Theoretical and Applied Genetics (2025) 138:5
https://doi.org/10.1007/s00122-024-04789-5

ORIGINAL ARTICLE

Genome-wide association study of haploid female fertility (HFF) and haploid male fertility (HMF) in BS39-derived doubled haploid maize lines

Mercy Fakude[1] · Ann Murithi[1] · Ursula K. Frei[1] · Paul M. Scott[1,2] · Thomas Lübberstedt[1]

Received: 28 July 2024 / Accepted: 19 November 2024 / Published online: 11 December 2024

Abstract

Key message **Restoration of haploid female and haploid male fertility without colchicine is feasible. Three SNPs and eight gene models for HFF, and one SNP and a gene model for HMF were identified.**

Abstract Doubled haploid (DH) breeding accelerates the development of elite inbred lines and facilitates the incorporation of exotic germplasm, offering a powerful tool for maize improvement. Traditional DH breeding relies on colchicine to induce haploid genome doubling. Colchicine is toxic, and its application is labor-intensive, with most genotypes recording low genome doubling rates (10–30%). This study investigates spontaneous haploid genome doubling (SHGD) as a safer and more efficient alternative to colchicine. We evaluated the effectiveness of SHGD in restoring haploid female fertility (HFF) and haploid male fertility (HMF) without colchicine. Using genome-wide association studies (GWAS), we identified genomic regions influencing HFF and HMF. The plant materials included the BS39-haploid isogenic lines (HILs) and BS39-SHGD-haploid isogenic lines (HILs). Our results revealed significant SNP associations for both traits, with candidate genes involved in cell cycle regulation, cytoskeletal organization, and hormonal signaling. Analysis of variance (ANOVA) revealed significant variation in HFF across haploids and two environments. Similarly, HMF showed substantial differences across haploids and between the two environments. Spearman correlation between HFF and HMF showed no correlation ($r = -0.03$) between the two traits. HFF showed high heritability (0.8), indicating strong genetic control, whereas HMF displayed moderate heritability (0.5), suggesting additional environmental influences. The findings underscore the potential of SHGD to enhance DH breeding efficiency and support the development of new maize varieties tailored to diverse agricultural needs.

Introduction

Maize (*Zea mays L.*) is a key cereal crop in global agriculture, supplying nutrition, feedstock, and industrial raw materials (Shiferaw et al. 2021). Its significance stems not only from its economic value but also from its remarkable genetic diversity and adaptability (Mikel 2018). To enhance maize productivity and resilience to the current and projected climate changes, doubled haploid (DH) breeding emerges as a powerful tool, offering accelerated breeding cycles and precise genetic manipulation (Prasanna 2012). DH breeding in maize allows breeders to accelerate the development of elite inbred lines for hybrid seed production. DH technology accelerates the development of new maize varieties tailored to the diverse needs of farmers and consumers by simplifying the breeding process and eliminating recurrent self-pollination and generations of backcrossing. Thus, DH breeding is gaining prominence in commercial breeding programs due to its benefits in accelerating the release of new hybrids and realizing increased genetic gain per year (Melchinger and Würschum 2010).

Unlike the traditional inbred line production protocol, the DH method facilitates development of homozygous inbred lines in two generations (one year). Generation 1 entails haploid induction. Generation 2 involves selfing fertile haploids to produce 100% homozygous DH lines (Trentin et al. 2020; Gupta et al. 2022). Currently, DH technology utilizes chemical doubling agents to induce haploid genome doubling, and

Communicated by Rod Snowdon.

✉ Thomas Lübberstedt
thomasl@iastate.edu

[1] Department of Agronomy, Iowa State University, Ames, IA 50011, USA

[2] United States Department of Agriculture, Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, IA 50011, USA

subsequently, haploid male fertility. The resulting plants are doubled haploids (DHs) (Chaikam et al. 2019a, b). In most cases, the haploid plants are chimeras comprising an assortment of both haploid and homozygous diploid cell lines but are still referred to as haploids or D0 generation (Xu et al. 2007).

Practically, the DH inbred line production protocol involves (1) crossing of two inbred parents to produce an F_1. (2) F_1 plants or other donor plants are planted and pollinated with an inducer genotype to obtain putative haploid kernels. (3) Kernels from induction crosses are harvested and sorted into hybrids and putative haploid kernels using a phenotypic marker (*R1-Navajo*). (4) Putative haploid seed is planted in seedling trays in the greenhouse, and seedlings are treated with the chemical colchicine to induce haploid genome doubling. (5) These haploid plants (D0 plants) are transplanted into the field, and false-positive plants (hybrids resulting from the cross between inducer and donor) are rogued. (6) Male and female fertile haploids are self-pollinated to produce 100% homozygous DH lines (DH1 seeds) (Chang and Coe 2009; Chaikam et al. 2019a, b). DH lines can be released as a variety in self-pollinating crops like wheat, barley, rapeseed, and rice (Hale et al. 2022).

Haploid plants are sterile by nature. They contain only one set of chromosomes (Li et al. 2021). For haploid plants to produce seeds, they must be artificially initiated by colchicine treatment to double their genomes to produce fertile reproductive organs. Success in genome doubling results in fertile male and female inflorescence, also called haploid fertility (Trampe et al. 2020). Haploid fertility encompasses haploid male fertility (HMF) and haploid female fertility (HFF) (Fig. 1). HMF is the ability of haploid plants to produce fertile anthers and viable pollen, while HFF is the ability of the ear to set at least one seed after pollination. With colchicine treatment, HMF is reported to be low, while HFF does seem not to be limiting, with approximately 97–100% of haploids producing seeds when pollinated by diploid plants (Chalyk 1994).

Colchicine can be harmful to humans; its application is laborious and time-consuming and is associated with high greenhouse costs and governmental restrictions on handling and disposal of colchicine (Melchinger et al. 2013). Moreover, colchicine treatment typically confers a genome doubling rate of approximately 10–30%. Thus, haploid genome doubling remains a bottleneck to large-scale DH production due to low genome doubling rates presented by colchicine treatment (Boerman et al. 2020).

Here, we examine a trait called spontaneous haploid genome doubling (SHGD). SHGD occurs when a haploid plant duplicates its genome without applying colchicine. While most haploids are sterile, some genotypes display doubling rates exceeding 50% (Boerman et al. 2020). It has been shown that SHGD is heritable, involving both major and minor genes (Molenaar et al. 2019; Trampe et al. 2020; Ren et al. 2020; Foster et al. 2024). Genome doubling can occur during mitosis or meiosis, leading to fertile female and male inflorescences (Wu et al. 2014a, b; Sugihara et al. 2013; Trampe et al. 2020). Unlike the colchicine-mediated DH protocol, the SHGD-mediated DH protocol entails (1) crossing the source inbred line (e.g. BS39) is crossed with a SHGD donor inbred line (e.g. A427) to produce an F_1 generation. Since the SHGD allele is inherited from A427, the resulting haploids do not require colchicine or other chemicals for genome doubling. (2) F_1 seeds are planted and pollinated with an inducer genotype. (3) These kernels are harvested, and putative haploid kernels are identified. (4) Putative haploid seed is planted directly in the field, and false-positive plants are rogued. (5) The male fertile haploids

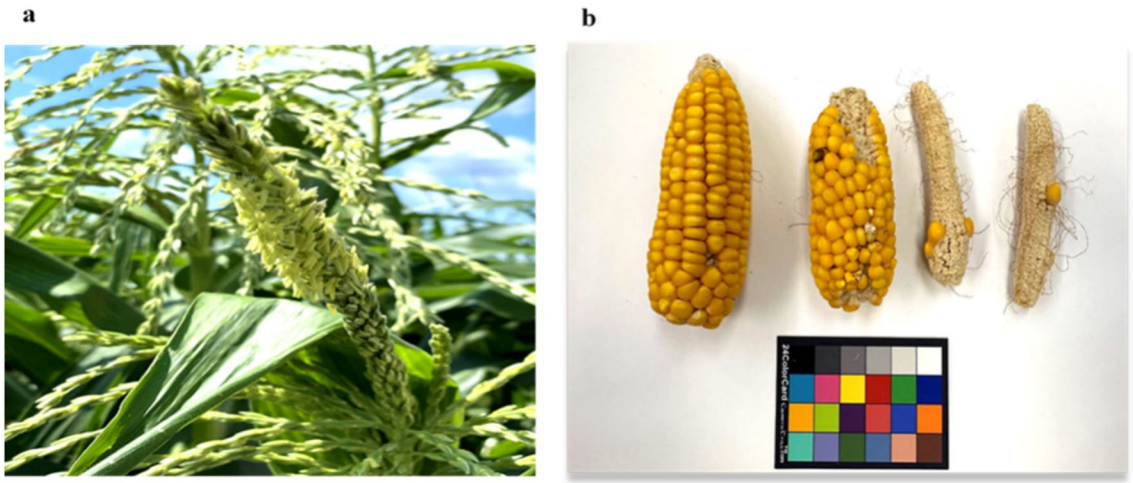

Fig. 1 Haploid fertility includes haploid male fertility and haploid female fertility **a** Haploid plant with functional "fat" anthers and viable pollen (HMF) **b** Ears harvested from one haploid line with varying levels of female fertility or seed set (HFF)

are self-pollinated to produce 100% homozygous DH lines. SHGD also presents the opportunity to breed seeds suitable for organic production since it allows breeders to omit colchicine in the protocol.

Inbred line A427 exhibits a high haploid male fertility (~78%) and carries a major QTL for SHGD (Trampe et al. 2020; Verzegnazzi et al. 2021). The major SHGD QTL, which has been shown to increase HMF, has been introgressed to over 200 BS39-derived lines (Santos et al. 2022). However, there exists a gap in knowledge regarding whether (1) haploid female fertility (HFF) restoration, without colchicine application, is adequate to yield ears with optimal seed set after pollination (2) and whether there are genomic regions, influencing both haploid female fertility and male fertility in the haploids derived through spontaneous haploid genome doubling (SHGD). Genome-wide association study (GWAS) was conducted to identify key loci controlling both HFF and HMF. We anticipated to reveal significant genetic variation for HFF between BS39 and BS39-SHGD lines, since the SHGD QTL may enhance HFF. Additionally, this study was conducted to better understand, whether HFF and HMF are due to the same or distinct loci as this will guide future breeding strategies aimed at improving both male and female fertility in haploids.

Thus, the objectives of the study were to (1) evaluate genetic variation for HFF in BS39 haploids with and without the major SHGD QTL allele *qshgd1* from A427, (2) determine the relationship between HFF and HMF, and (3) identify genomic regions for both HFF and HMF in a genome-wide association study (GWAS).

Materials and methods

Plant materials

The BS39 population was derived from five exotic Tusón accessions and has been adapted to Midwest photoperiod conditions through several generations of intercrossing (Hallauer and Carena 2016). BS39-SHGD-DH lines were derived from a cross between BS39 and inbred A427 at Iowa State University-DH Facility. Inbred A427 is a public non-stiff stalk inbred line developed by the University of Minnesota (Gerdes et al. 1993). It is notable for its HMF rate, which is approximately 78%. This inbred line was used as a source of SHGD donor. Both BS39 and BS39-SHGD-derived DH lines were genotyped by genotyping by sequencing (GBS) (Verzegnazzi et al. 2021; Santos et al. 2022).

In summer 2022, we induced 228 DH lines (85 derived from the original BS39 population (BS39-DHs) and 143 DH lines derived from the cross between BS39 population and the SHGD donor A427 (BS39-SHGD-DHs). The 228 DH lines were crossed to a haploid inducer line (BHI306) to produce haploid seed. After harvesting, haploid selection was manually done based on the expression of *R1-Navajo* in the embryo and aleurone. After haploid selection, we obtained haploids from 192 DH lines. Those constituted our 192 haploid isogenic lines (HILs), which were genetically uniform and perfectly isogenic to the original 192 DH lines. 66 HILs were derived from BS39-DHs, and 126 HILs were derived from BS39-SHGD-DHs. Those HILs were used in subsequent agronomic experiments (see below). Moreover, because these HILs were isogenic to respective genotypes DH lines, marker data produced for DH lines could be used in our HILs-based GWAS.

Experimental design

In summer 2023, the 192 HILs were directly sown at Iowa State University's Agricultural Engineering and Agronomy Farm in Boone, IA. The 192 HILs were planted in a completely randomized design in two environments. Twenty seeds were planted per HIL in a single-row plot with a within-row spacing of 10 cm and a between-row spacing of 0.7 m. At V2 stage throughout V6 stage, diploid plants, identified as false positives, were removed by observing leaf morphology. Compared to diploid plants, haploid plants tend to be shorter, have smaller internodes, narrower leaf area, and upright leaf morphology (Aboobucker et al. 2022; Dermail et al. 2024). BS39 and BS39-SHGD diploid plants, along with the Viking 60-01N hybrid, were planted around the haploid experiment to ensure sufficient bulk pollen for open pollination. At physiological maturity, ears were left to dry.

Phenotypic data collection

The trial was evaluated daily to capture HILs that may flower early. At flowering, haploid plants displaying fertile anthers and pollen shedding were marked with an orange tag and evaluated. Fertility was scored on a scale from 0 (no pollen shed) to 5 (good pollen shed). The haploid male fertility (HMF) percentage was calculated by dividing the number of pollen-shedding haploids by the total number of haploids in each plot as follows:

$$\mathrm{HMF}(\%) = \left[\frac{\text{Number of pollen shedding haploids}}{\text{Total number of haploids}}\right] \times 100 \quad (1)$$

The HMF percentages for each HIL were recorded and used in the association study for HMF. At harvest, the average number of kernels per plot were counted and recorded. The average number of kernels recorded per plot was used to represent HFF in the association study. To adjust the average number of kernels per plot, final plant stand was counted,

and the average number of kernels per plot was computed as follows

$$\text{Average HFF} = \left[\frac{\text{Number of kernels per plot}}{\text{Total number of haploids per plot}}\right] \quad (2)$$

To compare HFF and HMF between the BS39 and BS39-SHGD-derived HILs, permutation tests were conducted to assess the likelihood that there is no difference between groups.

Statistical analysis of phenotypic data

Linear mixed models for both HFF and HMF analysis were fitted in R using the *lme4* package (Bates et al. 2015) as follows

$$Yij = \mu + Bi + Gj + \varepsilon ij \quad (3)$$

where *Yij* is the response variable (i.e., HFF or HMF), μ is the overall mean. *Bi* is the random effect of the *i*th environment. *Gj* is the fixed effect of the *j*th genotype, and εij is the residual effect.

Violation of normality of HFF and HMF data was evaluated using Shapiro–Wilk tests, in R software (Shapiro and Wilk 1965). Transformation was performed on data for both traits, as their distributions did not follow a normal distribution. The HFF trait was transformed using the *'bestNormalize'* package in R. While HMF was transformed using the inverse logit function, given that HMF is a binary trait. Analyses of variance (ANOVA) for HMF and HFF were performed after both traits were transformed. Heritability on an entry-mean basis for the two traits was computed as

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\varepsilon^2}{r}} \quad (4)$$

where σ_g^2 is the genetic variance component. σ_ε^2 is the residual variance component. *r* is the effect of the environments. Three correlations were computed using Spearman correlation in *R*. Correlation for HFF and HMF in both BS39 and BS39-SHGD-HILs, correlation for HFF and HMF in BS39-HILs was computed, and correlation for HFF and HMF in BS39-SHGD-haploids was computed. Best linear unbiased estimates (BLUES) of the genotype effect for both traits were computed in R using the *lme4* package. The effect of genotype was treated as a fixed factor, while the effect of environments was treated as a random effect. The generated BLUES were used as phenotypic data in the GWAS.

Genotyping and SNP calling

Genotyping and SNP calling were previously performed by Verzegnazzi et al. (2021) and Santos et al. (2022). Genotyping by Sequencing (GBS), as described by Elshire et al. (2011), was conducted on 471 lines, comprising 153 BS39-DH lines and 318 BS39-SHGD-DH lines. This process yielded a total of 955,690 SNPs. For the present study, the genotypic data of the 192 DH lines were sampled from this larger dataset of 471 lines. Filtering of the SNP data was conducted using TASSEL 5.2.58 (Glaubitz et al. 2014). SNPs with a minor allele frequency (MAF) below 5% and a call rate below 50% were excluded. Additionally, lines exhibiting more than 5% heterozygosity were discarded. Any remaining heterozygous loci were considered as missing data. After applying these filtering criteria, 414,337 SNPs were retained, for the 66 BS39-DH and 126 BS39-SHGD-DH lines that were used to generate 192 HILs. The selected SNPs were annotated to represent the corresponding chromosome number and base pair position.

Linkage disequilibrium decay and population structure

Linkage disequilibrium decay, kinship, and principal component analysis were performed in R software using the GAPIT package (Tang et al. 2016). A genome-wide LD was quantified using the squared coefficient of correlation (R^2) values of alleles, which provide insights into the non-random association of alleles at different loci. A kinship matrix was computed using the VanRaden method (VanRaden 2008). A principal component analysis (PCA) was performed to account for population structure, relatedness, and dimensionality. By performing PCA on the genotype data matrix, GAPIT identified the major axes of genetic variation, represented by principal components (PCs), which capture distinct patterns of genetic diversity within the population. The top three PCs, explaining the most genetic variation, were interpreted as indicators of population structure. These PCs served as covariates in the subsequent association mapping analyses.

Marker-trait analysis

GAPIT was used for the association between SNP markers and HFF and HMF. FARMCPU and MLMM model was used for the association studies. FarmCPU was selected based on its ability to integrate genetic relatedness as a random effect to account for population structure. MLMM was selected based on its ability to account for kinship, control false positives, and handle a large number of markers (Kaler et al. 2020). SimpleM was used for multiple hypotheses correction. Visualization of Manhattan and QQ plots of GWAS was conducted using the R package *CMplot*.

Mapping of potential candidate genes

Candidate genes were identified within the 200 kb regions upstream or downstream of stable QTL loci, utilizing the B73 RefGen_v5 reference genome in MaizeGDB (http//www.maizegdb.org). To align SNP data developed using the B73v2 reference genome with the updated B73v5 reference genome, a coordinate conversion for the flanking regions surrounding each SNP was conducted. Approximately, 200 kb both upstream and downstream of each SNP were considered. The conversion process involved identifying a gene at the boundary of each 200 kb segment in the B73v2 genome and confirming its updated location in the B73v5 genome. This approach enabled the accurate redefinition of the SNP-flanking regions based on the positional shifts of these boundary genes in the newer genome version. Genes found directly on or close to each associated SNP were considered possible candidate genes for HFF and HMF.

Results

Trait distribution, correlations, and heritability

Statistical analysis revealed distinct distributions for the two traits under investigation. HFF displayed a positively skewed distribution, while HMF exhibited a bimodal distribution (Fig. 2). Analysis of variance (ANOVA) revealed significant variation among haploids and environments for HFF. For HMF, a binary logistic regression model was used to assess the effects of HILs and environments. Likewise, HMF showed significant differences among HILs and between the two environments. Heritability estimates on an entry-mean basis revealed a high heritability for HFF (0.8) and an intermediate heritability for HMF (0.5) (Table 1). Spearman correlation between HFF and HMF in the whole panel was $r=-0.03$ ($p>0.5$). Similarly, correlation coefficient between HFF and HMF for the BS39-HILs was $r=-0.01$ ($p>0.5$) and $r=0.01$ ($p>0.5$) for BS39-SHGD-HILs.

Comparative statistics between BS39 and BS39-SHGD HFF and HMF

The comparison of average HFF between BS39-HILs and BS39-SHGD-HILs revealed significant differences in their distributions. BS39-HILs revealed a lower mean (15.3 per plot) compared to BS39-SHGD-HILs (76.1 per plot). This difference is further highlighted by the standard deviation of 20.6 for BS39-HILs and 75.1 for BS39-SHGD-HILs. The Interquartile Range (IQR) for BS39-HILs (14.5) and BS39-SHGD-HILs (71.7) further highlights the wider spread of values in the BS39-SHGD subset. The 95% confidence interval for the mean for BS39-HILs (10.5–20.1) did not overlap with that for BS39-SHGD-HILs (58.4–93.7). The permutation test supports a significant difference in HFF ($p=6.4\times10^{-9}$) between BS39-HILs and BS39-SHGD-HILs.

For HMF, the BS39-HILs exhibited minimal fertility (min = 0.0%), with an average (mean) HMF of 46.2% and a maximum of 100%. The standard deviation (SD) was 38.9, indicating variability within this subset. The 95% confidence interval for the mean ranged from 39.5–52.9. In contrast, BS39-SHGD-HILs showed a minimum HMF of 60%, a mean HMF of 76.0%, and a maximum HMF of 100%. The standard deviation was significant at 12.8, reflecting the extensive range in observed values. The interquartile range was 22.7, with a 95% confidence interval for the mean fertility from 80.5 to 85.0. A permutation test comparing BS39-HILs to BS39-SHGD-HILs showed a highly significant p-value of 2.2×10^{-16}, supporting a higher HMF for BS39-SHGD-HILs (Table 1).

Linkage disequilibrium decay and population structure

A principal component analysis (PCA) of the two traits was conducted. The PCA identified two distinct subpopulations. PCA plot showed that the first two principal components explain a cumulative variance of 11.2% (Fig. 3), suggesting that while there is genetic diversity within the population, much of it remains unexplained by the first two principal components. The sample size in this analysis may be small, limiting the power of PCA to effectively summarize the variation in the data.

Marker-trait analysis

After filtering for SNPs with a minor allele frequency (MAF) below 1%, missing rate below 20%, and heterozygosity rate below 10%, 414,337 SNPs were retained. A multi-locus method (MLMM) was used in the association analysis (Table 2). Three significant SNPs associated with HFF were detected. The first SNP (S4_193345574) was detected on chromosome 4 at position 193,345,574 bp (4.7×10^{-6}). The second SNP (S5_135455180) was detected on chromosome 5 at position 135,455,180 bp (1.2×10^{-6}). The third SNP (S10_114316610) was detected on chromosome 10 at position 114,316,610 bp ($8.7\times10^{-6)}$. One significant SNP (S1_279763611) was detected for HMF. This SNP was detected on chromosome 1 at position 279,763,611 bp (7.3×10^{-6}). Based on the LD decay (220 kb) of the association panel, a total of eight gene models were identified for the three significant SNPs (Table 2).

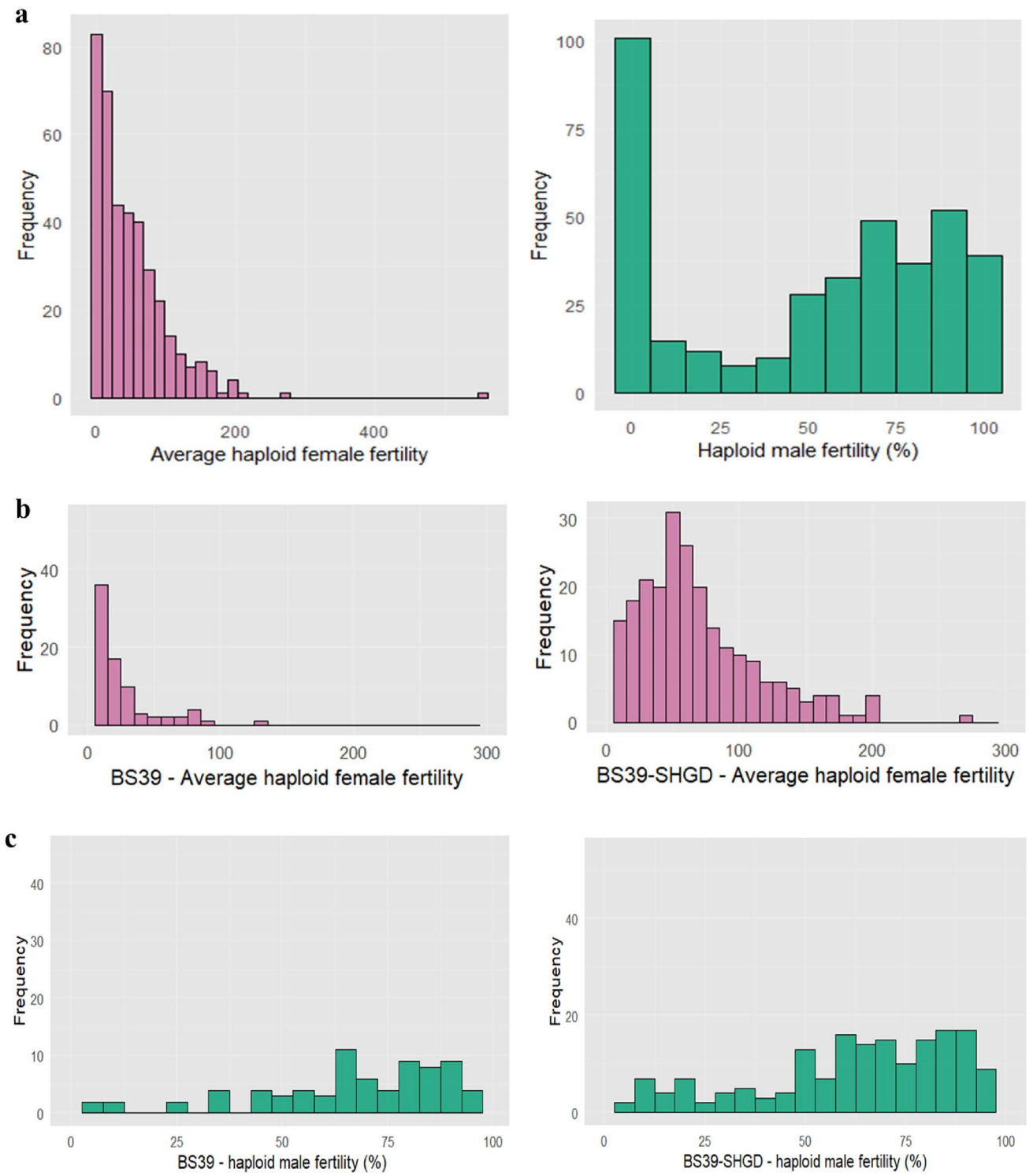

Fig. 2 **a** Distribution of average HFF (left) and HMF (right) in the whole panel **b** distribution of HFF in BS39-HILs (left) and in BS39-SHGD-HILs (right) **c** distribution of HMF in BS39-HILs (left) and BS39-SHGD-HILs (right)

Candidate gene mapping

Eight candidate genes were identified for haploid female fertility (HFF). Three candidate genes (Zm00001eb198220, Zm00001eb198240, and Zm00001eb198300) were associated with SNP S4_193345574. Zm00001eb198220 encodes the RETINOBLASTOMA-RELATED1 (RBR) protein (Desvoyes and

Table 1 Comparative statistics for HFF and HMF in BS39-HILs and BS39-SHGD-HILs

Trait	Source	Df	P-value	Heritability	Group	Min	Mean	Max	SD	IQR	95% CI	Permutation test
Average HFF	Haploids	191	2×10^{-16}	0.8	BS39	0.0	15.3	83.3	20.6	14.6	10.5–20.1	$p=6.4\times10^{-9}$
	Environments	1	0.03		BS39-SHGD	2.5	76.1	549.7	75.1	71.7	58.4–93.7	
HMF (%)	Haploids	191	2.2×10^{-16}	0.5	BS39	0.0	46.2	100	38.9	81.3	39.5–52.9	$p=2.2\times10^{-16}$
	Environments	1	2.9×10^{-5}		BS39-SHGD	60	82.7	100	12.8	22.7	80.5–85.0	

Df. degrees of freedom, *SD* standard deviation, *IQR* the interquartile range, *CI* confidence interval

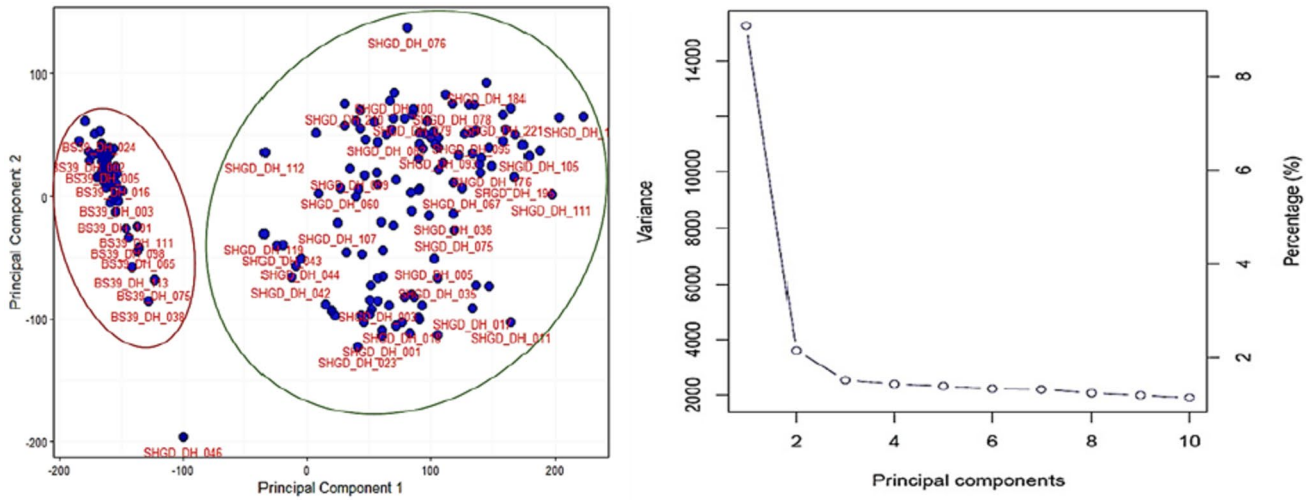

Fig. 3 Principal component analysis plot presenting two subpopulations (left) with a scree plot showing the first two PCs explaining variation (right)

Table 2 Summary of the three significant SNPs associated with haploid female fertility (HFF) and one SNP associated with haploid male fertility (HMF)

Trait	SNP	Chr	Position	*P*-value	MAF	Method
HFF	S4_193345574	4	193,345,574	4.7×10^{-6}	0.1	MLMM
	S5_135455180	5	135,455,180	1.2×10^{-6}	0.4	MLMM; FARMCPU
HMF	S10_114316610	10	114,316,610	8.7×10^{-6}	0.1	MLMM; FARMCPU
	S1_279763611	1	279,763,611	7.3×10^{-6}	0.3	MLMM; FARMCPU

Chr. Chromosome

Gutierrez 2020; Zaragoza et al. 2024). Zm00001eb198240 encodes a HORMA domain-containing protein (Prince and Martinez-Perez 2022). Zm00001eb198300 encodes Ankyrin repeat domain-containing protein 2A. Two candidate genes (Zm00001eb237460 and Zm00001eb237470) were identified within SNP S5_135455180. Zm00001eb237460 encodes Myb-related protein 3R-1 (Haga et al. 2007). Zm00001eb237470 encodes Rop guanine nucleotide exchange factor 9, which includes a PRONE domain (Zhang and McCormick 2007). Three candidate genes (Zm00001eb421910, Zm00001eb421980, and Zm00001eb421990) were identified for SNP S10_114316610. Zm00001eb421910 encodes Protein BIC1 (Yang et al. 2020). Zm00001eb421980 encodes an OTU-like cysteine protease (Li et al. 2023). Zm00001eb421990 encodes a putative F-box protein, SLEEPY1 (SLY1) in *Arabidopsis thaliana* (Ariizumi et al. 2011). For the SNP (S1_279763611) detected for haploid male fertility (HMF), only one gene, Zm00001eb058360, was identified. This gene encodes a GNAT transcription factor (Putative N-acetyltransferase HLS1) (Sakamota et al. 2022) (Table 3).

Table 3 Functional annotations of the candidate genes associated with significant SNPs

Trait	SNP	Candidate gene	Chr	Position (bp)	Functional annotations
HFF	S4_193345574	Zm00001eb198220	4	199,499,786–199,508,476	Retinoblastoma-related protein 1
		Zm00001eb198240	4	199,508,616–199,513,936	HORMA domain-containing protein
		Zm00001eb198300	4	199,899,922–199,925,239	Ankyrin repeat domain-containing protein 2A
	S5_135455180	Zm00001eb237460	5	137,986,868–138,015,382	Myb-related protein 3R-1
		Zm00001eb237470	5	138,168,145–138,171,296	Rop guanine nucleotide exchange factor 9; PRONE domain-containing protein
HMF	S10_114316610	Zm00001eb421910	10	116,630,994–116,631,964	Protein BIC1
		Zm00001eb421980	10	117,016,687–117,023,205	OTU-like cysteine protease family protein
		Zm00001eb421990	10	117,036,868–117,038,439	Putative F-box protein
	S1_279763611	Zm00001eb058360	1	286,695,758–286,698,656	GNAT transcription factor (Putative N-acetyltransferase HLS1)

SNP Single nucleotide polymorphism, *Chr.* Chromosome

Discussion

Spontaneous haploid genome doubling requires both HFF and HMF

Spontaneous haploid genome doubling (SHGD) presents an opportunity to eliminate colchicine treatment in the DH maize breeding protocol (Kleiber et al. 2012; Boerman et al. 2020). Eliminating colchicine not only reduces human exposure to this hazardous chemical but also qualifies DH breeding to meet requirements for organic seed production. To fully leverage the potential of SHGD in DH maize breeding, a thorough understanding of how HFF and HMF function is essential. While there is a considerable body of research focused on HMF and its underlying genetic mechanisms, studies exploring HFF are comparatively few (Geiger et al. 2009; Chalyk 1994). The few studies only confirm the ability of haploid plants to set seeds if pollinated with pollen from diploid plants but have not identified the genetic mechanism underlying HFF in order to leverage an SHGD DH protocol. Thus, this study seeks to understand whether haploid female fertility restoration without colchicine application is feasible and also identify genomic regions influencing female fertility restoration in haploids. In addition to the available HMF studies on various germplasm (De La Fuente et al. 2013; Wu et al. 2014a, b; Chaikam et al. 2019a, b; Trampe et al. 2020; Ma et al. 2018; Molenaar et al. 2019), this study focused on evaluating HMF and the genomic regions influencing HMF in the BS39 and BS39-SHGD germplasm by genome-wide association studies (GWAS). GWAS is conducted to identify genomic regions linked to specific traits. Identifying genomic regions is useful for elucidating the genetic architecture of a trait and designing successful breeding programs (Cortes et al. 2021). In this study, genotyping by sequencing (GBS) was selected based on its efficiency and ability to process large volumes of genetic data (Elshire et al. 2011). Figure 4a reveals that GBS-SNPs were fairly uniform across the entire genome, while some regions revealed higher SNP density.

Comparison of HFF and HMF between BS39 and BS39-SHGD

The comparative analysis of HFF and HMF between the two groups (BS39-HILs and BS39-SHGD-HILs) revealed significant differences in the performance of both traits. For HFF, BS39-HILs revealed a mean fertility of 15.3 kernels per plot, with instances of complete infertility (min = 0 kernels per plot). Contrary, BS39-SHGD-HILs showed a higher fertility mean of 76.1 kernels per plot, suggesting an enhanced female fertility performance. Furthermore, BS39-HILs showed a mean HMF of 46.2%, which is higher than the 12.72% reported by Santos et al. (2022) for BS39-HILs. Similarly, BS39-SHGD-HILs showed an HMF mean of 82.7%, which is higher than the 22.59 reported by Santos et al. (2022) for BS39-SHGD-HILs. The increase in HMF may be attributed to selection. The BS39-HILs were derived from BS39-DH lines, which likely have been inadvertently selected for HMF. The observed increase in HMF in BS39-HILs compared to Santos et al. (2022) suggests that selection for HMF has been effective, and potentially indicates the presence of other SHGD QTL that may be contributing to improved fertility rates. These findings align with Molenaar et al. (2019), reporting an increase in HMF and confirming that recurrent selection for HMF is effective. The high haploid male and female fertility in the BS39-SHGD-HILs lines can be attributed to the presence of the SHGD QTL, supporting the feasibility of HFF and HMF restoration without colchicine. In this study, BS39-SHGD-HILs revealed a higher mean (76.1 kernels per plot) than BS39-HILs (15.3 kernels per plot). These results are higher than the results obtained by Chase (1969) but conform to the trend seen in the female fertility results of Chalyk (1994) and Geiger et al. (2009). Chalyk (1994) reported an unexpected mean

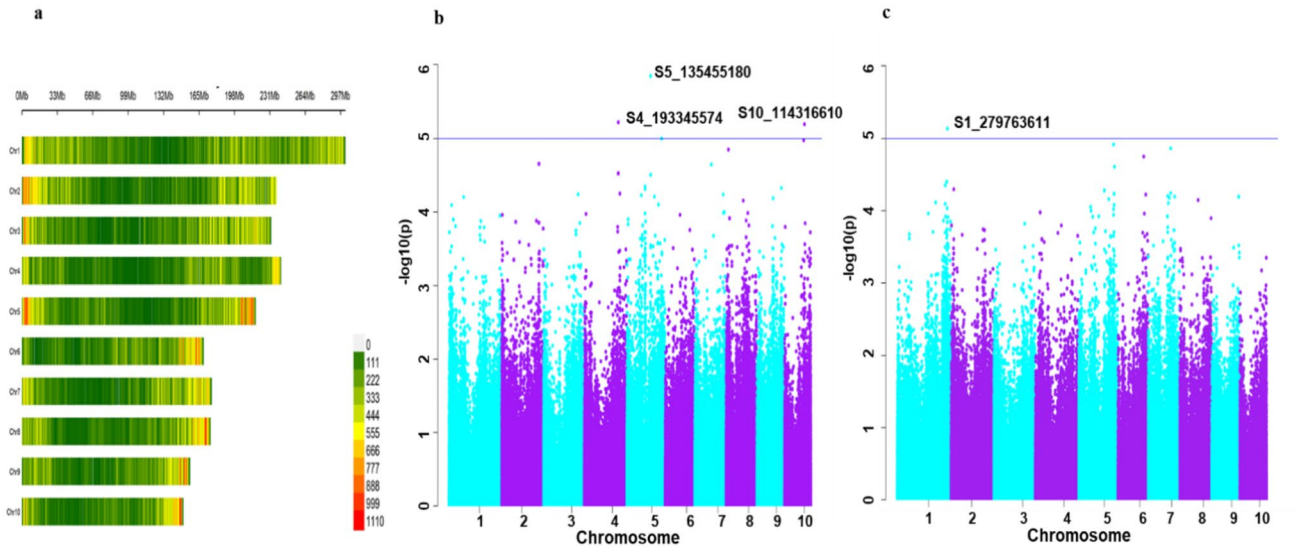

Fig. 4 Manhattan plots of HFF association analysis using the FARM-CPU and MLMM methods. The SNP density plot **a** shows marker density in each chromosome, where green to red indicates low to high marker density. The Manhattan plot **b** highlights the three most significant SNPs in chromosomes 4, 5, and 10 for HFF. The Manhattan plot **c** shows the most significant SNP in chromosome 1 for HMF. The Y-axis represents the − log10 *P*-value and the X-axis represents the chromosomes. The blue line represents the Simple M-corrected significance threshold

of 27.4 in 1991 and 26.2 in 1992, with the highest-yielding ears at 127 kernels. Differences between BS39 and BS39-SHGD-HILs with regard to HFF and HMF were significant (Table 1). These findings are consistent with our preliminary study conducted in 2022. One possible explanation for these results is the impact of the introgressed SHGD (QTL) not only on haploid male fertility (HMF) but also on haploid female fertility (HFF).

HFF and HMF heritability

Significant genetic variance and environmental effects ($p < 0.05$) were observed for both HFF and HMF. However, the heritability of these traits differed markedly. In line with findings from temperate germplasm by Kleiber et al. (2012), fertile haploids exhibited significant genetic variance ($p < 0.05$) with a high heritability of 0.8. In this study, HFF exhibited a high heritability of 0.8, indicating that phenotypic variation in HFF was largely attributable to genetic differences among individual lines. This heritability suggests good predictability in breeding programs focused on enhancing this trait (Visscher et al. 2008). In contrast, HMF revealed a moderately high heritability of 0.5. Ma et al. (2018) observed a similar heritability (0.65) for HMF in two genetic backgrounds. While this heritability confirms the importance of genetic factors, it also emphasizes the substantial influence of non-genetic factors on the expression of HMF (Riedelsheimer and Melchinger 2013). Nonetheless, the effect of the genotype was significant for HMF, showing the amenability to improve this trait through breeding. While the heritability of HMF is moderate (0.5) in our study, Molenaar et al. (2019), Trampe et al. (2020), and Foster et al. (2024) reported markedly higher entry-mean heritabilities for HMF ranging from 0.68–0.91. The difference in heritability estimates for HMF in our study compared to those reported by Molenaar et al. (2019), Trampe et al. (2020), and Foster et al. (2024) can be attributed to the lower number of plants used in our study. Molenaar et al. (2019) planted 40 plots, each with 60 plants, in two replicates and two environments. Similarly, Trampe et al. (2020) and Foster et al. (2024) planted 220 F_{23} families and 232 BC1F2 families, respectively, each laid out in four rows of 25 plants. In contrast, our study had 192 HILs in two environments, which were primarily planted to study the highly heritable trait plant height comparatively between HILs and their diploid parent lines. The higher plant numbers in the previous studies may have resulted in higher heritability estimates, as higher plant numbers reduce the impact of sampling error and environmental variability, enhancing the statistical power to detect genetic influences on a trait (Lynch and Walsh 1998).

HFF and HMF correlation

Haploid fertility can be influenced by meiotic or mitotic events. Aboobucker et al. (2023) demonstrated that mutations in two *Arabidopsis* genes, AtPS1 and AtJason/AtJAS, lead to atypical spindle formations during the second meiotic

division. Instead of forming the typical perpendicular orientation, these mutants produce parallel spindles, which result in dyad formation rather than the standard tetrad. This abnormality contributes to the production of unreduced gametes and consequently diploid (2n) pollen, impacting fertility in haploids. Wu et al. (2014a, b) demonstrated with flow cytometric analysis of the flag leaf that early doubling haploid plants were homozygous diploids, which resulted in performance comparable to diploid plants. Furthermore, a flow cytometric analysis of different leaves of line A427 haploids revealed a significant increase in diploid cells early in plant development (Yavuz et al. unpublished). Consequently, the presence of qshgd1 (Foster et al. 2024) likely acts during mitosis. In our study, the expectation that haploid plants (BS39-SHGD-HILs) with the qshgd1 would set more seed and have a higher chance to produce pollen, and that HMF and HFF would be positively correlated, was not supported by our findings. Despite the presence of qshgd1, no significant correlation between HFF and HMF was observed. Across the entire panel, we noted a weak negative correlation ($r = -0.03$, $p > 0.5$), indicating no direct relationship between these traits. This pattern persisted in subgroup analyses, with correlations of $r = -0.01$ for BS39-HILs and $r = 0.01$ for BS39-SHGD-HILs. The absence of correlation between HMF and HFF implies the potential influence of additional QTL besides qshgd1. The independence of these traits is likely due to additive genetic effects, where alleles at one locus exert their influence without interacting with alleles at other loci. This additive action suggests that improvements in one trait may not necessarily lead to improvements in the other (Hill 2010). This independence allows for the independent selection of HMF and HFF, which can be beneficial for breeding programs focused on each trait. However, it can pose challenges when the breeding objective includes simultaneous improvements in both traits.

Genomic regions and candidate genes of HFF

Within Chromosome 4, we identified three candidate genes, Zm00001eb198220, Zm00001eb198240, and Zm00001eb198300. The gene Zm00001eb198220, encoding RETINOBLASTOMA-RELATED1 (RBR) protein, is crucial for asymmetric cell division, stem cell maintenance, and the DNA damage response (DDR). It plays a key role in reproductive processes across both female and male gametes and influences embryo development (Desvoyes and Gutierrez 2020). In *Arabidopsis thaliana*, RBR is a central regulator of cell cycle progression, further emphasizing its significance across plant species (Zaragoza et al. 2024). The second candidate gene was Zm00001eb198240, which encodes the HORMA domain-containing protein. This protein plays a significant role in coordinating key events during cell division and ensuring genomic stability. One well-known example of a HORMA domain-containing protein is MAD2 (Mitotic Arrest Deficient 2), which is involved in the spindle assembly checkpoint and ensures accurate chromosome segregation during mitosis (Prince and Martinez-Perez 2022). The third gene, Zm00001eb198300, encodes Ankyrin repeat domain-containing protein 2A. *Arabidopsis* ankyrin repeat protein (AKRP), encoded by the gene AT5G66055, is a protein found within the plastids of Arabidopsis thaliana. It is believed to play a role in the differentiation and morphogenesis of plastids, the organelles responsible for tasks such as photosynthesis and pigment synthesis. Loss of function mutations in the AKRP gene leads to developmental arrest during embryo development, emphasizing its essential role in early plant development. The critical functions of these proteins suggest that their roles in maize may also be essential in influencing plant developmental processes, though further studies are needed to clarify the exact mechanisms. The involvement of RBR, MAD2, and AKRP in ensuring precise chromosome segregation and cell division suggests that disruptions in their function could lead to fertility issues due to the formation of defective gametes. This connection highlights the importance of these proteins in maintaining fertility by safeguarding the integrity of cell division processes.

Near the SNP (S5_135455180) identified on chromosome 5, two candidate genes, Zm00001eb237460 and Zm00001eb237470, were identified. Zm00001eb237460 encodes Myb-related protein 3R-1, a transcriptional regulator essential for controlling gene expression during the G2/M phase transition of the cell cycle in *Arabidopsis thaliana* (Haga et al. 2007). By binding to specific DNA sequences, the Myb proteins can modulate the activity of G2/M phase-specific genes, influencing processes such as mitosis and cytokinesis. Proper cell cycle regulation is crucial for cell proliferation and differentiation, which are fundamental processes in gametophyte development and overall fertility (Qu 2008). In maize, similar regulatory mechanisms are likely at play. The Myb-related protein 3R-1 could be critical in ensuring that cell division during gametophyte development proceeds accurately and efficiently, thereby contributing to the formation of viable pollen and ovules. Any disruptions in the cell cycle could lead to defects in gamete formation, ultimately impacting fertility. Furthermore, Zm00001eb237470, encodes Rop guanine nucleotide exchange factor 9 (RopGEF9), which includes a PRONE domain. This protein plays a significant role in cytoskeletal organization and polarized growth, processes that are crucial for pollen tube development in plants (Zhang and McCormick 2007). The cytoskeleton is integral to cell shape, division, and intracellular transport, all of which are essential for the growth and guidance of the pollen tube toward the ovule during fertilization. Effective pollen tube growth and

navigation are vital for successful fertilization in maize (Xu and Huan 2020). RopGEF9's role in regulating these processes suggests that it is essential for reproductive success. Disruptions in cytoskeletal organization or polarized growth could impair pollen tube development, reducing the efficiency of fertilization and subsequently affecting seed set and yield.

Three candidate genes were identified for the SNP (S10_114316610) identified on chromosome 10. The first gene, Zm00001eb421910, encodes Protein BIC1, which is known to regulate Brassinosteroids (BRs) signaling. BRs are essential growth-promoting steroid hormones that control a wide range of plant developmental processes, including cell elongation, and flowering in *Arabidopsis thaliana* (Yang et al. 2020). In maize, proper BRs signaling is likely crucial for the development of reproductive structures (Chen et al.2024). Disruptions in BRs pathways could lead to defects in these processes, thereby impacting fertility. The second gene, Zm00001eb421980, encodes an OTU-like cysteine protease. In maize, this protein has been implicated in pollen development, germination, and plant drought tolerance, highlighting its role in environmental adaptation (Li et al. 2023). Effective pollen development is vital for male fertility, as it directly influences the ability of pollen to fertilize ovules. Additionally, the ability to tolerate drought conditions ensures that reproductive processes can continue even under environmental stress, thereby maintaining fertility and yield. The third gene, Zm00001eb421990, encodes a Putative F-box protein. In *Arabidopsis thaliana*, F-box proteins like SLEEPY1 (SLY1) are positive regulators of gibberellin (GA) signaling. GA signaling influences various growth aspects, including seed germination, stem elongation, and flowering (Schwechheimer 2012). A loss of function in SLY1 results in phenotypes of GA insensitivity, such as dwarfism, reduced fertility, delayed flowering, and increased seed dormancy (Ariizumi et al. 2011). In maize, the regulation of GA signaling by F-box proteins is likely essential for normal reproductive development and fertility.

Genomic regions and candidate genes of HMF

In our genome-wide analysis, the SNP (S1_279763611) associated with HMF was found to be linked to a single gene, Zm00001eb058360, which encodes for the GNAT transcription factor (Putative N-acetyltransferase HLS1). This gene has been implicated in key developmental processes in *Arabidopsis thaliana*. HLS1 is reported to play an essential role in callus formation, which is a process essential for plant tissue culture and regeneration (Sakamota et al. 2022). The function of HLS1 in modulating histone acetylation levels suggests its involvement in the epigenetic regulation of gene expression, particularly during cell proliferation and differentiation. This mechanism is critical during callus induction, where cells must undergo significant physiological changes (Verdone et al. 2005). In *Arabidopsis*, HLS1's influence on histone acetylation emphasizes its broader role in genomic stability and gene expression, essential for effective cell division and differentiation (Liao et al. 2016). Given the similar foundational processes in maize, the association of HLS1 with HMF in our study suggests that modifications in histone acetylation could similarly impact male gametophyte development (Kumar et al. 2021). These findings open up potential avenues for enhancing fertility traits in maize through targeted manipulation of epigenetic regulators such as HLS1. Further investigations into HLS1's specific roles and mechanisms in maize will be crucial for validating these prospects and understanding how they can be applied to improve crop breeding techniques.

Limitations of the study and recommendations for future research

One limitation of this study is the relatively low number of haploid plants per isogenic line, which may have contributed to the moderate heritability estimates for HMF. Previous research has reported higher heritabilities, exceeding 0.6 (Molenaar et al. 2019; Trampe et al. 2020; Foster et al. 2024), suggesting that increasing the number of haploid plants per plot (e.g., from 20 to 40) could improve the precision of these estimates by minimizing environmental variation and sampling error. In this study, we focused on plot-to-plot variation by calculating HFF as the ratio between the total number of kernels per plot and the total number of haploid plants. Future studies should focus on plant-to-plant variation, calculating HFF on a per-ear basis to assess individual plant fertility. Additionally, to further explore the correlation between HFF and HMF, HFF could be expressed as a percentage, representing the ratio of haploid plants producing at least one kernel to the total number of haploid plants per plot. This approach could uncover relationships between HFF and HMF that were not observed in this study. Finally, the candidate genes identified in this study present a foundation for further research. Future studies should aim to functionally validate these genes using techniques such as gene editing or knockout experiments to confirm their roles in regulating fertility traits. This would deepen our understanding of the genetic mechanisms driving HMF and HFF and provide valuable tools for improving haploid fertility in maize breeding.

Conclusions

On the basis that highly heritable traits are expected to have a less complex genetic architecture, we can conclude that HFF holds a less complex genetic architecture

compared to HMF (Combs and Bernardo 2013). These findings highlight the contrasting genetic architectures of HFF and HMF and emphasize the need for tailored breeding approaches. For HFF, selection strategies can be more directly based on genetic markers due to their higher heritability. Meanwhile, improving HMF may require a more holistic approach that focuses on genetic selection and considers environmental and management optimizations to fully harness the potential of genetic improvements. These insights affirm the foundational role of genetics in influencing fertility traits and enhance our understanding of how these traits can be effectively manipulated through targeted breeding efforts.

Conducting a genome-wide association study for HFF and HMF traits holds promise for uncovering the genomic regions influencing these traits. Identifying these regions serves as a crucial step toward understanding the genetic architecture underlying both traits and paves the way for more targeted breeding in our subsequent novel rapid cycle breeding approach. This novel approach will integrate Genomic Selection (GS) and Spontaneous Haploid Genome Doubling (SHGD), enabling efficient recombination and seed increase within a single generation. One of the important steps in the rapid cycle breeding is the haploid by haploid crosses which will allow direct hybrid formation for the next breeding cycle. Thus, leveraging the knowledge from this association mapping will facilitate the success of the haploid by haploid crossing in the rapid cycle breeding.

Acknowledgements This research was supported in part by the U.S. Department of Agriculture, Agricultural Research Service (project number 5030-21000-073-00D) and by grants from the U.S. Department of Agriculture, National Institute of Food and Agriculture, OREI program (award numbers 2020-51300-32180 and 2021-51300-34896). The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Author contribution statement M.F, U.K.F, and T.L. designed the field experiments. M.F, U.K.F, P.S, A.M, and T.L. contributed to manuscript writing. M.F. analyzed data and performed GWAS. A.M. and M.F. performed candidate genes search. All authors read and approved the final manuscript.

Funding Funding for this work was provided by USDA's National Institute of Food and Agriculture (NIFA) Project, No. IOW04314, IOW01018, and IOW05510; NIFA award 2018-51181-28419; and the National Science Foundation under Grant No. DGE-1545453. 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.

Data availability Datasets generated and analyzed in this study are available from this GitHub repository: https://github.com/Mercyfakude/Data_Availability_Manuscript.

Declarations

Conflict of interest The authors declare they have no financial or non-financial interests to disclose.

References

Aboobucker SI, Jubery TZ, Frei UK, Chen YR, Foster T, Ganapathysubramanian B, Lübberstedt T (2022) Protocols for in vivo doubled haploid (DH) technology in maize breeding: From haploid inducer development to haploid genome doubling. In: Lambing Christophe (ed) Plant gametogenesis: methods and protocols. Springer US, New York, NY, pp 213–235. https://doi.org/10.1007/978-1-0716-2253-7_16

Aboobucker SI, Zhou L, Lübberstedt T (2023) Haploid male fertility is restored by parallel spindle genes in Arabidopsis thaliana. Nat Plants 9(2):214–218. https://doi.org/10.1038/s41477-022-01332-6

Ariizumi T, Lawrence PK, Steber CM (2011) The role of two F-box proteins, SLEEPY1 and SNEEZY, in arabidopsis gibberellin signaling. Plant Physiol 155:765–775. https://doi.org/10.1104/pp.110.166272

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

Boerman NA, Frei UK, Lübberstedt T (2020) Impact of spontaneous haploid genome doubling in maize breeding. Plants 9:369

Chaikam V, Gowda M, Nair SK, Melchinger AE, Boddupalli PM (2019) Genome-wide association study to identify genomic regions influencing spontaneous fertility in maize haploids. Euphytica Neth J Plant Breed 215(8):138. https://doi.org/10.1007/s10681-019-2459-5

Chaikam V, Willem M, Albrecht EM, Prasanna MB (2019b) Doubled haploid technology for line development in maize technical advances and prospects. Theor Appl Genet 132(12):3227–3243

Chalyk ST (1994) Properties of maternal haploid maize plants and potential application to maize breeding. Euphytica 79:13–18. https://doi.org/10.1007/BF00023571

Chang MT, Coe EH (2009) Doubled haploids. In: Kriz AL, Larkins BA (eds.) Molecular genetic approaches tomaize improvement. Biotechnology in Agriculture and Forestry, vol 63. Springer, Berlin, Heidelberg.https://doi.org/10.1007/978-3-540-68922-5_10

Chase SS (1969) Monoploids and monoploid-derivatives of maize (*Zea mays* L.). Botani Rev 35:117–168. https://doi.org/10.1007/BF02858912

Chen X, Hu X, Jiang J, Wang X (2024) Functions and mechanisms of brassinosteroids in regulating crop agronomic traits. Advance online publication, Plant and Cell Physiology. https://doi.org/10.1093/pcp/pcae044

Combs E, Bernardo R (2013) Accuracy of genome-wide selection for different traits with constant population size, heritability, and number of markers. The Plant Genome. https://doi.org/10.3835/plantgenome2012.11.0030

Cortes LT, Zhang Z, Yu J (2021) Status and prospects of genome-wide association studies in plants. The Plant Genome. https://doi.org/10.1002/tpg2.20077

De La Fuente GN, Frei UK, Lübberstedt T (2013) Accelerating plant breeding. Trends Plant Sci 18(12):667–672. https://doi.org/10.1016/j.tplants.2013.09.001

Dermail A, Mitchell M, Foster T, Fakude M, Chen Y-R, Suriharn K, Frei UK, Lübberstedt T (2024) Haploid identification in maize. Front Plant Sci 15:1378421. https://doi.org/10.3389/fpls.2024.137842

Desvoyes B, Gutierrez C (2020) Roles of plant retinoblastoma protein cell cycle and beyond. The EMBO J. https://doi.org/10.15252/embj.2020105802

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6(5):e19379. https://doi.org/10.1371/journal.pone.0019379

Geiger HH (2009) Doubled haploids. Maize handbook, vol 2. Springer Science & Business Media LLC, Stuttgart, Germany, pp 641–657

Gerdes JT, Behr CF, Coors JG, Tracy WL (1993) Compilation of North American maize breeding germplasm. Crop Sci Soc Am Madison. https://doi.org/10.2135/1993.compilationofnorthamerican.frontmatter

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS a high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9(2):e90346

Gupta M, Choudhary M, Kumar H, Kaswan V, Kaur Y, Choudhary JR, Yadav S (2022) Doubled haploid technology in maize (*Zea Mays*) status and applications. Indian J Agric Sci 92(3):283–291

Haga N, Kato K, Murase M, Araki S, Kubo M, Demura T, Suzuki K, Müller I, Voß U, Jürgens G, Ito M (2007) R1R2R3-Myb proteins positively regulate cytokinesis through activation of KNOLLE transcription in Arabidopsis thaliana. Plant J 50(5):699–709

Hale B, Ferrie AMR, Chellamma S, Samuel JP, Phillips GC (2022) Androgenesis-Based Doubled Haploidy: Past, Present, and Future Perspectives. Front Plant Sci 7(12):751230. https://doi.org/10.3389/fpls.2021.751230

Hill WG (2010) Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet 6(7):e1001008. https://doi.org/10.1371/journal.pgen.1000008

Kaler AS, Gillman JD, Beissinger T, Purcell LC (2020) Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. Front Plant Sci 10:1794. https://doi.org/10.3389/fpls.2019.01794

Kleiber D, Prigge V, Melchinger AE, Burkard F, San Vicente F, Palomino G, Gordillo GA (2012) Haploid fertility in temperate and tropical maize germplasm. Crop Sci 52:623–630. https://doi.org/10.2135/cropsci2011.07.0395

Kumar V, Thakur JK, Prasad M (2021) Histone acetylation dynamics regulating plant development and stress responses. Plant Mol Biol. https://doi.org/10.1007/s11103-021-01105-5

Li Y, Lin Z, Yue Y, Zhao H, Fei X, Lizhu E, Liu C, Chen S, Lai J, Song W (2021) Loss-of-function alleles of ZmPLD3 cause haploid induction in maize. Nat Plants 7(12):1579–1588

Li Y, Niu L, Zhou X, Liu H, Tai F, Wang W (2023) Modifying the expression of cysteine protease gene PCP affects pollen development, germination and plant drought tolerance in maize. Int J Molec Sci 24(8):7406. https://doi.org/10.3390/ijms24087406

Liao CJ, Lai Z, Lee S, Yun DJ, Mengiste T (2016) Arabidopsis HOOKLESS1 regulates responses to pathogens and abscisic acid through interaction with MED18 and acetylation of WRKY33 and ABI5 chromatin. Plant Cell 28(7):1662–1681. https://doi.org/10.1105/tpc.16.00105

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits, vol 1. Sinauer Associates, Sunderland, MA, pp 535–557

Ma H, Li G, Würschum T, Zhang Y, Zheng D, Yang X, Li J, Liu W, Yan J, Chen S (2018) Genome-wide association study of haploid male fertility in maize (*Zea mays* L.). Front Plant Sci. https://doi.org/10.3389/fpls.2018.00974

Melchinger AE, Würschum T (2010) Doubled haploid technology in maize breeding theoretical and practical considerations. Biotechnology in agriculture and forestry, vol 64. Springer, Berlin, Heidelberg, pp 139–154

Melchinger A, Schipprack W, Würschum T et al (2013) Rapid and accurate identification of in vivo-induced haploid seeds based on oil content in maize. Sci Rep 3(1):2129. https://doi.org/10.1038/srep02129

Mikel MA (2018) Improving Maize in Handbook of Maize Its Biology. Springer, Cham, pp 3–23

Molenaar WS, Schipprack W, Brauner PC et al (2019) Haploid male fertility and spontaneous chromosome doubling evaluated in a diallel and recurrent selection experiment in maize. Theor Appl Gen 132:2273–2284. https://doi.org/10.1007/s00122-019-03353-w

Prasanna BM (2012) Doubled haploid technology in maize breeding theory and practice in in vitro haploid production in higher plants. Springer, Berlin, Heidelberg, pp 181–204

Prince JP, Martinez-Perez E (2022) Functions and regulation of meiotic HORMA-domain proteins. Genes 13(5):777. https://doi.org/10.3390/genes13050777

Qu LJ (2008) Meiotic and mitotic cell cycle mutants involved in gametophyte development in Arabidopsis. Mol Plant 1(4):564–574. https://doi.org/10.1093/mp/ssn033

Ren J, Boerman NA, Liu R, Wu P, Trampe B, Vanous K, Frei UK, Chen S, Lübberstedt T (2020) Mapping of QTL and identification of candidate genes conferring spontaneous haploid genome doubling in maize (*Zea mays* L.). Plant Sci 293:110337. https://doi.org/10.1016/j.plantsci.2020.110337

Riedelsheimer C, Melchinger AE (2013) Optimizing the allocation of resources for genomic selection in one breeding cycle. Theor Appl Gen 126:2835–2848. https://doi.org/10.1007/s00122-013-2175-9

Sakamoto Y, Kawamura A, Suzuki T, Segami S, Maeshima M, Polyn S, De Veylder L, Sugimoto K (2022) Transcriptional activation of auxin biosynthesis drives developmental reprogramming of differentiated cells. Plant Cell 34(11):4348–4365. https://doi.org/10.1093/plcell/koac218

Santos IGD, Verzegnazzi AL, Edwards J, Frei UK, Boerman N, Zuffo LT, Pires LPM, de La Fuente G, Lübberstedt T (2022) Usefulness of temperate-adapted maize lines developed by doubled haploid and single-seed descent methods. Theor Appl Genet 135(6):1829–1841. https://doi.org/10.1007/s00122-022-04041-8

Schwechheimer C (2012) Gibberellin signaling in plants - the extended version. Front Plant Sci 2:107. https://doi.org/10.3389/fpls.2011.00107

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples), Biometrika 52(3–4):591–611. https://doi.org/10.1093/biomet/52.3-4.591

Shiferaw B, Prasanna BM, Hellin J et al (2011) Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. Food Sec 3:307–327. https://doi.org/10.1007/s12571-011-0140-5

Sugihara N, Higashigawa T, Aramoto D, Kato A (2013) Haploid plants carrying a sodium azide-induced mutation (fdr1) produce fertile pollen grains due to first division restitution (FDR) in maize (Zea mays L.). Theor Appl Genet 126:2931–2941

Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, Yan J (2016) GAPIT version 2 an enhanced integrated tool for genomic association and prediction. The Plant Genome 9(2):1–9. https://doi.org/10.3835/plantgenome2015.11.0120

Trampe B, Santos IGD, Frei UK, Ren J, Chen S, Lübberstedt T (2020) QTL mapping of spontaneous haploid genome doubling using genotyping-by-sequencing in maize (*Zea mays* L.). Theor Appl Gen 133(7):2131–2140. https://doi.org/10.1007/s00122-020-03601-8

Trentin HU, Frei UK, Lübberstedt T (2020) Breeding maize maternal haploid inducers. Plants 9(5):614. https://doi.org/10.3390/plants9050614

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Verdone L, Caserta M, Di Mauro E (2005) Role of histone acetylation in the control of gene expression. Biochem Cell Biol 83(3):344–353. https://doi.org/10.1139/o05-041

Verzegnazzi AL, dos Santos IG, Krause MD 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. https://doi.org/10.1007/s00122-021-03780-8

Visscher P, Hill W, Wray N (2008) Heritability in the genomics era — concepts and misconceptions. Nat Rev Gen. https://doi.org/10.1038/nrg2322

Wu P, Li H, Ren J, Chen S (2014) Mapping of maternal QTLs for in vivo haploid induction rate in maize (*Zea mays* L.). Euphytica; Neth J Plant Breed 196(3):413–421

Wu P, Ren J, Li L, Chen S (2014b) Early spontaneous diploidization of maternal maize haploids generated by in vivo haploid induction. Euphytica Neth J Plant Breed 200(1):127–138

Xu L, Najeeb U, Tang GX, Gu HH, Zhang GQ, He Y, Zhou WJ (2007) Haploid and Doubled Haploid Technology. In: Advances in Botanical Research. Advances in Botanical Research. Elsevier, pp. 181–216

Xu Y, Huang S (2020) Control of the Actin Cytoskeleton Within Apical and Subapical Regions of Pollen Tubes. Front Cell Dev Biol. https://doi.org/10.3389/fcell.2020.614821

Yang Z, Yan B, Dong H, He G, Zhou Y, Sun J (2020) BIC1 acts as a transcriptional coactivator to promote brassinosteroid signaling and plant growth. EMBO J. https://doi.org/10.15252/embj.202004615

Zaragoza JZ, Klap K, Heidstra R, Zhou W, Scheres B (2024) The dual role of the retinoblastoma-related protein in the DNA damage response is coordinated by the interaction with LXCXE-containing proteins. Plant J. https://doi.org/10.1111/tpj.16665

Zhang Y, McCormick S (2007) A distinct mechanism regulating a pollen-specific guanine nucleotide exchange factor for the small GTPase Rop in Arabidopsis thaliana. Proc Natl Acad Sci 104(47):18830–18835. https://doi.org/10.1073/pnas.0705874104