INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
20–1524/2021/25–3–659–666
DOI: 10.17957/IJAB/15.1714
http://www.fspublishers.org

Full Length Article

A Semi-Quantitative Bulked Segregant Analysis Preliminarily Localizes a Maize Male-Sterility Gene

Mingbo Wu[1†], Chenchen Ma[1†], Xiaoqing Wang[1], Xiaojing Sun[1], Lanjie Zheng[1], Zhanyong Guo[1], Zhiyuan Fu[1], Weihua Li[1], Jihua Tang[1,2*] and Xiangge Zhang[1*]

[1]*National Key Laboratory of Wheat and Maize Crops Science, College of Agronomy, Henan Agricultural University, Zhengzhou, 450018, China*

[2]*Hubei Collaborative Innovation Center for Grain Industry, Yangtze University, Jingzhou, 433200, China*

[*]For correspondence: tangjihua1@163.com; maizezxg@163.com

[†]Contributed equally to this work and are co-first authors

Received 22 October 2020; Accepted 04 December 2020; Published 25 January 2021

Abstract

Bulked segregant analysis (BSA) assists in map-based cloning of mutant genes. However, a traditional BSA requires many high-density markers for successful linkage analysis which is labor-intensive and time-consuming. In this study, a semi-quantitative DNA analysis program was optimized and combined it with BSA, resulting in a semi-quantitative BSA (sq-BSA). The sq-BSA approach allowed evaluation of the proportions of marker-defined individuals (dominant or recessive marker types) in bulks. The sq-BSA method was used to map a male-sterility (*ms*) gene, *ms2016*, in maize. Forty polymorphic markers were screened from one-third of each chromosome (from the head or tail) for mapping. Among these markers, seven were identified as candidate gene-linked markers, of which four markers (bnlg1046, umc1563, umc1171 and umc1722) were located on chromosome 5. Using group validation, *ms2016* was anchored on chromosome 5 and was most closely linked to bnlg1046. Furthermore, four new InDel markers located near bnlg1046 were screened to map the preliminary location of *ms2016.* The *ms2016* gene was mapped to an 8.7 Mb interval flanked by the InDel polymorphic markers I5-3 (chr5:14588060) and I5-12 (chr5:23308445). Thus, this improved BSA method (sq-BSA) requires only a small number of molecular markers to quickly localize a target gene, representing a high-efficiency tool for mutant gene mapping.

Keywords: Bulked segregant analyses; Maize; Male-sterility; Mapping; Molecular markers

Introduction

Map-based cloning is a valuable forward genetic approach for gene isolation. Based on linkage analyses, the strategy is used to identify genes controlling a particular trait or phenotype from segregating populations by using defined markers (Michelmore *et al.* 1991; Page and Grossniklaus 2002; Rallapalli *et al.* 2019). To avoid laboriously identifying marker types for superabundant individuals, bulked segregant analyses (BSA) are conducted to classify and bulk segregants according to their phenotypes and to compare marker allele frequencies in different bulks (Michelmore *et al.* 1991). In a recessive bulk, these unlinked markers are heterozygous for the two parental types owing to recombination, whereas linked markers are homozygous for the mutant parental type (Klein *et al.* 2018). However, a traditional BSA frequently requires many high-density markers to genotype bulks for successful linkage analysis (almost complete linkage), which is labor-intensive and time-consuming.

After the advent of next-generation deep sequencing, a novel BSA technique based on deep sequencing (BSA-seq) was developed (Jay and Hanlee 2008; Fekih *et al.* 2013; Takagi *et al.* 2013a, b, 2015; Huang *et al.* 2020). Using whole-genome high-throughput resequencing, the proportion of the number of short reads harboring SNPs (defined as SNP-index) is evaluated for each bulked DNA, and then the candidate region is defined based on the regional distribution of the SNP-index (Takagi *et al.* 2013b; Itoh *et al.* 2019). Without marker screening and genotyping, BSA-seq is a quick, straightforward method for gene mapping. This method has been used to identify genes in many different species, such as wheat (*Triticum aestivum* L.) (Trick *et al.* 2012), chickpea (*Cicer arietinum* L.) (Das *et al.* 2015), rice (*Oryza sativa* L.) (Zheng *et al.* 2016) and sunflower (*Helianthus annuus* L.) (Imerovski *et al.* 2019), and is proven to be effective for gene mapping. Nevertheless, there are several potential factors that affect the accuracy and validity, including the genome coverage

To cite this paper: Wu M, C Ma, X Wang, X Sun, L Zheng, Z Guo, Z Fu, W Li, J Tang, X Zhang (2021). A semi-quantitative bulked segregant analysis preliminarily localizes a maize male-sterility gene. *Intl J Agric Biol* 25:659–666

for sequencing (the higher the coverage, the more accurate the results) and the statistical methods (sequence alignment and allele calling algorithms) used to estimate allele frequencies in the bulks (Huang *et al.* 2020). Thus, for larger genomes, the required high genome coverage reduces the feasibility of the method.

Evaluation of the components in the bulk (marker types or SNP index) is central to all BSA methods. Namely, it is a key to evaluate the DNA quantities of each type in the bulked DNA sample. Semi-quantitative reverse-transcription PCR (RT-PCR) provides a rapid, sensitive method for quantifying gene expression (*i.e.*, mRNA quantities) in samples (Maddocks and Jenkins 2017). A similar program might be available to quantify the DNA in samples. In this case, a small number of molecular markers are sufficient to identify a candidate marker partially linked to the target gene in the bulks, which will facilitate rapid gene mapping. In the present study, a semi-quantitative DNA analysis program was established for use in combination with BSA for rapid localization of mutant genes. To test the effectiveness of semi-quantitative BSA (sq-BSA), we applied the method to the mapping of a male-sterile (*ms*) gene (*ms2016*) in maize (*Zea mays* L.).

Materials and Methods

Plant materials

The maize male-sterile mutant *ms2016* is a naturally occurring recessive mutant derived from breeding inbred line T9047, which is preserved in our laboratory (Henan Agricultural University). On account of the male-sterility trait, *ms2016* was preserved in the F_2 individuals [(*ms2016* × T9047)^2]. Chang 7-2 (C7-2) is an excellent inbred line with sufficient fertility and differs in genetic background from that of T9047 line. To map the male-sterility gene *ms2016*, we used C7-2 and *ms2016* to construct an isolated F_2 population [(*ms2016* × C7-2)^2]. Then, F_2 populations raised from four ears, each comprising ~400 individuals, were cultivated in the standard experimental field of Henan Agricultural University, Yuanyang, China in 2019. In addition, the parental lines C7-2 and T9047 (representative of *ms2016*) were also planted.

Identification of male sterile individuals and DNA extraction

Under environmental stress, the maize tassel readily shows delayed anther dehiscence but retains pollen fertility (Tazib *et al*. 2014; Matsui and Hasegawa 2019). However, *ms2016* is entirely sterile owing to anther indehiscence and absence of pollen grains in the anther. The isolated F_2 population may contain some fertile plants that show delayed anther dehiscence, which might be mistaken as male-sterile individuals. Thus, during the pollen shedding period, we rubbed open anthers twice at a 5-day interval to judge the pollen fertility. The individuals with no pollen grains were identified as sterile (recessive individuals).

Leaves of the recessive individuals, as well as the parental lines C7-2 and T9047, were excised with scissors and preserved at −70°C. DNA was rapidly extracted from 50 mg of each leaf sample using the sodium lauryl sulfate method. A NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) was used to determined DNA quantity and purity. All purified DNA samples were stored at −20°C.

Screening of SSR polymorphic molecular markers

A set of simple sequence repeat (SSR) molecular markers covering the complete maize genome was previously synthesized in our laboratory and was used to perform the primary mapping. In this study, we deliberately selected molecular markers located on one-third of each chromosome (from the head or tail), such as chromosome 1 (chr. 1) bin1.03-1.04 and bin1.06-1.07, respectively. Based on information for maize SSR markers (https://www.maizegdb.org/data_center/ssr), 120 SSR markers (12 for each chromosome, Table 1) were selected to screen for polymorphisms between the parents (C7-2 and T9047). Polymorphism screening was performed by polyacrylamide gel electrophoresis (PAGE). At least two polymorphic markers were required for each chromosomal region (upper or lower), ensuring that at least one marker was close to the target gene.

Optimization of a semi-quantitative DNA analysis program

A suitable DNA template concentration and PCR cycle number are crucial factors that influence semi-quantitative DNA analysis programs. Generally, 20–200 ng/μL DNA template and 35 PCR cycles are optimal for routine PCR. Here, a two-factor experiment involving DNA template concentration (20, 30 and 40 ng/μL) and PCR cycle number (25, 30, and 35) was conducted to develop a semi-quantitative DNA analysis program. Two polymorphic SSR markers were used to demonstrate the experimental process using C7-2 and T9047 DNA templates. The specific PCR system and procedure are presented in Table 2. Finally, the optimal criteria for the semi-quantitative DNA analysis program were determined for the subsequent experiments.

Semi-quantitative BSA

To improve reliability, two recessive bulks were prepared for the BSA. Thirty recessive DNA samples were randomly selected from a single F_2 ear-derived population. Of the samples, 15 and 30 DNA samples were equivalently mixed to form the bulks B1 and B2, respectively. Using the criteria for the semi-quantitative DNA analysis program, the total DNA quantity in each bulk and the quantity of each DNA sample in each bulk were fixed. The volume of each bulk

Table 1: Information of the 120 SSR molecular markers

Chromosome	SSR	bin	Chromosome	SSR	bin
Chr.1-1	bnlg182	1.03	Chr.6-1	phi389203	6.04
	bnlg2097	1.03		umc1014	6.04
	umc1397	1.03		umc2317	6.04
	umc2217	1.03		umc2006	6.04
	phi001	1.03		nc009	6.04
	umc2145	1.03		umc2318	6.04
Chr.1-2	bnlg2057	1.06	Chr.6-2	umc1520	6.06
	umc2151	1.06		umc2170	6.06
	umc1812	1.06		umc1912	6.06
	umc1122	1.06		umc2322	6.06
	mmc0011	1.06		umc1424	6.06
	umc2236	1.06		umc2389	6.06
Chr.2-1	umc2247	2.03	Chr.7-1	bnlg2233	7.02
	bnlg1064	2.03		umc2142	7.02
	mmc0231	2.03		bnlg434	7.03
	umc2193	2.03		umc1987	7.03
	umc2032	2.04		umc1877	7.03
	bnlg1175	2.04		bnlg1305	7.03
Chr.2-2	umc1080	2.06	Chr.7-2	umc2333	7.05
	umc1658	2.06		dupssr13	7.05
	umc2186	2.06		umc1407	7.05
	mmc0143	2.07		umc2334	7.05
	umc2178	2.07		umc1760	7.05
	bnlg1413	2.07		phi116	7.06
Chr.3-1	bnlg1447	3.03	Chr.8-1	phi060	8.03
	unc2259	3.03		umc1913	8.03
	umc1025	3.04		umc2154	8.03
	umc1717	3.04		umc1460	8.04
	bnlg1904	3.04		umc1202	8.04
	umc2033	3.04		bnlg2046	8.04
Chr.3-2	umc1400	3.06	Chr.8-2	umc2356	8.06
	umc1796	3.06		umc1997	8.06
	umc1674	3.06		umc1149	8.06
	umc2169	3.06		bnlg1152	8.06
	umc1985	3.06		umc2395	8.06
	bnlg197	3.06		umc1905	8.06
Chr.4-1	umc1550	4.03	Chr.9-1	umc1586	9.03
	umc2281	4.03		dupssr6	9.03
	umc2211	4.03		umc2370	9.03
	umc1963	4.04		umc1492	9.04
	phi026	4.04		bnlg1209	9.04
	umc1652	4.04		umc1519	9.04
Chr.4-2	bnlg2291	4.06	Chr.9-2	umc2095	9.05
	bnlg1023	4.06		umc1771	9.05
	bnlg252	4.06		phi448880	9.06
	umc1299	4.06		umc2207	9.06
	umc1194	4.07		umc1733	9.06
	umc1994	4.07		umc2346	9.06
Chr.5-1	umc1373	5.03	Chr.10-1	umc2017	10.03
	bnlg1046	5.03		umc1432	10.03
	umc1226	5.03		bnlg1712	10.03
	umc1850	5.03		bnlg1716	10.03
	umc1966	5.04		umc1179	10.03
	umc1563	5.04		umc1863	10.03
Chr.5-2	umc1171	5.05	Chr.10-2	phi057	10.06
	umc1524	5.06		bnlg1250	10.06
	bnlg1847	5.06		bnlg1028	10.06
	umc1722	5.06		umc2122	10.06
	phi085	5.06		umc1993	10.06
	umc1680	5.06		umc2212	10.06

was fixed to 400 μL, with the components of each bulk (Table 3) as follows:

For B1, $V_i = (C_n \times 400)/15/C_i$, supplemented with ddH_2O to 400 μL

For B2, $V_i = (C_n \times 400)/30/C_i$, supplemented with ddH_2O to 400 μL

Table 2: Semi-quantitative PCR-based DNA analysis system and procedure

PCR System	Volume	Concentration
2 × taq Master Mix for PAGE	5 μL	
Pi-L (10 μM)	0.5 μL	
Pi-R (10 μM)	0.5 μL	
ddH_2O	3 μL	
DNA	1 μL	20/30/40 ng/μL
Total	10 μL	
PCR Procedure	Cycle Number	Temperature conditions
	1	95°C 90 s
	25/30/35	95°C 20 s; TM°C 30 s; 72°C 40 s;
	1	72°C 5 min
	1	12°C Preservation

Note: Pi represents a polymorphic marker; TM represents annealing temperature of primer

Table 3: Standards of the mixed bulks as assessed by sq-BSA

Individual	Concentration (ng/μL)	Bulk1(μL)	Bulk2 (μL)
1	126	6.3	3.2
2	216	3.7	1.9
3	227	3.5	1.8
4	214	3.7	1.9
5	379	2.1	1.1
6	508	1.6	0.8
7	375	2.1	1.1
8	440	1.8	0.9
9	440	1.8	0.9
10	370	2.2	1.1
11	414	1.9	1.0
12	417	1.9	1.0
13	632	1.3	0.7
14	725	1.1	0.6
15	350	2.3	1.2
16	365		1.1
17	174		2.3
18	299		1.3
19	364		1.1
20	183		2.2
21	224		1.8
22	666		0.6
23	155		2.6
24	256		1.6
25	281		1.4
26	716		0.6
27	297		1.3
28	464		0.9
29	229		1.7
30	270		1.5

Where V_i represents the required volume of each DNA sample in each bulk, C_n represents the normal DNA template concentration of sq-BSA, and C_i represents the measured concentration of each DNA sample.

All SSR markers polymorphic between C7-2 and T9047 were used to detect the composition of the two bulks with PAGE. Generally, owing to chromosomal recombination events and exchanges in the isolated F_2 population, two marker types representing C7-2 and T9047 were detected in the bulks, and thus two bands were visualized in the electrophoretogram for each marker. The sq-BSA effectively evaluated the proportion of each type in each bulk by comparing the intensity (brightness) of their

corresponding bands. In the recessive bulks (male sterile), if the T9047 marker band was much brighter than the C7-2 marker band, individuals carrying the T9047 marker accounted for the majority, and thus this SSR marker was possibly linked to the target gene (*ms2016*).

Validation of candidate linked markers

Candidate markers were used to detect the marker type of every recessive individual in the isolated F_2 population (derived from one F_2 ear). Individuals with heterozygous (C7-2 and T9047) or homozygous C7-2 marker types were considered to be recombinants, whereas those with homozygous T9047 marker types represented non-recombinant individuals. The heterozygous types were marked with a black dot and number 3 on the electrophoretogram for each candidate marker, whereas the homozygous C7-2 marker types were marked with a black dot and number 1. The total number of black dots was scored as the number of recombinants and the recombination rate (r) of the isolated F_2 population was calculated using the formula: $(1 - r)^2$ = (total individuals − recombinant individuals)/total individuals. Markers with low recombination rates (r much less than 50%) were considered to be potentially linked to the *ms2016* gene. A low recombination rate indicates a short distance from the target gene on either side (Fig. 1). Once After approximate localization of *ms2016*, 20 insertion/deletion (Indel) markers flanking the candidate marker (those most closely linked to the target gene) were designed with 0–20 Mb intervals for verification and further clarification of the gene interval. The InDel markers were derived from a sequence analysis of natural populations (Liu *et al*. 2015). Finally, the primary location of the *ms2016* gene was determined by the flanking markers.

Results

Acquisition of recessive male-sterile individuals from an isolated F_2 population

The maize male-sterile mutant *ms2016* was isolated from the T9047 (wild type) inbred line in 2016. The *ms2016* × T9047 F_1 population showed restored fertility. The F_2 individuals, (*ms2016* × T9047)^2, showed a 3:1 segregation ratio of fertile to male-sterile plants, which indicated that the male-sterility phenotype of *ms2016* was genetically controlled by a single recessive gene. No difference in tassel morphology was observed between *ms2016* and T9047, but the anthers of *ms2016* failed to produce pollen (Fig. 2), resulting in the male-sterility phenotype.

Using the fertile C7-2 inbred line as a parent (Fig. 2), we constructed an isolated F_2 population, (*ms2016* × C7-2)^2. The tassel of (*ms2016* × C7-2) F_1 individuals produced fertile pollen, whereas the isolated F_2 population produced fertile and sterile tassels. Successful mapping of the *ms2016* gene required accurate identification of the male-sterile individuals. By rubbing open the anthers, plants that produced no pollen were identified as male-sterile individuals. Ultimately, 81, 100, 90, and 110 recessive male-sterile individuals were identified from the isolated F_2 populations raised from four ears, respectively. A χ^2 test determined that there was a 3:1 segregation ratio of fertile to male-sterile phenotypes in all four F_2 populations (Table 4). These data corroborated that the *ms2016* male-sterile phenotype is controlled by a single recessive gene.

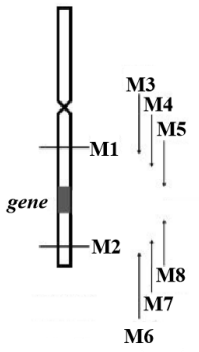

Fig. 1: Diagram of the locations of linked markers and the target gene. M denotes a molecular marker. The arrows' directions indicate closeness to the target gene, as determined by lower recombination rates

Fig. 2: Morphological comparison of maize tassels among *ms2016*, T9047 and C7-2. Bar = 5 cm

Table 4: Tassel sterility segregation ratio in the maize (*ms2016* × T9047)^2 F_2 population

	Fertile individuals	Sterile individuals	Ratio	χ^2	$\chi^2_{0.05}$
Ear-F_2-1	256	81	3.16	0.05	3.84
Ear-F_2-2	298	100	2.98	0.01	
Ear-F_2-3	268	90	2.98	0.01	
Ear-F_2-4	336	110	3.05	0.03	

Note: The χ^2 value less than $\chi^2_{0.05}$ (3.84) means that the segregation ratio (Fertile/Sterile) in F_2 population conforms to theoretical value (3: 1)

Establishment of semi-quantitative DNA analysis program

Molecular markers polymorphic between the parents are crucial for map-based cloning (Peters *et al*. 2003). Using PAGE, 40 SSR markers polymorphic between C7-2 and T9047 were selected (Fig. 3a and Table 5). In addition, two polymorphic markers were identified on one-third of each chromosome (from the head or tail) (Table 5), which enabled division of each chromosome into three equal portions.

Two polymorphic SSR markers (P33 and P28, random sampling) and the parental DNA templates (C7-2 and T9047) were used to determine the optimal criteria for semi-quantitative DNA analysis. To address the two key factors (DNA template concentration and PCR cycle number) influencing PCR amplification, we conducted a two-factor experiment involving DNA template concentration (20, 30, and 40 ng/μL) and PCR cycle number (25, 30, and 35). For marker P33, no electrophoresed bands were visible after 25 PCR cycles at any DNA template concentration, whereas after 30 and 35 PCR cycles distinct bands were observed (Fig. 3b). Notably, with 30 PCR cycles, the intensity of each band gradually increased together with the DNA template concentration, whereas with 35 cycles all bands displayed similar brightness (Fig. 3b). These data showed that the band intensity reflected the DNA template concentration after 30 PCR cycles. Indeed, after 30 PCR cycles, the band brightness with 30 ng/μL DNA template was significantly stronger than that with 20 ng/μL DNA template, but slightly weaker than that with 40 ng/μL DNA template. A similar scenario was observed for marker P28 (Fig. 3b). Taken together, the 30 ng/μL DNA template concentration and 30 PCR cycles were optimal for the semi-quantitative DNA analysis program.

Identification of the candidate markers linked to the *ms2016* gene using sq-BSA

Using the optimized semi-quantitative DNA analysis program, we randomly selected 30 recessive DNA samples to mix as bulks. Half (15) of these DNA samples were randomly selected and equivalently mixed as bulk B1, and all (30) of the samples were equivalently mixed as bulk B2. The identified polymorphic SSR markers were used to analyze the two bulks by sq-BSA. For most markers, such as P6, two bands of equal brightness corresponding to C7-2 and T9047 were amplified in B1 and B2 (Fig. 4). These data revealed that the number of markers for each parent (C7-2 and T9047) was similar in each bulk; therefore, these markers were probably not linked to the *ms2016* gene. However, seven markers (P17, P18, P19, P20, P22, P29, and P33), amplified only one bright band corresponding to T9047 in the two bulks (Fig. 4), which indicated that a higher proportion of individuals had the T9047 marker in each bulk. Therefore, these seven SSR markers were selected as candidate markers linked to the *ms2016* gene.

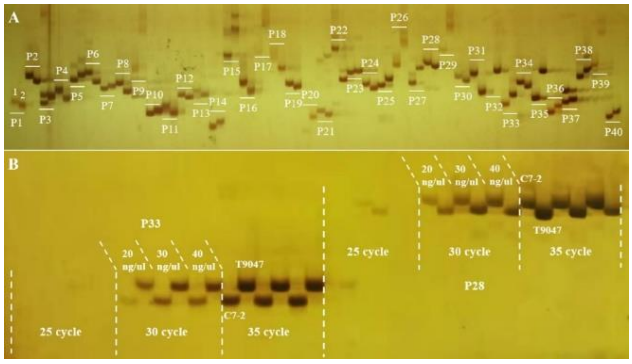

Fig. 3: Establishment of the semi-quantitative DNA analysis method. **a**) Screening for SSR polymorphic molecular markers. P1–40 represents the SSR markers, which are detailed in Table 5. The white lines over lanes indicate pairs of bands for each marker from the two parental lines. The band of parent C7-2 appears in the first lane (1) and that of parent T9047 appears in the second lane (2) for each marker. **b**) Optimization of the semi-quantitative DNA program using two SSR markers (P33 and P28)

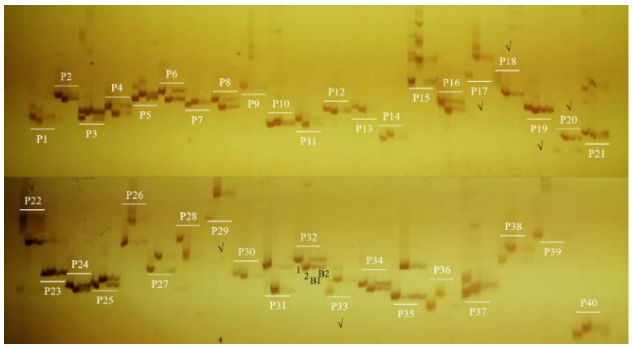

Fig. 4: Linkage analysis of SSR molecular markers based on the sq-BSA. P1 - P40 represent the SSR markers, which are detailed in Table 5. The white lines over lanes indicate bands for each marker from the two parental lines and the two bulks, B1 and B2. The band of parent C7-2 appears in the first lane (1) and that of parent T9047 appears in the second lane (2) for each marker. The bands for the two bulks, B1 and B2, appear in the third and fourth lanes, respectively, for each marker. √ represents candidate linked markers

Primary mapping of the *ms2016* gene

To validate the candidate markers, 81 recessive individuals of the F_2 population raised from one ear were classified using these markers. The candidate markers P17, P18, P19, and P20, identified as bnlg1046 (bin 5.03), umc1563 (bin 5.04), umc1171 (bin 5.06), and umc1722 (bin 5.06), respectively, were localized on chromosome 5. Thus, we focused on these four candidate markers for our analysis and validation. The number of black spots for P17 (bnlg1046) was much less than that for the other markers (Fig. 5), thus fewer recombinants were identified for the bnlg1046 marker in the F_2 population. Through counting the number of black spots on the electrophoretogram, we calculated the recombination

Table 5: Information of 40 polymorphic SSR molecular markers

	SSR	Bin	Position
P1	bnlg182	1.03	51850098
P2	umc2217	1.03	61783198
P3	umc1812	1.06	182705124
P4	umc2151	1.06	197052463
P5	umc2247	2.03	30966754
P6	umc2032	2.04	42496682
P7	umc1658	2.06	168257256
P8	umc2178	2.07	174320099
P9	umc1025	3.04	18218616
P10	umc1717	3.04	27971021
P11	umc1674	3.06	182057738
P12	umc1985	3.06	189146158
P13	umc1963	4.04	29284075
P14	phi026	4.04	38554782
P15	bnlg2291	4.06	172523019
P16	umc1299	4.06	162657074
P17	bnlg1046	5.03	19174453
P18	umc1171	5.05	173408957
P19	umc1563	5.04	102015409
P20	umc1722	5.06	192005879
P21	phi389203	6.04	122283399
P22	umc2006	6.04	117494887
P23	umc2170	6.06	163859885
P24	umc1424	6.06	160678118
P25	umc2142	7.02	27988241
P26	umc1877	7.03	33056045
P27	dupssr13	7.05	170316101
P28	phi116	7.06	180131154
P29	umc1913	8.03	19243133
P30	umc1202	8.04	111651663
P31	umc1997	8.06	166553719
P32	umc2395	8.06	168954107
P33	dupssr6	9.03	12898158
P34	umc2370	9.04	38917446
P35	umc1771	9.05	129497025
P36	umc2346	9.06	147802538
P37	umc1432	10.03	5439594
P38	umc1863	10.03	13503823
P39	bnlg1250	10.06	133411724
P40	umc1993	10.06	142891007

rates for each marker. The recombination rates of bnlg1046, umc1563, umc1171 and umc1722 were 5.7%, 19.9%, 18.2%, and 33.7%, respectively (Table 6). Thus, *ms2016* was localized on chromosome 5 and was most closely linked to the bnlg1046 marker. In addition, we analyzed the other candidate markers (P22, P29, and P33) located on different chromosomes. A large number of black spots were observed on the electrophoretogram (Fig. 5; only shown for P33). The recombination rates for P22, P29, and P33 were 45.8%, 43.6%, and 40.2%, respectively (Table 6), which indicated that these markers were not linked to *ms2016*.

We developed 20 InDel markers located near the bnlg1046 marker (Table 7). Four InDel markers (I5-1, I5-3, I5-12 and I5-16) were polymorphic between C7-2 and T9047. These four InDel markers were used to detect the marker type of each recessive individual. The recombination rates of I5-1, I5-3, I5-12, and I5-16 were 10.6%, 7.8%, 11.2%, and 12.5%, respectively (Table 6), which further corroborated that the *ms2016* gene was located on chromosome 5. On the basis of the recombination rates and marker locations, the site of the *ms2016* locus was localized to an 8.7 Mb interval between markers I5-3 and I5-12 (Fig. 6).

Discussion

Mutants are important materials for forward genetics, which is a powerful approach to explore the genetic basis of a particular trait or phenotype (Mueller 2006). Thus, rapid localization of mutant genes is an important prerequisite for identifying the genes responsible for specific traits.

Performing a BSA is an indispensable step in gene mapping because it enables identification of DNA markers tightly linked to the target gene by using several bulked DNA samples instead of a large number of individuals (Giovannoni *et al*. 1991; Michelmore *et al.* 1991). However, given the inability to quantitatively evaluate the components in DNA bulks, a traditional BSA requires many high-density markers to screen a candidate marker tightly linked to the target gene, which is labor-intensive and time-consuming. In the present study, we combined an optimized semi-quantitative DNA analysis program and BSA (termed sq-BSA) to map the *ms2016* gene. The sq-BSA approach allowed evaluation of the proportions of marker-defined individuals (dominant or recessive marker types) in bulks. A small number of molecular markers were sufficient to identify a candidate marker partially linked to the target gene in the bulks. Only 40 polymorphic markers were used to perform the linkage analysis in recessive bulks and seven of these markers were rapidly identified as candidate gene-linked markers, of which four markers (bnlg1046, umc1563, umc1171, and umc1722) were located on chromosome 5. Through group validation, *ms2016* was anchored on chromosome 5 and most closely linked to bnlg1046. The present results indicated that sq-BSA is a viable method for the primary gene mapping with a small number of molecular markers. A similar principle has been applied in BSA-seq for rapid gene mapping (Abe *et al*. 2012; Takagi *et al*. 2013a, b; Fekih *et al*. 2013). Based on calculation of the discrepant SNP index, BSA-seq allowed quantitative evaluation of the proportions of parental types in the DNA bulks. Using a BSA-seq strategy, several maize genes have been mapped rapidly, including *tb1 enhancer* to a 0–20 Mb interval on chromosome 1 (Klein *et al*. 2018), the dominant dwarfing gene *rht-DM* to a 111.07–124.56 Mb region on chromosome 9 (Chen *et al*. 2018) and a resistance locus for *Gibberella* stalk rot was mapped to a 161.0–170.6 Mb interval on chromosome 8 (Chen *et al.* 2017). Therefore, two approaches, BSA-seq and sq-BSA, sharing the similar principle, achieve similar outcomes.

Notably, the maize genome (2.3 Gb) has undergone several rounds of genome duplication, with almost 85% of the genome composed of multiple transposable elements dispersed non-uniformly across the genome (Paterson *et al.* 2004, 2009; Swigonova *et al*. 2004; Schnable *et al*. 2009). The large size, complexity, and high repeatability of the maize genome pose potential risks to BSA-seq, possibly

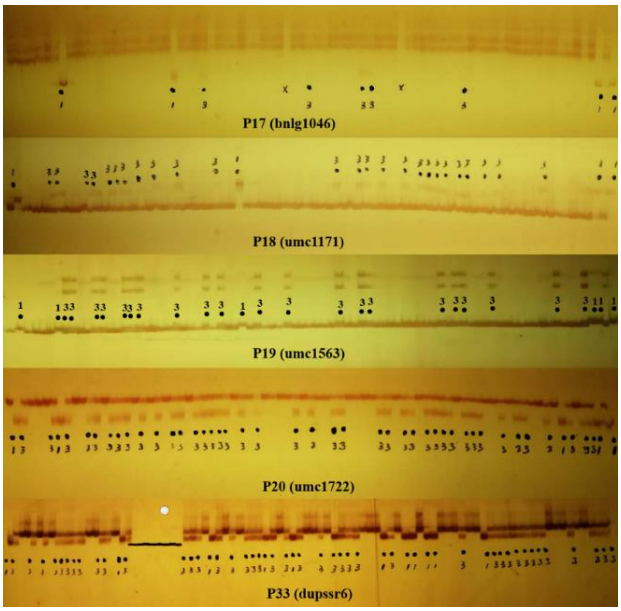

Fig. 5: Group validation of the key candidate markers. 1 represents the C7-2 band pattern, which indicates the dual gamete recombination type; while 3 represents the heterozygotic band pattern of parents (C7-2 and T9047), which indicates the single gamete recombination type. The others have the T9047 band pattern, which indicates the non-recombination type. A dot (·) indicates a recombinant individual; a cross (×) represents a missing band

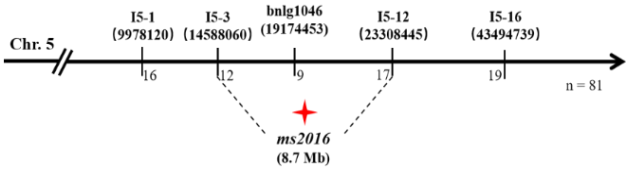

Fig. 6: Primary localization of *ms2016*. Labels above the vertical bars denote the linked markers. Numbers below the vertical bars indicate the numbers of recombinant individuals. n represents the total number of recessive individuals in the F_2 population. The red shuriken represents the target gene *ms2016*

disturbing the unbiased alignment of short reads and reducing the accuracy of evaluation of the SNP indices of bulks (Magwene *et al.* 2011). By contrast, sq-BSA relies on a set of specific markers, ensuring accurate evaluation of specific marker types in bulks. Indeed, we intentionally selected a small number of molecular markers located on one-third of each chromosome (from the head or tail), so that every chromosome was divided into three almost equal portions, which ensured that some markers should be close to the target gene. As a result, the target gene would be readily and rapidly identified by sq-BSA.

Conclusion

In this study, 30 ng/μL DNA template concentration and 30 PCR cycles were determined to be optimal for semi-quantitative DNA analysis. Using the sq-BSA approach, 40 polymorphic markers located on one-third of each chromosome (from the head or tail) were screened and used to map the *ms2016* gene. The gene was quickly anchored on chromosome 5 and was most closely linked to bnlg1046. Through verification using new InDel markers, *ms2016* was mapped to an 8.7 Mb interval (chr5:14588060-23308445).

Table 6: Linkage analysis of the candidate male sterility-related maize markers

	Chr.	Position	Total individuals	Recombinant individuals	Recombination rate
umc2006	6	117494887	81	57	45.8%
phi115	8	102415941	81	55	43.6%
dupssr6	9	12898158	81	52	40.2%
bnlg1046	5	19174453	81	9	5.7%
umc1563	5	102015409	81	27	18.2%
umc1171	5	173408957	81	29	19.9%
umc1722	5	192005879	81	45	18.2%
I5-1	5	9978120	81	16	10.6%
I5-3	5	14588060	81	12	7.8%
I5-12	5	23308445	81	17	11.2%
I5-16	5	43494739	81	19	12.5%

Table 7: Information of the 20 InDel molecular markers

InDel	Chromosome	Position	primer	InDel
I5-1	5	9978120	Left	CGTCCCATCTTTCACTGTGG
			Right	AGGAACGAATAGCAACGGC
I5-2	5	87517377	Left	GTCGGTGCTTACGATCAGTG
			Right	TTGCCTTCGTATTCTCCTCTG
I5-3	5	14588060	Left	TGCTCTCACAAGATGGTGGA
			Right	CCACAGGATAAAATCGGCTG
I5-4	5	15587601	Left	CTATGCCGAATTGCCGAT
			Right	GTTGCATTAGGCTAGGCCAT
I5-5	5	15737462	Left	TGGTTTTAAAGACGTGCTGC
			Right	TGCTCGAAGGAGGAGGAAA
I5-6	5	15738185	Left	TGCCTTTCTCTTTAGATGGC
			Right	GACTGGCGACGATACTGTCA
I5-7	5	17419900	Left	CTCTATTGAGCCACACAGGG
			Right	CTTTCGTGAGGCCTATTACTTC
I5-8	5	95116808	Left	TTGCCACCTTCTTTCACAA
			Right	AAGTTTGTTGACGCTGCTCC
I5-9	5	99418680	Left	CATCTCATACATGCGCCAAG
			Right	TGGTACCTGTTCTTGACCCG
I5-10	5	100285086	Left	CTGCAACCACTGGACAAAAA
			Right	TGCTCCTCCTGAAGAGTTGG
I5-11	5	21470105	Left	AAGCAGGTTATTGCACCCC
			Right	TTCTCCAAATTGAGAGCTTCC
I5-12	5	23308445	Left	TGCAACGCGTACGTAGAAGA
			Right	GCAGTAGCGTCTCTCGCAA
I5-13	5	26333998	Left	GCAGAAAGTGGACGAGATGG
			Right	CGGCCTTTCATCAGTTGTG
I5-14	5	19175223	Left	TTGGTGGTATGCATTAAGACG
			Right	CAATGATGGACATGGGCG
I5-15	5	19725680	Left	CGCAGGATCACCTATTCACC
			Right	GTCCGGTGTGAAAAGCCT
I5-16	5	43494739	Left	CTGATGTGGTGTGCCGATT
			Right	CAACCCTGCCTTTCTCTCTC
I5-17	5	37051650	Left	AACAGTGAGTGTCCGGTGC
			Right	AAGGTAGCCGTTGGGGAG
I5-18	5	41052694	Left	TGTTTCAAATGCAAAATGTCG
			Right	GACGCAGCTCAGAGACTTCA
I5-19	5	32774763	Left	TGGGCCTTTCTGAATTTTTG
			Right	CGAGAGGAAAACAGACGCA
I5-20	5	36878700	Left	GGTGTGGGAGAACGAACCTA
			Right	TGCAAGTAGGTGCGAGTTTG

Therefore, the sq-BSA method accelerates the gene mapping process in maize and provides a high-efficiency tool for forward genetic analysis.

Acknowledgements

We acknowledge the financial grant from Henan province young backbone teachers training plan (2016GGJS-034), Henan province basic and frontier technology research (162300410165) and Henan Natural Science Foundation (13A210469). We thank Robert McKenzie, PhD, from Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript.

Author Contributions

JT designed the research. MW, CM, XS and XW performed the experiments. XZ wrote the manuscript. LZ, ZG, ZF and WL revised the manuscript. All the authors read and approved the final manuscript.

References

Abe A, S Kosugi, K Yoshida, S Natsume, H Takagi, H Kanzaki, H Matsumura, K Yoshida, C Mitsuoka, M Tamiru, H Innan, L Cano, S Kamoun, R Terauchi (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol* 30:174–178

Chen Q, J Song, WP Du, LY Xu, Y Jiang, J Zhang, XL Xiang, GR Yu (2018). Identification and genetic mapping for *rht-DM*, a dominant dwarfing gene in mutant semi-dwarf maize using QTL-seq approach. *Genes Genom* 40:1091–1099

Chen Q, J Song, WP Du, LY Xu, Y Jiang, J Zhang, XL Xiang, GR Yu (2017). Identification, mapping, and molecular marker development for *Rgsr8.1:* A new quantitative trait locus conferring resistance to stalk rot in maize. *Front Plant Sci* 8; Article 1355

Das S, HD Upadhyaya, D Bajaj, A Kujur, S Badoni, Laxmi, V Kumar, S Tripathi, CL Gowda, S Sharma, S Singh, AK Tyagi, SK Parida (2015). Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. *DNA Res* 22:193–203

Fekih R, H Takagi, M Tamiru, A Abe, S Natsume, H Yaegashi, S Sharma, H Kanzaki, H Matsumura, H Saitoh, C Mitsuoka, H Utsushi, A Uemura, E Kanzaki, S Kosugi, K Yoshida, L Cano, S Kamoun, R Terauchi (2013). MutMap+: Genetic mapping and mutant identification without crossing in rice. *PLoS One* 8; Article e68529

Giovannoni JJ, RA Wing, MW Ganal, SD Tanksley (1991). Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucl Acids Res* 19:6553–6558

Huang L, W Tang, S Bu, W Wu (2020). A statistical method for QTL mapping based on bulked segregant analysis by deep sequencing. *Bioinformatics* 36:2150–2156

Imerovski I, B Dedić, S Cvejić, D Miladinović, S Jocić, G Owens, N Tubić, L Rieseberg (2019). BSA-seq mapping reveals major QTL for broomrape resistance in four sunflower lines. *Mol Breed* 39; Article 41

Itoh N, T Segawa, M Tamiru, A Abe, S Sakamoto, A Uemura, K Oikawa, H Kutsuzawa, H Koga, T Imamura, R Terauchi, H Takagi (2019). Next-generation sequencing-based bulked segregant analysis for QTL mapping in the heterozygous species *Brassica rapa*. *Theor Appl Genet* 132:2913–2925

Jay S, J Hanlee (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26:1135–1145

Klein H, Y Xiao, PA Conklin, R Govindarajulu, JA Kelly, MJ Scanlon, CJ Whipple, M Bartlett (2018). Bulked-segregant analysis coupled to whole genome sequencing (BSA-Seq) for rapid gene cloning in maize. *G3 Genes Genomics Genet* 8:3583–3592

Liu J, J Qu, C Yang, D Tang, J Li, H Lan, T Rong (2015). Development of genome-wide insertion and deletion markers for maize, based on next-generation sequencing data. *BMC Genomics* 16; Article 601

Maddocks S, R Jenkins (2017). *Quantitative PCR: Things to Consider; Understanding PCR*, pp:45–52. Academic Press, Boston, Massachusetts, USA

Magwene PM, JH Willis, JK Kelly (2011). The statistics of bulk segregant analysis using next generation sequencing. *PLoS Comput Biol* 7; Article e1002255

Matsui T, T Hasegawa (2019). Effect of long anther dehiscence on seed set at high temperatures during flowering in rice (*Oryza sativa* L.). *Sci Rep* 9; Article 20363

Michelmore RW, I Paran, RV Kesseli (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA* 88:9828–9832

Mueller R (2006). Ask the plant: Investigating and teaching plant structure. *Bot J Linn Soc* 150:73–78

Page DR, U Grossniklaus (2002). The art and design of genetic screens: *Arabidopsis thaliana*. *Nat Rev Genet* 3:124–136

Peters Jl, F Cnudde, T Gerats (2003). Forward genetics and map-based cloning approaches. *Trends Plant Sci* 8:484–491

Paterson AH, JE Bowers, R Bruggmann, I Dubchak, J Grimwood, H Gundlach, G Haberer, U Hellsten (2009). The sorghum bicolor genome and the diversification of grasses. *Nature* 457:551–556

Paterson AH, JE Bowers, BA Chapman (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci USA* 101:9903–9908

Rallapalli G, P Corredor-Moreno, E Chalstrey, M Page, D Maclean (2019). Rapid fine mapping of causative mutations from sets of unordered, contig-sized fragments of genome sequence. *BMC Bioinform* 20; Article 9

Schnable PS, D Ware, RS Fulton, JC Stein, F Wei, S Pasternak, C Liang, J Zhang, L Fulton, TA Graves, P Minx, AD Reily (2009). The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326:1112–1115

Swigonova Z, J Lai, J Ma, W Ramakrishna, V Llaca, Jl Bennetzen, J Messing (2004). On the tetraploid origin of the maize genome. *Compar Funct Genomics* 5:281–284

Takagi H, M Tamiru, A Abe, K Yoshida, A Uemura, H Yaegashi, T Obara, K Oikawa, H Utsushi, E Kanzaki, C Mitsuoka, S Natsume, S Kosugi, H Kanzaki, H Matsumura, N Urasaki, S Kamoun, R Terauchi (2015). MutMap accelerates breeding of a salt-tolerant rice cultivar. *Nat Biotechnol* 33:445–449

Takagi H, A Abe, K Yoshida, S Kosugi, S Natsume, C Mitsuoka, A Uemura, H Utsushi, M Tamiru, S Takuno, H Innan, LM Cano, S Kamoun, R Terauchi (2013a). QTL-seq: Rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J* 74:174–183

Takagi H, A Uemura, H Yaegashi, M Tamiru, A Abe, C Mitsuoka, H Utsushi, S Natsume, H Kanzaki, H Matsumura, H Saitoh, K Yoshida, LM Cano, S Kamoun, R Terauchi (2013b). MutMap-Gap: Whole-genome resequencing of mutant F_2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene Pii. *New Phytol* 200:276–283

Tazib T, Y Kobayashi, H Koyama, T Matsui (2014). QTL analyses for anther length and dehiscence at flowering as traits for the tolerance of extreme temperatures in rice (*Oryza sativa* L.). *Euphytica* 203:629–642

Trick M, NM Adamski, SG Mugford, CC Jiang, M Febrer, C Uauy (2012). Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol* 12; Article 14

Zheng WJ, Y Wang, LL Wang, ZB Ma, JM Zhao, P Wang, LX Zhang, ZH Liu, XC Lu (2016). Genetic mapping and molecular marker development for Pi65(t), a novel broad-spectrum resistance gene to rice blast using next-generation sequencing. *Theor Appl Genet* 129:1035–1044