Mol Breeding (2014) 34:1147–1158
DOI 10.1007/s11032-014-0106-3

Marker-assisted selection and evaluation of high oil in vivo haploid inducers in maize

Xin Dong · Xiaowei Xu · Liang Li · Chenxu Liu · Xiaolong Tian · Wei Li · Shaojiang Chen

Received: 5 December 2013 / Accepted: 30 April 2014 / Published online: 8 May 2014

Abstract Doubled haploid technology, which is used to rapidly purify genetic resources, is one of the key technologies in modern maize breeding. In a previous study, the major quantitative trait locus *qhir1*, which influences in vivo haploid induction, was narrowed down to a 243-kb region, which made it feasible to use marker-assisted selection (MAS) for inducer development. Recently, a new method was developed for haploid identification using oil content (OC). The objective of this study was to develop high oil inducer lines using MAS of the *qhir1* locus. We constructed an F_2 population, two backcross populations that were backcrossed to the inducer CAU5 (BC_1F_1-CAU5) and the high oil inbred line GY923 (BC_1F_1-GY923), respectively, which was derived from the cross GY923 × CAU5, and subjected continuous selfing to develop high oil inducer lines. In each cycle, three different parameters including kernel OC, marker genotype at *qhir1* and haploid induction rate (HIR) were used for pedigree selection. Three candidate high oil inducer lines were developed, with an OC of approximately 8.5 %, an HIR of approximately 8 % and superior agronomic performance, which are suitable values for the application of these lines to haploid identification by OC. Our results confirm the notion that HIR selection combined with MAS for *qhir1* is an effective approach to haploid inducer breeding. In addition, we determined that the accuracy of haploid identification by OC is influenced by the female germplasm resource and the high oil inducer and that appropriate critical points for OC can balance the false discovery rate and false negative rate.

Keywords Haploid induction rate · *qhir1* · Marker-assisted selection · Oil content · Haploid identification

Electronic supplementary material The online version of this article (doi:10.1007/s11032-014-0106-3) contains supplementary material, which is available to authorized users.

X. Dong · X. Xu · L. Li · C. Liu · X. Tian · W. Li · S. Chen (✉)
National Maize Improvement Center of China, China Agricultural University, Yuanmingyuan West Road, Haidian District, Beijing 100193, China
e-mail: shaoj@cau.edu.cn

S. Chen
Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Yuanmingyuan West Road, Haidian District, Beijing 100193, China

Introduction

The inbred-hybrid concept paved the way for the development of homozygous lines, an important part of commercial maize breeding programs (Hallauer et al. 2010). In contrast to conventional procedures used to produce homozygous lines, which take at least 6–8 generations to produce lines with approximately 99 % homozygosity, doubled haploid (DH) technology, which is based on in vivo haploid induction, can

be used to rapidly purify breeding resources in the course of only two generations. DH technology greatly reduces the time required for breeding and has gradually become a routine tool used in modern maize breeding (Geiger 2009).

Maize maternal haploids can be generated when female materials are pollinated by inducers with specific genotypes through in vivo haploid induction. The first maize inducer, Stock 6, produced a haploid frequency as high as 3.2 % (Coe 1959). Modern maize inducers, which have haploid induction rates (HIR) of approximately 8–10 % on average, have been produced through different cross-combinations with Stock 6 during the past 20 years, which has improved the development of DH technology (e.g., Chalyk 1994; Lashermes and Beckert 1988; Röber et al. 2005; Prigge et al. 2011; Xu et al. 2013).

Screening haploid kernels from testcross seeds produced by pollinating the testers is important for commercial scale use of DH technology (Geiger 2009). The most efficient marker for haploid identification is regulated by the dominantly inherited gene *R1-navajo* (*R1-nj*), conferring purple coloration expressed in both the scutellum and the aleurone (Nanda and Chase 1966; Neuffer et al. 1997). The *R1-nj* gene is highly expressed in most genetic backgrounds but is not always expressed in tropical materials or in European flint germplasm containing dominant genes such as *C1-I*, *C2-Idf* and *In1-D*, which inhibit the expression of *R1-nj* (Röber et al. 2005; Belicuas et al. 2007). Chen and Song (2003) found that diploid crossing seeds have more than 30 % higher oil content (OC) than haploid seeds in lines pollinated by the high oil inducer CAUHOI; the large difference in OC between crossing seeds and haploid seeds makes it possible to identify haploids based on OC. Rotarenco et al. (2007) also reported that the OC of haploids was significantly lower than that of crossing seeds in lines crossed with the normal inducer MHI. To meet the demands of large-scale application of DH technology, an automatic screening system based on nuclear magnetic resonance (NMR) was developed to identify haploids among testcross seeds based on OC (Liu et al. 2012). Melchinger et al. (2013) validated a method used to identify haploids from testcross seeds pollinated by a high oil inducer and discussed several factors that affect this method, which provides a more widely applicable and reliable high-throughput system for DH production.

The HIR is a quantitative trait controlled by a small number of genes (Lashermes and Beckert 1988). The major quantitative trait locus (QTL) *qhir1*, located on chromosome 1, was first detected in the inducer Stock 6 (Deimling et al. 1997; Röber 1999). A segregating population from the inducer line PK6 and the non-inducer line DH99 was then analyzed, revealing that this locus is located in bin 1.04 and segregation distortion against the inducer parent is present at this locus (Barret et al. 2008). Using the inducers CAUHOI and UH400, genome-wide QTL analysis of HIR led to the identification of eight QTLs, including the *qhir1* locus, which explained up to 66 % of the genotypic variance (Prigge et al. 2012). This major QTL was detected in several different studies using different haploid inducers from different backgrounds. We narrowed down the *qhir1* locus to a 243-kb region based on analysis of the B73 reference genome using a large F_2 population derived from the cross 1680 × UH400, and we developed numerous closely linked markers that are useful for marker-assisted selection (MAS) of new inducers (Dong et al. 2013).

There are no previous reports about MAS of new inducers, and our previous study involving fine mapping of the *qhir1* locus has given us the opportunity to perform MAS of new inducers. In addition, high oil inducers can provide us with new opportunities to identify haploids by examining the oil xenia effect in high oil maize, and the application of a high oil inducer to identify haploids from different genetic backgrounds has not previously been reported. Therefore, our objectives were to (1) develop high oil inducers using MAS of the *qhir1* locus, (2) compare different selection strategies under the condition of two trait selection, (3) evaluate the newly developed high oil inducers and (4) assess the method of haploid identification by OC.

Materials and methods

Breeding scheme and evaluation of high oil inducers

GY923 is a high oil inbred line from the Alexander high oil population; its kernel OC is 10.66 %, and it lacks induction ability. CAU5 (developed by China Agricultural University) is a haploid inducer with an HIR of 10 %, and its kernel OC is 3.50 % (Xu et al.

2013). Crosses were made between GY923 and CAU5 to generate F_1 hybrids. The F_1 hybrids were (a) selfed to produce the F_2 generation, (b) backcrossed to the inducer CAU5 to produce the first BC generation (BC_1F_1-CAU5) and (c) backcrossed to the high oil inbred line GY923 (as the female parent) to produce the first BC generation (BC_1F_1-GY923), and followed by continuous ear-to-row cultivation of selected progeny and assessment of three different parameters including OC, marker genotype at *qhir1* and HIR (Fig. S1). These three strategies were designated as the F_2 strategy, BC_1F_1-CAU5 strategy and the BC_1F_1-GY923 strategy, respectively, and details about the assessment of the three different parameters were as follows. First, the OC was measured, and high oil seeds were selected and planted in the field. Second, all of the plants were genotyped using markers X18 and X109, and located in the *qhir1* region. Plants carrying the *qhir1* locus were selfed and tested using the tester hybrid ZD958, which produces clear expression of the *R1-nj* marker in both embryos and endosperm when pollinated by inducers (Li et al. 2009; Prigge et al. 2012; Xu et al. 2013; Dong et al. 2013). Then, the lines with high HIRs (i.e., exceeding 5 %) were advanced to the next cycle. Using this selection method, a total of 1,117 F_2, 1,574 BC_1F_1-CAU5 and 689 BC_1F_1-GY923 plants were advanced to two $F_{4:5}$ and one BC_1F_4-GY923 lines which were chosen as candidate high oil inducer lines (CHOI), with OCs of approximately 8.5 % and HIRs of approximately 8 % (Table 2; Fig. S1).

Then, two candidate inducer lines, candidate high oil inducer 1 (CHOI1) and candidate high oil inducer 3 (CHOI3), were used to evaluate the HIR in Hainan, China, in Winter 2012. Since the *R1-nj* marker is clearly expressed in both the embryo and endosperm when pollinated by inducers, seven elite maize inbred lines were used as testers to check the HIR and the accuracy of haploid identification by OC. These seven inbred lines were Zheng58, Te7922, Mo17, Q319, Chang7-2, Jing24 and 9801, which were from different genetic backgrounds.

Genotyping

Markers flanking the *qhir1* region were tested for polymorphism between GY923 and CAU5 (Dong et al. 2013). Since the markers X18 and X109 cover the entire *qhir1* region, they were selected for MAS for HIR; the distance between these markers is 800 kb, based on the B73 physical map. All of the plants examined during the process of developing high oil inducers were genotyped using markers X18 and X109. DNA extraction was performed as described by Murray and Thompson (1980). Each DNA sample was analyzed by PCR amplification and electrophoretically on 1 % agarose gels or 6 % denaturing polyacrylamide gels to determine the genotypes of the plants.

Haploid identification and oil measurement

Two methods were used to determine the HIR. First, for the process of high oil inducer line development, the color of the embryo, which is affected by the dominant gene *R1-nj*, was examined to identify haploids having the same color in the aleurone layer (Nanda and Chase 1966; Neuffer et al. 1997). The inducer CAU5 has a deep purple pigmentation in the embryo and purple aleurone color in the endosperm. Every plant of the segregation populations of GY923-CAU5 containing the *R1-nj* gene was selected and used to pollinate three ears of the tester hybrid ZD958. The putative haploids had purple endosperm and colorless embryos, and the diploid crossing seeds had purple endosperm and embryos (Li et al. 2009). HIR was determined by examining the proportion of haploids in each testcross. The HIR was calculated using the following formula: HIR = No. of putative haploids/total No. of *R1-nj* normal kernels × 100 %. There were three replications for each plant of the segregation populations, and the HIR for each genotype was calculated as the mean HIR of three replications.

Second, to evaluate the candidate high oil inducers, ZD958 and seven inbred lines were pollinated, and all of the seeds were sorted using the above method. Then, all of the putative haploid and diploid kernels were examined using a nuclear magnetic resonance (NMR) mq20 instrument (Bruker, Germany) (Song and Chen 2004) to obtain information about OC. All putative haploids (based on OC) were grown in the field in the ShangZhuang experiment station in Beijing, Summer 2013, to confirm their ploidy status by visual scoring. The haploids displayed shorter stature, erect and narrow leaves and reduced growth rates. All seeds in the field that did not germinate were assumed to be haploids because they were distinctly

weaker than crossing seeds. (The embryo-less seeds with OC levels below 1 % were not planted in the field.) Based on the methods of Kebede et al. (2011), the HIR was calculated using the following formula: (No. of putative haploids/total No. of R1-nj normal kernels) × [1 − (No. of diploids in the field that were misclassified and selected/No. of putative haploids)] × 100 %. The false discovery rate (FDR) and false negative rate (FNR) were calculated according to the method of Melchinger et al. (2013): FDR = No. of haploids with OC exceeding the critical point/No. of true haploids; FNR = No. of crossing seeds with OC below the critical point/No. of kernels below the critical point.

Survey of agronomic traits

Ten plants of three candidate high oil inducer lines and the normal inducer line CAU5 were randomly chosen for phenotyping. A total of five traits were evaluated in the field, including tassel length, tassel branch number, plant height, ear height and days to anthesis. Plant height was measured from the soil surface to the apex of the tallest panicle. Ear height was measured from the soil surface to the upper ear attachment. Tassel length was measured from the non-branching node present below the lower most primary branch to the tip of the central spike. Tassel branch number was calculated as the number of primary branches. Days to anthesis was recorded as the number of days from sowing to the time when pollen shedding occurred for more than half of the individuals of a single entry. All phenotypic data analysis was performed with R i386 3.0.1 software.

Results

Marker-assisted selection (MAS) for *qhir1*

Among the two initial F_2 and BC_1F_1-GY923 populations, the proportion of plants with the *qhir1* genotype including homozygous and heterozygous for the inducer genotype in the *qhir1* region was significantly less than 75 and 50 %, respectively, which is inconsistent with Mendel's laws (Table S1). This result suggested that the *qhir1* locus can cause strong segregation distortion against the inducer genotype. We performed a Wilcoxon rank-sum test to identify the differences in HIR for two or three genotypes in different populations. The HIR for plants with class A genotypes (lacking the inducer haplotype from CAU5 at *qhir1)* was significantly ($P < 0.01$) lower than that for plants with class B (homozygous for the inducer haplotype at *qhir1*) or H (heterozygous) genotypes in the F_2, F_3, BC_1F_2-CAU5, BC_1F_1-GY923 and BC_1F_2-GY923 populations (Table 1). The HIR for plants with genotype H was significantly ($P < 0.01$) lower than that of plants with class B genotypes in the F_2, F_3, BC_1F_1-CAU5, BC_1F_2-CAU5 and BC_1F_2-GY923 populations (Table 1).

During pedigree selection for three parameters, the proportion of plants with class B and H genotypes was quickly improved, and the class B genotype was ultimately fixed in the population. The HIRs for this genotype improved to 8.59, 7.36 and 6.70 % in the F_4, BC_1F_3-GY923 and BC_1F_3-CAU5 populations, respectively (Table 1). Initially, the BC_1F_1-GY923 population had a lower HIR (on average) than BC_1F_1-CAU5 and that of F_2 was between the other values. The HIR significantly ($P < 0.01$) increased from F_3 to F_4, BC_1F_2-GY923 to BC_1F_3-GY923 and BC_1F_2-CAU5 to BC_1F_3-CAU5, but there was no significant difference between F_2 and F_3, BC_1-GY923 and BC_1F_2-GY923 or BC_1-CAU5 and BC_1F_2-GY923 at a confidence interval of 0.05 (Fig. S2).

OC variation under HIR selection

When the selection of two traits (OC and HIR) was combined, the OC showed a significant ($P < 0.01$) difference between the initial and last cycles in the F_2, BC_1F_1-CAU5 and BC_1F_1-GY923 strategies. The OC increased from 6.71 % in F_2 to 8.10 % in F_6, from 5.07 % in BC_1F_1-CAU5 to 5.58 % in BC_1F_4-CAU5 and from 7.90 % in BC_1F_1-GY923 to 8.22 % in BC_1F_5-GY923 (Table 2). The initial OC exhibited a significant ($P < 0.01$) difference among the F_2, BC_1F_1-GY923 and BC_1F_1-CAU5 populations. The F_2 strategy yielded intermediate initial OC and the largest selection response, while the BC_1F_1-GY923 strategy yielded a higher initial OC and the lowest selection response and the BC_1F_1-CAU5 strategy yielded a low initial OC and an intermediate selection response (Table 2; Fig. S1). Compared to the BC_1F_1-GY923 and BC_1F_1-CAU5 strategies, in the F_2 strategy, the OC showed a gradual, significant ($P < 0.01$) increase from F_2 to F_6, except during the F_3 to F_4

Table 1 Mean, standard deviation (SD), range for haploid induction rate (HIR), tester ears (TE) and testcross seeds (TS) produced by pollinating the tester ZD958 among different genotypes based on the markers X18 and X109 in the *qhir1* region for different generations derived from the cross GY923 × CAU5

Generation	Genotype[a]	HIR %			Test cross	
		Mean[b]	SD	Range	TE	TS
F_2	A	0.44a	0.35	0.00–1.23	18	11,552
	H	3.13b	1.66	0.00–7.00	42	16,425
	B	6.67c	3.61	1.77–12.50	8	3,280
F_3	A	0.42a	0.6	0.00–2.38	49	22,688
	H	2.43b	1.97	0.00–7.98	76	34,414
	B	5.87c	3.39	0.99–18.60	90	24,675
F_4	B	8.59	1.91	5.28–13.29	49	14,772
BC_1F_1-CAU5	H	3.46a	1.97	0.44–10.54	45	20,659
	B	5.75b	2.75	1.23–13.04	28	9,834
BC_1F_2-CAU5	A	0.77a	0.72	0.00–2.78	13	4,103
	H	2.24b	1.75	0.00–7.04	44	19,854
	B	5.62c	3.4	1.34–14.29	64	20,097
BC_1F_3-CAU5	B	6.70	1.19	5.61–8.70	7	2,992
BC_1F_1-GY923	A	0.42a	0.65	0.00–1.79	10	2,151
	H	2.07b	1.53	0.00–4.85	17	3,777
BC_1F_2-GY923	A	0.24a	0.27	0.00–0.86	15	7,089
	H	1.74b	1.48	0.00–4.69	16	7,539
	B	5.18	n.a.[c]	n.a.[c]	1	238
BC_1F_3-GY923	B	7.36	1.78	5.00–10.37	10	2,963

[a] All plants were classified into three different genotypes based on genotyping by markers X18 and X109 in the *qhir1* region. A (=homozygous for the GY923 haplotype), B (=homozygous for the CAU5 haplotype) or H (=heterozygous)

[b] For each generation, including F_2, F_3, BC_1F_1-CAU5, BC_1F_2-CAU5, BC_1F_1-GY923 and BC_1F_2-GY923, a Wilcoxon rank-sum test was performed to identify the difference for two or three genotypes. Numbers in a line followed by the same letter are not significantly different from each other at the 1 % probability level

[c] *n.a.* Not available

process. Finally, the OCs for F_6 and BC_1F_5-GY923 exceeded 8 %, which indicates that these lines are candidate high oil inbred lines (CHOI) for use in haploid induction.

Agronomic traits, HIR and OC for the CHOI

The potential inducers demonstrated a higher OC and HIR and superior agronomic performance compared with the normal inducer line CAU5 when grown in Beijing in Summer 2012 and in Hainan in Winter 2013 (Table 3; Fig. S3). The OC of three candidate inducers was greater than 8.5 %, which was significantly higher ($P < 0.001$) than that of CAU5. The HIR for three candidate inducers was higher than that of CAU5 in Beijing and lower than that of CAU5 in Hainan. Since there were more testcross seeds detected in Hainan than in Beijing, we deemed that the HIR was approximately 7 % on average, which is a bit lower than that of CAU5, based on analysis using the tester ZD958 (Table 3). Similarly, all the three candidate inducers carried the homozygous *R1-nj* marker with excellent purple embryo and endosperm coloration (Fig. S3). Compared with CAU5, all the three candidate inducers had a better seed set, germination rates and pollen production (data not shown). The anthesis date was delayed for the three candidate inducers compared to that of CAU5, especially for CHOI3, which was backcrossed to GY923 at both locations. The plant height, ear height and tassel size for the three candidate inducers were significantly higher than those of CAU5 (Table 3). The tassel

Table 2 Mean (±SD) of all kernels and selected kernels, selection difference and selection response for the different generations derived from the cross GY923 × CAU5

Generation	OC %			
	Mean of all kernels ± SD[a] (N^b)	Mean of selected kernels ± SD (N^b)	Selection difference	Selection response
F_2	6.71 ± 0.98a (1,117)	8.1 ± 0.55 (230)	1.39	n.a.[c]
F_3	6.96 ± 1.38b (992)	7.79 ± 0.60 (538)	0.83	0.25
F_4	6.86 ± 1.01b (1,161)	7.73 ± 0.57 (518)	0.87	−0.1
F_5	7.90 ± 0.82c (3,016)	9.38 ± 0.29 (271)	1.48	1.04
F_6	8.10 ± 1.87d (3,827)	n.a.[c]	n.a.[c]	0.20
BC_1F_1-CAU5	5.07 ± 0.80a (1,574)	6.48 ± 0.49 (186)	1.41	n.a.[c]
BC_1F_2-CAU5	6.02 ± 1.28c (713)	6.82 ± 0.60 (384)	0.80	0.95
BC_1F_3-CAU5	5.65 ± 0.81b (118)	6.47 ± 0.51 (42)	0.82	−0.37
BC_1F_4-CAU5	5.58 ± 0.76b (805)	n.a.[c]	n.a.[c]	−0.07
BC_1F_1-GY923	7.90 ± 1.06a (689)	9.12 ± 0.53 (194)	1.22	n.a.[c]
BC_1F_2-GY923	8.29 ± 1.82b (230)	9.52 ± 0.78 (112)	1.23	0.39
BC_1F_3-GY923	7.76 ± 0.87a (171)	8.92 ± 0.42 (35)	1.16	−0.53
BC_1F_4-GY923	8.24 ± 0.90b (587)	9.45 ± 0.40 (115)	1.21	0.48
BC_1F_5-GY923	8.22 ± 0.82b (1,600)	n.a.[c]	n.a.[c]	−0.02

[a] For the F_2, BC_1F_1-CAU5 and BC_1F_1-GY923 strategies, a Tukey's Honestly Significantly Different (HSD) test was performed to identify the difference among different selection generations. Numbers in a line followed by the same letter are not significantly different from each other at the 1 % probability level

[b] No. of kernels examined

[c] *n.a.* Not available

branch number of the three candidate inducers was significantly higher than that of CAU5 in Hainan but not in Beijing, except for CHOI1 (Table 3).

Haploid identification by oil content

We used 14 cross-combinations pollinated by two inducers (CHOI1 and CHOI3) to evaluate HIR, and we examined the accuracy of haploid identification via OC. The HIR for CHOI3 was higher than that for CHOI1, except in the evaluation using the tester Te7922; the HIR for CHOI1 and CHOI3 ranged from 5.38 and 12.84 to 4.06 and 19.12 %, respectively, using the seven testers (Table 4). There was a significant difference ($P < 0.001$) in OC between haploid and crossing seeds for all cross-combinations; the difference in the mean OC of haploid and crossing seeds ranged from 1.38 to 2.66 % (Table 4). For each female parent, there was no significant difference ($P < 0.01$) between the OC of haploids from two high oil inducers, while the OC of crossing seeds pollinated by inducer CHOI1 was significantly higher than that of crossing seeds pollinated by inducer CHOI3 (Table 4).

The female parents (9801, Zheng58, Mo17 and Q319) exhibited large differences in the mean OC of haploid versus crossing seeds, which often led to a lack of overlap between the OC distributions of haploid and crossing seeds (Fig. 1; Table 4). This difference in OC enabled better discrimination between haploid and crossing seeds, with FDR and FNR values of 0 % at the critical points between the maximum OC for haploids and the minimum OC for crossing seeds (Table S2). Other female germplasm (Te7922, Jing 24) exhibited minor differences in the mean OC of haploid versus crossing seeds, which caused partial overlap between the OC distributions of haploid and crossing seeds and revealed a negative relationship between FDR and FNR (Fig. 1; Table 4). If we controlled the FDR at a level of 10 %, the FNR for all four cross-combinations was below 11 % (Table S2). The difference in the mean OC of haploid and crossing seeds was significantly ($P < 0.001$) negatively correlated with the corresponding OC of

Table 3 The means (±SD) of kernel oil content (OC), haploid induction rate (HIR), days to anthesis (DTA), tassel size (TAS), tassel branch (TB), plant height (PH) and ear height (EH) among three candidate high oil inducers (CHOI1, CHOI2, CHOI3) and normal inducer CAU5 in Beijing (BJ) in Summer 2012 and Hainan (HN) in Winter 2013

Location	Inducer	Pedigree	Traits[a]						
			Mean of OC ± SD % (N^b)	HIR % (N^c)	DTA	TAS	TB	PH	EH
BJ	CK	CAU5	3.43 ± 0.38 (100)	8.30 (1,663)	58	26.3	11.3	146	35.1
	CHOI1	(GY923 × CAU5)-B-40-21-4	8.71 ± 0.67*** (83)	9.85 (132)	60	30.1***	14.6***	170***	59.5***
	CHOI2	(GY923 × CAU5)-B-40-6-27	8.88 ± 0.71*** (188)	7.93 (391)	62	33.6***	12.8	186***	60.6***
	CHOI3	(GY923 × (GY923 × CAU5))-57-67-13	8.88 ± 0.74*** (57)	11.32 (106)	66	29.7*	12.4	202***	79.6***
HN	CK	CAU5	3.14 ± 0.43 (100)	8.34 (1,427)	56	24.3	5.7	135	31.5
	CHOI1	(GY923 × CAU5)-B-40-21-4-1	8.75 ± 0.79*** (178)	7.56 (1,177)	59	29.3*	10.2***	170***	42.0***
	CHOI2	(GY923 × CAU5)-B-40-6-27-14	9.39 ± 0.70*** (84)	6.41 (608)	61	29.1*	7.7*	189***	61.9***
	CHOI3	(GY923 × (GY923 × CAU5))-57-67-13-7	8.56 ± 0.76*** (159)	7.27 (470)	65	28.9**	11.5**	175***	73.3***

[a] A two-tailed *t* test was used to test means of traits such as KOC, TAS, TB, PH and EH between candidate high oil inducer and normal inducer CAU5. The HIR was calculated based on the tester, maize hybrid ZD958

[b] No. of kernels examined

[c] No. of testcross kernels

*, **, *** Significant at $P < 0.05$, 0.01 and 0.001, respectively

haploids, which affected the distribution of OC for haploid and crossing seeds (r^2sp $= -0.73$).

Discussion

HIR selection combined with MAS for *qhir1*

MAS, an indirect selection process, is widely used in breeding to select genes that control complex traits that are difficult to phenotype (Collard and Mackill 2008; Timonova et al. 2013). The phenotype for HIR is difficult to detect using traditional pedigree selection, as observing this phenotype requires distinguishing haploids from diploid crossing seeds, which is not only time-consuming but also tester dependent. However, the use of MAS for haploid inducer breeding can overcome these problems. Normally, the main factors that determine the efficiency of MAS include the selection strategy, population size and availability and location of markers relative to the target gene (Frisch et al. 1999; Hospital 2005; Herzog and Frisch 2011). The advantage of MAS allows higher selection intensity, especially for haploid inducer breeding, for the following reasons: (1) It is subjected to large effect of *qhir1* locus, which can explain up to 66 % of the genotypic variance in the population derived from the cross 1680 × UH400 (Prigge et al. 2012); (2) the developed markers are closely linked with the *qhir1* locus within an 800-kb region (based on the B73 map), which greatly increase the reliability of the markers for predicting the HIR phenotype; and (3) the *qhir1* locus can cause strong segregation distortion against the inducer genotype. This segregation distortion, eliminating the favorable HIR-enhancing haplotype, has negative effects on maintaining the inducer genotype for haploid inducer breeding, As a result, there are far more phenotypes in the segregation populations with lower HIRs than phenotypes with higher HIRs between crosses of non-inducer and inducer lines (Barret et al. 2008; Prigge et al. 2012; Xu et al. 2013; Dong et al. 2013). In our study, there were 58 plants without the *qhir1* locus among 126 F_2 plants examined, and there were 93 plants without the *qhir1* locus among 127 BC_1GY923 plants examined; the HIRs of these plants were nearly 0 % (Table 1, Table S1). Phenotyping these plants is both time-consuming and costly. Therefore, in haploid inducer breeding, it is necessary to make full use of MAS to maintain *qhir1* locus to improve selection intensity.

Table 4 The mean (±SD), range of oil content (OC) for haploid and crossing (C) seeds, difference in the mean OC % of haploid and crossing seeds and haploid induction rate (HIR) for 14 cross-combination

Cross-combination	Traits					
	Mean of haploid[a] OC % ± SD	Range of haploid OC %	Mean of C seeds[b] OC % ± SD (N^c)	Range of C seeds OC %	Difference in the mean OC % of haploid and C seeds	HIR %
Zheng58/CHOI1	2.74 ± 0.34	2.13–3.32	5.40 ± 0.64*** (250)	4.00–7.27	2.66	6.37
Zheng58/CHOI3	2.68 ± 0.40	1.94–3.41	4.63 ± 0.45 (750)	3.57–6.32	1.95	7.52
9801/CHOI1	2.59 ± 0.26	1.91–2.96	5.10 ± 0.60*** (466)	3.23–6.47	2.51	6.80
9801/CHOI3	2.58 ± 0.32	1.94–3.23	4.85 ± 0.53 (563)	3.81–6.95	2.27	7.25
Mo17/CHOI1	3.23 ± 0.23	2.69–3.70	5.10 ± 0.40* (258)	4.18–6.28	1.87	12.84
Mo17/CHOI3	3.23 ± 0.21	2.62–3.69	5.03 ± 0.33 (255)	4.01–5.92	1.80	18.53
Chang7-2/CHOI1	3.18 ± 0.30	2.54–3.72	5.45 ± 0.68*** (434)	3.87–7.17	2.27	7.46
Chang7-2/CHOI3	3.10 ± 0.23	2.64–3.52	5.27 ± 0.55 (451)	3.54–6.91	2.17	7.01
Q319/CHOI1	4.04 ± 0.23	3.43–4.48	6.15 ± 0.44*** (530)	4.73–7.21	2.11	11.52
Q319/CHOI3	4.01 ± 0.32	2.92–4.66	6.08 ± 0.41 (567)	4.86–7.20	2.07	19.12
Jing24/CHOI1	4.10 ± 0.28	3.34–4.55	5.61 ± 0.51*** (507)	4.43–6.93	1.51	10.42
Jing24/CHOI3	4.02 ± 0.33	3.09–4.67	5.40 ± 0.52 (561)	4.24–7.30	1.38	12.21
Te7922/CHOI1	4.20 ± 0.28	3.70–4.74	5.90 ± 0.56*** (616)	4.62–7.25	1.70	5.38
Te7922/CHOI3	4.28 ± 0.36	3.53–4.71	5.79 ± 0.50 (591)	4.57–7.48	1.51	4.06

[a] For the same female, a two-tailed t test was performed to identify the significant difference in OC for haploids between two candidate inducers

[b] A one-tailed t test (H_0: $\mu_A = \mu_B$ vs. H_A: $\mu_{B<}\mu_A$) was also used to identify the significant difference in OC for crossing seeds between candidate inducer lines CHOI1 and CHOI3

[c] No. of kernels examined

*, **, *** Significant at $P < 0.05$, 0.01 and 0.001, respectively

However, the phenotypic variation for HIR among individuals homozygous plants at *qhir1* locus was still large, and it was larger than the phenotypic variation for HIR among homozygous individuals without the *qhir1* locus. In a previous study, two major QTLs (*qhir1*, *qhir8*) combined with six minor QTLs were detected and found to control HIR through four population using two inducers CAUHOI and UH400 (Prigge et al. 2012), Therefore, HIR is a quantitative trait and controlled by several loci. In spite of the large effect of the *qhir1* locus, selection for one locus was not sufficient for developing inducers with high HIRs. The major locus (*qhir1*) can be quickly fixed by MAS, and another major QTL (*qhir8*) and six small-effect modifier QTLs can be integrated together to further improve the HIR for inducers through phenotype selection. Therefore, phenotype selection combined with MAS selection for the *qhir1* locus is an effective way to further improve HIRs for inducers.

The strategy for developing high oil inducers

Maize OC is a quantitative trait controlled by more than 20 QTLs, which together explain more than 80 % of phenotypic variation (Cook et al. 2012; Li et al. 2012). OC can be improved through long-term selection, which enables the accumulation of many genes with small effects (Dudley and Lambert 2004; Song and Chen 2004; Laurie et al. 2004). Therefore, the best ways to select for oil traits include both phenotype and genomic selections. We previously found that it is feasible to improve kernel OC by continuously selecting high oil kernels using F_2 strategy, which demonstrates that the additive effect is important for oil accumulation (Li et al. 2012). Therefore, in the current study, continuous selection was the major strategy used to select OC phenotypes. Unlike OC, HIR is a quantitative trait that is influenced by two major QTLs and several minor QTLs (Prigge et al.

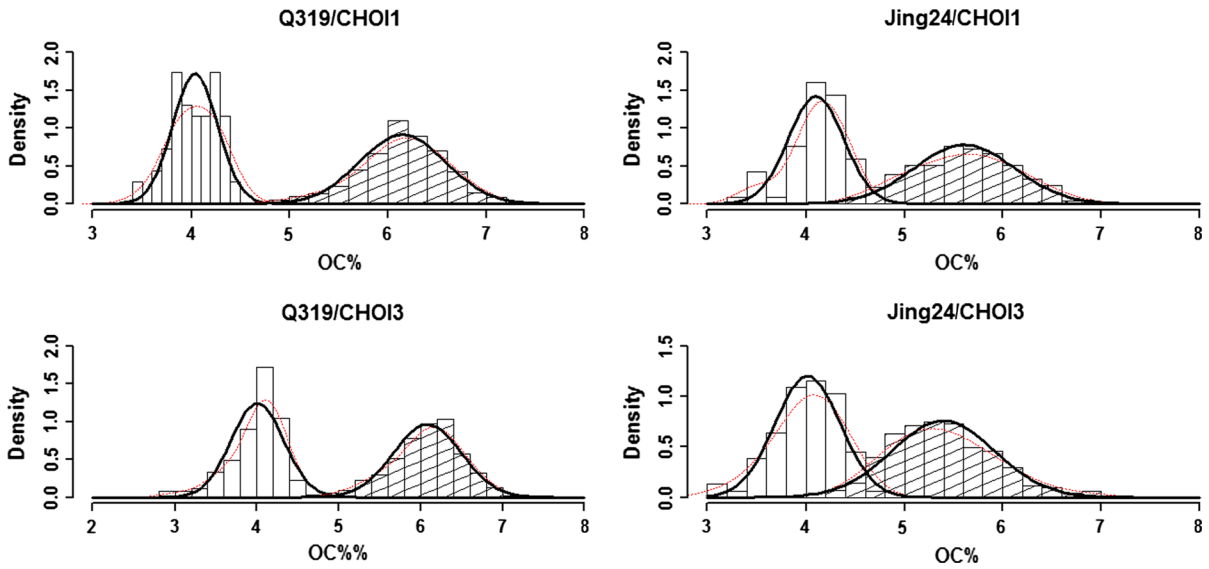

Fig. 1 The distribution density curve of OC for haploids and crossing seeds, the norm distribution curve based on the mean = mean (OC for haploids or crossing seeds) and standard deviation (SD) = SD (OC for haploids or crossing seeds) in 14 cross-combinations

2012). Our results show that HIR can be improved quickly via continuous HIR selection combined with MAS for the *qhir1* locus (Fig. S2). These results suggest that the two major genes quickly become fixed in the population and gains from selection that depend on small-effect modifier genes gradually become fixed through continuous phenotype selection. Strong selection pressure in the early generation appears to be more suitable for HIR than for OC. Our results also show that the CHOI3 line developed from backcrossing to the non-inducer line GY923 had a higher HIR than the CHOI1 line, supporting the notion that a high HIR in an inducer will not be sacrificed if the inducer is backcrossed to a non-inducer line, which is consistent with the conclusions of a previous report (Prigge et al. 2012). Therefore, the strategy of F_2 and BC_1-GY923 production is suitable for high oil inducer breeding, as this strategy can greatly increase OC without sacrificing HIR. Since these two quantitative traits are conferred by numerous QTLs, the strategy used to develop high oil inducers was to (1) construct a large F_2 population and a large population from a high oil inducer line to produce high HIR individuals and (2) integrate phenotype selection and MAS selection to improve HIR and increase oil by phenotype selection with high OC.

Agronomic traits of high oil inducers

Haploid inducers are difficult to maintain and have poor seed set because inducers have the common ancestor Stock 6 and the *sed1* locus, which influence kernel development after pollination (Xu et al. 2013). The high oil inducers developed in our study have taller plant height, larger tassel size and better pollen production than previous lines, and they provide sufficient pollen for haploid induction and enable the use of open pollination. Moreover, these inducer lines exhibit delayed maturity, with similar flowering times to those of most temperate germplasm resources, which precludes the need to stagger the planting of inducer lines multiple times to coincide their flowering times with that of conventional lines. Therefore, the newly developed high oil inducer lines will enhance the efficiency of large-scale haploid induction of haploid for temperate maize.

Methods for haploid identification

There are many methods used to identify haploids include cytogenetic methods (counting the chromosome number), flow cytometry (analyzing the amount of nuclear DNA), analysis of SSR marker for haploid

and crossing seeds, *R1-nj* marker system, and OC marker system (Antoine-Michard and Beckert 1997; Barret et al. 2008; Melchinger et al. 2013). In order for haploid identification approach to be used on a commercial scale, an efficient screening system must be in place that enables one to differentiate among haploids from crossing seeds at the seed or seedlings level (Geiger 2009). Since methods such as cytogenetic methods, flow cytometry, and analysis of SSR marker are costly, time-consuming and labor intensive, these methods are not suitable for haploid identification on a commercial scale. In general, the *R1-nj* marker-based haploid identification scheme is quite effective, and it is widely used in haploid identification. *R1-nj* causes deep pigmentation of the aleurone (endosperm tissue) in the crown region of the kernel and the scutellum (embryo tissue), which can be used to distinguish haploids from crossing seeds by visual scoring. However, this method has three limitations: (1) the pigmentation may vary in extent and intensity depending on the genetic background, and even some pigmentation may not appear in some genetic backgrounds such as European flint and tropical germplasm containing dominant genes such as *C1-I*, *C2-Idf* and *In1-D* (Röber et al. 2005; Belicuas et al. 2007); (2) the method requires highly trained workers with a good understanding of haploid detection through the color expression in endosperm and embryos and (3) the method is very labor intensive and has so far not been amenable to automation. These limitations can be overcome by the use of a novel haploid identification system based on OC. With the development of automated NMR and developed high oil inducers, assessment of oil concentration is feasible with single kernels at high throughput, making this method very promising for DH technology applications (Liu et al. 2012; Melchinger et al. 2013). Compared to *R1-nj* marker system, the OC marker system has three advantages: (1) the haploids produced via the OC marker system are more reliable because the critical points for OC can be regulated to reduce the FDR; (2) rather than visual scoring of the *R1-nj* marker, which relies on intensive labors, the OC marker system enables automated rapid screening of haploids using NMR. For example, an automated NMR can screen 21,600 kernels per day (in 24 h), while a well-trained worker can screen 10,000 kernels per day (in 8 h). Therefore, the efficiency of an automated NMR is nearly that of two well-trained workers; (3) for the OC marker system, the cost of an automated NMR equipment is $80,000, while for the *R1-nj* marker system, the cost of a trained workers is approximately $20 per day in China (and much higher in European and the USA). Thus, the cost of NMR screening is equivalent to that of two trained workers for 5 years. Although it NMR equipment is expensive, considering the expense and time required for long-term labor, it is more economical to employ OC to identify haploids in the long run. However, the proposed OC marker system is a new approach that requires improvements. A detailed analysis of the OC marker system is presented in next section.

Haploid identification by OC

The large difference in OC between haploid and crossing seeds makes it possible to successfully sort these two types of seeds when a normal maize resource is crossed with a high oil inducer. The accuracy of the OC marker system was mainly relies on the use of high oil inducers and is also affected by the genetic background to be induced (Miller and Brimhall 1951; Curtis et al. 1956). In the current study based on the developed high oil inducers, the difference in the mean OC of haploid and crossing seeds was significantly ($P < 0.001$) negatively correlated with the OC of the corresponding haploids, affecting the accuracy of haploid identification based on OC, which was strongly dependent on the genetic background to be induced. For different germplasm, the genetic background with the lower haploid OC often did not overlap between haploid and crossing seeds, making it easy to choose critical points for OC to completely distinguish haploids from crossing seeds (Table 4; Fig. 1). However, for germplasm with higher haploid OC that exhibited some overlap, the choice of critical points for OC was quite important, as the FDR and FNR must be balanced. Low critical points for OC can reduce the FDR, allowing us to obtain more reliable, true haploids, but they can also increase the FNR, leading to a loss of more true haploids; conversely high critical points for OC can reduce the FNR to allow us to retain more haploids, with less loss but they can also increase the FDR, leading to the selection of more false haploids. To solve the problem of partial overlap of haploids and crossing seeds using this method, it is advisable to choose a lower critical point for OC to reduce FDR (FDR <5 %) without sacrificing

too many true haploids (FNR <20 %). More ears of this genotype should then be pollinated to compensate for the loss of some true haploids, which would enable us to obtain enough haploids to produce adequate DH lines for use in maize hybrid breeding. Moreover, it is possible to further improve the OC of high oil inducers. If the OC of a high oil inducer is ≧15 %, it is possible to discrimination haploids from crossing seeds for germplasm of any genotype (the upper limit of OC is 7 %) except for high oil maize. To further develop a higher oil inducer, two elements should be considered, including the OC of the high oil inducers and the xenia effect from high oil inducers. Maize accessions such as the Illinois high oil populations and the Beijing high oil population have OCs of approximately 20 %, which can be used to further improve high oil inducers. Our results also demonstrate that high oil inducers with similar OCs can cause different OCs in crossing seeds based on the same testers, Therefore, different testers can be used to evaluate the xenia effect during the process of developing high oil inducers. Indeed, further studies of high oil inducers and improvement of the corresponding OC marker system should provide a new, effective identification system that paves the way for further development of DH technology.

Acknowledgments This work was supported by grants from the National High-Tech Program of China (863 Program, Nos.: 2011AA10A103, 2012AA10A305) and the Modern Maize Industry Technology System Foundation of China (Grant No.: CARS-02-09). The authors greatly appreciate the helpful comments from an anonymous reviewer.

References

Antoine-Michard S, Beckert M (1997) Spontaneous versus colchicine-induced chromosome doubling in maize anther culture. Plant Cell Tissue Organ Cult 48(3):203–207

Barret P, Brinkmann M, Beckert M (2008) A major locus expressed in the male gametophyte with incomplete penetrance is responsible for in situ gynogenesis in maize. Theor Appl Genet 117:581–594

Belicuas PR, Guimarães CT, Paiva LV et al (2007) Androgenetic haploids and SSR markers as tools for the development of tropical maize hybrids. Euphytica 156:95–102

Chalyk ST (1994) Properties of maternal haploid maize plants and potential application to maize breeding. Euphytica 79:13–18

Chen SJ, Song TM (2003) Identification haploid with high oil xenia effect in maize. Acta Agron Sin 29(4):587–590 (in Chinese)

Coe EH (1959) A line of maize with high haploid frequency. Am Nat 93:381–382

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc B Biol Sci 363(1491):557–572

Cook JP, McMullen MD, Holland JB et al (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158(2):824–834

Curtis JJ, Brunson AM, Hubbard JE, Earle FR (1956) Effect of the pollen parent on oil content of the corn kernel. Agron J 48(12):551–555

Deimling S, Röber FK, Geiger HH (1997) Methodology and genetics of in vivo haploid induction in maize. Vortr Pflanzenzüchtg 38:203–224 (in German)

Dong X, Xu X, Miao J et al (2013) Fine mapping of *qhir1* influencing in vivo haploid induction in maize. Theor Appl Genet 126:1713–1720

Dudley JW, Lambert RJ (2004) 100 generations of selection for oil and protein in corn. Plant Breed Rev 24(1):79–110

Frisch M, Bohn M, Melchinger AE (1999) Comparison of selection strategies for marker assisted backcrossing of a gene. Crop Sci 39:1295–1301

Geiger HH (2009) Doubled haploids. In: Bennetzen JL, Hake S (eds) Maize handbook, vol 2. Springer Press, New York, pp 641–657

Hallauer AR, Carena MJ, Miranda FJB (2010) Quantitative genetics in maize breeding. Springer Press, USA

Herzog E, Frisch M (2011) Selection strategies for marker assisted backcrossing with high-throughput marker systems. Theor Appl Genet 123:251–260

Hospital F (2005) Selection in backcross programmes. Philos Trans R Soc B 360:1503–1511

Kebede AZ, Dhillon BS, Schipprack W et al (2011) Effect of source germplasm and season on the in vivo haploid induction rate in tropical maize. Euphytica 180(2):219–226

Lashermes P, Beckert M (1988) Genetic control of maternal haploidy in maize (*Zea mays* L.) and selection of haploid inducing lines. Theor Appl Genet 76:405–410

Laurie CC, Chasalow SD, LeDeaux JR et al (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168(4): 2141–2155

Li L, Xu XW, Jin WW, Chen SJ (2009) Morphological and molecular evidences for DNA introgression in haploid induction via a high oil inducer CAUHOI in maize. Planta 230:367–376

Li H, Peng Z, Yang X, Wang W et al (2012) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet 45:43–50

Liu J, Guo TT, Yang PQ et al (2012) Development of automatic nuclear magnetic resonance screening system for haploid kernels in maize. Trans Chin Soc Agric Eng 28(2):233–236

Melchinger AE, Schipprack W, Würschum T, Chen S, Technow F (2013) Rapid and accurate identification of in vivo-induced haploid seeds based on oil content in maize. Sci Rep 3:2129

Miller PA, Brimhall B (1951) Factors influencing the oil and protein content of corn grain. Agron J 43(7):305–311

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4326

Nanda DK, Chase SS (1966) An embryo marker for detecting monoploids of maize (*Zea mays* L.). Crop Sci 6:213–215

Neuffer MG, Coe EH, Wessler SR (1997) Mutants of maize. CSHL Press, New York

Prigge V, Sánchez C, Dhillon BS, Schipprack W, Araus JL, Bänziger M, Melchinger AE (2011) Doubled haploids in tropical maize: I. Effects of inducers and source germplasm on in vivo haploid induction rates. Crop Sci 51:1498–1506

Prigge V, Xu XW, Li L, Babu R, Chen SJ, Atlin GN, Melchinger AE (2012) New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize. Genetics 190:781–793

Röber F (1999) Reproductive biology and genetic studies with RFLP markers in vivo haploid induction in maize. Dissertation, University of Hohenheim (in German)

Röber FK, Gordillo GA, Geiger HH (2005) In vivo haploid induction in maize: performance of new inducers and significance for doubled haploid lines in hybrid breeding. Maydica 50:275–283

Rotarenco VA, Kirtoca IH, Jacota AG (2007) Possibility to identify kernels with haploid embryo by oil content. Maize Genet Coop Newsl 81:11. http://www.agron.missouri.edu/mnl/81/32rotarenco.htm

Song TM, Chen SJ (2004) Long term selection for oil concentration in five maize populations. Maydica 49:9–14

Timonova EM, Leonova IN, Röder MS, Salina EA (2013) Marker-assisted development and characterization of a set of *Triticum aestivum* lines carrying different introgressions from the *T. timopheevii* genome. Mol Breeding 31: 123–136

Xu X, Li L, Dong X, Jin W, Melchinger AE, Chen S (2013) Gametophytic and zygotic selection leads to segregation distortion through in vivo induction of a maternal haploid in maize. J Exp Bot 64(4):1083–1096