Mapping Genes Encoding Microsomal Omega-6 Desaturase Enzymes and Their Cosegregation with QTL Affecting Oleate Content in Soybean.
Soybean [Glycine max (L.) Merr.] is an oilseed crop of major importance, providing much of the world's edible oil. According to the National Agricultural Statistics Service (www.nass.usda.gov; verified 18 Jan. 2008), soybean oilseeds produce 29.4% of the world's edible vegetable oil, while the United States represent 40.1% of the world's soybean production. Yet, soybean oil is lacking in some quality characteristics. Oil quality is determined by the relative composition of saturated and unsaturated fatty acids in seed triacylglycerols. The average fatty acid composition of soybean oil is 110 g kg\(^{-1}\) palmitic acid (16:0), 40 g kg\(^{-1}\) stearic acid (18:0), 230 g kg\(^{-1}\) oleic acid (18:1), 530 g kg\(^{-1}\) linoleic acid (18:2), and 90 g kg\(^{-1}\) linolenic acid (18:3) (Wilson, 2004). Increased concentration of the monounsaturated oleic acid would improve the nutritional value and oxidative stability of soybean oil. This would also reduce the need for hydrogenation that leads to undesirable trans fatty acids (Wilson, 2004). Thus, extensive research has been conducted to incorporate the high-oleic-acid trait into soybean germplasm using traditional plant breeding methods (Wilson et al., 1981; Burton et al., 1989, 2006).

Mapping Genes Encoding Microsomal \(\omega\)-6 Desaturase Enzymes and Their Cosegregation with QTL Affecting Oleate Content in Soybean


ABSTRACT

The microsomal \(\omega\)-6 desaturase enzymes, which catalyze the desaturation of oleic acid to linoleic acid during fatty acid biosynthesis, are encoded by the FAD2-1 and FAD2-2 genes in soybean [Glycine max (L.) Merr.]. Breeders aim to incorporate the high-oleate trait into soybean germplasm in order to improve the nutritional value and oxidative stability of soybean oil. The objectives of this study were to map the isoforms of the FAD2-1 and FAD2-2 genes and investigate the association of these genetic loci with the oleate phenotype in three populations segregating for oleate content. The populations were grown in replicated multi-environment field trials. According to linkage analysis conducted for two of the populations, FAD2-1A and FAD2-1B mapped on Linkage Groups O and I, respectively, while the closely linked FAD2-2A and FAD2-2B isoforms mapped on Linkage Group L. Oleate quantitative trait loci with minor effects were detected in the proximity of FAD2-1B and possibly FAD2-2B on Linkage Groups I and L. Quantitative trait loci affecting maturity were also detected on chromosomal regions adjacent to the FAD2 genes. The genotyping assays developed for the FAD2-1A, FAD2-1B, and FAD2-2B isoforms, as well as their linked simple sequence repeat markers, can be used in soybean breeding programs for the elevation of oleic acid seed content through marker-assisted selection.
and genetic engineering approaches (Kinney, 1995; Kinney and Knowlton, 1998; Buhr et al., 2002).

Fatty acids are the precursors for the synthesis of triacylglycerols (TAGs), which constitute the storage lipid reserves of the developing oilseeds. Polyunsaturated fatty acids of the seed TAGs are synthesized primarily by ω-3 and ω-6 desaturase enzymes located in the endoplasmic reticulum; however, a small percentage is also derived from comparable enzymes localized in the plastids. The desaturase enzymes of the endoplasmic reticulum act on the fatty acids esterified to the first and second backbone position of phosphatidylcholine (PC), whereas the plastid-localized desaturases predominantly utilize galactolipids as substrates (Ohlrogge and Browse, 1995).

Plant fatty acid desaturase genes were initially elucidated in Arabidopsis. The desaturation of oleic to linoleic acid in the endoplasmic reticulum was shown to be controlled by the fad2 gene, encoding the oleyl-PC ω-6 desaturase activity (Miquel and Browse, 1992; Okuley et al., 1994). Subsequently, genes designated FAD2-1 and FAD2-2 were reported to encode comparable activities in soybean (Heppard et al., 1996). The FAD2-1 gene was expressed only during seed development and FAD2-2 was constitutively expressed during seed development and vegetative growth. Therefore, it was suggested that the seed-specific FAD2-1 is primarily responsible for the desaturation of oleic to linoleic acid for fatty acids destined for seed storage lipids, while FAD2-2 is involved in the desaturation of membrane lipids in vegetative tissues, playing a lesser role in storage oil biosynthesis. Moreover, the existence of at least two copies of each of the FAD2-1 and FAD2-2 genes was proposed (Heppard et al., 1996).

Recently, the existence of two distinct isoforms of the soybean FAD2-1 gene (designated FAD2-1A and FAD2-1B), encoding microsomal ω-6 desaturase enzymes that differ in stability at high temperatures, was confirmed (Tang et al., 2005). The FAD2-1A and FAD2-1B genes map on Linkage Groups O and I, respectively, in the soybean genome. Both FAD2-1A and FAD2-1B produce alternatively spliced transcripts that do not change the amino acid residues of the encoded ω-6 desaturase enzymes (Schlueter et al., 2007). For FAD2-2 genes, which also encode microsomal ω-6 desaturase enzymes, three isoforms have been reported, designated FAD2-2A, FAD2-2B, and FAD2-2C. FAD2-2A and FAD2-2B are localized on Linkage Group L (Schlueter et al., 2007).

Environmental factors affect the unsaturated fatty acid content of soybean oilseeds. Temperature exerts an effect on the unsaturated fatty acid composition during the pod-filling stage that coincides with oil deposition. Planting and maturity dates lead to changes in the fatty acid profiles due to different temperatures during oil deposition (Burton et al., 1983; Wilcox and Cavin, 1992). For example, the elevated-oleic lines N98-4445A and N97-3363-3, which were used in this study, are unstable across environments and their oleate variation is largely attributed to growth temperature differences (Oliva et al., 2006; Wilson et al., 2002). Although temperature effects during oil deposition were implicated in the altered fatty acid composition of soybean oilseeds, the expression patterns of the isoforms encoding the ω-6 desaturase enzymes, with the exception of FAD2-2C, showed no change in the level of transcript accumulation at different temperatures during seed development that could explain the changes in the levels of unsaturated fatty acids in seed triacylglycerols (Heppard et al., 1996; Schlueter et al., 2007). A recent study of the ω-6 desaturase enzymes encoded by the FAD2-1A and FAD2-1B genes, however, provided evidence of differential enzymatic stability at high temperatures (Tang et al., 2005). Apart from temperature effects, coordination of metabolite flux through the reactions of glycerolipid biosynthesis in the endoplasmic reticulum coupled with the de novo fatty acid biosynthesis pathway in the plastids underlie the complexity of the factors (both genetic and environmental) that can ultimately determine oleate composition of the storage oil.

Although the location of the isoforms encoding the ω-6 desaturase enzymes was recently revealed, there is minimal information regarding whether the isoforms of FAD2-1 and FAD2-2 genes cause the observed genotypic variation for oleate content in populations in which this trait is segregating. As an initial step toward unraveling the complexity of oleate biosynthesis and accumulation in soybean, the effects of the FAD2-1 and FAD2-2 isoforms (and their alleles), as well as their epistatic and environmental interactions, need to be investigated (Cardinal, 2008). The objectives of this study were to: (i) identify single nucleotide polymorphisms (SNPs) in the coding or upstream regions of the isoforms of FAD2-1 and FAD2-2 genes for the soybean lines under study; (ii) develop isoform- and allele-specific markers; (iii) map the individual isoforms of the FAD2-1 and FAD2-2 genes in the soybean genome; and (iv) investigate whether segregation at FAD2-1 and FAD2-2 loci is associated with changes in oleic acid seed content in three soybean populations segregating for oleate content.

MATERIALS AND METHODS
Development of Soybean Populations

Three soybean populations, designated as FAE, FAF, and FAS, were developed by single-seed descent (Brim, 1966). Population FAE consists of 721 F2-derived lines from the cross of N98-4445A × N97-3525. Population FAF represents 118 F2-derived lines from the cross of N97-3363-3 × PI423893. Population FAS consists of 231 F2-derived lines from the cross of N98-4445A × PI423893. The N98-4445A line is a high oleic (563.1 g kg⁻¹), low linoleic acid line developed by the USDA-ARS in Raleigh, NC (Burton et al., 2006). The N98-4445A line originated as a plant selection from the cross N94-2473 × (N9
The PI423893 line is a plant introduction with mid-oleic acid (305.8 g kg⁻¹) seed oil content and unknown genetic background (USDA-ARS National Plant Germplasm System, www.ars-grin.gov/npgs/searchgrin.html, verified 12 Jan. 2008). The N97-3525 line is the low palmitate, low linolenate cultivar Satelite, which was derived from the cross of ‘Soyola’ × (‘Brim’ (2) × [N88–431(2) × (N90–2013 × C1726)]) (Cardinal et al., 2007). The FAF and FAS populations are segregating for oleate genes and the reduced-linolenate fan(P123440) allele, while the FAE population is segregating for oleate genes and the fapₜ reduced-palmitate allele.

Experimental Design
The FAE and FAF populations were planted separately in sets-within-replications experimental designs with two replications in each location. Fifteen sets of seven rows by seven columns were randomly assigned to each replication, location, and year for the FAE population. Five sets of five rows by five columns were randomly assigned in each replication, location, and year for the FAF population. In each population, before lines were randomly assigned to each set, they were divided into five arbitrary groups that differed in maturity according to the lines’ maturity phenotypes in 2004. Each maturity group was represented at least once in each set of the FAE and FAF populations, and the maturity groups that included the majority of the segregating lines were represented two to three times in each set. Once lines were randomly assigned to a set, they remained in that set in all replications, locations, and years. Within each set, the maturity groups were randomly assigned to a column to facilitate mechanical harvest. Then the lines within each maturity group were randomly assigned to plots within each column. The parental lines N97-3363-3 and PI423893 in the FAF population were randomly assigned to plots within each column. The parent lines N97-3363-3 and PI423893 in the FAF population were randomly assigned to one set, to a maturity group within that set depending on their maturity date, and, finally, to a plot within the set, as explained above. The N98–4445A parental line of the FAE population was assigned to a maturity group in all the sets and randomly assigned a plot within each set as explained above. The FAE population was planted in four-row plots at Clinton, Kinston, and Plymouth, NC, in 2005 and 2006. The FAF population was grown at Clinton and Kinston in 2005 and at Clinton, Kinston, Clayton, and Plymouth, NC, in 2006. The FAE population was planted in four-row plots, with the exception of Clayton in 2006, which was planted in one-row plots.

The FAS population was planted in five-seed hills at Clayton, NC, in a sets-within-replications design with three replications in 2006. Sets were randomly assigned to each replication, and lines within each set were randomly assigned to hills. The parental lines, N98–4445A and PI423893, were randomly assigned within each of the five sets in each replication.

Phenotypic Evaluation
Maturity date was recorded for the FAE, FAF, and FAS populations at the R8 reproductive stage as days after planting (Fehr and Caviness, 1977). For the FAE and FAF populations, mature soybean seed was harvested mechanically from the two middle rows of each experimental plot or each single-row plot for the FAF population at Clayton in 2006. Approximately 10 g of seed was subsampled from each plot for the evaluation of fatty acid composition. For the FAS population, 5 g of seed was subsampled from the seed harvested from each hill. Fatty acid composition was evaluated by gas chromatography, as described by Burkey et al. (2007).

Statistical Analysis
Statistical analysis of the sets-within-replications design was conducted using Proc MIXED in SAS 9.1 (SAS Institute, 2004). Environments, sets, replications, and lines, as well as their interactions, were considered as random effects. Best linear unbiased predictors (BLUPs) were obtained for all traits and lines of the FAE, FAF, and FAS populations. The degrees of freedom for FAF and FAS populations were calculated using the Satterthwaite approximation, while for the FAE population they were derived with the containment method to reduce the computing power required. Analysis of the FAE and FAS populations was conducted separately for each environment, as well as combined across all the environments for fatty acid traits and maturity date. The phenotypic data of the FAF population from Clinton in 2005 were discarded due to numerous missing data points and heterogeneity of variance in comparison with the other environments. The analysis of the FAS population for a single environment was performed as described above.

Simple sequence repeat (SSR) markers and allelic-specific markers were tested for association with fatty acid phenotypes, derived as BLUPs, in FAE, FAF, and FAS populations using single-factor analysis in Proc GLM of SAS 9.1. The additive genetic effects were obtained by the CONTRAST statement in Proc GLM of SAS 9.1. Dominance genetic effects were also estimated for the FAS population. The means of the homozygous genotypic classes for each trait in the FAE and FAF populations or the homozygous and heterozygous genotypes in the FAS population, as well as their standard errors, were obtained with the LSMEANS statement in Proc GLM of SAS 9.1. The difference of the means of the homozygous genotypic classes, as well as its standard error and its significance, were determined with the ESTIMATE statement in Proc GLM of SAS 9.1. To investigate whether maturity date affected the associations between fatty acid traits and molecular markers, single-factor analysis was also conducted, with maturity date included as a covariate in the model.

Genotypic Evaluation and Linkage Mapping
Genomic DNA was extracted from leaf tissue of approximately 5 to 10 plants for each of the lines of the FAF and FAS populations collected from Kinston (2005) and Clayton (2004), respectively. The DNA isolation was conducted in a 96-well plate format with the Genta PureGene DNA purification kit (Genta Systems, Minneapolis, MN). Genomic DNA was isolated from leaf tissue (Kem et al., 1988) of each line of the FAS population from all the plants in each hill grown at Clayton (2006). One hundred forty-seven polymorphic SSR markers of the consensus linkage map (Cregan et al., 1999; Song et al., 2004) were genotyped across the 20 linkage groups of the soybean genome in the FAF population. Only SSR markers flanking the genomic regions of interest, as discussed below, were genotyped in the FAE and FAS populations (Table 1). The reactions were performed in a 384-well PTC-200 Peltier Thermal Cycler (MJ Research, Waltham, MA). Thermocycling conditions were 95°C for 2 min and 38 cycles of 92°C for 1 min, 49°C for 1 min, and 68°C for 1 min and 30 s. The amplifica-
tion products were resolved on 4% super-fine resolution agarose gels (Amresco, Solon, OH) with ethidium bromide staining in 1x Tris-borate-EDTA buffer, or 6.5% polyacrylamide gels in a LICOR 4300 DNA analysis system (LICOR Biosciences, Lincoln, NE) using M13-tailed unmodified primers and IRD-700 or IRD-800 labeled M13 oligonucleotides.

Linkage analysis was performed with JoinMap 3.0 (Van Ooijen and Voorrips, 2001). The Kosambi’s mapping function was used for linkage map construction and linkage was declared with a minimum recombination frequency of 0.4 and a maximum logarithm of odds (LOD) score of 3.0. Linkage analysis was also conducted with MAPMAKER/EXP 3.0 (Lander et al., 1987), as previously described by Cardinal et al. (2001), to verify the insufficient linkage declared in JoinMap 3.0 and the differential placement of FAD2-2B in FAF and FAS populations, as explained below. Linkage analysis with MAPMAKER/EXP 3.0 was performed using the same LOD and recombination thresholds as linkage analysis with JoinMap 3.0; however, the heterozygote genotypes were discarded in order for the F₃-derived lines of the FAF population to be analyzed as recombinant inbred lines, while the F₃-derived lines of the FAS population were actually analyzed as F₂ lines. Therefore, linkage mapping results from both analyses are discussed here, but only the linkage analysis with JoinMap 3.0 is presented.

Database Search and Polymerase Chain Reaction Primer Design
The coding regions of FAD2-1A and FAD2-1B were initially amplified with isoform-specific primers designed from the GenBank (Benson et al., 2002) accessions AB188250 and AB188251, respectively. For the isoforms of FAD2-2, in silico analysis was conducted on the soybean expressed sequence tags (Shoemaker et al., 2002) deposited in GenBank using BLASTN searches. Three distinct isoforms were predicted and isoform-specific primers were designed. The specific primers that amplify the coding region of each isoform are: FAD2-1A_d_F (5′-GTG TGG CCA AAC TGG AAG TT-3′) and FAD2-1A_b_R (5′-CAA AGC TCC CCTT CAG CCA GT-3′); FAD2-1B_a_F (5′-GTG TGG CCA AAC TGG AAA T-3′) and FAD2-1B_b_R (5′-AAG AAC CCC TCG CCA AAC AAT-3′); and FAD2-2B_c_F (5′-CCT CAG TCT AGT CTC AGC CAG ATT-3′) and FAD2-2B_c_R (5′-TGG GCC TTA TTG CCT TTG TC-3′). The specific primers that amplify the upstream region of the FAD2-2B isoform are: FAD2-2B_up_1_F (5′-ACG CAG TGA ATC AAA TGA CAA-3′) and FAD2-2B_up_2_R (5′-CTT CCTT TAT CTG GGC AGT AAT CT-3′). Isoform and primer designations for the isoforms of FAD2-1 genes follow the nomenclature of Schlueter et al. (2007), while the isoforms of FAD2-2 genes follow the nomenclature of Tang et al. (2005), respectively. The amplified fragments of FAD2-1B and FAD2-2B isoforms for CAPS analysis were obtained with the following Table 1. Molecular assays designed for the isoforms FAD2-1A, FAD2-1B, and FAD2-2B as well as simple sequence repeat markers mapping on Linkage Groups I, O, and L that were tested for associations with the unsaturated fatty acids' content in the FAF (N97-3363-3 × PI423893), FAE (N98-4445A × N97-3525), and FAS (N98-4445A × PI423893) populations.

<table>
<thead>
<tr>
<th>Linkage group</th>
<th>FAF</th>
<th>FAE</th>
<th>FAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAD2-1B CAPS</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>satt571</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>satt496</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>satt354</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>sat_268</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sat_418</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sat_324</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sat_299</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group L</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAD2-2B CAPS</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>satt495</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>satt652</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sat_320</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>satt462</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>sat_340</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sat1006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group O</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAD2-1A ASPE/SBE</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>satt420</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>satt478</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>satt153</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>sat_108</td>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1CAPS, cleaved amplified polymorphic sequence; ASPE, allele specific primer extension; SBE, single-base extension.

Sequencing of FAD2 Genes, Single Nucleotide Polymorphism Detection, and Allele-Specific Marker Development
Genomic DNA was isolated from bulked leaf tissue (Keim et al., 1988) from approximately 10 plants of the N98-4445A, N97-3363-3, and PI423893 lines and Brim. The amplification of the various FAD2-1 and FAD2-2 genes was performed in a MyCycler thermal cycler (Bio-Rad, Hercules, CA). The polymerase chain reactions (PCRs) were performed using the Expand High Fidelity PCR system (Roche Applied Science, Indianapolis, IN). Each 50-μL reaction contained 0.8 U polymerase, 20 pmol of each of the forward and reverse primers, 187 μmol L⁻¹ deoxynucleotide triphosphates (dNTPs), and 100 ng of genomic template DNA. The cycling program included a 5-min step at 94°C, 30 amplification cycles of 30-s denaturation at 94°C, annealing at 55°C for 30 s, and extension at 72°C for 2 min, followed by an extension step for 7 min at 72°C. The amplification products were cloned into pCR 2.1 cloning vectors (Invitrogen, Carlsbad, CA) and sequenced (Iowa State DNA facility, Ames, IA). Pairwise alignments of the sequences obtained for the isoforms of FAD2-1 and FAD2-2 genes were performed for the detection of SNPs among the soybean lines N98-4445A, N97-3363-3, PI423893, and Brim. Single nucleotide polymorphisms leading to restriction site polymorphisms were identified, and cleaved amplified polymorphic sequence (CAPS) markers were developed for FAD2-1B and FAD2-2B using the HpyCH4III and Aul endonucleases, respectively.

The amplified fragments of FAD2-1B and FAD2-2B isoforms for CAPS analysis were obtained with the following...
primes: FAD2-1B_c_F (5’-ATC CAA AGT TGC ATG GTA CAC-3’), FAD2-1B_b_R (5’-AAG CAC CCC TCA GCC AAT-3’), FAD2-2B_UP_1_F (5’-ACG CAG TGA ATC AAA TGA CAA-3’), and FAD2-2B_UP_1_R (5’-TGC TAT CGT TAC CCA CAT ACG A-3’). Cycling conditions were the same as described above. Amplification of genomic DNA was conducted with a 12-μL PCR reaction containing 6 pmol each of the forward and reverse primers, 0.5 U Taq polymerase (New England Biolabs, Ipswich, MA), 10X buffer (10 mmol L⁻¹ Tris-HCl, 50 mmol L⁻¹ KCl, 1.5 mmol L⁻¹ MgCl₂, pH 8.3), 234 μmol L⁻¹ dNTPs, and 75 ng genomic DNA. Next, 10 μL of the amplified product was digested in a 22-μL reaction at 37°C for 2 h. For the digestion of FAD2-1B fragments, the reaction contained 2.5 U HpyCH4II, buffer (20 mmol L⁻¹ Tris-acetate, 50 mmol L⁻¹ potassium acetate, 10 mmol L⁻¹ magnesium acetate, 1 mmol L⁻¹ dithiothreitol, pH 7.9) and 5 μg bovine serum albumin (BSA). The FAD2-2B amplification products were digested in reactions containing of 5 U Alul, buffer (10 mmol L⁻¹ Tris-HCl, 50 mmol L⁻¹ NaCl, 10 mmol L⁻¹ MgCl₂, 1 mmol L⁻¹ dithiothreitol, pH 7.9) and 5 μg BSA. Restriction fragments were resolved in 4% superfine-resolution agarose gels, as described above.

In the FAF population, the FAD2-1A isoform was genotyped with the allele-specific primer extension (ASPE) assay (University of California-Davis Genome Center, Davis, CA) in a Luminex platform (Luminex Corp., Austin, TX) using the primers FAD2-1A_ASPE1 (5’-TCA ATT ACT TCA CTT TAA TTC TTT gcc aaa gtt gag gta cag g-3’) for N97-3363-3 and FAD2-1A_ASPE2 (5’-TCA TTC ATA TAC ATA CCA ATT CAT gcc aaa gtt gag gta cag c-3’) for PI423893. The primers include the SNP-specific sequence (lowercase letters) coupled to the oligonucleotides associated with the fluorescent microspheres (upercase letters). In the FAS population, FAD2-1A was genotyped with the single-base extension (SBE) assay in a Luminex platform (Luminex Corp., Austin, TX) according to Chen et al. (2000) (data not shown).

RESULTS AND DISCUSSION
Identification of the Isoforms of the FAD2-2 Gene
A database query indicated that there are three isoforms for the FAD2-2 gene represented within GenBank accessions AB188253 and AC166091 (data not shown). The AC166091 accession corresponds to a bacterial artificial chromosome (BAC) containing two independent FAD2-2 genes. Recently, Schlueter et al. (2007) designated the isoform in AB188253 (also found in a BAC corresponding to accession no. AC166742) as FAD2-2C and the two isoforms in AC166091 as FAD2-2A and FAD2-2B. Because FAD2-2A and FAD2-2B are located on the same BAC, they should map to the same chromosomal region in the soybean genome. Since FAD2-2A has an approximately 100-base pair (bp) deletion within the coding region of the enzyme, it probably encodes a nonfunctional enzyme. Therefore, FAD2-2A was not considered a candidate gene that could explain differences in the ω-6 desaturase activity in our populations and was not studied further. Indeed, transcript analysis of FAD2-2A recently provided evidence that this isoform may not be expressed (Schlueter et al., 2007). Also, gene-specific primers designed for FAD2-2C amplified only part of the isoform’s coding region, which did not reveal any SNPs (data not shown) among the genotypes used in this study. Consequently, further information for the FAD2-2C isoform could not be obtained.

Single Nucleotide Polymorphism Detection
The isoform-specific primers were used for the amplification of FAD2-1A, FAD2-1B, and FAD2-2B from genomic DNA of lines N98-4445A, N97-3363-3, and PI423893. Amplification products were cloned, sequenced, and aligned for the parental soybean lines of interest and the cultivar Brim, leading to the detection of SNPs in coding or upstream regions of each isoform. One SNP originating from PI423893 results in the change of the 22nd amino acid residue of the predicted protein of FAD2-1A from glycine to arginine due to a substitution of a guanine with a cytosine in the first position of the codon. This is one of the four amino acid residues that differ between the FAD2-1A and FAD2-1B enzymes in the N-terminal region. Interestingly, the location of the SNP in the coding region of FAD2-1A coincides with a region responsible for the instability of the enzyme at high temperatures (Tang et al., 2005).

For FAD2-1B, a total of five SNPs were detected in the coding region, two of which result in nonsynonymous amino acid changes. The substitution of a cysteine with a thymine in the second position of the 86th codon of FAD2-1B causes an amino acid change from serine to phenylalanine in PI423893. Also, the 126th amino acid residue of the FAD2-1B enzyme in N98-4445A and N97-3363-3 changes from valine to methionine due to the substitution of a guanine with an adenine at the first position of this codon. Three additional nucleotide polymorphisms that do not change the amino acid residues of the predicted FAD2-1B enzyme were also detected in PI423893 (Fig. 1). These are located in the third positions of codons 219, 223, and 228. All nonsynonymous SNPs in FAD2-1A and FAD2-1B were evaluated for potential implications in protein function using the SIFT program (Ng and Henikoff, 2003). None of the nonsynonymous amino acid changes, however, were predicted to be deleterious.

Amplification and sequencing of the coding regions of FAD2-2B revealed no polymorphisms among the soybean lines of interest. Therefore, primers were designed to amplify a region immediately upstream of FAD2-2B. One SNP was detected 641 bp upstream of the putative start codon of FAD2-2B, corresponding to a polymorphism of a thymine vs. cytosine.
Development of Allele-Specific Molecular Markers for FAD2-1B and FAD2-2B

The detected nucleotide substitutions were examined for possible restriction site polymorphisms coinciding with the location of the SNPs to develop molecular assays that could readily distinguish among the alleles found in N98-4445A, N97-3363-3, PI423893, and ‘Brim’. The substitution of a thymine with a cytosine in the 223rd codon of FAD2-1B generates an additional restriction site for the HpyCH4III endonuclease in PI423893 (Position 172, Fig. 1), and thus enabled the development of a CAPS marker. Digestion of a 395-bp FAD2-1B-specific fragment from PI423893 generates fragments of 172, 90, 70, and 63 bp, whereas the same product derived from N98-4445A or N97-3363-3 CAPS yields fragments of 235, 90, and 70 bp (data not shown). The unique SNP in the upstream region of FAD2-2B results in the loss of an AluI restriction site in the fragment amplified from PI423893. The primers used for CAPS analysis amplify a 494-bp fragment in the upstream region of FAD2-2B. The AluI digestion of this amplification product from N98-4445A and N97-3363-3 generates fragments of size 173, 147, and 174 bp, in contrast to fragments of 320 and 174 bp when genomic DNA from PI423893 is used as a template (data not shown).

The SNP identified in the coding region of FAD2-1A did not alter restriction endonuclease cleavage sites. Therefore, an ASPE assay was designed with SNP-specific primers to distinguish between the N97-3363-3 and PI423893 FAD2-1A alleles in the FAS population. In the FAS population, an SBE assay was designed according to Chen et al. (2000) to distinguish between the N98-4445A and PI423893 FAD2-1A alleles using fluorescent microspheres in a Luminex platform. The segregation
patterns of the CAPS markers for FAD2-1B and FAD2-2B, as well as of the ASPE and SBE assays for FAD2-1A, were utilized for linkage analysis and association tests with fatty acid content in the FAF and FAS populations.

Mapping the Isoforms of the FAD2-1 and FAD2-2 Genes

Linkage analysis was conducted for 147 polymorphic SSR markers genotyped in the FAF population, which cover the 20 linkage groups of the soybean genome, as well as the CAPS, ASPE, and SBE assays with both JoinMap 3.0 (Van Ooijen and Voorrips, 2001) and MAPMAKER/EXP 3.0 (Lander et al., 1987). Since, for the linkage analysis performed with MAPMAKER/EXP 3.0, the recombination frequency among markers and the map distances were slightly biased by analyzing the FAF F_5-derived population as a recombinant inbred line population, only the linkage analysis performed with JoinMap 3.0 is presented here (Fig. 2). It should be noted that linkage could not be declared among all markers on the same linkage group, probably due to the relatively small size of the FAF population. The FAS population was used to validate the localization of the FAD2-1A, FAD2-1B, and FAD2-2B isoforms.

Mapping of FAD2-1A at the end of Linkage Group O near the SSR marker sat_108, and of FAD2-1B in the interval of sat_268 and satt354 on Linkage Group I, for both FAF and FAS populations, coincide with the locations proposed by Schlueter et al. (2007). The congruency in the localization of FAD2-1A and FAD2-1B verifies the efficiency of the developed assays and the efficacy of linkage analysis as a mapping tool. However, a discrepancy in the linkage analysis of FAF and FAS populations was revealed on Linkage Group L, both with JoinMap 3.0 and MAPMAKER/EXP 3.0. The FAD2-2A and FAD2-2B isoforms mapped in the interval of the SSR markers sat_340 and satt462 for the FAF population, while for the FAS population they were localized in the interval of sat_340 and satt006. Considering the difference in FAF and FAS population sizes, which implies the limited mapping precision of the FAF population, and the location proposed by Schlueter et al. (2007), FAD2-2A and FAD2-2B probably map within the interval of sat_340 and satt006.

Associations between Molecular Markers and Fatty Acid Content or Maturity Date

Overview of Cosegregation Analysis

Segregation of the FAD2-1A, FAD2-1B, and FAD2-2B allele-specific markers, as well as SSR markers in the vicinity of the isoforms on Linkage Groups O, I, and L, was used to test for association with the content of unsaturated fatty acids in the oil and with maturity date (Table 1). Fatty acids and maturity date BLUPs were obtained for each line from the analysis of each environment separately and from the combined analysis across all environments using mixed models in SAS 9.1. Best linear unbiased predictors for each trait were selected by maturity. Effects were significant by multiple comparisons.

Cosegregation Analysis in the Proximity of FAD2-1A, FAD2-1B, and FAF2-2B

On Linkage Group I, segregation of the FAD2-1B CAPS marker was associated with changes in oleate and linoleate content in the FAS population. The FAD2-1B CAPS marker explained 3.3 and 3.5% of oleate and linoleate genetic variation, respectively, when maturity date was included as a covariate (Table 3). The genetic
locus coincided with a maturity quantitative trait locus (QTL), which explained 6.0% of the genetic variation for maturity. The high-oleate and low-linoleate alleles were inherited from PI423893 and the difference between the means of the homozygous genotypic classes was −13.6 and 11.6 g kg\(^{-1}\), respectively (Table 3). The allele for late maturity was also inherited from PI423893, as expected from the positive correlation between maturity and oleic acid seed content in the FAS population (data not shown). The SSR markers satt354 and sat_268, located 4.91 and 6.07 cM from the 

\[ \text{FAD2-1B} \]

isoform, were also supportive of the minor oleate QTL in the proximity of FAD2-1B in both FAS and FAE populations (Tables 3 and 4). No significant association was detected on Linkage Group I for the 

\[ \text{FAD2-1B} \]

CAPS marker or the flanking SSR markers with fatty acid phenotypes in the FAF population. Thus, the minor oleate QTL in the vicinity of FAD2-1B was detected in the FAS and FAE populations but was not validated by the FAF population, which had the smallest population size.

Analysis of the FAE population indicated the existence of a minor oleate QTL in the proximity of the FAD2-2B isoform on Linkage Group L. The oleate QTL, which mapped on sat_340 (13.86 cM from FAD2-2B according to the linkage map of the FAS population) explained 2.2% of oleate genotypic variation and was detected in all six environments tested for the FAE population (data not shown). The high-oleate allele at this locus was inherited from the N98-4445A parental line and the means of the homozygous genotypic classes differed by 18.8 g kg\(^{-1}\) (Table 4). The minor oleate QTL on Linkage Group L was not detected in the FAS or FAE populations, possibly due to their smaller population size, which impacts the power of QTL detection and the magnitude of the QTL effect (Beavis, 1994). Quantitative trait loci with minor effects for oleic, linoleic, and linolenic acid traits have been reported on Linkage Group L at the genomic region where the FAD2-2A and FAD2-2B isoforms map (Hyten et al., 2004); however, the population used for that study did not segregate for major oleic acid genes, since it was derived from the cross of normal-oleate parental lines, and maturity effects were not taken into consideration.

On Linkage Group O, segregation of the FAD2-1A ASPE and SBE markers was not associated with changes in unsaturated fatty acid content in the FAF and FAS populations, which suggests the lack of oleate QTLs in the proximity of FAD2-1A.

Cosegregation Analysis for Other Genetic Loci Mapped on Linkage Groups O, I, and L

On Linkage Group O, analysis suggests a minor oleate QTL near the SSR markers sat_108 and satt153, which explained 2.4 and 0.6% of oleate genotypic variation in FAS and FAE populations, respectively (Tables 3 and 4). The satt153 locus coincided with a QTL for maturity explaining 8.9% of maturity genotypic variation in the FAE population. The late-maturity and low-oleate alleles were inherited from N97-3525, as expected by the negative genetic correlation between maturity and oleic acid seed content in the late-maturing FAE population (data not shown). After accounting for maturity effects in the model, however, it was determined that N97-3525 contributed the high-oleate allele (Table 4), which favors the hypothesis of two linked QTLs controlling maturity and oleate content near the SSR marker satt153.

Apart from the minor oleate QTL near FAD2-2B on Linkage Group L, minor QTLs for oleate content were also mapped on the SSR markers satt652 and satt462. The SSR marker satt652, which is located 65.47 cM away from the N98-4445A parental line and the means of molecular markers significance, as well as the R\(^2\) for the difference of the means for molecular markers significantly associated with maturity, oleate, linoleate, and linolenate content in the FAF (N97-3363-3 × PI423893) population.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Marker (linkage group)</th>
<th>AA(^{\dagger})</th>
<th>BB(^{\ddagger})</th>
<th>Difference</th>
<th>R(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R8(^{\dagger})</td>
<td>sat_340 (L)</td>
<td>115.3 (0.5)</td>
<td>113.1 (0.6)</td>
<td>2.2 (0.8)**</td>
<td>6.3</td>
</tr>
<tr>
<td>FAD2-2B (L)</td>
<td>115.6 (0.6)</td>
<td>112.6 (0.6)</td>
<td>3.0 (0.8)**</td>
<td>10.8</td>
<td></td>
</tr>
<tr>
<td>satt1006 (L)</td>
<td>115.3 (0.5)</td>
<td>112.4 (0.7)</td>
<td>2.9 (0.9)**</td>
<td>9.1</td>
<td></td>
</tr>
<tr>
<td>sat_268 (L)</td>
<td>113.6 (0.6)</td>
<td>115.4 (0.7)</td>
<td>−1.8 (0.9)***</td>
<td>3.7</td>
<td></td>
</tr>
</tbody>
</table>

\( \text{R8}^{\dagger} \) — Significant at the 0.05 probability level.

\( \text{**} \) Significant at the 0.01 probability level.

\( \text{***} \) Significant at the 0.001 probability level.

\( \text{AA}, \text{lines homozygous for N97-3363-3 allele; BB, lines homozygous for PI423893 allele.} \)

\( \text{R8, maturity (days after planting).} \)

\( \text{Genotypic classes scored as dominant for N97-3363-3 allele are italicized. The mean and standard deviation for the dominant genotypic class corresponds to the pool of homozygous and heterozygous genotypes.} \)

\( \text{Maturity used as a covariate during single-factor analysis.} \)
as well as in the combined analysis across all environments (Table 2). The oleate QTL near sat_652 on Linkage Group L was further confirmed with the FAE population, where sat_652 explained 0.5% of the genetic variation for the oleate trait (Table 4). The significant association between sat_652 and oleate content was consistent across two of the six environments when maturity was used as a covariate (data not shown). The high-oleate allele of the detected QTL was inherited from the N97–3363–3 and N98–4445A alleles in the FAF and FAE populations, respectively. In addition, minor QTLs for linoleate and linolenate traits were detected near sat_652 after accounting for maturity effects. These results confirm other oleate QTL mapping studies currently in progress (H.R. Boerma, personal communication, 2007). In the FAS population, sat_652 was not associated with oleate content but explained 1.6% of linolenate genetic variation (Table 3).

Also, on Linkage Group L, the SSR marker sat_462 explained 4.6% of the genotypic variation for the oleate trait in the FAF population (Table 2). The segregation of sat_462 was significantly associated with oleate content at the \( P < 0.05 \) level in four of the five environments tested (data not shown), as well as in the overall environment analysis (Table 2). The FAE population provided additional evidence for an oleate QTL in the proximity of sat_462. When maturity effects were taken into account, sat_462 explained 1.2% of the genetic variation for oleic acid content in the seed oil (Table 4). The associations declared for the sat_462 marker and oleate content in the FAE population were consistent across all six environments tested when maturity was used as covariate (data not shown). The high-oleate alleles of the QTL near sat_462 on Linkage Group L were inherited from the N97–3363–3 and the N98–4445A parental lines in the FAF and FAE populations, respectively. The FAS population did not confirm the association of oleic acid content with the SSR markers sat_340 and sat_462. In the FAF population, sat_462 explained 7.9% of the genotypic variation for oleic acid content in the seed oil (Table 2), the QTls conditioning linolenate content could not be validated in the FAF or FAS populations.

### CONCLUSIONS

Oleate QTLs with minor effects were detected in the proximity of the FAD2-1B and possibly FAD2-2B oleate desaturase genes. Thus, the results of this study suggest that no major oleate QTLs cosegregate with the FAD2-1A, FAD2-1B, and FAD2-2B isoforms of the genes encoding the \( \omega-6 \) desaturase enzymes. Therefore, the FAD2-1A, FAD2-2B, and FAD2-2B isoforms do not appear to encode for the major oleate genes that contribute to the elevated oleic

---

**Table 3.** The means and standard errors (in parentheses) of homozygous and heterozygous genotypic classes for maturity (R8), oleate (18:1), linoleate (18:2), or linolenate (18:3) content, the difference between means of homozygous genotypic classes, its standard error (in parentheses), and its significance, as well as the \( R^2 \) for the difference of the means for molecular markers significantly associated with maturity, oleate, linoleate, and linolenate content in the FAS (N98–4445A × PI423893) population.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Marker (linkage group)</th>
<th>AA†</th>
<th>BB†</th>
<th>HH†</th>
<th>Difference</th>
<th>( R^2 )</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>R8†</td>
<td>sat_340 (L)</td>
<td>102.3 (0.4)</td>
<td>99.8 (0.4)</td>
<td>100.9 (0.6)</td>
<td>2.5 (0.6)**</td>
<td>6.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAD2-2B (L)</td>
<td>102.2 (0.4)</td>
<td>100.0 (0.4)</td>
<td>2.2 (0.5)**</td>
<td>7.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sat_462 (L)</td>
<td>102.0 (0.4)</td>
<td>100.5 (0.4)</td>
<td>1.5 (0.6)**</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sat_652 (L)</td>
<td>102.0 (0.4)</td>
<td>101.0 (0.6)</td>
<td>1.6 (0.6)**</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sat_340 (I)</td>
<td>100.4 (0.4)</td>
<td>102.0 (0.4)</td>
<td>-1.8 (0.6)**</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAD2-2B (I)</td>
<td>100.1 (0.4)</td>
<td>102.0 (0.4)</td>
<td>-2.0 (0.5)**</td>
<td>6.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sat_268 (L)</td>
<td>100.2 (0.4)</td>
<td>102.3 (0.4)</td>
<td>100.8 (0.6)</td>
<td>-2.1 (0.6)**</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAD2-2B (L)</td>
<td>102.1 (0.4)</td>
<td>102.9 (0.4)</td>
<td>1.8 (0.6)**</td>
<td>3.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sat_340 (O)</td>
<td>136.3 (3.5)</td>
<td>138.4 (3.4)</td>
<td>135.1 (5.6)</td>
<td>-1.2 (5.2)**</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAD2-2B (O)</td>
<td>136.4 (3.3)</td>
<td>137.5 (3.3)</td>
<td>136.2 (5.6)</td>
<td>-1.3 (5.3)**</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sat_268 (O)</td>
<td>136.4 (3.4)</td>
<td>137.5 (3.4)</td>
<td>136.2 (5.6)</td>
<td>-1.3 (5.3)**</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAD2-2B (O)</td>
<td>136.4 (3.3)</td>
<td>137.5 (3.3)</td>
<td>136.2 (5.6)</td>
<td>-1.3 (5.3)**</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sat_340 (L)</td>
<td>466.0 (2.9)</td>
<td>453.1 (3.2)</td>
<td>457.8 (4.8)</td>
<td>13.6 (4.5)**</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAD2-1B (L)</td>
<td>466.1 (2.8)</td>
<td>445.2 (2.7)</td>
<td>12.2 (3.9)**</td>
<td>4.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sat_268 (L)</td>
<td>466.0 (2.9)</td>
<td>451.6 (3.2)</td>
<td>461.9 (4.6)</td>
<td>14.4 (4.3)**</td>
<td>4.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAD2-1B (L)</td>
<td>466.1 (2.9)</td>
<td>451.6 (3.2)</td>
<td>461.9 (4.6)</td>
<td>14.4 (4.3)**</td>
<td>4.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sat_340 (O)</td>
<td>55.9 (1.3)</td>
<td>59.5 (1.2)</td>
<td>55.3 (1.8)</td>
<td>-3.6 (1.8)**</td>
<td>1.6</td>
<td></td>
</tr>
</tbody>
</table>

---

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.
***Significant at the 0.001 probability level.
†AA, lines homozygous for N98–4445A allele; BB, lines homozygous for PI423893 allele; HH, lines heterozygous for N97–3363–3 and PI423893 alleles.
‡Maturity used as a covariate during single-factor analysis.
acid content of the N98-4445A and N97-3363-3 lines. Further research on the FAD2-2C isoform as well as other candidate genes implicated in oleate biosynthesis, such as the plastidial desaturase enzymes, is critical to substantiate the results presented here and shed light on the genetic factors that control oleic acid content in soybean oilseeds.

Acknowledgments
We would like to thank William Novitzky at the USDA Nitrogen Fixation Laboratory, Raleigh, NC, for providing training and equipment for the fatty acid analysis; Perry Cregan and Song Qijian at the USDA Soybean Genomics and Improvement Laboratory, Beltsville, MD, for conducting the SBE assay; as well as Carol Griffin for her technical expertise. This project was funded by the USDA Cooperative State Research, Education and Extension Service National Research Initiative, Plant Genome Program, and by the United Soybean Board.

References


