Theor Appl Genet (2014) 127:1991–2003
DOI 10.1007/s00122-014-2355-2

ORIGINAL PAPER

Identification of a single gene for seed coat impermeability in soybean PI 594619

Hirut Kebede · James R. Smith · Jeffery D. Ray

Received: 10 April 2014 / Accepted: 13 July 2014 / Published online: 8 August 2014

Abstract

Key message **Inheritance studies and molecular mapping identified a single dominant gene that conditions seed coat impermeability in soybean PI 594619.**

Abstract High temperatures during seed fill increase the occurrence of soybeans with impermeable seed coat, which is associated with non-uniform and delayed germination and emergence. This can be an issue in soybean production areas with excessively high-temperature environments. The objectives of the present study were to investigate the inheritance of impermeable seed coat under a high-temperature environment in the midsouthern United States and to map the gene(s) that affect this trait in a germplasm line with impermeable seed coat (PI 594619). Crosses were made between PI 594619 and an accession with permeable seed coat at Stoneville, MS in 2008. The parental lines and the segregating populations from reciprocal crosses were grown in Stoneville in 2009. Ninety-nine $F_{2:3}$ families and parents were also grown at Stoneville, MS in 2011. Seeds were assayed for percent impermeable seed coat using the standard germination test. Genetic analysis of the F_2 populations and $F_{2:3}$ families indicated that seed coat impermeability in PI 594619 is controlled by a single major gene, with impermeable seed coat being dominant to permeable seed coat. Molecular mapping positioned this gene on CHR 2 between markers Sat_202 and Satt459. The designation of *Isc* (impermeable seed coat) for this single gene has been approved by the Soybean Genetics Committee. Selection of the recessive form (*isc*) may be important in developing cultivars with permeable seed coat for high-heat production environments. The single-gene nature of impermeable seed coat may also have potential for being utilized in reducing seed damage caused by weathering and mold.

Introduction

Environmental stress during the reproductive growth of soybean [*Glycine max* (L.) Merr.] can reduce seed germination and seedling vigor (Gibson and Mullen 1996; Spears et al. 1997; Egli et al. 2005). Exposure of soybean plants to excessively high temperatures during seed fill can increase the presence of hard seed (impermeable seed coat) and wrinkled/shriveled seed, which can lower the quality of the seed (Gibson and Mullen 1996; Spears et al. 1997; Egli et al. 2005; Smith et al. 2008). Seed with hard or impermeable seed coats may look normal, but do not imbibe water. Therefore, seed lots with impermeable seed coats have reduced germination and, where emergence does occur, there is often reduced seedling vigor (Spears et al. 1997). The terms "hard seed" and "impermeable seed coat" have been used interchangeably for permeability of the soybean seed coat to water based on tests of seed germination or imbibition (Keim et al. 1990; Sakamoto et al. 2004; Watanabe et al. 2004; Liu et al. 2007). "Hard seededness" has also been used as a term for whole seed hardness in evaluating food-grade soybeans, where pressure-cooked samples were tested for hardness using a texture analyzer (Zhang et al. 2008). In the present study, seed coat impermeability (Hard Seed) was defined as "seeds that remain hard at the end of the prescribed test period because they have not absorbed water due to an impermeable seed coat" (Association of Official Seed Analysts 2006).

Earlier studies conducted in humid areas in the southern United States indicated that seeds with impermeable seed

Communicated by Brian Diers.

H. Kebede · J. R. Smith (✉) · J. D. Ray
USDA-ARS-CGRU, P.O. Box 345, Stoneville, MS 38776, USA
e-mail: rusty.smith@ars.usda.gov

coat might be beneficial for reducing field deterioration of soybean seed and for improving its storage potential (Potts et al. 1978). However, delayed and non-uniform germination and emergence are problems associated with planting impermeable soybean seed because the degree of seed coat permeability can be highly variable within a given lot of seed (Gibson and Mullen 1996; Keith and Delouche 1999). Hence, although impermeability may provide some protection against field weathering (Potts et al. 1978; Hartwig and Potts 1987), it may not be a reliable positive characteristic for producers.

High-quality seed is needed by producers to ensure adequate plant stands with reasonable seeding rates (Egli et al. 2005). This can be a concern for seed bean production in the tropics and sub-tropics. It may become an issue for soybean producers in the midsouthern United States because of the shift to the early soybean production system (ESPS). Under the previous production system, the dominant maturity groups (MG) planted were VI and VII cultivars, with these being planted in May and June. Seed bean production for these cultivars occurred in the southern regions where they were intended for production. However, the ESPS consists of the early planting (March and April) of "early" (MG IV and V) soybeans that usually begin to mature in August when temperatures and humidity are usually high (Tyler 1999; Heatherly 1999; Mengistu and Heatherly 2006; Gillen et al. 2012). Although the ESPS was adopted for its improved overall profitability, production of seed with impermeable seed coats lowers the value of the seed for seed beans, as well as for grain intended for specialty markets (food-grade soybeans), where hard seededness reduces the value of the product (Geater et al. 2000; Zhang et al. 2008). Therefore, producers of seed beans for the midsouthern United States have mostly avoided the issue of seed coat impermeability by producing their seed beans in northern states such as Illinois, where seed fields mature later under cooler temperatures. This avoidance resolves the immediate issue, but also reduces the earnings of southern seed producers. However, the problem is not so easily avoided in many seed production areas of the world, such as in South America, where seed production fields are located in the same area as are grain production fields. Additionally, as global temperatures rise, current seed production locales may be impacted.

When considering genetic/environmental interactions on seed coat permeability it is important to understand the development of the seed coat (testa), which develops from integumentary tissue (Esau 1965) in the mother plant. At the time of fertilization of the pistil, the inner integument consists of two to three cell layers and the outer integuments consists of two to four cell layers, except in the hilum, where it is thicker (Carlson and Lersten 2004). The two levels of integument (inner and outer) develop to envelop the embryo sac prior to fertilization, and then continue to develop into the mature seed coat as the embryo and cotyledons develop and mature inside it (Carlson and Lersten 2004). Ordinarily in seed plants, the level of impermeability depends on the genetics of the species as well as on that of the specific cultivar. However, environmental conditions during seed maturation are also important (Hartman and Kester 1975). In general, high temperatures during senescence increase seed coat impermeability. This is especially true for legumes on the whole (Hartman and Kester 1975) and for soybeans in particular (Egli et al. 2005). Therefore, as the mother soybean plant experiences high heat during seed fill and senescence, the seed coat it forms to surround the newly developing seed becomes increasingly impermeable. Hence, it is the genotype of the mother plant that interacts with the environment to affect the level of impermeability of its progeny's seed coat. As such, in this study we evaluated the progeny seed to assess the genotype of the mother plant. In a similar way, Keim et al. (1990) compared the phenotypes from bulked F_4 progeny seed with their corresponding F_2 genotypes in a genetic study on hard seededness.

Studies have been conducted to identify genes conditioning seed coat permeability and to characterize its gene action (Kilen and Hartwig 1978; Keim et al. 1990; Sakamoto et al. 2004; Watanabe et al. 2004; Liu et al. 2007). Kilen and Hartwig (1978) studied the inheritance of impermeable seed coat in soybean and suggested that as few as three major genes may control the permeable–impermeable response with additional minor genes contributing to impermeablility. Six genomic regions on chromosomes (CHR) 2, 3, 6, 8, 19 and 20 (Linkage groups, D1b, N, C2, A2, L and I, respectively) have been found to influence seed coat permeability in soybean in other studies (Keim et al. 1990; Sakamoto et al. 2004; Watanabe et al. 2004; Liu et al. 2007). Smith et al. (2008) reported high and low levels of seed coat impermeability in PIs 594619 and 587982A, respectively, when planted in mid-April and harvested in mid-August in Mississippi. As the genetics of seed coat impermeability have not been previously examined in PI 594619, the objectives of the present study were to study the inheritance of impermeable seed coat under high-temperature environments in the midsouthern United States and to map the gene/s that affect this trait in this PI.

Materials and methods

Plant material and seed assays

Glycine max accessions PI 587982A (MG III) (USDA 2014a) and PI 594619 (MG IV) (USDA 2014b) were obtained from the USDA-ARS Soybean Germplasm

Collection located at Urbana, IL. PI 587982A has a permeable seed coat, whereas PI 594619 has an impermeable seed coat (Plant Anatomy Ontology—Seed Coat, PO:0009088; Plant Trait Ontology—Seed Coat Hardness, TO:0000889; Soybean Trait Ontology, Seed Coat Hardness, SOY:0001415) as observed by Smith et al. (2008) under conditions of the ESPS in Mississippi. Crosses were made between these two accessions as PI 587982A × PI 594619 (POP1) and as a reciprocal cross PI 594619 × PI 587982A (POP2) at the Jamie Whitten Delta States Research Center, Stoneville, MS in 2008. Hybrid seed was harvested, mechanically scarified by making a single fleck in each seed coat with a small hand-held knife, and then sent to the Tropical Agriculture Research Station in Mayaguez, Puerto Rico for F_2 seed production at Isabela, Puerto Rico. Seed from individual F_1 plants was harvested separately and shipped back to Stoneville in 2009. The parental lines, cultivar Williams 82 (Bernard and Cremeens 1988), F_1 seed from POP2, and F_2 seed from POP1 and POP2 were planted at Stoneville, MS in a Sharkey clay soil (very-fine, smectitic, thermic Chromic Epiaquert) on May 18, 2009. Prior to planting, seed of PI 594619 and the F_1 and F_2 generations were mechanically scarified. Seed of the F_1 was scarified as indicated above, while seed of the F_2 and PI 594619 were scarified as follows. Seed was placed inside a hollow tin cylinder (13.5 cm high and 10 cm in diameter) with jagged-edged perforations (1–3 mm in diameter with approximately four perforations cm^{-2}) on the inside of the cylinder. The cylinder was hand rotated in a circular motion for 15 revolutions, thereby producing small flecks in the seed coats.

Single-row plots (2.74 m long with 0.91 m between rows) were used with a seeding rate of 25 seed m^{-1}. PIs 594619 and 587982A and Williams 82 were arranged in a randomized complete block design (RCBD) with three replications, with F_1 plants grown in one row, POP1-F_2 plants grown in seven rows, and POP2-F_2 plants grown in two rows. Water was applied as needed during the growing season by furrow irrigation through poly-pipe to alleviate moisture-deficit stress that could potentially affect seed quality (Heatherly 1993). Each F_1 and F_2 plant was individually tagged with a unique number after flowering (R1, Fehr and Caviness 1977) for the purpose of associating leaf tissue/DNA sampling with impermeable seed coat rating. Air temperature and relative humidity were measured at the official weather station at Stoneville. Within 1–4 days after R8 (Fehr and Caviness 1977), individual parental, F_1, and F_2 plants were harvested by hand. All plants were individually threshed in a single-plant belt thresher (Almaco BT14-E, Nevada, IA) so as to reduce mechanical damage. Diseased and mechanically damaged seeds were removed after threshing and not tested because these could affect the measures of seed coat permeability. Seed from each individual F_2 plant was bulked to form F_2-derived F_3 ($F_{2:3}$) families. Threshed seed was stored at 21 °C and 60 % relative humidity until they were assayed for impermeability during the winter of 2009–2010.

Fifty seeds from each harvested plant (parental, F_1, and F_2) with sufficient seed were assayed for percent impermeable seed coat as part of the protocol for the standard germination test. Specific details are provided by the Association of Official Seed Analysts (2006), but a brief summary is provided here. Seed from each plant were placed in water-moistened germination paper towels, which were then rolled up and placed vertically in plastic containers on a shelf in a germinator (Percival Scientific, Inc., Perry, IA). The towels were maintained at near 100 % humidity for the 7 days duration of the assay. Daily temperatures were maintained at 20 °C for 16 h and then adjusted to 30 °C for 8 h. At the conclusion of the 7 days, seed that had not imbibed water were counted as "hard seed," and were recorded as the percentage of seed out of 50 total seed that had not imbibed water. All assays were conducted by the State Seed Testing Laboratory, Mississippi State, MS.

One hundred and one $F_{2:3}$ families and parents of POP1 were planted into a Sharkey clay soil at Stoneville, MS on April 13, 2011. Prior to planting, seed of each family and of PI 594619 were mechanically scarified in the perforated tin cylinder as previously described. Each family and parental line was planted in two replications in a RCBD. Single-row plots (2.74 m long with 0.91 m between rows) were used as experimental units with a seeding rate of 9 seed m^{-1}. The lower seeding rate was used due to limited seed supply. A minimum of 100 seeds per F_2 plant was needed to phenotype the F_2 plant and still have sufficient seed to test its corresponding F_3 family in replicated plots. Water was applied as needed during the growing season by furrow irrigation through polypipe to alleviate potential moisture-deficit stress. Air temperature and relative humidity were measured at the official weather station at Stoneville. Within 1–4 days after R8, ten plants were harvested from each plot. Individual plants that were earlier or later than the group of 10 were not harvested. Each harvested plant from each plot was individually threshed, giving a total of 20 plants per family. Assays for seed coat impermeability were conducted on the seed from each individual of each family using 50 seeds per plant and following the official protocols for standard germination (Association of Official Seed Analysts 2006) and as described above. All assays were again conducted by the State Seed Testing Laboratory. Based on the parental distributions of seed coat impermeability, F_2 phenotypes derived directly from the seed of F_2 plants were grouped into two classifications; impermeable seed coat and permeable seed coat. Those estimated from $F_{2:3}$ families were classified as homozygous impermeable seed coat: heterozygous: homozygous permeable seed coat.

Genotyping with SSR and SNP markers

Leaf samples were collected from individual plants for F_2-POP1 and F_2-POP2. The samples were freeze-dried in a Model 2400 freeze dryer (The Freeze Dry Company, Nisswa, MN 56468, USA) and ground to a fine powder using a tissue pulverizer (Garcia Manufacturing, Visalia, CA 93292, USA). DNA was isolated from the samples using a Maxwell 16™ automated DNA isolation machine (Promega, Madison, WI 53711, USA) following the manufacturer's protocols.

Primers for the SSR markers were manufactured with either a hexachloro-fluorescein (HEX) or 6-carboxy-fluorescein (FAM) 5′-fluorescent label (Integrated DNA Technologies, Coralville, IA 52241, USA) based on sequences obtained from SoyBase (http://soybase.org/resources/ssr.php). About 600 SSR primers were screened for polymorphism in the parental lines. PCR amplification was performed on a MJ Research PTC-225 Thermal Cycler (Bio-Rad, Hercules, CA 94547, USA) using conditions of 95 °C for 120 s; 35 cycles of 94 °C for 30 s; 46 °C for 30 s; 72 °C for 30 s; and one cycle of 72 °C for 300 s followed by maintenance at 4 °C until detection. Amplicons were detected on an ABI 3730 (Applied Biosystems, Foster City, CA 94404, USA) at the USDA-ARS Midsouth Area Genomics Facility at Stoneville, MS. The lengths (bp) of amplicons were determined using GeneMapper 3.7 software (Applied Biosystems, Foster City, CA 94404, USA). All polymorphic SSR markers identified in this study were co-dominant. The marker data were coded numerically into three classes (1, 2 or 3), where 1 = the homozygous marker allele from Parent 1, 2 = marker alleles from both parents (i.e. heterozygotes), and 3 = the homozygous marker allele from Parent 2.

KBioScience KASP genotyping assay (www.kbioscience.co.uk) was used for SNP markers. Three hundred and twenty-eight allele-specific primers and other assay components were purchased from KBioscience (KBioscience Ltd, Hoddesdon, EN11 0EX, United Kingdom) based on the supplied marker sequences. PCR reactions were carried out for 15 µL final volume reactions. The cycling conditions were as follows: 94 °C for 15 min, 94 °C for 20 s, touchdown over 65–57 °C for 60 s (10 cycles dropping 0.8 °C each cycle) and an extra 30 cycles at 57 °C followed by maintenance at 4 °C until detection. SNP genotyping was performed using LightCycler® 480 System (Roche Applied Bioscience, http://www.roche-applied-science.com) based on endpoint analysis method.

Genetic map construction and linkage analysis

A linkage map was constructed using JoinMap® 4.0 (Van Ooijen 2006) on marker data collected from 89 F_2 individuals of POP1with extreme ratings for seed coat permeability (45 with permeable seed coat and 44 with maximum seed coat impermeability). Chi square analysis was performed for goodness-of-fit to the expected Mendelian segregation ratio for each marker and skewed markers were identified using a threshold of $P < 0.05$. Linkage groups were created by omitting the skewed markers from the data. A LOD (Log of odds) of 5.0 and recombination fraction <0.40 were used to create linkage groups. Recombination values were converted to genetic distances using the Kosambi mapping function. Marker order determined in this study was compared against the soybean consensus map (http://soybase.org/MarkerDB/index.php) to identify any major discrepancies and confirm position. This map was used for a preliminary QTL analysis across all linkage groups. For those chromosomes showing regions with significant associations in the initial analysis (CHR 2 and 7) the chromosomal linkage groups were reconstituted using the full 199 F_2 individuals of POP1 and the QTL analysis recomputed.

Single marker analysis (SMA) was conducted using the SMA routine (regression of individual markers and trait values) of JMP Genomics (JMP®, Version 6. SAS Institute Inc., Cary, NC, 1989–2007). The significance of putative associations is reported as the −log(probability) and a conservative −log(2.0) level was used as a threshold of significance. Based on the results of the SMA analysis on the 89 extreme F_2 individuals of POP1, marker data were generated on the additional 110 F_2 individuals of POP1 on the chromosomes which showed significant marker-trait associations. The complete data (199 F_2 plants) were then reanalyzed. Additionally, more markers were analyzed on chromosomes where putative loci were identified using the extreme phenotypes.

Quantitative trait loci (QTL) analysis was conducted using the linkage map described above. Based on the results of this preliminary screen and the putative loci identified by SMA, linkage groups of chromosomes with regions of interest were reconstituted using the full 199 member F_2 population. The QTL analysis consisted of Interval mapping (IM) followed by Multiple-QTL model Mapping (MQM) analysis using MapQTL6.0 software (Van Ooijen 2009). Interval mapping analysis was performed to find putative QTL and to select markers significantly associated with the seed coat impermeability trait. Genome-wide LOD thresholds were determined for the trait using the Permutation test of MAPQTL with 10,000 iterations. Based on the permutation tests, a threshold LOD value of 3.5 ($P < 0.05$) was used to declare presence of a QTL. Automatic cofactor selection was run for each linkage group. Markers located in the vicinity of putative loci associated with seed coat impermeability were selected as cofactors. MQM analysis was used to more precisely locate the loci associated with

the trait. The positions with the maximum LOD score on the linkage groups were considered to be the loci associated with seed coat impermeability.

To confirm the locus that was identified in POP1, which is associated with seed coat impermeability, marker data were generated on 52 F_2 individuals of POP2 (a reciprocal cross of POP1). SSR and SNP markers on the chromosome which showed significant marker-trait association in POP1 were used to conduct SMA and QTL analyses using the methods indicated above.

Marker analysis on the $F_{2:3}$ families was conducted using the F_2 marker data and the phenotype inferred from 99 $F_{2:3}$ family classifications (As indicated above, data were originally collected on 101 $F_{2:3}$ families. However, the classification of two families could not be resolved and they were dropped from the analysis). The 99 F_2 plants from which the 99 $F_{2:3}$ families were derived, were reclassified based on the results of the $F_{2:3}$ family phenotyping as homozygous permeable seed coat, heterozygous, or homozygous impermeable seed coat. SMA and QTL analyses were conducted using the reclassified phenotypic data and marker data for the 99 F_2 plants.

Results

Air temperatures during seed fill

As indicated above, high temperatures during seed fill in the midsouthern US can increase the presence of impermeable seed coat. During the 2009 R7–R8 stages (Fehr and Caviness 1977) of the plants of POP1 (PI 587982A × PI 594619), POP2 (PI 594619 × PI 587982A), and the parents, daily maximum air temperature ranged from 27 to 36 °C. The R8 stage for the permeable seed coat parent (PI 587982A) ranged from August 14 through September 1 with a mean R8 date of August 23, while that of the impermeable seed coat parent (PI 594619) ranged from August 25 through September 8th with a mean R8 date of August 31. The R8 stage for the F_2 individuals of POP1and POP2 ranged from August 21 through September 8th, and the mean R8 date for both populations was August 30th. The average maximum temperature during the 2011 R7–R8 stages for the $F_{2:3}$ families of POP1 and the parental lines ranged 28–41 °C.

Inheritance and mapping of POP1 (PI 587982A × PI 594619)

In 2009, the seed coat impermeability ratings for the 199 F_2 individual plants of POP1 ranged 0–98 % with an overall mean of 37.9 % (Table 1). The frequency distribution of the F_2 plants in POP1 is shown in Fig. 1b. In this test, the seed coat impermeability ratings for the permeable seed coat parent, PI 587982A, ranged 0–4 % with an average of 0.3 %, with only one plant with a score of 2 % and three plants with a score of 4 % out of 46 total plants (Fig. 1a). The impermeable seed coat parent, PI 594619, had ratings ranging from 50 to 94 % with an average of 80.6 % (30 plants). On the basis of the parental impermeability ratings, a binary system of classifying plants as having either permeable or impermeable seed coats was developed. The cut-off point between the permeable and impermeable seed coat classes was set at 4 % (seed coats ≤4 % impermeability were rated as "permeable" and >4 % were rated as "impermeable," Fig. 1b) based on the maximum PI 587982A score of 4 %. Eighteen Williams 82 plants were also evaluated as a check, and all had similar ratings as the permeable seed coat parent (Table 1).

Using the binary classification system with 4 % as the break point between permeable and impermeable seed coat based on the parental distribution (described above and shown in Fig. 1a), the F_2 plants were classified as 147 "impermeable seed coat" to 52 "permeable seed coat" (Table 1). These values fit a 3:1 ratio ($\chi^2 = 0.14$, $P = 0.71$) expected for a single gene with dominance/recessive gene action and indicate that impermeable seed coat is dominant to permeable seed coat.

A total of 197 SSR and SNP molecular markers were used to genotype 89 individual plants of F_2-POP1. The seed coat impermeability ratings of these 89 plants represented the extremes of the distribution of the entire 199 plant F_2 population. Of the 89 plants, 45 had a permeable seed coat (all had a seed coat impermeability rating of 0 %) and 44 had an impermeable seed coat (seed coat impermeability rating of ≥66 %). Single marker associations were evaluated using the SMA routines of JMP-Genomics over the whole genome using the SSR and SNP marker data for the 89 F_2 plants with the extreme ratings for seed coat impermeability. Figure 2 shows a Manhattan plot of associations (negative log of probability) with percent seed coat impermeability across all markers analyzed. By far the strongest association detected was with markers located at the end of CHR 2 (Satt274, $-\log(P) = 20.8$ and Sat_192, $-\log(P) = 5.6$). The marker with the strongest association (Satt274) accounted for 65 % of the variation in seed coat impermeability. Additionally, there was a significant, but much less strong association with markers at the end of CHR 7 (Fig. 2). One marker (Sat_330), was significantly associated with seed coat impermeability ($-\log(P) = 2.4$). Sat_330 accounted for about 9 % of the variability in seed coat impermeability. Two other markers at the end of CHR 7 were just below the $-\log(P)$ threshold of 2.0. A somewhat stronger association with seed coat impermeability was also detected with one marker at the end of CHR 20 (Sat_155, $-\log(P) = 5.1$, Fig. 2). This marker accounted

Table 1 Descriptive statistics for seed coat impermeability in the parental lines, Williams 82, the F_2, and $F_{2:3}$ families of the cross PI 587982A × PI 594619 (POP1), and F_1 and F_2 individuals of the reciprocal cross PI 594619 × PI 587982A (POP2) from field studies at Stoneville, MS in 2009 and 2011

Genotype	N[a]	Mean (%)[b]	Impermeable seed (%) range[c]	Observed ratio	Expected ratio	χ^2	*P* value
2009							
PI 587982A (permeable seed coat parent)	46	0.3	0–4				
PI 594619 (impermeable seed parent)	30	80.6	50–94				
Williams 82 (check)	18	1.2	0–4				
F_2-POP1 PI 587982A × PI 594619	199	37.9	0–98	147:52	3:1	0.14	0.71
F_1-POP2 PI 594619 × PI 587982A	2	56	52–60				
F_2-POP2 PI 594619 × PI 587982A	52	32.5	0–88	35:17	3:1	1.64	0.20
2011							
PI 587982A (permeable seed coat parent)	27	0	0				
PI 594619 (impermeable seed parent)	20	73.7	46–96				
Williams 82 (check)	33	6.4	0–22				
$F_{2:3}$ families PI 587982A × PI 594619	99 families (20 plants each)		0–100	26:48:25	1:2:1	0.11	0.95

[a] Number of individual plants

[b] Mean value for seed coat impermeability scores in percent

[c] Seed coat impermeability with 0−4 % is classified as permeable and >4 % as impermeable

for about 22 % of the variability in seed coat impermeability. Very similar results were found for these same markers when the analysis was conducted using the binary classification system (data not shown).

For the three genomic regions identified above (CHRs 2, 7, and 20), the SSR and SNP markers on these chromosomes were run on the entire 199 plants of the F_2 population and additional polymorphic markers on CHR 2. Figure 3 shows the results of analyzing the SSR and SNP markers on the entire F_2 population. Seven markers at the end of CHR 2 were highly significantly associated with seed coat impermeability (Sat_069, Sat_183, Satt274, Satt459, Sat_202, Sat_415, and Sat_192). The marker most strongly associated with seed coat impermeability was SSR Sat_202 ($-\log(P) = 22.91$) which accounted for about 41 % of the variability in seed coat impermeability. Although much less strongly associated with seed coat impermeability than those on CHR 2, four markers (Sat_147, Satt336, Sat_330, and Sat_359) on CHR 7 were over the threshold for significance. All had $-\log(P)$ values between 2.17 and 2.84, slightly above the threshold for significance at 2.0. However, none of these markers individually accounted for more than 5 % of the variation in seed coat impermeability. The one marker on the end of CHR 20 that was significant in the analysis using the extreme phenotypes (Sat_155) was still significant when using the entire F_2 population ($-\log(P) = 3.71$). However, no other significant associations were detected on CHR 20. Additionally, Sat_155 only accounted for about 7 % of the variation in seed coat impermeability.

In addition to using SMA on percent seed coat impermeability, SMA was also conducted using phenotypic data split into two classes (0–4 % as permeable and 5–100 % as impermeable) as described above and shown in Table 1 and Fig. 1b. The results of both analyses (percent seed coat impermeability rating and rating the seed of each plant in a binary system as either permeable or impermeable) were very similar whether using both the extreme phenotype subset (data not shown) or the entire population on CHR 2, 7, and 20 (Fig. 3). On CHR 2, the same marker (Sat_202) had the strongest association in both classification systems. However, for the binary classification, the variation accounted for by Sat_202 increased from 41 to 60 % in the SMA. Similarly, Sat_330 had the strongest association among the markers evaluated on CHR 7 using the binary rating ($-\log(P) = 4.68$) as it did in the seed coat impermeability rating. The amount of variation accounted for by Sat_330 increased to about 9 %. The one marker on CHR 20 that was significantly associated with the seed coat impermeability rating (Sat_155) was not significant using the binary rating system ($-\log(P) = 1.95$, $r^2 = 0.03$).

A two-marker model was evaluated using the two markers with the strongest SMA associations on CHR 2 (Sat_202) and 7 (Sat_330) reported above. Although the contribution of both markers was significant ($P < 0.05$), using both markers only slightly increased the amount of variation accounted for by Sat_202 alone (less than a 0.01 difference in r^2 values). This was the case for both the percentage of seed coat impermeability rating and for the binary rating of either permeable or impermeable.

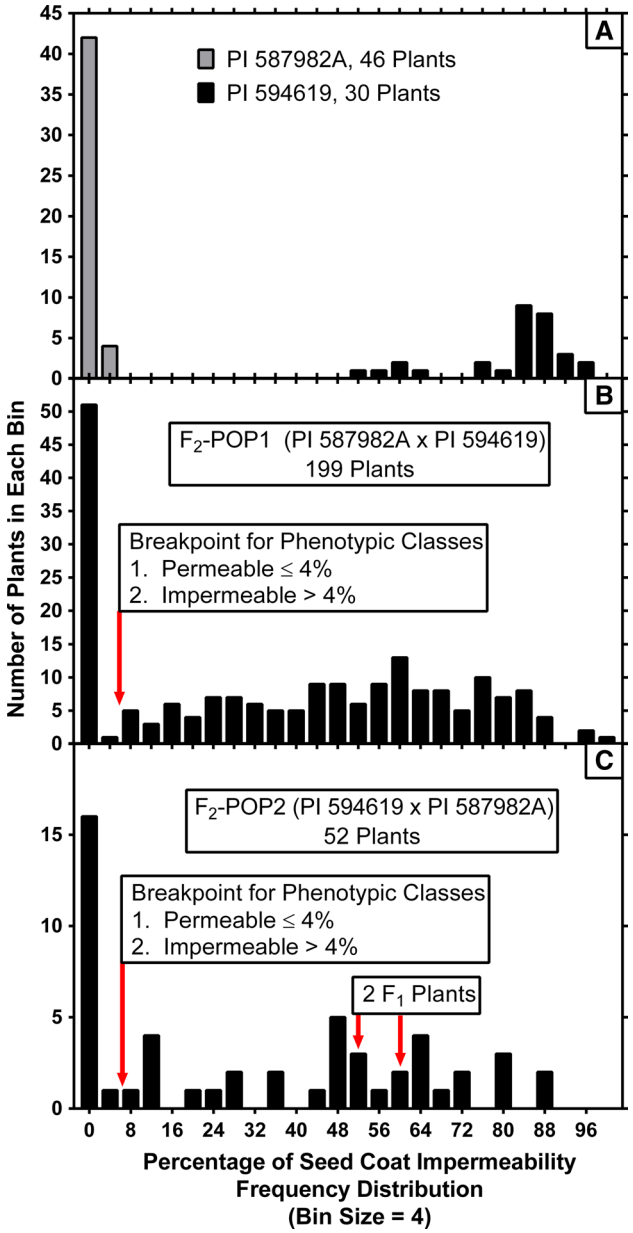

Fig. 1 Frequency distributions for seed coat impermeability in the (**a**) parental lines, PI 587982A (permeable seed coat) and PI 594619 (impermeable seed coat) (**b**) F_2 individuals of the cross PI 587982A × PI 594619 (F_2-POP1), and (**c**) F_2 and F_1 individuals of the reciprocal cross PI 594619 × PI 587982A (F_2-POP2). The break point for phenotypic classes is shown by an *arrow* in each panel, where seeds with impermeability percentages of 4 % or less were rated as permeable and seeds with percentages greater that 4 % were considered impermeable

Both the analysis of the percent seed coat impermeability and the binary classification identified the same primary locus on CHR 2 (near Sat_202). When this marker's allele patterns were examined relative to the binary classification system for the 199 F_2 plants (0–4 % as permeable and 5–100 % as impermeable), all but two plants had the permeable parental allele of Sat_202 in the 0–4 % class (one heterozygote and one impermeable parental allele) and only two plants had the permeable parental allele in the 5–100 % class. The rest of the plants were either heterozygous for Sat_202 or had the impermeable parental allele. Very similar allelic distributions were evident for the nearby marker Satt459. These allelic distributions provide a posteriori confirmation of the binary classification system.

Using the molecular marker data set and the 89 extreme genotypes, a linkage map was developed with the 201 SSR and SNP markers using JoinMap 4.0 software. The map covered 2,187 cM (Kosambi function) and represented all 20 linkage groups. The groupings and positions of almost all markers were consistent with the soybean consensus map (http://soybase.org/MarkerDB/index.php). This map was used in a preliminary QTL analysis to identify regions of interest (results not shown). Linkage groups with regions of interest identified from the preliminary QTL analysis and from the SMA analysis were then expanded to include all 199 F_2 plants and the QTL analysis re-conducted on the expanded linkage groups. The QTL analysis was conducted with MapQTL 6.0 software using both phenotypic classifications (percent seed coat impermeability rating and the binary classification of either permeable or impermeable seed coats). Interval Mapping (IM) was first applied for each single locus to identify putative loci. At a threshold value of LOD 3.5 ($P < 0.05$) as a genome-wide significance threshold, a locus associated with markers Sat_202 (LOD 70.6) and Satt459 (LOD 17.4) on CHR 2 (Fig. 4) showed a highly significant ($P < 0.0001$) association with seed coat impermeability just as in the SMA. Another putative genomic region detected by SMA was also detected by IM on CHR 7 (Sat_330, LOD 4.2; Satt336, LOD 3.6; Sat_147, LOD 3.5). To ensure that these were stable QTL, Multiple-QTL Model (MQM) analysis was performed by using markers close to these putative QTL regions as cofactors. The MQM results indicated that one significant and stable genomic region was associated with the seed coat impermeability ratings in this population, the same region on CHR 2 that showed a highly significant association with seed coat impermeability both in the SMA and IM. MQM showed that seed coat impermeability was highly linked to markers Sat_202 (LOD 88.1, $P < 0.0001$) and Satt459 (LOD 79.0, $P < 0.0001$) (Table 2) with a genomic region spanning 1.6 cM. The phenotypic variation in seed coat impermeability explained by these two markers was 65.6 and 63.3 % for Sat_202 and Satt459, respectively (Table 2). About 65 % of the total phenotypic variation in seed coat impermeability in this population was accounted for by this genomic region. The analysis also indicated that the allele from the impermeable seed parent, PI 594619, contributed to the expression of the impermeable seed coat phenotype in this genomic region.

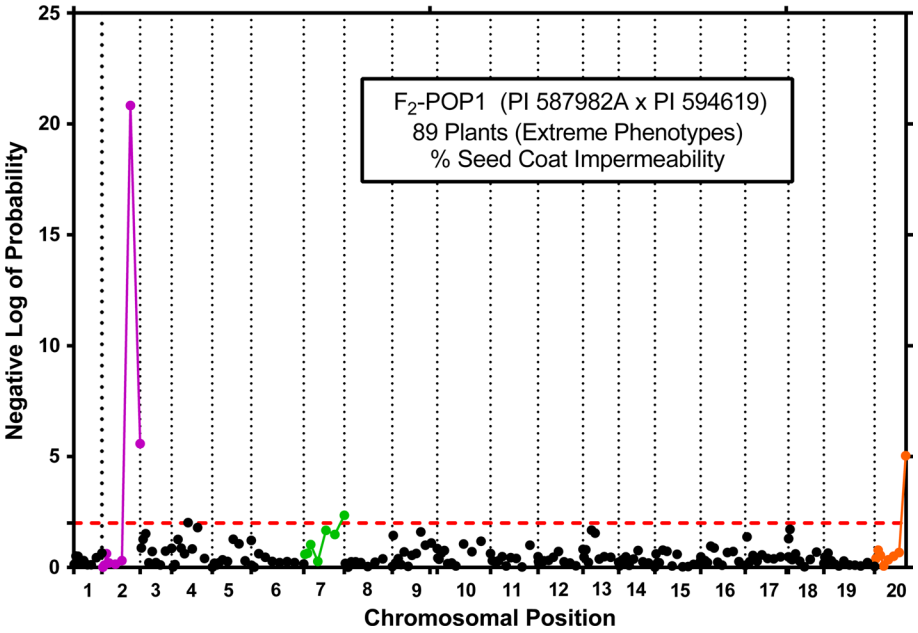

Fig. 2 Probability ($-\log(P)$) of single marker associations between markers and the seed coat impermeability ratings of 89 F_2 plants from POP1 (PI 587982A × PI 594619). The 89 plants represent the extremes of the population and consisted of 45 plants with a permeable seed coat (all had a seed coat impermeability rating of 0 %) and 44 plants with an impermeable seed coat (minimum seed coat impermeability rating of 66 %). A total of 197 SSR and SNP markers from across all 20 chromosomes were applied to the 89-plant subset. The red horizontal dashed line indicates the $-\log(0.01) = 2.0$ level of significance. Markers are arranged according to the order and distance reported in the soybean consensus map (Song et al. 2004)

Inheritance and mapping of $F_{2:3}$ families (POP1)

The $F_{2:3}$ families that descended directly from the individual F_2 plants assayed in POP1 were evaluated for seed coat impermeability in two replicates in 2011. Of the 101 families, two families were eliminated from the analysis because of phenotyping discrepancies that could not be resolved between replications. The remaining 99 $F_{2:3}$ families had a total of 1,980 plants (20 plants per family) that were individually scored for seed coat impermeability. Each family was then classified as having permeable seed coat, impermeable seed coat, or as segregating, using the same break point for each plant as was used for the F_2 plants (Table 1). These classifications provided an independent estimate for inferring the phenotype of the F_2 source plant and also allowed the determination of heterozygous F_2 plants. The parental lines and Williams 82 were also planted and evaluated along with the $F_{2:3}$ families in 2011. The seed coat impermeability ratings for the permeable seed coat parent (PI 587982A) were 0 % for all 27 plants evaluated (Table 1). Ratings for the impermeable seed parent (PI 594619) ranged 46–96 % with a mean of 73.7 %, where 20 plants were evaluated. The check, Williams 82, had a slightly higher seed coat impermeability rating than in 2009, with a range 0–20 and a mean of 6.4 % (Table 1). There was 97 % concordance for seed coat impermeability classifications between the F_2 individuals and their $F_{2:3}$ families. Data from the $F_{2:3}$ families indicated that the F_2 generation segregated 26:48:25 (permeable seed coat: heterozygote: impermeable seed coat), which very closely matched the expected 1:2:1 ratio expected for a single gene ($\chi^2 = 0.11$, $P = 0.95$) (Table 1). In addition, the impermeability/permeability ratings of 20 individual F_3 plants per family, with each family derived from a subset (70) of the 147 F_2 plants that were classified in the impermeable class (Table 1), were further evaluated to determine which of these families were homozygous impermeable and which were segregating. For a single-gene model with dominance gene action, one would expect a ratio of two segregating families:1 impermeable families. The impermeability ratings for these 70 F_2 individuals ranged 6–94 %. After evaluation of the individual F_3 plant data, it was determined that 46 families were segregating and 24 families were homozygous impermeable. This fits very well ($\chi^2 = 0.03$, $P = 0.87$) the above expected 2:1 ratio and confirms the above (Table 1) F_2 classification data that originally suggested a single major gene with dominance gene action.

Figure 5 shows the results of SMA using the corresponding F_2 marker data and the phenotype inferred from the 99 $F_{2:3}$ family classifications of homozygous permeable seed coat, heterozygous, or homozygous impermeable seed coat.

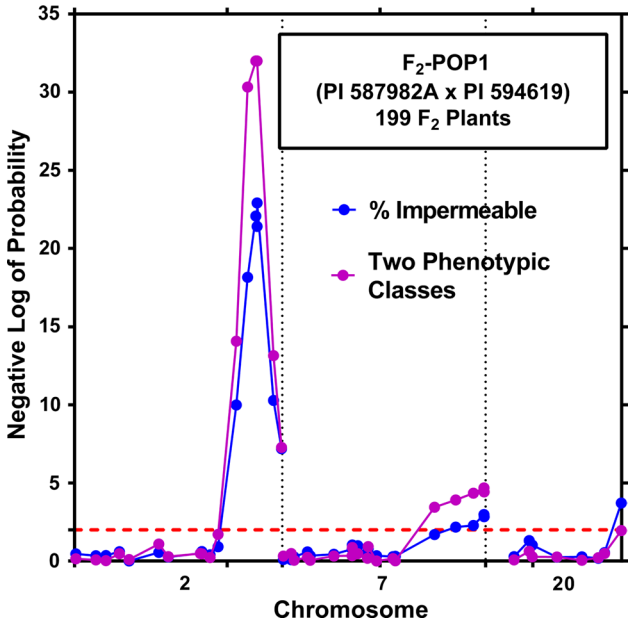

Fig. 3 Probability ($-\log(P)$) of single marker associations between markers located on CHR 2, 7, and 20 and the seed coat impermeability ratings (*blue closed circles*) of 199 F_2 plants from POP1 (PI 587982A × PI 594619). Also shown are the probabilities of associations between the two-class phenotypic ratings (represented by *purple closed circles*) and the markers. The *red horizontal dashed line* indicates the $-\log(0.01) = 2.0$ level of significance. Markers are arranged according to the order and distance reported in the soybean consensus map (Song et al. 2004)

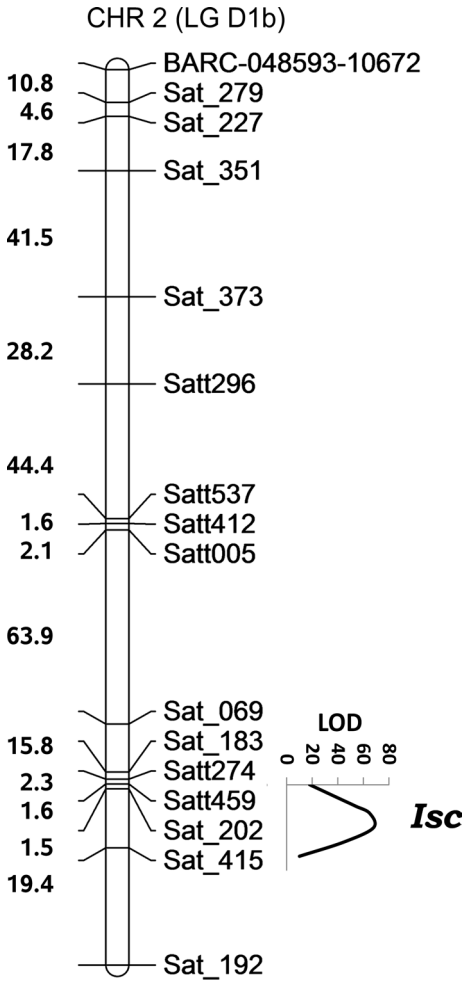

Fig. 4 Genomic region associated with seed coat impermeability on chromosome 2 (LG D1b) identified in the F_2 population of the cross PI 587982A × PI 594619 (POP1). A linkage map was constructed using JoinMap® 4.0 (Van Ooijen 2006). Distances between neighboring markers measured in centi Morgans (cM, Kosambi units) are shown on the left side of the map. Quantitative trait loci (QTL) analysis was conducted using MapQTL6.0 software (Van Ooijen 2009). QTL position and Logarithmic odds (LOD) score plots are shown to the right of the genetic map

That is to say, the corresponding 99 F_2 plants from which the $F_{2:3}$ families were derived, were reclassified based on the results of the $F_{2:3}$ family phenotyping and then reanalyzed. SMA of the $F_{2:3}$ family data detected a very strong association at the end of CHR 2 and a significant, but much less strong association detected on CHR 7 (3 markers each with a $(-\log(P) < 2.5$ and $r^2 \leq 0.08)$. These results were very similar to that of the SMA analysis of the 199 F_2 plants, although the most significant association on CHR 2 was with Satt274 ($-\log(P) = 29.1$, $r^2 = 0.74$) rather than with Sat_202 ($-\log(P) = 27.9$, $r^2 = 0.72$) These results were also confirmed by MQM analysis, as shown in Table 2.

Inheritance and mapping of POP2 (PI 594619 × PI 587982A a reciprocal cross of POP1)

In POP2, the reciprocal cross of POP1, 52 F_2 individuals were evaluated for seed coat impermeability. Seed coat impermeability ratings for these plants ranged 0–88 % with an average rating of 32.5 % (Table 1). The frequency distribution of this F_2 population is shown in Fig. 1c. Along with the F_2 population for this cross, the seed of two F_1 plants were also evaluated and both had impermeable seed coats (52–60 %, Table 1; Fig. 1), indicating that impermeable seed coat is dominant over permeable seed coat. The impermeable: permeable seed coat ratio for the F_2 plants was 35:17, which fits the 3:1 ratio ($\chi^2 = 1.64$, $P = 0.20$) expected for a single gene (Table 1) and so again supports impermeable seed coat as dominant to permeable seed coat. Further, that the F_2 data from the reciprocal crosses both segregate 3:1 indicates that the trait is not maternally inherited. That is, there was no evidence of a cytoplasmic effect on seed coat impermeability.

Twenty-one SSR and SNP markers on CHR 2 were used to genotype the F_2 individuals in POP2. As with POP1, markers at the end of CHR 2 were found to be significantly

associated with seed coat impermeability by SMA (Fig. 6). Of the markers tested, seven had a $-\log(P) > 2.0$ with SSR Sat_202 having the highest probability ($-\log(P) > 8.0$). Using the percent impermeability rating, Sat_202 accounted for 54 % of the variability, whereas using the binary system (either impermeable or permeable) it accounted for 72 % of the variability. Multiple QTL Model analysis of the POP2 F_2 data also detected highly significant associations between seed coat impermeability ratings and markers at the end of CHR 2 (Table 2), including Sat_202 (LOD 28.5, $P < 0.00001$) which accounted for 83.9 % of the phenotypic variability. However, from the MQM analysis two other markers (Satt274 and Satt459) tightly linked to Sat_202, accounted for a greater percentage of the phenotypic variation (92.1 and 92.0 %, respectively, Table 2). Although the size of POP2 was small, it was developed from a reciprocal cross of POP1 and the results independently confirm the presence and the position of the gene detected in POP1.

Discussion

High-quality seed for planting has become an issue in soybean production areas with excessively high temperatures during seed fill, such as in the midsouthern United States where the ESPS is practiced (Tyler 1999; Heatherly 1999; Mengistu and Heatherly 2006; Gillen et al. 2012). This environmental stress increases the occurrence of soybeans with impermeable seed coat, which is associated with non-uniform and delayed germination and emergence (Keith and Delouche 1999). In this study, an effort was made to investigate the inheritance of seed coat impermeability and to identify gene(s) that control this trait in a *Glycine max* soybean accession with impermeable seed coat (PI 594619) grown under field conditions in the midsouthern United States. Seed coat impermeability is caused by both genetic and environmental factors (Copland and McDonald 1999). The objective in the present study was to study the inheritance of this trait in a high temperature environment,

Table 2 Markers significantly associated with seed coat impermeability in the F_2 plants of the cross PI 587982A × PI 594619 (POP1), the $F_{2:3}$ families of POP1, and the F_2 plants of the reciprocal cross, PI 594619 × PI 587982A (POP2) based on MQM analysis

Chromosome (linkage group)	Marker	Position		LOD	Phenotypic variation explained (%)
		Gm consensus 2008a (cM)	Sequence position (bp)		
F_2-POP1: PI 587982A × PI 594619 (analysis of 199 F_2 plants)					
2 (D1b)	Satt459	119.5	48,389,888	79.0	63.3
2 (D1b)	Sat_202	119.6	48,576,403	88.1	65.6
PI 587982A × PI 594619 (analysis of 99 $F_{2:3}$ families derived from POP1, 20 plants per family)					
2 (D1b)	Satt274	118.6	48,345,840	29.4	74.2
2 (D1b)	Satt459	119.5	48,389,888	28.3	72.8
2 (D1b)	Sat_202	119.6	48,576,403	27.1	71.3
F_2-POP2: PI 594619 × PI 587982A (reciprocal of POP1, analysis of 52 F_2 plants)					
2 (D1b)	Satt274	118.6	48,345,840	28.7	92.1
2 (D1b)	Satt459	119.5	48,389,888	28.5	92.0
2 (D1b)	Sat_202	119.6	48,576,403	20.6	83.9

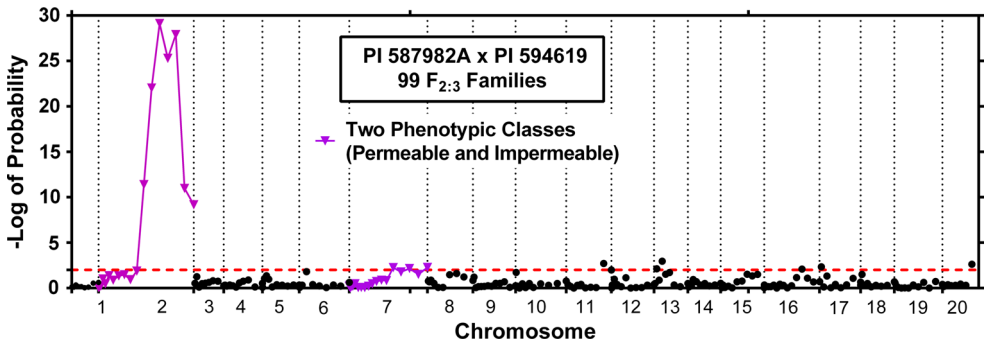

Fig. 5 Probability ($-\log(P)$) of single marker associations between markers and the seed coat impermeability ratings of 99 $F_{2:3}$ families from POP1 (PI 587982A × PI 594619). Single marker analysis was conducted using the corresponding marker data from 99 F_2 plants and the phenotype inferred from the 99 $F_{2:3}$ family classifications of homozygous permeable seed coat, heterozygous, or homozygous impermeable seed coat. The *red horizontal dashed line* indicates the $-\log(0.01) = 2.0$ level of significance

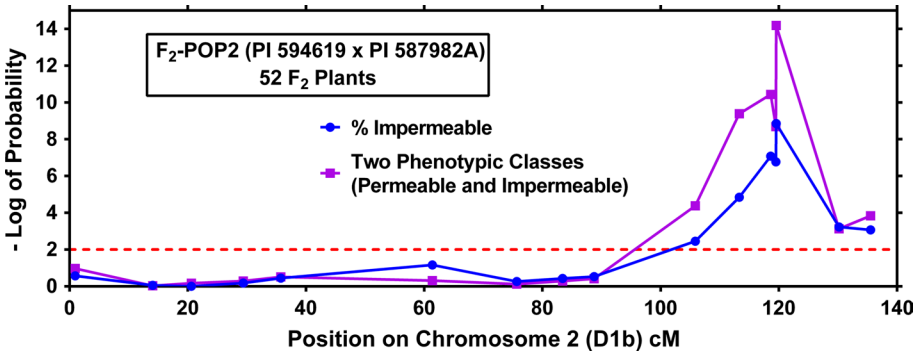

Fig. 6 Probability (−log(*P*)) of single marker associations between markers located on CHR 2 and the seed coat impermeability ratings (*blue filled-circles*) of 52 F_2 plants from POP2 (PI 594619 × PI 587982A). Also shown are the probabilities of associations between the two-class phenotypic ratings (represented by *purple filled-squares*). The *red horizontal line* indicates the −Log(0.01) = 2.0 level of significance

which is common in the production areas of the midsouthern United States. A high-temperature production environment combined with a genetic predisposition for impermeability results in a highly impermeable seed coat. Such seeds are unfit for planting due to their low and inconsistent germinability.

Genetic analysis of the F_2 individuals of the crosses between the impermeable seed coat line, PI 594619, and the permeable seed coat line, PI 587982A, indicated that seed coat impermeability in PI 594619 is controlled by a single major gene, with impermeable seed coat being dominant to permeable seed coat. Analysis of the binary classification of phenotypic ratings (0–4 % permeable and 5–100 % impermeable) on reciprocal F_2 populations and on $F_{2:3}$ families derived from one of the F_2 populations, showed that the segregation in all three evaluations fit expected ratios for a single gene. The segregation data and data on F_1 plants also indicated that impermeability was dominant over permeability.

Molecular mapping of the 199-member F_2 population indicated the most likely position of the putative locus associated with impermeable seed coat to be in the region of Satt274 to Satt459 on CHR 2. This location was confirmed using phenotypic results of 99 $F_{2:3}$ families and a smaller, but independent reciprocal F_2 population. Markers in this region accounted for 63–92 % of the phenotypic variation for seed coat impermeability (Table 2). The molecular marker results are consistent with the genetic segregation results and confirm that one major gene was involved in conditioning seed coat impermeability under the conditions of the present study.

In previous studies, six genomic regions on CHR 2, 3, 6, 8, 19 and 20 were reported to have effect on seed coat impermeability in soybean in four different mapping populations (Keim et al. 1990; Sakamoto et al. 2004; Watanabe et al. 2004; Liu et al. 2007). In each of these studies, more than one genomic region was found to be associated with variation in seed coat impermeability, but one genomic region on CHR 2 was common to all the studies. In the present study, this same genomic region on CHR 2 was found to be associated with seed coat impermeability using markers Satt459 and Sat_202 positioned at 119.5 and 119.6 cM, respectively, on the Gm Consensus40_D1b map (Soybase 2014). This indicates that this genomic region contains a gene that conditions seed coat impermeability in mapping populations created by *G. max* × *G. max* crosses (present study and Watanabe et al. 2004) as well as in mapping populations created by *G. max* × *G. Soja* crosses (Keim et al. 1990; Liu et al. 2007). However, in the present study the major gene identified on CHR 2 in PI 594619 had a larger effect on seed coat impermeability than did all the QTL reported in prior studies, accounting for 63–92 % of the phenotypic variation (Table 2).

In previous studies, the genomic region on Chr 2 (D1b) accounted for 13 % (Keim et al. 1990), 10.6 % (Watanabe et al. 2004), and up to 47.8 % (Liu et al. 2007). The larger effect observed in the expression of this gene in PI 594619 may be due to the uniqueness of this genotype and to its interaction with environmental factors. All four studies were conducted in different environments. The studies of Keim et al. (1990) and Watanabe et al. (2004) were conducted in the field in Iowa, USA, and Chiba, Japan, respectively. The study of Liu et al. (2007) was conducted in a greenhouse. Only the current study was purposefully conducted under conditions of high temperatures. In the other three studies, the effect of temperature was not reported, even though it is known that high temperatures can enhance seed coat impermeability (Gibson and Mullen 1996; Spears et al. 1997; Egli et al. 2005; Smith et al. 2008). The mapping populations derived from PI 594619 were grown under high temperatures and this likely enhanced the effect of the single major gene for expressing seed coat impermeability.

Even though one major gene was detected in the present study, the continuous nature of the seed coat impermeability scores in the progenies evaluated in this study is an indication that minor genes may have also been involved in determining a portion of the variation. The genomic region associated with seed coat impermeability that we detected with the SMA and IM on CHR 7 may have been overshadowed by the effect of the major gene on CHR 2. Possibly, the region on CHR 7 may contain another gene with a minor effect on seed coat impermeability. However, the other studies reporting genomic regions associated with seed coat impermeability did not report anything of significance on CHR 7 (Keim et al. 1990; Sakamoto et al. 2004; Watanabe et al. 2004; Liu et al. 2007). Possibly under other environmental conditions, its effect could be larger. Even so, this is the first report of a region on CHR 7 associated with seed coat impermeability in soybean. As such, further studies under different environments are needed to confirm the validity and impact of this minor QTL.

As shown in Soybase, within the genomic region for seed coat impermeability we identified on CHR 2, there are QTL for seven seed amino acids (Glycine, Threonine, Glutamine, Tyrosine, Phenylalanine, Leucine and Arginine) (Panthee et al. 2006). This region has also been reported to have a major seed protein QTL (Panthee et al. 2005). However, we have not found any report indicating a relationship between amino acid/protein composition and seed coat permeability.

The current study confirms that a region in the soybean genome near linked markers Sat_202 and Satt459 on CHR 2 (D1b) is associated with seed coat impermeability. Previous studies also reported this region as containing a QTL associated with seed coat impermeability. However, based on phenotypic segregation ratios in the current study, the locus acted as a single major gene. Possibly this locus acted as a single major gene as a result of the high-temperature environment in which the experiments were conducted, coupled with the wide phenotypic difference between the parents. It may also be a result of the specific allele present in PI 594619, which has not been previously tested in a genetic study. From the analysis of two independent F_2 populations, accompanying F_1 plants, and related $F_{2:3}$ families in independent assays, the current research determined that this major single gene was responsible for 63–92 % of the observed variation for seed coat impermeability and that it has dominance gene action (impermeability is dominant to permeability). As it was clear that a single major gene from PI 594619 is associated with impermeable seed coat, we proposed *Isc* (impermeable seed coat) as its designation, which was approved by the Soybean Genetics Committee. We preferred *Isc* rather than a derivative of "hard seed" to distinguish this locus from loci putatively involved in the physical hardness of seed, especially as the term is used in the food industry (Zhang et al. 2008).

The recognition of *Isc* as a major gene with known gene action has very useful implications for soybean breeding. Of immediate importance is the knowledge that PI 587982A has the recessive form of the gene (*isc*) and that its phenotype for seed coat impermeability is nearly zero under high-heat production conditions. PI 587982A and markers (Sat_202 and Satt459) linked to *isc* can be utilized in breeding programs to create and select new soybean lines with very low levels of seed coat impermeability. This could be valuable for the Tofu, Natto, Miso, and other food-grade markets (Watanabe et al. 2004), where there is near zero tolerance for seeds with impermeable seed coats, as well as for seed grade markets, where seeds are produced to be planted by producers for grain. In both cases, usable yields of seed lots will be increased by lowering the levels of seed with impermeable seed coats, thereby improving overall profits.

Another potential use for recognizing the single-gene nature of *Isc* is in its possible use for reducing seed damage caused by weathering and mold. Each year, producer earnings on grain soybeans are discounted at the time of sale to elevators due to seed damage caused by mold, weathering, stinkbugs, heat, discoloration, etc. In the past, the impermeable seed coat trait was proposed as a way to protect beans from field weathering (Potts et al. 1978; Hartwig and Potts 1987). For example, Hartwig and Potts (1987) reported higher seed viability after 6 weeks of post-maturity weathering for lines with impermeable versus permeable seed coats. However, this protection was never realized for producers because no large-scale effort to scarify hard seeds prior to planting was ever established. But understanding the mechanisms for impermeable seed coat may now allow alternatives to scarification. Identifying the genomic location of genes controlling impermeable seed coat could be an initial step in understanding the biochemical and physiological processes leading to an impermeable seed coat. This understanding coupled with modern recombinant DNA technology may allow improved protection against weathering. Although highly undesirable for food-grade soybean, the impermeable seed coat trait is not discounted for U.S. No. 1 grain-grade yellow soybean. And PI 594619, the source of *Isc*, contains the i^i allele for yellow seed coat, thereby making it more readily useful to soybean breeders. Hence, knowledge of the *Isc* locus and its alleles could potentially benefit both food grade and grain-grade markets in multiple ways.

Author contributions H. Kebede phenotyped populations, analyzed the data genotypically for molecular markers and mapping, and was the lead author in writing the manuscript. J. R. Smith conceived the experiment, created

and phenotyped the populations, analyzed the data for Mendelian segregation ratios and models, and co-wrote the manuscript. J. D. Ray analyzed the data for Mendelian segregation ratios and models, analyzed the data genotypically for molecular markers and mapping, and co-wrote the manuscript.

Acknowledgments This work was supported by United States Department of Agriculture–Agriculture Research Service (USDA-ARS) project number 6402-21220-012-00D. The authors would also like to thank the United Soybean Board for funding (Project numbers 9232 and 1232). The authors appreciate the dedicated work of the USDA-ARS personnel at Stoneville, MS, including Dr. Eduardo Benitez, Ms. Loria Courtney and Mr. Hans Hinrichsen in harvesting plants of the $F_{2:3}$ families in the field, and also Ms. Angelie Davis and Rebecca Simpson in assisting in the molecular work in the lab. In Puerto Rico, we were ably assisted in seed production by Dr. Ricardo Goenaga and Mr. Carlos Torres of the USDA-ARS. Mention of trademark or proprietary product does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture and does not imply approval or the exclusion of other products that may also be suitable. USDA is an equal opportunity provider and employer.

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

References

Association of Official Seed Analysts (2006) Rules for testing seed. Association of Official Seed Analysts, Las Cruces

Bernard RL, Cremeens CR (1988) Registration of 'Williams 82' soybean. Crop Sci 28:1027–1028

Carlson JB, Lersten NR (2004) Reproductive morphology. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production and uses, 3rd edn. American Society of Agronomy, Madison, pp 59–95

Copland LO, McDonald MB (1999) Principles of seed science and technology, 3rd edn. Springer, New York

Egli DB, TeKrony DM, Heitholt JJ, Rupe J (2005) Air temperature during seed filling and soybean seed germination and vigor. Crop Sci 45:1329–1335

Esau K (1965) Plant anatomy, 2nd edn. Wiley, New York, p 617

Fehr WR, Caviness CE (1977) Stages of soybean development. Iowa agric. Exp. Stn. Spec. Rep. 80. Iowa State Univ, Ames

Geater CW, Fehr WR, Wilson LA (2000) Association of soybean seed traits with physical properties of natto. Crop Sci 40:1529–1534

Gibson LR, Mullen RE (1996) Soybean seed quality reductions by high day and night temperatures. Crop Sci 36:1615–1619

Gillen AM, Smith JR, Mengistu A, Bellaloui N (2012) Effects of maturity and phomopsis on germination and vigor of soybean seed of near isogenic lines. Crop Sci 52:2757–2766

Hartman HT, Kester DE (1975) Plant propagation; principles and practices, 3rd edn. Prentice-Hall, Englewood Cliffs, pp 131–132

Hartwig EE, Potts HC (1987) Development and evaluation of impermeable seed coats for preserving soybean seed quality. Crop Sci 27:506–508

Heatherly LG (1993) Drought stress and irrigation effects on germination of harvested soybean seed. Crop Sci 33:777–781

Heatherly LG (1999) Early soybean production system (ESPS). In: Heatherly LG, Hodges HF (eds) Soybean production in the midsouth. CRC Press, Boca Raton, pp 103–118

Keim P, Diers BW, Shoemaker RC (1990) Genetic analysis of soybean hard seededness with molecular markers. Theor Appl Genet 79:465–469

Keith BC, Delouche JC (1999) Seed quality, production and treatment. In: Heatherly LG, Hodges HF (eds) Soybean production in the midsouth. CRC Press, Boca Raton, pp 197–239

Kilen TC, Hartwig EE (1978) An inheritance study of impermeable seed in soybeans. Field Crops Res 1:65–70

Liu B, Fujita T, Yan Z, Sakamoto S, Xu D, Abe J (2007) QTL Mapping of domestication—related traits in soybean (*Glycine max*). Ann Bot 100:1027–1038

Mengistu A, Heatherly LG (2006) Planting date, irrigation, maturity group, year, and environment effects on *Phomopsis longicolla*, seed germination, and seed health rating of soybean in the early soybean production system of the midsouthern USA. Crop Prot 25:310–317

Panthee DR, Pantalone VR, West DR, Saxton AM, Sams CE (2005) Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Sci 45:2015–2022

Panthee D, Pantalone VR, Saxton AM, West DR, Sams CE (2006) Genomic regions associated with amino acid composition in soybean. Mol Breed 17(1):79–89

Potts HC, Duangpatra J, Hairston WG, Delouche JC (1978) Some influences of hard seededness on soybean quality. Crop Sci 18:221–224

Sakamoto S, Abe J, Kanazawa A, Shimamoto T (2004) Marker-assisted analysis for soybean hard seededness with isozyme and simple sequence repeat loci. Breed Sci 54:133–139

Smith JR, Mengistu A, Nelson RL, Paris RL (2008) Identification of soybean accessions with high germinability in high-temperature environments. Crop Sci 48:2279–2288

SoyBase, the USDA-ARS Genetics and Genomics Database (2014) http://soybase.org. Accessed 29 Jul 2014

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128

Spears JF, TeKrony DM, Egli DB (1997) Temperature during seed filling and soybean seed germination and vigour. Seed Sci Technol 25:233–244

Tyler JM (1999) Note on variation in germinability among early-maturing soybean populations. Field Crops Res 62:177–180

USDA, ARS, National Genetic Resources Program (2014) Germplasm Resources Information Network—(GRIN). National Germplasm Resources Laboratory, Beltsville, Maryland. http://www.ars-grin.gov/cgi-bin/npgs/acc/search.pl?accid=PI+587982A. Accessed 13 Mar 2014

USDA, ARS, National Genetic Resources Program (2014) Germplasm Resources Information Network—(GRIN). National Germplasm Resources Laboratory, Beltsville, Maryland. http://www.ars-grin.gov/cgi-bin/npgs/acc/search.pl?accid=PI+594619. Accessed 07 Mar 2014

Van Ooijen JW (2006) JoinMap® 4.0. Software for the calculation of genetic linkage maps in experimental populations. Wageningen, The Netherlands

Van Ooijen JW (2009) MapQTL® 6.0. Software for the mapping of quantitative trait loci in experimental populations. Wageningen, The Netherlands

Watanabe S, Tajuddin T, Yamanaka N, Hayashi M, Harada K (2004) Analysis of QTLs for reproductive development and seed quality traits in soybean using recombinant inbred lines. Breed Sci 54:399–407

Zhang B, Chen P, Chen CY, Wang D, Shi A, Hou A, Ishibashi T (2008) Quantitative trait loci mapping of seed hardness in soybean. Crop Sci 48:1341–1349