INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Molecular marker analysis of seed size in soybean

by

Joseph Andrew Hoeck

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Plant Breeding

Program of Study Committee
Walter R. Fehr, Major Professor
E. Charles Brummer
Alicia L. Carriquiry
Rohan L. Fernando
Randy C. Shoemaker

Iowa State University
Ames, IA
2002

Graduate College
Iowa State University

This is to certify that the Doctoral dissertation of

Joseph Andrew Hoeck

has met the requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

TABLE OF CONTENTS

ABSTRACT

Seed size is an important attribute of soybean [*Glycine max* (L.) Merr.] for some food uses. The objectives of this study were to identify markers associated with quantitative trait loci for seed size (SSQTL), determine the influence of the environment on expression of the marker-SSQTL associations, and compare the efficiency of phenotypic selection and marker-assisted selection for the trait. Three small-seeded lines were crossed to a line or cultivar with normal seed size to form three two-parent populations. The parents of the populations were screened with 178 simple sequence repeat (SSR) markers to identify polymorphism. Population 1 (Pop 1) had 75 polymorphic SSR markers covering 1306 cM, population 2 (Pop 2) had 70 covering 1143 cM, and population 3 (Pop 3) had 82 covering 1237 cM. Seed size of each population was determined with 100 F_2 plants grown at Ames, IA, and their F_2-derived lines grown in two replications at three environments. Single-factor analysis of variance and multiple regression were used to determine significant marker-SSQTL associations. Pop 1 had 12 markers that individually accounted for 8 to 17% of the variation for seed size. Pop 2 had 16 markers that individually accounted for 8 to 38% of the variation, and Pop 3 had 22 markers that individually accounted for 8 to 29% of the variation. Four of the 12 markers in Pop 1, four in Pop 2, and one in Pop 3 had significant associations with SSQTL across four environments, while five loci in Pop 1, seven in Pop 2, and eight in Pop 3 had significant associations in more than one environment. Three marker loci that had significant SSQTL associations in this study also were significant in previous research, and 13 markers had unique SSQTL associations. The relative effectiveness of phenotypic and marker-assisted selection among F_2 plants varied for the three populations. On the average, phenotypic selection for seed size was as effective and less expensive than marker-assisted selection.

CHAPTER 1. GENERAL INTRODUCTION

Introduction

The cultivated soybean [*Glycine max* (L.) Merrill] is one of the major oilseed crops of the world (Fehr, 1987). Soybean also is a source of high quality protein for human and animal consumption. Seed size is an important trait for production of some soy food products. Seed size of *G. max* strains ranges from 40 to 550 mg seed^{-1} (Hartwig, 1973). Small-seeded soybeans with ≤ 80 mg seed^{-1} are used in the production of sprouts and natto, a fermented soybean. Large-seeded soybeans with ≥ 250 mg seed^{-1} are used in the production of miso, a paste made from soybean, a fungus, and grain such as rice or barley, and for edamame, a food dish for which the green soybean pods are boiled in water and the green seed is consumed as a vegetable. Soybean that possess large seed size ≥ 200 mg seed^{-1} and high protein are desired for the production of tofu. Tofu is made by coagulating soymilk to concentrate the solids.

Breeders are attempting to increase the yield of soybean cultivars of different seed sizes for the various food products. The traditional method of cultivar improvement for food-grade soybean involves the use of artificial hybridization to develop genetic variability followed by self-fertilization and screening of the offspring for the desired size.

Molecular markers may augment traditional methods of breeding for seed size in soybean. Once the molecular markers associated with seed size have been identified in multiple populations over multiple generations and in multiple environments, the plant breeder can use these data to decide which parents to cross to develop breeding populations (Dudley, 1993). This information also could be useful for screening offspring from a segregating population in any generation to identify suitable progeny for field evaluation (Lamkey and Lee, 1999). This will probably not decrease the time involved in developing new cultivars, but it may decrease the amount of

resources needed to breed for a particular trait which would make it possible to breed for additional traits with the same amount of resources (Lamkey and Lee, 1999).

The use of molecular markers, like simple sequence repeats (SSR) and restriction fragment length polymorphism (RFLP), provide a powerful tool for the analysis of plant genome structure and function (Shoemaker and Specht, 1995). The marker density of SSRs and RFLPs on molecular maps make them useful for genetic research purposes ranging from the detection of quantitative trait loci (QTL) to map-based cloning of agronomically important genes (Shoemaker and Specht, 1995). Molecular markers have no known effect on the phenotype of the plant making them ideal for studying quantitative traits (Stuber, 1992).

Several types of populations have been used to map the QTL for seed size in soybean. Mian et al. (1996) developed two *G. max* populations from four parents with normal seed size. Maughan et al. (1996) crossed a *G. max* parent with 240 mg seed^{-1} to an accession of wild soybean [*Glycine soja* (L.) Sieb. & Zucc.] with 15 mg seed^{-1}. Mansur et al. (1996) developed a recombinant inbred population from the cross between 'Minsoy' and 'Noir 1'. Orf et al. (1999) compared three populations derived from crossing 'Archer', Minsoy, and Noir 1 to each other. Sebolt et al. (2000) developed a backcross population derived from a *G. max* recurrent parent and an F_2-derived line from a cross between *G. max* and *G. soja*. My study is based on three single-cross populations between normal and small-seeded *G. max* parents. Populations from small-seeded and normal-seeded parents have segregation for seed size within the range of the two parents, which is ideal for detecting QTL (Dudley, 1993; Johnson et al., 2001).

The objectives of my study were to (i) estimate the number and distribution of QTL associated with seed size (SSQTL) in the three soybean populations developed at Iowa State University, (ii) determine the influence of the environment on expression of the marker-QTL associations, (iii) determine the effect of genetic background on SSQTL, and (v) compare the

efficiency of phenotypic selection, marker-assisted selection, and an index of phenotypic and molecular marker data to select among soybean plants for seed size.

Dissertation Organization

This dissertation has been organized into four chapters. Chapter 1 is a review of the literature on the inheritance of seed size in soybean, QTL detection and estimation, and previous studies that identified SSQTL. Chapter 2 is a manuscript submitted for publication in *Crop Science*. General conclusions will be discussed in Chapter 3. Additional data not contained in Chapter 2 will be found in the appendices in Chapter 4.

Literature Review

Quantitative Trait Loci

Quantitative genetic variation is attributed to the segregation of multiple genes with small individual effects. Quantitative traits are influenced by the environment, genotype of individuals, and genotype × environment interactions. With the advent of molecular markers and statistical software packages, the detection of QTL was possible. To study QTL, the properties of the genes individually need to be considered, including their frequencies and the magnitude of their effects on the trait of interest (Falconer and Mackay, 1996). The components that comprise a QTL experiment are (1) a population that is segregating for the trait of interest, (2) a linkage map, (3) quantitative data, including both phenotypic and molecular-maker data, and (4) a QTL analysis tool, such as single-factor analysis of variance, MAPMAKER, QTL Cartographer, or PLABQTL. The results that can be derived from QTL experiments include the number and location of QTL that control the trait, the amount of phenotypic variation accounted for by a putative QTL, and which parent possesses the favorable alleles for the trait. The information from QTL experiments

can be used in designing marker-assisted selection programs to improve parent selection, to classify germplasm, to facilitate map-based cloning, and to establish evolutionary relationships between species (Dudley, 1993). The utility and power of QTL analysis may be limited and conclusions may only be formed about genetic variation that exists within the segregating population being studied (Beavis et al., 1991). Therefore, mapping SSQTL requires replicated testing of lines over multiple environments to reliably determine their phenotype. Mapping several populations is necessary to find the majority of the SSQTL. Both of these requirements cause QTL mapping to be costly and time consuming.

The development of improved soybean cultivars depends on the genetic potential of available parents and the amount of genetic variability generated when they are mated. Iowa State University has been developing small-seeded cultivars since 1977 (Carpenter and Fehr, 1986). Superior small-seeded cultivars typically are derived from a single-cross between two small-seeded parents or from a three-way cross. For a three-way cross, a small-seeded parent is mated to a high-yielding parent with normal size and the F_1 from the mating is crossed to a second small-seeded parent. Screening the parents and offspring with markers associated with seed size may improve the efficiency and effectiveness of a breeding program. One current limitation for the use of markers in a breeding program is the cost of marker analysis. As the technology improves, the use of markers to facilitate breeding for seed size may be possible at a lower cost.

Inheritance of Seed Size

Seed size in soybean is inherited as a quantitative trait (Ting, 1946). Weber (1950) concluded that seed size was primarily controlled by additive gene action. In his study of a cross between the *G. max* parent 'Dunfield' (162 mg seed^{-1}) and the *G. soja* parent PI65569 (16 mg seed^{-1}), none of the F_2 plants had the same size as either parent.

Weber and Moorthy (1952) developed three populations with *G. max* parents of similar seed size. F_2-derived lines from the crosses 'Adams' (144 mg seed^{-1}) × 'Habaro' (181 mg seed^{-1}), Habaro × 'Mandel' (153 mg seed^{-1}), and Adams × 'Hawkeye' (169 mg seed^{-1}) were evaluated for seed size. The crosses Adams × Habaro and Habaro × Mandel had transgressive segregates with seed size smaller and larger than the parents, while the cross Adams × Hawkeye had transgressive segregates with seed size equal to or larger than Hawkeye. The heritabilities for seed size on a plot basis were 54%, 47%, and 62% for the three crosses.

Brim and Cockerham (1961) developed two *G. max* populations; N48-4860 (314 mg seed^{-1}) × 'Lee' (250 mg seed^{-1}) and 'Roanoke' (297 mg seed^{-1}) × Lee (263 mg seed^{-1}). They reported that the mean seed size of the population regressed toward the mid-parent value with successive generations of selfing. They reported that the mean seed size was 308 mg seed^{-1} for the F_1, 284 mg seed^{-1} for the F_2 and F_3, 281 mg seed^{-1} for the F_4, and 280 mg seed^{-1} for the F_5 for the cross N48-4860 × Lee. For the Roanoke × Lee cross, the mean seed size was 298 mg seed^{-1} for the F_1, 288 mg seed^{-1} for the F_2, 278 mg seed^{-1} for the F_3, 289 mg seed^{-1} for the F_4, and 276 mg seed^{-1} for the F_5. They concluded that genetic variability for seed size was primarily additive.

Buhr (1976) developed one population by crossing a *G. max* cultivar 'Hill' (198 mg seed^{-1}) with a *G. soja* strain PI245331 (7 mg seed^{-1}) and a second population by crossing the *G. max* cultivar 'Hardee' (218 mg seed^{-1}) with a *G. max* strain PI2276787 (66 mg seed^{-1}). The seed size of the $F_{2:3}$ lines ranged from 28 to 72 mg seed^{-1} for the first population and from 96 to 171 mg seed^{-1} for the second population. None of the $F_{2:3}$ lines possessed a seed size equal to that of their parents.

Carpenter and Fehr (1986) developed two populations from *G. max* and *G. soja* parents. One population was developed from the cross of the *G. max* cultivar 'Amsoy 71' (136 mg seed^{-1}) and the *G. soja* strain PI424001 (21 mg seed^{-1}). The second population was developed from the cross between the *G. max* cultivar 'Century' (168 mg seed^{-1}) and the *G. soja* strain PI326581 (12

mg seed^{-1}). The seed size for the F_2 plants ranged from 38 to 84 mg seed^{-1} and the $F_{2:3}$ lines ranged from 35 to 81 mg seed^{-1} for the Amsoy 71 × PI424001 cross. For the Century × PI326581 cross, seed size for the F_2 plants ranged from 32 to 81 mg seed^{-1} and for the $F_{2:3}$ lines ranged from 27 to 67 mg seed^{-1}. None of the F_2 plants or $F_{2:3}$ lines had seed size equal to that of their parents. Heritabilities for seed size on a single-plant basis were 72% and 81% for the two populations.

Cianzio and Fehr (1987) studied reciprocal crosses of the *G. max* cultivar Century (212 mg seed^{-1}) with the *G. soja* strain PI326581 (20 mg seed^{-1}) and the *G. max* cultivar Amsoy 71 (197 mg seed^{-1}) with the *G. soja* strain PI424001 (33 mg seed^{-1}). The mean size of the F_1 seeds for Century × PI326581 was 51 mg seed^{-1} and for PI326581 × Century was 49 mg seed^{-1}. The mean size of the F_1 seeds for Amsoy 71 × PI424001 and the reciprocal cross were both 60 mg seed^{-1}. They concluded that there was partial dominance for seed size in soybean.

Bravo et al. (1980) obtained heritabilities for seed size utilizing the populations developed in Bravo et al. (1981). Based on the evaluation of $F_{2:3}$ lines, the average heritabilities were 27% on a plant, 41% on a plot , and 71% on an entry-mean basis. Bravo et al. (1981) examined the segregation of seed size in two-parent and three-parent crosses between soybean cultivars and experimental lines that possessed normal and large seed size. Three sets of populations were developed, each consisting of a two-parent and three-parent cross. Set one consisted of a two-parent cross between A72-512 (172 mg seed^{-1}) and 'Prize' (282 mg seed^{-1}) and the three-parent cross of (A72-512 × Prize) × A74-201020 (224 mg seed^{-1}). Seed size of the $F_{2:3}$ lines ranged from 191 to 268 mg seed^{-1} for the two-parent cross and from 192 to 281 mg seed^{-1} for the three-parent cross. Set two consisted of a two-parent cross between A73-19084 (153 mg seed^{-1}) and 'Disoy' (256 mg seed^{-1}) and a three-parent cross of (A73-19084 × Disoy) × 'Vinton' (230 mg seed^{-1}). Seed size of the $F_{2:3}$ lines ranged from 173 to 234 mg seed^{-1} for the two-parent cross and from 177 to 244 mg seed^{-1} for the three-parent cross. The third set consisted of a two-parent cross between A74-

104030 (198 mg seed^{-1}) and Prize (282 mg seed^{-1}) and the three-parent cross of (A74-104030 × Prize) × A74-201026 (221 mg seed^{-1}). Seed size of the $F_{2:3}$ lines ranged from 208 to 287 mg seed^{-1} for the two-parent cross and from 184 to 268 mg seed^{-1} for the three-parent cross. In set one and two, the $F_{2:3}$ lines did not have seed size equal to greater than either parent for the two-parent populations, while there was transgressive segregation observed for seed size of the $F_{2:3}$ lines in set three. Transgressive segregation also was observed for seed size in the three-parent crosses from set one and three, while $F_{2:3}$ lines of set two did not possess seed size equal to or greater than any parents of the cross.

Leroy et al. (1991) calculated heritabilities based on $F_{2:3}$ lines developed from *G. max* × *G. soja* crosses. They reported average heritabilities estimates combined over three crosses of 35% on a plant, 52% on a plot, and 89% on an entry-mean basis.

Johnson et al. (2001) compared three population types including a small-seeded x small-seeded two-parent population, a small-seeded x normal-size two-parent population, and a (small-seeded x normal-size) x small-seeded three-parent population to determine which population type produced a sufficient number of small-seeded segregates. They reported that 90% of the lines in the small-seeded x small-seeded populations had seed size equal to or smaller than one of the parents in the cross, while only 4% of the lines from the small-seeded x normal-size populations and only 20% of the lines from the three-parent populations had seed size equal to or smaller than one of the small-seeded parents used to develop the populations. They also found that 10% of the lines from the small-seeded x small-seeded populations had significantly smaller seed size than either of the small-seeded parents used to develop the population. No transgressive segregation was observed in either the small-seeded x normal-seeded or three-parent populations. Johnson et al. (2001) concluded that to develop small-seeded cultivars with adequate seed size, small-seeded x small-seeded or three-parent populations would be preferred.

In summary, research on the heritability of seed size of soybean indicated that the range was from 27 - 89% depending on the unit of evaluation and the population studied. Marker-assisted selection will be most effective for traits with low to mid-range heritabilities, such as seed size, when large portions of their variability can be explained by the markers (Lamkey and Lee, 1999). The studies indicated that developing populations with parents that differ in seed size would result in segregates that have a size between that of the two parents, as required to associate the molecular makers with SSQTL.

DNA Marker Systems

Several DNA marker systems have been used to identify QTL in soybean. These systems include random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), single nucleotide polymorphism (SNP), and simple sequence repeats (SSR).

RAPD technology utilizes short oligonucleotide primers of 9 to 10 base pairs (bp) to amplify genomic regions by polymerase chain reaction (PCR) (Waugh and Powell, 1992). The number of PCR products generated depends on the length of the primer, size of the target genome, and the probability that the complementary sequences are present on both strands in opposite orientation. RAPD loci exhibit dominant rather than the codominant inheritance observed for RFLP and SSR alleles.

RFLP technology is based on the variation of DNA length between two restriction sites (Russell, 1996). Southern analysis is used to detect the size differences in RFLPs. RFLP markers have codominant inheritance that makes it possible to detect both alleles at a locus in a heterozygote.

AFLP markers combine elements from both RAPD and RFLP marker systems. Double stranded DNA is digested with two enzymes to create different fragment ends. Oligomer adapters are ligated onto the ends and the fragment is amplified by PCR. The fragments are either radioactively or fluorescently labeled, separated on a polyacrylamide gel, and scored for the presence or absence of the polymorphic fragments. The AFLP marker system requires small amounts of DNA, is highly repeatable, and can detect numerous loci per reaction, which makes AFLP markers well-suited for genomic map construction.

SNP technology is a relatively new compared with the other marker systems. SNPs are molecular makers that possess a single base pair variation between two otherwise identical DNA sequences. This variation can be expressed either as a deletion, an insertion, or a substitution. One of the potential benefits of SNPs is that they occur very frequently within the genome. This is beneficial when conducting molecular research because there may be a higher the likelihood of finding significant differences between individuals (Kwok and Gu, 1999). A second potential benefit is that the mutation rate of SNPs is low from generation to generation (Kwok and Gu, 1999). This allows scientists to conduct more accurate population studies when the goal is to map gene location. A third potential benefit is that SNPs are often linked to genes (Kwok and Gu, 1999).

The first step in developing SNPs is to sequence the DNA surrounding the SNP. This step is essential because the sequence is necessary to develop primers or oligonucleotide probes that can be used to create a sequence-tagged site (STS). An STS is a segment of DNA that can be amplified by PCR and is unique within the genome. To identify a SNP, the STSs of individuals expressing different alleles are compared. Once a single nucleotide polymorphism has been located, it must be mapped to a specific chromosomal location. Mapping can be done in a number of ways, including the linkage disequilibrium method. Researchers must develop a genotyping assay to use in

experiments involving the SNP because identifying a marker is relatively useless without the ability to easily screen for it in genetic studies (Kwok and Gu, 1999).

SSR markers are composed of a 1 to 6 bp DNA sequence that is repeated a variable number of times (Litt and Luty, 1989). Regions that flank the SSR are usually highly conserved, and complementary primers can be developed that amplify the SSR (Ashley and Dow, 1994). The variation in the number of tandem repeats results in the different PCR product length (Litt and Luty, 1989).

SSRs have advantages over RAPD, AFLP, and RFLP marker systems. They have codominant inheritance instead of the dominant inheritance of RAPDs. They utilize PCR to amplify the DNA, which makes it possible to extract smaller quantities of DNA from an individual than is required for RFLPs. They exhibit a higher level of polymorphism than RFLPs. As many as 26 alleles have been reported at a SSR locus, whereas RFLPs with more than two alleles are rarely identified (Cregan et al., 1994; Akkaya et al., 1995). AFLPs are useful in filling in gaps on the molecular map, however, it is difficult to compare AFLP markers across genetic maps, which is an advantage for using SSR markers. Because of the aforementioned advantages, SSRs were the logical marker of choice for my research.

Genetic Mapping

The reason for identifying the map location of genes is to allow researchers to study gene function, regulation, expression, and interactions. The first step in developing a map is population development. Parent selection is important for obtaining a broad range of segregation in a population. Ideally, the parents selected should be at opposite ends of the spectrum for the traits of interest to ensure adequate segregation. Second, the parents used to develop the population have to be genotyped. Because genetic maps are based on DNA polymorphism, markers that are

polymorphic between the parents will be used in the genetic analysis. Third, marker selection is very important. Markers should be polymorphic and low in copy number to facilitate scoring (Dudley, 1993). Finally, the appropriate population size must be determined. Larger population sizes provide more accuracy in calculating linkage estimates because more segregates are recovered for each genotypic class.

In addition to the aforementioned criteria needed to conduct a genetic mapping study, software packages are required to analyze the data. There are three main QTL software packages: MAPMAKER, QTL Cartographer, and PLABQTL. MAPMAKER is a computer package used to construct genetic linkage maps and the subsequent mapping of the gene(s) for the traits of interest using those linkage maps. MAPMAKER contains two programs MAPMAKER/EXP and MAPMAKER/QTL. MAPMAKER/EXP is the program that performs the linkage analysis to construct the primary linkage maps (Lander et al., 1987; Lincoln et al., 1992a). MAPMAKER/EXP conducts multi-point linkage analysis considering all of the raw genotypic data simultaneously in each computation to find map order and map distances. MAPMAKER/QTL is a program that utilizes the genetic linkage maps constructed in MAPMAKER/EXP to map genes controlling polygenic quantitative traits (Paterson et al., 1988; Lincoln et al., 1992b). MAPMAKER/QTL utilizes interval mapping, which uses maximum likelihood to map the genes underlying the quantitative traits segregating in the population. QTL Cartographer (Basten et al., 1994; Basten et al., 2001) and PLABQTL (Utz and Melchinger, 1996) are a suite of programs for mapping QTLs onto a genetic linkage map. These programs use linear regression, interval mapping, or composite interval mapping.

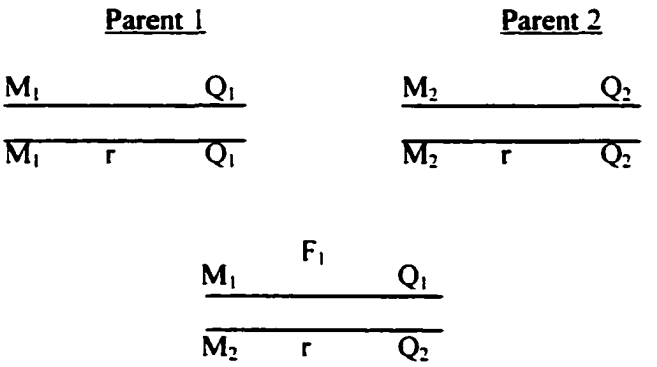
The premise behind using molecular markers to map QTLs is that by crossing two inbred lines, linkage disequilibrium is created between the loci that differ in the lines, which in turn creates associations between the marker loci and linked segregating QTLs (Lynch and Walsh,

1998). There are two main mating designs used to generate this disequilibrium, F_2 and backcross populations. The F_2 design examines the marker-trait associations in the progeny of a cross obtained by selfing the F_1 plants. The backcross design examines the marker-trait associations in the progeny formed by backcrossing the F_1 plants to one of the parents (Lynch and Walsh, 1998). Lynch and Walsh (1998) explained additional populations that can be derived, such as recombinant inbred lines and doubled haploid lines, which create a homozygous background from which to examine marker-trait associations. F_2 populations were used in my study. The main advantage of F_2 populations over the previously mentioned population types is that three genotypic classes are generated for each marker locus, which makes it possible for dominance effects to be estimated for the given QTL.

There are three methods used to generate marker-trait associations: single-factor analysis, interval mapping, and composite interval mapping. In single-factor analysis, the distribution of the phenotypic values are examined separately for each marker locus. Single-factor analysis is a good choice for the detection of a QTL linked to a marker; however, it is not as precise in the estimation of position, amount of phenotypic variation, and additive and dominance effects as the other two methods (Lynch and Walsh, 1998). To understand single-factor ANOVA, a genetic model has to be developed that describes the different marker loci and QTL genotype combinations. A simple genetic model assumes that there are two alternative alleles at each QTL that are segregating in a population, Q_1 and Q_2 (Falconer and Mackay, 1996). The genotypic values for each trait can calculated as follows:

QTL genotype	Genotypic value
Q_1Q_1	$m + a$
Q_1Q_2	$m + d$
Q_2Q_2	$m - a$

Using the notation given in Falconer and Mackay (1996), $+a$ and $-a$ are the additive gene effects that correspond to the deviations of the homozygotes from the mid-parent value at the QTL. Dominance deviations are denoted by the symbol d and refer to the deviation of the heterozygote from the mid-parent value. When developing an F_2 population, the assumption is made that the parent lines used to develop the population are completely inbred and that there is a QTL, Q, linked to a marker locus with a recombination frequency of r between the marker and the QTL. Considering a single locus, the parental genotypes are as follows:



The F_1s are selfed and the resulting F_2 population is as follows:

Genotype	Genotypic value	Frequency
$M_1M_1Q_1Q_1$	+a	$\frac{1}{4}(1-r)^2$
$M_1M_1Q_1Q_2$	+d	$\frac{1}{2}r(1-r)$
$M_1M_1Q_2Q_2$	−a	$\frac{1}{4}r^2$
$M_1M_2Q_1Q_1$	+a	$\frac{1}{2}r(1-r)$
$M_1M_2Q_1Q_2$	+d	$\frac{1}{4}[(1-r)^2+r^2]$
$M_1M_2Q_2Q_2$	−a	$\frac{1}{2}r(1-r)$
$M_2M_2Q_1Q_1$	+a	$\frac{1}{4}r^2$
$M_2M_2Q_1Q_2$	+d	$\frac{1}{2}r(1-r)$
$M_2M_2Q_2Q_2$	−a	$\frac{1}{4}(1-r)^2$

The information above is needed to derive the expected genotypic values for each of the three genotypes at a marker locus, as shown below.

Marker genotype	Genotypic value	Frequency
M_1M_1	$a[(1-r)^2 - r^2] + 2d[r(1-r)]$	¼
M_1M_2	$d[(1-r)^2 + r^2]$	½
M_2M_2	$-a[(1-r)^2 - r^2] + 2d[r(1-r)]$	¼

Contrasts can be made to determine if a QTL is present and estimates of their additive and dominance effects can be made.

The contrast between the two homozygous classes is

E_1: $E(M_1M_1 - M_2M_2) = 2a[(1-r)^2 - r^2] = 2a(1-2r)$

The contrast between the heterozygote and the mid-parent is

E_2: $E(M_1M_2 - ½ [M_1M_1 + M_2M_2]) = d[(1-r)^2 + r^2 - 2d[r(1-r)] = d(1-2r)^2$

If the marker is not linked to the QTL, $r = ½$ and the expected contrasts are $E_1 = 0$ and $E_2 = 0$.

There are two disadvantages associated with detecting QTL using single-factor ANOVA. Edwards et al. (1987) discovered that QTL estimates were underestimated and confounded with the recombination frequency between the marker and the QTL. As a result, it is difficult to distinguish between the effect of a small QTL located close to the marker or the effect of a large distant QTL. Another disadvantage is that single-factor ANOVA does not have the capability to pin-point the location of the QTL. This drawback can be lessened by having a densely populated molecular linkage map.

Lander and Botstein (1989) developed the maximum likelihood method for QTL detection, called interval mapping. Interval mapping utilizes two locus marker genotypes to derive marker

associations. A separate analysis is conducted on each individual pair of marker loci resulting in n-1 separate tests (Lynch and Walsh, 1998). The log-odds (LOD) analysis is used to provide an estimate of the QTL location and effect. This method estimates the location of the QTL from the distributions associated with the trait of interest within each marker genotype class and the mean differences between the genotype class of the flanking markers. The advantage of interval mapping over single-factor analysis is that interval mapping increases the power of QTL detection and estimation of position effects. However, interval mapping is only accurate when one QTL is segregating in the interval between the flanking markers (Lander and Botstein, 1989).

Composite interval mapping is similar to interval mapping in that a separate test is conducted on each interval pair; however, additional well-chosen markers around the interval in question also are included in the analysis (Zeng, 1994). Composite interval mapping utilizes maximum likelihood and multiple linear regression to locate QTL. The additional flanking markers included in the analysis decrease the bias that can be caused by multiple QTL linked to the marker interval under evaluation (Zeng, 1994). These additional flanking markers are called cofactors. Zeng (1994) used step-wise regression to select the important markers as cofactors to increase the power and precision with which to detect QTL. It is this increase in power and precision that is the main advantage in using composite interval mapping.

Past QTL Experiments Involving Seed Size

Mian et al. (1996) developed two *G. max* populations that were developed from four parents of normal seed size. F_4-derived lines of 'Young' (160 mg seed^{-1}) × PI416937 (174 mg seed^{-1}) (Pop 1) were mapped with 155 RFLP markers and lines from 'Coker' (147 mg seed^{-1}) × PI97100 (128 mg seed^{-1}) (Pop 2) were mapped with 153 RFLP markers. Pop 1 was grown in three environments during 1994 (Plains, Plymouth, and Windblow, GA) and Pop 2 was grown in one

environment during 1994 (Athens, GA) and two environments in 1995 (Athens and Blackville, GA). Based on single-factor analysis of variance (ANOVA), seven independent loci were associated with seed size for Pop 1 and explained 73% of the phenotypic variation. In Pop 2, nine independent loci were associated with seed size, which explained 74% of the phenotypic variation. Markers associated with seed size were highly consistent across environments and years indicating the potential effectiveness of marker-assisted selection for seed size.

Maughan et al. (1996) developed a population by crossing a *G. max* breeding line V71-370 (240 mg seed^{-1}) with a *G. soja* plant introduction PI407.162 (15 mg seed^{-1}). F_2-derived lines were analyzed with 91 polymorphic markers, including RFLPs, RAPDs, and SSRs. Markers were associated with seed size using single-factor ANOVA and the computer program Mapmaker/QTL. Three markers were associated with seed size and explained 50% of the phenotypic variation among the F_2 plants, while five markers were associated with seed size and explained 60% of the phenotypic variation among the $F_{2:3}$ lines.

Mansur et al. (1996) performed QTL analysis for agronomically important traits on 284 F_7-derived lines developed from the cross between Minsoy (130 mg seed^{-1}) and Noir 1 (140 mg seed^{-1}). They constructed a molecular map that was 1981 cM in length using RFLPs, SSRs, and classical markers. They used Mapmaker v. 3.0 to construct the linkage maps, and QTL positions were determined by analysis of variance using SAS. Three markers were associated with seed size and explained 23.1% of the phenotypic variation among the lines. They concluded that the majority of the traits they studied were controlled by a few loci with major effects instead of the traditionally held theory that quantitative traits are governed by a large number of loci having small effects.

Orf et al. (1999) performed QTL analysis on three populations derived from Archer, Minsoy, and Noir 1 (NA = population developed from the cross Noir 1 x Archer, MA = population

developed from the cross Minsoy x Archer, and MN = population developed from the cross Minsoy x Noir 1). The study focused mainly on important agronomic traits. They found that many of the traits clustered on three of the 20 linkage groups. They found seven markers associated with SSQTL in the NA population that accounted for 50% of the phenotypic variation, seven markers associated with SSQTL in the MN population that accounted for 50% of the phenotypic variation, and two markers associated with SSQTL in the MA population that accounted for 12% of the phenotypic variation. They found that only one QTL was detected in two populations. They concluded that genetic background was important for QTL expression.

One of the main objectives in the study of Sebolt et al. (2000) was to evaluate the effect of a *G. soja* QTL for increased seed protein on other seed traits in different genetic backgrounds. They developed a backcross population that was initially used to determine QTL position and effect. Test populations were developed to study different genetic background effects by crossing a line from the backcross population to three soybean genotypes; 'Parker', 'Kenwood', and C1914. According to the backcross data, the *G. soja* allele for IaSU-A144H-1 was associated across years with reduced seed size. Data from the test populations showed that seed size was significant across years and locations in two of the populations. This research showed that seed component traits can be modified through genetic mapping coupled with marker-assisted selection. They were able to backcross *G. soja* genes into a soybean genotype within 1 yr when it has typically taken much longer. To utilize marker-assisted selection in this manner, they indicated that the genes of interest must be mapped prior to backcrossing, which is not required in traditional backcrossing.

Significant QTL x Environment Interactions for SSQTL

The significance of marker genotype x environment interactions have been studied using analysis of variance procedures or by comparing the frequency of identification of significant marker-QTL

associations in different environments (Dudley, 1993). Maughan et al. (1996) reported that five of six RFLP markers were significantly associated with soybean seed size in the F_2 and $F_{2:3}$ generations. Mian et al. (1996) reported that six of seven RFLP marker loci were consistent for Pop 1 (Young × PI416937) across three locations during 1 yr. They also found that six of nine RFLP marker loci were consistent for Pop 2 (PI97100 × Coker 237) across two locations and two years. Different markers were used to identify the QTL associations in these two studies. Of the 21 RFLP markers that were associated with seed size QTL in these two studies, nine markers were located in close proximity to each other on three chromosomes. Maughan et al. (1996) identified four RFLP markers and Mian et al. (1996) identified five RFLP markers associated with seed size that were located on linkage group G, J, and L, suggesting that these markers are probably associated with the same seed size QTL. However, the other nine markers were associated with different seed size QTL. Small populations were used by Mian et al. (1996) ($N_{(Pop\ 1)} = 120$, $N_{(Pop\ 2)} = 111$) and by Maughan et al. (1996) ($N_{(F2\ and\ F2:3)} = 150$); therefore, only major QTL could be identified. Additional research should be conducted to identify QTL associated with seed size in different populations of soybean.

Marker-Assisted Breeding

Marker-assisted selection uses molecular information to assist in the selection of parents for crossing and in selection among segregates in a population. There are three potential benefits from utilizing molecular markers in a breeding program. First, individuals can be objectively screened using molecular markers and the subjective nature of phenotypic selection can be minimized. Second, the parents can be screened before hybridization, theoretically increasing the amount of genetic gain. Third, individuals from a segregating population can be screened at any generation to identify the best progeny for field evaluation (Lamkey and Lee, 1999).

For marker-assisted selection to be effective, four criteria must be met. First, the molecular map of the species of interest must be highly saturated (Dudley, 1993). Second, the markers must be easy to use and cost effective (Dudley, 1993). Third, the genetic variance explained by the markers should exceed the heritability of the trait. Fourth, the marker must be associated with QTL in different populations.

My study incorporated aspects from the previous studies by developing three single-cross populations from *G. max* parents with normal and small seed size. The cross between a normal-size and a small-seeded parent resulted in segregation for seed-size between the two parents (Johnson et al., 2001), which is ideal for detecting QTL (Dudley, 1993). In any QTL mapping study, a trade-off exists between identifying all QTL present in a single large population or identifying the major QTL in a number of small populations. In my study, 100 F_2-derived lines from each of three populations were used to identify the major SSQTL.

References

Akkaya, M.S., R.C. Shoemaker, J.E. Specht, A.A. Bhagwat, and P.B. Cregan. 1995. Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci. 35:1439–1445.

Ashley, M.V., and B.D. Dow. 1994. The use of microsattellite analysis in population biology: background, methods, and potential applications. p. 185–201. *In* B. Shierwater, B. Streit, and R. Daselle (ed.) Molecular ecology and evolution: approaches and applications. Verlag, Basel, Switzerland.

Basten, C.J., B.S. Weir, and Z.-B. Zeng. 1994. Zmap-a QTL cartographer. p. 65-66. *In* C. Smith, J.S. Gavora, B. Benkel, J. Chesnais, W. Fairfull, J.P. Gibson, B.W. Kennedy, and E.B. Burnside (ed.) Proceedings of the 5th World Congress on Genetics Applied to Livestock

Production: Computing Strategies and Software. 5th World Congress on Genetics Applied to Livestock Production, Guelph, Ontario, Canada.

Basten, C.J., B.S. Weir, and Z.-B. Zeng. 2001. QTL Cartographer, Version 1.15. Department of Statistics, North Carolina State University, Raleigh, NC.

Beavis, W.D., D. Grant, M. Albertsen, and R. Fincher. 1991. Quantitative trait loci for plant height in four maize populations and their associations with quantitative genetic loci. Theor. Appl. Genet. 83:141–145.

Bravo, J.A., W.R. Fehr, and S.R. Cianzio. 1980. Use of indirect selection of seed weight in soybeans. Crop Sci. 20:507–510.

Bravo, J.A., W.R. Fehr, and S.R. Cianzio. 1981. Use of small-seeded soybean parents for the improvement of large-seeded cultivars. Crop Sci. 21:430–432.

Brim, C.A., and C.C. Cockerham. 1961. Inheritance of quantitative characters in soybean. Crop Sci. 1:187–190.

Buhr, K.L. 1976. Inheritance of timing to flower, time to physiological maturity, and growth habit in soybeans grown at a tropical latitude. Ph.D. Dissertation. Iowa State Univ. Univ. Microfilms. Ann Arbor, Mich. (Diss Abstr. 37/02B:552).

Carpenter, J.A., and W.R. Fehr. 1986. Genetic variability for desirable agronomic traits in populations containing *Glycine soja* germplasm. Crop Sci. 26:681–686.

Cianzio, S.R., and W.R. Fehr. 1987. Inheritance of agronomic and seed composition traits in Glycine max x Glycine soja crosses. J. Agric. Univ. P. R. 71:53–63.

Cregan, P. B., A. A. Bhagwat, M. S. Akkaya, and J. Rongwen. 1994. Microsattellite fingerprinting and mapping of soybean. Meth. Mol. Cell. Biol. 5:49–61.

Dudley, J.W. 1993. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci. 33:660–668.

Edwards, M.D., C.W. Stuber, and J.F. Wendel. 1987. Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and type of gene action. Genetics 116:113–125.

Falconer, D.S., and T.F.C. Mackay. 1996. Quantitative trait loci. p. 356–378. *In* Introduction to quantitative genetics. Longman Group Ltd., Essex, England.

Fehr, W.R. 1987. Soybean. p. 533–576. *In* W.R. Fehr (ed.) Principles of cultivar development. Vol. 2. Macmillan Publishing Company, New York, NY.

Hartwig, E.E. 1973. Varietal development. p. 187–207. *In* B.E. Caldwell (ed.) Soybeans: Improvement, production, and uses. ASA, CSSA, and SSSA, Madison, WI.

Kwok, P.-Y., and Z. Gu. 1999. SNP libraries: why and how are we building them? *In* Molecular Medicine Today 12:538–543.

Johnson, S.L., W.R. Fehr, G.A. Welke, and S.R. Cianzio. 2001. Genetic variability for seed size of two- and three-parent soybean populations. Crop Sci. 41:1029–1033.

Lamkey, K.R., and M. Lee. 1999. Quantitative genetics, molecular markers, and plant improvement. http://corn2.agron.iastate.edu/publications/.

Lander, E., P. Green, J. Abrahamson, A. Barlow, M. Daly, S. Lincoln, and L. Newburg. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181.

Lander, E., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitaive traits using RFLP linkage maps. Genetics 121:185–199.

Leroy, A.R., S.R. Cianzio, and W.R. Fehr. 1991. Direct and indirect selection for small seed of soybean in temperate and tropical environments. Crop Sci. 31:697–699.

Lincoln, S., M. Daly, and E. Lander. 1992a. Constructing genetic maps with MAPMAKER/EXP 3.0. Whitehead Institute Technical Report. 3rd edition., Cambridge, MA.

Lincoln, S., M. Daly, and E. Lander. 1992b. Mapping genes controlling quantitative traits with MAPMAKER/QTL 1.1. Whitehead Institute Technical Report. 2rd edition., Cambridge, MA.

Litt, M., and J. A. Luty. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. 44:397–401.

Lynch, M., and J.B. Walsh. 1998. Genetics and analysis of quantitative traits. 1st ed. Sinauer Associates, Inc., Sunderland, MA.

Mansur, L.M., J.H. Orf, K. Chase, T. Jarvik, P.B. Cregan, and K.G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci. 36:1327–1336.

Maughan, P.J., M.A. Saghai Maroof, and G.R. Buss. 1996. Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species. Theor. Appl. Genet. 93:574–579.

Mian, M.A.R., M.A. Bailey, J.P. Tamulonis, E.R. Shipe, T.E. Carter Jr., W.A. Parrott, D.A. Ashley, R.S. Hussey, and H.R. Boerma. 1996. Molecular markers associated with seed weight in two soybean populations. Theor. Appl. Genet. 93:1011–1016.

Orf, J.H., K. Chase, T. Jarvik, L.M. Mansur, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Sci. 39:1642–1651.

Paterson, A., E. Lander, J. Hewitt, S. Peterson, S. Lincoln, and S. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726.

Russell, P.J. 1996. Genetics. 4th ed. HarperCollins College Publishers, New York, NY.

Sebolt, A.M., R.C. Shoemaker, and B.W. Diers. 2000. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci. 40:1438–1444.

Shoemaker, R.C. and J.E. Specht. 1995. Integration of the soybean molecular and classical genetic linkage groups. Crop Sci. 35:436–446.

Stuber, C.W. 1992. Biochemical and molecular markers in plant breeding. p. 37–61. *In* J. Janick (ed.) Plant breeding reviews. John Wiley & Sons, Inc., New York, NY.

Ting, C.L. 1946. Genetic studies on the wild and cultivated soybeans. J. Am. Soc. Agron. 38:381–393.

Utz, H.F. and A.E. Melchinger. 1996. PLABQTL: A program for composite interval mapping of QTL. J. Quantitative Trait Loci (http://probe.nalusda.gov:8000/otherdocs/jqtl/jqtl1996-01/utz.html; verified 13 May 1999).

Waugh, R. and W. Powell. 1992. Using RAPD markers for crop improvement. Trend. Biotech. 10:186–191.

Weber, C.R. 1950. Inheritance and interrelation of some agronomic and chemical characters in an interspecific cross in soybeans, *Glycine max* x *G. ussuriensis*. Iowa Agric. Exp. Stn. Bull. 374.

Weber, C.R. and B.R. Moorthy. 1952. Heritable and nonheritable relationships and variability of oil content and agronomic characters in the F_2 generation of soybean crosses. Agron. J. 44:202–209.

Zeng, Z.-B. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457–1468.

CHAPTER 2. MOLECULAR MARKER ANALYSIS OF SEED SIZE IN SOYBEAN

A paper submitted for publication in *Crop Science*

Joseph A. Hoeck, Walter R. Fehr,* Randy C. Shoemaker, Grace A. Welke,

Susan L. Johnson, and Silvia R. Cianzio

Abstract

Seed size is an important attribute of soybean [*Glycine max* (L.) Merr.] for some food uses. The objectives of this study were to identify markers associated with quantitative trait loci for seed size (SSQTL), determine the influence of the environment on expression of the marker-SSQTL associations, and compare the efficiency of phenotypic selection and marker-assisted selection for the trait. Three small-seeded lines were crossed to a line or cultivar with normal seed size to form three two-parent populations. The parents of the populations were screened with 178 simple sequence repeat (SSR) markers to identify polymorphism. Population 1 (Pop 1) had 75 polymorphic SSR markers covering 1306 cM, population 2 (Pop 2) had 70 covering 1143 cM, and population 3 (Pop 3) had 82 covering 1237 cM. Seed size of each population was determined with 100 F_2 plants grown at Ames, IA, and their F_2-derived lines grown in two replications at three environments. Single-factor analysis of variance and multiple regression were used to determine significant marker-SSQTL associations. Pop 1 had 12 markers that individually accounted for 8 to 17% of the variation for seed size, Pop 2 had 16 markers that individually accounted for 8 to 38% of the variation, and Pop 3 had 22 markers that individually accounted for 8 to 29% of the variation. Four of the 12 markers in Pop 1, four in Pop 2, and one in Pop 3 had significant associations with SSQTL across four environments, while five loci in Pop 1, seven in Pop 2, and eight in Pop 3 had significant associations in more than one environment. Three marker loci that

had significant SSQTL associations in this study also were significant in previous research, and 13 markers had unique SSQTL associations. The relative effectiveness of phenotypic and marker-assisted selection among F_2 plants varied for the three populations. Averaged across the three populations, phenotypic selection for seed size was as effective and less expensive than marker-assisted selection.

Introduction

Seed size is an important trait for production of some specialty soy food products, including tofu, natto, miso, and edamame. Seed size of *G. max* is quantitatively inherited and ranges from 40 to 550 mg seed^{-1} (Hartwig, 1973). Plant breeders select for improved yield and other desirable agronomic traits in developing soybean cultivars with different seed sizes for various food products. The traditional method of soybean breeding involves artificial hybridization to develop genetic variability followed by self-fertilization and phenotypic selection for seed size among the offspring. Molecular markers may improve traditional methods of breeding for seed size by increasing the reliability with which desirable progeny are selected.

Molecular marker associations with quantitative trait loci for seed size (SSQTL) of soybean have been reported. Mian et al. (1996) developed two *G. max* populations utilizing normal-size parent lines. They identified 16 independent marker loci that were significantly associated with SSQTL that together explained 73 to 74% of the phenotypic variation in each of the two populations. None of their marker loci was significantly associated with SSQTL across both populations. Twelve of the 16 marker loci were significantly associated with SSQTL in all environments, three were significant in two environments, and one was significant in only one environment. Maughan et al. (1996) developed a population by crossing a *G. max* line with a seed size of 240 mg seed^{-1} to an accession of wild soybean [*Glycine soja* (L.) Sieb. & Zucc.] with a seed

size of 15 mg seed^{-1}. Three molecular markers were associated with SSQTL in F_2 plants that explained 50% of the phenotypic variation, while five markers were associated with SSQTL for $F_{2:3}$ lines that explained 60% of the variation. Mansur et al. (1996) observed three molecular markers that explained 23.1% of the variation for seed size among $F_{2:7}$ lines developed from the cross between 'Minsoy' (130 mg seed^{-1}) and 'Noir 1' (140 mg seed^{-1}). Orf et al. (1999) found seven marker loci that accounted for 50% of the variation for seed size among $F_{2:7}$ lines in the cross of Noir 1 × 'Archer', seven in a Minsoy × Noir 1 population that accounted for 50%, and two in a Minsoy × Archer population that accounted for 12% of the variation. Only one molecular marker was significantly associated with SSQTL in the three populations.

This study was conducted to identify additional molecular markers associated with SSQTL, to evaluate simple-sequence-repeat (SSR) markers that were previously reported or located closely to previously reported loci that have been associated with SSQTL, and to compare the effectiveness of phenotypic selection and marker-assisted selection (MAS), for seed size among F_2 plants.

Materials and Methods

Three single-cross populations were developed from six *G. max* cultivars for this study. The parents with normal seed size were 'S12-49' developed by the Northrup King Co., Washington, IA, and A96-492041 and A96-492058 developed by Iowa State University. The small-seeded parents, A97-775019, A97-775006, and A97-775026, were developed by Iowa State University. Population 1 (Pop 1) was from the cross A97-775019 × A96-492041, population 2 (Pop 2) from A97-775006 × S12-49, and population 3 (Pop 3) from A97-775026 × A96-492058. The crosses were made in March 1998 at Iowa State University-University of Puerto Rico soybean breeding nursery in Isabela, Puerto Rico. The F_1 seeds were planted in May 1998 at the Agronomy

Research Center near Ames, IA. Pubescence color was used to confirm that F_1 plants of the populations originated from hybrid seed. The F_2 and parent seed of each population were planted in February 1999 at Isabela. The soil type is a Coto clay (Very-fine, koalinitic, isohyperthermic, Typic Haplorthox). The 200 F_2 seeds of each population and 40 seeds of each parent were planted ≈15 cm apart in rows 102 cm wide under artificial lights to extend the day length for increased seed production. The F_2 and parent plants were harvested individually. Seed size of 10 random parent plants and 100 random F_2 plants from each of the populations was measured in mg seed^{-1} by dividing the weight of all the seeds by the number of seeds.

For each population, a set of the 100 $F_{2:3}$ lines derived from the F_2 plants and 10 entries of each of the parents were evaluated in 1999 as a separate experiment. The 120 entries in a set were planted in a randomized complete-block design with two replications on 24 May 1999 at the Burkey Farm and on 26 May 1999 at the Agronomy Research Center near Ames. The soil type at both locations is a Nicollet loam (Fine-loamy, mixed, superactive, mesic Aquic Hapludoll). The entries were grown in single-row plots 76 cm long with 102 cm between rows and a 107-cm alley between the ends of the plots. The seeding rate was 12 seeds per plot. The plots were harvested in bulk with a self-propelled combine. The three sets of 120 entries were planted on 1 November 1999 at Isabela. Each set was planted under natural day length conditions in two replications of a randomized complete-block design. Each plot was a single row 61 cm long with 102 cm between rows and a 30-cm alley between the ends of the plots. The seeding rate was 16 seeds per plot. The plots were harvested by hand and threshed in bulk with a stationary belt thresher. Seed size was measured by weighing 400 random seeds from each plot in the three environments.

A 15 to 20 g sample of leaf material was collected at Isabela from at least 10 different plants of each entry. The leaf samples were placed in a plastic bag with an identification card and kept on ice until they were frozen in liquid nitrogen and dried in a vacuum for approximately 24 hr.

The dried leaf samples were stored at –20°C until DNA extraction. Dried leaf tissue was placed in 50-mL screw-cap polypropylene tubes containing ≈ 4 g of 1.5-mL glass beads. The leaf material was ground into a powder by agitation with a paint shaker. DNA was extracted from each sample using the CTAB protocol by Keim et al. (1988).

A total of 178 SSRs was used to evaluate the six parents of the three populations. Pop 1 had 75, Pop 2 had 70, and Pop 3 had 82 polymorphic markers. Each SSR marker had been mapped in soybean (Cregan et al., 1999). For each population, there was an average of four markers in each of the 20 linkage groups. Multiplex sets of nine markers were constructed based on the forward primer label and the allele size of the different markers as described by Narvel et al. (2000).

The multiplex sets were used to determine the marker alleles of the F_2-derived lines. All reagents for the marker analysis were obtained from Perkin-Elmer Applied Biosystems (PE/ABI, Foster City, CA). The final polymerase chain reaction (PCR) volume was 10 µL and consisted of 30 ng genomic DNA, 0.8 µl of 25 m*M* magnesium chloride, 0.8 µl of 10 m*M* dNTPs, 0.2 µl (1.0 unit) of AmpliTaq™ Gold DNA polymerase, 1.0 µl of GeneAmp® 10X PCR Buffer II, 1.0 µl of 5 p*M* forward/reverse primer, and 5.7 µl of sterile water. The quantity of primer used in each reaction was chosen to optimize PCR. PCR was conducted with GeneAmp® thermocyclers (PE/ABI) models 9600 or 9700. The PCR procedure was 95°C for 10 min followed by 35 cycles of 95°C for 25 s, 58°C for 25 s, and 72°C for 25 s, followed by a final extension at 72°C for 60 min (Narvel et al., 2000). A 1.5-µL sample from each PCR run was submitted to the DNA Facility at Iowa State University for analysis with a PE/ABI model 377 automated DNA sequencer. Electrophoresis was carried out at 3000 V for 2 hr. Data were collected using the DNA Sequencing Collection software version 2.5 (PE/ABI) and analyzed with GENESCAN™ Prism software version 2.1 (PE/ABI). SSR allele sizes were automatically estimated by GENOTYPER™

software version 2.0 (PE/ABI). Allele sizes not identified automatically were estimated manually from the electropherogram peaks.

The lines were scored based on the marker genotype of the parents. Lines that possessed a homozygous allele derived from the parent with normal seed size were scored as 0, lines that possessed alleles from both parents were scored as 1, and lines that possessed a homozygous allele derived from the small-seeded parent were scored as 2.

MAPMAKER/EXP v. 3.0 was used to test marker pairs for evidence of linkage, and two-point recombination values were calculated by maximum likelihood at a minimum LOD of 3.0 and a maximum recombination frequency of $\theta = 0.50$ using the GROUP command (Lander et al., 1987). The order of each group was determined using either the COMPARE or THREE POINT commands, and loci orders were confirmed using the RIPPLE command. Linkage maps were created using the Haldane map function.

Single-factor analysis of variance (GLM) was used to associate polymorphic markers with SSQTL (SAS Institute, 1992). Significant SSQTL associations for each population were identified when a marker at an individual environment was significant at $P \leq 0.01$ or significant at $P \leq 0.05$ across multiple environments. Interval mapping was not used because individual linkage groups were not fully saturated and many markers were unlinked (Lander and Botstein, 1989). Two-way analyses of variance were used to test for digenic interactions between markers significantly associated with SSQTL and all other marker loci using the program EPISTACY (Holland, 1998). Significant marker loci were combined in a multiple-locus regression model (REG) to determine their combined effect (SAS Institute, 1992).

Results and Discussion

There were significant ($P<0.01$) differences in seed size among the three environments at which the F_2-derived lines were evaluated (Table 1). The mean seed size of the six parents was 102 mg seed^{-1} at the Agronomy Research Center, 101 mg seed^{-1} at the Burkey Farm, and 137 mg seed^{-1} at Puerto Rico. Significant differences were present among the F_2-derived lines of each population at the three environments and combined across environments. The genotype × environment interactions were significant for each population. None of the F_2-derived lines had the same seed size as either of their parents based on the means combined across environments. The failure to recover lines with seed sizes similar to the parents was consistent with segregation reported for other small-seeded × normal-size soybean crosses (Weber, 1950; Buhr, 1976; Carpenter and Fehr, 1986; Johnson et al., 2001). The broad-sense heritabilities for the three populations ranged from 0.45 to 0.85 on the plot basis and from 0.76 to 0.93 on the entry-mean basis, which were consistent with previous heritability estimates for small-seeded × normal-size crosses (Bravo et al., 1980; Leroy et al., 1991; Johnson et al., 2001).

Of the 178 SSR markers used to evaluate the parents, 75 were polymorphic for Pop 1, 70 for Pop 2, and 82 for Pop 3. There were 60 markers in Pop 1 associated with 15 linkage groups for a coverage of 1306 cM, 60 markers in Pop 2 associated with 19 linkage groups for a coverage of 1143 cM, and 75 markers in Pop 3 associated with 19 linkage groups for a coverage of 1237 cM. Fifteen markers in Pop 1, 10 in Pop 2, and 7 in Pop 3 could not be associated with any of the previously established ISU-USDA linkage groups (Cregan et al., 1999).

In Pop 1, 12 SSRs had significant associations with SSQTL in one or more environments (Table 2). The 12 marker loci were on nine linkage groups. Four of the markers (Satt409, Satt322, Satt045, and Satt510) were significantly associated with SSQTL in each of the four environments and five markers (Satt070, Satt002, Satt154, Satt185, and Satt273) were present in more than one

Table 1. Mean seed size for 100 F_2 plants and their F_2-derived lines and variance component and broad-sense heritability estimates for the F_2-derived lines at three environments and combined across environments.

		Mean†				Variance component‡			Heritability	
Pop	Environment	NSD	SSD	Population	Range	$\sigma^2_g \pm$ SE	$\sigma^2_{ge} \pm$ SE	$\sigma^2_e \pm$ SE	Plot	Entry-mean
		----------mg seed^{-1}----------								
1	F_2 plants	157 ± 29	80 ± 10	110 ± 26	83 - 140					
	F_2-derived lines									
	Agronomy	134 ± 3	71 ± 3	102 ± 7	78 - 125	98 ± 15		21 ± 3	0.82 ± 0.13	0.90 ± 0.14
	Burkey	135 ± 4	66 ± 2	98 ± 7	71 - 123	113 ± 18		24 ± 3	0.82 ± 0.13	0.90 ± 0.14
	Puerto Rico	172 ± 5	101 ± 6	128 ± 8	96 - 171	153 ± 24		27 ± 3	0.85 ± 0.13	0.92 ± 0.14
	Combined	147 ± 2	79 ± 2	109 ± 4	81 - 137	110 ± 17	11 ± 3	24 ± 2	0.76 ± 0.11	0.93 ± 0.14
2	F_2 plants	163 ± 13	96 ± 11	124 ± 26	95 - 155					
	F_2-derived lines									
	Agronomy	132 ± 4	71 ± 3	101 ± 7	82 - 126	80 ± 13		25 ± 3	0.76 ± 0.13	0.87 ± 0.14
	Burkey	135 ± 4	66 ± 2	98 ± 6	77 - 129	94 ± 15		19 ± 3	0.83 ± 0.13	0.91 ± 0.14
	Puerto Rico	176 ± 6	97 ± 2	135 ± 9	105 - 170	123 ± 20		36 ± 5	0.77 ± 0.13	0.87 ± 0.14
	Combined	148 ± 3	78 ± 1	111 ± 4	88 - 135	86 ± 13	13 ± 3	27 ± 2	0.68 ± 0.11	0.91 ± 0.14

† Mean ± standard error of the mean for normal-size parent (NSD), small-seeded parent (SSD), and the population.

‡ Variance component estimates for genotypes (σ^2_g), genotype x environment interactions (σ^2_{ge}), and error (σ^2_e) and their standard errors (SE).

Table 1. Continued.

Pop	Environment	Mean			Range	Variance component			Heritability	
		NSD	SSD	Population		$\sigma^2_g \pm SE$	$\sigma^2_{ge} \pm SE$	$\sigma^2_e \pm SE$	Plot	Entry-mean
		----------mg seed^{-1}----------								
3	F_2 plants	180 ± 33	87 ± 14	124 ± 26	96 - 166					
	F_2-derived lines									
	Agronomy	135 ± 4	71 ± 2	98 ± 7	79 - 115	39 ± 7		22 ± 3	0.64 ± 0.12	0.78 ± 0.14
	Burkey	138 ± 3	67 ± 2	95 ± 6	75 - 124	79 ± 13		21 ± 3	0.79 ± 0.13	0.88 ± 0.14
	Puerto Rico	176 ± 6	101 ± 3	133 ± 10	96 - 159	76 ± 14		47 ± 7	0.62 ± 0.12	0.76 ± 0.15
	Combined	150 ± 4	80 ± 2	109 ± 4	89 - 129	45 ± 8	20 ± 4	30 ± 2	0.45 ± 0.09	0.80 ± 0.14

Table 2. Marker loci significantly associated with seed size of 100 F_2 plants grown at Ames, their F_2-derived lines grown at three environments, and combined across environments for population 1 using single-factor analysis of variance.

SSR locus	Linkage group†	Combined					Environment							
		Allelic mean‡					F_2 plants		Agronomy		Burkey		Puerto Rico	
		NN	NS	SS	P§	R^2#	P	R^2	P	R^2	P	R^2	P	R^2
		---------mg seed^{-1}---------				---%---		---%---		---%---		---%---		---%---
Satt409	A2	116	108	107	0.0030	11.6	0.0180	8.0	0.0050	10.6	0.0020	12.0	0.0100	9.2
Satt070	B2	114	110	105	0.0050	10.4	0.0050	10.3	0.0010	13.9	NS		0.0060	9.9
Satt322	C2	110	107	117	0.0010	14.6	0.0030	11.9	0.0010	14.0	0.0002	16.8	0.0070	10.1
Satt077	D1A				NS		NS		NS		0.0100	8.9	NS	
Satt002	D2	114	109	104	0.0170	8.1	NS		0.0030	11.1	0.0110	8.8	NS	
Satt154	D2	115	108	105	0.0020	12.5	NS		0.0010	12.8	0.0040	11.0	0.0060	10.0
Satt185	E	106	112	106	0.0170	8.5	0.0080	10.2	NS		0.0170	8.6	NS	
Satt045	E	106	113	106	0.0040	11.2	0.0010	12.9	0.0130	8.8	0.0020	12.0	0.0080	9.7
Satt510	F	115	110	104	0.0004	14.9	0.0060	9.9	0.0002	16.1	0.0030	15.2	0.0070	9.8
Satt001	K	107	109	117	0.0180	8.2	NS		NS		0.0150	8.6	NS	

† Linkage group as designated by the current USDA-ISU map.

‡ NN: homozygous normal-size parent, NS: heterozygous, SS: homozygous small-seeded parent.

§ Probability level, NS = not significant.

Percent phenotypic variation explained by the marker locus.

Table 2. Continued.

SSR locus	Linkage group	Combined					Environment							
		Allelic mean					F_2 plants		Agronomy		Burkey		Puerto Rico	
		NN	NS	SS	P	R^2	P	R^2	P	R^2	P	R^2	P	R^2
		---------mg seed^{-1}----------				---%---		---%---		---%---		---%---		---%---
Satt273	K	103	109	115	0.0010	13.5	NS		0.0010	14.0	0.0110	9.0	0.0010	13.7
Satt551	M				NS		NS		NS		NS		0.0100	8.8

Table 3. Marker loci significantly associated with seed size of 100 F_2 plants grown at Ames, their F_2-derived lines grown at three environments, and combined across environments for population 2 using single-factor analysis of variance.

SSR locus	Linkage group†	Combined					Environment							
		Allelic mean‡					F_2 plants		Agronomy		Burkey		Puerto Rico	
		NN	NS	SS	P§	R^2#	P	R^2	P	R^2	P	R^2	P	R^2
		---------mg seed^{-1}---------				---%---		---%---		---%---		---%---		---%---
Satt070	B2	117	111	105	0.0001	17.5	0.0002	16.2	0.0001	17.5	0.0001	21.0	0.0110	8.8
Satt534	B2				NS		0.0040	14.4	NS		0.0100	12.6	NS	
Satt565	C1	115	112	107	0.0130	9.0	NS		0.0100	9.8	NS		NS	
Satt227	C2				NS		NS		0.0170	8.1	0.0180	8.0	NS	
Satt277	C2	106	109	115	0.0020	12.5	NS		0.0070	10.1	0.0030	11.5	0.0060	10.4
Satt184	D1A				NS		NS		NS		0.0100	9.0	NS	
Sctt008	D2	108	114	107	0.0070	9.6	0.0060	9.9	NS		NS		0.0020	12.0
Satt135	D2				NS		0.0060	10.5	NS		NS		NS	
Satt185	E				NS		0.0160	8.5	0.0140	8.8	NS		NS	
Satt431	J				NS		NS		NS		NS		0.0040	10.9

† Linkage group as designated by the current USDA-ISU map.

‡ NN: homozygous normal-size parent, NS: heterozygous, SS: homozygous small-seeded parent.

§ Probability level, NS = not significant.

Percent phenotypic variation explained by the marker locus.

Table 3. Continued.

SSR locus	Linkage group	Combined					Environment							
		Allelic mean					F_2 plants		Agronomy		Burkey		Puerto Rico	
		NN	NS	SS	P	R^2	P	R^2	P	R^2	P	R^2	P	R^2
		---------mg seed^{-1}----------				---%---		---%---		---%---		---%---		---%---
Satt166	L	119	111	105	0.0001	23.2	0.0001	19.8	0.0001	20.1	0.0001	27.9	0.0010	13.4
Sat_099	L	122	111	104	0.0001	36.5	0.0001	22.4	0.0001	34.9	0.0001	37.7	0.0001	23.2
Satt006	L	120	110	105	0.0001	27.5	0.0020	11.8	0.0001	28.4	0.0001	30.4	0.0004	14.9
Satt373	L	115	111	107	0.0050	10.8	NS		0.0040	11.0	NS		0.0120	9.1
Satt336	M	113	113	106	0.0020	12.6	NS		0.0020	12.3	0.0080	9.4	0.0020	12.3
Satt173	O	111	113	106	0.0190	7.8	NS		0.0190	7.9	NS		NS	

Table 4. Marker loci significantly associated with seed size of 100 F_2 plants grown at Ames, their F_2-derived lines grown at three environments, and combined across environments for population 3 using single-factor analysis of variance.

SSR locus	Linkage group†	Combined					Environment							
		Allelic mean‡					F_2 plants		Agronomy		Burkey		Puerto Rico	
		NN	NS	SS	P§	R^2#	P	R^2	P	R^2	P	R^2	P	R^2
		---------mg seed^{-1}----------				---%---		---%---		---%---		---%---		---%---
Satt187	A2	111	109	104	0.0060	10.1	NS		NS		NS		0.0001	17.2
Satt304	B2	112	109	104	0.0002	16.0	NS		0.0030	11.4	NS		0.0001	17.1
Satt070	B2	112	109	104	0.0001	17.1	NS		0.0030	11.7	0.0100	9.2	0.0002	16.8
Sct_094	B2				NS		NS		NS		NS		0.0070	12.5
Satt565	C1	113	106	112	0.0010	15.1	NS		0.0010	14.5	0.0020	12.3	0.0100	9.4
Satt184	D1A				NS		NS		NS		NS		0.0100	9.4
Satt172	D1B				NS		NS		NS		NS		0.0090	9.2
Satt154	D2	111	111	104	0.0001	17.9	NS		0.0070	10.5	0.0010	15.1	0.0020	13.0
HSP176	F				NS		NS		NS		NS		0.0003	15.3
Satt114	F	110	111	105	0.0050	11.2	NS		NS		NS		0.0001	22.0

† Linkage group as designated by the current *USDA-ISU* map.

‡ NN: homozygous normal-size parent, NS: heterozygous, SS: homozygous small-seeded parent.

§ Probability level, NS = not significant.

Percent phenotypic variation explained by the marker locus.

Table 4. Continued.

SSR locus	Linkage group	Combined					Environment							
		Allelic mean					F_2 plants		Agronomy		Burkey		Puerto Rico	
		NN	NS	SS	P	R^2	P	R^2	P	R^2	P	R^2	P	R^2
		---------mg seed^{-1}---------				---%---		---%---		---%---		---%---		---%---
Satt334	F				NS		NS		NS		NS		0.0002	17.6
Satt510	F				NS		NS		NS		NS		0.0002	16.3
Satt072	F				NS		NS		NS		NS		0.0040	10.7
Sctt009	H	113	110	105	0.0010	13.3	NS		0.0050	10.6	0.0002	12.2	0.0080	9.6
Satt541	H	110	111	106	0.0090	9.7	NS		NS		0.0130	8.9	NS	
Satt314	H	113	110	106	0.0090	10.3	NS		NS		NS		0.0130	9.3
Satt302	H	112	107	107	0.0180	8.6	NS		0.0190	8.4	0.0160	8.7	NS	
Satt006	L	117	110	104	0.0001	28.8	NS		0.0001	20.9	0.0001	23.6	0.0002	17.8
Satt143	L	113	109	101	0.0001	20.0	NS		0.0070	9.8	0.0030	11.3	0.0001	22.6
Satt336	M	104	108	113	0.0002	18.3	0.0030	12.7	0.0004	16.9	0.0010	14.6	0.0080	10.8
Satt009	N				NS		NS		NS		0.0060	10.1	NS	
Satt237	N				NS		NS		NS		NS		0.0049	10.7

Table 5. Marker loci significantly associated with seed size of 100 F_2 plants, their F_2-derived lines at three environments, and combined across environments for three soybean populations using multiple regression.

					Environment							
			Combined		F_2 plants		Agronomy		Burkey		Puerto Rico	
	SSR	Linkage		Partial		Partial		Partial		Partial		Partial
Population	locus	group†	P‡	R^2§	P	R^2	P	R^2	P	R^2	P	R^2
				----%----		---%---		---%---		----%----		-----%-----
1	Satt409	A2	0.0020	7.7	NS		0.0180	4.0	0.0128	5.2	0.0038	6.9
	Satt070	B2	0.0329	3.0	0.0075	7.6	0.0030	7.1	NS		0.0181	4.1
	Satt322	C2	0.0164	4.0	0.0114	6.0	0.0220	3.5	0.0363	3.3	0.0193	4.2
	Satt077	D1A	NS		NS		NS		0.0454	2.9	NS	
	Satt154	D2	0.0009	9.8	NS		0.0166	4.3	0.0021	8.5	0.0026	8.2
	Satt045	E	NS		0.0085	6.9	NS		NS		NS	
	Satt510	F	0.0107	4.9	0.0136	5.4	0.0008	10.0	0.0001	15.1	NS	
	Satt273	K	0.0001	14.8	NS		0.0001	15.5	0.0274	3.9	0.0001	15.1
Total variation explained				44.2		25.9		44.4		38.9		38.5

† Linkage group as designated by the current USDA-ISU map.

‡ Probability value, NS = not significant.

§ Percent phenotypic variation explained by the marker locus.

Table 5. Continued.

			Environment									
			Combined		F_2 plants		Agronomy		Burkey		Puerto Rico	
Population	SSR locus	Linkage group	P	Partial R^2	P	Partial R^2	P	Partial R^2	P	Partial R^2	P	Partial R^2
				----%----		---%---		---%---		----%----		-----%-----
2	Satt070	B2	0.0134	4.0	0.0039	8.2	0.0122	4.2	0.0008	8.7	NS	
	Satt565	C1	0.0047	5.7	NS		0.0042	5.9	NS		NS	
	Satt227	C2	NS		NS		0.0491	2.4	0.0068	4.5	NS	
	Satt277	C2	0.0014	8.1	NS		0.0021	7.5	0.0016	6.8	0.0198	5.4
	Sctt008	D2	NS		0.0018	8.7	NS		NS		NS	
	Sat_099	L	0.0001	33.7	0.0001	18.4	0.0001	34.0	0.0001	35.0	0.0001	19.2
Total variation explained				51.5		35.3		54.0		55.0		24.6
3	Satt187	A2	0.0062	4.4	NS		NS		NS		0.0071	5.5
	Satt070	B2	0.0048	4.2	NS		0.0260	3.9	NS		0.0273	2.9
	Satt154	D2	0.0065	4.7	NS		NS		0.0405	3.3	0.0055	5.3
	Satt114	F	NS		NS		NS		NS		0.0001	15.0
	Sctt009	H	0.0037	5.9	NS		0.0233	4.3	0.0084	5.8	0.0268	2.8
	Satt006	L	0.0001	32.5	NS		0.0001	21.7	0.0001	24.6	0.0001	25.6
	Satt336	M	0.0001	11.7	0.0007	12.7	0.0002	13.3	0.0009	10.1	0.0201	3.5
Total variation explained				63.4		12.7		43.2		43.8		60.6

Table 6. Percentage of F_2 soybean plants selected for the smallest seed size based on the phenotypic (PS), marker (MAS), and index selection methods that also were the smallest as F_2-derived lines.

Data source	Method	Population 1	2	3	$\bar{x}$	Combined
		%				
SFAV†	PS	71	45	47	54	52
	MAS	48	50	37	45	47
	Index	76	45	47	56	53
	Random	23	22	21	22	22
MLR‡	PS	74	47	43	55	53
	MAS	42	53	47	47	50
	Index	74	47	43	55	54
	Random	21	21	23	22	26

† There were 21 F_2 plants selected in population 1, 20 in population 2, 19 in population 3, and 60 in the combined population using single-factor analysis of variance (SFAV).

‡ There were 19 F_2 plants selected in population 1, 19 in population 2, 21 in population 3, and 68 in the combined population using multiple-locus regression (MLR).

Table 7. Percentage of F_2 soybean plants selected for the largest seed size based on the phenotypic (PS), marker (MAS), and index selection methods that also were the largest as F_2-derived lines.

Data source	Method	Population 1	2	3	$\bar{x}$	Combined
		----------%----------				
SFAV†	PS	53	46	64	54	42
	MAS	42	64	46	51	46
	Index	53	50	64	56	48
	Random	21	25	25	24	22
MLR‡	PS	50	43	56	50	47
	MAS	50	62	56	56	54
	Index	50	48	64	54	49
	Random	17	23	29	23	26

† There were 19 F_2 plants selected in population 1, 22 in population 2, 22 in population 3, and 59 in the combined population using single-factor analysis of variance (SFAV).

‡ There were 16 F_2 plants selected in population 1, 21 in population 2, 25 in population 3, and 68 in the combined population using multiple-locus regression (MLR).

environment. The marker loci individually explained between 8 to 17% of the variation for seed size according to results derived from single-factor analysis of variance. Results from multiple-locus regression identified eight loci (Satt409, Satt070, Satt322, Satt077, Satt154, Satt045, Satt510, and Satt273) that marginally contributed 3 to 16% of the variation after accounting for the other marker loci in the model and together explained as much as 44% of the total variation for seed size at individual environments (Table 5). The small-seeded parent A97-775019 contributed alleles for small seed at five of the 12 loci (Satt409, Satt070, Satt002, Satt154, and Satt510) and for large seed at five loci (Satt322, Satt077, Satt001, Satt273, and Satt551), the normal-size parent A96-492041 contributed alleles for small seed at five loci (Satt322, Satt077, Satt001, Satt273, and Satt551) and for large seed at five loci (Satt409, Satt070, Satt002, Satt154, and Satt510) (Table 2). The remaining two marker loci (Satt185 and Satt045) varied in the estimate of the alleles that they contributed to seed size.

In Pop 2, 16 marker loci on 10 linkage groups were significantly associated with SSQTL in at least one environment (Table 3). Four marker loci were significant in the four environments (Satt070, Satt166, Sat_099, and Satt006). Seven of the remaining 12 marker loci were significant in more than one environment (Satt534, Satt227 , Satt277, Sctt008, Satt185, Satt373, and Satt336). The marker loci individually explained 8 to 38% of the variation for seed size according to results derived from single-factor analysis of variance. Six marker loci (Satt070, Satt565, Satt227, Satt277, Sctt008, and Sat_099) identified using multiple-locus regression marginally contributed 2 to 35% of the variation after accounting for the other marker loci in the model and together explained as much as 55% of the total variation for seed size at individual environments (Table 5). The small-seeded parent A97-775006 contributed alleles for small seed at 13 of the 16 marker loci (Satt070, Satt534, Satt565, Satt184, Sctt008, Satt135, Satt431, Satt166, Sat_099, Satt006, Satt373, Satt336, and Satt173) and large size at the remaining three loci (Satt227, Satt277, and Satt185)

(Table 3). The normal-size parent S12-49 contributed alleles for small size at three loci (Satt227, Satt277, and Satt185) and alleles for large size at 13 loci (Satt070, Satt534, Satt565, Satt184, Sctt008, Satt135, Satt431, Satt166, Sat_099, Satt006, Satt373, Satt336, and Satt173).

There were 22 marker loci identified on 11 linkage groups in at least one environment for Pop 3 (Table 4). Only one marker was significant in the four environments (Satt336); however, eight loci were significant in more than one environment (Satt304, Satt070, Sattt565, Satt154, Sctt009, Satt302, Satt006, and Satt143). The marker loci individually explained 8 to 29% of the variation for seed size according to results derived from single-factor analysis of variance. Results from multiple-locus regression identified seven loci (Satt187, Satt070, Satt154, Satt114, Sctt009, Satt006, and Satt336) that marginally contributed 3 to 33% of the variation after accounting for the other marker loci in the model and together explained as much as 63% of the total variation for seed size at individual environments (Table 5). The small-seeded parent A97-775026 contributed alleles for small size at 18 marker loci (Satt187, Satt304, Satt070, Sct_094, Satt184, Satt172, Satt154, HSP176, Satt114, Satt334, Satt510, Satt072, Sctt009, Satt541, Satt314, Satt302, Satt143, and Satt006) and alleles for large size at three of the remaining four loci (Satt336, Satt009, and Satt237) (Table 4). The normal-size parent A96-492058 contributed alleles for small size at three of the 22 loci (Satt336, Satt009, and Satt237) and alleles for large size at 18 loci (Satt187, Satt304, Satt070, Sct_094, Satt184, Satt172, Satt154, HSP176, Satt114, Satt334, Satt510, Satt072, Sctt009, Satt541, Satt314, Satt302, Satt143, and Satt006). The remaining marker loci (Satt565) varied in its estimate of the alleles for seed size.

The results from the three populations indicated that both the small-seeded and the normal-size parents could contribute alleles for small and large seed size to their progeny. Two-way analyses of variance revealed that there were no significant epistatic interactions between SSQTL in the three populations across the four environments.

Satt187 in Pop 3 and Satt277 and Sat_099 in Pop 2 were marker loci previously found to be associated with SSQTL in soybean populations by Orf et al. (1999). In both their study and ours, the smallest seeded parent had an allele at Satt187 and Sat_099 that decreased seed size. For Satt277, the small-seeded parent in their population contributed alleles for large seed, whereas it contributed alleles for small seed in our study. The difference between the studies may be due to the limited seed-size difference between the two parents in the population evaluated by Orf et al. (1999).

Across the three populations, Satt409 on linkage group A2; Satt304, Satt070, Sct_094, and Satt534 on linkage group B2; Satt565 on linkage group C1; Satt322 and Satt227 on linkage group C2; Sctt008, Satt154, and Satt135 on linkage group D2; Satt185 and Satt045 on linkage group E; Satt510, HSP176, Satt114, Satt334, and Satt072 on linkage group F; Satt431 on linkage group J; Satt001 and Satt273 on linkage group K; and Satt166, Satt006, Satt143, and Satt373 on linkage group L were within 1.4 to 36.4 cM of marker loci identified in previous studies (Mansur et al., 1996; Maughan et al., 1996; Mian et al., 1996; Orf et al., 1999; Sebolt et al., 2000). Thirteen marker loci located on linkage groups D1A, D1B, H, M, N, and O represent unique SSQTL associations not previously identified in other studies (Tables 2, 3, and 4). No significant marker loci were associated with seed size on linkage groups A1, B1, G, and I.

The effectiveness of MAS for small and large seed size using the molecular markers associated with the trait was compared with traditional phenotypic selection (PS) and an index based on ranking F_2 plants by MAS and PS. All significant markers identified using single-factor analysis of variance or multiple-locus regression at $P \leq 0.01$ or at $P \leq 0.05$ were used to determine the MAS score for the F_2 plants and the F_2-derived lines. All lines were scored based on their marker genotype. Individuals that possessed a homozygous allele associated with large seed size were scored as −1, individuals that possessed both alleles were scored as 0, and individuals that

possessed a homozygous allele associated with small seed size were scored as +1. Selection was practiced among the 100 F_2 plants of each population separately and among the 300 plants of the three populations without regard to the population from which they originated. The number of plants selected was ≈ 20% (Tables 6 and 7). The selection percentage was based on the marker score at which there was a separation among groups of lines. For PS and MAS, F_2 plants and F_2-derived lines were sorted based on their seed size and marker score. The plants and lines with the smallest and largest seed size and marker score were selected. For the index selection method, the F_2 plants and F_2-derived lines were given a rank score for PS and for MAS. The rank scores for each individual were added to determine the index score. F_2 plants and F_2-derived lines were sorted based on their index score, the plants and lines with the smallest and largest index score were selected.

The mean seed size of the F_2-derived lines across environments was used to determine if the F_2 plants were correctly selected by the three methods. An F_2 plant was correctly selected if its F_2-derived line was in the selected group of F_2-derived lines. For example, 21 F_2 plants were selected by each method for the smallest seed size in Pop 1 (Table 6). The number of selected F_2 plants represented in the 21 F_2-derived lines with the smallest size was determined and expressed as a percentage of 21. The effectiveness of the three methods was compared with random selection. Random selection was conducted by sampling without replacement. Using the previous example of selecting 21 lines, sampling without replacement was calculated as 1/100 + 1/99 + 1/98 ····· 1/82 + 1/81+ 1/80 = 0.23 * 100 = 23%.

The three methods varied in effectiveness across populations for selection of F_2 plants with small and large seed (Tables 6 and 7). All the methods were more effective than random selection. The method of choice for evaluation of individual plants from a population would primarily depend on the cost of conducting each method. There was not an advantage for index selection; therefore,

the cost of conducting both phenotypic and marker selection could not be justified. The cost of phenotypic selection was estimated to be $US 0.35 per plant, including harvesting and threshing the plants and counting and weighing their seeds to determine mg seed^{-1}. The current cost of MAS was estimated to be a minimum of $US 0.75 per plant, which assumed that six multiplexed markers run on one lane were used for each plant. Based on these estimates, phenotypic selection for seed size in soybean would be preferred.

Acknowledgments

Some of the molecular markers used in the study were kindly provided by H.R. Boerma, Dep. of Crop and Soil Sciences, and R.S. Hussey, Dep. of Plant Pathology, Univ. of Georgia. Facilities and thermocyclers were generously provided by G.F. Polking, DNA Sequencing and Synthesis Facility, Iowa State University.

References

Bravo, J.A., W.R. Fehr, and S.R. Cianzio. 1980. Use of pod width for indirect selection of seed weight in soybeans. Crop Sci. 20:507-510.

Buhr, K.L. 1976. Inheritance of timing to flower, time to physiological maturity, and growth habit in soybeans grown at a tropical latitude. Ph.D. Dissertation. Iowa State Univ., Ames (Diss Abstr. 37/02B:552).

Carpenter, J.A., and W.R. Fehr. 1986. Genetic variability for desirable agronomic traits in populations containing *Glycine soja* germplasm. Crop Sci. 26:681–686.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464–1490.

Hartwig, E.E. 1973. Varietal development. p. 187–207. *In* B.E. Caldwell (ed.) Soybeans: Improvement, production, and uses. ASA, CSSA, and SSSA, Madison, WI.

Holland, J.B. 1998. EPISTACY: A SAS program for detecting two-locus epistatic interactions using genetic marker information. J. of Heredity 89:374–375.

Johnson, S.L., W.R. Fehr, G.A. Welke, and S.R. Cianzio. 2001. Genetic variability for seed size of two- and three-parent soybean populations. Crop Sci. 41:1029–1033.

Keim, P., T.C. Olson, and R.C. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. Newsl. 15:150–152.

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, and L. Newburg. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181.

Lander, E.S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199.

Leroy, A.R., S.R. Cianzio, and W.R. Fehr. 1991. Direct and indirect selection for small seed of soybean in temperate and tropical environments. Crop Sci. 31:697–699.

Mansur, L.M., J.H. Orf, K. Chase, T. Jarvik, P.B. Cregan, and K.G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci. 36:1327–1336.

Maughan, P.J., M.A. Saghai Maroof, and G.R. Buss. 1996. Molecular-marker analysis of seed-weight: Genomic locations, gene action, and evidence for orthologous evolution among three legume species. Theor. Appl. Genet. 93:574–579.

Mian, M.A.R., M.A. Bailey, J.P. Tamulonis, E.R. Shipe, T.E. Carter Jr., W.A. Parrott, D.A. Ashley, R.S. Hussey, and H.R. Boerma. 1996. Molecular markers associated with seed weight in two soybean populations. Theor. Appl. Genet. 93:1011–1016.

Narvel, J.M., W.R. Fehr, W. Chu, D. Grant, and R.C. Shoemaker. 2000. Simple sequence repeat diversity among soybean plant introductions and elite genotypes. Crop Sci. 40:1452–1458.

Orf, J.H., K. Chase, T. Jarvik, L.M. Mansur, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Sci. 39:1642–1651.

SAS Institute. 1992. SAS guide for personal computers. 6th ed. SAS Inst., Cary, NC.

Sebolt, A.M., R.C. Shoemaker, and B.W. Diers. 2000. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci. 40:1438–1444.

Weber, C.R. 1950. Inheritance and interrelation of some agronomic and chemical characters in an interspecific cross in soybeans, *Glycine max* x *G. ussuriensis*. Iowa Agric. Exp. Stn. Bull. 374.

CHAPTER 3. GENERAL DISCUSSION

Mapping genes has become a tool in studying many basic areas of the plants biology, including evolution, and a potential tool for selecting superior progeny without collecting phenotypic data by utilizing marker-assisted selection (MAS).

For my study, three populations were developed for quantitative trait loci (QTL) mapping. The maps were highly conserved with only minor rearrangements when compared with the USDA-ISU molecular maps (Cregan et al., 1999). Additional markers, not used in Narvel et al. (2000), were synthesized to increase the saturation of the map and increase the power with which to detect QTL.

In the three populations, 12 to 22 marker loci were associated with seed size QTL (SSQTL). Of the SSQTL identified, one to four SSQTL were observed in the four environments, while five to eight of the remaining SSQTL were observed in more than one environment. Similar marker-SSQTL associations were detected across populations. Linkage groups B2, D1A, D2, and M possessed similar marker-SSQTL associations across the three populations. Linkage groups C2 and E produced similar marker-SSQTL associations in Pop 1 and Pop 2. Linkage groups C1 and L produced similar marker-SSQTL associations in Pop 2 and Pop 3, and linkage group F produced similar marker-SSQTL associations in Pop 1 and Pop 3. Because the consistency of detecting QTLs for seed size across environments and populations was relatively high, MAS may be possible.

For MAS to be effective in soybean, a molecular map for soybean must be highly saturated (Dudley, 1993). In my study only, 70 to 82 markers were used to construct the molecular maps of the three populations. Selection based on molecular markers or the index method were equally effective when compared with selection based solely on phenotypic data. This could be due to the

moderate to high heritability of seed size in soybean, the lack of QTL detection due to inadequate marker coverage, or the relatively small population size. Increasing the population size and the trait's heritability to improve the estimates of gene effects also improves phenotypic selection, leaving little room for improvement of selection efficiency via gene information (Bernardo, 2001). Therefore, until the cost of collecting molecular data decreases considerably, MAS will not be as effective as collecting phenotypic data for selecting superior individuals for seed size in soybean.

References

Bernardo, R. 2001. What if we knew all the genes for a quantitative trait in hybrid crops? Crop Sci. 41:1–4.

Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark, A. L. Kahler, N. Kaya, T. T. VanToai, D. G. Lohnes, J. Chung, and J. E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464–1490.

Dudley, J.W. 1993. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci. 33:660–668.

Narvel, J.M., W.R. Fehr, W. Chu, D. Grant, and R.C. Shoemaker. 2000. Simple sequence repeat diversity among soybean plant introductions and elite genotypes. Crop Sci. 40:1452–1458.

CHAPTER 4. APPENDICES

APPENDIX A

MEANS OF GENOTYPES AT INDIVIDUAL ENVIRONMENTS AND ACROSS ENVIRONMENTS

Table A1. Seed size of parents, their 100 F_2 plants grown at Ames, and their F_2-derived lines grown at three environments and combined across environments for population 1.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry†	Plant	Rep 529	Rep 530	$\bar{x}$	Rep 531	Rep 532	$\bar{x}$	Rep 579	Rep 580	$\bar{x}$	Combined‡
	mg seed^{-1}										
1	87	93	87	93	85	83	84	113	116	115	96
2	92	97	88	93	79	87	83	105	106	106	94
3	83	92	82	87	79	83	81	108	100	104	91
4	105	107	105	106	98	105	102	134	136	135	114
5	106	109	108	108	104	99	102	132	130	131	114
6	102	101	103	102	118	104	111	149	151	150	121
7	107	99	87	93	90	98	94	122	127	124	104
8	99	98	97	97	84	88	86	111	121	116	100
9	116	112	102	107	101	89	95	139	134	137	113
10	120	119	113	116	106	111	108	129	126	127	117
11	104	106	98	102	91	107	99	125	121	123	108
12	96	88	88	88	87	80	83	117	117	117	96
13	133	122	128	125	115	110	113	156	159	157	131
14	113	107	94	101	90	89	90	126	123	125	105
15	116	110	97	104	98	104	101	129	132	130	111
16	104	101	117	109	112	104	108	134	136	135	117

† Entries 1 to 100 were the lines of Population 1, entries 101-110 were A96-492041, and entries 111-120 were A97-775019.

‡ Mean of $F_{2:3}$ and $F_{2:4}$ lines across environments in 1999.

§ Standard error of the mean based on the error mean square.

¶ Least significant difference at the 0.05 and 0.01 probability levels based on the error mean square.

Table A1. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 529	Rep 530	$\bar{x}$	Rep 531	Rep 532	$\bar{x}$	Rep 579	Rep 580	$\bar{x}$	Combined
	mg seed^{-1}										
17	109	109	94	101	94	98	96	130	135	132	110
18	95	90	93	91	85	87	86	114	121	118	98
19	96	89	91	90	84	73	79	114	110	112	94
20	116	117	114	116	111	113	112	129	128	128	119
21	109	102	88	95	88	86	87	121	128	124	102
22	108	104	103	103	101	104	102	130	121	125	110
23	122	121	116	118	109	115	112	135	131	133	121
24	92	87	82	84	85	78	81	118	107	112	93
25	111	107	96	101	105	102	103	128	131	129	111
26	114	111	111	111	103	115	109	132	129	131	117
27	114	123	115	119	123	120	122	136	128	132	124
28	100	111	111	111	105	110	107	138	144	141	120
29	100	114	112	113	110	109	109	135	134	135	119
30	99	107	99	103	101	87	94	121	113	117	105
31	108	104	104	104	104	94	99	129	127	128	110
32	120	107	98	103	99	99	99	131	121	126	109
33	110	112	102	107	108	85	97	123	126	125	109
34	95	80	75	78	72	76	74	112	116	114	88
35	97	105	102	103	97	95	96	123	134	129	109
36	112	114	112	113	110	104	107	149	152	150	123

Table A1. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry†	Plant	Rep 529	Rep 530	$\bar{x}$	Rep 531	Rep 532	$\bar{x}$	Rep 579	Rep 580	$\bar{x}$	Combined
	mg seed^{-1}										
37	102	100	98	99	97	97	97	115	110	113	103
38	106	101	99	100	100	97	98	132	132	132	110
39	98	106	100	103	104	111	107	129	125	127	112
40	127	105	105	105	101	93	97	128	132	130	110
41	102	102	88	95	91	85	88	122	109	116	100
42	99	81	75	78	77	78	77	102	99	101	85
43	113	114	100	107	103	107	105	131	133	132	115
44	90	88	86	87	85	74	79	104	108	106	91
45	104	77	78	78	70	72	71	86	106	96	81
46	118	98	94	96	96	95	95	111	120	115	102
47	122	110	105	108	99	95	97	134	130	132	112
48	125	119	126	123	103	100	101	144	138	141	122
49	106	115	123	119	111	113	112	145	148	147	126
50	109	99	95	97	100	99	100	122	137	130	109
51	111	105	100	103	108	94	101	127	129	128	111
52	123	103	102	103	91	98	95	137	129	133	110
53	118	94	94	94	90	90	90	129	131	130	105
54	121	104	105	105	95	111	103	135	128	131	113
55	96	96	90	93	89	89	89	103	117	110	97
56	133	120	114	117	111	120	115	142	148	145	126

Table A1. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 529	Rep 530	$\bar{x}$	Rep 531	Rep 532	$\bar{x}$	Rep 579	Rep 580	$\bar{x}$	Combined
	mg seed^{-1}										
57	114	106	107	106	99	107	103	129	132	130	113
58	112	109	96	103	105	107	106	132	131	132	113
59	113	112	114	113	103	121	112	136	137	137	120
60	125	116	101	108	95	97	96	120	115	118	107
61	98	94	93	94	100	94	97	118	111	114	102
62	114	101	96	98	87	90	88	119	118	119	102
63	93	91	94	92	89	92	90	128	116	122	102
64	128	115	113	114	112	110	111	138	130	134	120
65	112	122	112	117	118	128	123	139	149	144	128
66	100	105	112	108	112	111	111	138	131	135	118
67	117	107	101	104	102	106	104	127	134	130	113
68	87	84	82	83	86	86	86	117	108	113	94
69	116	111	107	109	105	110	108	138	150	144	120
70	113	103	110	106	99	103	101	131	128	130	112
71	97	92	90	91	87	87	87	114	115	114	98
72	137	120	123	121	118	118	118	169	174	171	137
73	103	105	99	102	105	97	101	121	135	128	110
74	100	95	95	95	98	94	96	126	125	126	106
75	107	108	94	101	95	103	99	125	135	130	110
76	98	103	106	104	87	99	93	121	130	126	108

Table A1. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 529	Rep 530	$\bar{x}$	Rep 531	Rep 532	$\bar{x}$	Rep 579	Rep 580	$\bar{x}$	Combined
	mg seed^{-1}										
77	97	103	92	98	86	89	88	119	123	121	102
78	119	113	100	106	105	110	107	123	140	132	115
79	101	98	90	94	80	80	80	127	124	125	100
80	113	100	99	99	94	100	97	118	132	125	107
81	116	113	112	113	102	107	105	138	160	149	122
82	119	117	106	111	100	105	103	130	155	142	119
83	111	107	103	105	99	108	104	125	126	125	111
84	88	86	92	89	77	85	81	117	115	116	95
85	128	110	118	114	118	112	115	152	163	157	129
86	140	102	94	98	101	102	102	123	124	123	108
87	140	122	117	119	113	117	115	158	167	162	132
88	128	103	90	96	96	91	94	125	125	125	105
89	132	120	116	118	113	125	119	145	147	146	128
90	118	95	94	95	84	90	87	132	130	131	104
91	138	96	92	94	90	90	90	118	118	118	101
92	120	96	99	97	98	98	98	120	117	118	105
93	122	109	102	105	98	99	98	137	140	138	114
94	108	102	115	108	96	96	96	125	136	130	111
95	125	110	92	101	94	93	93	128	130	129	108
96	104	100	104	102	99	96	98	124	132	128	109

Table A1. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 529	Rep 530	$\bar{x}$	Rep 531	Rep 532	$\bar{x}$	Rep 579	Rep 580	$\bar{x}$	Combined
	mg seed^{-1}										
97	93	79	85	82	88	94	91	120	123	121	98
98	130	114	117	116	115	110	112	149	150	150	126
99	106	91	99	95	88	80	84	126	122	124	101
100	105	90	83	86	78	85	81	121	113	117	95
101	158	138	140	138	134	142	138	170	169	169	149
102	166	137	135	136	132	122	127	173	176	175	146
103	163	135	125	130	132	137	135	172	174	173	146
104	138	129	139	134	128	136	132	170	166	168	145
105	170	131	126	129	128	142	135	170	164	167	143
106	164	137	131	134	133	137	135	166	182	174	148
107	169	134	141	137	143	138	140	168	180	174	151
108	159	143	139	141	135	143	139	178	184	181	154
109	161	135	128	132	122	138	130	171	180	175	146
110	124	134	128	131	141	138	139	169	155	162	144
111	74	67	65	66	68	62	65	93	102	98	76
112	75	71	73	72	69	65	67	104	99	101	80
113	71	72	83	78	68	69	69	127	95	111	86
114	76	74	67	71	66	66	66	95	100	98	78
115	84	74	72	73	66	68	67	97	101	99	80

Table A1. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 529	Rep 530	$\bar{x}$	Rep 531	Rep 532	$\bar{x}$	Rep 579	Rep 580	$\bar{x}$	Combined
	mg seed^{-1}										
116	83	66	71	68	64	65	65	93	102	98	77
117	84	70	75	72	71	72	72	105	106	105	83
118	83	69	67	68	71	66	68	98	106	102	79
119	84	74	70	72	67	65	66	94	100	97	78
120	82	69	73	71	58	63	61	103	103	103	78
$\bar{x}$	110	104	100	102	98	98	98	127	129	128	109
SE§				3.1			2.8			3.7	2.4
$LSD_{0.05}$ ¶				8.6			7.8			10.3	6.5
$LSD_{0.01}$				11.4			10.3			13.7	8.8

Table A2. Seed size of parents, their 100 F_2 plants grown at Ames, and their F_2-derived lines grown at three environments and combined across environments for population 2.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry†	Plant	Rep 533	Rep 534	$\bar{x}$	Rep 535	Rep 536	$\bar{x}$	Rep 581	Rep 582	$\bar{x}$	Combined‡
	mg seed^{-1}										
1	110	90	97	93	88	88	88	130	128	129	103
2	110	90	91	90	87	85	86	115	120	117	98
3	141	104	104	104	104	110	107	129	134	131	114
4	109	104	103	104	99	97	98	130	145	138	113
5	112	98	99	98	93	88	90	129	126	128	105
6	120	111	111	111	103	103	103	147	153	150	121
7	128	115	133	124	116	107	111	141	146	143	126
8	119	104	101	102	100	101	100	145	142	143	115
9	107	92	104	98	86	90	88	125	129	127	104
10	119	108	108	108	107	101	104	139	127	133	115
11	128	115	106	110	96	98	97	138	146	142	117
12	131	110	96	103	109	100	104	142	161	151	120
13	102	89	84	87	77	77	77	129	134	132	98
14	132	115	113	114	104	118	111	149	150	149	125
15	115	99	96	97	91	91	91	120	124	122	103
16	131	118	119	118	117	118	117	158	165	162	132

† Entries 1 to 100 were the lines of Population 2, entries 101-110 were S12-49, and entries 111-120 were A97-775006.

‡ Mean of $F_{2:3}$ and $F_{2:4}$ lines across environments in 1999.

§ Standard error of the mean based on the error mean square.

¶ Least significant difference at the 0.05 and 0.01 probability levels based on the error mean square.

Table A2. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 533	Rep 534	$\bar{x}$	Rep 535	Rep 536	$\bar{x}$	Rep 581	Rep 582	$\bar{x}$	Combined
	mg seed^{-1}										
17	118	97	97	97	89	94	91	129	128	129	105
18	128	93	111	102	106	95	101	134	139	137	113
19	119	92	103	97	88	99	93	124	136	130	107
20	108	101	101	101	97	96	96	134	133	133	110
21	125	88	95	91	88	90	89	136	130	133	104
22	122	99	108	103	101	94	97	129	177	153	118
23	139	111	113	112	103	106	104	137	150	143	120
24	131	101	102	102	98	98	98	129	137	133	111
25	135	103	110	106	106	113	109	131	136	133	116
26	135	103	93	98	92	93	92	134	135	134	108
27	124	92	99	95	100	102	101	138	144	141	113
28	115	96	95	95	90	94	92	121	129	125	104
29	146	103	112	108	109	117	113	144	141	143	121
30	135	105	100	102	100	112	106	139	150	144	117
31	122	111	104	108	100	103	102	132	132	132	114
32	118	89	93	91	81	85	83	127	125	126	100
33	121	97	85	91	91	89	90	125	130	128	103
34	136	121	120	120	118	115	117	147	153	150	129
35	123	96	96	96	100	96	98	138	122	130	108
36	148	116	116	116	134	124	129	160	160	160	135

Table A2. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 533	Rep 534	$\bar{x}$	Rep 535	Rep 536	$\bar{x}$	Rep 581	Rep 582	$\bar{x}$	Combined
	mg seed^{-1}										
37	125	106	105	106	118	119	119	129	134	131	119
38	107	86	86	86	81	82	82	122	129	125	98
39	155	109	107	108	101	107	104	135	138	137	116
40	97	97	93	95	84	92	88	119	130	125	102
41	132	100	104	102	103	102	102	142	148	145	116
42	124	103	103	103	103	106	104	131	136	133	114
43	118	101	105	103	95	100	97	153	132	143	114
44	150	114	123	118	106	118	112	166	173	170	133
45	125	107	102	105	98	99	98	137	139	138	114
46	101	82	81	82	80	74	77	97	113	105	88
47	115	86	91	89	88	89	88	123	126	124	100
48	143	113	126	120	102	121	112	139	145	142	124
49	105	94	111	103	91	97	94	103	136	120	105
50	132	112	107	110	99	98	99	137	142	140	116
51	121	100	103	101	98	97	98	130	138	134	111
52	132	120	107	114	102	112	107	148	156	152	124
53	135	94	93	93	94	93	93	130	135	133	106
54	133	118	122	120	121	110	116	152	159	155	130
55	134	114	104	109	109	98	103	146	159	152	121
56	116	100	93	97	95	102	99	121	126	124	106

Table A2. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 533	Rep 534	$\bar{x}$	Rep 535	Rep 536	$\bar{x}$	Rep 581	Rep 582	$\bar{x}$	Combined
	mg seed^{-1}										
57	110	92	104	98	98	95	96	125	121	123	106
58	134	89	87	88	85	83	84	119	123	121	98
59	133	110	118	114	101	107	104	150	156	153	124
60	132	95	92	93	93	91	92	121	126	123	103
61	132	104	109	107	104	97	101	140	138	139	115
62	132	92	94	93	85	95	90	125	130	128	103
63	118	97	95	96	86	85	85	117	122	120	100
64	110	89	85	87	91	93	92	114	121	118	99
65	131	104	116	110	115	107	111	139	145	142	121
66	117	101	96	99	103	107	105	121	128	125	109
67	115	97	102	99	95	97	96	120	124	122	106
68	139	115	109	112	111	114	113	125	139	132	119
69	142	108	111	110	101	98	100	131	135	133	114
70	120	97	102	99	87	83	85	128	133	130	105
71	100	95	100	97	90	94	92	127	114	121	103
72	116	86	86	86	83	85	84	128	123	126	99
73	121	104	111	107	102	95	99	137	151	144	117
74	126	97	93	95	97	96	96	119	123	121	104
75	148	101	105	103	95	111	103	136	141	138	115
76	138	113	119	116	116	99	107	150	156	153	125

Table A2. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 533	Rep 534	$\bar{x}$	Rep 535	Rep 536	$\bar{x}$	Rep 581	Rep 582	$\bar{x}$	Combined
	mg seed^{-1}										
77	112	102	95	98	89	91	90	138	142	140	109
78	113	105	103	104	104	101	102	130	135	132	113
79	139	87	98	92	98	90	94	135	129	132	106
80	123	103	89	96	102	99	101	130	135	132	110
81	116	88	87	88	87	88	88	134	139	136	104
82	95	86	87	86	77	80	79	101	111	106	90
83	126	125	126	126	112	104	108	142	154	148	127
84	128	95	100	98	87	91	89	138	148	143	110
85	123	102	107	104	100	95	98	136	141	138	113
86	133	113	110	111	113	109	111	156	140	148	123
87	114	92	89	91	93	92	92	120	129	124	102
88	132	92	104	98	95	97	96	127	131	129	108
89	152	109	99	104	95	94	94	138	138	138	112
90	117	93	95	94	91	87	89	132	138	135	106
91	129	116	108	112	113	111	112	150	155	152	125
92	107	86	97	92	81	80	81	116	120	118	97
93	120	79	91	85	86	84	85	115	121	118	96
94	96	93	100	96	94	93	94	127	132	130	106
95	131	95	93	94	104	101	102	126	131	129	108
96	115	87	90	88	94	79	86	119	126	123	99

Table A2. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 533	Rep 534	$\bar{x}$	Rep 535	Rep 536	$\bar{x}$	Rep 581	Rep 582	$\bar{x}$	Combined
	mg seed^{-1}										
97	141	102	102	102	101	105	103	148	154	151	118
98	121	92	104	98	97	94	95	114	118	116	103
99	115	90	99	94	88	91	89	118	146	132	105
100	97	92	80	86	83	77	80	131	136	133	100
101	164	122	130	126	139	129	134	186	182	184	148
102	148	131	123	127	138	139	139	170	145	157	141
103	159	136	149	142	139	131	135	190	184	187	155
104	159	128	129	129	126	138	132	172	185	178	146
105	171	131	137	134	138	130	134	175	185	180	149
106	165	134	137	136	142	138	140	177	171	174	150
107	171	139	141	140	124	124	124	182	175	179	148
108	164	133	118	125	133	137	135	173	166	169	143
109	165	132	143	137	135	138	136	178	163	171	148
110	166	129	125	127	137	143	140	192	168	180	149
111	99	63	71	67	64	67	66	96	97	97	77
112	96	71	69	70	61	66	63	91	91	91	75
113	101	74	79	76	68	64	66	94	94	94	79
114	95	68	59	64	63	65	64	99	91	95	74
115	102	72	70	71	66	65	65	98	97	98	78

Table A2. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 533	Rep 534	$\bar{x}$	Rep 535	Rep 536	$\bar{x}$	Rep 581	Rep 582	$\bar{x}$	Combined
	---mg seed^{-1}---										
116	87	74	80	77	66	66	66	99	99	99	81
117	96	68	75	71	63	68	65	96	105	101	79
118	97	72	67	69	67	68	67	91	91	91	76
119	100	73	74	73	72	64	68	98	103	100	80
120	87	72	70	71	68	67	68	101	101	101	80
$\bar{x}$	124	100	102	101	98	98	98	132	137	135	111
SE§				3.4			2.8			4.3	2.5
$LSD_{0.05}$ ¶				9.4			7.8			12	6.9
$LSD_{0.01}$				12.4			10.3			15.9	9.1

Table A3. Seed size of parents, their 100 F_2 plants grown at Ames, and their F_2-derived lines grown at three environments and combined across environments for population 3.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry†	Plant	Rep 537	Rep 538	$\bar{x}$	Rep 539	Rep 540	$\bar{x}$	Rep 583	Rep 584	$\bar{x}$	Combined‡
	mg seed^{-1}										
1	137	93	95	94	91	98	94	126	126	126	105
2	122	87	85	86	77	90	83	131	132	131	100
3	137	103	93	98	93	103	98	138	144	141	112
4	113	103	94	99	108	116	112	132	134	133	115
5	128	99	117	108	95	109	102	140	141	140	117
6	144	114	99	107	114	133	124	140	140	140	123
7	118	113	99	106	93	103	98	123	139	131	112
8	127	95	94	95	95	99	97	146	147	146	112
9	126	97	86	91	85	98	92	125	141	133	105
10	134	111	104	108	96	111	103	132	136	134	115
11	116	95	99	97	90	98	94	139	137	138	110
12	135	108	100	104	91	102	96	145	145	145	115
13	115	101	104	102	93	100	96	125	130	127	109
14	96	102	98	100	98	98	98	140	140	140	113
15	162	108	120	114	119	123	121	152	153	153	129
16	115	77	81	79	76	74	75	109	115	112	89

† Entries 1 to 100 were the lines of Population 3, entries 101-110 were A96-492058, and entries 111-120 were A97-775026.

‡ Mean of $F_{2:3}$ and $F_{2:4}$ lines across environments in 1999.

§ Standard error of the mean based on the error mean square.

¶ Least significant difference at the 0.05 and 0.01 probability levels based on the error mean square.

Table A3. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 537	Rep 538	$\bar{x}$	Rep 539	Rep 540	$\bar{x}$	Rep 583	Rep 584	$\bar{x}$	Combined
	mg seed^{-1}										
17	141	104	94	99	100	96	98	147	148	147	115
18	122	105	103	104	93	96	94	134	142	138	112
19	113	96	93	94	87	93	90	135	131	133	106
20	132	96	93	94	101	94	97	128	129	129	107
21	133	96	89	93	83	83	83	129	130	129	102
22	107	96	86	91	84	71	77	116	124	120	96
23	113	96	97	96	96	92	94	131	132	132	107
24	125	97	103	100	97	95	96	125	126	125	107
25	123	103	88	96	102	105	104	134	135	134	111
26	116	101	101	101	95	98	96	137	138	138	112
27	123	97	95	96	99	101	100	126	132	129	108
28	116	94	88	91	79	77	78	129	129	129	99
29	128	90	101	95	82	91	86	133	139	136	106
30	108	97	90	94	96	91	94	132	132	132	106
31	130	107	99	103	97	91	94	141	137	139	112
32	112	100	98	99	87	89	88	140	133	136	108
33	113	99	88	93	90	100	95	132	120	126	105
34	111	108	105	106	89	94	92	135	138	137	111
35	103	93	89	91	89	87	88	121	118	120	100
36	166	107	106	107	99	104	101	141	142	142	116

Table A3. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 537	Rep 538	$\bar{x}$	Rep 539	Rep 540	$\bar{x}$	Rep 583	Rep 584	$\bar{x}$	Combined
	mg seed^{-1}										
37	137	107	106	106	101	101	101	142	143	143	117
38	107	95	88	91	88	84	86	127	125	126	101
39	112	103	101	102	90	87	89	140	141	140	110
40	113	100	93	96	93	96	94	130	137	134	108
41	115	100	102	101	102	94	98	146	131	138	112
42	127	98	94	96	83	95	89	120	121	121	102
43	122	95	90	93	81	86	84	125	140	133	103
44	139	104	104	104	95	93	94	136	133	134	111
45	131	94	93	94	98	94	96	138	127	133	107
46	111	102	92	97	106	102	104	133	135	134	112
47	114	94	101	97	88	79	83	126	127	126	102
48	119	94	91	93	91	95	93	126	127	126	104
49	131	103	97	100	85	89	87	148	135	142	110
50	111	99	98	99	89	94	92	123	126	124	105
51	125	85	97	91	85	89	87	126	137	132	103
52	131	103	103	103	108	103	106	132	133	132	114
53	148	112	102	107	95	108	101	144	147	145	118
54	120	96	98	97	92	88	90	147	132	140	109
55	146	106	111	108	103	106	104	138	139	138	117
56	139	100	100	100	95	110	102	149	140	144	116

Table A3. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 537	Rep 538	$\bar{x}$	Rep 539	Rep 540	$\bar{x}$	Rep 583	Rep 584	$\bar{x}$	Combined
	---mg seed^{-1}---										
57	134	115	101	108	100	108	104	138	132	135	116
58	109	95	98	97	89	87	88	131	127	129	104
59	129	90	101	95	91	95	93	143	136	140	109
60	127	93	91	92	79	81	80	120	122	121	98
61	120	102	97	99	92	90	91	119	120	119	103
62	110	81	83	82	76	77	76	123	119	121	93
63	130	106	109	108	104	92	98	150	151	151	119
64	105	86	98	92	89	95	92	119	120	120	101
65	128	106	109	108	108	117	112	99	146	123	114
66	140	116	110	113	104	117	110	163	154	159	127
67	126	102	112	107	97	105	101	129	135	132	113
68	106	94	101	98	94	97	96	133	132	132	109
69	127	92	89	90	76	85	80	139	140	140	104
70	107	94	95	94	92	94	93	138	131	134	107
71	127	91	92	92	97	98	97	147	138	142	110
72	120	93	98	96	91	94	92	122	136	129	106
73	133	85	94	89	90	98	94	136	143	140	108
74	128	93	96	95	90	92	91	128	127	128	104
75	132	100	88	94	84	95	90	137	130	133	106
76	112	93	98	95	93	85	89	127	120	124	103

Table A3. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 537	Rep 538	$\bar{x}$	Rep 539	Rep 540	$\bar{x}$	Rep 583	Rep 584	$\bar{x}$	Combined
	mg seed^{-1}										
77	111	89	102	95	85	88	86	118	116	117	99
78	135	115	114	115	117	121	119	150	146	148	127
79	107	102	108	105	91	93	92	178	134	156	117
80	129	91	94	92	89	88	88	110	130	120	100
81	152	103	92	97	105	110	107	96	96	96	100
82	122	94	91	92	98	94	96	132	120	126	105
83	109	91	91	91	88	87	88	126	126	126	101
84	119	103	104	103	92	100	96	132	133	133	111
85	113	80	84	82	83	79	81	116	118	117	93
86	111	84	86	85	83	77	80	101	128	114	93
87	112	92	90	91	93	95	94	125	122	124	103
88	126	103	101	102	117	110	113	145	140	143	119
89	125	98	101	100	102	102	102	139	127	133	112
90	134	103	110	106	96	99	97	133	133	133	112
91	124	100	98	99	106	103	105	131	132	132	112
92	118	96	96	96	91	96	94	135	136	136	108
93	128	103	108	105	103	92	98	139	159	149	117
94	152	94	95	94	97	94	95	147	147	147	112
95	126	109	113	111	115	111	113	141	147	144	123
96	143	104	101	102	109	112	111	146	144	145	119

Table A3. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 537	Rep 538	$\bar{x}$	Rep 539	Rep 540	$\bar{x}$	Rep 583	Rep 584	$\bar{x}$	Combined
	mg seed^{-1}										
97	125	111	103	107	95	105	100	142	139	141	116
98	107	90	90	90	87	87	87	121	128	124	101
99	113	86	85	85	91	84	88	131	126	128	100
100	146	90	95	92	90	89	89	135	145	140	107
101	175	122	138	130	135	130	133	179	168	174	145
102	162	130	136	133	133	137	135	160	162	161	143
103	177	122	132	127	139	133	136	124	121	123	128
104	166	138	129	134	135	135	135	171	185	178	149
105	182	142	143	142	133	148	141	199	169	184	156
106	170	138	140	139	143	144	143	187	183	185	156
107	177	137	138	137	142	149	145	198	193	195	159
108	176	138	132	135	141	143	142	193	188	190	156
109	214	133	142	137	138	134	136	189	191	190	154
110	205	131	134	133	134	143	139	178	173	175	149
111	90	66	66	66	62	67	65	101	100	101	77
112	91	66	71	68	68	72	70	112	104	108	82
113	93	74	73	73	63	72	67	101	103	102	81
114	85	68	70	69	68	67	67	106	105	105	80
115	88	68	70	69	68	66	67	104	110	107	81

Table A3. Continued.

	F_2	$F_{2:3}$						$F_{2:4}$			
	Individual	Agronomy Farm			Burkey Farm			Puerto Rico			
Entry	Plant	Rep 537	Rep 538	$\bar{x}$	Rep 539	Rep 540	$\bar{x}$	Rep 583	Rep 584	$\bar{x}$	Combined
	mg seed^{-1}										
116	98	73	79	76	64	64	64	109	103	106	82
117	89	69	72	70	72	69	70	91	86	89	76
118	84	71	69	70	63	64	64	96	96	96	77
119	72	75	70	72	67	71	69	99	109	104	82
120	83	73	73	73	66	60	63	98	98	98	78
$\bar{x}$	124	98	97	98	94	96	95	133	134	133	109
SE§				3.2			2.8			4.8	2.8
$LSD_{0.05}$ ¶				8.9			5.5			13.4	7.7
$LSD_{0.01}$				11.8			10.3			17.7	10.2

APPENDIX B

ANALYSIS OF VARIANCE FOR SEED SIZE ACROSS ENVIRONMENTS

Table B1. Analysis of variance of seed size $F_{2:3}$ and $F_{2:4}$ lines grown at three environments.

		Mean squares		
Sources of variation	Df	Pop 1	Pop 2	Pop 3
Environments (E)	2	53757.8 **	83701.1 **	91306.1 **
Replications (R)/E	3	236.3 **	452.1 **	101.8 *
Genotypes (G)†	99	706.4 **	567.0 **	339.5 **
G x E	198	46.6 **	53.1 **	69.4 **
Error	297	24.2	26.7	30.0

*,** Significant at the 0.05 and 0.01 probability levels, respectively.

† Parent lines were excluded from the analysis of variance.

APPENDIX C

ANALYSIS OF VARIANCE OF SEED SIZE AT INDIVIDUAL ENVIRONMENTS

Table C1. Analysis of variance of seed size $F_{2:3}$ lines grown at the Agronomy Farm in 1999.

		Mean squares		
Sources of variation	Df	Pop 1	Pop 2	Pop 3
Replication (R)	1	570.9**	84.5	34.7
Genotypes (G)†	99	216.6**	184.9**	101.0**
Error	99	21.3	24.7	22.3

*,** Significant at the 0.05 and 0.01 probability levels, respectively.

† Parent lines were excluded from the analysis of variance.

Table C2. Analysis of variance of seed size $F_{2:3}$ lines grown at the Burkey Farm in 1999.

		Mean squares		
Sources of variation	Df	Pop 1	Pop 2	Pop 3
Replication (R)	1	22.1	0.1	247.1**
Genotypes (G) †	99	249.2**	206.9**	178.3**
Error	99	23.9	19.1	20.9

*,** Significant at the 0.05 and 0.01 probability levels, respectively.

† Parent lines were excluded from the analysis of variance.

Table C3. Analysis of variance of seed size $F_{2:3}$ lines grown at Isabella, PR in 2000.

		Mean squares		
Sources of variation	Df	Pop 1	Pop 2	Pop 3
Replication (R)	1	116.1*	1271.5**	23.6
Genotypes (G) †	99	333.9**	281.5**	199.0**
Error	99	27.5	36.5	46.8

*,** Significant at the 0.05 and 0.01 probability levels, respectively.

† Parent lines were excluded from the analysis of variance.

APPENDIX D

LINKAGE MAP CONSTRUCTION FOR THE THREE POPULATIONS USING SSR MARKERS

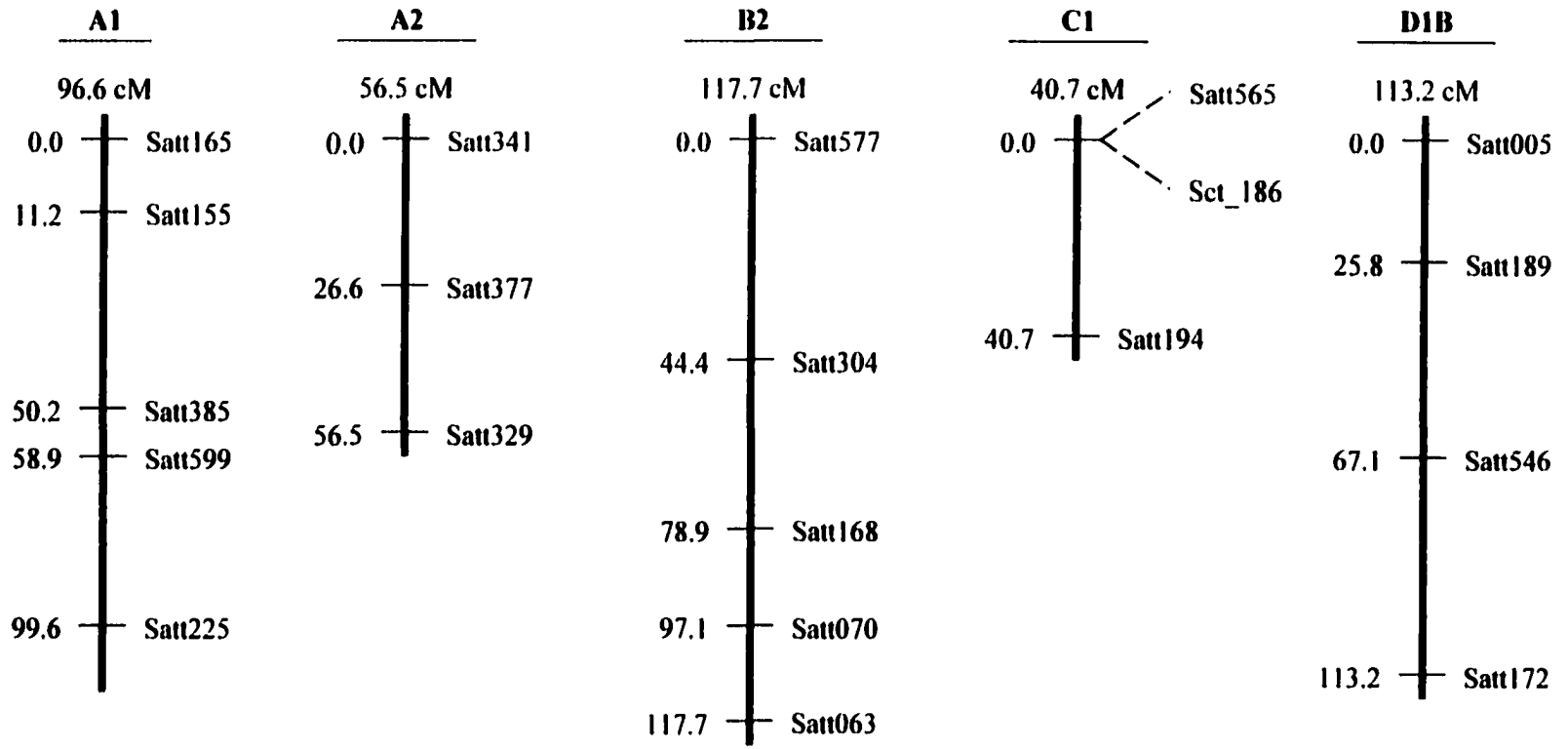


Figure D1. Linkage map constructed for Population 1 using SSR markers.

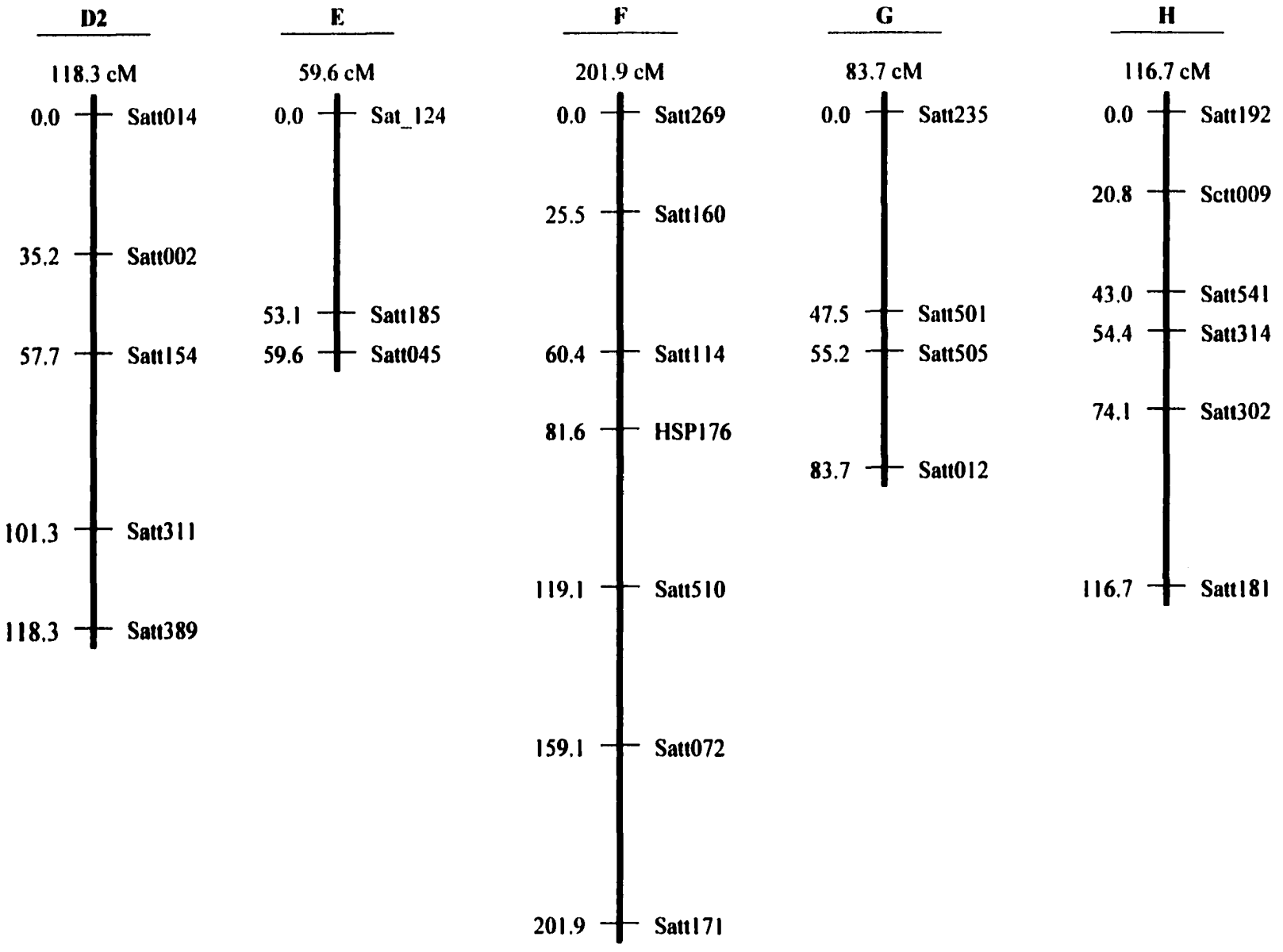
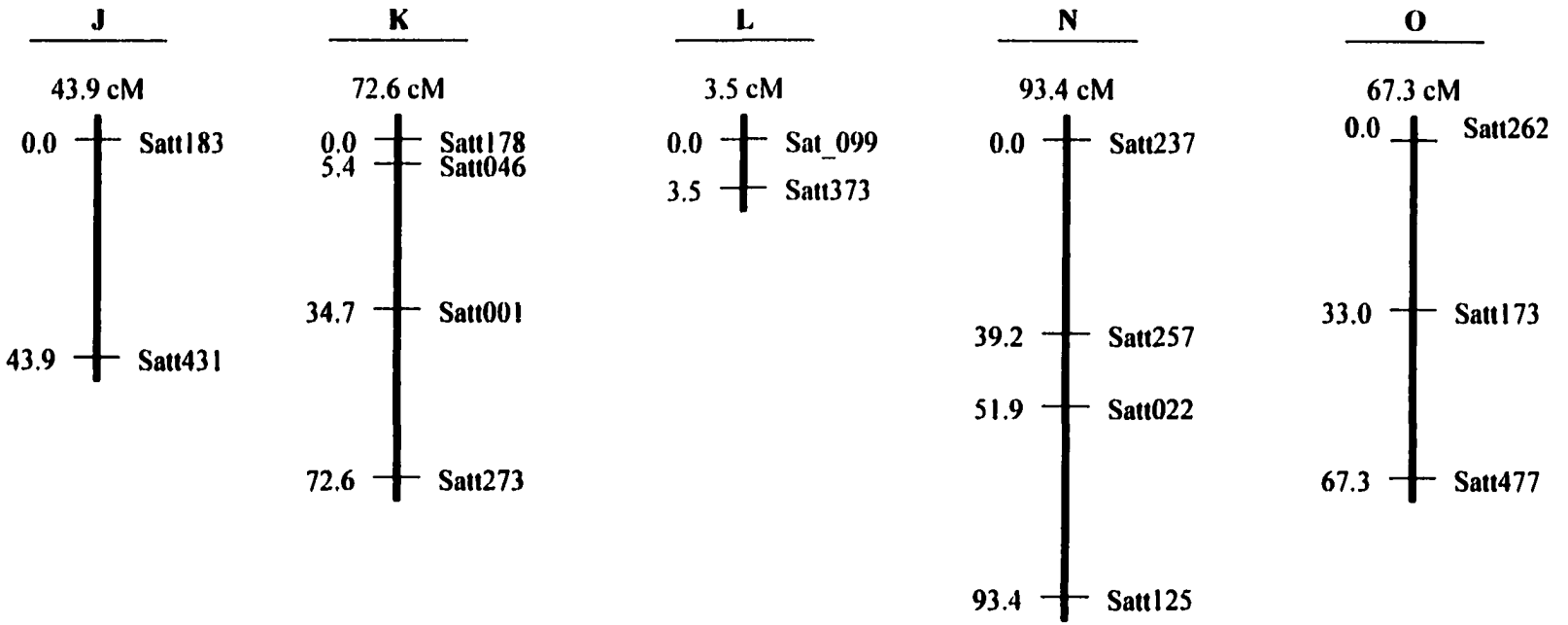


Figure D1. Continued.



Unlinked Markers – Population 1

Satt409	Satt369
Satt453	Satt367
Satt164	Satt148
Satt277	Satt249
Satt184	Satt260
Satt077	Satt551
Satt129	Satt358

Figure D1. Continued.

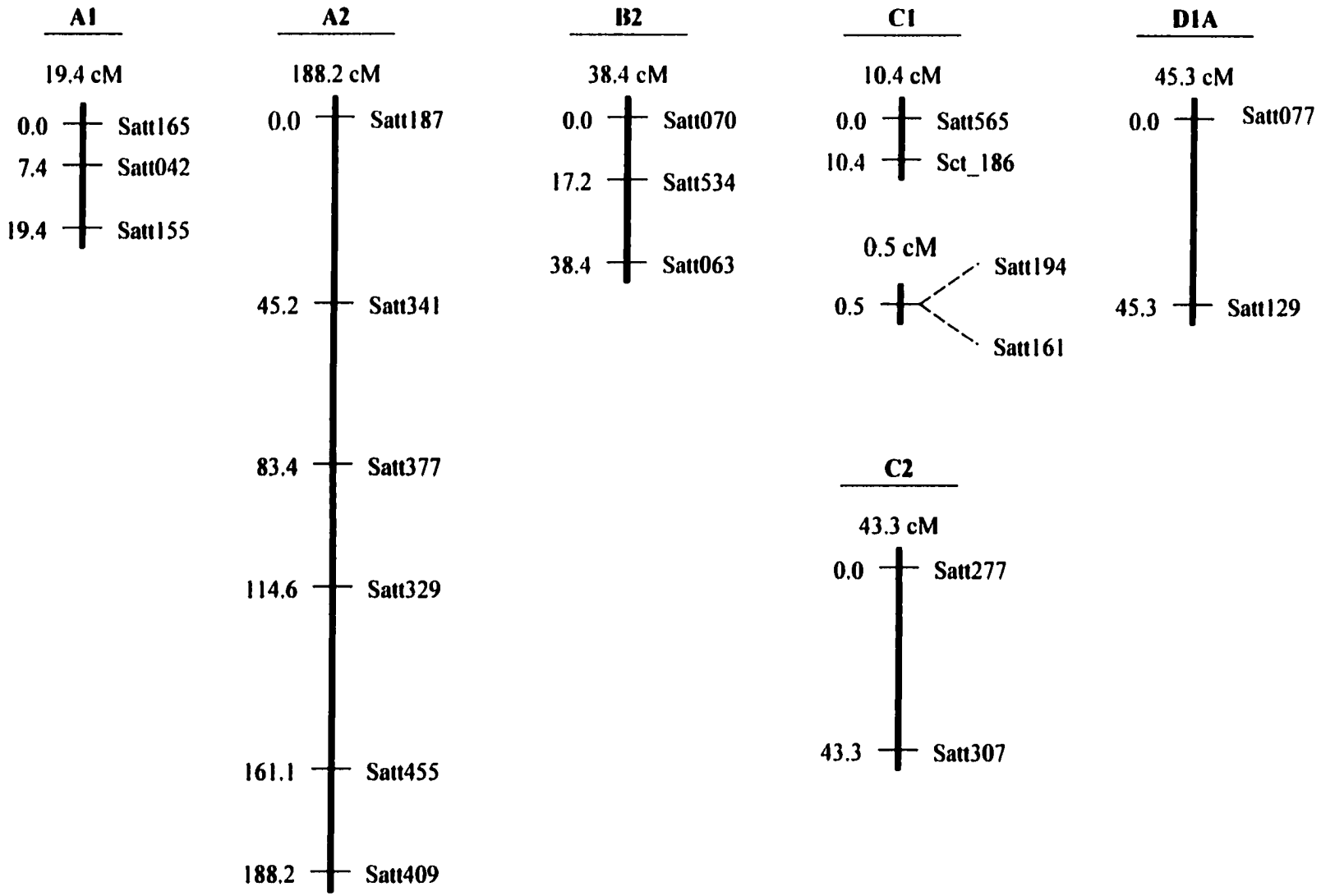


Figure D2. Linkage map constructed for Population 2 using SSR markers.

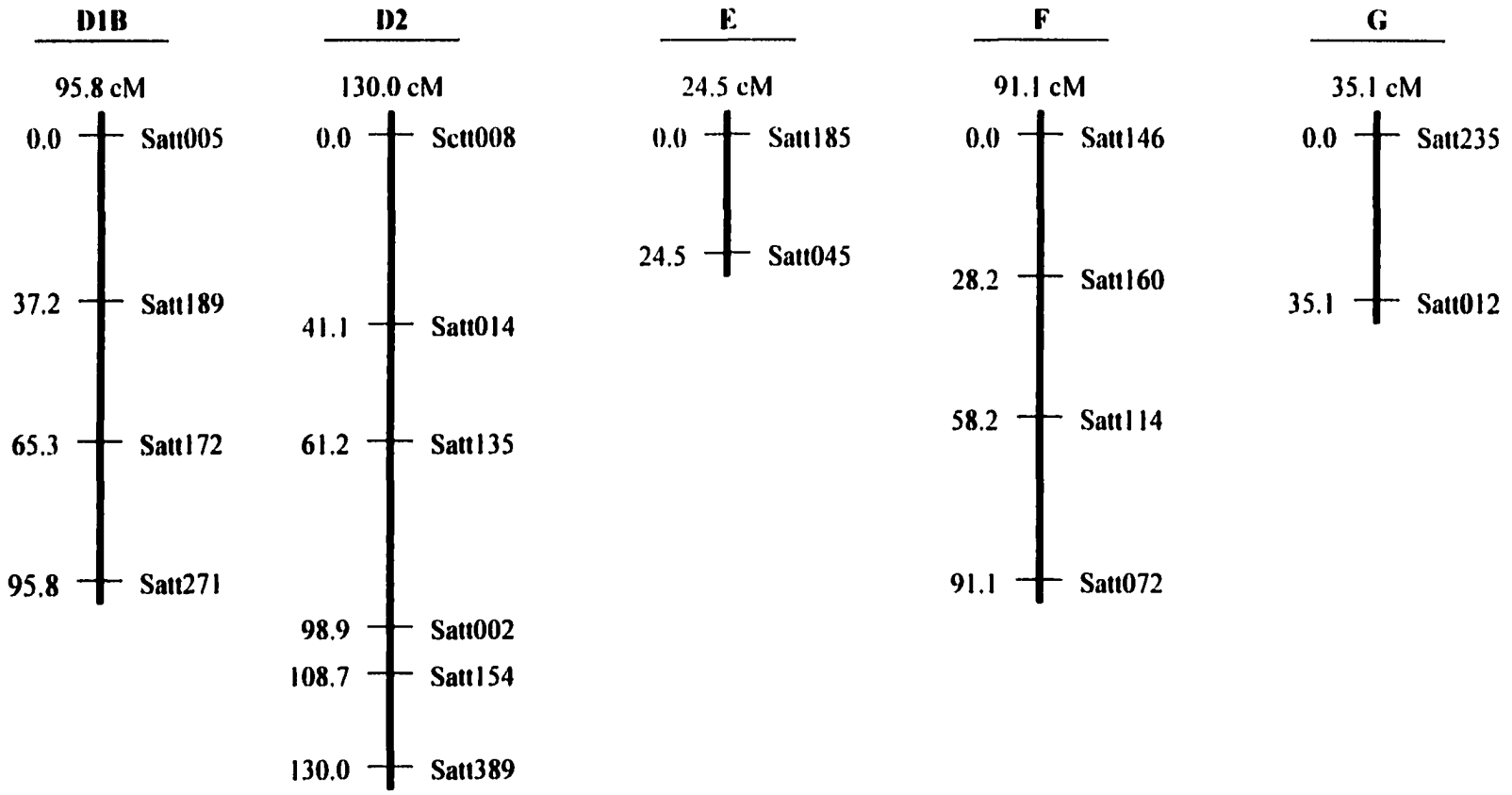


Figure D2. Continued.

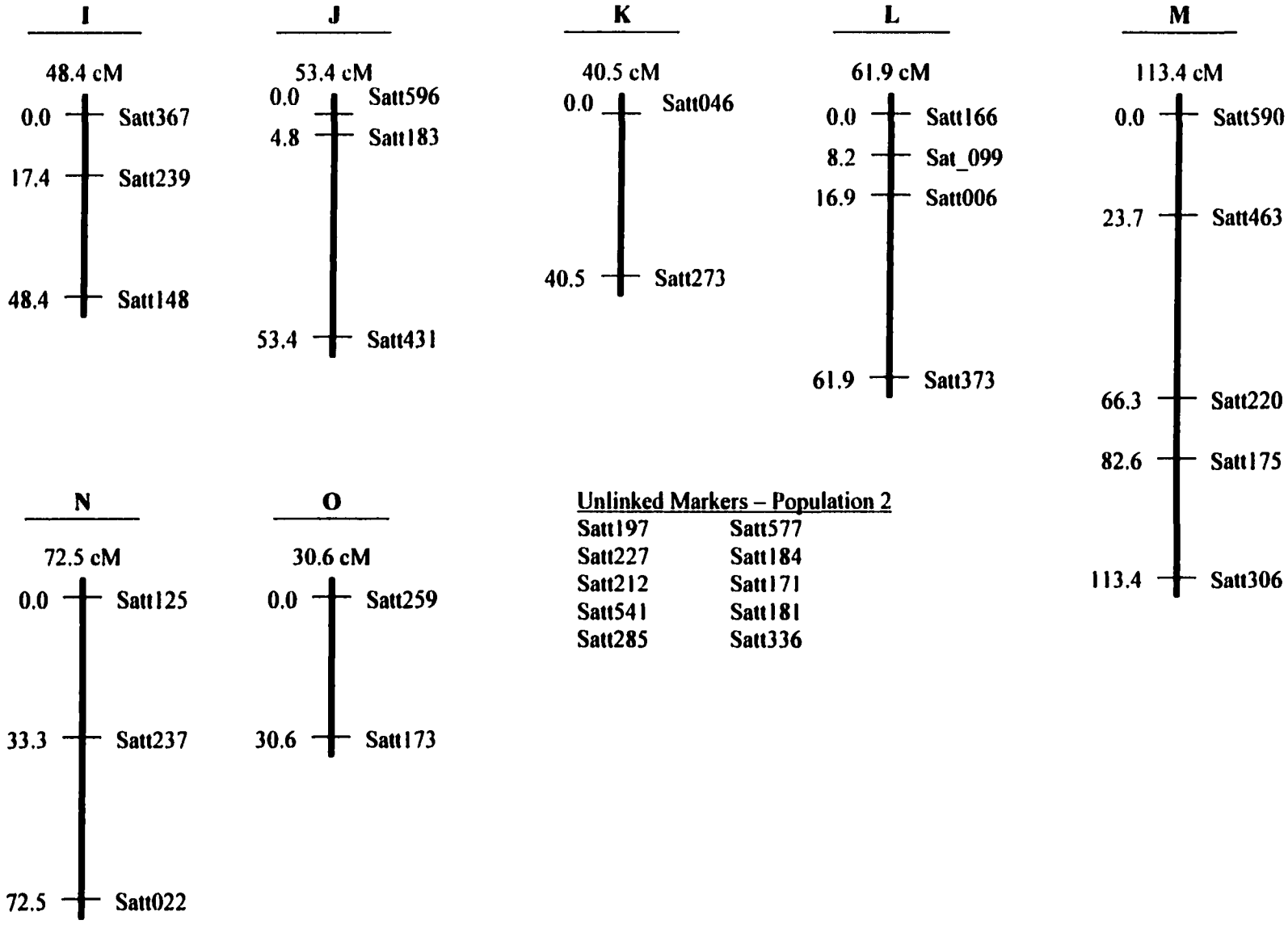


Figure D2. Continued.

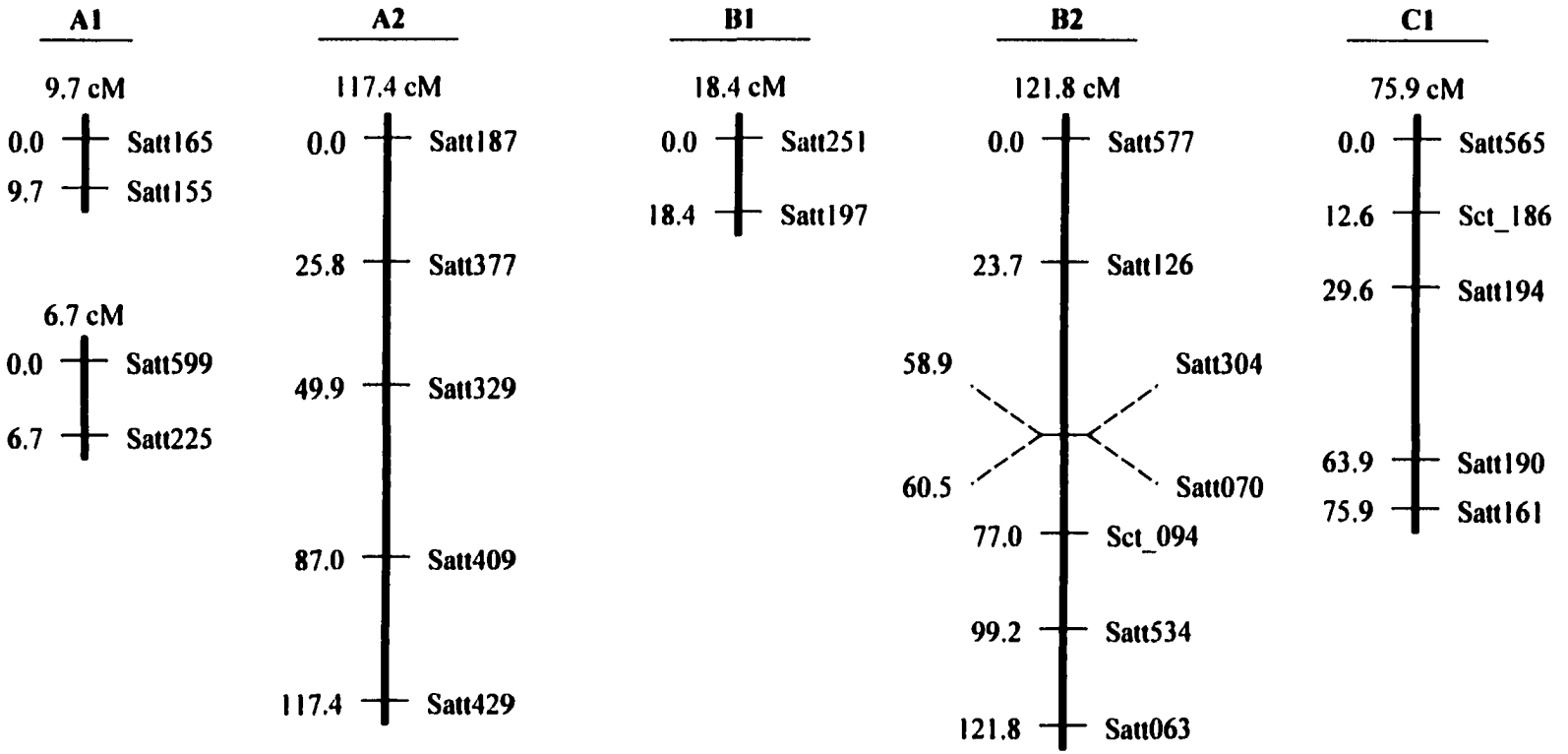


Figure D3. Linkage map constructed for Population 3 using SSR markers.

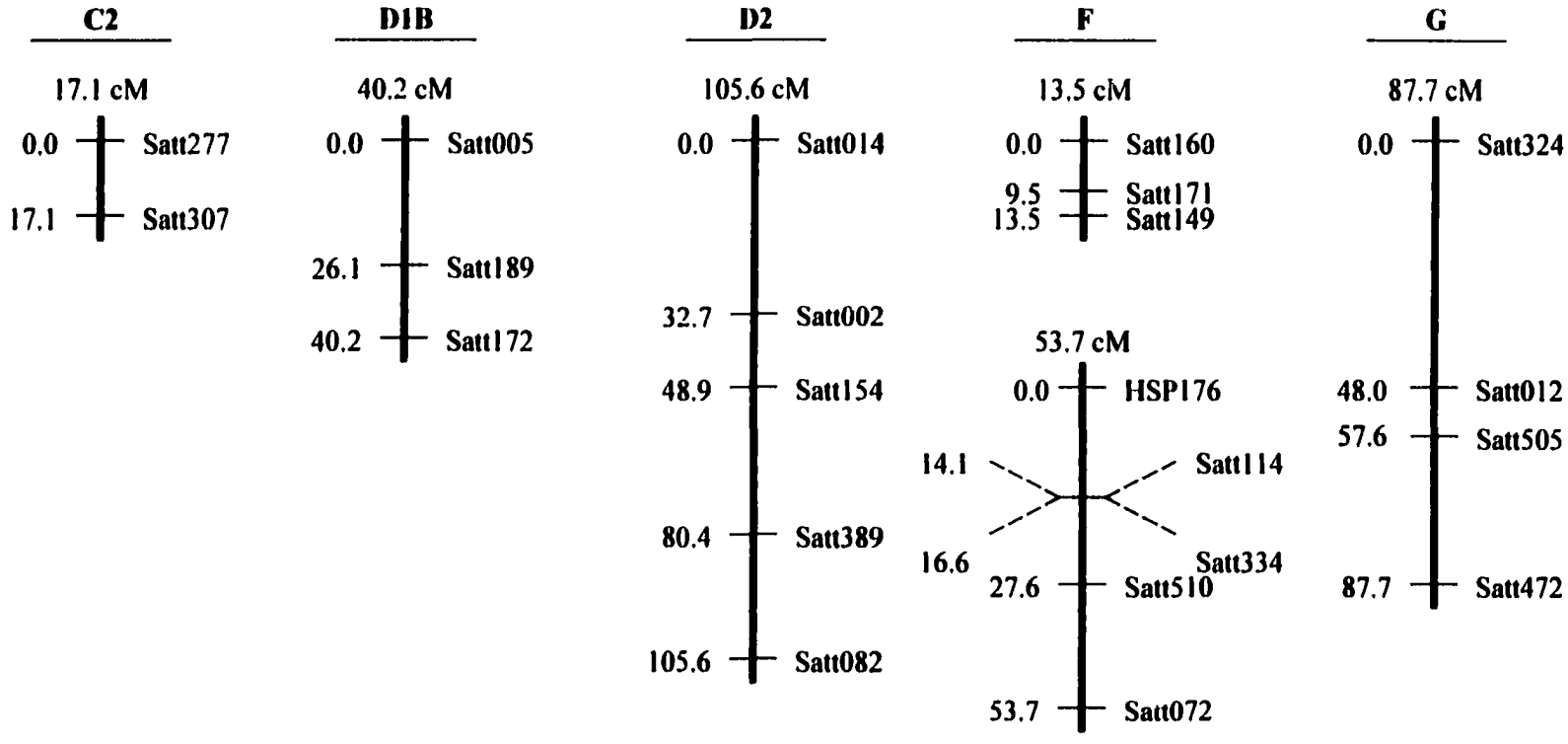


Figure D3. Continued.

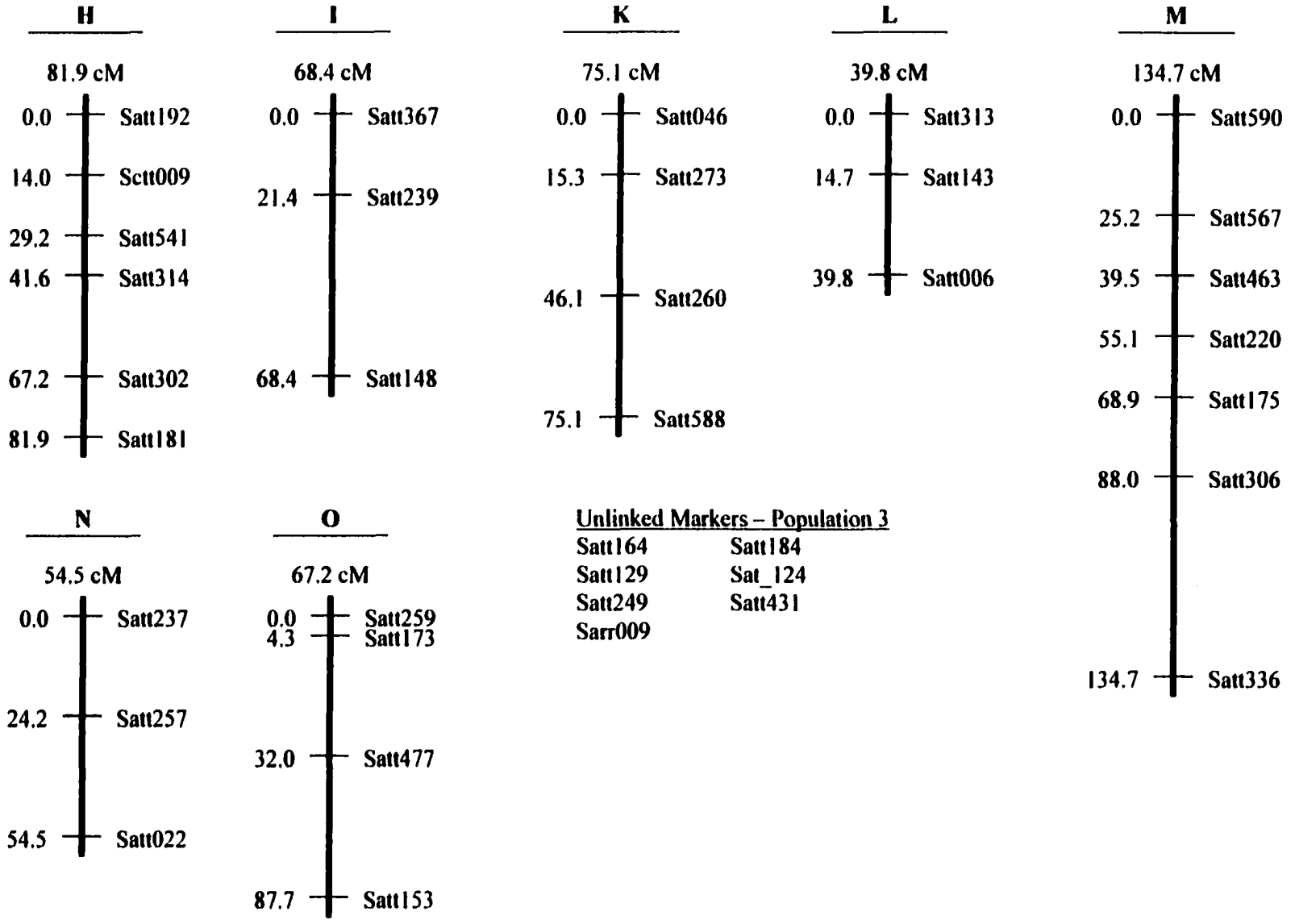


Figure D3. Continued.

APPENDIX E

MARKER LOCI SIGNIFICANTLY ASSOCIATED WITH SEED SIZE FOR EACH POPULATION ACROSS ENVIRONMENTS USING SINGLE-FACTOR ANALYSIS OF VARIANCE

Table E1. Marker loci significantly associated with seed size for 100 F_2 plants grown and Ames, their F_2-derived lines grown at three environments, and combined across environments for population 1 using single-factor analysis of variance.

Marker	Environment	LG†	Allelic mean NN‡	NS	SS	P§	R^2#	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt409	F_2 plants	A2	117	108	108	0.0180	8.0	-4.35 *	-4.63
	F_2-derived lines								
	Agronomy		108	101	99	0.0046	10.6	-4.59 **	-1.18
	Burkey		105	96	96	0.0022	12.0	-4.88 **	-2.11
	Puerto Rico		135	127	125	0.0098	9.2	-5.34 **	-0.68
	Mean		116	108	107	0.0027	11.6	-4.93 **	-------
Satt070	F_2 plants	B2	114	112	104	0.0050	10.3	-5.08 **	2.94
	F_2-derived lines								
	Agronomy		107	102	97	0.0007	13.9	-3.95 **	0.04
	Burkey					NS			
	Puerto Rico		134	128	123	0.0063	9.9	-5.45 **	0.07
	Mean		114	110	105	0.0049	10.4	-4.69 **	-------
Satt322	F_2 plants	C2	111	107	119	0.0028	11.9	4.00 *	-7.07 **
	F_2-derived lines								
	Agronomy		104	100	109	0.0009	14.0	2.57	-3.42 **
	Burkey		100	95	106	0.0002	16.8	3.28 *	-3.99 **
	Puerto Rico		128	126	136	0.0072	10.1	4.07 *	-1.55 *
	Mean		110	107	117	0.0006	14.6	3.33 *	-------

*,** Significant at the 0.05 and 0.01 probability levels, respectively.

† Linkage group as designated in the current USDA-ISU map.

‡ NN-homozygous normal-seeded parent; NS-heterozygous; SS-homozygous small-seeded parent measured as mg seed^{-1}.

§ Probability value, NS = not significant.

Percent phenotypic variation explained by the marker locus.

Table E1. Continued.

Marker	Environment	LG	Allelic mean NN	Allelic mean NS	Allelic mean SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt077	F_2 plants	D1A				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey		96	97	105	0.0100	8.9	4.47**	-1.87
	Puerto Rico					NS			
	Mean					NS			
Satt002	F_2 plants	D2				NS			
	F_2-derived lines								
	Agronomy		107	102	96	0.0033	11.1	-5.60**	0.41
	Burkey		103	98	93	0.0112	8.8	-5.43**	-0.01
	Puerto Rico					NS			
	Mean		114	109	104	0.0167	8.1	-5.05**	-------
Satt154	F_2 plants	D2				NS			
	F_2-derived lines								
	Agronomy		108	101	97	0.0013	12.8	-5.07**	-0.92
	Burkey		103	96	94	0.0036	11.0	-4.84**	-1.20
	Puerto Rico		134	126	124	0.0060	10.0	-5.18**	-0.78
	Mean		115	108	105	0.0016	12.5	-5.03**	-------
Satt185	F_2 plants	E	111	114	106	0.0075	10.2	-2.47	5.96*
	F_2-derived lines								
	Agronomy					NS			
	Burkey		95	101	95	0.0166	8.6	0.04	3.28*
	Puerto Rico					NS			
	Mean		106	112	106	0.0172	8.5	-0.10	-------

Table E1. Continued.

Marker	Environment	LG	Allelic mean NN	NS	SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt045	F_2 plants	E	110	115	105	0.0014	12.9	-2.28	7.63 **
	F_2-derived lines								
	Agronomy		99	106	100	0.0128	8.8	0.61	3.24 **
	Burkey		95	102	94	0.0023	12.0	-0.25	3.81 **
	Puerto Rico		126	132	124	0.0080	9.7	-0.71	1.90 **
	Mean		106	113	106	0.0035	11.2	-0.13	-------
Satt510	F_2 plants	F	113	113	104	0.0062	9.9	-4.56 **	3.82
	F_2-derived lines								
	Agronomy		108	103	97	0.0002	16.1	-5.52 **	0.27
	Burkey		104	99	92	0.0030	15.2	-5.76 **	0.29
	Puerto Rico		133	1239	123	0.0066	9.8	-5.26 **	0.34
	Mean		115	110	104	0.0004	14.9	-5.52 **	-------
Satt001	F_2 plants	K				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey		96	98	106	0.0145	8.6	5.02 **	-1.65
	Puerto Rico					NS			
	Mean		107	109	117	0.0181	8.2	4.77 **	-------
Satt273	F_2 plants	K				NS			
	F_2-derived lines								
	Agronomy		95	102	107	0.0008	14.0	5.99 **	0.48
	Burkey		93	98	103	0.0111	9.0	5.12 **	0.13
	Puerto Rico		120	128	109	0.0009	13.7	7.42 **	0.23
	Mean		103	109	115	0.0010	13.5	6.15 **	-------

Table E1. Continued.

Marker	Environment	LG	Allelic mean NN	Allelic mean NS	Allelic mean SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt551	F_2 plants	M				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		124	132	126	0.0100	8.8	0.67	1.89 **
	Mean					NS			

Table E2. Marker loci significantly associated with seed size for 100 F_2 plants grown and Ames, their F_2-derived lines grown at three environments, and combined across environments for population 2 using single-factor analysis of variance.

Marker	Environment	LG†	Allelic mean NN‡	NS	SS	P§	R^2#	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt070	F_2 plants	B2	132	124	116	0.0002	16.2	-7.97 **	-0.16
	F_2-derived lines								
	Agronomy		107	101	95	0.0001	17.5	-6.11 **	0.02
	Burkey		105	96	91	0.0001	21.0	-7.08 **	-0.20
	Puerto Rico		140	135	129	0.0114	8.8	-5.34 **	0.04
	Mean		117	111	105	0.0001	17.5	-6.18 **	-------
Satt534	F_2 plants	B2	128	123	114	0.0037	14.4	-6.92 **	2.52
	F_2-derived lines								
	Agronomy					NS			
	Burkey		99	99	90	0.0077	12.6	-4.71 **	2.02
	Puerto Rico					NS			
	Mean					NS			
Satt565	F_2 plants	C1				NS			
	F_2-derived lines								
	Agronomy		105	102	97	0.0083	9.8	-4.03 **	0.36
	Burkey					NS			
	Puerto Rico					NS			
	Mean		115	112	107	0.0126	9.0	-3.83 **	-------

*,** Significant at the 0.05 and 0.01 probability levels, respectively.

† Linkage group as designated in the current USDA-ISU map.

‡ NN-homozygous normal-seeded parent; NS-heterozygous; SS-homozygous small-seeded parent measured as mg seed^{-1}.

§ Probability value, NS = not significant.

Percent phenotypic variation explained by the marker locus.

Table E2. Continued.

Marker	Environment	LG	Allelic mean NN	NS	SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt227	F_2 plants	C2				NS			
	F_2-derived lines								
	Agronomy		102	98	105	0.0167	8.1	1.46	-2.48**
	Burkey		96	96	102	0.0176	8.0	2.96*	-1.89
	Puerto Rico					NS			
	Mean					NS			
Satt277	F_2 plants	C2				NS			
	F_2-derived lines								
	Agronomy		97	99	104	0.0069	10.1	3.89**	-0.51
	Burkey		93	96	102	0.0033	11.5	4.30**	-0.78
	Puerto Rico		129	133	139	0.0058	10.4	4.82**	-0.36
	Mean		106	109	115	0.0019	12.5	4.33**	-------
Satt184	F_2 plants	D1A				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey		103	97	94	0.0100	9.0	-4.59**	-0.82
	Puerto Rico					NS			
	Mean					NS			
Sctt008	F_2 plants	D2	124	126	116	0.0064	9.9	-3.83*	6.40*
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		131	138	129	0.0021	12.0	-1.18	2.02**
	Mean		108	114	107	0.0073	9.6	-0.67	-------

Table E2. Continued.

Marker	Environment	LG	Allelic mean NN	NS	SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt135	F_2 plants	D2	125	127	118	0.0055	10.5	-3.79 *	5.60 *
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico					NS			
	Mean					NS			
Satt185	F_2 plants	E	111	125	124	0.0160	8.5	6.25 *	7.67 *
	F_2-derived lines								
	Agronomy		91	102	102	0.0136	8.8	5.02 **	2.72 *
	Burkey					NS			
	Puerto Rico					NS			
	Mean					NS			
Satt431	F_2 plants	J				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		139	136	129	0.0037	10.9	-5.14 **	0.49
	Mean					NS			
Satt166	F_2 plants	L	130	126	115	0.0001	19.8	-7.50 **	3.85
	F_2-derived lines								
	Agronomy		108	101	95	0.0001	20.1	-6.35 **	-0.06
	Burkey		106	98	90	0.0001	27.9	-7.96 **	-0.19
	Puerto Rico		141	135	129	0.0011	13.4	-6.29 **	-0.11
	Mean		119	111	105	0.0001	23.2	-6.87 **	-------

Table E2. Continued.

Marker	Environment	LG	Allelic mean NN	NS	SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Sat_099	F_2 plants	L	134	126	116	0.0001	22.4	-9.29 **	0.74
	F_2-derived lines								
	Agronomy		111	101	94	0.0001	34.9	-8.50 **	-0.68
	Burkey		110	98	90	0.0001	37.7	-9.66 **	-0.95
	Puerto Rico		146	135	128	0.0001	23.2	-8.80 **	-0.49
	Mean		122	111	104	0.0001	36.5	-8.99 **	-------
Satt006	F_2 plants	L	130	124	117	0.0023	11.8	-6.69 **	1.04
	F_2-derived lines								
	Agronomy		110	100	95	0.0001	28.4	-7.77 **	-1.07
	Burkey		108	97	91	0.0001	30.4	-8.51 **	-1.09
	Puerto Rico		143	134	130	0.0004	14.9	-6.83 **	-0.71
	Mean		120	110	105	0.0001	27.5	-7.70 **	-------
Satt373	F_2 plants	L				NS			
	F_2-derived lines								
	Agronomy		104	101	97	0.0043	11.0	-3.85 **	0.38
	Burkey					NS			
	Puerto Rico		139	135	130	0.0116	9.1	-4.41 **	0.09
	Mean		115	111	107	0.0050	10.8	-3.88 **	-------
Satt336	F_2 plants	M				NS			
	F_2-derived lines								
	Agronomy		102	104	96	0.0017	12.3	-3.25 *	2.19 *
	Burkey		98	100	93	0.0084	9.4	-2.80 *	2.19 *
	Puerto Rico		139	136	129	0.0017	12.3	-5.36 **	0.52
	Mean		113	113	106	0.0015	12.6	-3.80 **	-------

Table E2. Continued.

Marker	Environment	LG	Allelic mean NN	Allelic mean NS	Allelic mean SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt173	F_2 plants	O				NS			
	F_2-derived lines								
	Agronomy		101	103	96	0.0187	7.9	-2.31	2.28*
	Burkey					NS			
	Puerto Rico					NS			
	Mean		111	113	106	0.0194	7.8	-2.42	-------

Table E3. Marker loci significantly associated with seed size for 100 F_2 plants grown and Ames, their F_2-derived lines grown at three environments, and combined across environments for population 3 using single-factor analysis of variance.

Marker	Environment	LG†	Allelic mean NN‡	NS	SS	P§	R^2#	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt187	F_2 plants	A2				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		137	134	126	0.0001	17.2	-5.45 **	0.64
	Mean		111	109	104	0.0063	10.1	-3.22 **	-------
Satt304	F_2 plants	B2				NS			
	F_2-derived lines								
	Agronomy		101	98	94	0.0029	11.4	-3.28 **	0.12
	Burkey					NS			
	Puerto Rico		137	135	126	0.0001	17.1	-5.27 **	0.73
	Mean		112	109	104	0.0002	16	-4.05 **	-------
Satt070	F_2 plants	B2				NS			
	F_2-derived lines								
	Agronomy		101	98	94	0.0027	11.7	-3.14 **	0.13
	Burkey		98	95	91	0.0100	9.2	-3.69 **	-0.02
	Puerto Rico		137	135	127	0.0002	16.8	-4.94 **	0.73
	Mean		112	109	104	0.0001	17.1	-3.92 **	-------

*,** Significant at the 0.05 and 0.01 probability levels, respectively.

† Linkage group as designated in the current USDA-ISU map.

‡ NN-homozygous normal-seeded parent; NS-heterozygous; SS-homozygous small-seeded parent measured as mg seed^{-1}.

§ Probability value, NS = not significant.

Percent phenotypic variation explained by the marker locus.

Table E3. Continued.

Marker	Environment	LG	Allelic mean NN	Allelic mean NS	Allelic mean SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Sct_094	F_2 plants	B2				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		135	134	127	0.0071	12.5	-4.02 **	0.77
	Mean					NS			
Satt565	F_2 plants	C1				NS			
	F_2-derived lines								
	Agronomy		101	96	101	0.0006	14.5	-0.26	2.73 **
	Burkey		97	92	100	0.0021	12.3	1.39	-3.20 **
	Puerto Rico		139	131	135	0.0095	9.4	-1.95	-1.45 **
	Mean		113	106	112	0.0005	15.1	-0.28	-------
Satt184	F_2 plants	D1A				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		136	135	127	0.0100	9.4	-4.42 **	0.89
	Mean					NS			
Satt172	F_2 plants	D1B				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		132	136	129	0.0094	9.2	-1.09	1.37 **
	Mean					NS			

Table E3. Continued.

Marker	Environment	LG	Allelic mean NN	NS	SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt154	F_2 plants	D2				NS			
	F_2-derived lines								
	Agronomy		99	99	94	0.0069	10.5	-2.62 **	1.30
	Burkey		98	97	89	0.0006	15.1	-4.33 **	1.97 *
	Puerto Rico		135	136	127	0.0019	13.0	-3.86 **	1.22 *
	Mean		111	111	104	0.0001	17.9	-3.60 **	-------
HSP176	F_2 plants	F				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		137	134	126	0.0003	15.3	-5.50 **	0.55
	Mean					NS			
Satt114	F_2 plants	F				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		137	136	126	0.0001	22.0	-0.80	0.82 *
	Mean		110	111	105	0.0050	11.2	-5.87 **	-------
Satt334	F_2 plants	F				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		135	137	127	0.0002	17.6	-4.01 **	1.47 **
	Mean					NS			

Table E3. Continued.

Marker	Environment	LG	Allelic mean NN	NS	SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt510	F_2 plants	F				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		136	135	126	0.0002	16.3	-4.97 **	1.20 *
	Mean					NS			
Satt072	F_2 plants	F				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		136	134	126	0.0042	10.7	-5.15 **	0.71
	Mean					NS			
Sctt009	F_2 plants	H				NS			
	F_2-derived lines								
	Agronomy		101	99	95	0.0049	10.6	-3.21 **	0.50
	Burkey		102	95	91	0.0002	12.2	-5.15 **	-0.72
	Puerto Rico		135	135	129	0.0083	9.6	-3.11 *	0.90
	Mean		113	110	105	0.0012	13.3	-3.82 **	-------
Satt541	F_2 plants	H				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey		97	97	91	0.0130	8.9	-3.13 *	1.30
	Puerto Rico					NS			
	Mean		110	111	106	0.0090	9.7	-2.16	-------

Table E3. Continued.

Marker	Environment	LG	Allelic mean NN	NS	SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt314	F_2 plants	H				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		138	135	129	0.0134	9.3	-4.14 **	0.30
	Mean		113	110	106	0.0085	10.3	-3.43 **	-------
Satt302	F_2 plants	H				NS			
	F_2-derived lines								
	Agronomy		101	96	96	0.0191	8.4	-2.45 *	-1.12
	Burkey		99	92	94	0.0164	8.7	-2.69 *	-2.09 *
	Puerto Rico					NS			
	Mean		112	107	107	0.0176	8.6	-2.55 *	-------
Satt006	F_2 plants	L				NS			
	F_2-derived lines								
	Agronomy		105	98	94	0.0001	20.9	-5.50 **	-0.40
	Burkey		104	96	89	0.0001	23.6	-7.53 **	-0.09
	Puerto Rico		142	134	128	0.0002	17.8	-7.05 **	-0.23
	Mean		117	110	104	0.0001	28.8	-6.68 **	-------
Satt143	F_2 plants	L				NS			
	F_2-derived lines								
	Agronomy		100	98	93	0.0071	9.8	-3.64 **	0.81
	Burkey		99	95	88	0.0032	11.3	-5.30 **	0.65
	Puerto Rico		138	134	123	0.0001	22.6	-7.63 **	0.88
	Mean		113	109	101	0.0001	20.0	-5.53 **	-------

Table E3. Continued.

Marker	Environment	LG	Allelic mean NN	NS	SS	P	R^2	Additive Estimate	Dominant Estimate
			----------mg seed^{-1}---------				---%--		
Satt336	F_2 plants	M	117	123	129	0.0033	12.7	6.42 **	-0.27
	F_2-derived lines								
	Agronomy		94	97	102	0.0004	16.9	3.92 **	-0.51
	Burkey		90	93	100	0.0013	14.6	4.76 **	-0.70
	Puerto Rico		128	133	137	0.0082	10.8	4.70 **	0.16
	Mean		104	108	113	0.0002	18.3	4.46 **	-------
Satt009	F_2 plants	N				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey		90	98	94	0.0060	10.1	1.92	2.99 **
	Puerto Rico					NS			
	Mean					NS			
Satt237	F_2 plants	N				NS			
	F_2-derived lines								
	Agronomy					NS			
	Burkey					NS			
	Puerto Rico		129	137	131	0.0049	10.7	0.62	1.64 **
	Mean					NS			

APPENDIX F

COST COMPARISON BETWEEN COLLECTING PHENOTYPIC AND MARKER DATA

Table F7. Cost comparison between collecting phenotypic and marker data.

Phenotypic Analysis		Molecular Marker Analysis	
Item	Amount/EU†	Item	Amount/EU
Equipment‡	0.05	Equipment	0.03
Supplies§	---------	Supplies	0.35
Labor¶	0.30	Gel#	1.50/.50/.25
		Labor	0.12
Total	0.35	Total	2.00/1.00/.75

† Amount is calculated based on cost (dollars) per experimental unit (EU) tested.

‡ Equipment costs were based on the price of the item divided by the number of EUs processed in 1 yr over a 20 yr operating life.

§ Includes supplies needed to collect data on EUs (DNA extraction reagents, Taq Gold, DNTPs, etc.)

¶ Labor costs were calculated based on the number of EUs collected in one hour based on a $10.00 per hour labor fee.

Gel costs were calculated based on the number of molecular markers multiplexed per gel lane (one marker per lane/three markers per lane/six markers per lane).

ACKNOWLEDGMENTS

First, I thank my wife Tiffany for her love, support, and encouragement during the completion of my graduate study. I would also like to thank my parents for their guidance and for instilling in me a goal oriented work ethic. I thank Dr. Walter Fehr for his instruction and leadership in matters both professional and personal. Finally, I would like to thank Grace Welke, Susan Johnson, Kevin Scholbrock, Bruce Voss, and all the graduate and undergraduate students affiliated with the soybean project for making graduate school an enjoyable experience.