Putative Quantitative Trait Loci Associated with Calcium Content in Soybean Seed

BO ZHANG, PENGYIN CHEN, AINONG SHI, ANFU HOU, TETSUAKI ISHIBASHI, AND DECHUN WANG

From the Department of Crop, Soil, and Environmental Sciences, 115 Plant Science Building, University of Arkansas, Fayetteville, AR 72701 (Zhang, Chen, Shi, Hou, and Ishibashi) and the Department of Crop and Soil Sciences, A384-E Plant and Soil Science Building, Michigan State University, East Lansing, MI 48824 (Wang). A. Hou is now at the Agriculture and Agri-Food Canada, Unit 100-101 Route 100, Morden, Manitoba R6M 1Y5, Canada.

Address correspondence to P. Chen at the address above, or e-mail: pchen@uark.edu.

Seed calcium content is an important quality attribute of specialty soybean [Glycine max (L.) Merr.] for soyfoods. However, analyzing seed for calcium content is time consuming and labor intensive. Knowing quantitative trait loci (QTL) for seed calcium will facilitate the development of elite cultivars with proper calcium content through marker-assisted selection (MAS). The objective of this study was to identify major QTL associated with calcium content in soybean seed. Calcium content was tested in 178 F2:3 and 157 F2:4 lines derived from the cross of SS-516 (low calcium) × Camp (high calcium). The F2:3 lines were genotyped with 148 simple sequence repeat markers in a previous study on seed hardness, and the genotypic data were used in the QTL analysis of the current study. Four QTL designated as Ca1, Ca2, Ca3, and Ca4 on linkage groups (LGs) A2, I, and M were identified by both single-marker analysis and composite-interval mapping, and the QTL accounted for 10.7%, 16.3%, 14.9%, and 9.7% of calcium content variation, respectively. In addition, multiple-interval mapping analysis revealed a significant dominant-by-dominant interaction effect between Ca1 and Ca3, which accounted for 4.3% calcium content variation. These QTL will facilitate the implementation of MAS for calcium content in soybean-breeding programs.

Key words: calcium, quantitative trait loci, soybean (Glycine max)

Soyfoods have been consumed for more than 1000 years in Asia and are getting popular in Western diet because of their nutritional values and health benefits (Gollitz 1995). Soybean sprouts and natto (a fermented soyfood) are generally made using small-seeded soybean. However, hard or stone seeds, which are the seeds that do not absorb water during the soaking process, cause problems for soybean sprouting and natto manufacturing. For example, yield and quality of soybean sprouts are reduced because stone seeds decrease germination rates and seedling vigor (Mullin and Xu 2001). Likewise, seeds with nonuniform water absorption affect the fermentation process and ultimate texture of natto. Additional cost will occur if stone seeds have to be removed (Mullin and Xu 2001). Texture is another factor affecting quality attribute and consumer acceptability of soyfoods; seed hardness is an essential measurement and a good indicator for texture (Trumbull 1992).

Calcium content in the seed coat was found to be positively correlated with water absorption (Saio et al. 1973; Saio 1976). High calcium content appears to be the main cause of stone seeds. In texture studies, calcium content in soybean seed was found positively correlated with seed hardness, but this correlation was not always consistent because of environmental effects such as soil type and temperature variations (Chen et al. 1993, 2001).

The stone seed problem can be partially alleviated by using improved varieties with low calcium content. Finding quantitative trait loci (QTL) and associated molecular markers for soybean seed calcium content will help accelerate the selection process and improve the breeding efficiency. Recent developments in molecular marker techniques have allowed for the identification of many QTL underlying the seed quality of soybean. A QTL related to protein, oil, and yield was located close to randomly amplified polymorphic DNA marker OPASW13a on LG I (Chung et al. 2003). A very strong QTL associated with protein content was also identified between restriction fragment length polymorphism (markers A407 and A144 on LG I (Diers et al. 1992; Mansur et al. 1993; Sebolt et al. 2000; Chung et al. 2003). Seven significant marker loci were found to account for 53% of the total sucrose variation, and some of the loci were significantly associated with an increase of sucrose and oil content but a decrease of protein content (Maughan et al. 2000). In another study, simple sequence repeat (SSR) markers, Satt002 and Satt184, were
verified, and a novel marker, Satt147, has been detected to be associated with seed size (Panthee et al. 2005).

Studies on several plant species have explored the putative QTL for calcium to increase calcium uptake from seeds. In common bean (*Phaseolus vulgaris*), an F_{2:3} population derived from a cross between a wild and a cultivated common bean was evaluated for calcium QTL identification (Guzman-Maldonado et al. 2003). Two QTL with amplified fragment length polymorphism markers were found to explain about 25% of the calcium content variation. Calcium uptake is an important parameter of salt tolerance in *Helianthus annuus*. Three QTL for calcium uptake were identified accounting for 51.8% of the phenotypic variation, and 3 survivorship QTL were found to explain 78% of the calcium uptake (Lexer et al. 2003). Calcium content as one of the important mineral components for seed quality was studied in *Arabidopsis thaliana*, and 5 QTL were found to explain 36.4% of the variation (Vreugdenhil et al. 2004).

Little information on QTL analysis for natto traits is available. Keim et al. (1990) studied progenies from a cross *Glycine max* × *Glycine soja* and reported that 5 independent genomic regions contained putative QTL for seed-coat hardness (stone seeds), and these regions did not have equal genetic influence on seed-coat hardness. No QTL for calcium content have been mapped in soybean. It should be feasible to manipulate the calcium concentration in soybean seeds through molecular breeding if QTL for calcium are available. This study was performed to identify QTL associated with calcium content in soybean seed in a plant population derived from a low × high calcium cross.

**Materials and Methods**

**Plant Material**

Two natto soybean cultivars, SS-516 and Camp, were selected for their contrasting differences in seed calcium content and hardness. SS-516, released by the Southern States seed company (pedigree unavailable), contains 0.15–0.25% of calcium in seed. Camp, derived from the cross of Essex company (*pedigree unavailable*), contains 0.35–0.40% of calcium in seed. Camp, derived from the cross of Essex company (*pedigree unavailable*), contains 0.15–0.25% of calcium in seed. SS-516, released by the Southern States seed company (*pedigree unavailable*), contains 0.35–0.40% of calcium in seed.

Two natto soybean cultivars, SS-516 and Camp, were selected for their contrasting differences in seed calcium content and hardness. SS-516, released by the Southern States seed company (pedigree unavailable), contains 0.15–0.25% of calcium in seed. Camp, derived from the cross of Essex × an unknown *G. soja* and released by Virginia Polytechnic Institute and State University, Blacksburg, VA, contains 0.35–0.40% of calcium content in seed. The cross SS-516 × Camp was made in the greenhouse in spring 2002. Four F_1 plants were space planted in the field in summer 2002. Flower color and leaf shape were used as markers to verify the true hybrids. F_2 plants were grown in a winter nursery in Costa Rica and harvested individually to derive F_{2:3} lines. A total of 259 F_{2:3} lines were developed from the cross of SS-516 × Camp. One hundred and seventy-eight progeny lines with adequate amount of seed were randomly selected as a source population and were planted to produce F_{2:4} lines. In summer 2003 and 2004, 178 F_{2:3} and 157 F_{2:4} lines were evaluated in a field experiment with complete randomized block design with 2 replications at the University of Arkansas Agricultural Experiment Station, Fayetteville, AR, respectively. All lines were grown in single-row plot with 1.5 m in length and 0.95 m row spacing. Both SS-516 and Camp were included as checks in the test.

**Calcium Analysis**

Calcium content was determined by the HNO_3 method (Huang et al. 1985; Zarcinas et al. 1987; Campbell and Plank 1991). Briefly, 10 g of seeds from each plot were ground by a Knifetec 1095 sample mill (Foss Inc., Eden Prairie, MN), and 0.25 g of each ground sample was digested by 2.5 ml HNO_3. Samples were slowly heated to 60 °C for 45 min and gradually increased to 120 °C for 1 h after adding 1 ml H_2O_2. Cool samples were then mixed with deionized water, filtered through # 41 quantitative paper, and analyzed for calcium content using a Spectro Cirois with simultaneous inductively coupled plasma (Spectro Analytical Instruments, Inc., Mahwah, NJ).

**Genotyping with SSR Markers**

The F_{2:3} lines were genotyped with 148 SSR markers on 19 LGs on the consensus map, and the genotyping methods and results were reported in a previous study (Zhang et al. 2008). The genotypic data and linkage map constructed in the earlier study were used in the QTL analysis of this study.

**Calcium Content Data and QTL Analysis**

Phenotypic data on calcium content were analyzed using JMP 7.0 software (SAS Institute Inc., Cary, NC). Analysis of variance was used to assess the difference of calcium content among lines within each year and over 2 years (2003 and 2004). Shapiro—Wilk’s (W) test was used to test the populations for normal distribution. Broad-sense heritability of calcium content was estimated using $H^2 = \frac{\sigma^2_g}{\sigma^2_e + \left(\frac{\sigma^2_g}{\sigma^2_e}\right) + \left(\frac{\sigma^2_g}{\sigma^2_e}\right)}$ (Nyquist 1991), where $H^2$ is heritability, $\sigma^2_g$ is genotypic variance, $\sigma^2_g/\sigma^2_e$ is genotype × environment interaction variance, $\sigma^2_e$ is error variance, $\sigma^2_e/\sigma^2_g$ is number of replications, and $\sigma^2_e$ is number of environments, which is the number of years in this study.

The methods of single-marker analysis (SMA), composite-interval mapping (CIM), and multiple-interval mapping (MIM) implemented in WINDOWS QTL CARTOGRAPHER V2.5 (Wang et al. 2006) were used to identify QTL associated with calcium content. In SMA, $P < 0.0001$ was used as the threshold for significant markers. In the CIM analysis, the empirical significance threshold was determined by 1000 permutations with a walk speed of 1 cM and a significant level of 0.05. MIM analysis was used to estimate the optimum positions and effects of QTL as well as QTL interactions. The MIM model $c(n) = \ln(n)$ was selected with a walk speed of 1 cM. Mapchart (Voorrips 2002) was used to create the likelihood of odds (LOD) plots according to the data from QTL CARTOGRAPHER.

**Results**

**Seed Calcium Content Variation and Heritability**

SS-516 contained 0.21% calcium, whereas Camp had 0.38% calcium on average in 2003 and 2004. Calcium content
between 2 replications of 178 F2:3 lines and 157 F2:4 lines were not significantly different at \( a = 0.05 \), respectively. Therefore, average calcium content of each line was used to represent the phenotypic data of each population. Calcium content of 178 F2:3 lines ranged from 0.19% to 0.42% with an average of 0.297%. The calcium content of 157 F2:4 lines ranged from 0.26% to 0.39% with an average of 0.314% (Figure 1). The distribution of calcium content of both F2:3 lines and F2:4 lines was not different from a normal distribution with a \( P \) value of 0.20 and 0.73, respectively (Figure 2). The broad-sense heritability \( (H^2) \) estimate using variance components for calcium content in 2 years with 2 replications was 0.63.

Analysis of QTL

The genetic linkage map (148 polymorphic SSR markers on 19 LGs) previously constructed using genotypic data of 178 F2:3 lines derived from SS-516 × Camp (Zhang et al. 2008) combined with calcium content data sets was used to identify calcium QTL. Twenty-two markers on 8 LGs were found to be highly significantly associated with calcium content in the SMA \( (P < 0.0001) \) in 2003, and 11 of them were also significant in 2004 \( (P < 0.01, \text{Table 1}) \). Six markers on each of LGs A2 and I, 3 markers on LG L, 2 markers on each of LGs B1 and M, and 1 marker on each of LGs D2, F, and G were significantly associated with calcium content at \( P < 0.0001 \) in 2003. One marker on each of LGs D2, I, and M was significantly associated with calcium content at \( P < 0.001 \) in 2004.

In 2003, the LOD threshold for experimentwise \( P < 0.05 \) was 3.5 as determined by 1000 permutations. CIM analysis confirmed the presence of 3 major QTL \( (Ca_1, Ca_2, \text{and } Ca_3) \) on LGs A2, I, and M (Table 2, Figure 3). \( Ca_1 \) with LOD score of 4.3 was flanked by Satt228 and Sat_377 \( (P < 0.001) \) and explained 10.7% of phenotypic variation. \( Ca_2 \) with LOD score of 6.8 was flanked by Sat_174 and Satt354 and explained 16.3% of phenotypic variation. \( Ca_3 \) with LOD score of 4.5 was flanked by Satt677 and Sat_391 and explained 14.9% of phenotypic variation. In all cases, parent SS 516 contributed the positive alleles that significantly reduced seed calcium content. In 2004, the LOD threshold for experimentwise \( P < 0.05 \) was 3.8 as determined by 1000 permutations. CIM analysis only confirmed 1 QTL for calcium content \( (Ca_4) \) on LG M. \( Ca_4 \) flanked by Satt677 and Sat_391 had LOD score of 4.9 and explained 9.7% of phenotypic variation. Parent SS 516 contributed the desired allele that significantly reduced seed calcium content for this QTL. \( Ca_1 \) and \( Ca_2 \) were not confirmed in the F2:4 population, but they had the third (2.5) and second (2.9) highest LOD score among all the markers in the CIM. Sat_377, flanking marker of \( Ca_1 \), and Sat_174, flanking marker of \( Ca_2 \), were also significantly associated with calcium content in the F2:4 population \( (P < 0.001) \).

MIM analysis optimized the location of \( Ca_1, Ca_2, Ca_3, \text{and } Ca_4 \) and estimated interactions among \( Ca_1, Ca_2, \text{and } Ca_3 \) on

---

**Table 1.** SSR markers highly significantly associated with calcium content in F2:3 \( (P < 0.0001) \) and \( P \) value for the presence in F2:4 lines derived from soybean cross SS-516 × Camp

<table>
<thead>
<tr>
<th>LG</th>
<th>Marker</th>
<th>Position (cM)</th>
<th>( P ) value in F2:4 lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>Sat_382</td>
<td>0</td>
<td>0.112</td>
</tr>
<tr>
<td>A2</td>
<td>Satt429</td>
<td>3.087</td>
<td>0.11</td>
</tr>
<tr>
<td>A2</td>
<td>Satt228</td>
<td>14.433</td>
<td>0.006</td>
</tr>
<tr>
<td>A2</td>
<td>Sat_392</td>
<td>34.069</td>
<td>0.004</td>
</tr>
<tr>
<td>A2</td>
<td>Satt089</td>
<td>49.584</td>
<td>0.005</td>
</tr>
<tr>
<td>A2</td>
<td>Satt133</td>
<td>57.244</td>
<td>0.034</td>
</tr>
<tr>
<td>B1</td>
<td>Satt359</td>
<td>6.438</td>
<td>0.1</td>
</tr>
<tr>
<td>B1</td>
<td>Sat_095</td>
<td>27.2</td>
<td>0.088</td>
</tr>
<tr>
<td>B1</td>
<td>Sat_095</td>
<td>27.2</td>
<td>0.088</td>
</tr>
<tr>
<td>D2</td>
<td>Sctt_008</td>
<td>0</td>
<td>0.001</td>
</tr>
<tr>
<td>F</td>
<td>Satt252</td>
<td>43.908</td>
<td>0.045</td>
</tr>
<tr>
<td>G</td>
<td>Satt610</td>
<td>101.458</td>
<td>0.026</td>
</tr>
<tr>
<td>I</td>
<td>Satt496</td>
<td>0</td>
<td>0.023</td>
</tr>
<tr>
<td>I</td>
<td>Sat_174</td>
<td>3.141</td>
<td>0.0001</td>
</tr>
<tr>
<td>I</td>
<td>Satt354</td>
<td>17.048</td>
<td>0.004</td>
</tr>
<tr>
<td>I</td>
<td>Sat_170</td>
<td>36.737</td>
<td>0.03</td>
</tr>
<tr>
<td>I</td>
<td>Satt330</td>
<td>38.565</td>
<td>0.01</td>
</tr>
<tr>
<td>I</td>
<td>Satt292</td>
<td>40.039</td>
<td>0.011</td>
</tr>
<tr>
<td>L</td>
<td>Satt462</td>
<td>49.318</td>
<td>0.028</td>
</tr>
<tr>
<td>L</td>
<td>Satt229</td>
<td>72.316</td>
<td>0.005</td>
</tr>
<tr>
<td>L</td>
<td>Satt737</td>
<td>80.381</td>
<td>0.042</td>
</tr>
<tr>
<td>M</td>
<td>Satt150</td>
<td>66.025</td>
<td>0.009</td>
</tr>
<tr>
<td>M</td>
<td>Satt677</td>
<td>90.667</td>
<td>0.001</td>
</tr>
</tbody>
</table>
calcium content (Table 2). \textit{Ca1} was located at 16.4 cM on LG A2, 2 cM from Satt228, and 5.4 cM from Sat_377. \textit{Ca2} was located at 9.2 cM on LG I, 6.1 cM from Sat_174, and 7.9 cM from Satt354. \textit{Ca3} was located at 96.7 cM on LG M, 6.0 cM from Satt677, and 7.6 cM from Sat_391. \textit{Ca4} was located at 92.7 cM on LG M, 2.0 cM from Satt677, and 11.6 cM from Sat_391. There were no significant interactions between \textit{Ca2} and \textit{Ca1} or \textit{Ca3}. However, \textit{Ca1} and \textit{Ca3} had significant dominant-by-dominant interaction, which explained 4.3% of phenotypic variation.

### Discussion

Low calcium soybeans are desired by soyfood manufacturers for producing natto with acceptable texture. Genetic variation in calcium content is a prerequisite for breeders to excise selection in a breeding population. In this study, there were significant differences in calcium content among lines in the population derived from SS-516. In the population, we would have a better chance to confirm \textit{Ca1} and \textit{Ca2}. The 4 QTL identified in this study might be useful in MAS for proper calcium content in a specialty soybean-breeding program, but they need to be tested and confirmed in different genetic background and/or environment because calcium content was found to be significantly different among natto soybean lines (Cober et al. 1997). In a previous study, environment factors were found less important than genotypic effects on quality traits of food-grade soybean (Taira 1990). Similarly, it was reported that genotype × year × location interactions had no significant effects on calcium content (Cober et al. 1997). In this study, calcium content was higher in 2004 than in 2003 likely due to a drought stress. The effect of genotype × year interaction was also significant on calcium content. Chen et al. (2001) reported a 63% heritability of calcium content in soybean seeds, which is similar to the calcium heritability found in this study.

Seed size and hardness were reported to be associated with calcium content (Chen et al. 1993, 2001). Comparison of \textit{Ca1} on LG A2 and QTL, previously reported for seed size indicated that seed size and calcium content share a common genetic region. Sat_377 was significantly associated with calcium content and mapped on downstream of \textit{Ca1}. Satt470 (116.6 cM), associated with a QTL \((R^2 = 6\%)\) for seed size, was tightly linked to Sat_377 (116.7 cM) in the public map (Orf et al. 1999). Satt508 (108.9 cM), another QTL \((R^2 = 8\%)\) marker for seed size on LG A2, was also located between Satt228 and Sat_377 (Specht et al. 2001). In the summary of QTL for seed size (Hyten et al. 2004),

### Table 2. QTL associated with calcium content in soybean seed by CIM and MIM analyses

<table>
<thead>
<tr>
<th>QTL</th>
<th>LG</th>
<th>Optimum position on calcium genetic map</th>
<th>Flanking markers</th>
<th>Position of flanking markers</th>
<th>On calcium genetic map (cM)</th>
<th>On consensus map (cM)</th>
<th>LOD</th>
<th>R² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca1</td>
<td>A2</td>
<td>16.4</td>
<td>Satt228–Sat_377</td>
<td>14.4–21.8</td>
<td>154.1–116.6</td>
<td>4.3</td>
<td>10.7</td>
<td></td>
</tr>
<tr>
<td>Ca2</td>
<td>I</td>
<td>9.2</td>
<td>Sat_174–Satt354</td>
<td>3.1–17.1</td>
<td>36.6–46.2</td>
<td>6.8</td>
<td>16.3</td>
<td></td>
</tr>
<tr>
<td>Ca3</td>
<td>M</td>
<td>96.7</td>
<td>Satt677–Sat_391</td>
<td>90.7–104.3</td>
<td>75.6–1.0</td>
<td>4.5</td>
<td>14.9</td>
<td></td>
</tr>
<tr>
<td>Ca4</td>
<td>M</td>
<td>92.7</td>
<td>Satt677–Sat_391</td>
<td>90.7–104.3</td>
<td>75.6–1.0</td>
<td>4.9</td>
<td>9.7</td>
<td></td>
</tr>
<tr>
<td>Ca1 × Ca3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.3</td>
</tr>
</tbody>
</table>
10 QTL were mapped on LG L and 7 QTL were mapped on LGs F and G. In the present study, SMA showed that LGs L, F, and G contain 7–8 markers that are significantly associated with calcium content in 2003, 2004 or 2-year data combined. However, additional QTL may be identified and mapped on these LGs using populations with different genetic background than SS-516/C2 Camp.

Ca2 on LG I was mapped onto an interval flanked by Sat_174 (3.1 cM) and Satt354 (17.1 cM) in this study. Several QTL have been reported to be associated with soybean protein, oil, and yield within this region. There are 3 significant markers, Satt496 (36.4 cM), Sat_174 (36.6 cM), and Satt239 (36.9 cM) between Sat_174 (36.6 cM) and Satt354 (46.2 cM) on the public consensus map that have been identified as QTL markers for soybean protein, oil, and yield (Chung et al. 2003). Csanadi et al. (2001) also reported that the genomic region near Sat_174 contains a QTL for oil ($R^2 = 5.7$) and seed size ($R^2 = 11.6$). Sebolt et al. (2000) detected QTL for protein, oil, and seed size in the region where Satt562 was located, just about 13.7 cM upstream of Sat_174 on LG I. Seed size is generally correlated with calcium content and hardness (Chen et al. 1993, 2001).
Usually, large seeds contain less calcium content and are softer than small seeds. It was assumed that QTL for related seed traits were colocated in the 10- to 20-cM interval as a cluster because a single gene or several tightly linked genes may control the correlated seed traits (Khavin and Coe 1997; Sene et al. 2001).

LG M is an important LG for soybean where QTL for seed quality traits such as protein and seed size as well as agronomic traits such as maturity were mapped (Csanadi et al. 2001; Zhang et al. 2004). In our map, Satt677 was mapped at 90.7 cM on LG M, 6 cM above Ca3 (96.7 cM), and 2 cM above Ca4 (92.7 cM). QTL region (Satt306) for seed size was also near Satt677 on LG M in the public map (Csanadi et al. 2001). A study on 10 traits of soybean indicated that Satt150 was the flanking marker of QTL for maturity and yield (Zhang et al. 2004). Satt150 was significantly associated with calcium content in SMA ($P < 0.0001$ in 2003 and $P < 0.01$ in 2004) of our study, but its effect was decreased in CIM analysis. Maturity has been considered as an intersection for traits including yield, plant height, lodging, protein, oil, and seed size. Seed weight is one important attribute for yield. It is not surprising that QTL for seed size and yield are tightly linked or clustered together. The mineral content in seed affects the generative part of plants (Tyler and Zohlen 1998). Therefore, the genetic relationship between calcium and other traits could be explored by comparing QTL mapping results when available, and fine mapping of each trait would provide accurate genetic information. In practice, stable or verified QTL are required for MAS (Kearsey 1998). The QTL found in this study can be considered as preliminary work on MAS for low calcium content in natto soybean breeding. Research is under way to confirm identified QTL for calcium content in 2 additional populations with different genetic background.

**Funding**

This research was partially supported by a grant from the United States Department of Agriculture–National Research Initiative program.

**References**


Chen P, Buss GR, Dichl KC. 1993. Physical and chemical characteristics associated with hardness of small-seeded soybean for natto. ASA-CSSA-SSSA International Annual Meetings


Sene M, Thevenot C, Hoffmann D, Benetrix F, Causse M, Prioul JL. 2001. Identification of a cluster because a single gene or several tightly linked genes may control the correlated seed traits (Khavin and Coe 1997; Sene et al. 2001).


Received May 29, 2008; Revised September 26, 2008; Accepted October 2, 2008

Corresponding Editor: Reid Palmer