Soybean [Glycine max (L.) Merr.] cultivars have been developed that contain about half the concentration of saturated fatty acids in the oil as conventional cultivars (Fehr, 2007). Lowering the amount of saturated fatty acids in the human diet is desirable for reducing the risk of coronary heart disease (Hu et al., 1997). Reduction of the saturated fatty ester concentration in soybean oil has been achieved by the use of alleles that lower the concentration of palmitate, the predominant saturated fatty ester found in the seed oil. The allele fap1(C1726) was developed by treatment of seeds of the cultivar Century with ethyl methanesulfonate (Erickson et al., 1988). This allele decreased the palmitate concentration in the line C1726 from about 120 to about 86 g kg^{-1}. Another allele was identified in soybean line N79-2077 that lowered palmitate concentration to about 60 g kg^{-1} (Burton et al., 1994). This allele, designated originally as fapnc, was considered to be a natural mutation. Genetic studies determined fapnc to be allelic to fap3(A22), and it was given the designation fap3nc (Fehr, 2007). Cardinal et al. (2007) found that soybean lines containing the fap3nc allele had a deletion in a gene encoding the 16:0–acyl carrier protein (ACP) thioesterase (TE) enzyme that is responsible for generating free palmitate in the seed oil. The allele fap1(C1726) was developed by treatment of seeds of the cultivar Century with ethyl methanesulfonate (Erickson et al., 1988). This allele decreased the palmitate concentration in the line C1726 from about 120 to about 86 g kg^{-1}.

Another allele was identified in soybean line N79-2077 that lowered palmitate concentration to about 60 g kg^{-1} (Burton et al., 1994). This allele, designated originally as fapnc, was considered to be a natural mutation. Genetic studies determined fapnc to be allelic to fap3(A22), and it was given the designation fap3nc (Fehr, 2007). Cardinal et al. (2007) found that soybean lines containing the fap3nc allele had a deletion in a gene encoding the 16:0–acyl carrier protein (ACP) thioesterase (TE) activity that is responsible for generating free palmitate that ultimately becomes incorporated.

Molecular characterization of the mutant fap3(A22) allele for reduced palmitate concentration in soybean

Brian D. De Vries, Walter R. Fehr,* Grace A. Welke, and Ralph E. Dewey

ABSTRACT

Reduction of the palmitate concentration in soybean [Glycine max (L.) Merr.] oil is desirable for reducing the amount of saturated fat in the human diet. Chemical mutagenesis was used to develop the line A22 with the mutant allele designated fap3(A22) that reduces palmitate concentration in the seed oil. The objective of our study was to determine the molecular basis of the fap3(A22) mutation and develop a corresponding molecular marker to assist in future efforts for developing soybean cultivars with low saturated fat. DNA sequence analysis of the GmFATB1a gene of soybean revealed a single nucleotide polymorphism (SNP) resulting in a nonconservative amino acid substitution that was likely to be detrimental to the function of the 16:0–acyl carrier protein (ACP) thioesterase (TE) enzyme. An association analysis was conducted using F₂-derived lines from a cross between the cultivar Archer (Fap3Fap3) and A22 (fap3fap3) that had been analyzed for their palmitate concentration by gas chromatography. Molecular genotyping of these lines established a perfect correlation between lines phenotypically classified as homozygous for the Fap3 allele or homozygous for the fap3(A22) allele based on their palmitate concentration. The polymorphism in the GmFATB1a gene was used to develop a functional, codominant marker that could be used to distinguish the Fap3 and fap3(A22) alleles in segregating populations. This marker will be useful for breeders who are developing low-saturated cultivars with the fap3(A22) allele.
into the seed storage triacylglycerols. Four distinct 16:0-ACP TE–encoding gene isoforms were identified in the soybean genome, and the fap3 allele was associated with a deletion in GmFATB1a, the isoform that appears to be the most highly transcribed (Cardinal et al., 2007).

The mutant allele fap3(A22) was developed by treatment of the cultivar A1937 with N-nitroso-N-methylurea (Fehr et al., 1991; Schnebly et al., 1994). This allele in the line A22 is responsible for lowering the palmitate concentration to about 70 g kg⁻¹. By crossing C1726 to A22, the palmitate concentration of lines homozygous for both fap1(C1726) and fap3(A22) was lowered to about 44 g kg⁻¹ (Fehr et al., 1991). This combination of alleles has been used in the cultivars with low saturated fat that currently are in commercial production.

The identification of a difference in the DNA sequence of the Fap3 and fap3(A22) alleles would make it possible to develop a molecular marker that could be used in a breeding program. When soybean breeders cross cultivars with the fap1(C1726) and fap3(A22) alleles to conventional cultivars with the Fap1 and Fap3 alleles, the only method available at present to identify those with the fap3(A22) allele is to analyze seed by gas chromatography to select those with reduced palmitate. If a molecular marker for the allele could be developed, tissue of individual plants in the F2 or later generations of breeding could be analyzed and those homozygous for the allele could be saved. This would be particularly important for identifying plants with the allele when backcrossing it into a conventional cultivar. Molecular markers for other seed traits in soybean are used in cultivar development programs (Bilyeu et al., 2006; Gillman et al., 2009; Lenis et al., 2010).

The objective of our study was to identify a molecular difference between the Fap3 and fap3(A22) alleles and develop a functional marker that could be used for selection in breeding programs for the development of low-saturate cultivars.

MATERIALS AND METHODS

Plant Materials

The soybean cultivar Archer that is homozygous for the wild-type allele Fap3 and the mutant line A22 that is homozygous for the mutant allele fap3(A22) were crossed at the Agricultural and Agronomy Research Center near Ames, IA. The F1 seeds were planted at the Iowa State University–University of Puerto Rico soybean breeding nursery at Isabela, PR. Individual F2 plants were harvested and five individual seeds from each plant were evaluated by gas chromatography as described by Hammond (1991) to confirm hybrid plants by segregation for palmitate concentration. There were 220 F2 seeds and seeds of the two parents planted at Isabela. Individual F2 and parent plants were harvested.

The two parents and the F3 progeny of 102 random F2 plants were planted in a randomized complete-block design with one replication at the Agronomy Farm and one replication at the Burkey Farm of Iowa State University near Ames, IA. This was an adequate number of F2–derived lines to obtain approximately 25 individuals that were homozygous for Fap3 and fap3(A22) based on the known inheritance of the gene (Fehr et al., 1991; Schnebly et al., 1994). The soil type at both locations is a Nicotlet loam (fine-loamy, mixed, superactive, mesic Aquic Hapludoll). For each plot, 20 seeds were planted in rows 0.76 m long with a spacing of 1.02 m between rows and an alley 1.07 m wide between the ends of plots. Single pods harvested from the individual plants in each plot were threshed in bulk.

Twenty individual F3 seeds from each F2–derived line and 80 individual seeds of each parent were evaluated for palmitate concentration by gas chromatography. The range in palmitate concentration among individual seeds of A22 was 56 to 84 g kg⁻¹ and for Archer was 90 to 111 g kg⁻¹. The 20 lines for which all of the 20 seeds had <84 g kg⁻¹ palmitate were considered to be homozygous for the fap3(A22) allele and the 21 lines for which all seeds had >90 g kg⁻¹ palmitate were considered to be homozygous for the Fap3 allele. The mean palmitate concentration plus or minus the SEM was ±1 ± 1 g kg⁻¹ for the 20 fap3(A22)fap3(A22) lines and ±1 ± 1 g kg⁻¹ for the 21 Fap3Fap3 lines. Lines that had one or more seeds <84 and >90 g kg⁻¹ were assumed to be from F3 plants with the genotype Fap3fap3(A22) and were not used for molecular analysis.

Amplification and Sequence Analysis of the GmFATB1a Complementary DNA from A22

To characterize the GmFATB1a complementary DNA (cDNA) from A22, total cellular RNA was isolated from 1 g of young leaf tissue about 14 d after germination using the TRizol reagent, according to the manufacturer’s protocol (Invitrogen, Carlsbad, CA). Poly (A)+ RNA was recovered from 5 μg RNA using the Messagemaker system (Invitrogen, Carlsbad, CA). First-strand cDNA was synthesized from poly (A)+ RNA using an oligo-dT primer and SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA). A full-length cDNA was amplified from 20 ng of first-strand A22 cDNA template using the GmFATB1a–specific polymerase chain reaction (PCR) primers described by Cardinal et al. (2007). The amplification product was cloned into the TA vector (Taq or other type of polymerase cloning vector) pCR2.1 (Invitrogen, Carlsbad, CA) and sequenced at the North Carolina State University Genome Sciences Laboratory (Raleigh, NC) using an Applied Biosystems 3730 (Applied Biosystems, Carlsbad, CA) capillary sequencer.

Molecular Genotyping of F2–Derived Lines from Archer × A22 Cross

Single leaf disks 1 cm in diameter were harvested from individuals plants of Archer and A22 and individual F1 plants of the 20 F2–derived lines phenotypically classified as homozygous for the fap3(A22) allele and 21 F2–derived lines phenotypically classified as Fap3Fap3. Analysis of a single F1 plant was considered sufficient because the homogeneity of each line for the Fap3 or fap3(A22) alleles was known based on the palmitate concentration of the 20 individual seeds evaluated by gas chromatography. In all cases, tissue was harvested about 14 d after germination. Leaf disks were ground with liquid N in 1.5-mL microfuge tubes using plastic pestles. Genomic DNA was isolated from the ground material by adding 325 μL extraction buffer (200 mM Tris HCl pH 7.5, 250 mM NaCl, 25 mM ethylenediaminetetraacetate [EDTA], and 0.5% sodium dodecyl sulfate [SDS]) and vortexing for 20 sec. After a 10 min incubation at 65°C, 60 μL of protein precipitation solution
(Qiagen, Germantown, MD) was added, and the samples were vortexed briefly before being placed on ice for 5 min. The samples were centrifuged at 16,000 × g for 4.5 min and the supernatants were transferred to fresh tubes. DNA precipitation was mediated by the addition of 250 μL isopropanol followed by a 10 min incubation at room temperature and centrifugation at 16,000 × g for 7.5 min. After being washed with 70% ethanol, the pellets were dried and resuspended in 50 μL Tris EDTA buffer.

Polymerase chain reaction assays were conducted in 96-well plates using a MyCycler Thermal Cycler System (BioRad, Hercules, CA). Individual reactions contained 1 μL genomic DNA (typically about 20 ng), 1.5 μL of 10x Taq buffer (New England Biolabs, Ipswich, MA), 6 pmol of each primer, 200 μM deoxyribonucleotide triphosphates (dNTPs) (Roche Applied Science, Indianapolis, IN), and 1.5 U of Taq DNA polymerase (New England Biolabs, Ipswich, MA) in a total volume of 15 μL. The PCR primers used were the same as those described by Cardinal et al. (2007) to amplify a 312 bp GmFATB1a-specific molecular marker designated FATB1a-312. The thermal cycling profile consisted of denaturation at 94°C for 5 min followed by 30 cycles of 94°C for 30 sec, 52°C for 30 sec, and 72°C for 45 sec, with a final extension at 72°C for 7 min. Excess dNTPs and primers were removed by adding 1 U of Shrimp Alkaline Phosphatase (Promega, Madison, WI) and 2 U Exonuclease 1 (New England Biolabs, Ipswich, MA) to 7 μL of the PCR product in a total volume of 10 μL. The tubes were incubated at 37°C for 45 min followed by heat inactivation of the enzymes at 80°C for 15 min. Samples were stored at −20°C until used for sequence analysis.

The PCR products were directly sequenced using either the forward or reverse PCR primers as the sequencing primer in a cycle-sequencing-labeling reaction using BigDye Terminator version 3.1 (Applied Biosystems, Carlsbad, CA). Each sequencing reaction contained 5 μL PCR DNA, 0.7 μL Big Dye, 1.6 μL of 5x buffer and 3.2 pmol of primer in final volume of 10 μL. Cycling conditions were 95°C for 30 sec followed by 25 cycles of 96°C for 10 sec, 50°C for 5 sec, and 60°C for 4 min. Removal of the dye terminator was accomplished using Per-3 B, according to the manufacturer’s protocol (EdgeBio, Gaithersburg, MD). Samples were loaded on an ABI 3730 DNA analyzer (Applied Biosystems, Carlsbad, CA) for sequence determination at the North Carolina State University Genome Sciences Laboratory (Raleigh, NC).

SimpleProbe Marker Development

Ten soybean plants of A22, ‘1A1025’, Archer, and ‘Williams 82’ were grown and fresh leaf tissue was harvested in bulk from 10-d-old seedlings. 1A1025 was a low-saturate cultivar homozygous and homogeneous for both the fap3 (A22) alleles that reduce palmitate concentration. Williams 82 was used because its genome is publicly available and the sequence of the GmFATB1a gene of Williams 82 is identical to the GmFATB1a gene of Century. The fresh plant tissue was lyophilized for 24 h by the Iowa State University DNA Facility (Ames, IA) and genomic DNA was isolated using an AutoGen 740 (Autogen, Holliston, MA) and a Geno/Grinder 2000 (SPEX CertiPrep Inc., Metuchen, NJ).

A SimpleProbe molecular marker for the fap3 (A22) allele was developed corresponding to the single nucleotide polymorphism (SNP) identified in the GmFATB1a gene (Roche Applied Sciences, Indianapolis, IN). The SimpleProbe for the fap3 (A22) assay, designed by Jason Gillman of the USDA-Agricultural Research Service at the University of Missouri (Columbia, MO), was 5'-AAGCCATCAGCCGATAC-phosphate–3'. Asymmetric PCR was performed to maximize single stranded DNA synthesis for probe binding. The assay was performed with an asymmetric mixture of the forward primer (5'-TGACATAGTTCAAGTGACACT-3') at 0.2 μM and reverse primer (5'-GCAAAATGCAAATGTACCCTG-3') at 0.5 μM final concentration. Reactions were performed in a total volume of 20 μL containing template, primers, SimpleProbe (0.4 μM final concentration), buffer (40 mM Tricine-KOH [pH 8.0], 16 mM KCl, 3.5 mM MgCl2, 3.75 μg mL−1 bovine serum albumin (BSA), 200 μM dNTPs, 5% dimethyl sulfoxide (DMSO), and 0.2X Titanium Taq polymerase (BD Biosciences, Sparks, MD). Genotyping reactions were performed using a Lightcycler 480 II real time PCR instrument (Roche Applied Sciences, Indianapolis, IN) with the following PCR parameters: 95°C for 5 min followed by 40 cycles of 95°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec, and a melting curve from 48 to 72°C. Fluorescence was read every 0.1°C during the melting curve analysis.

RESULTS AND DISCUSSION

Sequence Analysis of GmFATB1a in A22

Cardinal et al. (2007) concluded that a naturally occurring deletion mutation of the GmFATB1a gene was the molecular basis of the fap3n allele that originated from germplasm line N79-2077-12. Given the allelic nature of fap3(A22) and fap3n, GmFATB1a was considered the prime target of the mutagen that generated the fap3(A22) allele. To determine whether any deleterious mutations could be found in the GmFATB1a gene from A22, DNA sequence analysis of the full-length cDNA was conducted using messenger RNA (mRNA) isolated from developing seeds. Comparison of the GmFATB1a cDNA sequence from A22 with the sequence previously obtained from the normal-palmitate cultivar Century (Cardinal et al., 2007) revealed only one polymorphism that would alter the predicted amino acid sequence of the enzyme. The G to T substitution at nucleotide position 693 (with respect to the initiation codon) of the cDNA changed a Trp residue at amino acid position 231 to a Leu. Homology searches using the BLASTP algorithm (Altschul et al., 1990) of the nonredundant protein database of GenBank (National Center for Biotechnology Information, 2010) with the predicted amino acid sequence of GmFATB1a from A22 revealed that every plant protein annotated as a 16:0-ACP TE possessed a Trp residue at amino acid position 231 to a Leu. Homology searches using the BLASTP algorithm (Altschul et al., 1990) of the nonredundant protein database of GenBank (National Center for Biotechnology Information, 2010) with the predicted amino acid sequence of GmFATB1a from A22 revealed that every plant protein annotated as a 16:0-ACP TE possessed a Trp residue at amino acid position 231 to a Leu. Homology searches using the BLASTP algorithm (Altschul et al., 1990) of the nonredundant protein database of GenBank (National Center for Biotechnology Information, 2010) with the predicted amino acid sequence of GmFATB1a from A22 revealed that every plant protein annotated as a 16:0-ACP TE possessed a Trp residue at amino acid position 231 to a Leu.
As shown in Fig. 1, the G to T mutation found in A22 is located at position 63 within the 312 bp PCR amplification product of FATB1a-312. Using the same PCR primers described by Cardinal et al. (2007), a 312 bp PCR product was amplified and sequenced from the parental lines Archer and A22, and from a single F4 plant from each of the 20 F2-derived lines that were classified as \textit{fap3} (A22) and each of the 21 lines classified as \textit{Fap3Fap3}. Figure 1 shows the DNA sequence of this 312 bp region from Archer and A22, as well as Williams 82 and Century that were used as additional reference genomes. In addition to the G63T SNP responsible for the Trp231Leu mutation, a cultivar-specific SNP that does not alter the protein sequence was observed at position 55. In Archer, a C nucleotide resides at this location, whereas in Century, Williams 82, and A22, a T nucleotide is found (Fig. 1).

DNA sequence analysis showed that the 20 F2-derived lines classified phenotypically as homozygous for the \textit{fap3} (A22) allele were homozygous for the T at position 63 within the FATB1a-312 fragment, as well as a C at position 55 that originated from A22. All of the 21 F2-derived lines classified phenotypically as homozygous for the \textit{Fap3} allele were homozygous for a G at position 63 and a T at position 55 inherited from Archer. This perfect correlation provided strong support for the hypothesis that the SNP giving rise to the Trp231Leu amino acid substitution in the \textit{GmFATB1a} gene originating from A22 is the causative basis for the \textit{fap3} (A22) allele.

**Molecular Marker Assay**

The direct DNA sequence analyses of the 312 bp PCR products of the FATB1a-312 marker were effective in establishing a correlation between the G63T SNP and the \textit{fap3} (A22) allele. This sequence information for the G63T SNP and its surrounding genomic region was used to develop a SimpleProbe molecular marker assay to differentiate the \textit{Fap3} and \textit{fap3} (A22) alleles. SimpleProbe assays are based on the dissociation kinetics of a fluorescent oligonucleotide probe that binds perfectly to a wild-type sequence but has a single mismatch base pair when bound to a mutant sequence, or vice versa (Gillman et al., 2009). The oligonucleotide probe changes from a fluorescent state when bound to the corresponding strand to a nonfluorescent state when unbound. The difference in the fluorescence as the probe dissociates from the mutant or wild-type sequence with increasing temperature allows for determination of what nucleotide is present at the polymorphic site.

The specific \textit{GmFATB1a} sequence that was the target of the designed SimpleProbe is shown in Fig. 1. In addition to being complementary to the G63T SNP associated with the \textit{fap3} (A22) locus, the SimpleProbe also spans the nearby C55T SNP observed in Archer. Therefore, the SimpleProbe would be expected to hybridize perfectly to DNAs isolated from cultivars such as Williams 82 and Century, have a single mismatch at position 55 of the FATB1a-312 fragment from cultivars such as Archer, and a single mismatch at position 63 using A22 DNA.
To establish the effectiveness of the SimpleProbe assay, genomic DNA preparations from A22, IA1025, Archer, and Williams 82 were evaluated. Polymerase chain reaction amplification products from A22 and IA1025 genomic DNA samples had a SimpleProbe melting temperature of 58°C, while Archer had a melting temperature of 62.5°C (Fig. 2). A heterozygous individual displayed melting curves at both 58 and 62.5°C. Figure 2 shows the expected melting curves for the three possible genotypes when the C55T polymorphism is present in addition to the G63T SNP. Due to the perfect complimentarity between the SimpleProbe and the Williams 82 sequence, assays conducted using this source of DNA had the highest melting temperature of 66.5°C. Figure 3 shows the observed melting curves for the three possible genotypes when comparing individuals with only the G63T SNP present. The observed temperature curves were consistent with that expected based on the dissociation kinetics of the probe that was designed over
the region. The intermediate temperature curve of Archer suggested that the mismatched base pair is nearer the 3’ end of the probe, which allowed it to bind to the sequence at a higher temperature than the \( fap3(A22) \) mutation.

In summary, a SNP was identified within the \( GmFATB1a \) gene that differentiated the \( Fap3 \) from the \( fap3(A22) \) alleles. A SimpleProbe marker was developed for the SNP that can be used to select individuals with the \( fap3(A22) \) allele in a breeding program for development of low-saturate cultivars.

**Acknowledgments**
The authors would like to thank Silvia Cianzio for growing and harvesting the F\(_1\) and F\(_2\) plants and Daniel N. Duvick for assistance with analysis of fatty ester contents of oil samples.

This journal paper of the Iowa Agric. and Home Econ. Exp. Stn., Ames, IA, Project No. 5103 was supported by the Hatch Act, State of Iowa, Iowa Soybean Association, United Soybean Board, and the Raymond F. Baker Center for Plant Breeding.

**References**


