2014

Mapping the low palmitate \textit{fap1} mutation and validation of its effects in soybean oil and agronomic traits in three soybean populations

Andrea J. Cardinal  
\textit{North Carolina State University}

Rebecca Whetten  
\textit{North Carolina State University}

Sanbao Wang  
\textit{North Carolina State University}

Jérôme Auclair  
\textit{La Coop fédérée, Montreal, QC, Canada}

D. L. Hyten  
\textit{USDA-ARS, Soybean Genomics and Improvement Laboratory, Beltsville, Maryland, david.hyten@unl.edu}

See next page for additional authors  
Follow this and additional works at: \textit{http://digitalcommons.unl.edu/agronomyfacpub}

Part of the \textit{Agricultural Science Commons, Agriculture Commons, Agronomy and Crop Sciences Commons, Botany Commons, Horticulture Commons, Other Plant Sciences Commons, and the Plant Biology Commons}

Cardinal, Andrea J.; Whetten, Rebecca; Wang, Sanbao; Auclair, Jérôme; Hyten, D. L.; Cregan, P. B.; Bachlava, Eleni; Gillman, Jason; Ramirez, Martha; Dewey, Ralph; Upchurch, Greg; Miranda, Lilian; and Burton, Joesph W., "Mapping the low palmitate \textit{fap1} mutation and validation of its effects in soybean oil and agronomic traits in three soybean populations" (2014). \textit{Agronomy & Horticulture -- Faculty Publications}. 816.  
\textit{http://digitalcommons.unl.edu/agronomyfacpub/816}

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Mapping the low palmitate fap1 mutation and validation of its effects in soybean oil and agronomic traits in three soybean populations

Andrea J. Cardinal · Rebecca Whetten · Sanbao Wang · Jérôme Auclair · David Hyten · Perry Cregan · Eleni Bachlava · Jason Gillman · Martha Ramirez · Ralph Dewey · Greg Upchurch · Lilian Miranda · Joseph W. Burton

Received: 2 May 2013 / Accepted: 22 September 2013 / Published online: 17 October 2013 © Springer-Verlag Berlin Heidelberg 2013

Abstract

Key message fap1 mutation is caused by a G174A change in GmKASIIIA that disrupts a donor splice site recognition and creates a GATCTG motif that enhanced its expression.

Abstract Soybean oil with reduced palmitic acid content is desirable to reduce the health risks associated with consumption of this fatty acid. The objectives of this study were: to identify the genomic location of the reduced palmitate fap1 mutation, determine its molecular basis, estimate the amount of phenotypic variation in fatty acid composition explained by this locus, determine if there are epistatic interactions between the fap1 and fapnc loci and, determine if the fap1 mutation has pleiotropic effects on seed yield, oil and protein content in three soybean populations. This study detected two major QTL for 16:0 content located in chromosome 5 (GmFATB1a, fapnc) and chromosome 9 near BARCSOYSSR_09_1707 that explained, with their interaction, 66–94 % of the variation in 16:0 content in the three populations. Sequencing results of a putative candidate gene, GmKASIII, revealed a single unique polymorphism in the germplasm line C1726, which was predicted to disrupt the donor splice site recognition between exon one and intron one and produce a truncated KASIIIA protein. This G to A change also created the GATCTG motif that enhanced gene expression of the mutated GmKASIII gene. Lines homozygous for the GmKASIII mutation (fap1) had a significant reduction in 16:0, 18:0, and oil content; and an increase in unsaturated fatty acids content. There were significant epistatic interactions between GmKASIII (fap1) and fapnc for 16:0 and oil contents, and seed yield in two populations. In conclusion, the fap1 phenotype is caused by a single unique SNP in the GmKASIII gene.

Abbreviations

16:0 Palmitic acid
18:0 Stearic acid

Communicated by I. Rajcan.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2204-8) contains supplementary material, which is available to authorized users.

A. J. Cardinal (✉) · R. Whetten · S. Wang · R. Dewey
Department of Crop Science, North Carolina State University, Raleigh, NC 27695-7620, USA
e-mail: andrea.cardinal@syngenta.com

J. Auclair
La Coop fédérée, Montreal, QC, Canada

D. Hyten · P. Cregan
Soybean Genomics and Improvement Laboratory, U.S. Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705, USA

D. Hyten
DuPont Pioneer, P.O. Box 1000, Johnston, IA 50131-0184, USA

E. Bachlava
Monsanto Vegetable Seeds, 37437 State Hwy 16, Woodland, CA 95695, USA

J. Gillman
Plant Genetics Research Unit, USDA-ARS, University of Missouri, Room 204, Curtis Hall, Columbia, MO 65211, USA

M. Ramirez · G. Upchurch · L. Miranda · J. W. Burton
Soybean and Nitrogen Fixation Research Unit, USDA-ARS, Raleigh, NC 27607, USA
Soybean oil quality can be improved through genetics. Several lines with reduced palmitic acid concentration were developed by ethyl-methyl sulfonate mutagenesis (C1726, homozygous for the \textit{fap1} allele and ELLP2, homozygous for the \textit{fap}\textsuperscript{*} or \textit{fapx} alleles); \textit{N}-nitroso-\textit{N}-methyl urea mutagenesis (A22, homozygous for the \textit{fap3} allele); and X-radiation (J3 homozygous for the \textit{sop1} allele) (Fehr et al. 1991; Stojšin et al. 1998; Wilcox and Cavins 1990; Rahman et al. 1996). In addition, one natural mutation (\textit{fapnc}) was discovered in line N79-2077-11 and its sister line N79-2077-12, both derived from the fifth cycle of a recurrent selection program for high oleic concentration (Burton et al. 1983; Burton et al. 1994b). Among the five alleles conferring low palmitic acid in these lines, three are at independent loci (\textit{fap1}, \textit{fap3}, and \textit{fap}\textsuperscript{*}) (Schnebly et al. 1994; Stojšin et al. 1998; Wilcox et al. 1994; Kinoshita et al. 1998; Primomo 2000; Primomo et al. 2002). \textit{fapnc} is allelic to \textit{fap3} (Primomo 2000; Primomo et al. 2002). \textit{sop1} is independent of the \textit{fap1} locus, but its allelic relationship to the other loci is unknown (Kinoshita et al. 1998). In addition to the major genes mentioned above, minor/modifier genes influence the low palmitic acid content in soybean oil (Horejsi et al. 1994; Li et al. 2002; Rebetzke et al. 1998b, Rebetzke et al. 2001; Stojšin et al. 1998).

Several steps in the TAG biosynthetic pathway could affect palmitate concentration. In the case of low palmitic acid mutants, changes in activity or substrate specificity in 3-ketoacyl-ACP synthase II (KAS II), 16:0-ACP thioesterase, lysophosphatidic acid acyltransferase, glycerol-3-phosphate acyltransferase, or diacylglycerol acyl transferase enzymes could be responsible for their phenotype. Northern analysis of C1726 (\textit{fap1fap1}), N79-2077-12 (\textit{fapncfapnc}) and the cultivar Dare revealed that \textit{fapncfapnc} genotypes have reduced accumulation of transcripts corresponding to a 16:0-ACP thioesterase (\textit{GmFATB}) gene and no difference in transcript accumulation for genes encoding 18:1-ACP thioesterase (\textit{FATA}) or 18:0-ACP desaturase activities (Wilson et al. 2001b). No differences in transcript accumulation for \textit{fap1-fap1} genotypes were observed for any of the genes assayed (Wilson et al. 2001b). Southern analysis using the \textit{GmFATB} cDNA as a probe revealed that multiple copies of the 16:0-ACP thioesterase gene are present in the soybean genome and that the \textit{fapncfapnc} genotype has a deletion in one of those copies (Wilson et al. 2001a). The mutation in N79-2077-12 (\textit{fapnc} allele) is caused by a deletion in one of the copies of the 16:0-ACP thioesterase gene (\textit{GmFATB1a}) and allele-specific markers were developed to be used in marker-assisted selection schemes (Cardinal et al. 2007). The mutation in A22 (\textit{fap3} allele) is caused by a non-conservative amino acid substitution in the \textit{GmFATB1a} gene (Vries et al. 2011).

Correlated changes in stearic acid and unsaturated fatty acids in the soybean oil have been observed in low palmitic acid lines that carry the major genes described above. Lines homozygous for the \textit{GmFATB1a} deletion (homozygous \textit{fapnc} mutants) had decreased stearic acid (18:0) and increased linoleic acid (18:2) in two populations (Cardinal et al. 2007). Similarly, progenies that inherited the major \textit{fapnc} allele from the cross of N87-2122-4 to two high yielding normal palmitate lines had a 6–15 % reduction in 18:0 (Rebetzke et al. 1998b, 2001), a 4–10 % increase in oleic acid (18:1) content in both crosses (Rebetzke et al. 1998a, 2001), and an increase in linolenic acid (18:3) content in one cross (Rebetzke et al. 1998a). No differences in 18:2 were observed between progenies that inherited the major \textit{fapnc} allele and those that inherited the normal allele (Rebetzke et al. 1998a). In addition, when lines homozygous for the \textit{fap1} and \textit{fap3} alleles were compared to normal lines derived from the same population, the low 16:0 lines have reduced 18:0, and increased 18:1, 18:2, and 18:3 contents (Ndzana et al. 1994). Palmitate content in soybean oil was negatively correlated with 18:2 and positively correlated with 18:1 in a population that was segregating for the \textit{fap1} and \textit{fan} loci (Nickell et al. 1991). In summary, homozygous lines carrying \textit{fapnc} or both \textit{fap1} and \textit{fap3} alleles, not only have reduced 16:0 content but also have reduced 18:0 content and the effect on unsaturated fatty acid contents varies among populations.

Also, correlated changes in agronomic traits have been observed in low 16:0 lines that carry the major genes described above. Several studies have shown that lines homozygous for major \textit{fap1}, \textit{fap3} or \textit{fapnc} alleles have
Reduced seed yield (Ndzana et al. 1994; Rebetzke et al. 1998a; Cardinal et al. 2008; Cherrak et al. 2003). Reduced 16:0 lines homozygous for the \textit{fap1} and \textit{fap3} alleles had reduced oil content, were taller, and had greater protein content in two out of three populations, when compared to normal lines derived from the same population (Ndzana et al. 1994). Progenies homozygous for the \textit{fapnc} allele had no differences in protein content in all populations studied and an increase in oil content in only one cross (Rebetzke et al. 1998a). Lines homozygous for the \textit{GmFATB1a} deletion (homozygous \textit{fapnc} mutants) had decreased plant height and increased protein content in two populations but had a significantly decreased oil content in only one population (Cardinal et al. 2008). Small but significant positive correlations between palmitate and seed oil were observed in a population derived from the cross of a low-palmitate, low-linolenate line homozygous for the \textit{fapnc}, \textit{fap1}, and \textit{fan} alleles to the cultivar Anand (Cherrak et al. 2003).

The objectives of this study were to: (1) identify the genomic location of the \textit{fap1} mutation, (2) determine its molecular basis, (3) estimate the amount of phenotypic variation in the 16:0, 18:0, 18:1, 18:2 and 18:3 contents that is explained by this locus, (4) determine if there are epistatic interactions between the \textit{fap1} and \textit{fapnc} loci and (5) determine if the \textit{fap1} mutation has pleiotropic effects on seed yield, oil and protein content; in three adapted populations grown in replicated trials.

Materials and methods

Population development

Three populations, FAE, FAHH00 and FADD00 were developed by single seed descent (Brim 1966). The FAE population consisted of 746 F\textsubscript{5}-derived lines from the cross of N98-4445 \times Satelite. N98-4445 is a F\textsubscript{3}-derived high-oilate (563.1 g kg\textsuperscript{-1}), low-linoleate (26.3 g kg\textsuperscript{-1}), homozygous for \textit{fan PI 123440} line that is a sister line of the released germplasm N98-4445A developed by the USDA-ARS in Raleigh, NC (Burton et al. 2006). Satelite is a low-palmitate (39.8 g kg\textsuperscript{-1}; homozygous for \textit{fapnc} and \textit{fap1}), low-linolenate cultivar (\textit{fan PI 123440}) (Cardinal et al. 2007; Bachlava et al. 2008a, b, 2009b; Cardinal et al. 2011). FAHH00 (F\textsubscript{4} derived lines) was derived from the cross of a high yielding line, ‘Corsica’ (Kenworthy 1996), and N97-3681-11, a low linolenic (\textit{fan PI 123440}) and low palmitic acid (homozygous for \textit{fapnc} and \textit{fap1}) line. FADD00 (F\textsubscript{4} derived lines) was derived from the cross of a high yielding line, Brim (Burton et al. 1994a), and N97-3708-13, a low linolenic (\textit{fan PI 123440}), low palmitic acid line (homozygous for \textit{fapnc} and \textit{fap1}).

N97-3681-11, N97-3708-13 and the cultivar Satelite are sister lines homozygous for the low-palmitate mutations \textit{fapnc} and \textit{fap1} and they are highly related to Brim (coefficient of parentage = 0.777).

Experimental design

The FAE population was planted in a sets-within-replications experimental design (Hallauer and Miranda 1988) with two replications in each location. The experimental design was described by Bachlava et al. (2008b). Seven hundred nineteen experimental lines were initially divided into five groups according to maturity. Next, experimental lines of each maturity group were randomly assigned to sets. The maturity groups were randomly assigned to columns in each set and the lines of each maturity group were randomly assigned to plots within each column. Lines of the same maturity were planted in each column of each set to facilitate mechanical harvest. Each maturity group was included at least in one column within each set. Maturity groups with most of the segregating lines were allocated to two or three columns within each set. The parental line Satelite was also randomly assigned to one plot within a single set in each replication according to its maturity group. The other parental line, N98-4445, was randomly assigned, based on its maturity group, to one plot within each of the 15 sets in each replication of the experiment. Once the experimental and parental lines were assigned to a set, they remained in the same set in all replications, locations, and years. The order of sets and the lines within each set were randomized in each replication, location and year. Overall, the sets-within-replications design for the FAE population consisted of 15 seven-row by seven-column sets. Each plot consisted of four 4.88 m rows with 0.97 m spacing between rows. Only the two middle rows, end-trimmed to 3.96 m were harvested.

The experimental design used for FADD00 and FAHH00 was described by Cardinal et al. (2007). Briefly, 98 F\textsubscript{5}-derived lines and the parents of the FADD00 population and 99 F\textsubscript{5}-derived lines and one parent (N97-3681-11) of the FAHH00 population were grown in Plymouth (NC) in 2002 and Caswell and Plymouth (NC) in 2003. Within each environment, the experimental design for each population was a 10 \times 10 lattice with two replications.

All yield plots for the three populations were sown at a seeding rate of 30 seed per meter.

Phenotypic evaluation

Flowering and maturity dates were recorded for FAE, FAHH00, and FADD00 at the R2 and R8 reproductive stages, respectively, as days after planting (Fehr and Caviness 1977). Plant height was also measured for FAE,
FAHH00, and FADD00 populations as the distance from the soil to the plant apex in cm. Lodging was measured on a 1–5 scale (scored 1 if plant 90 % erect to 5 if plants 90 % procumbent). Yield and moisture content were evaluated on seed harvested from the two middle rows of each plot after end-trimming and mechanical harvest. Yield was adjusted to 13 % moisture content.

Approximately 10 and 30 g of seed were sub-sampled from each plot for the evaluation of oil content and fatty acid composition in the FAE and FADD00/FAHH00 populations, respectively. Seed oil content was determined by near infrared reflectance spectroscopy at the USDA North Regional Research Center, Peoria, IL, in 2003. Fatty acid composition was evaluated by gas chromatography, as described by Burkey et al. (2007). Fatty acid content is reported in g kg\(^{-1}\) of total lipids.

Statistical analysis

Statistical analyses of the sets-within-replications designs for FAE were conducted using PROC MIXED in SAS 9.1 (SAS Institute 2004). Environments, replication (environments), sets, sets*environments, sets*replication (environments), lines (set), environments*lines (set), and the error terms were considered random effects. Best linear unbiased predictors (BLUPs) were obtained for all traits as the sum of the overall mean and the random effect for each line of the FAE population (Littell et al. 1996). Degrees of freedom for FAE population were derived with the containment method in order to reduce the computing power required. Analysis of the FAE population was conducted separately for each environment, as well as combined across all the environments for all measured traits. In order to investigate whether maturity date affected the associations between fatty acid traits and molecular markers, maturity date was included as a covariate in the analysis conducted for each environment as well as across environments and BLUPs from the models with the covariate were obtained. Overall mean, range of BLUPs, heritabilities, genetic and phenotypic correlations were previously reported (Bachlava et al. 2008b). The lower bound of the pooled genetic correlations among environments for each trait, that did not account for heterogeneity of genotypic variances among environments were calculated by the ratio of \(\sigma_g^2/(\sigma_g^2 + \sigma_{gxe}^2)\) (Cooper and DeLacy 1994; Bachlava et al. 2008b).

The statistical analysis of the FADD00 and FAHH00 populations was described by Cardinal et al. (2007, 2008). The fatty acid data from each population were analyzed as a lattice design across environments with PROC MIXED, SAS 9.1. (SAS Institute 2004). Locations, replications, incomplete blocks, and lines were considered random effects. A BLUP for each line was obtained by adding the overall mean effect to the random effect of each line (Littell et al. 1996).

Genotypic evaluation and linkage mapping

Genomic DNA was extracted from leaf tissue of approximately 10–20 plants for each of 746 lines of the FAE population, collected from plants grown from seed increases grown in Clayton, NC in 2004. DNA isolation was conducted in a 96-well plate format with the Genta PureGene DNA purification kit (Gentra Systems, Minneapolis, MN). Ninety-eight polymorphic simple sequence repeat (SSR) markers of the consensus linkage map (Cre- gan et al. 1999; Song et al. 2004) were genotyped across the 20 linkage groups of the soybean genome in 746 lines of the FAE population following the protocol described by Bachlava et al. (2008a). SSR markers very tightly linked to fatty acid candidate genes (GmFAD2-2C, GmFAD2-2D, GmFAD6, GmFAD2-1B, GmAAPT1b) were developed and genotyped as described by Bachlava et al. (2009a). The same 746 lines were assayed with the GmFATB1a allele-specific marker developed to detect homozygote mutant lines carrying the low-palmitate \(fa_p \) mutation Cardinal et al. (2007). In addition, the parents and 562 F\(_5\),7 derived lines were genotyped with the GoldenGate 1,536 SNP panel (1536 Universal Soy Linkage Panel 1.0) (Hyten et al. 2010). The simple sequence repeat marker (SSR) BARCSOYSSR_09_1707 (Sat_1707) (Song et al. 2010) located at approximately Gm09:46.3 Mb was genotyped in the same 562 lines to more precisely define the location of a low-palmitate QTL on Gm09. The PCR conditions for this SSR marker were 95 °C for 2 min, 13 cycles of 94 °C for 20 s, a touchdown step of 63 °C for 15 s reducing 1 °C per cycle, and 72 °C for 20 s, then 18 cycles of 94 °C for 20 s, 50 °C for 20 s, and 72 °C for 20 s, and final step of 72 °C for 5 min.

DNA extraction, genotyping protocols and methods for FADD00 and FAHH00 were described by Cardinal et al. (2007). Briefly, an allele-specific marker for the deletion of GmFATB1a observed in lines homozygous for the \(fa_p \) mutation was genotyped in these populations. Both populations were also genotyped with SSR marker BARCSOYSSR_09_1707, to validate its effect using these populations.

Linkage analysis of the FAE population was conducted with MAPMAKER/EXP 3.0 (Lander et al. 1987) according to Cardinal et al. (2001), using a minimum likelihood of odds (LOD) score of 3.0 and a maximum Haldane distance.
of 50 cM. Heterozygous genotypes were coded as missing data in order for the F5-derived lines to be analyzed as recombinant inbred lines. Loci that could not be uniquely placed on a linkage map or that had no recombination with another locus in a linkage group were not included in the final maps.

Mapping the *fap1* locus and estimating its effect in three populations

QTL analysis was conducted using BLUPs derived from analyses across environments for all traits and 712 experimental lines of the FAE population that had both phenotypic and genotypic data. The marker genotypes were tested for association with palmitate content using single factor analysis, a simple linear regression approach (data not shown). Based on those results, new markers were developed in the vicinity of one major 16:0 QTL on chromosome 9 (LG K) and candidate genes in the region were sequenced.

To determine if the two major QTLs (*fap*<sub>nc</sub> and *fap*<sub>1</sub>) in the FAE population associated with palmitic acid content were also associated with other fatty acids and agronomic traits, an ANOVA was performed with the *GmFATB1a* and *GmKASIIIA* markers, and their interaction as independent variables was analyzed. BLUPs for each fatty acid and agronomic traits for each line were dependent variables in the analysis. The genotypic class means, the difference between means, significance of those differences, and the R-square for the model were obtained using SAS PROC GLM, SAS 9.2. (SAS Institute Inc. 2010).

In the FAHH00 and FADD00 populations an ANOVA was performed with *GmFATB1a* and the BARCOSYSSR_09_1707 marker classes and their interaction as independent variables was analyzed. BLUPs for each fatty acid and agronomic traits for each line were dependent variables in the analysis. The genotypic class means, the difference between means, significance of those differences, and the R-square for the model were obtained using SAS PROC GLM, SAS 9.2. (SAS Institute Inc. 2010).

Sequencing of candidate genes in the vicinity of major QTL on chromosome 9

Genomic DNA from cultivar Century (Wilcox et al. 1980) (WT) and its EMS induced mutant C1726 (*fap1*) was isolated from ~20 mg of mature seed or ~100 mg of leaf tissue, using the DNeasy Plant Mini kit (Qiagen, Inc., Valencia, CA). Gene sequences for target genes near the major QTL on chromosome 9 and potential homologous genes were downloaded from [http://www.phytozome.net/soybean](http://www.phytozome.net/soybean) and aligned using Vector-NTI (Invitrogen). PCR primers to amplify candidate genes Glyma09g39610 and Glyma09g41380 were designed based on gene sequences downloaded from the Williams 82 whole genome shotgun sequence ([http://www.phytozome.net/soybean](http://www.phytozome.net/soybean)) to be gene specific after alignment with homologous genes using AlignX (Invitrogen). A minimum of two differences per primer were required between the target gene and any gene homologs.

PCR amplification with genomic DNA was performed with 5–50 ng of genomic DNA with Ex taq according to manufacturer’s recommendations (Takara, Otsu, Shiga, Japan) in a PTC-200 thermocycler (MJ Research/Bio-Rad, Hercules, CA), under the following conditions: 95 °C for an initial 5 min denaturation, followed by 40 cycles of (95 °C for 30 s, then 60 °C for 30 s, and an extension step at 72 °C for 1 min per kilobase pair of expected product (7 min for Glyma09g39610 and 5 min for Glyma09g41380). PCR products were separated on 1 % agarose gels to ensure appropriate size and were purified using a QIAquick PCR purification kit (Qiagen). Following purification, products were Sanger sequenced at the DNA Core Facility at the University of Missouri-Columbia. Details on primers used for amplification and sequencing are presented in Table 1.

Genotyping the *GmKASIIIA* KASPAR allele-specific marker

A KASPar marker that distinguishes the *GmKASIIIA* allele inherited from Satelite or N98-4445 (Glyma09g41380) was genotyped in the same 562 lines described above. A pre-amplification step was performed to amplify only *GmKASIIIA* (Glyma09g41380) with the following primers (KASIIIA_pcr F 5′-CCGAAATGTTGCCCTT-TT-3′, KASIIIA_pcr R 5′-CCACTGTACCTATCTGTCTTAGCACACTAGT-3′) and conditions: 8 µl reaction containing 5–10 ng genomic DNA, 1× NEB standard taq reaction buffer, 234 µM dNTPs, 0.75 units NEB taq polymerase, 3.5 pmol KASIIIA_pcr F, 3.5 pmol KASIIIA_pcr R. The PCR was performed using a touchdown program with an initial 2 min at 95 °C, followed by 13 cycles of 94 °C for 20 s, 63 °C for 15 s (−1 °C/cycle), 72 °C for 20 s; followed by 18 cycles of 94 °C for 20 s, 50 °C for 20 s, 72 °C for 20 s, with a final extension at 72 °C for 2 min. The KASPar protocol to genotype the KASIIIA SNP used the following primers: (KASI/A_pcr F 5′-gaaggctgcagtcagcaccggttcctcccaacctcaaatccaca3′, KASIIIA_pcr C: 5′-CTTGGATGCGAACAACAACTGA GTA-3′).

One ul of the KASIIIA pre-amplified product was dried on a PCR plate. The KASPar PCR had a total volume of...
4 μl, which included 2 μl of 2× KASPar reagent (KBio-
science, KBS-1004-002), 0.66 pmol of allele-specific pri-
mer KASIIIa_C1726snp_G, 0.66 pmol of allele-specific
primer KASIIIa_C1726snp_A, 1.65 pmol of common
primer KASIIIa_C1726snp_C3, a final concentration of
0.2 mM Tris pH 8.3, and a final concentration of 2.7 mM
MgCl2.

The KASPar PCR was performed using a touchdown
program with an initial 15 min at 94 °C, followed by 10
cycles of 94 °C for 20 s, 65 °C for 60 s (~0.8 °C/cycle),
followed by 26 cycles of 94 °C for 20 s, 57 °C for 60 s.
Results were read on a Roche LightCycler480.

qPCR expression analysis of fatty acid related genes

<table>
<thead>
<tr>
<th>Gene locus, primers and expected amplicon size of candidate genes located in the vicinity of the major QTL on Chromosome 9</th>
<th>Reverse primer (5′ → 3′)</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glyma09g39610 (Thioesterase)</td>
<td>CTCGGCCTACGCTTTGACTGCGACAACAA</td>
<td>7,139</td>
</tr>
</tbody>
</table>
| Glyma09g41380 (ketoacyl-
synthase III) | GCGACCAACAAAAGGACCAAT | 4,110 |

The KASPar PCR was performed using a touchdown
program with an initial 15 min at 94 °C, followed by 10
cycles of 94 °C for 20 s, 65 °C for 60 s (~0.8 °C/cycle),
followed by 26 cycles of 94 °C for 20 s, 57 °C for 60 s.
Results were read on a Roche LightCycler480.

qPCR expression analysis of fatty acid related genes

Glyma09g39610

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer (5′ → 3′)</th>
<th>Reverse primer (5′ → 3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glyma09g39610 (Thioesterase)</td>
<td>CTCGGCCTACGCTTTGACTGCGACAACAA</td>
<td>GCGACCAACAAAAGGACCAAT</td>
</tr>
</tbody>
</table>
| Glyma09g41380 (ketoacyl-
synthase III) | GCGACCAACAAAAGGACCAAT | 4,110 |

These lines were homozygous or heterozygous for the
GmFATB1a allele (Fapnc,Fapnc or Fapnc,Fapnc genotypes)
but two of those lines were homozygous for the GmKA-
SIII allele inherited from N98-4445, one line was het-
erygous and two lines were homozygous for the
GmKASIII allele inherited from Satelite. Seed obtained
from each sample were pooled and ground under liquid
nitrogen and total RNA was isolated using the RNeasy
Plant mini kit and protocol from Qiagen (Valencia, CA). Traces of genomic DNA were eliminated using on-column
DNase digestion (Qiagen). First strand cDNA synthesis
was carried out using the High Capacity cDNA Reverse
Transcription kit (Applied Biosystems) with MultiScribe reverse transcriptase (RT) and random primers. 2.5 μg of
RNA in 12.5 μl of nuclease free water were mixed with the
same volume of a 2× RT master mix. The reaction con-
ditions were: (1) 10 min at 25 °C, (2) 120 min at 37 °C, and
(3) 5 s at 85 °C. Quantitative RT-PCR assays with SYBR green were performed with the i-Cycler

Springer
thermocycler (BioRad, Hercules, CA) for GmFATB1a, GmFATB1b, GmFATB2a, GmFATB2b and GmKASIIIa studied in expression analysis in soybean lines, using qRT-PCR

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward/reverse primers (5' → 3')</th>
<th>Gene locus</th>
<th>Ta (°C)</th>
<th>Amplicon (bp)</th>
<th>Efficiency (%)</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmFATB1a</td>
<td>CAAGTGGACACTTGGTTTTC GCTCGGTGTTAGCATTTC</td>
<td>Glyma05g08060</td>
<td>61</td>
<td>136</td>
<td>93.3</td>
<td>Acyl-ACP thioesterase</td>
<td>This study</td>
</tr>
<tr>
<td>GmFATB1b</td>
<td>AGTAGACACCTGGGCTTG GCTCGGTGTTAGCATTTC</td>
<td>Glyma17g12940</td>
<td>62</td>
<td>134</td>
<td>100</td>
<td>Acyl-ACP thioesterase</td>
<td>This study</td>
</tr>
<tr>
<td>GmFATB2a</td>
<td>TCTCTGGGTGCTAGTAGGA AAAGGGGGGCTGCATTTC</td>
<td>Glyma04g21910</td>
<td>60</td>
<td>121</td>
<td>101</td>
<td>Acyl-ACP thioesterase</td>
<td>This study</td>
</tr>
<tr>
<td>GmFATB2b</td>
<td>AGATATTTGCTAGGTTTAGG GACTCTCACCACTGAGG</td>
<td>Glyma06g23560</td>
<td>57</td>
<td>197</td>
<td>109</td>
<td>Acyl-ACP thioesterase</td>
<td>This study</td>
</tr>
<tr>
<td>GmKASIIIa</td>
<td>ATTTCCAATGAGCAGCCTTTGC GCCTCTATGCTAAACTTC</td>
<td>Glyma09g41380</td>
<td>57</td>
<td>126</td>
<td>98</td>
<td>3-ketoacyl-ACP synthase</td>
<td>This study</td>
</tr>
<tr>
<td>Soy57 actin</td>
<td>5' GAGCTATGAAATTGCTGAGG</td>
<td>Glyma15g005570</td>
<td>58</td>
<td>118</td>
<td>100</td>
<td>Cytoskeletal structures (Reference gene)</td>
<td>Moniz de Sa and Drouin 1996</td>
</tr>
<tr>
<td>GmEF1a</td>
<td>5' CCGGCTGAATCCCAGGAGC  CAGTCAAGGTTTGGACCT</td>
<td>Glyma17g23900</td>
<td>61</td>
<td>176</td>
<td>98</td>
<td>Translation elongation factor (Reference gene)</td>
<td>Aguilar et al. 1991</td>
</tr>
</tbody>
</table>

a Annealing temperature  
b PCR amplification efficiency

Table 2 Primers, gene locus, annealing temperature, amplicon and PCR amplification efficiency of fatty acid related genes GmKASIIIa, GmFATB1a, GmFATB1b, GmFATB2a and GmFATB2b studied in expression analysis in soybean lines, using qRT-PCR

The amounts of transcripts were estimated from standard curves obtained using PCR products amplified with the respective primer pairs and applying a normalization factor that was obtained based on the analysis of soybean actin 57 (Moniz de Sa and Drouin 1996) and elongation factor EF1α (Aguilar et al. 1991) and using geNorm software (Vandesompele et al. 2002). To ensure that the correct gene transcript sequences were amplified, amplicons were sequenced at the Iowa State University Biotechnology Center, Ames, IA.

Sequencing of fragment amplified from wildtype and GmKASIIIa mutant cDNA

Primers were selected to specifically amplify a fragment from cDNA of KASIIIa on Gm09 (Table 3). Forward primer 5'-AAATTCCCTTGTTGGTTACAT-3' is located in the 5' UTR of Glyma09g41380, and reverse primer 5'-CAGTCAAGGTTTGGACCT-3' is located in exon 4 of Glyma09g41380. PCR conditions included an initial extension step of 2 min at 94 °C, followed by 30 cycles of (94 °C for 14 s, 57 °C for 30 s, 72 °C for 1 min), and 5 min at 72 °C. The fragments were excised from a 1 % agarose gel and purified using a Qiaquick Gel Extraction Kit (Qiagen #28704). The purified fragments were cloned into a pCR2.1 vector using the TA Cloning Kit (Invitrogen #28704), and DNA from these clones was submitted to GeneWiz for sequencing. Sequence analysis was performed using VectorNTI software (Invitrogen).

Northern analysis of GmKASIIIa

Pods and unifoliate leaves were harvested from N95-4445A and Satelite to isolate RNA and perform Northern analysis. Pods were harvested 35 days and 50 days after flowering from plots grown at Clayton in 2012 and leaves were harvested from seedlings germinated in the laboratory. Total RNA was isolated using the RNeasy Plant mini

kit and protocol from Qiagen (Valencia, CA). Northern blots were performed as described by Sambrook and Russell (2001). A 1,208 bp fragment GmKASIIIA probe was amplified from cDNA synthesized for qPCR expression analysis. The probe was labeled using the Random Primed DNA Labeling Kit from Roche and [32P] dCTP from Perkin Elmer. Hybridization was done overnight at 65°C using PerfectHyb Plus Hybridization Buffer from Sigma. The autorad was developed after 5 days of exposure.

Results

Linkage map

Linkage analysis was conducted using MAPMAKER/EXP 3.0 with 98 polymorphic SSR markers and 420 polymorphic SNPs in the FAE population, which cover the 20 linkage groups (LGs) of the soybean genome. A total of 451 markers were uniquely placed on the final map of 31 LGs and 3 unlinked loci because some markers were redundant (no recombination) or could not be placed uniquely on the map and due to the lack of polymorphisms among the parents across large genomic areas, linkage could not be declared among all markers on the same chromosome (Supplementary Table 1). Chromosomes 1, 2, 4, 7, 8, 13, 14, 15, 16, 17, and 20 were split into 2 LGs each. In addition, there were 14 large gaps (26–50 cM) in the final map. The total length of the linkage map was 2,351 cM, equivalent to published consensus maps (Hyten et al. 2010). The order of marker loci within each linkage group was the same as the order presented in Hyten et al. (2010) with a few exceptions: in chromosome 4, BARC062503-17863 and BARC042719-08395; in chromosome 5, BARC021775-04203; in chromosome 9, BARC061897-17586; in chromosome 10, BARC038327-10019; in chromosome 13, BARC064707-18874 and BARC041141-07916; in chromosome 14, BARC017589-02630; in chromosome 15, BARC030171-06819 and BARC061277-17149; in chromosome 17, BARC047903-10427; in chromosome 18, BARC063581-18909 and BARC064177-18581; in chromosome 19, BARC064609-18739, BARC021733-04193, BARC035235-07156, and BARC047496-1294; are all in slightly different local orders (Supplementary Table 1).

Loci were declared to have significant segregation distortion if the Chi-test had a p value <0.0016 (0.05/31 LGs) (Supplementary Table 1). Nine markers encompassing a large genomic region (~45 cM) in chromosome 5 are severely distorted. The low-palmitate fapnc mutation caused by a deletion of the GmFATB1a gene maps to the end of this chromosome (Cardinal et al. 2007; Li et al. 2002; and this study).

There are 9 marker loci on chromosome 6 (~20 cM), 4 markers on chromosome 17 (~7.8 cM), 17 markers on chromosome 18 (~23.6 cM), 15 markers in chromosome 19 (~43.4 cM) that show significant segregation distortion.

Mapping the fap1 locus

Only BLUPs derived from analyses across environments were utilized for QTL analysis because the lower bound of the pooled genetic correlations among environments for each trait (Cooper and DeLacy 1994), were higher than 0.8 (except for oil content and seed yield, data not shown). Single factor marker analysis for palmitate content detected two major QTL, one in chromosome 5 (GmFATB1a) and the other at the end of chromosome 9. Based on those results, BARCSOYSSR_09_1707 (46.31 Mb, Supplemental Table 1) was the closest marker to the QTL on chromosome 9.

In order to identify candidate genes for reduced palmitic acid content due to fap1, we examined the Williams 82 reference sequence (Schmutz et al. 2010), corresponding to the genomic location near BARCSOYSSR_09_1707. Two genes were annotated as involved in fatty acid biosynthesis and were chosen as putative candidate genes: Glyma09g39610, annotated as a putative Acyl-CoA thioesterase gene and Glyma09g41380, annotated as a putative ketoacyl-ACP synthase III gene (KASIII) (http://www.phytozome.net/soybean). We PCR amplified and Sanger sequenced both genes using genomic DNA from Century (WT) and its EMS induced mutant C1726 (fap1). For Glyma09g39610, assembled gene sequences were identical.
between Century and C1726 (and Williams 82). In contrast, examination of sequencing results for Glyma09g41380 revealed a single unique polymorphism in C1726 (G174A, relative to the start codon) shown in Fig. 1, which was predicted to disrupt the donor splice site recognition between exon one and intron one (Hebsgaard et al. 1996; Long and Deutsch 1999). The resulting KASIIIa protein would have a truncation of the open reading frame of the Williams 82 version of 397 predicted residues to:

\[
\text{MANASGFTPSVPHKVRFKPLNTAIGFSAKVVF} \\
\text{GNIEQAEKASTVPSQSPIPRSLFLLCLIF*}
\]

In addition, the G to A change in C1726 creates the GATCTG motif that has been associated with enhanced gene addition, the G to A change in C1726 creates the GATCTG

\[
\text{GAAAAACATGCTTCCACCGTTTCCCCTTCTTCCCAATCTCCAATACCCAG} \\
\text{TCGCTTTTTCCCTGCTTCATCTTTCTCCGCTTTTCTCCTGCTTCATCTTTCTCCGCTTAAAGTTTTGATT-3'}
\]

A KASPar marker was developed for the SNP in *GmKASIIIA* gene and genotyped in the FAE population. *GmKASIIIA* mapped to Chromosome 9 between BARC-GmKASIIIA and genotyped in the FAE population. (Rose et al. 2008).

![Fig. 1 Region of KASIIIA genomic sequence Glyma09g41380 in Century and in the EMS induced mutant C1726 (fap1). A portion of Glyma09g41380 exon 1 is underlined, and an arrow indicates position of splice site at junction between exon 1 and intron 1. SNP in C1726 is located at the first base of intron 1](image)

In the FAHH00 and FADD00 populations an ANOVA was performed with *GmFATB1a* and the BARCSOYSSR_09_1707 (Sat_1707, closely linked to *GmKASIIIA*) and their interaction as independent variables and BLUPs for each fatty acid and agronomic traits for each line as dependent variables. The fap$_{nc}$, Sat_1707 (*GmKASIIIA*) and their interaction (in FADD00) explained 86–94 % of the variation for 16:0 content (Table 5). Lines homozygous for the Satelite allele at both loci had on average 44–45 g of 16:0 per kg of total lipids. Both alleles significantly reduced stearic acid content in both populations and the Sat_1707 allele inherited from the fap1 mutant parent (Satelite) significantly increased linoleic acid content in one population. The Sat_1707 allele from Satelite and its interaction decreased yield in FADD00 (Table 7). The Sat_1707 allele from Satelite and its interaction decreased oil content in both populations (Table 7). Lines homozygous for the Satelite allele at both loci matured significantly later in both populations.

qPCR expression of fatty acid related genes

*GmKASIIIAa, GmFATB1a, GmFATB1b, GmFATB2a* and *GmFATB2b* in soybean lines from FAE population

In order to test if the SNP that creates a GATCTG motif in the *GmKASIIIA* gene (from C1726) for enhanced gene expression as predicted from studies in Arabidopsis, RNA was obtained from seeds harvested 35 days after flowering from five lines of the FAE population of similar flowering date and maturity date and also from the parental line Satelite. cDNA was synthesized to perform quantitative RT-PCR analysis. Lines homozygous or heterozygous for the *GmFATB1a* allele (Fap$_{nc}$Fap$_{nc}$ or Fap$_{nc}$/Fap$_{nc}$ genotypes) and homozygous for the *GmKASIIIA* allele inherited from Satelite (mutant allele) had 1,247 Attamoles more of...
GmKASIIIA RNA per mg of total RNA than lines homozygous for the GmKASIIIA allele inherited from N98-4445 (wild type) (Table 8), indicating that the G to A change present in the first intron in C1726 and Satellite that creates the GATCTG motif enhanced gene expression. However, this transcript would produce a truncated protein. Sequencing analysis confirmed that the amplified products corresponded to the GmKASIIIA (Glyma09g41380) homolog as expected.

Satellite (mutant fapnc and GmKASIIIA genotype) had a 10-fold reduction in the amount of GmFATB1a transcript and significant increases of GmFATB1b, GmFATB2a, and GmFATB2b transcripts.

### Table 4
Palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid content BLUPs (g kg\(^{-1}\)) of seed oil of parental lines and genotypic means of progenies homozygous fapnc fapnc and homozygous wild type for GmKASIIIA (AA), homozygous fapnc fapnc and homozygous mutant for KASIIIA (BB), homozygous or heterozygous Fapnc/C0 and homozygous wild type for KASIIIA (AA), and homozygous or heterozygous Fapnc/C0 and homozygous mutant for KASIIIA (BB) in FAE population (Satelite × N98-4445)

<table>
<thead>
<tr>
<th>Genotype</th>
<th>No. of individuals</th>
<th>Palmitic acid (g kg(^{-1}))</th>
<th>Stearic acid (g kg(^{-1}))</th>
<th>Oleic acid (g kg(^{-1}))</th>
<th>Linoleic acid (g kg(^{-1}))</th>
<th>Linolenic acid (g kg(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>fapnc fapnc KASIIIA (AA)</td>
<td>76</td>
<td>63.3</td>
<td>36.6</td>
<td>428.1</td>
<td>439.4</td>
<td>32.1</td>
</tr>
<tr>
<td>fapnc fapnc KASIIIA (BB)</td>
<td>49</td>
<td>49.0</td>
<td>35.5</td>
<td>419.7</td>
<td>460.9</td>
<td>34.0</td>
</tr>
<tr>
<td>Fapnc KASIIIA (AA)</td>
<td>157</td>
<td>89.0</td>
<td>37.9</td>
<td>427.5</td>
<td>413.8</td>
<td>32.4</td>
</tr>
<tr>
<td>Fapnc KASIIIA (BB)</td>
<td>186</td>
<td>69.8</td>
<td>37.0</td>
<td>432.7</td>
<td>426.6</td>
<td>33.7</td>
</tr>
<tr>
<td>fapnc fapnc vs Fapnc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KASIIIA (AA vs BB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epistatic interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R(^2) model (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satellite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N98-4445</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5
Genotypic means of seed yield, date of flowering (R2), maturity date (R8), oil concentration, and height of lines homozygous fapnc fapnc and homozygous wild type for GmKASIIIA (AA), lines homozygous fapnc fapnc and homozygous mutant for GmKASIIIA (BB), lines homozygous or heterozygous Fapnc/C0 and homozygous wild type for GmKASIIIA (AA), and lines homozygous or heterozygous Fapnc/C0 and homozygous mutant for GmKASIIIA (BB) in FAE population (Satelite × N98-4445)

<table>
<thead>
<tr>
<th>Genotype</th>
<th>No. of individuals</th>
<th>Yield (kg ha(^{-1}))</th>
<th>R2 (DAP(^a))</th>
<th>R8 (DAP(^a))</th>
<th>Oil (g kg(^{-1}))</th>
<th>Height (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fapnc fapnc KASIIIA (AA)</td>
<td>76</td>
<td>1,474</td>
<td>53</td>
<td>130</td>
<td>196.3</td>
<td>89</td>
</tr>
<tr>
<td>fapnc fapnc KASIIIA (BB)</td>
<td>49</td>
<td>1,384</td>
<td>55</td>
<td>133</td>
<td>188.9</td>
<td>88</td>
</tr>
<tr>
<td>Fapnc KASIIIA (AA)</td>
<td>157</td>
<td>1,535</td>
<td>53</td>
<td>130</td>
<td>193.5</td>
<td>99</td>
</tr>
<tr>
<td>Fapnc KASIIIA (BB)</td>
<td>186</td>
<td>1,533</td>
<td>53</td>
<td>130</td>
<td>188.9</td>
<td>99</td>
</tr>
<tr>
<td>fapnc fapnc vs Fapnc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KASIIIA (AA vs BB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epistatic interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R(^2) model (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satellite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N98-4445</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Differences between genotypic means and their significance and the R\(^2\) for the full model in Proc GLM are included

<table>
<thead>
<tr>
<th>Significance</th>
<th>(p) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n.s.</td>
<td></td>
</tr>
<tr>
<td>Significant at ****</td>
<td>(p &lt; 0.0001)</td>
</tr>
<tr>
<td>Significant ***</td>
<td>(p &lt; 0.001)</td>
</tr>
<tr>
<td>Significant **</td>
<td>(p &lt; 0.01)</td>
</tr>
<tr>
<td>Significant *</td>
<td>(p &lt; 0.05)</td>
</tr>
</tbody>
</table>

### Notes:

- \(a\) Days after planting
- Significant at **** \(p < 0.0001\); significant *** \(p < 0.001\); significant ** \(p < 0.01\); and significant * \(p < 0.05\)

GmKASIIIA RNA per mg of total RNA than lines homozygous for the GmKASIIIA allele inherited from N98-4445 (wild type) (Table 8), indicating that the G to A change present in the first intron in C1726 and Satellite that creates the GATCTG motif enhanced gene expression. However, this transcript would produce a truncated protein. Sequencing analysis confirmed that the amplified products corresponded to the GmKASIIIA (Glyma09g41380) homolog as expected.

Satellite (mutant fapnc and GmKASIIIA genotype) had a 10× reduction in the amount of GmFATB1a transcript and significant increases of GmFATB1b, GmFATB2a, and GmFATB2b transcripts.

Sequencing of fragment amplified from wildtype and GmKASIIIA mutant cDNA

A fragment of 579 bp was amplified from cDNA of the wildtype GmKASIIIA sample, and a fragment of 799 bp was amplified from cDNA of the KASIIIA mutant sample, which would result from the presence of the unspliced 220 bp intron 1 in the KASIIIA mutant. Alignment of fragment sequences from 3 clones from each genotype ID shows that introns 1, 2, 3 are not present in the wildtype fragment. The mutant fragment did not contain introns 2 and 3 but it contained the entire sequence of intron 1.
Northern analysis of GmKASIIIA

In order to prove if the single unique polymorphism in C1726 (G174A, relative to the start codon) and Satelite, disrupted the donor splice site recognition between exon one and intron one, Northern Analysis was performed. Northern blots showed the presence of a larger GmKAS-III mRNA isolated from pods and leaves of Satelite, confirming the presence of intron 1 and stability of the mRNA. The GmKASIIIA mRNA isolated from pods harvested 50 days after flowering clearly shows a smaller band in N98-4445 that corresponds to the spliced mRNA lacking intron 1 (Fig. 2).

Discussion

In most regions, the linkage map developed had the same order expected from the Williams 82 Glycine max genome sequence (Mb) (v 1.01, verified 7/2012), but there are some important discrepancies in chromosomes 4, 5, 8, 9, 11, 14, 15, and 20 (Supplementary Table 1). Because N98-4445 (female parent) was an F3-derived line, it was not completely homozygous. Since the RILs developed in this study were derived from 3 different F1 plants, there were instances in which a marker was polymorphic in a subset of RILs but not in another subset of RILs derived from a different F1 plant. In such instances, the marker was scored as missing data in the subset of RILs that showed no segregation (monomorphic). There were a few marker loci in almost all LGs that showed this pattern of segregation (Supplementary Table 1), however, there were large genomic regions in chromosomes 8, 9, 12, 14, and 17 in which many markers segregated in a subset of RILs. Marker order and QTL (and their effect) observed in those regions would probably be biased due to the small sample of RILs that provided genotypic and phenotypic information.

The severe segregation distortion observed in the vicinity of the low-palmitate fapnc mutation had been reported before (Cardinal et al. 2007; Li et al. 2002) and the allele-specific marker designed for this mutation was
the most severely distorted. The other markers on the same chromosome (LG5) become less distorted the farther away they map from this locus.

Other chromosomal regions that also showed segregation distortion in this study could possibly be related to fitness loci. For example, on chromosome 6, in the genomic region where markers displayed segregation distortion, QTL for abnormal seedling, pod maturity and flowering date have been reported (http://soybeanbreederstoolbox.org; verified 8/2012). On chromosome 18, in the genomic

Table 7 Genotypic means of seed yield, date of flowering (R2), maturity date (R8), protein and oil concentrations, lodging rating, and height lines homozygous *fapnc* and homozygous AA for *Sat_1707* (AA), lines homozygous *fapnc*fapnc and homozygous BB for *Sat_1707* (BB) in two F1-derived populations

<table>
<thead>
<tr>
<th>Genotype</th>
<th>No. of individuals</th>
<th>Yield (kg ha(^{-1}))</th>
<th>R2 (DAP(^{a}))</th>
<th>R8 (DAP(^{a}))</th>
<th>Protein (g kg(^{-1}))</th>
<th>Oil (g kg(^{-1}))</th>
<th>Lodging (1-5(^{b}))</th>
<th>Height (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fapncfapnc, <em>Sat_1707</em> (AA)</td>
<td>6</td>
<td>2,588</td>
<td>56</td>
<td>136</td>
<td>437.5</td>
<td>199.9</td>
<td>2.6</td>
<td>87</td>
</tr>
<tr>
<td>fapncfapnc, <em>Sat_1707</em> (BB)</td>
<td>15</td>
<td>2,402</td>
<td>61</td>
<td>149</td>
<td>438.5</td>
<td>186.6</td>
<td>2.5</td>
<td>97</td>
</tr>
<tr>
<td>Fapnc, <em>Sat_1707</em> (AA)</td>
<td>37</td>
<td>2,736</td>
<td>58</td>
<td>143</td>
<td>430.0</td>
<td>197.1</td>
<td>2.6</td>
<td>107</td>
</tr>
<tr>
<td>Fapnc, <em>Sat_1707</em> (BB)</td>
<td>32</td>
<td>2,731</td>
<td>57</td>
<td>142</td>
<td>429.8</td>
<td>193.6</td>
<td>2.5</td>
<td>107</td>
</tr>
<tr>
<td>fapncfapnc vs Fapnc</td>
<td>n.s.</td>
<td>n.s.</td>
<td>0.8**</td>
<td>n.s.</td>
<td>n.s.</td>
<td>-1.5***</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Sat_1707</em> (AA vs BB)</td>
<td>n.s.</td>
<td>n.s.</td>
<td>-6*</td>
<td>n.s.</td>
<td>8.4****</td>
<td>n.s.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epistatic interaction</td>
<td>n.s.</td>
<td>n.s.</td>
<td>-13**</td>
<td>n.s.</td>
<td>9.8**</td>
<td>n.s.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2 model (%)</td>
<td>27.0</td>
<td>6.9</td>
<td>11.3</td>
<td>12.7</td>
<td>28.2</td>
<td>4.3</td>
<td>15.1</td>
<td></td>
</tr>
</tbody>
</table>

Table 8 RT-PCR results (Attomoles/mg RNA) from developing seeds RNA of 5 F5-derived lines from FAE population and ‘Satelite’ harvested at 35 days after flowering in Clayton 2010. Assays were done in triplicates (subsamples)

<table>
<thead>
<tr>
<th>Genotype</th>
<th>N(^{a})</th>
<th>GmKASIIIA</th>
<th>GmFATB1a</th>
<th>GmFATB1b</th>
<th>GmFATB2a</th>
<th>GmFATB2b</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>fapncfapnc</em>, KASIIIA (BB) (Satelite)</td>
<td>1</td>
<td>2,788.5</td>
<td>187.5</td>
<td>568.9</td>
<td>181.4</td>
<td>32.8</td>
</tr>
<tr>
<td><em>Fapnc</em>, KASIIIA (H)</td>
<td>1</td>
<td>1,072.6</td>
<td>1,547.5</td>
<td>306.5</td>
<td>60.3</td>
<td>19.2</td>
</tr>
<tr>
<td><em>Fapnc</em>, KASIIIA (AA)</td>
<td>2</td>
<td>953.5</td>
<td>1,591.3 (23)</td>
<td>381.5 (55.1)</td>
<td>84.5 (5.7)</td>
<td>16.9 (2.4)</td>
</tr>
<tr>
<td><em>Fapnc</em>, KASIIIA (BB)</td>
<td>2</td>
<td>19.06.8</td>
<td>1,425 (141.4)</td>
<td>383.1 (51.6)</td>
<td>91.5 (21.7)</td>
<td>17.5 (4.7)</td>
</tr>
<tr>
<td>T test <em>fapnc</em>,<em>fapnc</em> vs <em>Fapnc</em></td>
<td>n.s.</td>
<td>-1,328.5 (121.7)**</td>
<td>201.8 (55.6)*</td>
<td>99.0 (18.7)**</td>
<td>15.3 (3.1)**</td>
<td></td>
</tr>
<tr>
<td>T test KASIIIA (AA vs BB)(^{c})</td>
<td>-1,247.2*(397)</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td></td>
</tr>
</tbody>
</table>

| n.s. not significant |

Differences between genotypic means and their significance and the R2 for the full model in Proc GLM are included

n.s. not significant

a Days after planting

b Scored 1 if plants are 90 % erect to 5 if plants are 90 % procumbent

Significant at ****p < 0.0001; significant ***p < 0.001; significant **p < 0.01; and significant *p < 0.05
region where markers displayed segregation distortion, QTL for flowering time have been reported (http://soybeanbreederstoolbox.org; verified 8/2012). On chromosome 19 a seed set QTL was reported in the genomic region near markers with segregation distortion. If QTL for all these traits were segregating in our population, unintended selection against alleles with undesirable effects may have occurred during inbreeding causing the observed segregation distortion. Standard likelihood approaches for linkage mapping are not best suited for markers displaying segregation distortion (Lorieux et al. 1995), therefore the markers order and distance in the genomic regions described above should be taken with caution. It should also be noted that the recombination frequency among markers and map distances are slightly upwardly biased by the recombination frequency with standard likelihood methods (Lorieux et al. 1995) or problems in the assembly of the Williams 82 whole genome shotgun sequence (v1.01) in this region.

The results from the sequence analysis of the two candidate genes on chromosome 9 were validated with the gene expression and Northern blot assays, confirming that the fap1 low-palmitate phenotype is due to a single unique polymorphism in C1726 (G174A, relative to the start codon), which was predicted to disrupt the donor splice site recognition between exon one and intron one and also creates the GATCTG motif. The gene expression assay demonstrated that this motif enhanced gene expression of the GmKASIIIA locus, as had been observed before in Arabidopsis (Rose et al. 2008). The increases in GmFATB1b, GmFATB2a, and GmFATB2b transcripts observed in Satelite (fapnc and GmKASIIIA mutations) showed that when the transcription of the most highly transcribed GmFATB1a homolog is compromised (Wilson et al. 2001b; Cardinal et al. 2007), the soybean plant up regulates the transcription of the other three GmFATB genes to try to maintain a “normal” fatty acid biosynthesis. We speculated that since the 3-ketoacyl-ACP synthase enzyme III catalyzed the first condensation reaction in the biosynthesis of C16 and C18 fatty acids in developing seeds, a significant reduction in 16:0, 18:0 and total oil content would be observed in lines homozygous for the GmKASIIIA mutation. This was confirmed in all three populations and is in agreement with previous studies with the same mutation (Ndzana et al. 1994; Horejsi et al. 1994).

As we indicated previously (Cardinal et al. 2007), the GmFATB1a gene, which encodes the palmate thioesterase in soybean and is deleted in the fapnc allele, is putatively active toward stearoyl acyl carrier protein (ACP) substrates and, therefore, the mutations in this gene would reduce both 16:0 and 18:0 contents in the seed oil. This effect was observed in all the populations. The fapnc locus also caused a small increase in total oil content.

Because our study includes the FAE population with 712 F5-derived lines, our estimates of the genetic parameters have excellent precision. Lines homozygous for the fapnc mutation showed significantly reduced yield in all of the populations evaluated and this decreased yield was fairly consistent across the environments. A significant reduction in plant height was also observed for this mutation. In contrast, lines homozygous for the fap1 and fap3 alleles have increased plant height (Ndzana et al. 1994), and therefore, since fap3 (amino acid substitution) and fapnc (deletion of unknown size) are mutations in the same GmFATB1a gene, it is highly likely that the deletion per se may include another gene that is causing the plant height reduction observed in fapnc mutants. Yield reductions associated with the fap1 mutation were of smaller

Fig. 2 Northern-blot analysis of GmKASIIIA mRNA isolated from pods harvested 50 days after flowering in the germplasm line N98-4445 (lane 1) and in the cultivar ‘Satelite’ (lane 2)
magnitude and not always consistent across environments. This observation is in agreement with previous studies on the effect of the fap1 mutation on agronomic traits (Rebetzke et al. 1998a; Cardinal et al. 2008). Whether this small negative effect on yield is due to the mutation per se or to negative linkage drag is something that cannot be conclusively determined in these populations because none of the parents was a high yielding line, comparable to modern cultivars. Studying the agronomic effects of this mutation on a high yielding-high oil genetic background will help to reveal its potential as a source of low palmitate that will not compromise agronomic performance.

Acknowledgments The authors would like to thank William Novitzky at the USDA-ARS Soybean and Nitrogen Fixation Research Unit, Raleigh, NC, for providing training and equipment for the fatty acid analysis; James B. Holland for providing technical expertise in QTL analysis and discussion of results; Martha Ramirez for isolating RNA; and Carol Griffin for performing Northern analysis. This project was funded by the United Soybean Board (Grant # 59-6645-2-071).

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

Ethical Standards The experiments performed for this publication comply with the current laws of the United States of America.

References

Aguilar F, Montandon PE, Stutz E (1991) Two genes encoding the soybean translation elongation factor eEF-1 alpha are transcribed in seedling leaves. Plant Mol Biol 17:351–360


Burton JW, Wilson RF, Brim CA (1994b) Registration of N79-2077-12 and N87-2122-4, two soybean germplasm lines with reduced palmitic acid in seed oil. Crop Sci 34:313


Fehr WR, Caviness CE (1977) Stages of soybean development. Special Report. Agriculture and Home Economics Experiment Station. Iowa State University 80:11


Primomo VS (2000) Inheritance and stability of palmitic acid alleles in soybeans (Glycine max L. Merr.). MS Thesis, Department of Plant Agriculture, University of Guelph, Guelph, Ontario, Canada


