Journal of Applied Genetics
https://doi.org/10.1007/s13353-026-01051-0

PLANT GENETICS • ORIGINAL PAPER

Genetic mapping of fertility restoration (*Rf*) genes and development-validation of linked SNPs for A_1 cytoplasm of sorghum (*Sorghum bicolor* (L.) Moench)

Narkhede Gopal Wasudeo[1] · **Laavanya Rayaprolu**[3,4] · **Usha Kiranmayee KNS**[3] · **SP Mehtre**[2] · **Hirakant Kalpande**[2] · **SP Deshpande**[3,5] · **Shabir H. Wani**[5,6]

Received: 24 May 2025 / Revised: 31 January 2026 / Accepted: 19 February 2026

Abstract

Sorghum (*Sorghum bicolor* L. Moench) is one of the most important cereal crops in which cytoplasmic genetic male sterility (CGMS) is widely exploited for hybrid breeding. Commercial sorghum hybrids globally are predominantly based on the A1 cytoplasmic genetic male sterility (CGMS) system because of its operational ease and cytoplasmic stability. To map fertility restoration (*Rf*) loci, hybrids were developed by crossing the CMS line 296 A with a recombinant inbred line (RIL) population derived from the cross 296B × IS18551, which segregates for fertility restoration genes. The RIL population consisted of advanced-generation lines ($F_{6:7}$), ensuring high homozygosity suitable for genetic mapping. In addition, F_2 and BC_1F_1 populations were developed from selected fertile RIL-derived hybrids to validate fertility restoration loci.

The RIL population was genotyped using a genotyping-by-sequencing (GBS) approach and employed for genome-wide genetic mapping of fertility restoration genes. Based on significant associations, a set of 19 single nucleotide polymorphism (SNP) markers associated with fertility restoration was selected and converted into Kompetitive Allele-Specific PCR (KASP) assays. These 19 KASP markers were first validated in the RIL population and subsequently evaluated using a diverse panel of 150 sorghum lines, comprising 72 maintainer (B) lines and 78 restorer (R) lines.

Of the 19 validated KASP markers, six markers were associated with the major fertility restoration locus *Rf1* on chromosome SBI-08, while the remaining markers were linked to loci located on chromosomes SBI-01, SBI-03, SBI-07, SBI-09, and SBI-10. The results clearly demonstrate the ability of specific SNP markers to discriminate fertility restoration responses between B- and R-lines. The identified polymorphic SNP markers provide robust molecular tools for fast-tracking line conversion and marker-assisted selection in sorghum hybrid breeding programs. Overall, this study enhances the current understanding of fertility restoration genetics in sorghum by identifying key *Rf* loci, improving hybrid breeding efficiency, and providing new insights into cytoplasm–nuclear interactions governing male fertility restoration.

Keywords CGMS · SNP · KASP · GWAS · *Rf*

Communicated by: Izabela Pawłowicz

✉ Shabir H. Wani
shabirhwani@skuastkashmir.ac.in

[1] Department of Genetics and Plant Breeding, School of Agriculture, SR University, Warangal 506371, Telangana, India

[2] Vasantrao Naik Marathwada Krishi Vidyapeeth (VNMKV), Parbhani 431402, Maharashtra, India

[3] International Crops Research Institute for Semi-Arid Tropics (ICRISAT), Patancheru 502324, Telangana, India

[4] Wageningen University & Research, Droevendaalsesteeg 4, Wageningen 6708 PB, Netherlands

[5] Hytech Seed India Pvt. Ltd. Medchal Mandal, Medchal Malkajgiri District, Hyderabad, Telangana 501 401, India

[6] National Seed Project Laboratory, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Shalimar, Srinagar 190025, J&K, India

Introduction

Sorghum (*Sorghum bicolor* L. Moench) is the fifth most important cereal crop globally and a staple food for over 500 million people in the semi-arid and arid regions of Africa and Asia. It is highly valued for its resilience to drought, high temperatures, and low soil fertility, making it a vital crop under changing climatic conditions. Beyond food security, sorghum is gaining increasing importance as a source of bio-energy, fodder, and industrial products, thereby enhancing its global agricultural and economic significance.

In developed countries such as the United States, the exploitation of heterosis through F_1 hybrid cultivars forms the cornerstone of grain sorghum breeding programs. The cytoplasmic–nuclear male sterility (CMS) system, first elucidated in sorghum by Stephens and (Holland 1954), provides a reliable and cost-effective platform for large-scale hybrid seed production. This system relies on the interaction between a sterile cytoplasm, such as the widely used A1 or "*milo*" cytoplasm, and dominant nuclear fertility restorer genes of kafir origin. In the absence of these restorer genes, plants remain male sterile while retaining normal female fertility. The successful development and deployment of restorer lines (R-lines) carrying dominant restorer-of-fertility (*Rf*) genes have enabled the widespread commercial adoption of hybrid sorghum.

Despite the extensive use of CMS-based hybrid breeding systems, the genetic and molecular mechanisms underlying fertility restoration in sorghum remain incompletely understood. Classical genetic studies indicate that fertility restoration in the A1 cytoplasm is governed by two to three major *Rf* genes, along with the influence of several minor modifier loci (Maunder and Pickett 1959; Miller and Pickett 1964). Further complicating this genetic framework, environmental factors—particularly temperature during flowering—have a pronounced effect on fertility expression (Downs and Marshall 1971; Brooking 1976, 1979). As a result, fertility restoration is not a strictly binary trait but rather exhibits a continuous phenotypic spectrum ranging from complete sterility, characterized by pale and non-dehiscent anthers, to full restoration with viable pollen production.

At the molecular level, pentatricopeptide repeat (PPR) proteins, which play essential roles in RNA processing within mitochondria and chloroplasts, have emerged as key determinants of fertility restoration across several crop species (Lurin et al. 2004; Schmitz-Linneweber and Small 2008). Nearly all *Rf* genes cloned to date belong to this large PPR gene family, with the notable exception of *Rf2* in maize (Saha et al. 2007). These RNA-binding proteins are thought to suppress the expression of aberrant or chimeric mitochondrial open reading frames associated with male sterility, thereby restoring normal pollen development and fertility.

However, the limited number of functionally validated *Rf* genes in sorghum, coupled with inconsistent fertility expression across environments, presents a significant bottleneck for sorghum breeders. At present, the restorer status of breeding lines can be reliably confirmed only through test crosses with male-sterile lines, followed by labor-intensive phenotypic evaluation of F_1 progenies across multiple environments. This conventional approach is time-consuming, resource-intensive, and restricts the efficient introgression of restorer alleles from diverse or exotic germplasm sources.

To date, research efforts in sorghum have largely focused on classical genetic analyses, phenotypic evaluations, and partial molecular characterization of the CMS–restorer system. Although several candidate *Rf* regions have been proposed through quantitative trait locus (QTL) mapping and genome-wide association studies, only a limited number have been validated at the functional or marker level. There is therefore a critical need to identify and characterize candidate *Rf* genes, particularly those encoding PPR proteins, and to elucidate the regulatory mechanisms and environmental modulation governing fertility restoration. Such advances would greatly facilitate the development of robust molecular markers for early identification of restorer lines and accelerate marker-assisted breeding of stable, climate-resilient sorghum hybrids.

To address these challenges, the present study was undertaken to identify and characterize the genetic factors controlling fertility restoration in sorghum. Particular emphasis was placed on the discovery of restorer-of-fertility (*Rf*) loci, especially those belonging to the pentatricopeptide repeat (PPR) gene family, which are known to play pivotal roles in male fertility restoration across diverse crop species. In addition, the study aimed to evaluate the phenotypic expression of fertility restoration in F_1 hybrids under contrasting environmental conditions to better understand the extent to which environmental variables influence the effectiveness of restorer alleles. Finally, associations between fertility restoration traits and molecular markers were established to support the development of reliable molecular tools for early identification of restorer lines through marker-assisted selection.

By integrating genome-wide molecular approaches with rigorous field-based phenotyping, this study seeks to advance the understanding of the CMS–Rf system in sorghum and contribute to the development of more efficient and precise hybrid breeding strategies.

Materials and methods

Plant materials

A recombinant inbred line (RIL) population comprising 238 advanced-generation lines ($F_{6:7}$) derived from the cross 296B × IS18551 was used along with the A1 cytoplasm-based CMS line 296 A corresponding to an expected homozygosity of ~98.4%, making it suitable for high-resolution genetic mapping. F_1 hybrids were developed by crossing each RIL with the CMS line (296 A) and evaluated during the rainy (R) and post-rainy (PR) seasons of 2016 for percent seed set and pollen fertility score at the All India Coordinated Research Improvement Project on Soybean, Vasantrao Naik Marathwada Krishi Vidyapeeth, Parbhani, Maharashtra, India.

The experimental material was sown in an alpha lattice design with randomized plots and evaluated during both seasons in single rows of 3 m length, with two replications and a spacing of 45 × 15 cm. A basal dose of 100 kg di-ammonium phosphate (18% N and 46% P) was applied at sowing, followed by a top dressing of 100 kg urea (46% N) four days after thinning. The plots were irrigated at intervals of 7–10 days, and all recommended agronomic practices were followed to ensure optimal crop growth. Observations were recorded on a whole-plot basis for percent seed set and pollen fertility score using standard procedures (Jordan et al. 2010, 2011).

For estimation of percent seed set, prior to flowering, three representative panicles from each plot in each replication were bagged to prevent contamination by foreign pollen. Approximately 40 days after flowering, the Percent seed set was estimated by counting the number of filled grains relative to the total number of florets on each bagged panicle. Three panicles per plot were evaluated and mean values were used for analysis.

Pollen fertility was assessed at anthesis using a qualitative visual rating system. The pollen fertility score was assigned on a 1–9 scale, developed by DEEDI, based on visual evaluation of anther size, color, and morphology. The characteristics corresponding to each score are provided in Supplementary Table S1 (Jordan et al. 2011). Although pollen fertility score is ordinal in nature (1–9 scale), exploratory analysis showed an approximately continuous distribution with homogeneous variance. Therefore, consistent with previous quantitative genetic and GWAS studies in sorghum, this trait was treated as a quasi-continuous variable for linear mixed model and association analyses. Fully fertile F_1 hybrids were selected and advanced to generate F_2 populations and were simultaneously backcrossed to the CMS line 296 A to develop BC_1F_1 populations. These segregating populations were evaluated at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, for percent seed set. Percent seed set in individual F_2 and BC_1F_1 plants was recorded using the same standardized procedures described above.

The overall experimental workflow, including RIL development, F_1 phenotyping, validation using F_2 and BC_1F_1 populations, genome-wide association analysis, and KASP marker development and validation, is illustrated in Fig. 1.

Genotyping

The sorghum RIL population, along with the F_2 and BC_1F_1 populations, was genotyped using the genotyping-by-sequencing (GBS) methodology. GBS libraries were constructed using the ApeKI restriction enzyme, followed by single nucleotide polymorphism (SNP) discovery and filtering. The TASSEL 5 (Trait Analysis by Association, Evolution and Linkage) GBS v2 pipeline (Glaubitz et al. 2014) was used to extract high-quality and informative SNPs from the raw sequencing data. Sequence tags were aligned to the *Sorghum bicolor* reference genome version 3 (McCormick et al. 2018) using the Burrows–Wheeler Aligner (BWA) tool (Li and Durbin 2009).

After quality filtering, a total of 37,245 SNPs were retained for the RIL population and 100,887 SNPs for the F_2 and BC_1F_1 populations, applying a minor allele frequency (MAF) threshold of > 1% and allowing < 50% missing data. Missing genotypic data were imputed using the default imputation procedure implemented in the TASSEL GBS pipeline prior to association analysis. The physical positions of SNP markers were determined based on the reference genome annotation using standard procedures implemented in Phytozome (Goodstein et al. 2012).

Genome-wide association analysis

The genome-wide association study (GWAS) was conducted using best linear unbiased predictors (BLUPs) estimated separately for each genotype–environment combination. BLUP values were calculated using GenStat software by fitting a linear mixed model in which environment (season) was treated as a fixed effect, while replication, block nested within replication, and genotype were considered random effects.

GWAS analyses were performed in the recombinant inbred line (RIL), F_2, and BC_1F_1 populations using multiple statistical models, including the generalized linear model (GLM), mixed linear model (MLM), multi-locus mixed model (MLMM), and the Fixed and Random Model Circulating Probability Unification (FarmCPU). Among these, the FarmCPU model implemented in the R statistical environment was selected for final inference because of its

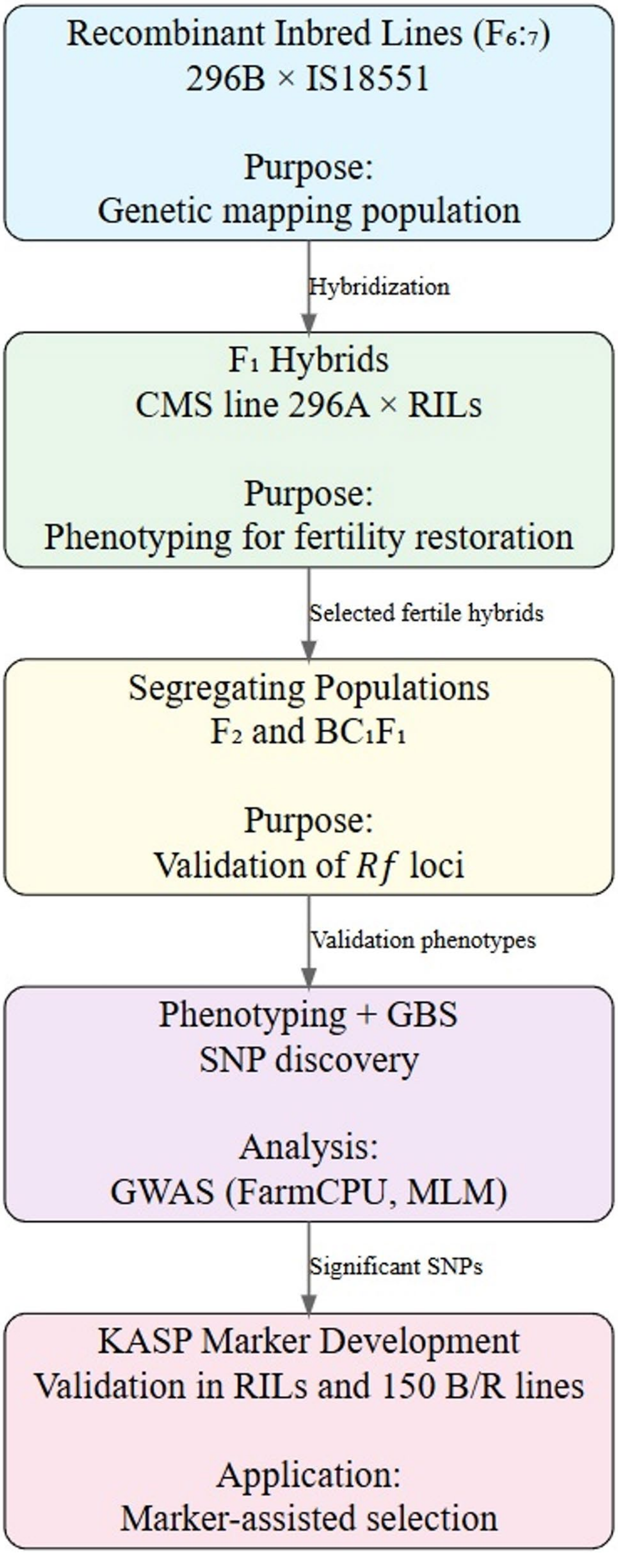

Fig. 1 Experimental workflow showing population development, phenotyping, GWAS, and KASP marker validation for fertility restoration in sorghum

superior ability to control both false positives and false negatives through iterative fitting of fixed and random effects, thereby improving statistical power and result consistency (Liu et al. 2016).

Although a Bonferroni-corrected threshold is highly conservative, it was initially applied to control for multiple testing. However, because strict Bonferroni correction can reduce the detection of biologically meaningful associations in complex traits, a relaxed significance threshold was adopted for candidate locus discovery, consistent with previous GWAS studies in sorghum. SNPs exceeding the predefined significance threshold ($P \leq 1 \times 10^{-4}$) were considered significant and were used consistently for tabulation (Tables 2, 3, 4 and 5) and manhattan plots. Functional annotation of significant SNPs was performed using the snpEff software package (Cingolani et al. 2012). Pairwise linkage disequilibrium (LD) analysis in the RIL population showed rapid LD decay, with r^2 values declining below 0.2 within a short physical distance, indicating limited long-range LD. Linkage disequilibrium (LD) among SNP markers was not explicitly pruned prior to GWAS because the FarmCPU model inherently accounts for marker dependencies through iterative fitting of fixed and random effects. In addition, the use of a biparental recombinant inbred line (RIL) population substantially reduces long-range LD compared with natural populations, as reported in earlier sorghum GWAS studies.

KASP marker development

The SNP markers identified in the present study were converted into Kompetitive Allele-Specific Polymerase Chain Reaction (KASP) assays. For each SNP, flanking DNA sequences were used to design allele-specific primers, and details of primer sequences and PCR conditions are provided in the Supplementary Table. All KASP assays were established and performed on a Roche LightCycler® 480 II platform using a total reaction volume of 10 µL and the KASP Master Mix with low ROX supplied by LGC Genomics (www.lgcgroup.com), following the manufacturer's recommended protocols (Kante et al. 2018).

KASP marker validation

The recombinant inbred lines (RILs) were genotyped using the validated KASP markers, and the phenotypic variance explained by each marker was evaluated using a linear regression model with fertility score as the dependent variable. This analysis enabled assessment of the individual contribution of each SNP marker to fertility restoration phenotypes within the RIL population.

In addition to validation within the RIL population, the KASP markers were further evaluated using a diverse panel of 150 sorghum lines, previously classified as fertility restorers (R-lines, $N = 78$) and maintainers (B-lines, $N = 72$). This independent validation was performed to assess the discriminatory power and practical utility of the markers for differentiating restorer and maintainer genotypes.

Results

Phenotypic variation

A total of 238 F_1 hybrids, developed by crossing recombinant inbred lines (RILs) derived from the cross 296B × IS18551 with the CMS line 296 A as the female parent, were evaluated along with two parental hybrids during the rainy (R) and post-rainy (PR) seasons of 2016 at Vasantrao Naik Marathwada Krishi Vidyapeeth, Parbhani, Maharashtra, India. Fertility restoration was assessed by recording key traits, including seed set percentage, pollen fertility score, and overall fertility percentage.

Mean performance of fertility restoration traits

The mean seed set percentage was 87.15% in the post-rainy season (PR2016), whereas it was 79.63% in the rainy season (R2016). The mean pollen fertility score was 8.16 in R2016 and 8.11 in PR2016. Similarly, the mean fertility percentage was 80.12% in R2016 and 82.45% in PR2016.

A higher magnitude of both genotypic and phenotypic coefficients of variation was observed for pollen fertility score in the rainy season (58.50% and 65.05%, respectively) as well as in the post-rainy season (53.95% and 62.04%, respectively), indicating substantial variability for this trait. In the rainy season, the highest heritability was recorded for fertility percentage (87%), followed by seed set percentage (85%) and pollen fertility score (80%). In contrast, during the post-rainy season, the maximum heritability was observed for seed set percentage (94%), followed by pollen fertility score (75%) (Table 1; Supplementary Tables S2 and S3).

The segregating F_2 and BC_1F_1 populations exhibited wide variation for seed set, with values ranging from complete sterility (0%) to full fertility (100%) at the individual plant level (Supplementary Table S4).

Genome-wide association mapping

Genome-wide association analysis was performed for fertility restoration traits using genotype–environment combinations analyzed independently. To control for multiple testing, a Bonferroni-based significance threshold was initially applied; however, the conventional Bonferroni correction ($P < 1 \times 10^{-6}$) proved overly stringent and failed to detect biologically meaningful associations for these complex traits. Consequently, a relaxed threshold of $P < 1 \times 10^{-4}$ was adopted for the identification of significant SNPs across all traits in the present study. Functional annotation of significant SNPs and associated candidate genes was conducted using publicly available genome annotation resources derived from the reference sequence assembly.

Genome-wide association mapping revealed several genomic regions significantly associated with fertility restoration traits across seasons and populations. During the rainy season (R2016), GWAS conducted on the RIL population identified one significant SNP (S9_46616891) associated with seed set percentage. Association analyses for pollen fertility score across the rainy (R2016) and post-rainy (PR2016) seasons identified a total of five significant SNPs. In addition, one significant SNP was detected for fertility percentage during the post-rainy season (PR2016) (Tables 2 and 3) (Figs. 2, 3 and 4).

Further association analyses conducted on segregating populations demonstrated the robustness of the detected loci. GWAS performed on the F_2 and BC_1F_1 populations identified five and two significant SNPs, respectively, associated with seed set percentage (Tables 4 and 5) (Figs. 5 and 6).

Table 1 Mean and genetic parameters for fertility restoration traits of sorghum RIL population in R and PR2016 at VNMKV, Parbhani, Maharashtra, India

	Traits	Mean	Genotypic Variance (VG)	Phenotypic Variance (VP)	Genotypic Coefficient of Variation (GCV)	Phenotypic Coefficient of Variation (PCV)	Heritability (%)	Genetic Advance	Genetic Advance % of mean
PR2016	Seed Setting Percent	87.15	0.03	0.03	13.29	13.67	94.0	0.32	26.53
	Pollen Fertility Score	8.11	0.02	0.03	53.95	62.04	75.0	0.27	95.29
	Fertility Percent	82.45	0.01	0.02	12.56	15.79	62.0	0.17	20.19
R2016	Seed Setting Percent	79.63	0.07	0.08	24.46	25.61	85.0	0.49	44.78
	Pollen Fertility Score	8.16	0.03	0.04	58.50	65.05	80.0	0.31	106.85
	Fertility Percent	80.12	0.03	0.03	21.05	22.51	87.0	0.32	40.29

Table 2 SNPs and candidate genes associated with fertility restoration traits in rainy season (R2016)

SN	Year	SNP	Chr.	Pos.	Allele	P value	Marker effect	Gene	Annotation
				Seed setting percent					
1	Rainy 2016	S2_13025961	2	13,025,961	A/G	4.71E-05	8.88		Intergenic
2		S2_13025955	2	13,025,955	C/A	7.08E-05	-8.72		Intergenic
3		S6_10452341	6	10,452,341	A/T	9.08E-05	8.51		Intergenic
4		S6_35994961	6	35,994,961	A/T	6.82E-05	-12.11		Intergenic
5		S9_46616891	9	46,616,891	G/A	3.81E-05	-11.80	Sobic.009G119000.1	Tetratricopeptide repeat containing protein, putative, expressed.
6		S10_6103308	10	6,103,308	C/T	7.81E-05	-33.50		Intergenic
				Pollen fertility score					
7		S1_62663034	1	62,663,034	A/G	1.52E-05	-0.16	Sobic.001G338900.1	Cytochrome p450, putative, expressed
8		S3_31348794	3	31,348,794	A/G	5.81E-05	-0.23		Intergenic
9		S4_10905289	4	10,905,289	T/C	7.54E-05	-0.21	Sobic.004G111100.1	Similar to Os02g0250400 protein
10		S4_10905290	4	10,905,290	C/T	7.54E-05	0.21	Sobic.004G111100.1	Similar to Os02g0250400 protein
11		S4_49818461	4	49,818,461	A/T	1.45E-05	0.13	Sobic.004G157100.1	Similar to putative uncharacterized protein
12		S8_55189444	8	55,189,444	C/G	1.01E-05	-0.16	Sobic.008G127100.1	CTP synthase, putative, expressed
13		S8_60934182	8	60,934,182	C/G	2.24E-05	0.18	Sobic.008G175032.1	Leucine Rich Repeat family protein
				Fertility percent					
14		S3_16936262	3	16,936,262	G/A	2.29E-05	-0.06		Intergenic
15		S9_49554151	9	49,554,151	A/C	49,554,151	0.06	Sobic.009G138500.1	DUF260 domain containing protein, putative, expressed

Table 3 SNPs and candidate genes associated with fertility restoration traits in post-rainy season (PR2016)

SN	Year	SNP	Chr.	Pos.	Allele	P value	Marker effect	Gene	Annotation
				Seed setting percent					
1	Post-rainy 2016	S1_18252916	1	18,252,916	C/T	1.63E-05	-2.93	--	Intergenic
2		S2_75152585	2	75,152,585	C/Gd	4.14E-05	-2.39	Sobic.002G400800.1	Protein of unknown function, DUF288
3		S2_76499055	2	76,499,055	A/C	1.04E-05	2.90	Sobic.002G417700.1	DUF623 domain containing protein, expressed.
4		S9_54956455	9	54,956,455	C/T	4.58E-05	3.22	Sobic.009G199500.1	Ribosomal protein L27, putative expressed.
				Pollen fertility score					
5		S3_24003464	3	24,003,464	C/T	1.93E-07	0.28	--	Intergenic
6		S3_57215830	3	57,215,830	C/G	6.97E-05	0.14	Sobic.003G233300.1	Expressed protein
7		S4_54186761	4	54,186,761	C/G	8.44E-07	-0.40	Sobic.004G189900.1	1-associated receptor kinase 1 precursor, putative, expressed
8		S7_1799521	7	1,799,521	A/C	2.50E-08	-0.29	Sobic.007G018900.1	RNA recognition motif containing protein, putative, expressed
9		S8_15472384	8	15,472,384	T/G	5.49E-06	-0.23	--	Modifier
10		S10_52241482	10	52,241,482	G/C	9.11E-05	0.16	Sobic.010G182800.1	Lectin like receptor kinase 1, putative, expressed
				Fertility percent					
11		S4_1921447	4	1,921,447	C/G	8.06E-05	0.03	Sobic.004G023600.1	Expressed protein (DUF1677)
12		S9_4248350	9	4,248,350	G/A	4.91E-06	0.05	Sobic.009G044400.1	1-aminocyclopropane-1-1Carboxylate oxidase, putative, expressed

Development and verification of KASP assays

In the present study, a set of 31 SNPs associated with fertility restoration was selected for validation using the Kompetitive Allele-Specific PCR (KASP) methodology. These 31 SNPs comprised 13 SNPs identified through genome-wide association analysis in the present study and 18 SNPs previously reported in the literature for fertility restoration in sorghum. The complete set of 31 KASP assays was used to screen a panel of 94 samples, which included four pairs of A-, B-, and R-line parents, six F_1 hybrids, one F_2 population, and two BC_1F_1 populations (Supplementary Table S5).

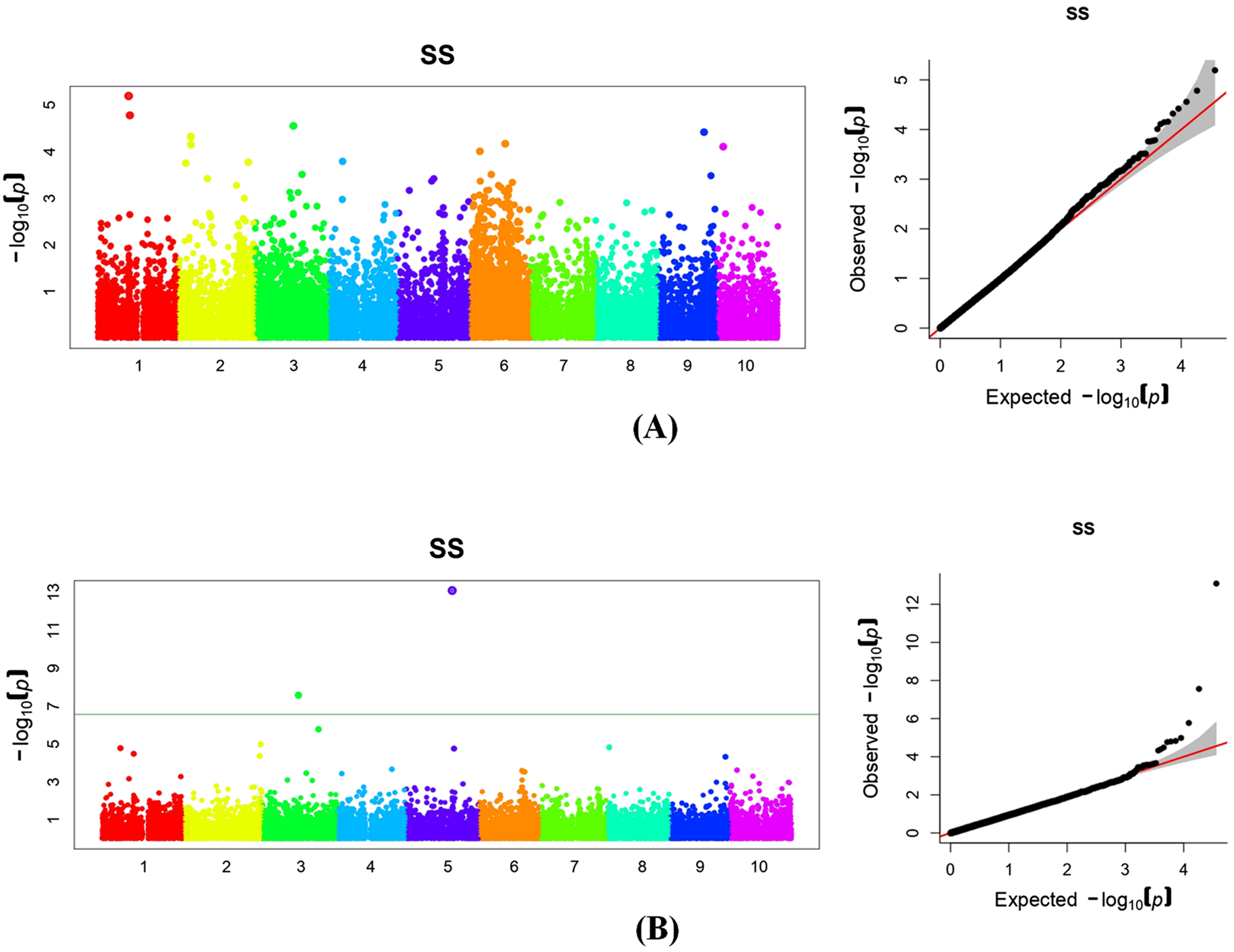

Fig. 2 Manhattan and QQ plots of GWAS for the seed setting percent in sorghum RIL population for R2016 (**A**) PR2016 (**B**)

Of the 31 KASP genotyping assays evaluated, 12 assays (38.7%) were uncallable across the 94 samples and were excluded from further analysis. Among the remaining 19 assays (61.3%), five SNPs were monomorphic across the tested panel. These monomorphic SNPs were retained for exploratory purposes to assess whether polymorphism might be detected in future or broader germplasm sets. Consequently, a total of 19 KASP assays were selected for downstream analyses in this study.

Of the 19 selected KASP assays, 14 produced reliable and reproducible amplification patterns, characterized by well-defined fluorescence clustering consistent with high-quality KASP reactions. Co-dominant polymorphic SNP assays generated three distinct clusters corresponding to the two homozygous classes and the heterozygous class, whereas dominant but polymorphic markers produced two clear clusters. In contrast, monomorphic SNP assays displayed a single homozygous cluster. Representative clustering patterns are illustrated in Fig. 7, and detailed assay performance is provided in Supplementary Table S6.

Validation of KASP assays

A set of 19 KASP markers was validated using 213 individuals from the recombinant inbred line (RIL) population developed from the cross 296B × IS18551. In this validation analysis, 11 KASP markers clearly differentiated the parental lines of the RIL population and consistently discriminated among individual RILs, demonstrating their robustness and reliability.

Among these 11 validated SNP markers, six SNPs—S8_653850, S8_912985, S8_912987, S8_913023, S8_62486248, and S8_60934182—were mapped to chromosome SBI-08 and were linked to the major fertility restoration locus *Rf1*. The remaining five SNPs were mapped to different chromosomes, namely S1_62663034 on

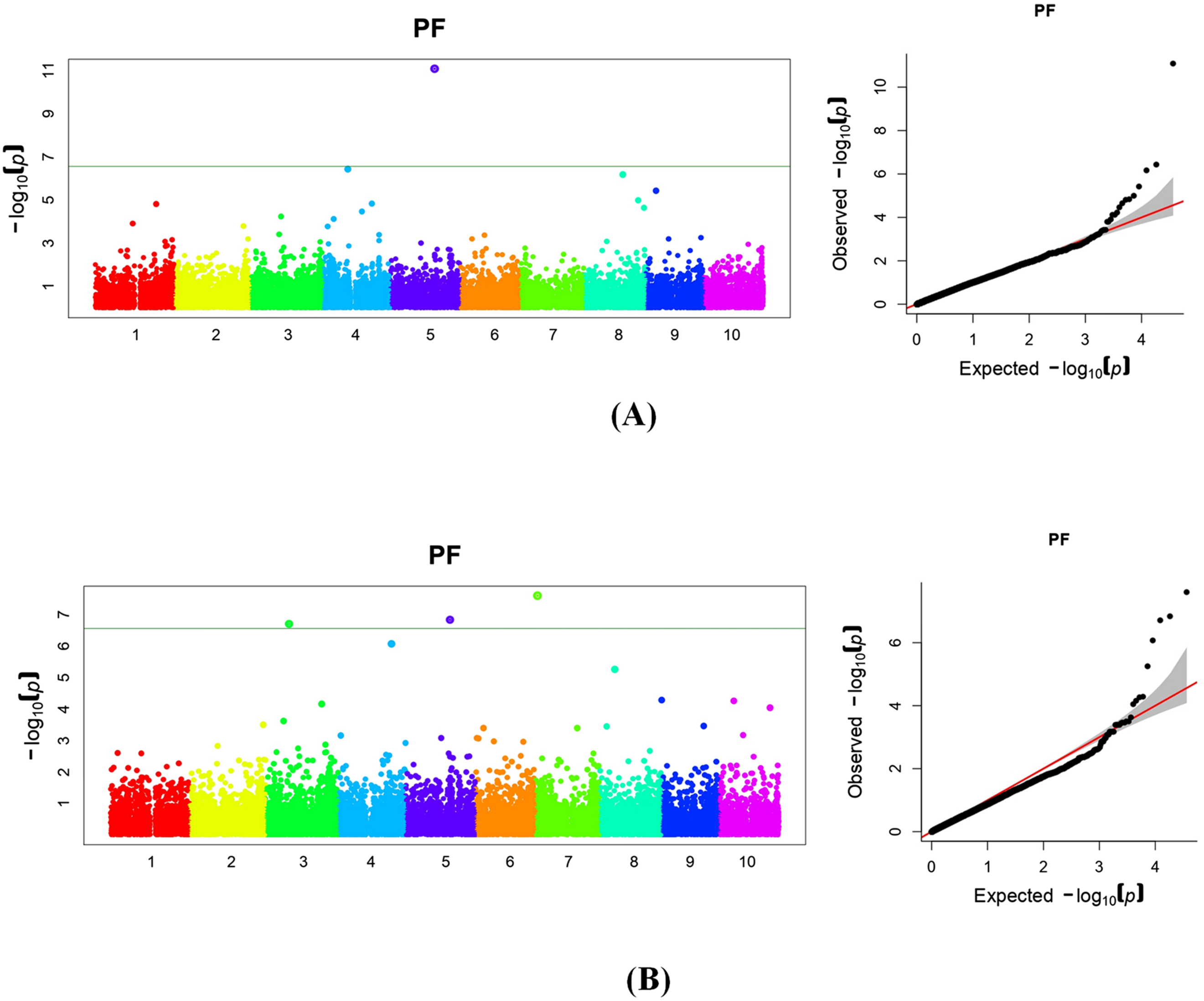

Fig. 3 Manhattan and QQ plots of GWAS for the pollen fertility percent in sorghum RIL population for R2016 (**A**) PR2016 (**B**)

chromosome SBI-01, S7_1799521 on chromosome SBI-07, S9_46616891 on chromosome SBI-09, and S10_52241482 on chromosome SBI-10.

To further evaluate the practical utility of these markers, the complete set of 19 KASP assays was screened across a diverse panel of 150 sorghum lines, comprising fertility restorers (R-lines) and maintainers (B-lines). Among these, the markers S5_1608322, S7_1799521, S8_62486248, and S9_4248350 demonstrated the highest accuracy in differentiating R-lines from B-lines (Fig. 7, Supplementary Figure S1). All remaining KASP markers showed limited or inconsistent discrimination between restorer and maintainer lines (Supplementary Table S7).

Discussion

Although fertility restoration is influenced by environmental factors, evaluation across two contrasting seasons within the same year provided sufficient environmental variation for initial genetic dissection. Multi-year validation will be required for confirming environmental stability of identified loci. High heritability coupled with high genetic advance was recorded for most fertility restoration traits during the rainy and post-rainy seasons, respectively. This combination indicates that the observed variation in these traits is largely governed by additive genetic effects, suggesting that effective improvement through selection is feasible. The

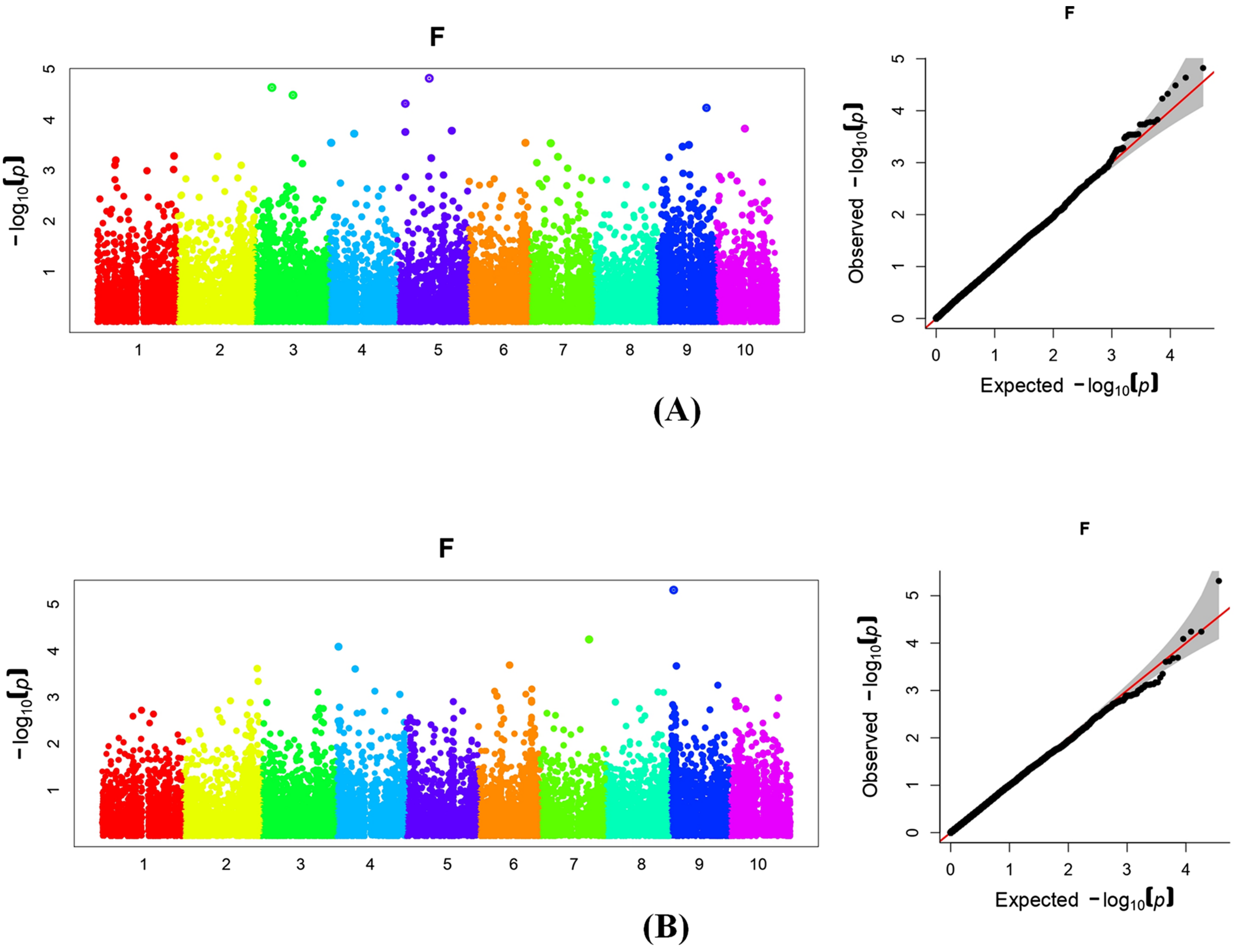

Fig. 4 Manhattan and QQ plots of GWAS for the fertility percent in sorghum RIL population for R2016 (**A**) PR2016 (**B**)

Table 4 SNPs and candidate genes associated with seed setting percent in F_2 population

SN	SNP	Chr.	Position (bp)	Allele	*P*. value	Marker effect	Gene	Annotation
1	S2_34869785	2	34,869,785	A/G	4.25E-06	0.14	--	Intergenic
2	S2_34869800	2	34,869,800	C/T	4.25E-06	0.14	--	Intergenic
3	S2_34869811	2	34,869,811	T/C	4.25E-06	0.14	--	Intergenic
4	S5_62015336	5	62,015,336	G/C	4.58E-06	-0.12	--	Intergenic
5	S3_43103180	3	43,103,180	G/A	2.45E-05	-0.21	--	Intergenic
6	S7_42064650	7	42,064,650	C/A	5.85E-05	0.21	--	Intergenic
7	S6_21453960	6	21,453,960	C/A	6.37E-05	0.13	--	Intergenic
8	S8_653850	8	653,850	T/C	7.30E-05	-0.16	Sobic.008G008400.1	Leucine-rich repeat (LRR)
9	S10_24735649	10	24,735,649	C/T	7.43E-05	0.18	--	Intergenic
10	S10_24735666	10	24,735,666	T/C	7.43E-05	0.18	--	Intergenic
11	S8_912985	8	912,985	G/A	7.49E-05	-0.23	Sobic.008G011000.1	RED-like protein
12	S8_912987	8	912,987	G/A	7.49E-05	-0.23	Sobic.008G011000.1	RED-like protein
13	S8_913023	8	913,023	T/C	7.49E-05	-0.23	Sobic.008G011000.1	RED-like protein
14	S3_10606777	3	10,606,777	C/T	8.58E-05	0.22	--	Intergenic

Table 5 SNPs and candidate genes associated with seed setting percent in BC$_1$F$_1$ population

SN	SNP	Chr.	Position (bp)	Allele	*P* value	Marker effect	Gene	Annotation
1	S1_49298939	1	49,298,939	G/A	8.27E-06	5.74	--	Intergenic
2	S2_43824988	2	43,824,988	T/C	1.94E-05	-7.25	--	Intergenic
3	S2_43824990	2	43,824,990	T/A	1.94E-05	-7.25	--	Intergenic
4	S2_43824996	2	43,824,996	T/C	1.94E-05	-7.25	--	Intergenic
5	S2_43824991	2	43,824,991	A/G	1.94E-05	-7.25	--	Intergenic
6	S2_43824992	2	43,824,992	A/T	1.94E-05	-7.25	--	Intergenic
7	S2_43825028	2	43,825,028	C/T	1.94E-05	-7.25	--	Intergenic
8	S5_47872291	5	47,872,291	G/T	2.12E-05	6.60	--	Intergenic
9	S5_16426809	5	16,426,809	A/G	2.15E-05	-7.45	Sobic.005G099800.1	gibberellin receptor GID1L2
10	S5_16426820	5	16,426,820	C/G	2.15E-05	-7.45	Sobic.005G099800.1	gibberellin receptor GID1L2
11	S5_16426865	5	16,426,865	C/G	2.15E-05	-7.45	--	Intergenic
12	S3_30310892	3	30,310,892	T/C	3.74E-05	-5.02	--	Intergenic
13	S3_30310883	3	30,310,883	G/T	3.74E-05	-5.02	Sobic.003G170200.1	CLP protease family protein
14	S8_62486248	8	62,486,248	A/C	9.92E-05	4.54	Sobic.008G190700.1	PPR repeat family protein

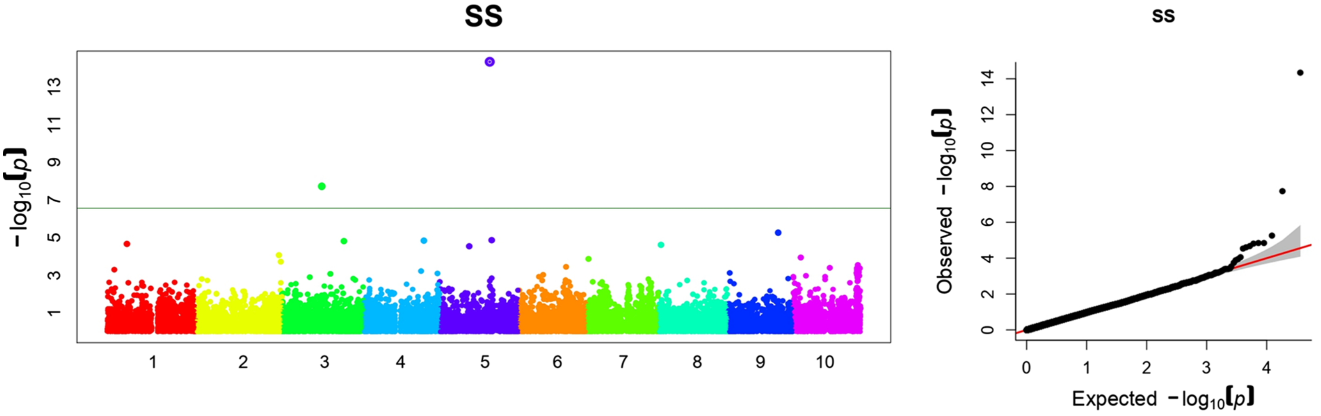

Fig. 5 Manhattan (**A**) and QQ (**B**) plots of GWAS for the seed setting percent in sorghum F$_2$ population

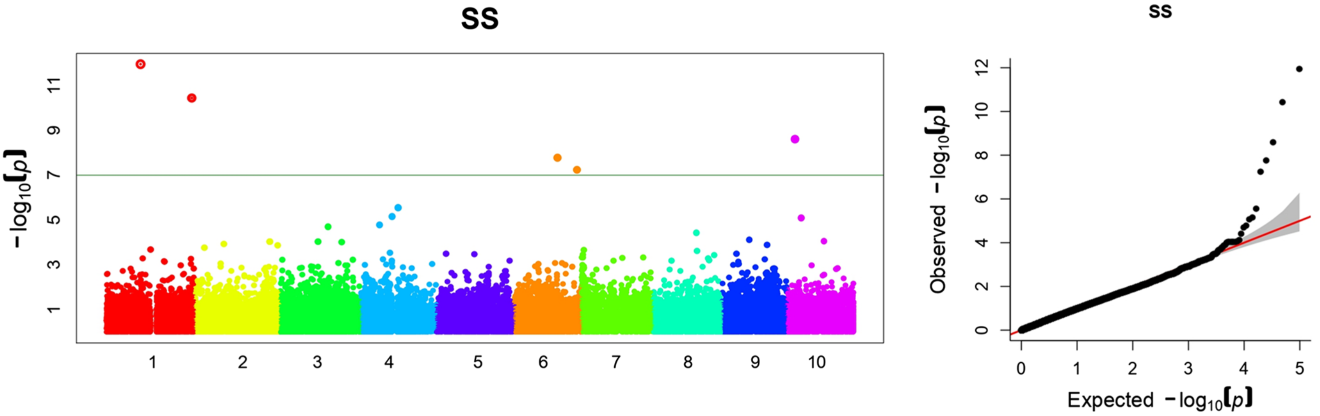

Fig. 6 Manhattan and QQ plots of GWAS for the seed setting percent in sorghum BC$_1$F$_1$ population

phenotypic coefficient of variation (PCV) was consistently higher than the genotypic coefficient of variation (GCV) for all traits studied, reflecting the influence of environmental factors on trait expression.

A relatively low magnitude of both genotypic and phenotypic coefficients of variation was observed for seed set percentage and fertility percentage during the rainy and post-rainy seasons, respectively, indicating that although

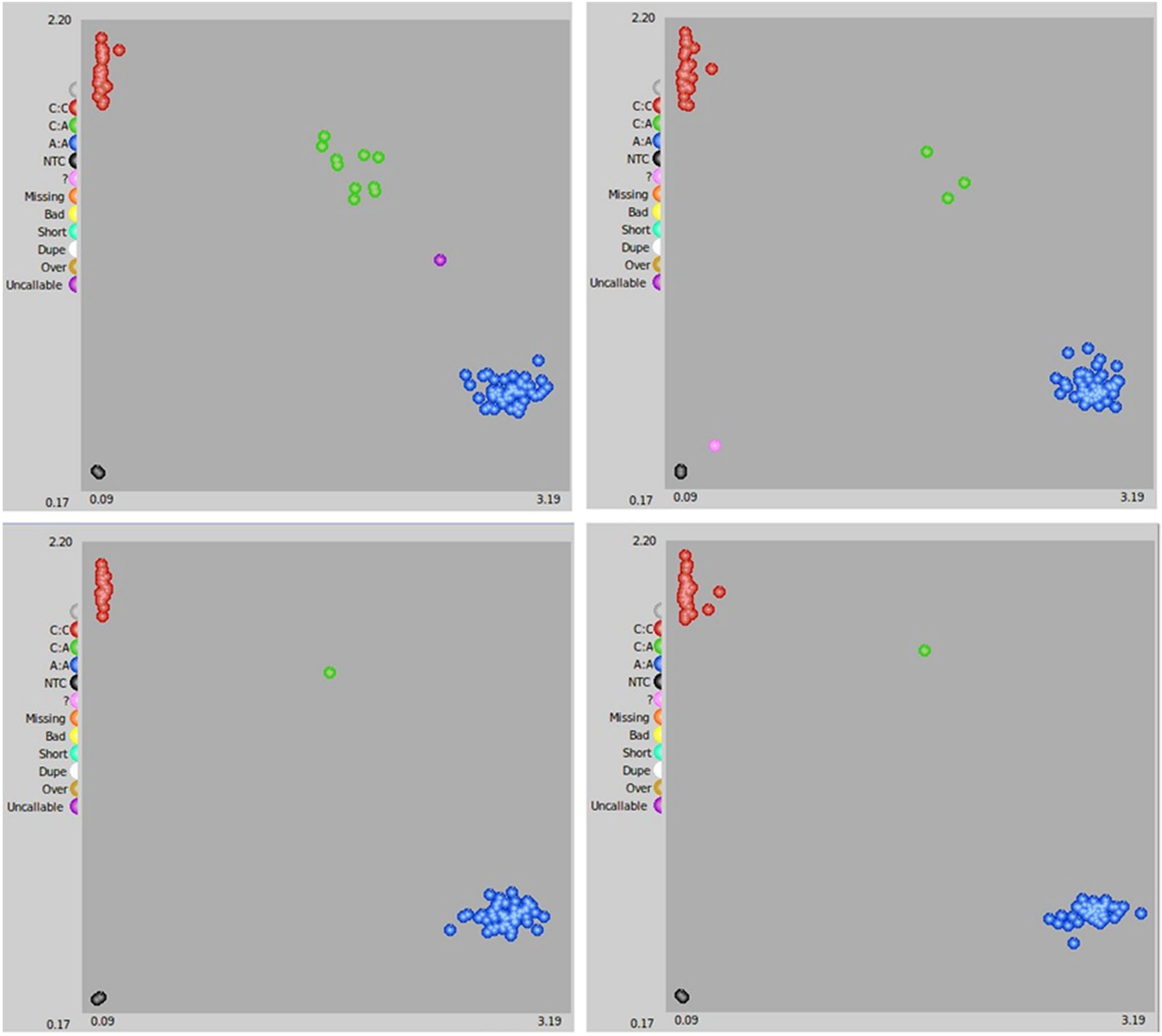

Fig. 7 KASP validation results of marker S7_1799521 for the fertility restoration traits. The scatter plots with axes x and y represents allele discrimination of genotypes. The red and blue dot represents the homozygous alleles and the green dot represents heterozygous alleles

variability exists, the scope for genetic improvement of these traits may be moderate (Arunkumar 2004).

Genome-wide association analysis conducted on the recombinant inbred line (RIL) population across rainy (R) and post-rainy (PR) seasons identified multiple SNPs significantly associated with fertility restoration traits. For seed set percentage, six significant SNPs were detected during the rainy season, whereas four significant SNPs were identified during the post-rainy season. In the rainy season, one SNP, S9_46616891, was associated with the gene *Sobic.009G119000.1*, annotated as a tetratricopeptide repeat protein belonging to the pentatricopeptide repeat (PPR) family, while the remaining five SNPs were located in intergenic regions. In the post-rainy season, one SNP was intergenic, whereas the remaining three SNPs were located within functional genomic regions but were annotated as proteins of unknown function.

Earlier research indicates that fertility restoration in sorghum is under relatively complex genetic control, involving a combination of major and minor gene effects (Maunder and Pickett 1959; Miller and Pickett 1964). The continuum of variation in seed set observed in the present study is consistent with previous reports (Jordan et al. 2011). It is noteworthy that the range and distribution of fertility restoration phenotypes can vary substantially depending on the parental lines involved, with some populations exhibiting a distinct bimodal distribution, while others display a broad and continuous spectrum of restoration phenotypes.

Such variation in fertility restoration is likely attributable to the segregation of multiple partial restorer genes, with the number and effect of minor loci being strongly influenced by the underlying genetic background (Jordan et al. 2011). The presence of effective restorer genes that enable the production of fully fertile F_1 hybrids through cytoplasmic genetic male sterility systems is therefore critical for the cost-effective production of hybrid sorghum seed. At present, commercial hybrid sorghum seed production relies almost exclusively on the A1 cytoplasm, underscoring the importance of understanding and efficiently manipulating fertility restoration mechanisms within this system (Jordan et al. 2011).

In the present study, genome-wide association analyses conducted across the RIL, F_2, and BC_1F_1 populations identified seven SNPs significantly associated with seed set percentage. Among these, five SNPs were linked to the major fertility restoration locus *Rf1*, while one SNP was located on chromosome SBI-09 and another on chromosome SBI-03. In addition, four SNPs were found to be significantly associated with pollen fertility score, of which one SNP was linked to *Rf1*, and the remaining three SNPs were mapped to chromosomes SBI-01, SBI-07, and SBI-10, respectively. For fertility percentage, a single significant SNP was identified on chromosome SBI-09.

Notably, all SNPs identified in association with fertility restoration traits were annotated to genes encoding proteins belonging to the pentatricopeptide repeat (PPR) gene family, which are widely recognized for their central role in cytoplasmic–nuclear interactions and fertility restoration mechanisms in sorghum and other crop species (Jordan et al. 2010, 2011; Praveen et al. 2018; Kante et al. 2018).

The capacity of fertility restorer (*Rf*) genes to encode mitochondria-targeted pentatricopeptide repeat (PPR) proteins was first demonstrated through the cloning of the *Rf-PPR592* gene in petunia (Bentolila et al. 2002). Subsequent studies have established that, with the exception of *Rf2* in maize—which belongs to the aldehyde dehydrogenase gene family (Cui et al. 1996; Liu et al. 2001)—nearly all *Rf* genes identified to date across crop species encode PPR proteins. This includes *Rf* genes characterized in rice (Kazama and Toriyama 2003; Komori et al. 2004; Inagaki et al. 2004) and sorghum (Klein et al. 2001, 2005; Kante et al. 2018).

Rf-like PPR genes are known to evolve rapidly through a "birth-and-death" evolutionary process, characterized by frequent gene duplication and divergence, which enables rapid functional diversification and compensatory correction of mitochondrial gene expression (Geddy et al. 2007; Dahan and Mireau 2013). Such evolutionary dynamics likely underlie the diversity and complexity of fertility restoration mechanisms observed in CMS-based hybrid systems.

Maintaining complete male sterility in A-lines while multiplying their seed is essential for efficient commercial hybrid seed production. Maintainer (B) lines must be devoid of restorer alleles to prevent unintended fertility restoration during A/B line maintenance. A-lines must therefore exhibit near-complete male sterility to ensure high genetic purity and uniformity of hybrid seed production. Partial fertility restoration in A-lines can compromise hybrid seed quality and reduce yield potential in field crops (Kante et al. 2018).

Very recently, the 236 kb *Rf2* locus in sorghum was fine-mapped to a 10.32 kb interval, within which *Sobic.002G057050* was identified as the sole PPR gene (Praveen et al. 2018). This gene was shown to be preferentially expressed in sorghum inflorescences compared to vegetative tissues and exhibited approximately twelvefold higher expression in male-fertile (restorer) lines than in male-sterile lines (Kante et al. 2018). Furthermore, (Praveen et al. 2018) identified 44 and 43 SNPs between the CMS line 296 A and two restorer lines (RS29 and DSV1, respectively), along with a three-base-pair deletion in the coding region of 296 A. Collectively, studies by (Jordan et al. 2010), (Praveen et al. 2018), and (Kante et al. 2018) concluded that *Sobic.002G057050* represents the underlying gene at the *Rf2* locus, with a SNP at position 1090 bp proposed as the causative mutation. The availability of reliable molecular markers for fertility restoration and sterility maintenance would enable marker-assisted assessment of restorer and maintainer status, offering a faster, more cost-effective alternative to conventional field-based evaluation. Such markers would be particularly valuable for eliminating partial restorer alleles from A- and B-lines, as these alleles can remain cryptic over multiple generations and become phenotypically expressed only under specific environmental conditions (Jordan et al. 2010, 2011). Moreover, combinations of major *Rf* genes with partial restorer loci can result in complete fertility restoration in F_1 hybrids.

KASP-based marker validation for the *Rf2* locus in a West African sorghum mapping population was previously reported by Kante et al. (2018), and similar SNP-based approaches have been successfully applied in other crops, including maize (Su et al. 2017). In the present study, a total of 31 SNPs—comprising 13 SNPs identified through GWAS and 18 SNPs derived from earlier reports of fertility restoration loci in sorghum (Jordan et al. 2010, 2011; Praveen et al. 2015, 2018; Kante et al. 2018)—were selected for Kompetitive Allele-Specific PCR (KASP) assay development.

Of the 31 KASP assays developed, 12 assays failed to amplify consistently, resulting in an assay success rate of 61.3%. The remaining 19 assays were evaluated further, of which 13 markers were polymorphic and six were monomorphic across a panel of 94 samples. These monomorphic markers were retained for exploratory evaluation in broader germplasm sets. The set of 19 KASP markers was subsequently validated in 213 individuals of the RIL population derived from the cross 296B × IS18551, where 11 markers effectively differentiated parental lines and reliably discriminated among RILs.

Furthermore, screening of a diverse panel of 150 restorer and maintainer lines demonstrated that the markers S5_1608322, S7_1799521, S8_62486248, and S9_4248350 most accurately differentiated R-lines from B-lines, while the remaining markers showed limited discriminatory ability. These findings highlight the practical value of selected KASP markers for routine application in sorghum hybrid breeding programs.

The primary requirement of any breeding program is the assurance of accurate crosses (Glaszmann et al. 2010). In sorghum, the small flower size and large number of florets per panicle make physical emasculation difficult and inefficient. The present study demonstrates that sorghum hybrids can be efficiently developed and verified using molecular markers, reducing reliance on labor-intensive phenotypic methods. Compared with simple sequence repeat (SSR) markers, which are limited in number and relatively costly, SNP-based KASP markers provide a high-throughput, cost-effective, and scalable alternative for fertility restoration diagnostics (Graves et al. 2016).

Conclusion

This study demonstrated substantial variability for fertility restoration traits in the sorghum recombinant inbred line (RIL) population, accompanied by high heritability, indicating strong genetic control of these traits. The set of SNP markers validated in the present investigation was shown to be directly associated with fertility restoration of the A1 cytoplasm in sorghum, confirming their biological relevance and practical utility.

The findings of this research significantly advance the understanding of the genetic basis underlying fertility restoration in sorghum and expand the molecular toolkit available for precise manipulation of this commercially important trait. Knowledge of the genomic location of restorer-of-fertility (*Rf*) genes will enable breeders to design more effective breeding strategies by identifying superior alleles and ensuring their presence or absence in elite parental lines through marker-assisted selection.

Overall, the validated molecular markers generated in this study provide breeders with a robust and efficient set of tools to accurately manipulate fertility restoration, thereby enhancing the efficiency, reliability, and genetic purity of sorghum hybrid breeding programs.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13353-026-01051-0.

Authors' contribution GN: Executed the experiment, carried out statistical analysis and prepared manuscript, LR, UK: contributed in statistical analysis, SM, SD: Planned, supervised the research, Contributed in preparing manuscript and review, SHW: Contributed in preparing manuscript and review.

Declarations

Conflict of interest The 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.

References

Arunkumar B, Biradar B, Salimath P (2004) Inheritance of fertility restoration on *milo* and *maldandi* sources of male sterility in *rabi* sorghum [*Sorghum bicolor* (L.) Moench]. Indian J Genet Plant Breed 64:325–326

Bentolila S, Alfonso AA, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic malesterile plants. Proc Natl Acad Sci USA 99:10887–10892. https://doi.org/10.1073/pnas.102301599

Brooking IR (1976) Male sterility in Sorghum bicolor (L.) Moench induced by low night temperature. I. Timing of the stage of sensitivity. Aust J Plant Physiol 3:589–596

Brooking IR (1979) Male sterility in Sorghum bicolor (L.) Moench induced by low night temperature. II. Genotypic differences in sensitivity. Aust J Plant Physiol 6:143–147

Cingolani P, Platts A, Wang l, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6(2):80–92. https://doi.org/10.4161/fly.19695

Cui X, Wise RP, Schnable PS (1996) The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. Science 272:1334–1336. https://doi.org/10.1126/SCIENCE.272.5266.1334

Dahan J, Mireau H (2013) The *Rf* and *Rf*-like PPR in higher plants, a fast-evolving subclass of PPR genes. RNA Biol 10:1469–1476. https://doi.org/10.4161/rna.25568

Downs RW, Marshall DR (1971) Low temperature induced male sterility in grain sorghum. Aust J Agric Res Anim Husb 11:352–356

Geddy R, Brown GG, Small I et al (2007) Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. BMC Genom 8:130. https://doi.org/10.1186/1471-2164-8-130

Glaszmann JC, Kilian B, Upadhyaya HD, Varshney RK (2010) Accessing genetic diversity for crop improvement. Curr Opinion Plant Biol 13 (2010): 167–173. https://doi.org/10.1016/j.pbi.2010.01.004

Glaubitz JC, Casstevens TM, Lu F et al (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9:e90346. https://doi.org/10.1371/journal.pone.0090346

Goodstein DM, Shu S, Howson R et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–D1186. https://doi.org/10.1093/nar/gkr944

Graves H, Rayburn AL, Gonzalez-Hernandez JL, Nah G, Kim D-S, Lee DK (2016) Validating DNA Polymorphisms Using KASP Assay in Prairie Cordgrass (*Spartina pectinata* Link) Populations in the U.S. Front Plant Sci 6:1271. https://doi.org/10.3389/fpls.2015.01271

Inagaki A, Akagi H, Nakamura A et al (2004) Positional cloning of the rice Rf-1 gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein. TAG Theor Appl Genet 108:1449–1457. https://doi.org/10.1007/s00122-004-1591-2

Jordan DR, Klein RR, Sakrewski KG et al (2011) Mapping and characterization of Rf_5: a new gene conditioning pollen fertility restoration in A1 and A2 cytoplasm in sorghum (*Sorghum bicolor* (L.) Moench). Theor Appl Genet 123:383–396. https://doi.org/10.1007/s00122-011-1591-y

Jordan DR, Mace ES, Henzell RG, Klein PE, Klein RR (2010) Molecular mapping and candidate gene identification of the Rf2 gene for pollen fertility restoration in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 120(7):1279-1287. https://doi.org/10.1007/s00122-009-1255-3

Kante M, Rattunde HFW, Nébié B et al (2018) QTL mapping and validation of fertility restoration in West African sorghum A_1 cytoplasm and identification of a potential causative mutation for Rf_2. Theor Appl Genet 131:2397–2412. https://doi.org/10.1007/s00122-018-3161-z

Kazama T, Toriyama K (2003) A pentatricopeptide repeat-containing gene that promotes the processing of aberrant atp6 RNA of cytoplasmic male-sterile rice. FEBS Lett 544:99–102. https://doi.org/10.1016/S0014-5793(03)00480-0

Klein RR, Klein PE, Chhabra AK et al (2001) Molecular mapping of the rf_1 gene for pollen fertility restoration in sorghum (*Sorghum bicolor* L). Theor Appl Genet 102:1206–1212. https://doi.org/10.1007/s001220100575

Klein RR, Klein PE, Mullet JE et al (2005) Fertility restorer locus Rf1 [corrected] of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. Theor Appl Genet 111:994–1012. https://doi.org/10.1007/s00122-005-2011-y

Komori T, Ohta S, Murai N et al (2004) Map-based cloning of a fertility restorer gene, *Rf*-1, in rice (*Oryza sativa* L). Plant J 37:315–325. https://doi.org/10.1046/j.1365-313X.2003.01961.x

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Liu F, Cui X, Horner HT et al (2001) Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. Plant Cell 13:1063–1078

Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet 12(2): e1005767. https://doi.org/10.1371/journal.pgen.1005767 PMID: 26828793

Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L, Small I (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 16:2089–2103

Maunder AB, Pickett RC (1959) The genetic inheritance of cytoplasmic-genetic male sterility in grain sorghum1. Agron J 51:47. https://doi.org/10.2134/agronj1959.00021962005100010016x

McCormick RF, Truong SK, Sreedasyam A et al (2018) The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. Plant J 93:338–354. https://doi.org/10.1111/tpj.13781

Miller D, Pickett R (1964) Inheritance of partial male-fertility in Sorghum vulgare Pers. Crop Sci 4:1–4

Praveen M, Anurag Uttam G, Suneetha N et al (2015) Inheritance and molecular mapping of Rf_6 locus with pollen fertility restoration ability on A1 and A2 cytoplasms in sorghum. Plant Sci 238:73–80. https://doi.org/10.1016/j.plantsci.2015.05.020

Praveen M, Uttam AG, Tonapi VA, Ragimasalawada M (2018) Fine mapping of Rf_2, a major locus controlling pollen fertility restoration in sorghum A 1 cytoplasm, encodes a PPR gene and its validation through expression analysis. Plant Breed 137:148–161. https://doi.org/10.1111/pbr.12569

Saha D, Prasad AM, Srinivasan (2007) Pentatricopeptide repeat proteins and their emerging role in plants. Plant Physiol Biochem 45:521–534

Schmitz-Linneweber C, Small I (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci 13:663–670

Stephens JC, Holland RF (1954) Cytoplasmic male-sterility for hybrid sorghum seed production. Agron J 46:20–23. https://doi.org/10.2134/agronj1954.00021962004600010006x

Su C, Wang W, Gong S, Zuo J, Li S, Xu S (2017) High Density Linkage Map Construction and Mapping of Yield Trait QTLs in Maize (*Zea mays*) Using the Genotyping-by-Sequencing (GBS) Technology. Front Plant Sci 8:706. https://doi.org/10.3389/fpls.2017.00706