A transposable element insertion within ZmGE2 gene is associated with increase in embryo to endosperm ratio in maize

Pan Zhang · William B. Allen · Nobuhiro Nagasawa · Ada S. Ching · Elmer P. Heppard · Hui Li · Xiaomin Hao · Xiaowei Li · Xiaohong Yang · Jianbing Yan · Yasuo Nagato · Hajime Sakai · Bo Shen · Jiansheng Li

Received: 2 March 2012 / Accepted: 16 June 2012 © Springer-Verlag 2012

Abstract Most of the maize kernel oil is located in the embryo while the majority of starch is located in the endosperm. Maize kernel composition and value are affected significantly by the ratio of the embryo size to the endosperm size; however, the genetic regulation of embryo to endosperm ratio (EER) in maize is unknown. Here we identified ZmGE2 gene, which encodes a cytochrome p450 protein, as a gene associated with EER variation in maize. We first expressed rice Giant Embryo (GE) gene driven by oleosin promoter in maize and detected a 23.2 % reduction in EER in transgenic seeds, demonstrating the existence of evolutionarily conserved mechanisms for EER determination in rice and maize. We next identified maize GE2, a homolog of rice GE sharing 70 % identity in amino sequence, as a candidate based on the similar expression pattern and co-localization with a previously detected QTL for EER. Followed by linkage and association mapping, a 247-bp transposable element (TE) insertion in 3′-untranslated region of ZmGE2 gene was identified to be associated with increase in EER and kernel oil content. Expression level of the favorable ZmGE2 allele containing the 247-bp TE insertion was strongly reduced. In addition, the 247-bp TE insertion site was a selection target during the artificial long-term selection for the high EER trait in a high oil population. This is the first report that demonstrates an association of ZmGE2 with EER variation in maize and identifies ZmGE2 gene as a promising target for manipulation of EER and grain composition by either transgenic approach or molecular breeding in maize.

Introduction

A maize kernel generally consists of embryo, endosperm, and pericarp. The mature embryo is typically 10 % of the total kernel mass, but accumulates about 85 % of the kernel lipids (Val et al. 2009). Maize oil extracted from the embryo is generally considered as high quality oil due to the high proportion of polyunsaturated fatty acids (Lambert 2001). Escalating demand for maize oil has spurred interest in the enlargement of the embryo to produce more oil. Breeding efforts undertaken as early as in 1896 have resulted in the creation of high oil maize germplasm which have extraordinarily high embryo to endosperm ratio (EER) (Moose et al. 2004). In contrast,
maize endosperm which contains majority of starch is the staple carbohydrate sources for human beings and livestock as well as the important feedstock for ethanol production. This dichotomy has led to alternative breeding objectives to either increase or decrease EER in maize. In spite of the importance of EER, knowledge about the genes associated with the genetic regulation of the balance between the development of embryo and endosperm is limited.

The cytochrome P450 super family (officially abbreviated as CYP) is a large and diverse group of enzymes. In plants, P450 s are involved in a wide range of biosynthetic reactions, leading to various fatty acid conjugates, plant hormones, defensive compounds, or medically important drugs (Bakb et al. 2011). In Arabidopsis, the RETARDED GROWTH OF EMBRYO1 (RGE1) gene has previously been shown to control embryo growth. As a member of the bHLH transcription factor family, RGE1 positively regulates the expression level of at least six genes, including a gene encoding a cytochrome P450 protein (Kondou et al. 2008). In rice, the GIANT EMBRYO mutant was first reported by Satoh and Omura (1981). The corresponding gene GIANT EMBRYO (GE), which encodes a cytochrome P450 protein, was fine mapped (Koh et al. 1996) and subsequently map-based cloned (Cahoon et al. 2003). Park et al. (2009) reported a new allele of GE gene, ge', induced from somaclonal variation derived by another culture. Loss-of-function of the GE gene leads to a significant enlargement of embryonic tissue at the expense of endosperm tissue. These reports highlighted that cytochrome P450 s play an important role in regulating EER in plants. In maize, research on EER is still at the primary QTL mapping level. Yang et al. (2012) identified ten quantitative trait loci (QTLs) for EER in B73 and By804. The coding region of ZmGE2 gene in B73 and By804 is identical. A 247-bp Insertion/Deletion site was identified in the 3' UTR region, and an InDel marker named TFI (All the primers used in this study were listed in Supplementary Table S2) was developed using Primer Premier Version 5.0 (Premier Biosoft International, Palo Alto, CA, USA). RIL129 is a line identified from B73 × By804 RIL population and contains 44.5% of B73 genetic background. A line BC4CSSL-2009BJ-5-3 (BC₄ near isogenic line which shares >95% background of B73) was determined to contain a short chromosomal segment harboring ZmGE2 from By804 in B73 genetic background. This line was used to construct BC₃ NIL populations for linkage analysis. A total of 1,350 BC₃F₁ individuals derived from BC4CSSL-2009BJ-5-3 were genotyped by two SSR markers OC53 and UMC2228 covering target fragment with physical distance of 1,270 Kb. Four overlapping recombinants in the marker interval of OC53-UMC2228 were identified. In 2009 the 4 BC₃F₁ recombinants were self-pollinated to harvest BC₄F₂ ears in the Beijing summer breeding nursery of China Agricultural University (Beijing, BJ, E 116°46, N 39°29). BC₄F₂ single kernels from each ear were genotyped using the three markers OC53, TFI, and UMC2228. An improved alkaline-heating method was employed to extract DNA from a very small piece of the endosperm of single kernel (Gao et al. 2011). +/+ (By804 homozygous) and −/− (B73 homozygous) seeds were planted and selfed.

Identification of maize GE homologs and phylogenetic analysis

Both the nucleotide sequence and amino sequence of the OsGE gene were used in BLAST searches against the commonly used databases containing maize sequence information (www.ncbi.nlm.nih.gov, www.maizesequence.org). Two putative maize homologs of OsGE were identified on Chromosome 7 (GRMZM2G138080, designated as ZmGE1), and Chromosome 1 (GRMZM2G470442, designated as ZmGE2). ZmGE2 gene on Chromosome 1 is located in qEER1-1, a minor QTL for EER detected in B73 × By804 RIL population, which is flanked by marker UMC1598 and UMC1884 (Yang et al. 2012). The phylogenetic tree of OsGE genes and its maize homologs as well as some members in CYP78A subfamily were generated using MEGA, version 5.1 (Tamura et al. 2011) with the neighbor-joining method. Robustness of the constructed phylogenetic tree was tested with 1,000 bootstrap repetitions of the informative polymorphisms.

Construction of NIL populations

To develop intragenic molecular markers for construction of NIL populations, the genomic region covering ZmGE2 gene was sequenced in two lines, B73 and By804. The coding region of ZmGE2 gene in B73 and By804 is identical. A 247-bp Insertion/Deletion site was identified in the 3' UTR region, and an InDel marker named TFI (All the primers used in this study were listed in Supplementary Table S2) was developed using Primer Premier Version 5.0 (Premier Biosoft International, Palo Alto, CA, USA). RIL129 is a line identified from B73 × By804 RIL population and contains 44.5% of B73 genetic background. A line BC4CSSL-2009BJ-5-3 (BC₄ near isogenic line which shares >95% background of B73) was determined to contain a short chromosomal segment harboring ZmGE2 from By804 in B73 genetic background. This line was used to construct BC₃ NIL populations for linkage analysis. A total of 1,350 BC₃F₁ individuals derived from BC4CSSL-2009BJ-5-3 were genotyped by two SSR markers OC53 and UMC2228 covering target fragment with physical distance of 1,270 Kb. Four overlapping recombinants in the marker interval of OC53-UMC2228 were identified. In 2009 the 4 BC₃F₁ recombinants were self-pollinated to harvest BC₄F₂ ears in the Beijing summer breeding nursery of China Agricultural University (Beijing, BJ, E 116°46, N 39°29). BC₄F₂ single kernels from each ear were genotyped using the three markers OC53, TFI, and UMC2228. An improved alkaline-heating method was employed to extract DNA from a very small piece of the endosperm of single kernel (Gao et al. 2011). +/+ (By804 homozygous) and −/− (B73 homozygous) seeds were planted and selfed.

Materials and methods

Vector construction and maize transformation

Standard restriction fragment preparation and ligation techniques were used to position the OsGE coding sequence behind the embryo-preferred promoter from the 16-kDa oleosin gene of maize (GenBank no. BD235503, including the 81-bp 5'-untranslated region of Oleosin, U13701) and before an NOS terminator. Vector construction and maize transformation were conducted as described previously (Zheng et al. 2008).
in the Hainan winter nursery (Hainan, HN, E 108°56', N 18°09') in 2009. The genetic effect of the substituted chromosome segment was evaluated by comparing the EER of +/+ (By804 homozygous) and −/− (B73 homozygous) genotypes using a standard t test.

Phenotype quantification

For measuring EER of an ear, 20 uniform kernels from the middle of the ear were soaked in deionized water for 24 h at 45 °C and dissected into embryo and endosperm. Embryos and endosperms were dried to the same moisture level of kernels after separation and then weighed, and the ratio of embryo-to-endosperm weight was calculated. To determine kernel oil content, maize kernels were dried for 60 h at 45 °C and then weighed, and oil content of a single kernel was measured by pulsed nuclear magnetic resonance (NMR) on a Minispec PC 20 NMR (Bruker, US). Measurements were made in duplicate, and the average was taken. Twenty uniform kernels from the middle of the ear were measured and averaged to determine the oil content of each ear.

Sequencing and genotyping

Two overlapping primer pairs covering the full length of ZmGE2 gene were used to sequence the parent lines B73 and By804 as well as the CAM155 association mapping panel. A total of 133 lines in the CAM155 association panel were successfully sequenced by the two primer pairs. The alignment of all the sequences for polymorphism identification was done using the multiple sequence alignment program MUSCLE (Edgar 2004), and was refined manually using BioEdit (Hall 1999). The aligned sequence was exported to Phylip format (Felsenstein 1989) for further analysis, and nucleotide polymorphisms including SNPs and InDels with minimum allele frequency ≥0.05 were extracted from the aligned sequence in TASSEL 2.0.1 (Bradbury et al. 2007). The marker TFI which encompasses the 247-bp InDel was used to follow the change in allele frequency in the ASK high oil population. DNA extractions and PCR conditions were described previously (Ching et al. 2002). The amplicons were then sequenced and the sequences were aligned using the software Sequencher (Gene Codes Corporation, Ann Arbor, MI USA http://www.gene codes.com).

Statistical analysis

Association analysis was performed in TASSEL 2.0.1 (Bradbury et al. 2007) using the mixed linear model (MLM) controlling population structure (\( Q \)) and relative kinship (\( K \)) (Yu et al. 2006). For significance testing of two genotypes, two-tail t tests were performed in MS Excel 2007; the r value for the correlation was calculated also by MS Excel 2007. The genetic effects explaining the phenotypic variation were calculated using the analysis of variance (ANOVA).

TE identification and structure prediction

The 247-bp insertion sequence was submitted to CENSOR (http://www.girinst.org) (Kohany et al. 2006) for transposable element prediction. The secondary structure and energy parameter of the structure were predicted by RNA Structure 4.2 (Reuter and Mathews 2010) and imported to RnaViz2.0 (De Rijk et al. 2003) for annotation.

RNA extraction and transcript profiling

Fresh plant tissues were frozen in liquid nitrogen and then stored at −70 °C until use. Total RNA was prepared using the Bioteke RNA extraction kit (Bioteke, Beijing, China) and digested with RNase-free DNase (Promega) following manufacturers’ instruction. RNA was subjected to complementary DNA (cDNA) synthesis using AMV reverse transcriptase and an oligo (dt) primer (Promega). Real-time quantitative PCR for ZmGE2 expression profiling was conducted with Ex Taq premix (Takara Shuzo). The \( 2^{-\Delta\Delta CT} \) method (Livak and Schmittgen 2001) with three replicates was used to calculate average expression level of each sample. A maize housekeeping gene (actin) was used as an internal control. A 20 DAP embryo tissue sample of the inbred line B73 was used as the reference tissue in the qRT-PCR for ZmGE2 expression profiling.

ZmGE2 sequence: GenBank JQ408671.

Results

ZmGE2 is a candidate gene for EER in maize

Recent research shows that loss-of-function of the OsGE gene in rice leads to increase of EER, and over expression of OsGE gene leads to reduced embryo phenotype (Cahoon et al. 2003). To understand if OsGE affects EER in maize, we introduced constructs containing OsGE cDNA driven by an embryo-preferred 16-kDa oleosin promoter into maize. The constructs also contained a DS-RED2 gene driven by an aleurone-specific lipid-transfer protein 2 (LTP2) promoter to facilitate identification of transgenic and null kernels for phenotypic analysis. Analysis of 23 independent transgenic maize lines revealed average decreases in T1 seeds’ EER by 23.2 %, whereas no
A significant kernel weight change was detected (Fig. 1a). These results demonstrate that OsGE can regulate EER in maize and suggest the existence of evolutionarily conserved mechanisms for EER determination in rice and maize.

Based on high sequence identity with OsGE, two putative maize homologs of OsGE were identified on chromosome 7 and 1, designated as ZmGE1 and ZmGE2, respectively (see Table 1). ZmGE1 shares 76% identity and 84% similarity with OsGE, whereas ZmGE2 shares 70% identity and 77% similarity with OsGE (Supplementary Fig. S1). Examination of the expression patterns of the two genes shows that ZmGE1 is mainly expressed in anther, immature ear, and stem whereas ZmGE2 is strongly expressed in embryo (Supplementary Fig. S2). Also, ZmGE2 was mapped to the interval of a QTL for EER in B73 × By804 RIL population (Yang et al. 2012). Therefore, we hypothesized that the ZmGE2 gene might affect EER in maize.

**Linkage analysis of ZmGE2 alleles**

Sequence analysis of the ZmGE2 gene in the parent lines B73 and By804 demonstrates the presence of a 247-bp insertion/deletion marker site, and the TFI marker was developed using this site for further linkage analysis (see “Materials and methods”). Using this marker a line BC4CSSL-2009BJ-5-3 (a BC4 near isogenic line which shares >95% background of B73) from our B73 × RIL129 chromosomal segment substitution lines (CSSLs) library was determined to contain a short chromosomal segment harboring the ZmGE2 gene from By804 in B73 genetic background. When compared with the B73 genotype parental line, the EER of line BC4CSSL-2009BJ-5-3 was significantly increased (Fig. 2a). To further determine the genetic effect of ZmGE2 alleles, 4 BC3F2 NIL families overlapping in a 1,270-Kb marker interval harboring TFI site were constructed. According to the maize reference sequence (Schnable et al. 2009), this marker interval contains 17 genes among which only ZmGE2 has been predicted to associate with EER. EER of B73 homozygous and By804 homozygous were measured in each of the four BC3F2 NIL families. No significant difference was detected in the two NIL populations (NIL-1, NIL-4) with the B73 genotype fixed at the TFI site, whereas significant difference was detected in the two populations (NIL-2, NIL-3) which segregated at the TFI site, with 4.3–5.3% increase of EER by By804 genotype (Fig. 2a). These results suggest the association of ZmGE2 gene with EER variation in maize.

**Association analysis of ZmGE2**

Candidate gene association analysis was performed to identify the association between the nucleotide polymorphisms within change in kernel weight %. b The x axis is the change in EER %, and the y axis is the change in oil content %. The r value indicates the correlation between the change in EER % and the change in oil content %.

**Table 1** Putative OsGE homologs in maize genome

<table>
<thead>
<tr>
<th>Homolog</th>
<th>Genomic position</th>
<th>Expression pattern</th>
<th>QTL co-location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZmGE1</td>
<td>Chromosome 7: 163, 986, 813–163, 988, 952 forward strand</td>
<td>Expressed in anther, immature ear, and stem</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>ZmGE2</td>
<td>Chromosome 1: 73, 374, 836–73, 376, 998 reverse strand</td>
<td>Strongly expressed in embryo; expressed in root</td>
<td>qEEWR1-1</td>
<td>Yang et al. (2012)</td>
</tr>
</tbody>
</table>

* Genomic position according to the B73 genome, version 5b.60(AGPv2) (www.maizesequence.org)
* See supplementary Figure 1
ZmGE2 gene and EER in maize kernel. Based on sequence analysis of ZmGE2 gene in the CAM155 panel developed by Yang et al. (2010) (Fig. 2b), the 247-bp InDel (TFI site) at position 1,821 (the first base of start codon is defined as position 1, see Fig. 2b) in 3' UTR region was strongly associated with EER, with a \( P \) value of 1.63E-07 and 7.11E-06 in two environments, respectively (Fig. 2c). Except for the TFI site, no other site within ZmGE2 gene was shown to associate with EER at \( P \) \leq 0.05 significance level. The sequence analysis of ZmGE2 gene also revealed that the ZmGE2 allele with the TE insertion is rare among the Chinese elite maize inbred lines, with the allele frequency of 0.11 (14 of 133) in the CAM155 panel. Interestingly, all the commercial inbred lines in CAM155 panel are homozygous TFI-/- genotype without the 247-bp TE insertion and the 14 lines with homozygous TFI+/+ genotype containing the 247-bp TE insertion are high oil lines derived from Beijing High Oil (BHO) or Alexho \( \times \) Illinois High oil (AIHO) population (Song and Chen 2004). The 247-bp inserted fragment of TFI+ allele was identified as a Mutator Distance Relative (MuDR), which forms a tight hairpin structure with predicted \( \Delta G \) of \(-164.5 \text{ kcal/mol} \) (Fig. 2d). As has been previously demonstrated (Slotkin et al. 2005; Slotkin and Martienssen 2007), this kind of hairpin structure opens the possibility that during transcription the corresponding mRNA precursor might spontaneously form double-strand RNA. This could target the ZmGE2 transcript for degradation by RNAi, and lead to the decreased expression of ZmGE2 transcript containing TE

---

**Fig. 2** Functional site mining of ZmGE2. **a** Linkage analysis of ZmGE2 alleles. Marker names and positions on the B73 \( \times \) By804 genetic map are listed on the top. The four BC\( _3 \) NIL populations were derived from a BC\( _3 \)F\( _1 \) chromosomal segment substitution line named BC4CSSL-2009BJ-5-3. Each line represents an independent NIL family. Black color indicates By804 homozygous regions, white color indicates B73 homozygous regions. For each pair of NILs, only the line containing fragments from the By804 genome is shown. Precise locations of crossovers were not known and were arbitrarily designated at the midpoint with “//” between two adjacent markers. The average EER % of each pair of NILs, Change % [(//)/(-/-)] \( \times 100 \) % and \( P \) value from the \( t \) test are shown at the right. 17 genes are in the 1,270-Kb marker region according to the MaizeSequence Zea mays version 5b.60(AGPv2), including ZmGE2.

**b** Gene structure of ZmGE2. Filled black boxes represent exons, and gray dash-dot boxes mark the regions sequenced in this study. The triangle indicates the insertion site of 247-bp transposable element. c Association between polymorphisms within ZmGE2 and EER % in 2009, 2010 two environments. In total 36 nucleotide polymorphisms were identified within ZmGE2 gene. The \( x \) axis is the location of ZmGE2 polymorphisms (the first base of start codon is defined as position 1) and the \( y \) axis is the \(-\log_{10} P \) values from the mixed linear model (results from 2009 presented as black dots and results from 2010 presented as gray dots). d The secondary structure formed by the 247-bp TE insertion at TFI site. The 247-bp insertion fragment is identified as a mutator distance relative (MuDR), which has 77-bp terminal inverted repeats, forming a tight hairpin structure with predicted \( \Delta G \) of \(-164.5 \text{ kcal/mol} \).
insertion. Based on these results, TFI site was considered to be a potential functional site for EER variation in maize.

Expression profiles of ZmGE2

Expression analysis of ZmGE2 gene was performed to examine the normal expression pattern of the gene in B73 inbreds as well as to examine changes in expression between B73 and By804 alleles. In B73 ZmGE2 gene is strongly expressed in the embryo, whereas the expression level in endosperm, tassel, silk, and husk is very low (Fig. 3a). Expression profiles of ZmGE2 revealed highest expression level in the embryo of 15 days after pollination (DAP); thus this period of tissue was chosen to examine whether the TFI/−/− and TFI+/+/ genotype exhibit differential expression levels of the gene. The results showed that the average transcript abundance of ZmGE2 in RIL129 and BC3F2 NILs with TFI+/+/ genotype was as low as one-twentieth of that in B73 with the TFI/−/− genotype (Fig. 3b). Additional expression analysis was performed in 20 diverse inbred lines from the CAM155 association panel (Supplementary Table S1). The average transcript abundance of inbred lines with TFI+/+/ genotype was significantly lower than the inbred lines with TFI/−/− genotype at TFI site. Also, EER is negatively correlated with the expression level of ZmGE2 gene among the 20 lines (Fig. 3c). These results suggest that the TE insertion at TFI site could substantially down-regulate the transcript abundance of ZmGE2 gene during the development of the embryo.

TFI site affects kernel oil content

Oil content in the maize kernel is determined by EER and embryo oil density (Zheng et al. 2008). Yang et al. (2012) reported that EER is significantly correlated with oil content in maize kernel, with $r$ value of 0.81. In this study, significant positive correlation between EER and kernel oil content was also detected in the OsGE transgenic lines, with $r$ value of 0.89. In conjunction with the decrease of EER, oil content of the OsGE transgenic lines was decreased by an average of 18.1 % (Fig. 1b). The TFI site was determined to significantly associate ($P \leq 0.001$) with oil content in CAM155 association panel in two environments (Table 2). Moreover, kernel oil content analysis in the four BC3F2 NIL families revealed that the TFI+/+ genotype could increase oil content by 4.5−4.8 % when compared with TFI/−/− genotype (Table 3).

TFI site is a selection target for increasing EER during high oil maize development

The high EER phenotype has been observed among high oil maize which has been developed by long-term artificial selection. It is believed that during the development of high oil maize, favorable alleles of the genes associated with EER were selected and accumulated, resulting in high frequency of the favorable alleles. ASK high oil population derives from a base population which consists of 56 open-pollinated varieties (Lambert et al. 2004). In spite of the rarity of TFI+/+ allele demonstrated in the CAM155 association panel, 96 % of the high oil inbred lines derived from the ASK high oil population have the TFI+/+ genotype. A high frequency of TFI+ allele has also been observed in Beijing high oil population which derives from Zhongzong-2 synthetic (Song and Chen 2004), in which 90 % of high oil lines have the
Table 2 Association between TFI site and oil content in maize kernel

<table>
<thead>
<tr>
<th>Environments</th>
<th>Frequency(^a)</th>
<th>(P) value(^b)</th>
<th>(R^2) (%)(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007 Beijing</td>
<td>13/121</td>
<td>5.22E−04</td>
<td>44.8</td>
</tr>
<tr>
<td>2007 Hainan</td>
<td>13/118</td>
<td>1.00E−03</td>
<td>44.6</td>
</tr>
</tbody>
</table>

\(^{a}\) Allele frequency at TFI site: insertion (247 bp)/deletion (0 bp), favorable alleles (higher oil content) are in boldface type

\(^{b}\) \(P\) value from association analysis carried out using the mixed model incorporating population structure and kinship

\(^{c}\) \(R^2\) values from analysis of variance (ANOVA) showing percentage phenotypic variation explained

TFI+/+ genotype. We further sampled C0, C5, C20, and C28 recurrent selection cycles of ASK high oil population to investigate the frequency change of TFI+ allele during the selection process. In C0 cycle, TFI+ allele was not detected (ND) by randomly genotyping 100 samples, suggesting the extremely low frequency of TFI+ allele in the base population. However, under the selection pressure for high oil trait during the increased selection cycles, the frequency of TFI+ allele was increased in the population and had become a major allele (>50 %) after 20 cycles of selection (see Fig. 4). These results are consistent with the notion that the TFI site has been a selection target in the history of high oil maize development, and this site has been an important determinant in the development of the high EER trait in high oil maize.

Discussion

Embryo to endosperm ratio is an important agronomic trait which represents the balance between embryo and endosperm tissues in cereal kernel. In maize, most of the oil is contained in the embryo while most of the starch is located in the endosperm; therefore, EER should also be considered as a trait which describes the kernel composition. In spite of the importance of EER, there is presently limited knowledge about the genetic regulation of EER in maize. In rice, however, the OsGE gene, which encodes a cytochrome P450 protein, has been identified as a negative regulator of EER (Cahoon et al. 2003; Park et al. 2009). In this study, we performed transformation experiments to investigate the function of the OsGE gene in maize. As the OsGE gene expressed in maize significantly decreased EER, we inferred that the endogenous maize homolog of OsGE gene might play an important role in regulating EER. We found that ZmGE2 gene, which shares 70 % identity in amino sequence with OsGE gene, is strongly

Table 3 Effect evaluation of TFI site on oil content in four overlapping BC\(_2\)F\(_2\) NIL populations

<table>
<thead>
<tr>
<th>NILs Genotype(^a)</th>
<th>Kernels Oil content ± SD</th>
<th>(P) (^c)</th>
<th>Change (%)(^d)</th>
<th>Ears Oil content ± SD</th>
<th>(P) (^c)</th>
<th>Change (%)(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSR1 (72.70(^b)) TFI (73.38(^b)) SSR2 (73.97(^b))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (-/-) (-/-) (-/-)</td>
<td>3.78 ± 0.23</td>
<td>29</td>
<td>0.51</td>
<td>1.1</td>
<td>3.73 ± 0.19</td>
<td>20</td>
</tr>
<tr>
<td>(+/+) (-/-) (-/-)</td>
<td>3.82 ± 0.23</td>
<td>36</td>
<td></td>
<td></td>
<td>3.80 ± 0.17</td>
<td>24</td>
</tr>
<tr>
<td>2 (-/-) (-/-) (-/-)</td>
<td>3.77 ± 0.19</td>
<td>35</td>
<td>3.6E−3</td>
<td>4.5</td>
<td>3.68 ± 0.15</td>
<td>20</td>
</tr>
<tr>
<td>(+/+) (+/+) (-/-)</td>
<td>3.94 ± 0.25</td>
<td>31</td>
<td></td>
<td></td>
<td>3.89 ± 0.16</td>
<td>22</td>
</tr>
<tr>
<td>3 (-/-) (+/+) (+/+)</td>
<td>3.72 ± 0.20</td>
<td>34</td>
<td>3.3E−4</td>
<td>4.8</td>
<td>3.66 ± 0.19</td>
<td>22</td>
</tr>
<tr>
<td>4 (-/-) (-/-) (+/+)</td>
<td>3.90 ± 0.21</td>
<td>36</td>
<td></td>
<td></td>
<td>3.90 ± 0.19</td>
<td>31</td>
</tr>
<tr>
<td>(-/-) (-/-) (+/+)</td>
<td>3.79 ± 0.20</td>
<td>35</td>
<td>0.66</td>
<td>0.5</td>
<td>3.82 ± 0.16</td>
<td>23</td>
</tr>
<tr>
<td>(-/-) (-/-) (+/+)</td>
<td>3.81 ± 0.22</td>
<td>37</td>
<td></td>
<td></td>
<td>3.80 ± 0.18</td>
<td>23</td>
</tr>
</tbody>
</table>

\(^{a}\) \(-/-\) represents the B73 type homozygous, \(+/+\) (boldface type) represents the By804 type homozygous

\(^{b}\) Marker location in the Zea mays AGP V2 physical map (Mb) of chromosome 1

\(^{c}\) \(P\) value from \(t\) tests of the two homozygous genotypes in each NIL populations

\(^{d}\) Oil content change in percentage comparing By804 type homozygous to B73 type homozygous in each NIL populations
expressed in the embryo and co-located with a previously mapped QTL for EER in maize. We further investigated the \textit{ZmGE2} gene by linkage and association analysis and identified a 247-bp TE insertion site in the 3'UTR associated with EER variation in maize. The modest genetic effect of \textit{ZmGE2} alleles agrees with the original mapping results that demonstrated the \textit{qEEWR1-1} is a minor QTL which explains only 3.8 % of the total EER variance in the B73 × By804 RIL population (Yang et al. 2012). Next, when we analyzed the expression level of lines with and without the 247-bp TE insertion in NILs and association populations, we found that the TE insertion significantly down-regulates the expression of \textit{ZmGE2} gene. In addition, we investigated the selection-response of the 247-bp TE insertion site under selection pressure for high EER during the development of a high oil population. We observed continuously increased allele frequency of the TE insertion genotype. Based on these results we concluded that (1) \textit{ZmGE2} gene functions as a negative regulator for EER, similar to the \textit{OsGE} homolog; (2) the TFI site within \textit{ZmGE2} gene significantly associates with EER variation in maize.

During the course of this study we have observed no negative correlations between the TFI site and important agronomic traits (kernel weight, row number, and plant height, etc., data not shown). Therefore, the TFI site could be used directly as a molecular marker for increasing EER for maize. Based on these analyses, we proposed that \textit{ZmGE2} gene might be involved in a signal transduction pathway which could regulate the embryo growth. It is also possible that the substrates of \textit{ZmGE2} gene could be short-chain fatty acids, and it could be important to examine whether the presence of the TE insertion at TFI site affects fatty acid-derived signals in future studies.

A common system for the nomenclature of P450 genes from all organisms has been set up on the basis of protein sequence identity and phylogeny (Nelson 2006). P450 s in the same family usually share at least 40 % identity and at least 55 % identity within a subfamily. Under these criteria the protein encoded by \textit{ZmGE2} gene could be included into CYP78A subfamily. The members in the CYP78A subfamily present conserved function. For example, \textit{CYP78A5}, \textit{CYP78A7}, and \textit{CYP78A9} were reported to regulate organ size via generating mobile growth signals which stimulates cell proliferation (Anastasiou et al. 2007; Wang et al. 2008; Adamski et al. 2009; Ito and Meyerowitz 2000); \textit{CYP78A1}, \textit{CYP78A5}, \textit{CYP78A7}, and \textit{CYP78A10} were reported as short-chain fatty acid hydroxylases (Imaishi et al. 2000; Kai et al. 2009). The phylogenetic relationship (see Fig. 5) shows that the \textit{OsGE} + \textit{ZmGE1} + \textit{ZmGE2} clade can be united with \textit{CYP78A5} + \textit{CYP78A10} clade with 97 % bootstrap support. BLAST searches revealed that the protein encoded by \textit{ZmGE2} gene has 56 % identity with \textit{CYP78A5} and \textit{CYP78A10}, respectively. Both \textit{CYP78A5} and \textit{CYP78A10} are short-chain fatty acid hydroxylases (Kai et al. 2009), and \textit{CYP78A5} regulates seed size by a novel signaling pathway (Adamski et al. 2009). Based on these analyses, we proposed that \textit{ZmGE2} gene might be involved in a signal transduction pathway which could regulate the embryo growth. It is also possible that the substrates of \textit{ZmGE2} gene could be short-chain fatty acids, and it could be important to examine whether the presence of the TE insertion at TFI site affects fatty acid-derived signals in future studies.

Acknowledgments This research was supported by the National Natural Science Foundation of China (31101156), the National Hi-Tech Research and Development Program of China (2012AA10A307), and a collaborative research project with Pioneer Hi-Bred International Inc. We thank Dr Lin Li, Dr Qing Li and Dr Jiming Li for their helpful suggestions on experiment design. We thank Dr Kent Wood and Dr J. Antoni Rafalski for their valuable comments.

Conflict of interest The authors declare that they have no conflict of interest.

References


