). The induction ability of an inducer is estimated by the rate of seeds with haploid embryo over the total number of seeds produced in cross-pollinations with a donor genotype. This rate is referred to as haploid induction rate (HIR)

and is usually expressed as a percentage. In paternal inducers, where HIRs of up to 6% were observed (Kindiger and Hamann, 1993), haploid induction is attributed to a single gene, called *indeterminate gametophyte1* (*ig1*) (Kermicle, 1969). In maternal inducers, where breeding efforts raised HIRs from 3.2% to 14.5% (Coe, 1959; Rotarenco et al., 2010), HIR is under polygenic control (Lashermes and Beckert, 1988; Deimling et al., 1997; Röber, 1999; Prigge et al., 2012a; Prigge et al., 2012b; Liu et al., 2015). The first study designed to investigate the genetic nature of maternal haploid induction was performed by Lashermes and Beckert (1988), who observed variation in the induction ability of different segregating populations derived from crosses between inducer Stock 6 (Coe, 1959) and various noninducer lines.

Two quantitative trait loci (QTL), located on chromosomes 1 and 2, were identified in a F₃ population derived from Stock 6, jointly explaining 17.9% of the phenotypic variance (Deimling et al., 1997). Studies conducted by Röber (1999) and Barret et al. (2008) confirmed the presence of a major QTL on chromosome 1 and reported strong segregation distortion against the inducer allele at this locus. A QTL mapping study conducted by Prigge et al. (2012b), which used multiple filial generations derived from four different bi-parental crosses, located this major QTL (*qhir1*) to bin 1.04 and found additional QTL on chromosomes 3 (*qhir2* and *qhir3*), 4 (*qhir4*), 5 (*qhir5* and *qhir6*), 7 (*qhir7*) and 9 (*qhir8*). *qhir8* explained more than 20% of the genetic variance in all filial generations of the cross between inducers CAUHOI and UH400 (Prigge et al., 2012b). Dong et al. (2013) subsequently fine mapped *qhir1* to a 243 kb region and observed large variation in HIR among F_{2,3} progeny homozygous for this locus. Their findings are in agreement with observations made by Prigge et al. (2012b), who noted that while *qhir1* is required for the formation of seeds with haploid embryos, multiple segregating alleles affect HIR. A conditional haplotype extension (CHE) test designed to detect selective sweeps under the assumption of confounding trait expression and population structure, identified two distinct loci (*qhir11* and *qhir12*) within *qhir1* (Hu et al., 2016). Subsequently, Nair et al. (2017) demonstrated the importance of *qhir11* only in haploid induction by genetically delineating the *qhir1* locus.

Subsequently independent researchers discovered that a frameshift mutation in a gene underlying *qhir11*, and coding for a patatin-like phospholipase, triggers the formation of haploid seeds (Gilles et al., 2017b; Kelliher et al., 2017; Liu et al., 2017). This frameshift mutation is caused by a 4-bp insertion in the last exon of gene Zm00001d029412, which leads to an early stop codon and results in a truncated protein, losing its proper localization at the pollen endo-plasma membrane (Gilles et al., 2021). The almost simultaneous discovery of the function of this gene by different researchers resulted in different names: *MATRILINEAL* (*MTL*) (Kelliher et al., 2017), *Zea mays Phospholipase A1* (*ZmPLA1*) (Liu et al., 2017) and *NOT LIKE DAD* (*NLD*) (Gilles et al., 2017b). By inducing mutations close to the 4-bp insertion site in a noninducer, Kelliher et al. (2017) observed HIRs ranging from 4.0% to 12.5% in T1 and derived plants. This suggest that mutations in *MTL* are capable to generate inducers with high HIRs. However, the same 4-bp insertion was found in all inducers sequenced by Kelliher et al. (2017); Liu et al. (2017) and Gilles et al. (2017b), which have HIRs

ranging from 2.0% to 10.0%. This is consistent with the presumed polygenic control of HIR (Lashermes and Beckert, 1988; Deimling et al., 1997; Röber, 1999; Prigge et al., 2012b; Liu et al., 2015). Zhong et al. (2019) identified Zm00001d044822, which encodes for a DUF679 domain membrane protein, to be the underlying cause of *qhir8*. The authors named this gene *ZmDMP* and confirmed its ability to enhance the HIR of inducers possessing *mtl/zmpla1/nld*, as previously reported by Liu et al. (2015). They also observed that in genotypes homozygous for the wild type *MTL/ZmPLA1/NLD* allele, knockout mutations in *ZmDMP* can disrupt haploid induction at very low frequencies (0.1-0.3%). This was an interesting observation because it indicated that mutations in different genes can promote and/or have synergetic effect on maternal haploid induction (Jacquier et al., 2020 and Jacquier et al., 2021).

The savings from employing inducers with higher HIRs can be substantial and are important to breeding programs where haploid selection is visually performed through the *R1-nj* color marker (Chase and Nanda, 1965). This dominant allele leads to anthocyanin production in the scutellum and aleurone layer of seeds in which double fertilization and embryonic development occurred normally. However, since embryos of haploid seeds do not contain inducer chromosomes, they also lack anthocyanin pigmentation. This difference in embryo pigmentation allows differentiation of haploid and diploid seeds. The savings from employing more efficient inducers can be significant, since the number of donor seeds that will be planted, and whose plants would be detasseled, harvested, dried, and screened for haploid seeds directly depends on the inducers' HIR and the number of DH lines that one wants to extract from that donor population. To gain better understanding of the genetic basis of HIR, we designed a genome-wide association study (GWAS) using a diverse set of haploid inducers adapted to Midwestern U.S. conditions. The objectives of this study were to (i) identify promising maternal inducers to be used as parents of new inducer breeding populations, (ii) locate and assess the effect of QTL affecting HIR, and (iii) gain better understanding of the biological processes underlying haploid induction by identifying candidate genes within the detected QTL.

Materials and methods

Germplasm

We used a diverse panel of North American and European haploid inducers to identify QTL affecting HIR. The North American panel contains inducers in the background of public and Ex-PVP lines, such as A637, B73, DK78004, LH82, Mo17, PHG50 and Va35 (Table S1). All lines from the North American panel have the cross between RWS (Röber et al., 2005) and RWK (Röber et al., 2005) as their source of *mtl/zmpla1/nld*. The European panel contains lines such as MHI (Chalyk, 1999), PHI-3 (Rotarenco et al., 2010), RWS, RWK and eight F₇ progeny derived from the cross of RWS and PHI-3, which were selected for high HIR. In total, 159 inducers were evaluated for HIR. The HIR of all inducers was evaluated in crosses with the commercial hybrid Viking 60-01N,

from Albert Lea Seed Company (Minnesota, US). This hybrid was chosen as a donor because it possesses good inducibility and allows clear expression of the *RI-nj* marker. Inducibility is the ability of the donor parent to generate haploid seeds, and multiple studies indicated that the source germplasm impacts HIR (Randolph, 1940; Chase, 1952; Lashermes and Beckert, 1988; Eder and Chalyk, 2002; De La Fuente et al., 2018; Nair et al., 2020; Ren et al., 2022; Trentin et al., 2022).

A panel of 793 non-inducer genotypes was considered for a case-control GWAS (cc-GWAS) (Table S1). This panel was composed mainly of DH or highly homozygous lines derived from various genetic backgrounds, such as B73, BS39 (Verzegnazzi et al., 2021; Santos et al., 2022), several cycles of the Iowa Stiff Stalk Synthetic Population (BSSS) (Ledesma et al., in preparation), BGEM (Iowa State University/USDA-ARS Germplasm Enhancement Maize project) and Ex-PVP lines. No haploid plants were observed when these genotypes were grown in the field, which is in accordance with Chase (1947) observation that the occurrence of haploid plants in elite inbred lines happens at very low frequencies (<0.1%).

Field plot design

The inducers and donor used for this experiment were sown at the Iowa State University, Agronomy and Agricultural Engineering Farm, located in Boone (Iowa), during the summer of 2018. The trial was grown under rainfed conditions, following the recommended practices for maize production in Central Iowa. Pre- and post-emergence herbicides along with manual weeding were used to control invasive species. Urea ammonium nitrate was applied in the area before sowing. Two blocks of inducers and donor were sown at different planting dates to ensure that enough seeds would be generated to obtain reliable estimates of HIR. The first block of donor seeds was sown on May 8th, and the first block of inducers seeds was sown on May 21st. The second block of donor seeds was sown along with the first block of inducers seeds, while the second block of inducers seeds was sown on May 31st. Sowing of inducer blocks was delayed because most inducers have a significantly earlier maturity than the donor.

With the delayed planting of the haploid inducers, it was possible to pollinate the first donor block exclusively with the first inducer block, and the second donor block exclusively with the second inducer block. Each inducer and donor block was composed of subblocks containing 16 plots. Inducers were not randomized among subblocks because the great difference in vigor among them would adversely affect other traits for which data were collected for a companion study (Trentin et al., 2023). This rationale also justified sowing closely related inducers side-by-side within each subblock. For instance, if hybrid and inbred inducers were randomized, the shading caused by differences in plant height would be detrimental to the growth and development of inbred inducers. Plots were 5.5 meters long, 0.75 meters wide, and were sown with 25 seeds. Inducer and donor blocks were sown side-by-side, and pollen from inducers in a given subblock was carried to the adjacent donor subblock. Multicolored tags with easy-to-match

codes were used to ensure that pollen from each inducer plot was placed in the corresponding donor plot. Bulk pollen was collected in tassel bags and used to pollinate at least 10 ears of the donor, which were covered before silk emergence using wax bags. Ears were harvested when seeds reached the black layer stage and were air-dried for one week.

Data collection

Visual haploid selection was performed using the *RI-nj* marker, and the number of putative haploid and diploid seeds of each ear was recorded. Embryo and endosperm abortion, which are correlated with HIR (Prigge et al., 2012b; Xu et al., 2013; Nair et al., 2017), occur at different stages of seed development (Xu et al., 2013). This makes the identification of embryo and endosperm aborted seeds quite subjective, and for this reason, these two classes of seeds were not considered as a category in the statistical analysis of the data.

Ploidy determination of the putative haploid seeds

Putative haploid seeds produced by each inducer at each planting date were bulked in a single envelope, whereas putative diploid seeds were discarded. The ploidy of the putative haploid seeds was verified by cutting seeds in half and observing the presence of anthocyanin pigmentation in the embryonic region. With the true number of haploid seeds, a weighted correction factor (CF_i) for the i^{th} inducer was calculated as follows:

$$CF_i = \frac{n_1 \frac{TH_1}{PH_1} + n_2 \frac{TH_2}{PH_2}}{n_1 + n_2}$$

Where the indices 1 and 2 stand for the first and second planting dates, respectively, TH is the true number of haploid seeds, PH is the putative number of haploid seeds, and n is the number of harvested ears. The CF_i was then used to correct the number of putative haploid seeds for the i^{th} inducer in each planting date, where the difference between the counts before and after the correction was re-classified as diploid seeds.

Statistical analysis of phenotypic data

Two modeling strategies were implemented to analyze and understand the variability of the phenotypic data by considering planting dates as blocks (Silva, 2017; Couto et al., 2019). In both cases, the corrected number of haploid seeds for the i^{th} inducer, in the j^{th} planting date, at the k^{th} harvested ear, was modeled as $Y_{ijk} \sim \text{Binomial}(m_{ijk}, \pi_{ijk})$, where Y_{ijk} , m_{ijk} , and π_{ijk} are the number of haploid seeds, the total number of seeds, and the probability of successful production of haploid seeds, respectively.

The first strategy consisted of fitting a generalized linear model (GLM) with a *logit* link function, expressed:

$$\text{logit}(\pi_{ijk}) = \log\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right) = \mu + \beta_j + \tau_i \quad (\text{Model 1})$$

Where μ is the intercept, β_j is the fixed effect of the j^{th} planting date, and τ_i is the fixed effect of the i^{th} inducer. The dispersion parameter ϕ from Model 1 was estimated by:

$$\phi = \frac{\sum_{k=1}^n r_k^2}{n - p}$$

Where r_k is the Pearson residual of the k^{th} observation, and $n-p$ is the degree of freedom for the deviance residuals. If Pearson's Chi-Square Statistic suggests a lack of fit from Model 1, then a Quasi-Likelihood inferential approach that accounts for overdispersion was assumed.

The second strategy consisted of fitting a generalized linear mixed model (GLMM) with a *logit* link function, expressed by the following linear predictor:

$$\text{logit}(\pi_{ijk}) = \log\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right) = \mu + \beta_j + \tau_i + \epsilon_{ij} \quad (\text{Model 2})$$

Where μ is the intercept, β_j is the random effect of the j^{th} planting date with $\beta_j \sim N(0, \sigma_\beta^2)$, τ_i is the fixed effect of the i^{th} inducer, and ϵ_{ij} is the random effect of the plot with $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. The following two models account for the random effects of individual observations ($\omega_{k(ij)}$, i.e., harvested ears) and both individual observations and plot variation, respectively, as follows:

$$\text{logit}(\pi_{ijk}) = \log\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right) = \mu + \beta_j + \tau_i + \omega_{k(ij)} \quad (\text{Model 3})$$

$$\text{logit}(\pi_{ijk}) = \log\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right) = \mu + \beta_j + \tau_i + \epsilon_{ij} + \omega_{k(ij)} \quad (\text{Model 4})$$

Where $\omega_{k(ij)} \sim N(0, \sigma_\omega^2)$. All the other model terms were previously defined.

Three criteria were used to select the best-fit model: (a) the goodness-of-fit via the Akaike Information Criterion (AIC; Akaike, 1974), (b) the Bayesian Information Criterion (BIC; Schwarz, 1978); and (c) visual inspection using half-normal plots with a simulated envelope (Demétrio et al., 2014) using the R package *hnp* (Moral et al., 2017). Plots of the Pearson residuals were also compared between models. Maximum likelihood estimates for the fixed effects of inducers ($\hat{Y}_{i..}$) from the selected model were further used for GWAS analysis. Asymptotic confidence intervals were obtained using the package *emmeans* (Lenth, 2020). All statistical analyzes were carried out using the *lme4* package (Bates et al., 2015) in the R environment (R Core Team, 2022).

Genotyping and quality control

Leaf samples were collected when seedlings were at the V3 stage. After lyophilization, they were sent to the International Maize and Wheat Improvement Center (CIMMYT), where they were

genotyped with the Diversity Arrays Technology (DARtSeq) platform (Kilian et al., 2012). SNP calling was performed using the DARtsoft analytical pipeline (<https://www.diversityarrays.com>) and the version 4 of B73's reference genome (AGPv4) as a reference. A total of 32,929 SNP markers were initially identified, and 9,007 remained after filtering for a call rate of at least 70% and minor allele frequency (MAF) of 1% (<https://dr.lib.iastate.edu/home>). The software Beagle 5.0 (Browning et al., 2018) was subsequently used for the imputation of missing data.

Genome-wide association analyses

We performed marker-trait association using quantitative and binary response variables. The quantitative GWAS was implemented in GAPIT (Lipka et al., 2012) with the multiple loci models (i) BLINK (Huang et al., 2019), (ii) FarmCPU (Liu et al., 2016), and (iii) MLM (Segura et al., 2012). The estimated HIR from the best-fit phenotypic model was normalized (i.e., transformed) using the *bestNormalize* package in R (Peterson, 2021). The binary GWAS was a case-control and included the 159 inducers considered "cases", and the panel of 793 noninducers considered "controls" (Table S1). In total, 952 genotypes were included. The phenotypes, a binary random variable, were coded as ones for the cases and zeros for the controls, and a mixed logistic regression model was implemented in the R package *milorGWAS* (Milet et al., 2020). The reader is referred to the cited references for details on these GWAS models.

For the quantitative GWAS, population/family structure was accounted by the additive genomic relationship G (VanRaden, 2008), and the first seven principal components (PCs) obtained from principal component analysis (PCA, built-in R function *prcomp*) of the numerical matrix of SNP markers, calculated via singular value decomposition (Figures 1A, B). For the cc-GWAS, the cases and controls grouping of genotypes were completely confounded by the first two PCs (Figure 1C), and hence population structure was only modeled by the matrix G. It is unlikely such correction can fully adjust for the population structure; hence, a permutation analysis was performed as follow: the phenotypes (zeros and ones) were shuffled under the null hypothesis of no association between markers and phenotypes, and the lowest *p-value* recorded. This process was repeated 10,000 times, and the 95% quantile of the distribution of the $-\log_{10}(p\text{-values})$ was used as a reference threshold. In addition to the permutation analysis for the cc-GWAS, SNPs were declared statistically significant for both GWAS analyses at an FDR-adjusted *p-value* of less than 0.05.

Lastly, we calculated the genome-wide (i.e., across chromosomes) pairwise linkage disequilibrium (LD) using the function *LD.decay* from the *sommer* package (Covarrubias-Pazarán, 2016). The LD decay was determined by locally-weighted polynomial regression (LOESS, built-in R function) with the physical distance between markers as a function of the coefficient of determination (r^2). The threshold r^2 value when estimating LD decay was calculated with the 95th percentile of the LD distribution between unlinked markers [*LD.decay*(..., *unlinked* = *TRUE*, *gamma* = 0.95)]. The search browser in MaizeGDB (<https://>

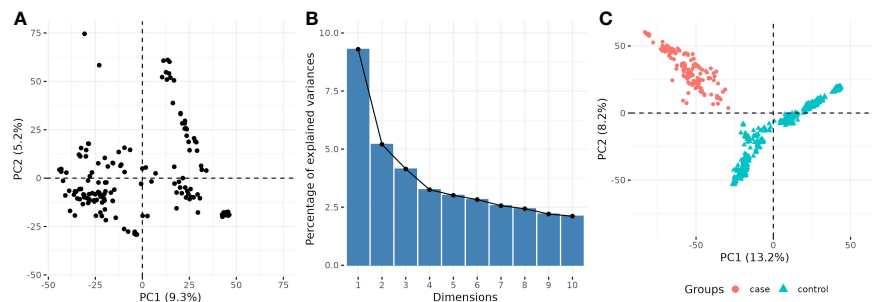


FIGURE 1

Scatterplot of two first principal components (PCs) from principal component analysis (PCA) of SNP matrix for (A) quantitative and (C) binary GWAS. The percentage of explained variance for each PC in the quantitative GWAS is shown in (B).

www.maizegdb.org/gbrowse) was used to identify the putative candidate genes in the interval of significant SNPs. Maize B73 RefGen_v4 version was used to locate the candidate genes.

Results

Phenotypic analyses

Estimates of HIR with asymptotic confidence intervals were obtained with Model 4 (Figure 2; Table S2). For the selection of the best-fit phenotypic model, the AIC values ranged from 14702.24 (Model 4) to 16932.52 (Model 1), and for the BIC criterion, from 15668.51 (Model 4) to 17886.93 (Model 1). In addition, Model 1 was adjusted with the Quasi-likelihood inference due to overdispersion (Table 1). Thus, both selection criteria suggested that Model 4, which included random effects for planting dates, plots, and individual observations (i.e., harvested ears), was the best-fit model for the observed data. The visual inspection of the half-normal and Pearson residuals plots also showed the superior fit of Model 4 (Figures S1 and S2). Prior to GWAS analyses, the estimated HIR values were transformed to ensure normality (Figure 3).

Genome-wide association analyses

The average rate of LD decay at the r^2 estimated threshold of 0.096 was of 113 Kb (Figure 4). The reference threshold obtained from the permutation analysis for the cc-GWAS was of 4.89,

whereas the $-\log_{10}(p\text{-value})$ of the statistically significant SNP marker on chromosome 10 was of 5.30, with an adjusted FDR value of 0.04 (Figure 5D). In the quantitative GWAS, significant SNPs were detected on chromosomes 1, 2, 3, 7 and 8 (Figures 5A–C). The significant SNPs detected in chromosome 1 were located in positions 66.6 Mb (S1_66636144), 69.3 Mb (S1_69321282) and 76.1 Mb (S1_76160596), flanking the region of *MTL/ZmPLA1/NLD*, which is located in position 69.4 Mb. The significant SNP detected on chromosome 2 (S2_220376487) is located at position 220.3 Mb and does not overlap with the QTL detected by Deimling et al. (1997) on the same chromosome. The significant SNP detected on chromosome 3 (S3_13701123) is located at position 13.7 Mb and does not overlap with any of chromosome 3 QTL's previously identified by Prigge et al. (2012b). The significant SNP detected on chromosome 7 (S7_131226237) is located at position 131.2 Mb, and also does not overlap with the chromosome 7 QTL previously identified by Prigge et al. (2012b). The two significant SNP detected on chromosome 8 (S8_138779408 and S8_174792234) are located at positions 138.7 and 174.7 Mb, respectively. There are no reports of HIR QTL on this chromosome in the literature. The cc-GWAS detected a significant SNP (S10_141729953) at position 141.7 Mb on chromosome 10. To the best of our knowledge, there are no reports of HIR QTL on chromosome 10 in the literature. A summary overview of the statistically significant SNPs is given in Table 2, and the allelic distribution as a function of the estimated HIR in Figure 6. For the binary GWAS, a classification table differentiating the number of cases and controls according to the genotyped allele is presented (Table 3).

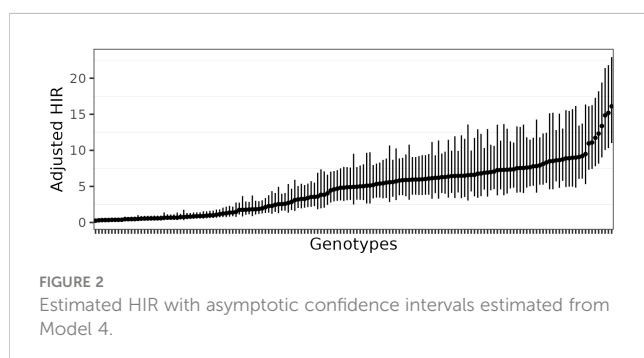


FIGURE 2

Estimated HIR with asymptotic confidence intervals estimated from Model 4.

Identification and annotation of candidate genes

The physical locations of the nine significant SNPs identified were recorded using the B73 RefGen_v4 based on the LD decay distance. The candidate genes were identified by the genome browser search available at Maizegdb (<https://www.maizegdb.org/gbrowse>). A total of 41 candidate genes with descriptions were found, out of which, 37 were protein-coding, and the remaining 4 were not annotated. The chromosomal distribution of these candidate genes included 12 located on chromosome 1, 8 on

TABLE 1 Goodness-of-fit for Models 1 to 4. Model 4 is the best-fit model.

Model	Log-likelihood	AIC	BIC	Variance components			Dispersion parameter*
				Block	Plot	Ears	
1	-8305.26	16932.52	17886.93	-	-	-	3.05
2	-7889.48	16102.95	17063.30	0.01	0.10	-	-
3	-7257.02	14838.03	15798.38	0.01	-	0.26	-
4	-7188.12	14702.24	15668.51	0.01	0.08	0.20	-

*Model fitted with Quasi-likelihood approach, "-" symbol indicates no data.

chromosome 2, 5 on chromosome 3, 5 on chromosome 7, 3 on chromosome 8, and 8 on chromosome 10. The functional domain of each of the candidate genes is provided in Table S3.

Discussion

The widespread use of the DH technique in maize was influenced by the increase of HIRs to manageable levels and the addition of anthocyanin-based phenotypic markers in maternal inducers. These two fields have recently seen remarkable advancements. For instance, Chen et al. (2022) recently developed an efficient and accurate (99.1% accuracy) haploid selection marker through co-expression of two transcription factor genes (*ZmC1* and *ZmR2*) in novel haploid inducer line called 'Maize Anthocyanin Gene Inducer 1 (MAGIC1)'. Shortly after, Wang et al. (2023) showed that expression of *RUBY* enabled 100% accuracy of haploid identification based on deep betalain pigmentation. The development of a group of inducers with numerous anthocyanin markers and HIRs above 12.0% allowed for effective haploid generation and screening (Rotarengo et al., 2010). The identification of the *MTL/ZmPLA1/NLD* (Gilles et al., 2017b; Kelliher et al., 2017; Liu et al., 2017) *DUF679* (Zhong et al., 2019), *ZmPLD3* (Li et al., 2021) and *ZmPOD65* (Jiang et al., 2022) genes contributed to a better understanding of the molecular mechanisms of haploid induction (Jacquier et al., 2020). Previous research (Lashermes and Beckert, 1988; Deimling et al., 1997; Röber, 1999; Prigge et al., 2012b; Liu et al., 2015) revealed evidence for the polygenic regulation of HIR and identified

various QTL controlling this trait. To increase the effectiveness of inducer breeding and, as a result, lower the cost of DH line production, it is critical to identify and validate the influence of QTL affecting HIR. To achieve this goal, we designed and performed a GWAS with a large and diverse panel of inducers. In a previous GWAS for HIR, the induction ability of 53 inducers was evaluated in distinct donors (Hu et al., 2016). Comparing the induction ability obtained through crosses with different donors is not ideal, since inducibility and HIR effects might be confounded (Randolph, 1940; Chase, 1952; Lashermes and Beckert, 1988; Eder and Chalyk, 2002; De La Fuente et al., 2018). In our study, 159 inducers were crossed to the same donor at two different planting dates.

Genome-wide association analyses

To this date, just a few studies were performed to identify QTL affecting the HIR of maize maternal haploid inducers (Deimling et al., 1997; Barret et al., 2008; Prigge et al., 2012b; Hu et al., 2016). However, multiple studies have been performed to fine-map or validate the function of previously identified QTL (Dong et al., 2013; Liu et al., 2015; Gilles et al., 2017b; Kelliher et al., 2017; Liu et al., 2017; Nair et al., 2017; Zhong et al., 2019; Li et al., 2021).

In this study, we used four different GWAS methods, including BLINK, FarmCPU, MLMM, and cc-GWAS, to identify potential

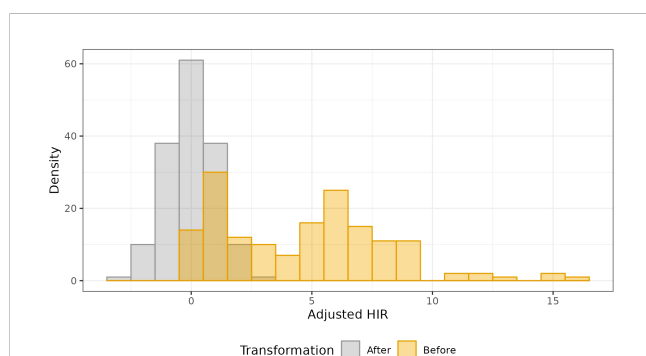


FIGURE 3 Data transformation of the estimated HIR from Model 4 to perform the quantitative GWAS.

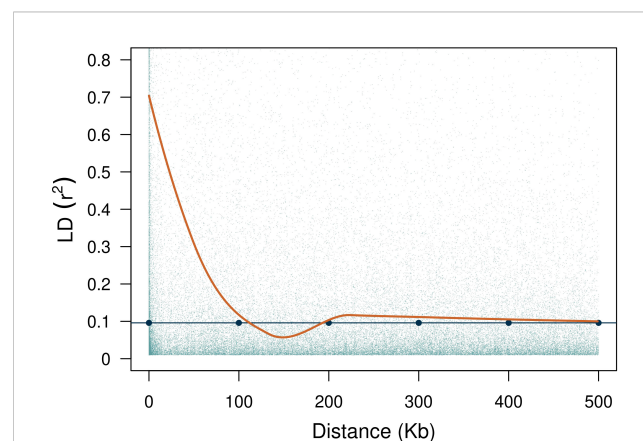


FIGURE 4 LD decay with an estimated threshold value of 0.096. The fitted LOESS curve is shown in orange.

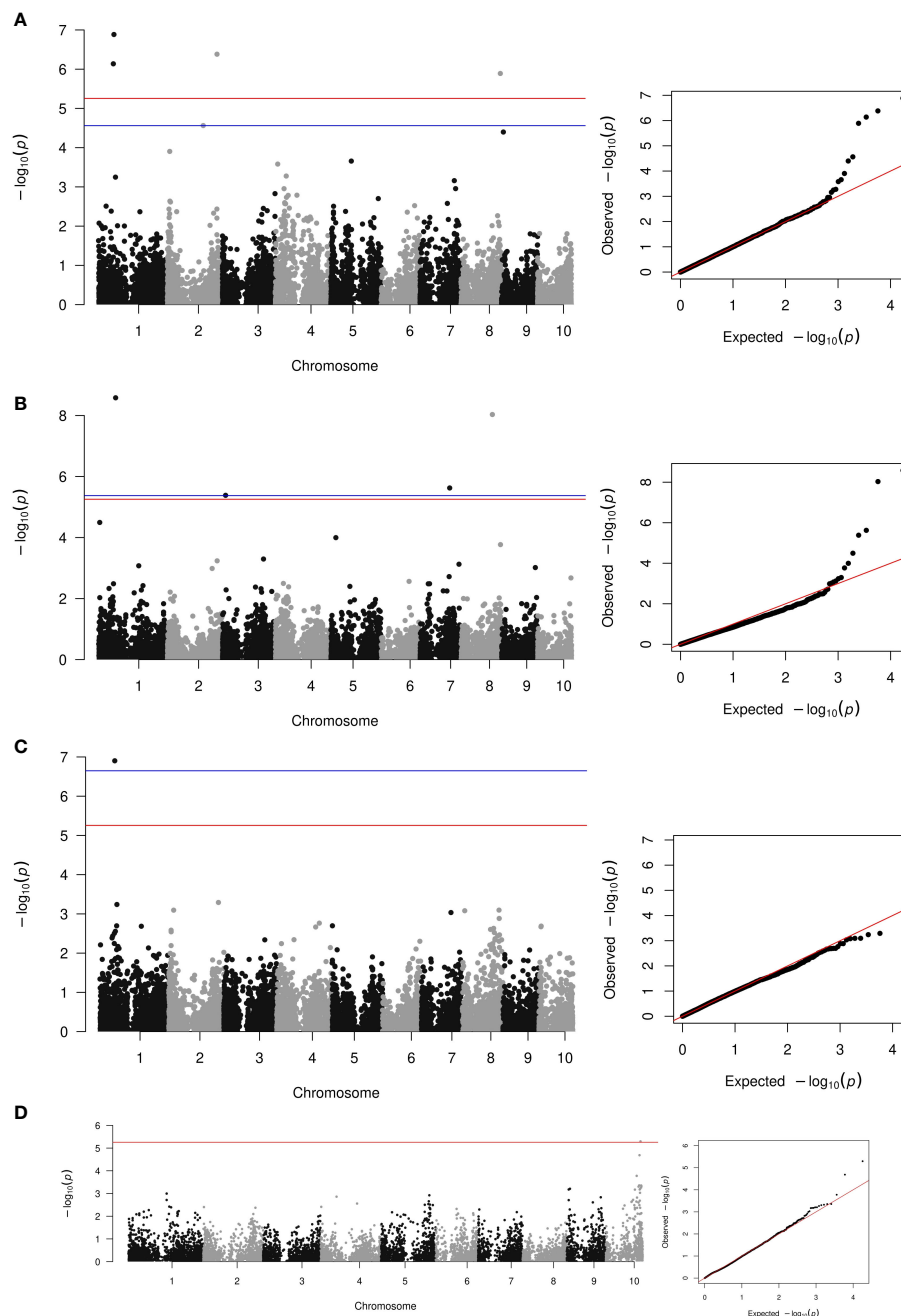


FIGURE 5
Manhattan and qq-plot of the quantitative GWAS using Blink (A), FarmCPU (B), MLMM (C), and the case-control GWAS (D).

associations with single nucleotide polymorphisms (SNPs). We identified a total of nine SNPs that were significantly associated with the trait. They correspond to different chromosomal origins, validating the polygenic control of HIR in maize, as previously reported (Lashermes and Beckert, 1988; Deimling et al., 1997; Röber, 1999; Prigge et al., 2012b; Liu et al., 2015). For instance, two SNPs (S8_138779408 and S8_174792234) were identified on chromosome-8 and a SNP (S10_14172995) chromosome-10. No QTLs were previously reported on these chromosomes. However, some of the significant SNPs that we identified are closely linked to the previously discovered QTLs for HIR. The marker S1_69321282,

at position 69.3 Mb on chromosome 1, is closely linked to the *MTL/ZmPLA1/NLD* gene at position 69.4 Mb (identified in the region *qhir11*). Similarly, marker S1_76160596 (position 76.1 Mb), is located inside the *qhir12* region which Hu et al. (2016) found a higher CHE score for *qhir12* than for *qhir11*. Interestingly, Hu et al. (2016) also reported that the *qhir12* haplotype was exclusively found in inducers, whereas the *qhir11* haplotype could be found in 2.7% of the non-inducers. Nair et al. (2017) showed the significance of *qhir11* region compared *qhir12* for inducing haploid embryo in maize. At the same time, it was discovered that *qhir11* harbours the *MTL/ZmPLA1/NLD* allele that possesses

TABLE 2 Summary of the statistically significant SNPs containing name, chromosome, position in base pairs, alleles, minor allele frequency (MAF), and models' name. C-C stands for cc-GWAS.

SNP marker	Chromosome	Position	Alleles	MAF ^a	BLINK	FarmCPU	MLMM	C-C
S1_66636144	1	66636144	A/G	0.21	x		x	
S1_69321282	1	69321282	C/G	0.13	x			
S1_76160596	1	76160596	G/T	0.22		x		
S2_220376487	2	220376487	A/G	0.18	x			
S3_13701123	3	13701123	A/T	0.08		x		
S7_131226237	7	31226237	C/T	0.03		x		
S8_138779408	8	138779408	A/G	0.08		x		
S8_174792234	8	74792234	A/C	0.06	x			
S10_141729953	10	141729953	G/T	0.13				x

^aMAF was calculated respective to the population used (i.e., 159 or 952 genotypes).

major effect on haploid induction in maize. The differences in germplasm, genotyping method, statistical models and sample size might explain the divergence between earlier studies and our results. The inbred panel used in this study essentially carried the mutant *MTL/ZmPLA1/NLD* allele and therefore it is obvious that the gene would not be detected in our quantitative GWAS.

We also found two significant markers, S2_220376487 at 220.3 Mb and S7_131226237 at 132.2 Mb on chromosome 2 and chromosome 7, respectively, in regions that does not overlap QTL identified on chromosome 2 (Deimling et al., 1997) and chromosome 7 (Prigge et al., 2012b). Although on the same chromosome, we report these new genomic regions that could influence haploid induction. PVE (phenotypic variance explained) is an important metric used in GWAS to quantify the contribution of SNP for trait value and indicator of the biological relevance of the SNP. PVE for the significant SNPs varied from 5.9% to 37% explaining substantial variance for HIR detected in our study. SNPs S1_66636144, located at position 66.6 Mb on chromosome

1 (linked to *MTL/ZmPLA1/NLD* gene), S8_174792234, located at position 174.7 Mb on chromosome 8 and S7_131226237, located at position 131.2 Mb on chromosome 7, with higher PVE values of 37%, 25% and 21%, respectively, are likely to be more important for the trait and are more likely to have a direct impact on the HIR. Notably, detection of S1_66636144 at position 66.6 Mb on chromosome 1 by BLINK and MLMM models makes it a potential region for investigation, in addition to SNP S8_174792234, at position 174.7 Mb on chromosome 8 which is reported first time for the HIR. A smaller p-value suggests that the association is less likely to be due to chance and more likely to be real. Apart from S1_66636144 at 66.6 Mb on chromosome 1, the marker S8_174792234 at position 174.7 Mb on chromosome 8 possessed the lowest p-value followed by S1_69321282 on chromosome 1 at 69.3 Mb (tightly linked to *MTL/ZmPLA1/NLD* gene) compared to the two other markers (S1_66636144 and S1_76160596 at 66.6 and 76.1 Mb, respectively) detected on chromosome 1. Given their high significance, it is interesting that

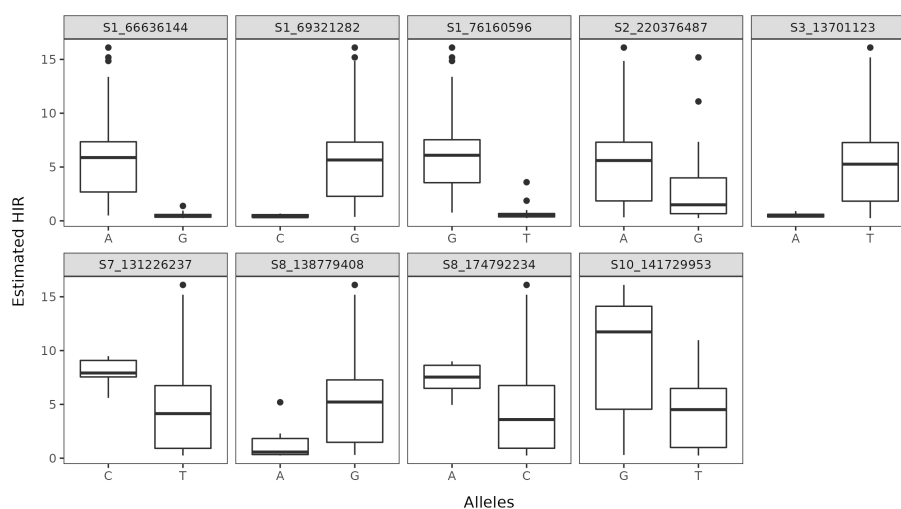


FIGURE 6 Allelic distribution as a function of the estimated HIR at each statistically significant SNP.

TABLE 3 Counts of number of genotypes for SNP marker S10_141729953 from cc-GWAS analysis.

Allele	Cases	Controls
G	11	789
T	146	3

they have not been detected in other mapping studies (Deimling et al., 1997; Barret et al., 2008; Prigge et al., 2012b; Hu et al., 2016). This failure of detection might have been caused by the absence of genetic variation at these QTL in the germplasm used in other studies.

In a cc-GWAS, we compared inducers (cases) to noninducers (controls) to identify genetic variants that are associated with the haploid induction. As a result, we identified a significant novel genetic region, close to marker S10_141729953, which is located at position 141.7 Mb of chromosome 10. Therefore, the identification of S10_141729953 as associated with HIR can help improve our understanding of the genetic factors involved in this trait. The failure of detection of this region might have been due to the absence of genetic variation at these QTL in the germplasm used in other studies (Lashermes and Beckert, 1988; Deimling et al., 1997; Röber, 1999; Prigge et al., 2012b; Liu et al., 2015). For marker S10_141729953, this seems to be a plausible explanation since the G allele was almost exclusively found in MHI and in its progeny. Another explanation might be that their effect is dependent on the presence of the *mtl/zmpla1/nld* allele, similarly to what is described for the *zmdmp* allele. This also seems to be the case for inducers carrying the G allele at marker S10_141729953. From those, the only two which do not have the common inducer haplotype at the *MTL/ZmPLA1/NLD* region (HUT45 and HUT71), display an HIR < 1.0% at all planting dates. The population structure for the cc-GWAS analysis was only controlled with the additive genomic relationship matrix due to the confounding between the cases and controls grouping of genotypes with the first two principal components. Thus, further research is needed to ensure that marker S10_141729953 was not a spurious association.

Candidate genes

We searched the putative candidate genes in the interval of 56kb on both sides of nine significant markers identified in the study. The aldose 1-epimerase enzyme-coding gene Zm00001d029341 has been shown to be expressed in meiotic tassel (Stelpflug et al., 2016). This could be one of the candidate genes for inducing haploid in maize. The gene Zm00001d029558, which encodes for a pathogenesis-related protein 1, contains the marker S1_76160596. Pollen allergen properties of pathogenesis-related proteins have been reported in a variety of species (Huang et al., 1997 and Huang et al., 2016). However, there was no evidence of Zm00001d029558 expression in pollen (Walley et al., 2016). S1_76160596, the same marker, is found 19.9 kb downstream of the gene Zm00001d029559, which encodes an EID1-like F-box protein 2. EID1 (Empfindlicher im Dunkelroten Licht 1) is a F-box

protein that belongs to the SKP1-Cullin1-F-box (SCF) proteomic complexes that target protein degradation (Marrocco et al., 2006). F-box-like protein 17 (FBL17) in Arabidopsis targets the degradation of two CDKA;1 inhibitors, KRP6 and KRP7 (Kim et al., 2008). Arabidopsis *fbl17* mutants occasionally fail to divide and produce pollen grains with a single sperm cell, resulting in single fertilization events (Kim et al., 2008). A-type cyclin-dependent kinase 1 (CDKA;1) is a homolog of the human serine/threonine protein kinase *cdc2*. Arabidopsis CDKA;1 mutation delay pollen generative cell division (Aw et al., 2010), resulting in the delivery of a single sperm cell to the ovary. It is possible that gene Zm00001d029559 interferes with maize's CDKA;1 via the same pathway described in Arabidopsis, influencing HIR by increasing the frequency of single fertilization events, which is one of two processes thought to cause haploid induction.

The marker S10_141729953 is in the intragenic region of gene Zm00001d026242, which was reported to be either exclusively or highly expressed in pollen grains (Davidson et al., 2011; Walley et al., 2016). Zm00001d026242 currently does not have an annotation, but the National Center for Biotechnology Information (NCBI) nucleotide Basic Local Alignment Search Tool (BLAST) shows that its nucleotide sequence is 88.9% and 83.0% identical to the *Sorghum bicolor*'s and maize's *kokopelli* (*KPL*) mRNAs, respectively. *Kokopelli* was first described in *Arabidopsis thaliana*, where it was named after the Native American deity of male fertility (Ron et al., 2010). In maize, *kokopelli* is located on chromosome 2 (NCBI's sequence ID: XP_008671112.1), whereas in sorghum it is located on chromosome 6 (NCBI's sequence ID: XP_021319809.1). *Kokopelli* encodes for a natural cis-antisense siRNA (cis-nat-siRNA), which is a class of small regulatory RNAs found in various eukaryotic species. Even though two apparently normal sperms are delivered to embryos sacs by the pollen tubes of the Arabidopsis *kpl* mutant, single fertilization was observed in approximately 40% of the ovules fertilized (Ron et al., 2010). In Arabidopsis, *KPL* forms a sperm-specific nat-siRNA pair with the inversely transcribed gene *ARIADNE14* (*ARI14*), which encodes for a putative ubiquitin E3 ligase, and whose overexpression decreases seed set (Ron et al., 2010). Protein degradation occurs during gametogenesis in plants and animals and is required for successful fertilization (Dickins et al., 2002; Suzumori et al., 2003; Roest et al., 2004; Ron et al., 2010; Wang et al., 2013). Altogether, these results suggest a possible involvement of Zm00001d026242 in fertilization and haploid induction. Recently, Jacquier et al. (2023) validated the role of *Atkpl* (KOKOPELLI, AT5G63720) mutant for triggering *in planta* maternal haploid induction.

Applicability of the findings to inducer breeding

The detection of candidate genes in six maize chromosomes is in agreement with other studies that suggested a polygenic control of HIR (Lashermes and Beckert, 1988; Deimling et al., 1997; Röber, 1999; Prigge et al., 2012b; Liu et al., 2015; Nair et al., 2017). The presence of genes with strong effects, like *MTL/ZmPLA1/NLD*, *ZmDMP*, *ZmPLD3*, *ZmPOD65* and the QTL identified in this

study on chromosomes 1, 8 and 10, suggest that some form of MAS, such as F_2 -enrichment or marker-assisted backcrossing, might be efficient breeding strategies for inducer development. Employing genomic selection or predicted cross value (Han et al., 2017) during or after the fixation of these major genes, can help to capture the best combination of small effect QTL (Almeida et al., 2020). Phenotypic selection for HIR is very laborious and time-consuming, since multiple cross-pollinations need to be performed to obtain reliable estimates of HIR. Moreover, visual discrimination of haploid and diploid seeds is time-consuming and error-prone, since *R1-nj* expression is affected by environmental conditions, seed morphology and anthocyanin inhibitor genes present in the donor background (Burr et al., 1996; Della Vedova et al., 2005; Röber et al., 2005; Kebede et al., 2011; Prigge et al., 2011; Prigge et al., 2012a; Prigge et al., 2012b).

The development of high oil content (OC) inducers allows automated haploid selection, which greatly facilitates accurate estimation of HIRs. However, OC is a quantitatively inherited character (Moreno-Gonzalez et al., 1975; Berke and Rocheford, 1995; Laurie et al., 2004; Môro et al., 2012), as are other traits important to inducers, such as plant height, tassel size and disease tolerance (Gallavotti et al., 2004; Satoh-Nagasawa et al., 2006; Peiffer et al., 2014; Tembo et al., 2014; Zhang et al., 2016; Almeida et al., 2020). Employing genomic selection to simultaneously improve multiple polygenic traits using optimization criteria (Akdemir et al., 2019) dilutes the costs of genotyping and becomes more cost-effective as the number of traits that otherwise would have to be phenotyped increase. Nevertheless, visual discrimination of haploid seeds through the *R1-nj* anthocyanin marker is still performed in the majority of medium to small scale breeding programs employing the DH technique. Therefore, the economic benefits of using inducers with high HIR is still important.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. The genetic polymorphism data is available at <https://doi.org/10.25380/iastate.24185784.v2>. Further inquiries can be directed to the corresponding author.

Author contributions

HT, UF, and TL designed this project; HT, UF, VR, and TL developed the inducer panel used in this study; HT, VA, and EP performed most of the HIR phenotypic selection; MK and WB analyzed the phenotypic and genomic data; HT and RZ executed the candidate gene analysis; HT, MK, RZ and TL wrote this manuscript; TL supervised the research work; All authors contributed to the article and approved the submitted version.

Funding

This research was funded by USDA's National Institute of Food and Agriculture (NIFA) Project, No. IOW04314, IOW01018, and IOW05510; and NIFA award 2018-51181-28419. Funding for this work was also provided by the R.F. Baker Center for Plant Breeding, Plant Sciences Institute, and K.J. Frey Chair in Agronomy, at Iowa State University. The Brazilian National Council for Scientific and Technological Development (CNPq) also provided funding for this work.

Acknowledgments

RZ gratefully acknowledges the funds provided by SERB-International Research Experience Fellowship, Department of Science and Technology, India (File No. SIR/2022/000112). We appreciate Dr. Thomas Widiez for his constructive and insightful remarks on the manuscript.

Conflict of interest

Author VR was employed by VAR BREEDING Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1218042/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Half-normal plots with a simulated envelope for (A) Model 1, (B) Model 1 with Quasi-likelihood inference, (C) Model 2, (D) Model 3, and (E) Model 4.

SUPPLEMENTARY FIGURE 2

Index Pearson residual plots for (A) Model 1, (B) Model 2, (C) Model 3, and (D) Model 4.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control* 19, 716–723. doi: 10.1109/TAC.1974.1100705
- Akdemir, D., Beavis, W., and Fritsche-Neto, R. (2019). Multi-objective optimized genomic breeding strategies for sustainable food improvement. *Heredity* 122, 672–683. doi: 10.1038/s41437-018-0147-1
- Almeida, V. C., Trentin, H. U., Frei, U. K., and Lübberstedt, T. (2020). Genomic prediction of maternal haploid induction rate in maize. *Plant Genome* 13 (1), e20014. doi: 10.1002/tpg2.20014
- Aw, S. J., Hamamura, Y., Chen, Z., Schnittger, A., and Bergeret, F. (2010). Sperm entry is sufficient to trigger division of the central cell but the paternal genome is required for endosperm development in Arabidopsis. *Development* 137, 2683–2690. doi: 10.1242/dev.052928
- Barret, P., Brinkmann, M., and Beckert, M. (2008). A major locus expressed in the male gametophyte with incomplete penetrance is responsible for *in situ* gynogenesis in maize. *Theor. Appl. Genet.* 117, 581–594. doi: 10.1007/s00122-008-0803-6
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Berke, T. G., and Rocheford, T. R. (1995). Quantitative trait loci for flowering, plant and ear height, and kernel traits in maize. *Crop Sci.* 35, 1542–1549. doi: 10.2135/cropsci1995.0011183X003500060004x
- Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103, 338–348. doi: 10.1016/j.ajhg.2018.07.015
- Burr, F., Burr, B., Scheffler, B. E., Blewitt, M., Wienand, U., and Matz, E. C. (1996). The maize repressor-like gene intensifier1 shares homology with the r1/b1 multigene family of transcription factors and exhibits missplicing. *Plant Cell* 8, 1249–1259. doi: 10.1105/tpc.8.8.1249
- Chalyk, S. T. (1999). Creating new haploid-inducing lines of maize. *Maize Genet. Coop. Newsl.* 73, 53–54.
- Chase, S. S. (1947). Techniques for isolating haploid plants. *Am. J. Bot.* 34, 579–609.
- Chase, S. S. (1952). Production of homozygous diploids of maize from monoloids 1. *Agron. J.* 44, 263–267. doi: 10.2134/agronj1952.00021962004400050010x
- Chase, S. S., and Nanda, D. K. (1965). Screening for monoloids of maize by use of a purple embryo marker. *Maize Genet. Coop. Newsl.* 39, 59–60.
- Chen, C., Liu, X., Li, S., Liu, C., Zhang, Y., Luo, L., et al. (2022). Co-expression of transcription factors ZmC1 and ZmR2 establishes an efficient and accurate haploid embryo identification system in maize. *Plant J.* 111, 1296–1307. doi: 10.1111/tpj.15888
- Coe, E. H. (1959). A line of maize with high haploid frequency. *Am. Nat.* 93, 381–382. doi: 10.1086/282187
- Couto, E. G. O., Curry, M. N., Bandeira, Souza, M., Granato, I. S. C., Vidotti, M. S., Domingos, Garbuglio, D., et al. (2019). Effect of F1 and F2 generations on genetic variability and working steps of doubled haploid production in maize. *PLoS One* 14, e0224631. doi: 10.1371/journal.pone.0224631
- Covarrubias-Pazarán, G. (2016). Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS One* 11, e0156744. doi: 10.1371/journal.pone.0156744
- Davidson, R. M., Hansey, C. N., Gowda, M., Childs, K. L., Lin, H., Vaillancourt, B., et al. (2011). Utility of RNA sequencing for analysis of maize reproductive transcriptomes. *Plant Genome* 4, 191–203. doi: 10.3835/plantgenome2011.05.0015
- Deimling, S., Röber, F. K., and Geiger, H. H. (1997). Methodology and genetics of *in vivo* haploid induction in maize. *Vorträge für Pflanzenzüchtung* 38, 203–224.
- De La Fuente, G. N., Frei, U. K., Trampe, B., Nettleton, D., Zhang, W., and Lübberstedt, T. (2018). A diallel analysis of a maize donor population response to *in vivo* maternal haploid induction: I. inducibility. *Crop Sci.* 58, 1830–1837. doi: 10.2135/cropsci2017.05.0285
- Della Vedova, C. B., Lorbiecke, R., Kirsch, H., Schulte, M. B., Scheets, K., Borchert, L. M., et al. (2005). The dominant inhibitory chalcone synthase allele C2-Idf (inhibitor diffuse) from *Zea mays* (L.) acts via an endogenous RNA silencing mechanism. *Genetics* 170, 1989–2002. doi: 10.1534/genetics.105.043406
- Demétrio, C. G. B., Hinde, J., and Moral, R. A. (2014). *Models for overdispersed data in entomology bt - ecological modelling applied to entomology*. Eds. C. P. Ferreira and W. A. C. Godoy (Cham: Springer International Publishing), 219–259.
- Dickins, R. A., Frew, I. J., House, C. M., O'Bryan, M. K., Holloway, A. J., Haviv, I., et al. (2002). The ubiquitin ligase component Siah1a is required for completion of meiosis I in male mice. *Mol. Cell. Biol.* 22, 2294–2303. doi: 10.1128/MCB.22.7.2294-2303.2002
- Dong, X., Xu, X., Miao, J., Li, L., Zhang, D., Mi, X., et al. (2013). Fine mapping of qhir1 influencing *in vivo* haploid induction in maize. *Theor. Appl. Genet.* 126, 1713–1720. doi: 10.1007/s00122-013-2086-9
- Eder, J., and Chalyk, S. (2002). *In vivo* haploid induction in maize. *Theor. Appl. Genet.* 104, 703–708. doi: 10.1007/s00122-001-0773-4
- Gallavotti, A., Zhao, Q., Kyozuka, J., and Meeley, R. B. (2004). The role of barren stalk1 in the architecture of maize. *Nature* 432, 1–6. doi: 10.1038/nature03122.1
- Gilles, L. M., Calhau, A. R. M., Padula, V. L., Jacquier, N. M.A., Lionnet, C., Martinant, J., et al. (2021). Lipid anchoring and electrostatic interactions target NOT-LIKE-DAD to pollen endo-plasma membrane. *J. Cell Biol.* 220 (10), e20210077. doi: 10.1083/jcb.202010077
- Gilles, L. M., Khaled, A., Laffaire, J., Chaignon, S., Gendrot, G., Laplaige, J., et al. (2017b). Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. *EMBO J.* 36, 707–717. doi: 10.15252/embj.201796603
- Gilles, L. M., Martinant, J., Rogowsky, P. M., and Widiez, T. (2017a). Haploid induction in plants. *Curr. Biol.* 27 (20), R1095–R1097. doi: 10.1016/j.cub.2017.07.055
- Han, Y., Cameron, J. N., Wang, L., and Beavis, W. D. (2017). The predicted cross value for genetic introgression of multiple alleles. *Genetics* 205 (4), 1409–1423. doi: 10.1534/genetics.116.197095
- Hu, H., Schrag, T. A., Peis, R., Unterseer, S., Schipprack, W., Chen, S., et al. (2016). The genetic basis of haploid induction in maize identified with a novel genome-wide association method. *Genetics* 202, 1267–1276. doi: 10.1534/genetics.115.184234
- Huang, J. C., Chang, F. C., and Wang, C. S. (1997). Characterization of a lily tapetal transcript that shares sequence similarity with a class of intracellular pathogenesis-related (IPR) proteins. *Plant Mol. Biol.* 34, 681–686. doi: 10.1023/A:1005824306560
- Huang, L. F., Lin, K. H., He, S. L., Chen, J., Jiang, J., Chen, B., et al. (2016). Multiple patterns of regulation and overexpression of a ribonuclease-like pathogenesis-related protein gene, OsPR10a, conferring disease resistance in rice and Arabidopsis. *PLoS One* 11, e0156414. doi: 10.1371/journal.pone.0156414
- Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* 8, 154. doi: 10.1093/gigascience/giy154
- Jacquier, N. M. A., Calhau, A. R. M., Fierlej, Y., Martinant, J., Rogowsky, P. M., Gilles, L. M., et al. (2023). In planta haploid induction by kokopelli mutants. *Plant Physiol.* 193 (1), 182–185. doi: 10.1093/plphys/kiad328
- Jacquier, N. M. A., Gilles, L. M., Martinant, J. P., Rogowsky, P. M., and Widiez, T. (2021). Maize *In Planta* haploid inducer lines, a cornerstone for doubled haploid technology. *Methods Mol. Biol.* 2288, 25–48. doi: 10.1007/978-1-0716-1335-1_2
- Jacquier, N. M., Gilles, L. M., Pyott, D. E., Martinant, J., Rogowsky, P. M., and Widiez, T. (2020). Puzzling out plant reproduction by haploid induction for innovations in plant breeding. *Nat. Plants* 6 (6), 610–619. doi: 10.1038/s41477-020-0664-9
- Jiang, C., Sun, J., Li, R., Yan, S., Chen, W., Guo, L., et al. (2022). A reactive oxygen species burst causes haploid induction in maize. *Mol. Plant* 15 (6), 943–955. doi: 10.1016/j.molp.2022.04.001
- Kebede, A. Z., Dhillon, B. S., Schipprack, W., Araus, J. L., Bänziger, M., Semagn, K., et al. (2011). Effect of source germplasm and season on the *in vivo* haploid induction rate in tropical maize. *Euphytica* 180, 219–226. doi: 10.1007/s10681-011-0376-3
- Kelliher, T., Starr, D., Richbourg, L., Chintamanani, S., Delzer, B., Nuccio, M. L., et al. (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* 542, 105–109. doi: 10.1038/nature20827
- Kermicle, J. L. (1969). Androgenesis conditioned by a mutation in maize abstract. *Science* 166 (3911), 1422–1424. doi: 10.1126/science.166.3911.1422
- Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., et al. (2012). Diversity Arrays Technology: A generic genome profiling technology on open platforms. *Methods Mol. Biol.* 888, 67–89. doi: 10.1007/978-1-61779-870-2_5
- Kim, H. J., Oh, S. A., Brownfield, L., Brownfield, L., Hong, S. H., Ryu, H., et al. (2008). Control of plant germline proliferation by SCF^{FBL17} degradation of cell cycle inhibitors. *Nature* 455, 1134–1137. doi: 10.1038/nature07289
- Kindiger, B., and Hamann, S. (1993). Generation of haploids in maize: A modification of the indeterminate gametophyte (ig) system. *Crop Sci.* 33, 342–344. doi: 10.2135/cropsci1993.0011183X003300020027x
- Lashermes, P., and Beckert, M. (1988). Genetic control of maternal haploidy in maize (*Zea mays* L.) and selection of haploid inducing lines. *Theor. Appl. Genet.* 76, 405–410. doi: 10.1007/BF00265341
- Laurie, C. C., Chasalow, S. D., Ledeaux, J. R., McCarroll, R., Bush, D., Hauge, B., et al. (2004). The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* 2155, 2141–2155. doi: 10.1534/genetics.104.029686
- Lenth, R. (2020). *Emmeans: estimated marginal means, aka least-squares means, R package version 1.5.1*. Available at: <https://cran.r-project.org/src/contrib/Archive/emmeans/>.
- Li, Y., Lin, Z., Yue, Y., Zhao, H., Fei, X., Lizhu, E., et al. (2021). Loss-of-function alleles of ZmPLD3 cause haploid induction in maize. *Nat. Plants* 7, 1579–1588. doi: 10.1038/s41477-021-01037-2
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444
- Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12, e1005767. doi: 10.1371/journal.pgen.1005767

- Liu, C., Li, X., Meng, D., Zhong, Y., Chen, C., Dong, X., et al. (2017). A 4-bp insertion at ZmPLA1 encoding a putative phospholipase generates haploid induction in maize. *Mol. Plant* 10, 520–522. doi: 10.1016/j.molp.2017.01.011
- Liu, C., Li, W., Zhong, Y., Li, L., Zhang, D., Mi, X., et al. (2015). Fine mapping of qhir8 affecting *in vivo* haploid induction in maize. *Theor. Appl. Genet.* 128, 2507–2515. doi: 10.1007/s00122-015-2605-y
- Marrocco, K., Zhou, Y., Bury, E., Dieterle, M., Funk, M., Genschik, P., et al. (2006). Functional analysis of EID1, an F-box protein involved in phytochrome A-dependent light signal transduction. *Plant J.* 45, 423–438. doi: 10.1111/j.1365-313X.2005.02635.x
- Milet, J., Courtin, D., Garcia, A., and Perdry, H. (2020). Mixed logistic regression in genome-wide association studies. *BMC Bioinfo.* 21, 536. doi: 10.1186/s12859-020-03862-2
- Moral, R. A., Hinde, J., and Demétrio, C. G. B. (2017). Half-normal plots and overdispersed models in R: The hnp Package. *J. Stat.* 81, 1–23. doi: 10.18637/jss.v081.i10
- Moreno-Gonzalez, J., Dudley, J. W., and Lambert, R. J. (1975). Design III study of linkage disequilibrium for percent oil in maize1. *Crop Sci.* 15, 840–843. doi: 10.2135/cropsci1975.0011183X001500060027x
- Môro, G. V., Santos, M. F., Bento, D. A. V., Aguiar, A. M., and de Souza, C. L. Jr. (2012). Genetic analysis of kernel oil content in tropical maize with design III and QTL mapping. *Euphytica* 185, 419–428. doi: 10.1007/s10681-011-0604-x
- Nair, S. K., Chaikam, V., Gowda, M., Hindu, V., Melchinger, A. E., and Boddupalli, P. M. (2020). Genetic dissection of maternal influence on *in vivo* haploid induction in maize. *Crop J.* 8 (2), 287–298. doi: 10.1016/j.cj.2019.09.00
- Nair, S. K., Molenaar, W., Melchinger, A. E., Boddupalli, P. M., Martinez, L., Lopez, L. A., et al. (2017). Dissection of a major QTL qhir1 conferring maternal haploid induction ability in maize. *Theor. Appl. Genet.* 130 (6), 1113–1122. doi: 10.1007/s00122-017-2873-9
- Peiffer, J. A., Romay, M. C., Gore, M. A., Flint-Garcia, S. A., Zhang, Z., Millard, M. J., et al. (2014). The genetic architecture of maize height. *Genetics* 196, 1337–1356. doi: 10.1534/genetics.113.159152
- Peterson, R. A. (2021). Finding optimal normalizing transformations via best Normalize. *R J.* 13, 310. doi: 10.32614/RJ-2021-041
- Prigge, V., Sánchez, C., Dhillon, B. S., Schipprack, W., Araus, J. L., Bänziger, M., et al. (2011). Doubled haploids in tropical maize: I. effects of inducers and source germplasm on *in vivo* haploid induction rates. *Crop Sci.* 51, 1498–1506. doi: 10.2135/cropsci2010.10.0568
- Prigge, V., Schipprack, W., Mahuku, G., Atlin, G. N., and Melchinger, A. E. (2012a). Development of *in vivo* haploid inducers for tropical maize breeding programs. *Euphytica* 185, 481–490. doi: 10.1007/s10681-012-0657-5
- Prigge, V., Xu, X., Li, L., Babu, R., Chen, S., Atlin, G. N., et al. (2012b). New insights into the genetics of *in vivo* induction of maternal haploids, the backbone of doubled haploid technology in maize. *Genetics* 190, 781–793. doi: 10.1534/genetics.111.133066
- Randolph, L. F. (1940). Note on haploid frequencies. *Maize Genet. Coop. News.* 14, 23–24.
- R Core Team. (2022). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: <http://www.r-project.org>.
- Ren, J., Zhang, X., Li, Z., and Wu, P. (2022). Genetic analysis of maternal haploid inducibility for *in vivo* haploid induction in maize. *Agriculture* 12, 845. doi: 10.3390/agriculture12060845
- Röber, F. (1999). *Fortpflanzungsbiologische und genetische Untersuchungen mit RFLP-Markern zur in-vivo-Haploideninduktion bei Mais* (Stuttgart, Germany: Universität Hohenheim). Ph.D. Thesis.
- Röber, F. K., Gordillo, G. A., and Geiger, H. H. (2005). *In vivo* haploid induction in maize - Performance of new inducers and significance of doubled haploid lines in hybrid breeding. *Maydica* 50, 275–283. doi: 10.1534/genetics.111.133066
- Roest, H. P., Baarends, W. M., de Wit, J., van Klaveren, J. W., Wassenaar, E., Hoogerbrugge, J. W., et al. (2004). The ubiquitin-conjugating DNA repair enzyme HR6A is a maternal factor essential for early embryonic development in mice. *Mol. Cell. Biol.* 24, 5485–5495. doi: 10.1128/MCB.24.12.5485-5495.2004
- Ron, M., Alandete Saez, M., Eshed Williams, L., Fletcher, J. C., and McCormick, S. (2010). Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in Arabidopsis. *Gen. Dev.* 24, 1010–1021. doi: 10.1101/gad.1882810
- Rotarencu, V., Georgeta, D., State, D., and Fuiia, S. (2010). New inducers of maternal haploids in maize. *Maize Genet. Coop. News.* 84, 36–50.
- Santos, I.G.d., Verzegnazzi, A. L., Edwards, J., Fletcher, J. C., and McCormick, S. (2022). Usefulness of temperate-adapted maize lines developed by doubled haploid and single-seed descent methods. *Theor. Appl. Genet.* 135, 1829–1841. doi: 10.1007/s00122-022-04075-2
- Satoh-Nagasawa, N., Nagasawa, N., Malcomber, S., Sakai, H., and Jackson, D. (2006). A trehalose metabolic enzyme controls inflorescence architecture in maize. *Nature* 441, 227–230. doi: 10.1038/nature04725
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136
- Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, U., Long, Q., et al. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44, 825–830. doi: 10.1038/ng.2314
- Silva, A.J.da. (2017). *Modelos para análise de dados superdispersos de indução de haploidia em milho [dissertação]* (Piracicaba: Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz). doi: 10.11606/D.11.2017.tde-20072017-133448
- Stelpflug, S. C., Sekhon, R. S., Vaillancourt, B., Hirsch, C. N., Buell, C. R., de Leon, N., et al. (2016). An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome* 9 (1). doi: 10.3835/plantgenome2015.04.0025
- Suzumori, N., Burns, K. H., Yan, W., and Matzuk, M. M. (2003). RFPL4 interacts with oocyte proteins of the ubiquitin-proteasome degradation pathway. *Proc. Natl. Acad. Sci.* 100, 550–555. doi: 10.1073/pnas.0234474100
- Tembo, L., Asea, G., Gibson, P. T., and Okori, P. (2014). Quantitative Trait Loci for Resistance to *Stenocarpella maydis* and *Fusarium graminearum* Cob Rots in Tropical Maize. *J. Crop Improv.* 28, 214–228. doi: 10.1080/15427528.2013.872750
- Trentin, H., Batiru, G., Frei, U. K., Dutta, S., and Lübberstedt, T. (2022). Investigating the effect of the interaction of maize inducer and donor backgrounds on haploid induction rates. *Plants (Basel Switzerland)* 11, 1527. doi: 10.3390/plants11121527
- Trentin, H. U., Yavuz, R., Dermal, A., Frei, U. K., Dutta, S., and Lübberstedt, T. (2023). Comparison between inbred and hybrid maize haploid inducers. *Plants* 12, 1095. doi: 10.3390/plants12051095
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Verzegnazzi, A. L., dos Santos, I. G., Krause, M. D., Hufford, M., Frei, U. K., Campbell, J., et al. (2021). Major locus for spontaneous haploid genome doubling detected by a case-control GWAS in exotic maize germplasm. *Theor. Appl. Genet.* 134, 1423–1434. doi: 10.1007/s00122-021-03780-8
- Walley, J. W., Sartor, R. C., Shen, Z., Schmitz, R. J., Wu, K. J., Urich, M. A., et al. (2016). Integration of omic networks in a developmental atlas of maize. *Science* 353, 814–818. doi: 10.1126/science.aag1125
- Wang, H., Lu, Y., Jiang, T., Berg, H., Li, C., and Xia, Y. (2013). The Arabidopsis U-box/ARM E3 ligase AtPUB4 influences growth and degeneration of tapetal cells, and its mutation leads to conditional male sterility. *Plant J.* 74, 511–523. doi: 10.1111/tj.12146
- Wang, D., Zhong, Y., Feng, B., Qi, X., Yan, T., Liu, J., et al. (2023). The RUBY reporter enables efficient haploid identification in maize and tomato. *Plant Biotechnol. J.* 21, 1707–1715. doi: 10.1111/pbi.14071
- Xu, X., Li, L., Dong, X., Jin, W., Melchinger, A. E., and Chen, S. (2013). Gametophytic and zygotic selection leads to segregation distortion through *in vivo* induction of a maternal haploid in maize. *J. Exp. Bot.* 64, 1083–1096. doi: 10.1093/jxb/ers393
- Zhang, Y., Cui, M., Zhang, J., Zhang, L., Li, C., Kan, X., et al. (2016). Confirmation and fine mapping of a major QTL for aflatoxin resistance in maize using a combination of linkage and association mapping. *Toxins (Basel)* 8, 1–15. doi: 10.3390/toxins8090258
- Zhong, Y., Liu, C., Qi, X., Jiao, Y., Wang, D., Wang, Y., et al. (2019). Mutation of ZmDMP enhances haploid induction in maize. *Nat. Plants* 5, 575–580. doi: 10.1038/s41477-019-0443-7