A Microsatellite Map Developed from Late Maturity Germplasm 'Essex' by 'Forrest' Detects Four QTL for Soybean Seed Yield Expected from Early Maturing Germplasm.

Kassem My Abdelmajid 1*, Khalid Meksem 2,5, Andrew J. Wood3,5, David A. Lightfoot4,5

1* Department of Natural Sciences, Fayetteville State University, Fayetteville, NC 28301-4252, USA.
2 Plants and Microbes Genomics and Genetics Lab, Department of Plant Soil, and Agricultural Systems, Southern Illinois University, Carbondale, IL 62901-4415, USA.
3 Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901-6509, USA.
4 Biotechnology and Genomics Core-Facility, Department of Plant Soil, and Agricultural Systems, Southern Illinois University, Carbondale, IL 62901-4415, USA.
5 Center for Excellence in Soybean Research, Teaching and Outreach, Southern Illinois University, Carbondale, IL 62901-4415, USA.

Summary

Early maturing (maturity groups 0-4), indeterminate or semi-determinate, soybean cultivars are genetically distinct from late maturing (maturity groups 5-8), determinate germplasm in the US and Canada. Marker assisted introgression of yield QTL between the germplasm pools has failed for reasons that are unclear. Previously, we had identified 3 QTL for yield in a population derived from determinate, late maturing 'Essex' by 'Forrest' that were not found in several early maturing populations. No QTL in intervals expected from analyses of early germplasm were detected. We aimed to identify additional seed yield QTL using a more dense map of the intervals associated with yield QTL made with replicated scores from co-dominant microsatellite markers. The population seed yield data used (from four environments and two years) was normally distributed. We compared against seed yield, two hundred thirty one (231) polymorphic microsatellite markers using ANOVA, MAPMAKER/QTL, and WINQTL CART. The three QTL previously reported and four new regions associated with seed yield were identified (P < 0.0075). The first region was detected on LG C2 by the markers SATT363 (P = 0.0027, R² = 10.2%) and SATT277 (P = 0.0066, R² = 8.2%), the second region by SATT009 (P = 0.0075, R² = 8%) on LG N, and the third region by SATT285 (P = 0.0067, R² = 10%) on LG J. In a separate data analysis using QTL CART, a fourth QTL for seed yield was identified by the marker SAT_162 on LG A2. Forrest provided the beneficial allele at all four regions. Two of the four additional loci were in regions associated with seed yield, plant morphology and other major yield determinants in early maturity germplasm, two in later maturity germplasm. Correspondence of the intervals of the four QTL with independent analyses suggested they were not type II errors and that low P values resulted from a relatively sparse map. Yield determinates in common intervals will complicate marker assisted introgression of yield QTL between germplasm pools.

Keywords : soybean, seed, yield, QTL, genomics, microsatellite, marker.

Introduction

Seed yield increases in soybean have been partly due to genetic advances resulting from inter-crossing existing varieties (Specht et al., 1999). Plant breeders identify cultivars with increased yield potential during recurrent selection. Yield is a multigenic trait and therefore inheritance is complex (Mansur et al., 1996). The restricted genetic base of soybean, incomplete linkage disequilibrium and small number of haplotypes (Gizlice et al., 1993; 1994; Zhu et al., 2003) does not prevent wide segregation for yield during intercrossing related cultivars (Yuan et al., 2002). Therefore, it was difficult to predict yield potential of intercross-derived lines without field tests. Components of soybean seed yield, such as maturity, growth habit, harvest index, seed weight, leaf area duration, nitrogen fixation, lodging and water deficit tolerance are identifiable (Mansur et al., 1993, 1996; Mian et al., 1996; Specht et al., 1999; Orf et al., 1999a,b; Concibido 2002; Wang et al., 2004). QTL in soybean that could affect yield in addition to the above listed include those for disease resistance (Yue et al., 2001; Arahana et al., 2001; Njiti et al., 2002), yield conditioned by disease resistance (Njiti et al., 1998; Hnetkovsky et al., 1996), and mineral nutrition (Kassem et al., 2004a). DNA markers for these loci can aid in the selection of yield from non-adapted germplasm (Specht et al., 1999; Mansur et al., 1996; Orf et al., 1999a,b).

In soybean recombinant inbred lines, or backcrossed lines where morphological segregation is restricted or absent, the physiological and biochemical basis of yield improvement may be studied (Njiti et al., 1997; Concibido et al., 2002; Wang et al., 2004). Some correspondence between seed yield QTL and loci that determine nitrogen fixation (Njiti et al., 1997), trigonelline content (Cho et al., 2002) and seed isoflavone content (Kassem et al.,...
Materials and Methods

Plant Material:

The Essex x Forrest RIL population (Lightfoot et al., 2005) shows little segregation for maturity, morphology or growth habit (Hnetkovsky et al., 1996). The distribution of the mean seed yield among the ExF RILs across four tested environments was unimodal and three seed yield QTL were previously inferred (Yuan et al., 2002). The ExF RIL genetic map currently contains 368 linked markers, but 231 are high quality microsatellite markers that have been scored repeatedly (Kassem et al., 2004b; Lightfoot et al., 2005). The average distance between markers is 17 cM, and there are about 23 markers per linkage group and 79 markers unlinked. The polymorphic markers include 90 RAPD markers, 27 RFLP markers, 20 AFLP markers, and 231 microsatellite markers (Lightfoot et al., 2005). Here, (1) additional seed yield QTL are detected and three previously detected QTL further delimited using an expanded set of independently replicated microsatellite marker scores, and (2) a summary of all seed yield QTLs identified in the ExF RIL population from 1996-2003 across four environments is presented.

DNA Isolation, PCR Amplifications, and Genetic Map Construction:

DNA purification, markers' amplifications, and genetic map construction were described earlier (Iqbal et al., 2001; Meksem et al., 2001a,b; Yuan et al., 2002; Kassem et al., 2004a,b, 2006a; Lightfoot et al., 2005).

QTL Mapping:

Markers and traits data were analyzed using MAPMAKER/QTL VER. 1.1 (Lander et al., 1987) and a simple interval mapping (SIM) was performed because we judged the marker set too sparse for an effective composite interval mapping (CIM). However, CIM using WINQTL CART. VER. 2.5 (Wang et al., 2005) was also performed and the model 6 was adopted. The control marker number and window size were 5 and 10 cM, respectively. A walk speed of 2 cM and the forward regression method were selected. LOD scores = 2.0 indicated the existence of QTL for seed yield reported in this study (Table 1). Experiment-wise threshold level to declare linkage was calculated from 1,000 permutations of each genotype marker against the phenotype in the population. Linkage was reported as significant if the two statistics for a marker were greater than the critical value at P = 0.05.

Results

Four additional QTLs for seed yield were found on different L.Gs. in this EXF population. The first QTL located on LG. C2 was identified by the marker SATT363 (P=0.0027, R2 = 10.2%), and derived the beneficial allele from Forrest. The linked marker SATT277 (4.7 cM; P = 0.0077, R2 = 8.2%) was also seed yield. The QTL spanned approximately 4.5 cM, had a peak LOD score of 2.7 (Figures 1a and 2), and explained approximately 18.4% of

Table 1: DNA markers associated with the QTLs that underlie seed yield in the Essex x Forrest RIL population in the mean of four Illinois environments Desoto 1997, Ridgway 1996, Carbondale 1996 and Desoto 1996.

<table>
<thead>
<tr>
<th>Marker</th>
<th>LG.</th>
<th>R2</th>
<th>P&gt;F</th>
<th>LOD</th>
<th>QTL Var.</th>
<th>Yield Means (Kg. ha⁻¹)</th>
<th>±SEMs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>New QTLs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BARC-SAT_162</td>
<td>A2</td>
<td>12.1</td>
<td>-</td>
<td>2.9</td>
<td>12.1</td>
<td>3.28 ± 0.02</td>
<td>3.39  ± 0.02</td>
</tr>
<tr>
<td>BARC-SATT277</td>
<td>C2</td>
<td>8.2</td>
<td>0.0066</td>
<td>2.0</td>
<td>13.8</td>
<td>3.28 ± 0.02</td>
<td>3.39  ± 0.02</td>
</tr>
<tr>
<td>BARC-SATT363</td>
<td>C2</td>
<td>10.2</td>
<td>0.0027</td>
<td>2.7</td>
<td>18.4</td>
<td>3.28 ± 0.02</td>
<td>3.39  ± 0.02</td>
</tr>
<tr>
<td>BARC-SATT285</td>
<td>I</td>
<td>10.0</td>
<td>0.0066</td>
<td>2.0</td>
<td>4.02</td>
<td>3.30 ± 0.02</td>
<td>3.40  ± 0.02</td>
</tr>
<tr>
<td>BARC-SATT009</td>
<td>N</td>
<td>8.0</td>
<td>0.0075</td>
<td>2.0</td>
<td>13.1</td>
<td>3.29 ± 0.02</td>
<td>3.38  ± 0.02</td>
</tr>
<tr>
<td><strong>Confirmed QTLs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BARC-SATT337</td>
<td>K</td>
<td>10.1</td>
<td>0.0442</td>
<td>2.1</td>
<td>10.7</td>
<td>3.40 ± 0.03</td>
<td>3.29  ± 0.03</td>
</tr>
<tr>
<td>BARC-SATT167</td>
<td>K</td>
<td>11.6</td>
<td>0.0023</td>
<td>1.9</td>
<td>9.6</td>
<td>3.39 ± 0.03</td>
<td>3.27  ± 0.03</td>
</tr>
<tr>
<td>BARC-SATT294</td>
<td>C1</td>
<td>9.5</td>
<td>0.0056</td>
<td>2.0</td>
<td>12.1</td>
<td>3.29 ± 0.02</td>
<td>3.41  ± 0.03</td>
</tr>
<tr>
<td>BARC-SATT440</td>
<td>I</td>
<td>10.2</td>
<td>0.0071</td>
<td>1.6</td>
<td>10.3</td>
<td>3.28 ± 0.03</td>
<td>3.39  ± 0.03</td>
</tr>
</tbody>
</table>

2004b) has been noted. However, there are more than 300 known loci conditioning more than 80 traits in soybean and QTL appear to cluster. About 85% of the known QTL have not yet been confirmed in a second study. Moreover, attempts to introgress seed yield QTL from early maturing 'Archer', 'Minsoy and Noir (Orf et al., 1999a; 1999b) to late maturing cultivars has failed (J. Specht and C Sneller, pers. comm.). Errors in assigning QTL or overestimating QTL effect may be responsible. In addition many studies contain errors in maps due to the use of RAPD, AFLP and even RFLP markers without sufficient replication to be error free and to detect all residual heterogeneity. Alternately, yield determinates may be in similar locations in both early and late germplasm so that introgression is balanced by both positive and negative effects.
variation in seed yield (Tables 1 and 2). The locus was associated with seed yield in Desoto 1997 (P = 0.0011) and Ridgway 1996 (P = 0.013) but not Carbondale 1996 or Desoto 1996. The second QTL located on LG J was identified by the marker SATT285 (P = 0.0067, R2 = 10%), and also derived its beneficial allele from Forrest. This locus was associated with seed yield in Desoto 1997 (P = 0.0011) and Ridgway 1996 (P = 0.013) but not Carbondale 1996 or Desoto 1996. The second QTL located on LG J was identified by the marker SATT009 (P = 0.0067, R2 = 12.1%), and also derived its beneficial allele from Forrest. This locus was associated with seed yield in Desoto 1996 (P = 0.002) but not Carbondale 1996, Ridgway 1996 or Desoto 1997. The third QTL located on LG N was identified by the marker SATT079 (P = 0.0075, R2 = 8%). This locus was associated with seed yield in Desoto 1997 (P = 0.0089) and Ridgway 1996 (P = 0.021) but not Carbondale 1996 or Desoto 1996 and also derived its beneficial allele from Forrest. The fourth QTL located on LG A2 was identified by the marker SAT_162 (LOD = 2.9, R2 = 12.1%), spanned <2.0 cM, had a peak LOD score of 2.9, and explained approximately 12.1% of total seed yield variation (Figures 1a and 2; Tables 1 and 2).

We further delimited (with new markers) the three QTLs associated with seed yield previously reported by Yuan et al. (2002) using the same EXF RIL population (Figures 1b and 2; Tables 1 and 2). The three QTLs were located on LGs. K (SATT337; P = 0.0042, R2 = 10% and SATT167; P = 0.0023, R2 = 12%), C1 (SATT294; P = 0.0056, R2 = 10%), and I (SATT440; P = 0.0071, R2 = 10%). The two QTLs on LGs. C1 and I derived the beneficial allele from Essex (Figures 1b and 2; Tables 1 and 2). Previously identified QTL Interval sizes were delimited and reduced from 10-12 cM to 3-5 cM regions. Summing R2 terms showed the seven regions explained less than 56% of the total variation in seed yield in this population. None of the additional loci reported here were associated with maturity date, plant height or lodging. However, the seed yield QTL on LG A2 was mapped 17.9 cM from a QTL for internodes length (qIN1) and the QTL on LG C2 underlie both seed yield and internodes length (qIN4a; Jacobson et al., 2006).

We further delimited (with new markers) the three QTLs associated with seed yield previously reported by Yuan et al. (2002) using the same EXF RIL population (Figures 1b and 2; Tables 1 and 2). The three QTLs were located on LGs. K (SATT337; P = 0.0042, R2 = 10% and SATT167; P = 0.0023, R2 = 12%), C1 (SATT294; P = 0.0056, R2 = 10%), and I (SATT440; P = 0.0071, R2 = 10%). The two QTLs on LGs. C1 and I derived the beneficial allele from Essex (Figures 1b and 2; Tables 1 and 2). Previously identified QTL Interval sizes were delimited and reduced from 10-12 cM to 3-5 cM regions. Summing R2 terms showed the seven regions explained less than 56% of the total variation in seed yield in this population. None of the additional loci reported here were associated with maturity date, plant height or lodging. However, the seed yield QTL on LG A2 was mapped 17.9 cM from a QTL for internodes length (qIN1) and the QTL on LG C2 underlie both seed yield and internodes length (qIN4a; Jacobson et al., 2006).

**Discussion**

Four additional QTL underlying the mean seed yield were detected and three QTL for seed yield previously reported (Yuan et al., 2002) were further delimited. In a recent study, two new seed yield QTLs were identified by CIM (Table 2; Figure 2; Kassem et al., 2006). The total variation of all nine QTL was probably close to explaining all the heritable variation (47%) but cannot be measured jointly in a population of only 100 lines. Therefore, more QTL for seed yield may not be detected in this population without considering additional locations. However, additional DNA markers in the QTL intervals should increase the P and R2 values (Meksem et al., 2001).

Comparing the new QTL with reports of intervals determining yield variation in other populations showed significant correspondences. A new analysis of Hartwig by Flyer (HXF) RIL population (semi-determinate, maturity group 4-5) showed that the new QTL on LG N identified by the marker SATT009 was associated with seed yield (P<0.04) in one environment (Kazi, 2005). However, the QTL on LG C2 identified by the markers SATT363 and SATT277 was not associated with yield in the HxF population in 1998 and 1999. The two QTLs on LGs. A2 (SAT_162) and J (SATT285) were not tested yet. Interestingly, the seed yield QTL on LG C2 mapped 40.7 cM from another new seed yield QTL identified by the marker SAT_079 (LOD = 4.0) in the ExF population using CIM (Kassem et al., 2006a).

SAT_162 identified a QTL for seed yield on LG A2. Interestingly, QTL for leaf area (Mian et al., 1998a) were detected within 3 cM of this interval in later maturing determinate germplasm (also resistance to SCN; Skorupska, 1995). On LG C2, QTL were detected within the 10 cM interval indicated in both earlier and later maturing germplasm. QTL conditioned disease resistance, specific leaf weight, leaf area, reproductive period, oil, seed pod maturity, plant height, maturity date, seed filling period, seed

**Table 2: A summary of the nine seed yield QTLs identified to date in the ExF RIL population in four Illinois environments: Desoto 1997, Ridgway 1996, Carbondale 1996, 2003, and Desoto 1996.**

<table>
<thead>
<tr>
<th>Marker</th>
<th>LG.</th>
<th>QTL#</th>
<th>LOD</th>
<th>Position (cM)</th>
<th>New QTL/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BARC-SAT_162</td>
<td>A2</td>
<td>1</td>
<td>2.9</td>
<td>9.4</td>
<td>New</td>
</tr>
<tr>
<td>BARC-SATT294</td>
<td>C1</td>
<td>2</td>
<td>2.0</td>
<td>12.6</td>
<td>Yuan et al., 2002</td>
</tr>
<tr>
<td>BARC-SATT079</td>
<td>C2</td>
<td>3</td>
<td>4.0</td>
<td>66.0</td>
<td>Kassem et al., 2006</td>
</tr>
<tr>
<td>BARC-SATT277</td>
<td>C2</td>
<td>4</td>
<td>2.0</td>
<td>106.7</td>
<td>New</td>
</tr>
<tr>
<td>BARC-SATT363</td>
<td>C2</td>
<td>2.7</td>
<td>111.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BARC-SATT398</td>
<td>G</td>
<td>5</td>
<td>2.9</td>
<td>2.0</td>
<td>Kassem et al., 2006</td>
</tr>
<tr>
<td>BARC-SATT440</td>
<td>I</td>
<td>6</td>
<td>1.6</td>
<td>0.0</td>
<td>Yuan et al., 2002</td>
</tr>
<tr>
<td>BARC-SATT285</td>
<td>J</td>
<td>7</td>
<td>2.0</td>
<td>4.0</td>
<td>New</td>
</tr>
<tr>
<td>BARC-SATT167</td>
<td>K</td>
<td>8</td>
<td>1.9</td>
<td>32.0</td>
<td>Yuan et al., 2002</td>
</tr>
<tr>
<td>BARC-SATT337</td>
<td>K</td>
<td>2.1</td>
<td>38.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BARC-SATT009</td>
<td>N</td>
<td>9</td>
<td>2.0</td>
<td>14.8</td>
<td>New</td>
</tr>
</tbody>
</table>
Figure 1: Locations of DNA markers and the QTL conditioning seed yield. Linkage groups A2, C1, C2, I, J, K and N are shown as black bars. The QTL were assigned to linkage groups on the soybean genetic map (Cregan et al., 1999, 2003). Panel (a) The additional QTL on linkage group A2, C2, J and N. Panel (b) The three QTL reported previously on linkage groups C1, I, and K (Yuan et al., 2002). These three QTLs were further delimited with new markers. The new markers are underlined and followed by a star (*). Genetic distances were from the recombinant inbred line function of MAPMAKER/EXP 3.0. The distance representing 10 cM is shown by the black rectangle. The estimated position of the interval containing the QTL (LOD > 2.0) is shown as a short bar based on interval mapping using MAPMAKER/QTL 1.1. The QTL LOD score is the peak LOD score of the interval showing association with seed yield. Linkage groups continue beyond the regions shown.
weight, seed yield, and yield (Mansur et al., 1993; 1996; Mahalingham and Skorupska, 1995; Mian et al., 1996, 1998a,b; Orf et al., 1999a,b; Rector et al., 1999; Specht et al., 2001; Yue et al., 2001; Wang et al., 2004). On LG J, QTL for plant height, leaf width, leaf area, pod dehiscence, lodging, pod maturity, reproductive maturity and yield were mapped in earlier maturing populations within the same interval reported here (Lark et al., 1995; Mansur et al., 1996; Orf et al., 1999a; Specht et al., 2001). In a recent study, we found that the seed yield QTL on LG A2 mapped only 17.9 cM from a QTL for internodes length (qIN1) and the QTL on LG C2 underlie both seed yield and internodes length (qIN4a) (Jacobson et al., 2006). On LG N, QTL for seed-coat hardness, plant height, leaf length, disease resistance, seed yield, and yield were mapped in several earlier maturing populations close to the seed yield QTL reported here (Keim et al., 1990; Lark et al., 1995; Mansur et al., 1996; Orf et al., 1999a; Arahana et al., 1999).
A plant height QTL was identified at the SATT440 locus in an earlier maturing population (Yuan et al., 2002). Correspondence between QTL locations in diverse germplasm may indicate clustering of QTL (in gene rich regions). Alternately the QTL detected may be allelic and reflect the small number of haplotypes in soybean (Zhu et al., 2003). Distinguishing these hypotheses will require fine maps, physical maps, genome sequence and candidate gene mutation and/or complementation (Lightfoot et al., 2003). Some of these loci lie in regions with many marker anchors on the physical map (Wu et al., 2003). On linkage group C2 in the 8-10 cM (4-5 Mbp) region encompassed by the QTL there are 17 contigs and 45 BACs with 89 BAC end sequences that detect 12 predicted genes (P=e-15) suggesting this region is gene rich (http://bioinformatics.siu.edu). NIL populations have been identified for each interval to assist with fine map

Figure 2 part 2
The new Essex by Forrest genetic map (Kassem et al., 2006) and the locations of the nine seed yield QTLs identified to date.
development and positional cloning in future.

Mean yield was enhanced by 0.10-0.12 Mg/ha per QTL in this population indicating the potential value of the QTLs to provide a 0.6 Mg/ha difference in yield potential between the best and worst cultivars. Data analysis from SIUC soybean breeding program showed that alleles are still segregating at the five markers linked to the seed yield QTLs on L.Gs. C1, C2, I, K, and N identified by SATT294, (SATT363 and SATT277), SATT440, (SATT337 and SATT167), and SATT009 (unpublished data). However, the presence of allelic disequilibrium is not yet clear (Meksem et al., 2001; Zhu et al., 2003). The QTLs discovered here are useful in genome recovery during backcrossing; however, their usefulness is restricted by the observation that each of the seven loci was associated with seed yield in one to three locations. Six of the seven seed yield QTLs reported here derived their beneficial alleles from Forrest which is the low yielding but disease resistant parent. Many physiological processes that can influence yield may be altered by the introduction of a disease resistance from Forrest.

Alternately, the historic observations of poor combining ability of Forrest derived germplasm may be related to dominance relations or interactions among seed yield QTL from other genotypes. The poor combining ability of Forrest may have caused Peking derived SCN resistance to be associated with linkage drag on yield. In turn this has contributed to the common use of 'PI88788' and 'Bedford' as the sources of SCN resistance combined with yield (Mahalingham and Skorupska, 1995). Determination of the basis of the combining ability is important as many new SCN resistance loci are being introgressed from PI437654 into soybean germplasm such as 'Hartwig' and 'Ina' using Forrest derived backcross lines (Prabhu et al., 1999).

Our results showed that QTLs for seed yield can be detected in soybean crosses using microsatellite markers and some QTLs were effective in multi-environments. Furthermore, these QTLs can be used in MAS programs to generate high yielding varieties if the allelic state of the recipient germplasm is known.

Acknowledgments

The authors acknowledge Dr. Paul Gibson and Dr. Oval Myers, Jr., Dept. of Plant, Soil, and Agricultural Systems (PSAS), Southern Illinois University, for developing the recombinant inbred line populations used in this study. Special thanks to Dr. Victor Njiti, Michael E Schmidt and J.H. Klein, III and the rest of the soybean field crew in the Dept. of PSAS for participating in planning and managing of the yield plots. This research was supported by Illinois Soybean Program Operating Board (ISPOB) grants 95-222-24-3, 98-222-25-2, and NSF 9872635.

References


