SSR-Marker Analysis of the Intracultivar Phenotypic Variation Discovered within 3 Soybean Cultivars

JENNIFER L. YATES, H. ROGER BOERMA, AND VASILIA A. FASOULA

From the Monsanto Corp, St. Louis, Missouri (Yates); and the Center for Applied Genetic Technologies, University of Georgia, 111 Riverbend Road, Athens, GA 30602-6810 (Boerma and Fasoula).

Address correspondence to Dr Vasilia A. Fasoula at the address above, or e-mail: vfasoula@uga.edu.

Abstract

Genetic variation within homogeneous gene pools in various crops is assumed to be very limited. One objective of this study was to use 144 simple sequence repeat (SSR) markers to determine if the single-plant lines selected at ultra-low plant density in honeycomb designs within the soybean cultivars Benning, Haskell, and Cook had unique SSR genetic fingerprints. Another objective was to investigate if the variation found was the result of residual genetic heterozygosity that could be detected in the original gene pool where selection initiated. Our results showed that the phenotypic variation for seed protein content and seed weight has a genotypic component identified by the SSR band variation. The 7 lines from Haskell had a total of 63 variant alleles, the 5 lines from Benning had 34 variant alleles, and the 7 lines from Cook had 34 variant alleles, therefore, possessing unique genetic fingerprints. Most of the intracultivar SSR band variation discovered was the result of residual heterozygosity in the initial plant selected to become the cultivar. More specifically, 82% of the SSR variant alleles were traced in the Benning Foundation seed source, 93% in the Haskell seed source, and 82% in the Cook seed source. The remaining variant bands (18% for Benning, 7% for Haskell, and 18% for Cook) could not be detected in the Foundation seed source and were likely the result of mutation or some other mechanism generating de novo variation. These results provide evidence that genetic variation among individual plants is present even in homogeneous gene pools and can be further utilized in breeding programs.

Key words: genomics, intracultivar variation, soybean, SSR markers, ultra-low plant density

Most breeding programs in soybean, Glycine max (L.) Merr., and other self-pollinated crops limit the use of parents to those improved for a variety of traits. Despite the fact that the genetic base for major crops continues to narrow, sizable genetic gains continue for all the major crops (Rasmusson and Phillips 1997). Data from selection experiments within fairly homogeneous gene pools and from long-term selection experiments and double haploid studies (Russell et al. 1963; Byth and Weber 1968; Dudley and Lambert 1992) suggest that the genome is more variable than previously assumed. Research on the continued selection within elite cultivars for agronomic and other seed traits has been restricted by the belief that released cultivars are highly homogeneous and, therefore, selection within cultivars is likely ineffective. Although elite cultivars are fairly homogeneous, it is also documented that latent genetic variation among the single plants of a cultivar exists and mechanisms that generate newly created variation may also be present. Fasoula and Boerma (2005, 2007) were able to find a significant amount of intracultivar variation for agronomic and seed traits by performing single-plant selection at ultra-low plant density within plants that were grown in honeycomb selection designs (Fasoulas and Fasoula 1995).

Honeycomb breeding uses advanced experimental designs and wide interplant spacing to allow selection of superior individual plants under conditions that exhibit their full genetic potential (Fasoula VA and Fasoula DA 2000). This planting pattern eliminates interplant competition, which in certain cases has been negatively correlated with yielding ability (Fasoula DA and Fasoula VA 1997), and it therefore allows plant breeders to evaluate effectively the yield potential of individual plants. When selection at a low plant density is applied to individual plants within a previously released cultivar, the goal is both to avoid yield deterioration and to exploit the outcome of favorable mutations or other genetic modifications that have occurred within the cultivar over time (Fasoula 1990; Fasoula VA and Fasoula DA 2000, 2002) as well as the role of adaptive variation (Fasoula 2008).
Honeycomb selection has been used effectively to select within elite cultivars of various crops (Fasoula 1990; Fasoulas 2000; Christakis and Fasoulas 2002; Tokatlidis et al. 2004, 2005; Fasoula and Boerma 2005, 2007). In soybean, significant intracultivar variation was discovered for seed protein and oil, seed weight, maturity, and plant height within the cultivars Benning, Haskell, and Cook (Fasoula and Boerma 2005, 2007). In the case of seed protein content, the single-plant selected lines varied from 13 to 24 g kg⁻¹ (Fasoula and Boerma 2005), whereas the seed weight varied from 22 to 45 mg seed⁻¹ (Fasoula and Boerma 2007). Some of these lines provided unique genetic material for seed and agronomic traits and were released as soybean germplasm lines (Fasoula et al. 2007a, 2007b, 2007c). All 3 cultivars revealed a significant amount of intracultivar variation at the single-plant level, with Haskell consistently exhibiting more variation than the other 2 cultivars, Benning and Cook. Evidence from selection experiments within fairly homogeneous genetic pools suggests that the genome is more flexible and plastic than previously assumed. Rasmussen and Phillips (1997) reported that elite gene pools have inherent mechanisms to provide a continuing source of new genetic variation and hypothesized that selection gain occurs due to variation present in the original gene pool as well as due to de novo generated variation.

Molecular markers are increasingly used to expedite the breeding process and can be used to further investigate the nature of intracultivar variation. Current developments in genome sequencing and the relative ease with which sets of molecular markers can now be developed have allowed researchers to understand even single-base pair variations among plants. Many studies have explored the genetic variation across a taxonomic level by evaluating close relatives of a cultivated species for genes enhancing yield or disease resistance (Hawkes 1977; Singh and Hymowitz 1999). Others have studied diversity among cultivars within a single species, such as the analysis of differences between Korean and Chinese soybean cultivars by Abe et al. (2003). At the molecular level, studies have reported the presence of allelic variation at simple sequence repeat (SSR) loci in inbred material of sunflower (Helianthus annuus L.; Zhang et al. 1995), rice (Oryza sativa L.; Olufowote et al. 1997; Cho et al. 1998; Nandakumar et al. 2004), chickpea (Cicer arietinum L.; Udupa and Baum 2001), melon, (Cucumis melo; Pépin et al. 2002), and maize (Zea mays L.; Senior et al. 1998; Berry et al. 2002; Mutsuoka et al. 2002; Banthe and Prasanna 2003). A previous study evaluating the use of SSRs in soybean reported SSR band variation within inbred material (Diwan and Cregan 1997). Lately, Haun et al. (2011) reported high rates of intracultivar structural variation within the soybean reference cultivar “Williams 82.” They concluded that soybean haplotypes can possess a high rate of structural and gene content variation and the impact of intracultivar genetic heterogeneity may be significant.

The main objective of this study was to use SSR markers to determine if the single-plant lines selected within the elite cultivars Benning, Haskell, and Cook at ultra-low plant density in honeycomb designs (Fasoula and Boerma 2005, 2007) had unique SSR genetic fingerprints. The second objective was to investigate if the variation found in these single-plant selected lines was the result of residual heterozygosity that could be detected in the original gene pools where selection was initiated. For this purpose, SSR markers were used to analyze the marker profiles of the single-plant progeny lines selected within Benning, Haskell, and Cook and compared with the SSR profiles of individual seeds from the seed source used in the initial honeycomb field trial. The parents of the 3 cultivars Benning, Haskell, and Cook were also analyzed using SSR markers to determine if some amount of the intracultivar variation can be found in the parents of the cultivars.

### Materials and Methods

#### Genotyping the Single-Plant Selected Lines

Details of the honeycomb selection, honeycomb field trials of the single-plant selections, and the subsequently developed soybean lines that we evaluated in this study have been described by Fasoula and Boerma (2005, 2007). Nineteen of these lines were chosen for analysis with SSR markers based on their significant divergence in seed protein content and seed weight from their corresponding source cultivar (Table 1). Cook and Haskell are F₅-derived lines that were released in 1991 and 1993, respectively (Boerma et al. 1992, 1994), whereas Benning is an F₄-derived line that was released in 1995 (Boerma et al. 1997). The single-plant selected lines developed from within these cultivars are hereafter referred to by the abbreviations “B” (Benning), “H” (Haskell), or “C” (Cook), according to the original cultivar from which the single plant was selected, followed by a unique numeric identifier.

DNA was extracted from a bulked sample of 30 seeds from each of the 19 single-plant selected lines and from each source cultivar. The seed source of the cultivars that was used for DNA extraction was from the original seed lot used in the honeycomb trial (1994 Foundation seed of each cultivar). A total of 30 seeds from each single-plant selected line or source cultivar were pulverized in a coffee grinder, and approximately 0.1 mg of the ground tissue was transferred to a 1.5-ml microcentrifuge tube. Genomic DNA was extracted from the seed tissue following the protocol of Edwards et al. (1991). The DNA was diluted to approximately 50 ng µl⁻¹ and used in a standard PCR for microsatellites as described in Diwan and Cregan (1997).

The SSR markers used in this study were chosen for 1) their proximity to previously reported seed weight or protein quantitative trait loci (QTL) in soybean as reported in the online database Soybase (http://soybase.org), 2) their high polymorphism index when used in diverse soybean accessions (also known as the PIC score) (Akkaya et al. 1992), and 3) their locations to achieve a distribution across the soybean genome (Song et al. 2004). A total of 144 primers were used in this study, with at least 3 SSR markers and in some cases up to 11 SSR markers tested from each plant source.
The markers had a blue (6-FAM), yellow (NED), or green (HEX) fluorescent tag label (PE-ABI, Foster City, CA). PCR cycling conditions were the same as those used in Diwan and Cregan (1997) with a PE-ABI 9700 thermocycler (PE-ABI, Foster City, CA). GeneScan® 3.1.2 software (PE-ABI, Foster City, CA) was used to determine the size of the products according to the local Southern option with the ROX 500 ladder (PE-ABI, Foster City, CA) as the size standard.

### Genotyping the Individual Seeds in the Original Cultivar Seed Source

DNA was extracted from a bulked sample of 30 seeds from each source cultivar as detailed before. We evaluated SSR band variation in remnant seeds from the original seed lot that was used to plant the honeycomb trials and make the single-plant selections. In order to test efficiently a large number of seeds, DNA from 4 seeds were combined and genotyped as one pooled sample. The number of seeds to be combined in this manner was chosen based on dilution tests of DNA mixed in known proportions from different cultivars. A mixture of 3 parts of DNA from one cultivar to one part of DNA from a different cultivar reliably detected the SSR band of the minority cultivar. This mimicked the situation where 3 seeds might contain the original cultivar band, whereas the other seed could have a different band. Following this logic, 372 seeds of each cultivar were divided into 93 pools of 4 seeds each. A ¼ seed chip was cut from the side of each soybean seed opposite the embryo, and four ¼ seed chips were placed in each well of a 96-well deep-well plate. The remaining ¼ seeds were kept in separate labeled envelopes for later analysis. The seed chips were soaked overnight in 100 μL of sterilized distilled and deionized water and were then macerated either with a glass stirring rod or a custom-made crushing device with 96 steel pins. DNA extraction followed the protocol in Edwards et al. (1991). PCR conditions and the visualization of PCR products proceeded as described above. A total of 75 SSR markers were tested on the pools.

### Genotyping the Parents of the 3 Cultivars

The parents of each of the 3 cultivars, Haskell, Benning, and Cook, were genotyped to determine whether the SSR band variation found in a given cultivar could have originated from its parents. Benning originated from a cross of “Hutcheson” × “Coker 6738,” Haskell from “Braxton” × “Johnston,” and Cook from “Braxton” × “Young” (Boerma et al. 1992, 1994, 1997).

The actual seed source of each parent cultivar originally used to create these crosses was no longer available. Instead, the most current seed source available from the University of Georgia soybean breeding program for each parent was genotyped. The seed source for Johnston was harvested in 2000, whereas the other 4 parents came from seeds harvested in 2003. Four seeds of each parent were planted in commercially available planting soil mix in large Styrofoam cups in the greenhouse. Approximately 2 weeks after germination, a small trifoliolate leaflet was collected from each of the plants in each cup and was freeze-dried.

### Table 1

| Source cultivar and selections | Protein content (g kg⁻¹) | Seed weight (mg seed⁻¹) |
|-------------------------------|--------------------------|-------------------------|
| Benning cultivar              | 410                      | 156                     |
| Benning selections            |                          |                         |
| B-335                         | 411                      | 167                     |
| B-1818                        | 406                      | 137                     |
| B-2227                        | 410                      | 143                     |
| B-4048                        | 407                      | 159                     |
| B-4123                        | 409                      | 146                     |
| LSD (0.05)                    | 3                        | 5                       |
| Cook cultivar                 | 402                      | 154                     |
| Cook selections               |                          |                         |
| C-319                         | 422                      | 145                     |
| C-1035                        | 412                      | 143                     |
| C-1346                        | 431                      | 158                     |
| C-2116                        | 413                      | 156                     |
| C-3038                        | 430                      | 157                     |
| C-3614                        | 430                      | 157                     |
| C-3746                        | 429                      | 162                     |
| LSD (0.05)                    | 4                        | 4                       |
| Haskell cultivar              | 406                      | 156                     |
| Haskell selections            |                          |                         |
| H-425                         | 424                      | 157                     |
| H-515                         | 419                      | 165                     |
| H-736                         | 404                      | 149                     |
| H-1112                        | 400                      | 156                     |
| H-1536                        | 409                      | 186                     |
| H-2646                        | 403                      | 154                     |
| H-4243                        | 414                      | 194                     |
| LSD (0.05)                    | 3                        | 5                       |

* Least significant difference for comparisons of the cultivar mean to the mean of a selected line.
overnight. Altogether, a single leaflet from each of 76 individual plants, or 19 pools of 4 leaflets each, were tested from each of the 5 parents. Four small leaflets were placed in each well of a 96-well deep-well plate, along with a zinc-plated 4.5-mm BB (Daisy Outdoor Products, Rogers, AR). Each plate was placed in a custom-made leaf grinder and shaken (exact revolution per minute unknown) for approximately 1 min to pulverize the tissue. DNA extraction proceeded as described in Keim et al. (1988). Subsequent PCR and gel electrophoresis conditions were the same as those described above. A total of 78 SSRs were used to genotype the parents of their cultivars.

Results and Discussion

SSR Band Variation Detected in the Single-Plant Selected Lines

The 19 single-plant selected lines developed from within the cultivars Benning, Haskell, and Cook were analyzed molecularly using 144 SSR markers, and the results are shown in Figure 1. Of these 144 primers tested, 83 SSRs, or 58% of the markers evaluated, showed at least one allelic difference between the single-plant selected line and the original cultivar. In some cases, the single-plant selected line was heterogeneous, with one band that was identical to that in the original cultivar, and a second band that was from an unknown source. In other cases, the original seed source from the cultivar was heterogeneous, whereas the single-plant selected lines were fixed for one of the 2 bands that were amplified in the original cultivar seed source. These data signify that all the 19 single-plant selected lines have unique SSR band profiles and therefore unique fingerprints. The 7 single-plant selected lines from Haskell had a total of 63 variant alleles, the 5 lines from Benning had 34 variant alleles, and the 7 lines from Cook had also 34 variant alleles. Therefore, as a group, the Haskell-selected lines had noticeably more SSR band variation than the Benning- or the Cook-selected lines (Figure 1).

These results are in good agreement with the phenotypic data for seed protein content and seed weight for the 19 single-plant progeny lines presented in Table 1. The intracultivar protein and seed weight variation found within the cultivar Haskell is larger than the phenotypic variation found in either Cook or Benning (Fasoula and Boerma 2005, 2007), and this is shown in the number of SSR allele variants among the selected lines within a cultivar (Figure 1). It is interesting to note that line H-425 has the largest number of variant SSR bands among the 7 single-plant selected lines from Haskell, and it was the most divergent from Haskell in protein content with 18 g kg⁻¹ significantly higher seed protein than Haskell as shown in Table 1. The other line H-4243, with the second largest number of variant SSR bands among the single-plant selected lines from Haskell, was the most divergent from Haskell in seed weight, and it had 17 mg seed⁻¹ larger seed weight than Haskell (Table 1). Line B-4123, with the second largest number of variant SSR bands among the 5 single-plant selected lines from Benning, was the most divergent from Benning in seed weight with 38 mg seed⁻¹ significantly larger seed weight than Benning (Table 1). Line C-3614, with the second largest number of variant SSR bands among the 7 single-plant selected lines from Cook, had 8 g kg⁻¹ significantly lower seed protein content and 12 mg seed⁻¹ significantly smaller seed weight than Cook (Table 1).

These results provide direct evidence that the phenotypic variation for seed protein content and seed weight found in
the 19 single-plant selected lines, developed from single-plant selection in a honeycomb field trial grown at ultra-low plant density, has a genotypic component identified by the SSR band variation. Therefore, not only are these lines unique phenotypically as described by Fasoula and Boerma (2005, 2007) and Fasoula et al. (2007a, 2007b, 2007c) but they also possess unique genetic fingerprints as shown in Figure 1.

SSR Band Variation Detected in the Original Cultivar Seed Source

The original foundation seed source used to make the single-plant selections shown in Figure 1 and Table 1 was subsampled to ascertain whether these seeds contained the same SSR band variation found in the single-plant selected lines. DNA was extracted from 93 pools of 4 single-seed samples from each cultivar. The pooled samples were tested with the SSR markers that had shown band size differences between the cultivars and their single-plant selected lines. If that variant band were identified in a pool, DNA was extracted from the remaining ¾ seed portion of each member of the pool and the SSR primers producing the variant band assayed individually on each member of the pool to confirm the presence of the variant band.

The results indicate that most of the variant alleles discovered in the single-plant selected lines were also found in the original cultivar seed source from which the lines were derived (Figure 2). More specifically, 82% of the SSR variant alleles discovered in the 5 single-plant progeny lines selected within the cultivar Benning were found in the pools of the Benning Foundation seed source, whereas 18% of SSRs found in the selected lines could not be detected in the Benning seed source. Similarly, 93% of the SSR variant alleles discovered in the 7 single-plant progeny lines selected within Haskell were found in the pools of the Haskell Foundation seed source, whereas 7% of them could not be detected in the Haskell seed source. Finally, 82% of the SSR variant bands discovered in the 7 single-plant progeny lines selected within the cultivar Cook were found in the pools of the Cook Foundation seed source, whereas 18% of them could not be detected in the Cook seed source.

These data provide evidence that most of the SSR band variation discovered in the single-plant selected lines from within the soybean cultivars Benning, Haskell, and Cook was found in the original gene pool (Foundation seed) of the cultivars where selection originated. This suggests that it is likely the result of residual heterozygosity in the F4 or F5 generation where a seed from a single plant was composited to become a cultivar.

To understand these results, we need to consider the method that the cultivars and their resultant breeder seed were developed. Cook and Haskell are F5-derived lines that were released in 1991 and 1993, respectively (Boerma et al. 1992, 1994), whereas Benning is an F4-derived line that was released in 1995 (Boerma et al. 1997). As an F4-derived cultivar, Benning is expected to maintain residual heterozygosity at approximately 12.5% of its loci that were polymorphic between its parents, whereas Cook and Haskell, as F5-derived cultivars, are expected to contain 6.25% heterozygosity. This means that if there were 100 polymorphic loci between the 2 inbred parents on the average, we would expect that a random F5 plant will still possess 6 heterozygous loci. If we inbred the progeny of that plant to homozygous loci, the resulting population will contain 64 unique genotypes for these 6 loci. Given that
only 35–40 F10:11 rows were bulked to create the actual breeder seed, one can expect a wide range in frequency among these 64 unique genotypes in the breeder seed lot. Therefore, our results are in agreement with the hypothesis and provide evidence that most of the SSR band variation discovered in the single-plant selected lines within Benning, Haskell, and Cook could be explained by residual heterozygosity in the initial F2 or F3 plant selected to become a cultivar and the creation of the breeder seed.

As shown in Figure 1, some of the single-plant selected lines contained a larger amount of variation when compared with the others within the group. In particular, C-1035 and H-425 have noticeably more SSR band variation than did the other lines (Figure 1).

This can be attributed to undetected residual heterozygosity or a possible seed contamination event. Extreme care had been taken to exclude obvious seed mixtures within the cultivars on the basis of flower color, pod wall color, and seed hilum color (Fasoula and Boerma 2005). In addition, Foundation seed sources were used, which are produced following the strict guidelines of seed certification (http://www.certifiedseed.org/). Therefore, C-1035 and H-425 are most likely rare selections with unique fingerprint profiles that could be the result of undetected residual heterozygosity in the original cultivar or the result of some kind of seed contamination event that happened in the process of the breeding cycle or the result of some mechanism generating de novo variation. Regardless, their value in a breeding program is great given their unique phenotype as reported by Fasoula and Boerma (2005, 2007).

We also investigated the presence of the variant bands in the parents of the 3 cultivars. Benning has “Hutcheson” and “Coker 6738” as its parents, Haskell has “Braxton” and “Johnston,” and Cook has “Braxton” and “Young.” DNA pools from individual plants from the parents of the 3 cultivars were tested in order to examine whether the variation that was presumed to be residual heterozygosity was also present in either parent. Seventy-six plants, 19 single-plant pools of 4 plants each, of Coker 6738 and Hutcheson, Braxton and Johnston, and Braxton and Young were analyzed with 78 SSR markers that also were tested on the 19 single plant-selected lines and the single-seed pools from the 3 cultivars. A significant number of the SSR band variants found in the single-plant selected lines were also found in one or the other parent of the cultivar they were selected from. Specifically, 94% of the variant alleles among the single-plant progeny lines selected within Benning were found in one or the other parent of Benning. Similarly, 50% of the variant alleles found in the Haskell selections and 24% of the variant alleles found in the Cook selections were also discovered in one or the other parent. This simply suggests that a large amount of residual heterozygosity discovered in the cultivars also was present in the inbred lines used as their parents and further indicates that inbred lines are not permanent genetic stocks but material that contains genetic variation that can be further exploited in a plant breeding or genomics project.

### SSR Band Variation not Detected in the Original Cultivar Seed Source

As shown in Figure 2, 18% of SSR variant bands found in the single-plant selected lines derived from within Benning could not be detected in the original Benning seed source. However, a percentage of these undetected variant bands in Benning were detected in one or the other parent of Benning, suggesting a lineal origin. Similarly, 7% of SSR variant bands found in the Haskell selections could not be detected in the Haskell seed source, and 18% of the variant bands found in the Cook selections could not be detected in the Cook seed source. Some of these undetected variant bands were found in the unique lines C-1035 and H-425. When the contributions from Benning’s parents and from these 2 lines were removed, there were a total of 2 unexplained variant Benning alleles, 6 variant Haskell alleles, and 2 variant Cook alleles. These variant bands could be the result of mutation or de novo variation.

Mutation at certain loci within the original seed lot could also result in low-level variation. Potential sources of mutation and newly created variation include mechanisms such as intragenic recombination, unequal crossing over, DNA methylation, and paramutation (Rasmusson and Phillips 1997). Mutation was suspected as a key source of variation in landmark studies with *N. tabacum* (East 1936) and Atlas barley (*Hordeum vulgare* L.; reviewed by Allard 1960). In more recent studies, mutation was suspected as the cause of variation in chickpea cultivars (Udupa and Baum 2001) and in certain soybean cultivars (Diwan and Cregan 1997; Priolli et al. 2002). With specific reference to microsatellites, the rate of mutation at an SSR locus is influenced by its genomic position, the number of microsatellite repeats, and the type of microsatellite repeat (Deka et al. 1999; Ellegren 2000; Schlötterer 2000; Santibáñez-Koref et al. 2001). Studies indicate that 65–96% of the mutations that occur in microsatellite loci involve a gain or loss of one repeat unit (Jin et al. 1996; Udupa and Baum 2001; Whittaker et al. 2003). This is believed to be the consequence of the DNA polymerase skipping over repeat units during replication, otherwise known as replication slippage (Tautz and Schlötterer 1994).

Mutation is a credible source of low-level variation, and it is not the objective of this study to investigate the potential sources of mutation and newly created variation, such as intragenic recombination, unequal crossing over, DNA methylation, and paramutation. However, previously calculated mutation rates can give some idea of how often mutant alleles may be expected in this study. Mutation rates for humans, mice, and fruit flies are estimated to range from 5.2 × 10−7 to 1 × 10−2 per generation (Kruglyak et al. 1998; Schlötterer 2000; Vigouroux et al. 2002). The calculated mutation rates in plants ranged from 2.4 × 10−2 per allele per generation in wheat (Thuillet et al. 2002) to 1 × 10−2 per allele per generation in chickpea (Udupa and Baum 2001).

Diwan and Cregan (1997) studied the use of SSR loci in soybean and calculated a mutation rate for soybean based on a previously reported mutation rate in humans and on their
own knowledge of cultivar development. The range of mutation rates that they calculated, $2.3 \times 10^{-4}$ to $5.6 \times 10^{-3}$ per allele per generation, falls within the range of mutation rates that have been calculated for plants. This is the expected mutation rate for a cultivar that is derived from a single $F_2$ or $F_3$ plant, which is the progeny of 2 cultivars that were themselves derived from single $F_4$ or $F_5$ plants. This does not include the additional generations of selfing it takes to arrive at a foundation seed-level cultivar. Because this mutation rate is for a single locus, the mutation rate we would expect for the 144 loci in this study, using the more conservative of the mutation estimates and not accounting for additional generations of selfing, is $144 \times 2.3 \times 10^{-4}$, or 0.033. The least number of new alleles we would predict for each pedigree is 4.75 new alleles per pedigree (per 144 loci). Thus, at least 4 of the new alleles that are seen within each pedigree may be feasibly attributed to mutation. Mutation leading to the creation of de novo variation therefore appears to be a feasible explanation for the variant alleles that could not be found in the original gene pool where selection was initiated.

Our results corroborate those of Rasmusson and Phillips (1997) who reported that selection gain in elite gene pools occurs due to variation present in the original gene pool as well as due to newly created variation. Moreover, molecular biologists are finding that the genome undergoes constant remodeling and restructuring by utilizing an array of different mechanisms like intragenic recombination, unequal crossing over, DNA methylation, excision or insertion of transposable elements, gene duplication, genetic restoration events, etc. (Walbot and Cullis 1985; Cullis 1990; Rasmusson and Phillips 1997; Brunner et al. 2005; Lolle et al. 2005; Springer and Stupar 2007). Our data are in agreement with those of Haun et al. (2011) who explored the genetic basis of intracultivar variation in soybean by investigating the nucleotide, structural, and gene content variation of different individuals from the reference cultivar Williams 82. They reported high rates of intracultivar structural variation within Williams 82 and noted that the impact of intracultivar genetic heterogeneity may be significant. Our study provides crucial data and additional evidence that inbred lines are not permanent genetic stocks but material that contains genetic variation that can be further utilized in a breeding program.

Conclusion

The results of this study provide direct evidence that the phenotypic variation for seed protein content and seed weight found in the 19 single-plant lines developed from honeycomb selection at ultra-low plant density has a genotypic component identified by the SSR band variation. Therefore, these lines are unique phenotypically as described by Fasoula and Boerma (2005, 2007), and they also possess unique genetic fingerprints. Consequently, single-plant selection at low plant densities was effective in identifying heritable variation at the single-plant level within a cultivar as shown by the SSR marker analysis presented in this study. Some of the single-plant selected lines discussed in this study have been registered and released as germplasm lines and are available for additional research (Fasoula et al. 2007a, 2007b, 2007c). Interestingly, the SSR variation was highest in lines showing the greatest phenotypic divergence in seed protein content and seed weight.

Most of the SSR band variation discovered in the single-plant selected lines from within the soybean cultivars Benning, Haskell, and Cook was also found in the Foundation seed source of the cultivars where selection originated, indicating that it was the result of residual heterozygosity in the initial plant selected to become the cultivar. In addition, we showed that some variant SSR bands discovered in the selected lines could not be detected in the cultivar seed source. More specifically, 18% of variant SSR bands found in the single-plant selected lines derived from within Benning could not be detected in the Benning seed source. Similarly, 7% of variant SSR bands found in the Haskell selections could not be detected in the Haskell seed source. And 18% of the variant bands found in the Cook selections could not be detected in the Cook seed source. Whereas some of this variation may have resulted from undetected residual heterozygosity, a few of the variant bands are likely the result of mutation or some kind of mechanism generating de novo variation. These results are very important because in addition to phenotypic evidence, they also provide molecular evidence that inbred lines and other homogeneous gene pools are not permanent genetic stocks but material that contains genetic variation which can be further exploited in a conventional or molecular breeding project.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/. Data were also deposited in the Dryad Digital Repository (Yates et al. 2012).

Funding

Memorandum of Agreement from the Georgia Agricultural Commodity Commission for Soybeans; Georgia Agricultural Experiment Stations Hatch Project No. GEO884; NSF Graduate Research Assistantship.

References

Abe J, Xu D, Suzuki Y, Kanazawa A, Shimamoto Y. 2003. Soybean germplasm pools in Asia revealed by nuclear SSRs. Theor Appl Genet. 106:445–453.
Akkaya MS, Bhagwat AA, Cregan PB. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics. 132:1131–1139.
Allard RW. 1960. Principles of plant breeding. New York: J. Wiley & Sons.
Bantte K, Prasanna BM. 2003. Simple sequence repeat polymorphism in Quality Protein Maize (QPM) lines. Euphytica. 129:337–344.
Reg. 1:160–161.

within the cultivar "Haskell" differing in seed and agronomic traits. J Plant

Boerma HR, Hussey RS, Phillips DV, Wood ED, Fennerty SL. 1992. Registration of HK soybean. Crop Sci. 32:97.

Boerma HR, Hussey RS, Phillips DV, Wood ED, Fennerty SL. 1994. Registration of Haskell soybean. Crop Sci. 34:541.

Boerma HR, Hussey RS, Phillips DV, Wood ED, Rowan GB, Fennerty SL. 1997. Registration of Benning soybean. Crop Sci. 37:1982.

Brunner S, Tingeys R, Rafalski A, Fengler K, Morgante M. 2005. Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell. 17:343–360.

Byth DE, Weber CR. 1968. Effects of genetic heterogeneity within two soybean populations. I. Variability within environments and stability across environments. Crop Sci. 8:44–47.

Cho YG, McCouch SR, Kuiper M, Kang M-R, Pot J, Groenew JTM, Eun MY. 1998. Integrated map of AFLP, SSLP and RFLP markers using a recombinant inbred population of rice (Oryza sativa L). Theor Appl Genet. 97:370–380.

Christakis PA, Fasoulas AC. 2002. The effects of the genotype by environment interaction on the fixation of heterosis in tomato. J Agric Sci. 138:1–5.

Cullis CA. 1990. DNA rearrangements in response to environmental stress. Adv Genet. 28:73–97.

Deka R, Guangun S, Smelser D, Zhong Y, Kimmel M, Chakraborty R. 1999. Rate and directionality of mutations and effects of allele size constraints at anonymous, gene-associated, and disease-causing trinucleotide loci. Mol Biol Evol. 16:1166–1177.

Dowan N, Cregan PB. 1997. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor Appl Genet. 95:723–733.

Dudley JW, Lambett RJ. 1992. Ninety generation of selection for oil and protein in maize. Maydisca. 37:81–87.

East EM. 1936. Genetic aspects of certain problems of evolution. Am Nat. 70:143–158.

Edwards K, Johnstone C, Thompson C. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. 19:1349.

Ellegren H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet. 16:551–558.

Fasoula DA. 1990. Correlations between auto-, allo-, and nil-competition and their implications in plant breeding. Euphytica. 50:57–62.

Fasoula DA, Fasoula VA. 1997. Competitive ability and plant breeding. Plant Breed Rev. 14:89–138.

Fasoula VA. 2008. Two novel whole-plant field phenotyping equations to maximize selection efficiency. In: Prohens J, Badenes ML, editors. Proceedings of the 18th Eucarpia General Congress, Modern Variety Breeding for Present and Future Needs; 2008 Sept 9–12; Valencia, Spain: Universidad Politecnica de Valencia. p. 361–365.

Fasoula VA, Boerma HR, Yates JI, Walker DR, Fennerty SL, Rowan GB, Wood ED. 2007a. Registration of six soybean germplasm lines selected within the cultivar “Haskell” differing in seed and agronomic traits. J Plant Reg. 1:158–159.

Fasoula VA, Boerma HR, Yates JI, Walker DR, Fennerty SL, Rowan GB, Wood ED. 2007c. Registration of five soybean germplasm lines selected within the cultivar “Benning” differing in seed and agronomic traits. J Plant Reg. 1:156–157.

Fasoula VA, Fasoula DA. 2000. Honeycomb breeding principles and applications. Plant Breed Rev. 18:177–250.

Fasoula VA. 2002. Principles underlying genetic improvement for high and stable crop yield potential. Field Crops Res. 75:191–209.

Fasoulas AC. 2000. Building up resistance to Verticillium wilt in cotton through honeycomb breeding. In: Gillham FM, editor. New Frontiers in Cotton Research. Proceedings of Second World Cotton Research Conference; 1998 Sept. 6–12; Athens, Greece. Thessaloniki, Greece: P. Petrakis. p. 120–124.

Fasoula AC, Fasoula VA. 1995. Honeycomb selection designs. Plant Breed Rev. 13:87–139.

Haan WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, Jeddlehof A, Jia G, Springer NM, Vince CP, et al. 2011. The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. Plant Physiol. 155:645–655.

Hawkes JG. 1977. The importance of wild germplasm in plant breeding. Euphytica. 26:615–621.

Jin L, Mucablas C, Hallmayer J, Kimura A, Mignot E. 1996. Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. Proc Natl Acad Sci U S A. 93:15285–15288.

Keim P, Olson TC, Shoemaker RC. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet Newsletter. 15:150–152.

Kruglyak S, Durret RT, Schug MD, Aquadro CF. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc Natl Acad Sci U S A. 95:10774–10778.

Lolle S, Victor JL, Young JM, Pruitt RE. 2005. Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis. Nature. 434:505–509.

Mutsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebely J. 2002. Microsatellites in Zea—variability, patterns of mutations, and use for evolutionary studies. Theor Appl Genet. 104:436–450.

Nandakumar N, Singh AK, Sharma RK, Mohapatra T, Prabhu KV, Zaman FU. 2004. Molecular fingerprinting of hybrids and assessment of genetic purity of hybrid seeds in rice using microsatellite markers. Euphytica. 136:257–264.

Oluafowote JO, Xu Y, Chen X, Park WD, Reachell HM, Dilday RH, Goto M, McCouch SR. 1997. Comparative evaluation of within-cultivar variation of rice (Oryza sativa L) using microsatellite and RFLP markers. Genome. 40:370–378.

Périn C, Hagen L, De Conto V, Katzir N, Danin-Poleg Y, Portnoy V, Brudarakco-Amos S, Chadoeuf J, Dagimont C, Pirat M. 2002. A reference map of Caenorhabditis elegans based on two recombinant inbred line populations. Theor Appl Genet. 104:1017–1034.

Pröll RHG, Mendes-Junior CT, Arantes NE, Contel EPB. 2002. Characterization of Brazilian soybean cultivars using microsatellite markers. Genet Mol Biol. 25:185–193.

Rasmusson DC, Phillips RL. 1997. Plant breeding progress and genetic diversity from de novo variation and elevated epistasis. Crop Sci. 37:303–310.

Russell WA, Sprague GF, Penny LH. 1963. Mutations affecting quantitative characters in long-time inbred lines of maize. Crop Sci. 3:175–178.

Santhiárez-Koréf MF, Gangeswaran R, Hancock JM. 2001. A relationship between lengths of microsatellites and nearby substitution rates in mammalian genomes. Mol Biol Evol. 18:2119–2123.
Schlotterer C. 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma. 109:365–371.

Senior ML, Murphy JP, Goodman MM, Stuber CW. 1998. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Sci. 38:1088–1098.

Singh RJ, Hymowitz T. 1999. Soybean genetic resources and crop improvement. Genome. 42:605–616.

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

Springer NM, Stupar RM. 2007. Allelic variation and heterosis in maize: how do two halves make more than a whole? Genome Res. 17:264–275.

Tautz D, Schlotterer C. 1994. Simple sequences. Cur Opin Genet Dev. 4:832–837.

Thuillet A-C, Bru D, David J, Roumet P, Santoni S, Sourdille P, Batillon T. 2002. Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp durum desf. Mol Biol Evol. 19:122–125.

Tokatlidis IS, Tsialtas JT, Xynias IN, Tamoutsidis E, Irakli M. 2004. Variation within a bread wheat cultivar for grain yield, protein content, carbon isotope discrimination and ash content. Field Crops Res. 86:33–42.

Tokatlidis IS, Xynias IN, Tsialtas JT, Papadopoulos II. 2005. Single-plant selection at ultra-low density to improve stability of a bread wheat cultivar. Crop Sci. 46:90–97.

Udupa SM, Baum M. 2001. High mutation rate and mutational bias at (TAA)n microsatellite loci in chickpea (*Cicer aritinum* L.). Mol Genet Genomics. 265:1097–1103.

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, Doebley J. 2002. Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol. 19:1251–1260.

Wallbot V, Cillo CA. 1985. Rapid genomic change in higher plants. Ann Rev Plant Physiol. 36:367–396.

Whittaker JC, Harbourn RM, Boxall N, Mackay I, Dawson G, Sibly RM. 2003. Likelihood-based estimation of microsatellite mutation rates. Genetics. 164:781–787.

Zhang YX, Gentzbittel L, Vear F, Nicolas P. 1995. Assessment of inter- and intra-inbred line variability in sunflower (*Helianthus annuus*) by RFLPs. Genome. 38:1040–1048.

Yates JL, Boerma HR, Fasoula VA. 2012. Data from: SSR-marker analysis of the intra-cultivar phenotypic variation discovered within three soybean cultivars. Dryad Digital Repository. doi:10.5061/dryad.7dn526gq.

Received September 2, 2011; Revised January 24, 2012; Accepted February 13, 2012

Corresponding Editor: Reid Palmer