Mol Breeding
DOI 10.1007/s11032-011-9644-0

Validation of *DGAT1-2* polymorphisms associated with oil content and development of functional markers for molecular breeding of high-oil maize

Yuchao Chai · Xiaomin Hao · Xiaohong Yang · William B. Allen · Jiming Li · Jianbing Yan · Bo Shen · Jiansheng Li

Received: 15 February 2011 / Accepted: 17 September 2011

Abstract The gene encoding acyl-CoA:diacylglycerol acyltransferase (*DGAT1-2*) is a key quantitative trait locus that controls oil content and oleic acid composition in maize kernels. Here we re-sequenced the *DGAT1-2* region responsible for oil variation in a maize landrace set and in 155 inbred lines (35 high-oil and 120 normal lines). The high-oil *DGAT1-2* allele was present in most Northern Flint and Southern Dent populations but was absent in five of eight Corn Belt Dent open-pollinated populations and in most of the earlier inbred lines. Loss of the high-oil *DGAT1-2* allele possibly resulted from genetic drift in the early twentieth century when a few Corn Belt Dent populations were selected for the development of high-grain-yield inbred lines. Association analysis detected significant effects of two PCR-based functional markers (HO06 and DGAT04; developed based on *DGAT1-2* polymorphisms) on kernel oil content and oleic acid composition using the 155 inbred lines. Zheng58 and Chang7-2, the parent inbred lines of elite hybrid Zhengdan958, were used to transfer the favorable allele from the high-oil line By804 using marker-assisted backcrossing with the two functional markers. In BC5F2:3 populations, oil content of the three genotypes (−/−, +/−, and +/+) was, respectively, 3.37, 4.20, and 4.61% (Zheng58 recipient line) and 4.14, 4.67, and 5.25% (Chang7-2 recipient line). Oil content of homozygous kernels containing the high-oil *DGAT1-2* allele increased by 27–37% compared with recurrent parents. Hence, these functional markers can be used to re-introduce the high-oil *DGAT1-2* allele into modern inbred lines for increased oil content through marker-assisted backcrossing.

Keywords High-oil maize · Functional marker · Oil content · Molecular breeding

Yuchao Chai and Xiaomin Hao contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11032-011-9644-0) contains supplementary material, which is available to authorized users.

Y. Chai · X. Hao · X. Yang · J. Yan · J. Li (✉)
National Maize Improvement Center of China, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China
e-mail: lijiansheng@cau.edu.cn

W. B. Allen · J. Li · B. Shen (✉)
Pioneer Hi-Bred International Inc., Johnston, IA 50131–1004, USA
e-mail: bo.shen@pioneer.com

Introduction

Maize oil is an important resource for biofuel production and human and livestock daily consumption. In 1896, artificial selection for maize oil content was initiated in Illinois with the open-pollinated variety Burr's White (Dudley and Lambert 1992,

Published online: 11 October 2011

2004). After long-term selection for almost 100 generations, the kernel oil content of Illinois High Oil (IHO) reached ~20%. Chinese high-oil maize selection began in the early 1980s. After 18 selection cycles, Beijing High Oil (BHO) oil content has increased to 15.55% (Song and Chen 2004).

Maize oil is a mixture of fatty acids (FAs) and their natural ester derivatives. Palmitic (16:0), stearic (18:0), oleic (18:1), linoleic (18:2), and linolenic (18:3) acids account for 99% of the total FAs. The percentage of linoleic acid is highest (61.9%), followed by 24.1% oleic acid (Lambert 2001). Maize oil content is a quantitative trait that is affected by a number of quantitative trait loci (QTL; Goldman et al. 1994; Alrefai et al. 1995; Mangolin et al. 2004; Song et al. 2004; Wassom et al. 2008). Recently, 42 individual QTL that are associated with FA composition and oil concentration were detected in 21 maize genomic regions in a recombinant inbred population derived from a cross between the normal line B73 and the high-oil line By804 (Yang et al. 2010a). A few major QTL with large additive effects have an important role in altering FA composition and increasing kernel oil content. A larger number of minor QTL and some epistatic QTL, all with additive effects, also contribute to these oil traits. The major QTL on chromosome 6 (*qHO6*) in bin 6.04 that affects maize grain oil content and oleic acid concentration encodes DGAT1-2, an acyl-CoA:diacylglycerol acyltransferase, that catalyzes the final step of triacylglycerol synthesis. A phenylalanine insertion in DGAT1-2 at position 469 (F469) causes increased oil content and oleic acid concentration. The high-oil *DGAT1-2* allele with F469 was present in maize wild relatives and was subsequently lost in the modern inbred lines (Zheng et al. 2008). This polymorphism provides an opportunity for the construction of appropriate markers for improving the oil trait in maize.

The objectives of the current study were (1) to determine when the high-oil *DGAT1-2* allele was lost during the development of modern inbred lines, (2) to screen for *DGAT1-2* polymorphisms associated with oil content by re-sequencing using a diverse association mapping panel, (3) to develop PCR markers based on the validated polymorphisms, and (4) to improve the oil content of elite inbred lines by backcrossing using the molecular markers developed.

Materials and methods

Experimental materials

An association mapping panel was composed of 155 inbred lines (35 high-oil lines, 120 normal lines), mostly selected from previously developed temperate germplasm (Yang et al. 2010b). This panel was planted in a randomized complete block design with two replications in the following four experiments: Changping Agronomy Farm, China Agricultural University, Beijing in 2005 and 2007; Shangzhuang Agronomy Farm, China Agricultural University, Beijing in 2006; and Winter Nursery, Sanya Agronomy Farm, China Agricultural University, Hainan in 2007. Each line was grown in a single 3-m row, with rows spaced 0.67 m apart and a density of 45,000 plants/ha. More than six plants in each row were self-pollinated. Pollinated ears were harvested at maturity and air-dried to reach ~13% (w/w) seed moisture content. The pedigree detail of this panel has been summarized previously (Yang et al. 2010b). Maize landraces were obtained from the Plant Introduction Station, United States Department of Agriculture, Ames, Iowa, USA.

Favorable allele introgression

High-oil inbred By804, which was selected from a BHO Cycle13 high-oil population (Song and Chen 2004), contains the favorable *DGAT1-2* allele. Zheng58 and Chang7-2 (both of which have normal oil content) are the parents of the well-known hybrid Zhengdan958, which has been the dominant hybrid planted in China in the last 5 years. By804 was chosen as the donor, and Zheng58 and Chang7-2 as recurrent parents. Two BC5F2:3 populations were developed by continuous backcrossing with Zheng58 and Chang7-2, respectively, and the developed functional markers were used to confirm the existence of the favorable allele. In the BC3F1 population, 100 simple sequence repeat (SSR) markers, evenly spread across the 10 maize chromosomes, were used to perform a background screen. Ninety percent of these markers were supported by expressed sequence tags. The individuals with the highest ratio of recurrent parental genetic background were backcrossed continuously to obtain the next generation. In 2010, BC5F2 individuals were planted at the Shangzhuang Agronomy Farm, China Agricultural University, Beijing. About 150 ears from

each population were harvested and air-dried for the kernel oil assay. The three parental lines were planted and harvested, and the phenotype was measured using BC5F2 individuals as reference samples.

Assay of FA composition and kernel oil concentration

Ten fully developed kernels randomly chosen from the bulked grain of each landrace and 50 kernels from each inbred line were collected and dried for 60 h at 45°C and then were crushed to powder by a Universal Disintegrator (Tianjin Taisite Instrument Co. Ltd., Beijing, China) to measure FA composition. The FA components were extracted by hexane in the form of FA methyl esters as described by Sukhija and Palmquist (1988) and then were measured with an HP6890 gas chromatograph (GC; Agilent Technologies, USA) as described by Yang et al. (2010a). The kernel oil content for each line was calculated by summing up all identified FA concentrations.

The ears from the two BC5F2:3 populations were divided into three groups according to their plant genotypes. Thirty fully developed kernels from the middle part of each ear were chosen to measure single kernel oil content with a Minispec PC 20 (20 MHz) nuclear magnetic resonance (NMR) machine (Bruker Corporation, Germany; Song 1989); these data were used with the kernel weight to calculate the kernel oil percentage. For each ear, the calculated sample average represented the average kernel oil content of the ear. NMR measurements of single-kernel oil content were taken for ~80 kernels from each of the five maize F2 populations. The phenotypic data for each F2 population were separated into three groups of corresponding genotypes to calculate the effect of the F469 site in different genetic backgrounds. Single-factor variance analysis was performed to reveal the phenotypic differences among the different population groups.

Primer design, PCR, and sequence analysis

DNA was extracted from bulked leaves from at least six individuals for each line by a modified CTAB procedure (Murry and Thompson 1980). Three pairs of oligonucleotide primers were designed to obtain the *DGAT1-2* sequence in the 5′ untranslated region (UTR), in the 3′UTR, and in several exons. Two functional markers were developed based on the validated functional polymorphisms (Fig. 1b; Electronic Supplementary Material Table 1). Fragments were amplified via PCR in a 20-μl reaction volume containing 30 ng genomic DNA, 3 μl of 10× PCR reaction buffer (containing Mg^{2+}), 2.4 μl of 2.5 mM dNTPs, 0.6 μl of 10 μM of each primer, and 3 U of *Taq* polymerase (TransGen Biotech Inc., Beijing, China). PCR was performed under the following conditions: 95°C for 5 min; 35 cycles of 95°C for 45 s, optimal annealing temperature for 45 s, and 72°C for 1 min; a final extension of 10 min at 72°C. The PCR fragments were purified for sequencing using an ABI 3730 sequencer.

Primer Premier Version 5.0 (Premier Biosoft International, Palo Alto, CA, USA) was used to design oligonucleotide primers. The protein sequence of maize DGAT1-2 (Zheng et al. 2008) was used to BLAST against the maize high-throughput genomic sequences database (www.ncbi.nlm.nih.gov) to obtain the B73 genomic sequence. The gene structure was predicted using online tools from the Softberry web site (www.softberry.com). ClustalW was used to carry out the sequence alignment on the EBI web site (http://www.ebi.ac.uk).

RNA extraction and real-time quantitative reverse transcription PCR (qRT-PCR)

Expression analysis was conducted using RNA from embryos at 20 days after pollination (DAP) from 20 inbred lines (Supplementary Table 2). All materials were frozen in liquid nitrogen and stored at −80°C until used. Total RNA was extracted using the RNA Extraction Kit RP3202 (Bioteke Corporation, Beijing, China) and digested with RNase-free DNase (Promega). Complementary DNA (cDNA) was synthesized using MMLV retrotranscriptase and an oligo(dT) primer (Promega). qRT-PCR was performed with Ex Taq premix (Takara Shuzo, Kyoto, Japan) with three replicates. All experiments were performed according to the manufacturer's instructions. The $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001) was used to calculate *DGAT1-2* expression levels in different inbred lines. Throughout the experiments, actin was used as the endogenous control, and 20 DAP embryos from the B73 line served as the reference tissue.

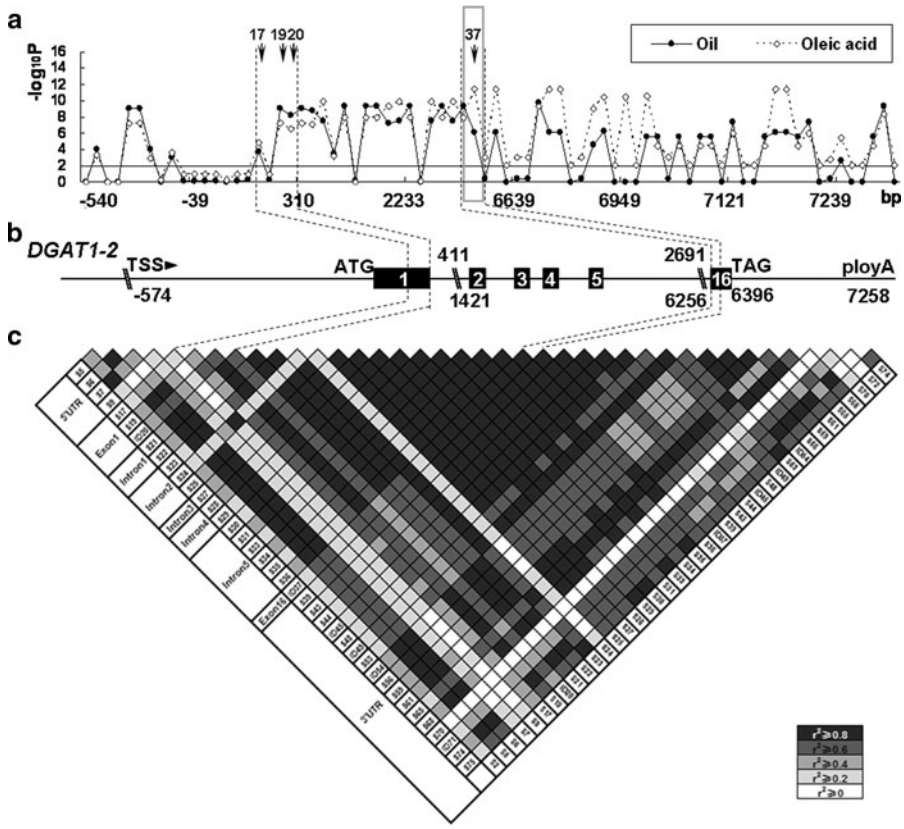

Fig. 1 Association analysis of polymorphisms in *DGAT1-2*. **a** The result of association analysis carried out using the mixed model incorporating population structure and kinship from 2006 data. The three other replicates showed similar results with respect to these polymorphisms (data not shown). The *abscissa* denotes the position of the polymorphic site. The positions of four polymorphisms that cause amino acid variations are marked and their positions outlined by *dotted lines* that join the three panels. The 3-bp InDel that leads to the F469 InDel in DGAT1-2 protein is denoted by a *gray open box*. **b** Three regions amplified by the sequence primers showing exons (*filled boxes*) enumerated in order. *Double slanted lines* represent gaps between fragments, with nearby *black numbers* representing the fragment start and termination positions. Start (ATG) and stop (TAG) codons are marked, with the adenine nucleotide of ATG marked as position 1. TSS, transcription start site. **c** Linkage disequilibrium results of the significant polymorphisms determined from the association analysis

Statistical analysis

The variance of FA composition and oil concentration was estimated using PROC GLM in the Statistical Analysis System v.9.13 (SAS Institute Inc.). The model for variance analysis was $Y = \mu + G_i + E_y + GE_{iy} + R_{yr} + \varepsilon_{iyr}$, where G_i is the effect of the ith line, E_y is the effect of the yth year, GE_{iy} is the line $\times$ year interaction, R_{yr} is the year $\times$ replicate interaction, and ε_{iyr} is the residual. All effects were considered to be random. Broad-sense heritability (H^2) was estimated as $H^2 = \sigma_i^2/(\sigma_i^2 + \sigma_{iy}^2/r + \sigma_e^2/yr)$ (Knapp et al. 1985), where σ_i^2 is the genetic variance, σ_{iy}^2 is the interaction of genotype with year, σ_e^2 is the residual error, r is the number of replications, and y is the number of years. All the variances were acquired from the variance analysis. PROC GLM was also used to perform single-factor variance analysis with the phenotypic data in different populations and analysis of variance (ANOVA) for the *DGAT1-2* polymorphic sites. The oil content data from the four experiments were integrated with PROC MIXED, and PROC CORR was used to quantify the relationships among relative expression levels and phenotypes using SAS v.9.13 software.

Association analysis was carried out using a mixed model incorporating kinship and population structure in TASSEL 2.1 (Bradbury et al. 2007). The kinship and population structure have been shown previously (Yang et al. 2010b). Linkage disequilibrium (LD)

analysis was performed using TASSEL 2.1 with the *DGAT1-2* sequence. The allele frequencies of each polymorphic site were counted from all lines in the association panel.

Results

Loss of high-oil *DGAT1-2* allele during breeding of modern inbred lines

To determine when the high-oil *DGAT1-2* allele was lost, the *DGAT1-2* region containing F469 was sequenced, and seed oleic acid concentration was measured in a set of landraces and early inbred lines (Supplementary Table 3). The high-oil *DGAT1-2* allele was present in the Northern Flint germplasm (except the Winnebago Mixed population) and in the Southern Dent and Southwestern populations (except Gourdeseed Dent). The low-oil *DGAT1-2* allele with deletion of F469 and high-oil allele segregated in half of the populations (Supplementary Table 3). Open-pollinated populations of Corn Belt Dent were developed in the early nineteenth century by crosses between Northern Flint and Southern Dent. The high-oil *DGAT1-2* allele was lost in five of eight Corn Belt Dent populations, such as Reid Yellow Dent and Lancaster Sure Crop. In 9 of 10 inbred lines developed in the early twentieth century, the high-oil *DGAT1-2* allele was lost. Only C103 still contained the high-oil *DGAT1-2* allele (Supplementary Table 3). The loss of the high-oil *DGAT1-2* allele was clearly associated with decreased seed oleic acid concentration. The landrace populations with the high-oil *DGAT1-2* allele showed much higher seed oleic acid concentrations than inbred lines with the low-oil *DGAT1-2* allele (Supplementary Table 3).

Association analysis of *DGAT1-2* with seed oil content and oleic acid concentration in inbred lines

The genomic DNA sequence of *DGAT1-2* comprised 7,832 bp, with 16 exons and 15 introns, as well as the 5′ and 3′ flanking sequences. To detect sequence diversity in the *DGAT1-2* gene, we re-sequenced the 5′UTR, 3′UTR and part of the coding regions of 155 inbred lines (Fig. 1b; Supplementary Table 1). Comparison of the sequences from 155 inbred lines revealed 76 polymorphisms in three regions with nine insertion/deletion polymorphisms (InDels) and 67 single nucleotide polymorphisms (SNPs). Among the 76 polymorphisms, 14 were located in the 5′UTR, 23 in the coding region, and 39 in the 3′UTR.

In total, 40 polymorphisms in one LD block were significantly associated with kernel oil content and oleic acid concentration (Fig. 1a, c): five in the 5′UTR region, 17 in the 3′UTR, and 18 in the coding region. Only four polymorphisms in the exon regions, all of which led to amino acid changes, were detected and found to be related to oil content and oleic acid concentration in the four experiments. These were identified as SNP17, SNP19, InDel20, and InDel37 (Table 1). InDel37 is the real codon coding the F469 amino acid. These four sites, the first three in the first exon of *DGAT1-2* and InDel37 in the last exon, were in the same LD block but were not in complete LD, with r^2 values between 0.53 (InDel20-InDel37) and 0.78 (SNP19-InDel20).

InDel37 in the last exon is responsible for the presence or absence of F469, a functional site reported by Zheng et al. (2008), and accounts for 46% (kernel oil content) and 40% (kernel oleic acid concentration) of the phenotypic variation among the 155 inbred lines measured (Table 1). If the 35 high-oil lines were excluded from the association panel, the phenotypic variation of these four polymorphisms was sharply reduced, and the oil content differences between the two genotypes of each site were not significant. Haplotype analysis indicated that the combination of the four polymorphisms could explain 55% (oil content) and 42% (kernel oleic acid concentration) of the phenotypic variation (Supplementary Table 4). When the 35 high-oil lines were excluded from the association panel, the difference in kernel oil content among the haplotypes was not significant ($\alpha = 0.05$), but the oleic acid difference was significant ($P = 0.0002$), and the 20.9% phenotypic variation could be explained.

One of the three polymorphisms, InDel20, was further analyzed. InDel20 had three different alleles in the association panel, whereas only two were present in Zhang's linkage mapping population (Zheng et al. 2008). The variation in *DGAT1-2* protein leads to glutamine deletion (-), one glutamine insertion (Q), or three glutamine insertions (QQQ), the last of which was a new allele found in our inbred lines. Single-factor variance analysis indicated that the difference between - and Q was great enough to be significant ($\alpha = 0.05$) when all the inbred lines were considered.

Table 1 *DGAT1-2* polymorphisms associated with oil traits in the association panel

Polymorphic site[a]	Frequency[b]	Experiment (average sample size)								
		Characteristic	2005BJ (75)[c]	2006BJ (129)[c]	2007BJ (130)[c]	2007HN (126)[c]	BLUP (142)[c]	R^2 (142)[d] (%)	P value[e]	r^{2f}
SNP17(G/T)	62/81	Oil	9.82×10^{-4}	1.91×10^{-4}	5.54×10^{-4}	3.44×10^{-4}	1.87×10^{-4}	36	4.78×10^{-15}	0.69
		Oleic acid	0.002	4.01×10^{-6}	9.21×10^{-5}	1.74×10^{-6}	1.18×10^{-5}	25	4.19×10^{-10}	
SNP19(T/C)	43/100	Oil	2.57×10^{-4}	1.27×10^{-7}	6.60×10^{-8}	2.15×10^{-9}	9.63×10^{-10}	53	1.68×10^{-24}	0.71
		Oleic acid	1.92×10^{-4}	2.78×10^{-7}	1.52×10^{-5}	2.12×10^{-7}	5.56×10^{-8}	33	1.27×10^{-13}	
InDel20(0/3/9)	43/91/9	Oil	0.0011	8.35×10^{-7}	4.78×10^{-7}	1.20×10^{-8}	6.66×10^{-9}	53	1.89×10^{-23}	0.53
		Oleic acid	8.43×10^{-4}	1.75×10^{-6}	7.00×10^{-5}	1.51×10^{-6}	2.51×10^{-7}	34	5.49×10^{-13}	
InDel37(3/0)	54/89	Oil	5.44×10^{-5}	4.04×10^{-6}	6.11×10^{-5}	3.86×10^{-6}	8.03×10^{-7}	46	1.36×10^{-20}	—
		Oleic acid	4.19×10^{-6}	1.61×10^{-11}	1.62×10^{-9}	4.32×10^{-10}	3.67×10^{-12}	40	2.87×10^{-17}	

BJ Beijing, *HN* Hainan, *BLUP* best linear unbiased prediction

[a] Only significant polymorphic sites that caused amino acid variation are shown. Polymorphisms shown in parentheses are presented as DNA sequences

[b] The allele frequencies counted in the association panel

[c] P values from the association analysis were determined using the mixed model incorporating population structure and kinship

[d] R^2 values from the ANOVA of BLUP data showing the percentage of phenotypic variation explained

[e] P values from ANOVA analysis of BLUP data

[f] r^2, a measure of LD with InDel37

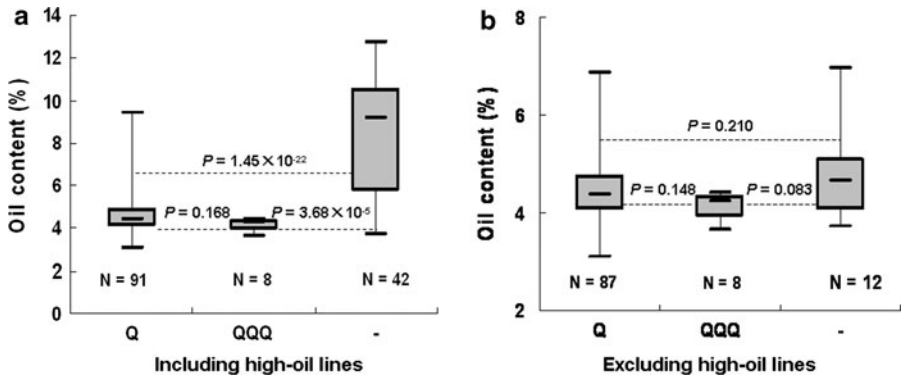

Fig. 2 Phenotypic comparison based on the polymorphism of InDel20. The symbols Q, QQQ, and – represent the three alleles at this site; *N*, the observation number; *P*, *P* value from a *t* test showing the difference between the two groups connected by *dotted lines*

The difference between Q and QQQ was not significant whether or not the high-oil lines were involved in the analysis (Fig. 2). Only eight inbred lines contained the QQQ genotype, and the two additional glutamine insertions may not have a real influence on oil content.

Development of functional PCR-based markers and their application for high-oil molecular breeding

A functional PCR-based marker, HO06, was developed based on InDel37 in *DGAT1-2* (Supplementary Table 1). The amplified fragment showed a 3-bp difference between high-oil lines and normal-oil lines. The 3-bp length difference was detected by polyacrylamide gel electrophoresis (PAGE) using the DNA silver staining method. The 155 maize inbred lines were divided into two groups determined by the length of the PCR fragments for the HO06 marker. One group contained 58 inbred lines, including all 35 high-oil inbred lines and the 23 normal lines that had the high-oil *DGAT1-2* allele with F469; the other group included the 97 inbred lines that had the low-oil *DGAT1-2* allele with the F469 deletion. To avoid the complexity of the PAGE and silver staining method, another three pairs of primers covering different sites with length polymorphisms were subsequently developed based on re-sequencing data. The DGAT04 pair covered an InDel in the sixth intron, with 1,002- and 728-bp amplified segments (Supplementary Table 1). This 274-bp InDel was in complete LD with InDel37 in our association panel. The results for the PCR products from the DGAT04 primer pair corresponded precisely with the results from the HO06 primer pair.

To determine whether the *DGAT1-2* DNA polymorphism affected *DGAT1-2* expression, real-time qRT-PCR for *DGAT1-2* was performed in inbred lines with different oil contents. *DGAT1-2* expression in high-oil inbred lines was approximately equal to that in the normal lines ($P = 0.7395$; Fig. 3a). Similar results were obtained when the groups were divided based on the HO06 marker (Fig. 3b). Thus, InDel37 did not affect *DGAT1-2* expression. In addition, there was no significant correlative relationship between the relative expression level and the phenotype for grain oil and oleic acid concentration ($r = 0.012$ and 0.013, respectively; Fig. 3c). These results suggested that the functional site in *DGAT1-2* was not affected by the gene expression level.

To determine whether InDel37 was correlated with oil content in different genetic backgrounds, the effect of InDel37 was measured on seed oil content in five F2 populations derived from crosses between four high-oil inbred lines containing the high-oil *DGAT1-2* allele with F469 (Qi205, Mo113, By4960, and Ry732) and five normal inbred lines containing the low-oil *DGAT1-2* allele with the deletion of F469 (Zong3, Hai014, P178, Mo17, and Xi502). In all five populations, the kernel oil content in the homozygote of the high-oil *DGAT1-2* allele (+/+) and the homozygote for the low-oil *DGAT1-2* allele (−/−) were significantly different ($\alpha = 0.05$). Lines with the +/+ genotype showed a 10–23% increase in kernel oil content (Table 2). Hence, the InDel37 site in *DGAT1-2* showed a consistent effect on kernel oil content in different genetic backgrounds, although the magnitude of the effect may vary depending on the presence of other polymorphisms.

The newly developed markers were then used to select for high-oil content in our breeding population. By804, a high-oil donor, was backcrossed to two

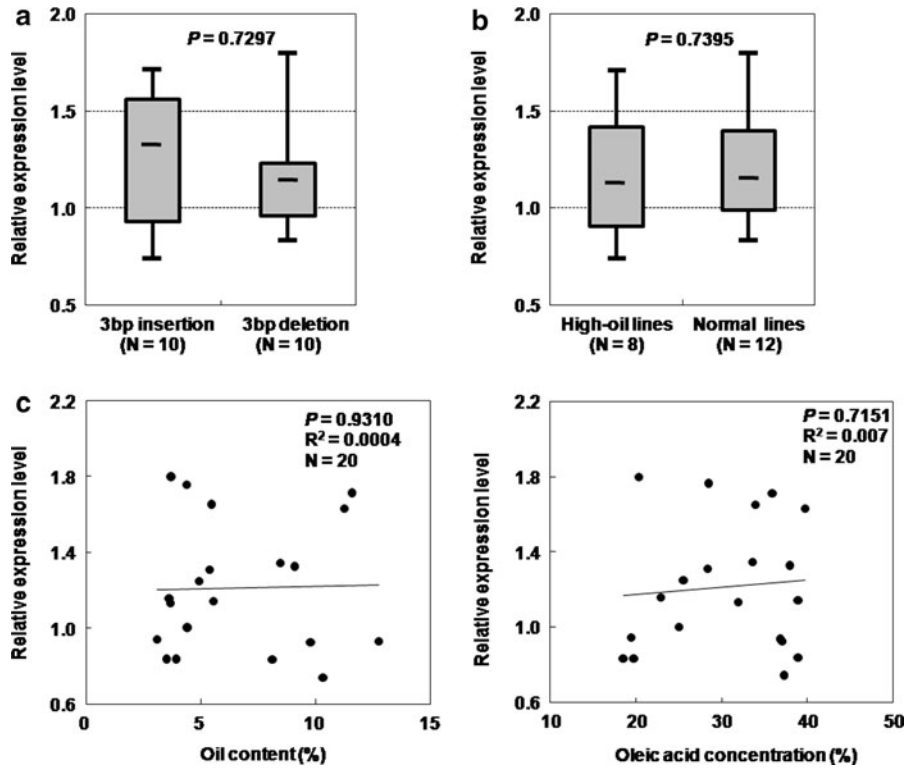

Fig. 3 Expression pattern of *DGAT1-2* and oil characteristics in different inbred lines. **a** Expression pattern of *DGAT1-2* in inbred high-oil and normal lines. **b** The expression level comparison between groups separated by marker HO06; *P*, *P* value from *t* test showing the difference between the two groups. **c** Correlation analysis of the relative expression level versus the phenotype (oil content and oleic acid percentage). *P*, correlation from PROC CORR showing the relationship between expression level and phenotypes; R^2, the Pearson correlation coefficient for expression level and phenotype; *N*, the number of lines analyzed

Table 2 The kernel oil content of three genotypes in five F2 populations

Parents[a]	*N*[b]	+/+ [c]	+/− [c]	−/− [c]	Change (%)[d]	*P* value[e]
Zong3 × Qi205	76	3.92 ± 0.41	3.68 ± 0.46	3.19 ± 0.46	0.73 (23.0%)	1.04×10^{-6}
Mo113 × Hai014	85	4.36 ± 0.41	4.51 ± 0.55	3.98 ± 0.50	0.38 (9.6%)	5.31×10^{-4}
Mo113 × P178	75	4.11 ± 0.57	4.01 ± 0.44	3.58 ± 0.42	0.54 (14.9%)	5.67×10^{-4}
By4960 × Mo17	93	7.24 ± 0.91	7.02 ± 1.02	6.61 ± 1.06	0.68 (10.2%)	4.48×10^{-2}
Xi502 × Ry732	91	5.43 ± 0.65	5.62 ± 0.85	4.95 ± 0.83	0.47 (9.6%)	6.63×10^{-3}

+/+, lines homozygous for the high-oil *DGAT1-2* allele (3-bp insertion); −/−, lines homozygous for the low-oil *DGAT1-2* allele (3-bp deletion); ± , the heterozygous lines

[a] The parents of each F2 population are shown. The *underlined* inbred lines have the high-oil *DGAT1-2* allele with F469

[b] Numbers of kernels sampled in each F2 population

[c] Average oil content (%) ± SD in the kernels

[d] The increase in oil content of +/+ versus −/− kernels

[e] *P* value from a single-factor variance analysis of the three groups

normal inbred lines, Zheng58 and Chang7-2. When each line was grown in 2010 in Beijing, the kernel oil contents were 10.96% (By804), 3.12% (Zheng58), and 4.06% (Chang7-2). In BC5F2:3 populations using Zheng58 as a recurrent parent, the homozygote for the By804 *DGAT1-2* allele (+/+) had an oil content of 4.61%, a 36.8% increase relative to the oil content of the homozygote for the *DGAT1-2* allele of the recurrent parent (−/−; Fig. 4a). Variance analysis and multiple comparisons indicated that the phenotype differences were significant ($P < 0.01$) among the three genotypes. Similarly, in the Chang7-2 BC5F2:3 population, individuals with the high-oil *DGAT1-2* allele (+/+) showed a 26.8% increase in seed oil content as compared with individuals with the low-oil allele (−/−; Fig. 4b).

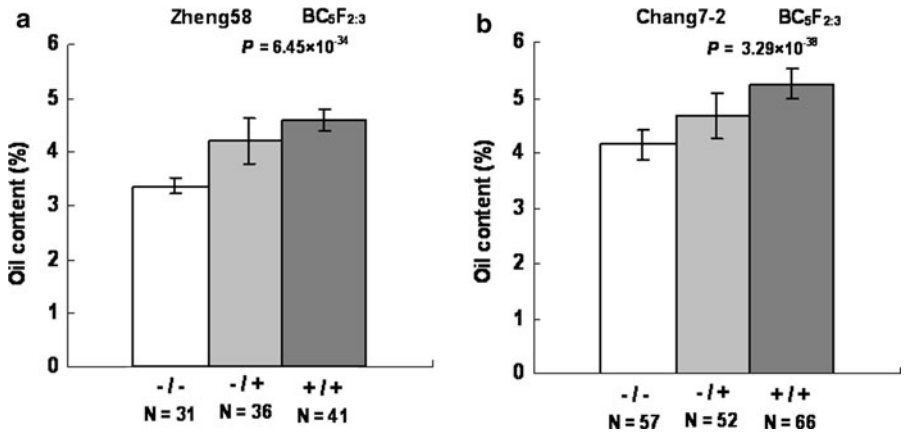

Fig. 4 Oil content in different genotypes of two BC5F2:3 populations (+/+, homozygote for the By804 *DGAT1-2* allele; −/−, lines homozygous for the *DGAT1-2* allele of the receptor line; +/−, the heterozygous form; *P*, *P* value from single-factor variance analysis of the three groups; *N*, the sample number)

Discussion

Loss of the high-oil *DGAT1-2* allele in modern inbred lines

The high-oil *DGAT1-2* allele is present in all teosinte lines but has been lost in modern inbred lines. To understand when the high-oil *DGAT1-2* allele was lost, we determined seed oleic acid composition and re-sequenced the *DGAT1-2* region that contains F469 in a set of maize landraces. The low-oil *DGAT1-2* allele with the 3-bp deletion in the last exon first appeared in a few Northern Flint populations and later in Southern Dent and Southwestern populations, which can be traced back a few thousand years (Troyer 2004). The low-oil allele was then fixed in five of eight Corn Belt Dent populations including the most important Reid Yellow Dent and Lancaster Sure Crop populations. Open-pollinated Corn Belt Dent populations were developed in the early nineteenth century by crosses of Northern Flint and Southern Dent, followed by mass selection. During the nineteenth century, there were ~800 open-pollinated cultivars of maize grown in North America, most of which are now extinct. In the early twentieth century, inbred lines were developed widely from open-pollinated populations for hybrid production. A potential genetic bottleneck of US maize occurred when only a few open-pollinated populations, including Reid Yellow Dent and Lancaster Sure Crop, were selected for inbred development. For example, Reid Yellow Dent alone accounts for 50% of the germplasm in modern hybrids, and Lancaster Sure Crop accounts for ~13% of the genetic background of US hybrid corn. All modern elite inbred lines were derived from a few early inbred lines. About six inbred lines and their close relatives are represented in 70% or more of US hybrids (Troyer 2004). Loss of the high-oil *DGAT1-2* allele could be caused by genetic drift that occurred in the early twentieth century**.** Alternatively, breeders may have selected against the high-oil *DGAT1-2* allele if that allele is associated with poor agronomic traits, such as grain quality and yield. Currently we have not detected any negative effects on grain quality and yield associated with the high-oil *DGAT1-2* allele in near isogenic lines.

Validation of the functional site in *DGAT1-2*

In total, we found 76 polymorphisms within *DGAT1-2*; of these, only four polymorphisms in the coding region were significantly associated with grain oil content and oleic acid concentration. Of the four sites, InDel37 was located in the last exon and was validated in different genetic backgrounds by both linkage mapping and association mapping analysis. Its effect on grain oil content was consistent across different genetic backgrounds. The other three polymorphisms were located in the first exon and were in LD with InDel37. For the population structure effect, the function of the first three polymorphisms could not be confirmed in our association panel. Their

significant association with grain oil content and oleic acid concentration may have been generated by the LD effects between them and InDel37. A molecular marker, HO06, was developed based on InDel37; however, labor-intensive electrophoresis and silver staining were needed to separate the favorable and unfavorable alleles when using this marker. Based on a large InDel in the sixth intron that showed complete LD with InDel37, another marker was developed, DGAT04, that was easily separated on a 1% (w/v) agarose gel.

Interestingly, about half of the detected polymorphisms within *DGAT1-2* were located in the 3′UTR. As mentioned in Ching et al. (2002), the frequencies of SNP and InDel polymorphisms are very high in the maize 3′UTR. The expression analysis did not, however, indicate that grain oil content and oleic acid concentration were affected by the level of *DGAT1-2* expression.

Marker-assisted selection (MAS) in high-oil maize breeding

Breeding selection for increased oil content in grain has been very successful, but commercialization of high-oil hybrids has been limited by reduction in grain yields and some poor agronomic traits associated with high-oil germplasm (Lambert 2001). The loss of grain yield in high-oil maize is mainly caused by decreased starch content, which is a major grain storage component and is negatively correlated with oil content. According to recent investigations using molecular markers (Goldman et al. 1994; Mangolin et al. 2004; Wassom et al. 2008; Yang et al. 2010a), oil content in maize grain is associated with ~10–50 QTL. Zhang et al. (2008) reported that many of the mapped QTL for major maize grain traits (such as chemical composition and oil, protein, and starch content) were located in identical positions on the genetic linkage map, with one or two other related QTL. Some oil QTL have a negative effect with starch QTL (e.g., *oilc1* and *stc1* on chromosome 1), whereas others are not correlative, such as the major oil content QTL (*oilc6* on chromosome 6), which was the focus of this study. Based on the analysis by Zheng et al. (2008), the major QTL, which encodes DGAT1-2, influences only the oil concentration in maize embryo. During phenotype selection to increase oil content, selection pressure favors all loci associated with oil content regardless of their correlation with starch content. In contrast, genotype selection based on molecular markers can focus only on specific loci, which may not correlate with starch content. The centralization of these QTL through marker-assisted selection should result in increased oil content in maize grain without significantly decreasing the starch content, so that grain yield would be unaffected.

Zhengdan958 is a superior-performing normal hybrid that has been widely extended in the Chinese Corn Belt with 3.2 million hectares in 2009 and has a grain oil content reaching ~4%. After the high-oil *DGAT1-2* allele was transferred from the high-oil donor line By804 to its two parental inbred lines (Zheng58 and Chang7-2) by marker-assisted backcrossing, the oil content of the homozygotes for the high-oil *DGAT1-2* allele was 27–37% higher than that in their recurrent parents in BC5F2:3 populations. These results demonstrate that re-introduction of the high-oil *DGAT1-2* allele into modern inbred lines by a developed functional marker could be effective for increasing kernel oil content in a high-oil breeding program. This provides a new strategy for increasing grain oil content significantly without significantly decreasing grain yield, by transferring a favorable QTL directly for high-oil molecular breeding. In this study, the effect of the functional marker for grain oil content and oleic acid concentration was stable across multiple genetic backgrounds (Table 2), suggesting that these markers could be widely used among different genetic backgrounds.

Acknowledgments This research was supported by the National Natural Science Foundation of China and Chinese High Technology Project, and a collaborative research project from Pioneer Hi-Bred International Inc. We also thank Li Lin and Li Qing for their guidance of statistic methods.

References

Alrefai R, Berke TG, Rocheford TR (1995) Quantitative trait locus analysis of fatty acid concentration in maize. Genome 38:894–901

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19–32

Dudley JW, Lambert RJ (1992) Ninety generations of selection for oil and protein in maize. Maydica 37:81–87

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

Goldman IL, Rocheford TR, Duldey JW (1994) Molecular markers associated with maize kernel oil concentration in an Illinois High Protein × Illinois Low Protein cross. Crop Sci 34:908–915

Knapp SJ, Stroup WW, Ross WM (1985) Exact confidence intervals for heritability on a progeny mean basis. Crop Sci 25:192–194

Lambert RJ (2001) High-oil corn hybrids. In: Hallau AR (ed) Special corn. CRC Press, Boca Raton, pp 131–153

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25:402–408

Mangolin CA, de Souza Jr CL, Garcia AAF, Garcia AF, Sibov ST, de Souza AP (2004) Mapping QTLs for kernel oil content in a tropical maize population. Euphytica 137:251–259

Murry MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Song TM (1989) Rapid determination of oil content of crop seed by plused NMR. Acta Agron Sin 15:160–166

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

Song XF, Song TM, Dai JR, Rocheford TR, Li JS (2004) QTL mapping of kernel oil concentration with high-oil maize by SSR markers. Maydica 49:41–48

Sukhija PS, Palmquist DL (1988) Rapid method for determination of total fatty acid content and composition of feedstuffs and feces. J Agric Food Chem 36:1202–1206

Troyer AF (2004) Persistent and popular germplasm in seventy centuries of corn evolution. In: Smith W (ed) Corn: origin, history, technology, and production. Wiley, Hoboken, pp 133–231

Wassom JJ, Wong JC, Martinez E, King JJ, DeBaene J, Hotchkiss JR, Mikkilineni V, Bohnh MO, Rocheford TR (2008) QTL associated with maize kernel oil, protein, and starch concentrations; kernel mass; and grain yield in Illinois High Oil 9 B73 backcross-derived lines. Crop Sci 48:243–252

Yang XH, Guo YQ, Yan JB, Zhang J, Song TM, Rocheford T, Li JS (2010a) Major and minor QTL and epistasis contribute to fatty acid composition and oil content in high-oil maize. Theor Appl Genet 120:665–678

Yang XH, Yan JB, Shah T, Warburton ML, Li Q, Li L, Gao YF, Chai YC, Fu ZY, Zhou Y, Xu ST, Bai GH, Meng YJ, Zheng YP, Li JS (2010b) Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theo Appl Genet 121:417–431

Zhang J, Lu XQ, Song XF, Yan JB, Song TM, Dai JR, Rocheford T, Li JS (2008) Mapping quantitative trait loci for oil, starch, and protein concentrations in grain with high-oil maize by SSR markers. Euphytica 162:335–344

Zheng PZ, Allen WB, Roesler K, Williams ME, Zhang SR, Li JM, Glassman K, Ranch J, Nubel D, Solawetz W, Bhattramakki D, Llaca V, Deschamps S, Zhong GY, Tarczynski MC, Shen B (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nat Genet 40:367–372