

Identification of Loci Underlying Seed Yield in Recombinant Inbred and Near Isogenic Soybean Lines Derived from Flyer by Hartwig

Samreen Kazi^{1,2}, Jeffry L. Shultz^{1,3}, Ahmed Jawaad Afzal^{1,2,4}, Yi-Chen Lee^{1,2}, and David A. Lightfoot^{1,2*}

¹ Plant Biotechnology and Genomics Core-Facility, Department of Plant, Soil, and Agricultural Systems, Illinois Soybean Center, Southern Illinois University, Carbondale, IL 6290, USA; ² Molecular Biology and Medical Biochemistry Program, Southern Illinois University, Carbondale, IL 62901, USA; ³ Present Address: School of Biological Sciences, Louisiana Tech University, Ruston LA, USA; ⁴ Present Address: LUMS, Lahore, Pakistan.

Received: January 6, 2017 / Accepted: February 19, 2017

Abstract

Two major determinants of soybean [*Glycine max* (L.) Merr.] seed yield were resistances to the soybean cyst nematode (SCN) and sudden death syndrome (SDS). Two loci were identified *rhg1/Rfs2* and *rhg3/rfs5* (for resistance to SCN, *Heterodera glycines* (L.) HG Type 1.3- (race 14), HG Type 0 (race 3) and SDS caused by *Fusarium virguliforme* (Roy & Rupe)). The aim of this study was to identify quantitative trait loci (QTL) underlying seed yield. Used were 142 microsatellite markers and the recombinant inbred line population (RIL) 'Flyer' × 'Hartwig' (F × H; n=92). Flyer (F) was high yielding but SCN and SDS susceptible. Hartwig (H) was lower yielding but resistant to all SCN Hg Types and SDS. Four regions on 3 chromosomes were associated with seed yield. The first region on chromosome 9 (Gm9, qYld09.1), identified by the microsatellite marker Satt539-Satt242 (LOD 2.9, 13% variation) derived the beneficial allele from Hartwig (F allele 2.76 ± 0.06 Mg/Ha; H allele 2.98 ± 0.03 Mg/ha). The second region on Gm9 (qYld09.2) between Satt337 and Satt326 spanned 1.4 cM (LOD of 5.31, 20.2% variation) and the beneficial allele derived from Flyer (0.22 Mg/ha F allele 2.98 ± 0.03, H allele 2.77 ± 0.04 Mg/Ha). The third and fourth QTL were identified in genetic linkage groups D2 (qYld19.1) and G (qYld18.1) in regions previously associated with resistance to SCN. The region encompassing *rhg1/Rfs2* on Gm18 between the microsatellite marker TMD1 and Satt610 spanned 15.5 cM (LOD 3.05, 15.8 % variation, F allele 2.37 ± 0.035; H allele 2.91 ± 0.058 Mg/Ha). The region on linkage group D2 between Satt514 and Satt488 spanned 32.6 cM (LOD 2.57,

13.3% variation, F allele 2.79 ± 0.049; H allele 3.1 ± 0.043 Mg/Ha). The QTL detected will allow marker assisted selection to stack seed yield, with pest resistance traits (*rhg1/Rfs2/qYld18.1*; H/H/H allele) and recombinant loci (*Rhg5/Rfs2/qYld19.1*; H/F/H alleles).

Introduction

The main goal of plant breeders is to increase crop yield (Fehr, 1987). Overtime yield improvement has two major inputs; genetic gain and production efficiency (Specht et al., 1999). Genetic gain expressed as seed yield potential, largely derives from improvement in stress and disease resistances. Farmers are responsible for production efficiency by choosing the best varieties and the best production technology.

Soybean seed yield is a multigenic trait (Mansur et al., 1993). Genotype by environment (G × E) interactions, genomic duplications and epistasis complicates phenotypic selection for soybean seed yield (Lark et al., 1996). Further, seed yield at harvest maturity (R8) is a composite trait composed of loci that control plant performance during the growing season (Sun et al., 2006; Palomeque et al., 2009). Hence a wide set of contributing loci are expected. To date the yield potential of lines derived by intercrossing has been difficult to predict without extensive field tests that are expensive and time consuming (Maughan et al., 1996). Components of yield such as maturity, lodging, growth habit (Fehr 1987; Specht et al., 1999) and disease resistance in pathogen infested environments (Concibido et al., 2004; Kas-

* Corresponding author: ga4082@siu.edu

sem et al., 2006) have major effects on yield potential in many segregating populations.

DNA molecular markers have been used extensively to identify and screen for QTL that underlie soybean seed yield and yield components (Mansur et al., 1994; Mansur et al., 1996; Orf et al., 1999b; Yuan et al., 2002; Kassem et al., 2006). The yield QTL identified to date are large genomic regions where at least two distinct alleles have been shown to influence yield. Individual yield QTL regions can explain up to 25% of yield variability and can be detected by bulked segregant analysis. Some have been stable in different genetic backgrounds and environments (Yuan et al., 2002) and some have not (Reyna and Sneller 2001). Some seed yield loci have been shown to be clusters of loci active at different states in plant and seed development (Palomeque et al., 2009)

Molecular markers closely linked to accurately mapped seed yield QTL allow the possibility of early selection of lines by genotype in soybean (Palomeque et al., 2009). Toward this end, hundreds of seed yield QTL have been mapped in diverse germplasms. Several seed yield QTL have been identified in every linkage group of the soybean genome. For example, Orf et al. (1999a,b) identified QTL for seed yield in different linkage groups A2, B1, B2, D1a+Q, D1b+W, D2, F, J, L, and M of the soybean genome. Other authors identified QTL for yield on linkage groups C2, H, M (Specht et al., 1999), F (Reyna and Sneller 2001), J, L, M (Mansur et al., 1993, 1996), and N (Kassem et al., 2006) of the soybean genome. Three QTL for yield were found on linkage groups C1, I, and K using 136 SSR markers and a RIL population of Essex \times Forrest (E \times F) (Yuan et al., 2002). Two genomic regions associated with yield were also identified and mapped on linkage group K using a RIL population of Flyer \times Hartwig (F \times H) (Yuan et al., 2002).

Here we describe further analysis of seed yield loci in the recombinant inbred line (RIL) population derived from F \times H (Yuan et al., 2002; Kazi et al., 2008). The population was used because it segregates for growth habit (determinate to semi-determinate) and about 100 -114 days in maturity. Yield QTL could be examined in the presence of these potentially confounding factors. Here, we report seed yield QTL identified in this populations based on a more extensive map than reported previously.

Materials and Methods

Plant Material

The genetic material used in this study consisted of the F \times H RILs (n = 92; Yuan et al., 2002; Kazi et al., 2008). Populations were advanced to the F_{5:14} from 2003 to 2005 and seed were released in 2007 (Kazi et al., 2008). The cross was selected because it showed segregation for seed yield and many pest resistances, including well characterized reactions to SDS, and SCN. Hartwig was lower yielding but consistently resistant to SDS in most locations (Njiti et al., 1997, 2001; Mueller et al., 2003) and strongly resistant to most HG Types of SCN (Anand 1992). Flyer was higher yielding but susceptible to most SCN HG Types and to SDS (Yuan et al., 2002; Kazi et al., 2007, 2008, 2010).

Methods of Field Trait Measurements

Field trait measurements were measured from 1997 to 2010 by methods described in Yuan et al., (2002). Briefly, the F \times H RIL population along with two parents was grown for increase in non-infested fields. Locations were planted with 4 row plots 6.1 m long and row spacing was 0.75 m apart in a randomized complete block design with three replications. Weeds were controlled with pre- and post- plant herbicides and hand weeding. Rows were examined to remove contaminants or off-types at flowering. Only the center two of the four rows were harvested. Seed yield was measured with cleaned seed at 13 % (w/v) moisture without discarding broken seed. Lodging was score on a 1-5 scale before harvest. Days after planting to maturity was measured at R6 and R8.

SDS Analysis

The population was planted in SDS infested environments as described in Yuan et al. (2002) and Kazi et al. (2007). Briefly at the Southern Illinois University Carbondale Agronomy research center (ARC), Ullin (U) and Ridgway (R) disease incidence (DI), disease severity (DS) and root infection severity (IS) were measured from 1997 to 2000. For disease rating RILs were planted as 2 row plots in a randomized complete block design with 3 replications.

SCN Scores

SCN resistance was determined as described by Yuan et al. (2002) with the following modifications. The complete F_{5:11} F \times H RIL (n=92) population was tested for reactions for SCN Hg Type 0 (PA-3; race 3). Of the HG Type 0 resistant lines the reactions to HG types 1.3.6.7 (PA-14, race 14) and HG Type 1.2.3.5.6.7 (PA2, race 2) were determined (Kazi et al., 2007). Five, three and two single-plant replications were used respectively per RIL line. Cultivar Hutcheson was used as the susceptible control while Peking, Pickett, PI 88788 and PI 90763 were used as the standard differentials to determine Hg biotype.

DNA Marker Analysis

DNA was extracted and used for microsatellite amplifications as in Yuan et al. (2002) as modified by Kazi et al. (2007). Briefly, more than 350 BARC-Satt markers spaced at 5-10 cM intervals (Song et al., 2004) were used. In addition, 140 SIUC-BES-SSR primers from the build 2 MTP BES clones (Shultz et al., 2006ab, 2007) were used chosen to be spaced at 10,000 kbp intervals from the soybean physical map (Shultz et al., 2006ab, 2007). Amplification reactions in 10 μ l, used 15 ng of genomic DNA, 0.75 μ l Taq polymerase (1 unit), 0.75 mM dNTPs, and 15 nM of forward and reverse primers. The PCR products were separated on 4 % (w/v) agarose gels by electrophoresis at 92 V for 4 hours. Gels were stained with ethidium bromide, DNA visualized under UV light and photographed. Flyer and Hartwig amplified DNA samples were included as controls to facilitate manual scoring.

Heritability Estimation

The heritability (h^2) estimates, a ratio of genotypic variation over phenotypic variation of seed yield were calculated using variance components obtained through analysis of variance (ANOVA; SAS Institute Inc. Cary, NC), as described in Fehr (1987). Narrow sense heritability for each trait was calculated. All correlations were calculated using the PROC CORR function of SAS.

Construction of the Genetic Linkage Map

Following Kazi et al. (2007) a linkage map was created using MAPMAKER/EXP 3.0 (Lander et al., 1987). Briefly, centimorgans (cM, Haldane units), LOD 3.0, and maximum distance 50 cM were used. Heterogenous scores for lines were excluded. The recombinant inbred line (RI-selfing genetic model) was used with error detection. Conflicts among the positions of linked markers in $F \times H$ were resolved in favor of experimental evidence when the maps generated at LOD 3.0 disagreed with the composite map of Song et al. (2004).

Construction of QTL Maps

Following Kazi et al. (2008) three methods were used to detect QTL, single point analysis; interval mapping and composite interval mapping.

A. Single Point Analysis was accomplished by line mean comparisons by ANOVA, with mean separation by LSD as described by Njiti et al. (1998). A significant difference ($P < 0.005$) was considered to be a preliminary indication of an association between a marker and a QTL for the trait in question. A value of $P \leq 0.0005$ was suggested by an approximate Bonferroni correction. However, at genomic regions where gaps between adjacent markers were greater than 10 cM in the map associations $0.005 > P > 0.0005$ were accepted as a potentially significant association. Precedents (Hnetkovsky et al., 1996; Chang et al., 1997; Njiti et al., 1997; Kassem et al., 2006) have shown these criteria to be valid in later maps (Meksem et al., 2001a; Njiti et al., 2002; Yuan et al., 2002; Triwitaykorn et al., 2005; Ruben et al., 2006).

B. Interval Maps of QTL were first analyzed by Mapmaker/QTL 1.1 using the F_2 -backcross genetic model for trait segregation (Kazi et al., 2008). Putative QTL were inferred when LOD scores exceeded 2.0, equivalent (but not equal) to a single marker $P < 0.005$ in one-way ANOVA. The position of the QTL was inferred from the LOD peaks at individual loci detected by maximum likelihood tests at positions every 2 cM between adjacent linked markers.

C. Composite Interval Maps of QTL (CIM) used WinQTL Cartographer (version 2.5; Jansen and Stam 1994; Basten et al., 2001). Following Kazi et al. (2008) a walk speed of 2 cM and the forward regression method were selected. QTL were inferred when LOD score peaks exceeded 2.0. To confirm linkage, experiment-wise threshold level 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 significant. The use of lower LOD values has been justified previously (Hnetkovsky et al., 1996; Prabhu et al., 1999; Kazi et al., 2007, 2008, 2010).

Results

Seed Yield Trait Distributions

Grand seed yield ($Mg\ ha^{-1}$) across four environments and 2 years among the 92 recombinant inbred lines from Flyer \times Hartwig showed no significant departure from normality ($P > 0.0001$; Figure 1). The frequency distribution of mean yield was skewed (-0.54) toward low yield limiting parent Hartwig. The distribution was continuous, uni-modal and a significant kurtosis (0.82) that reflected a peaked distribution. There was no line whose yield was higher than Flyer and 18 lines had a mean lower than Hartwig. Nine of the extreme lines were significant transgressive segregants (standard deviation 0.29; $P < 0.0001$).

The mean seed yield at Nashville in 1998 (N98) showed uni-modal and relatively intermediate distribution (Figure 2A). The distribution is skewed towards low yield like Hartwig. Fifteen lines with a mean yield higher than Flyer were significant transgressive segregants (at standard deviation 0.31; $P < 0.0001$). Twelve lines had a N98 mean lower than Hartwig and were significant transgressive segregants.

As shown in Figure 2B, the frequency distribution of the seed yield at Ridgway in 1998 (R98; $Mg\ ha^{-1}$) showed a bimodal distribution. There was evidence for peaks at 3.51 and 4.01. The distribution was skewed towards low yield. A negative kurtosis (-0.08) reflected a platykurtic distribution as these curves have fewer extreme scores than found in a normal distribution. Four lines with a mean yield higher than Flyer (Figure 1) were three were significant transgressive segregants (at standard deviation 0.50; $P < 0.0001$). Nineteen lines had a mean lower than

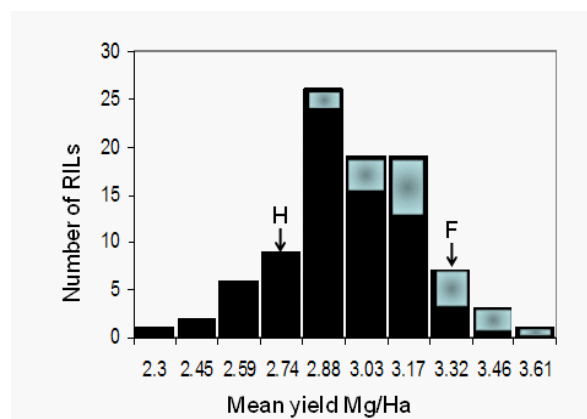


Figure 1. Frequency distributions of the mean seed yield ($Mg\ ha^{-1}$) across four field locations among 92 recombinant inbred lines from a cross between cultivars Flyer and Hartwig. The mean yield score for the parents were arrowed. The least significant difference between parent and line means was 0.015 ($P < 0.05$). The number of SCN resistant RILs in each yield class is shown by light colored boxes. The highest yielding RILs were also SCN resistant. RIL 95 was resistant to both races 3 and 14, whereas RIL 78 was only resistant to race 3.

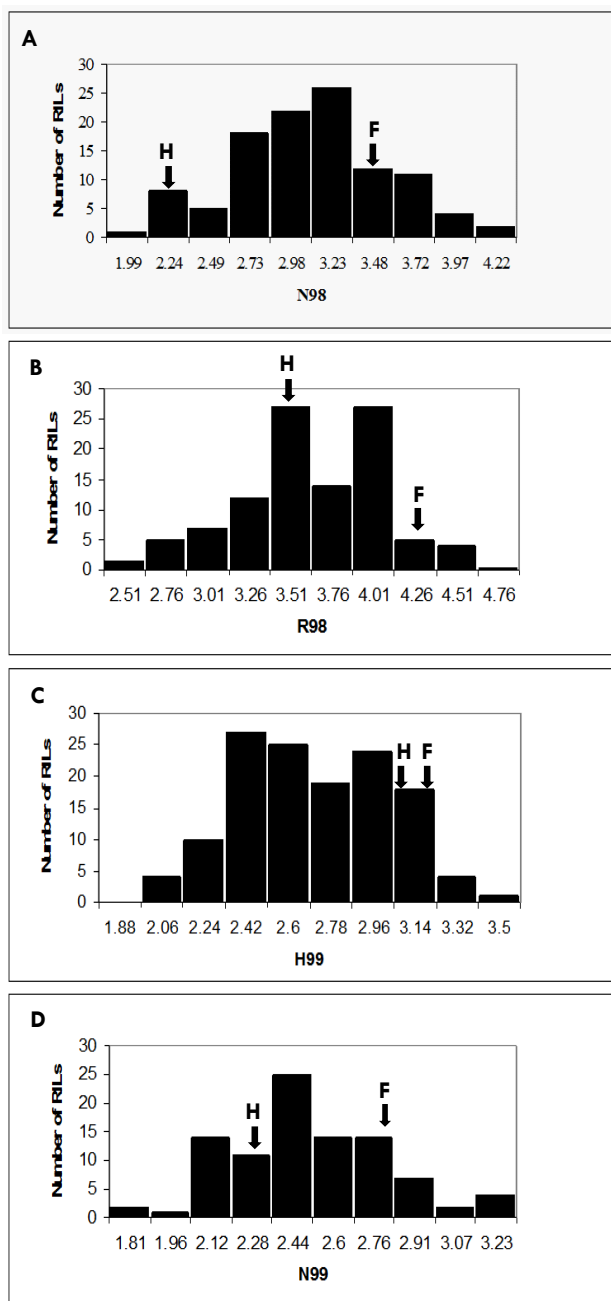


Figure 2. Frequency distributions of the mean yield (Mg ha^{-1}) at four field locations (Nashville 98, Ridgway 98, Harrisburg 99 and Nashville 99) among 92 recombinant inbred lines from the cross $F \times H$. The mean yield for each parent Flyer, Hartwig and lines mean were arrowed.

Hartwig and nine were significantly transgressive.

The frequency distribution at Harrisburg in 1999 (H99; Mg ha^{-1}) showed unimodal distribution as there was peak below 2.78 (Figure 2C). A significant positive kurtosis (1.20) reflected a leptokurtic ('lepto' means slender or narrow) distribution and skewed towards lower yield (left). Here, Hartwig yielded well too. There were five lines that have higher mean yield than Flyer. ($P < 0.0001$).

As shown in Figure 2D, the frequency distribution of the seed yield at Nashville in 1999 (N99; Mg ha^{-1}) showed an approxi-

mately normal ($P = 0.001$) and continuous distribution, although there is evidence for a peak below 2.28. Its distribution was skewed towards low yield Hartwig. A significant kurtosis (0.15) reflected a mesokurtic distribution. Thirteen lines with a mean yield higher than Flyer (Figure 1) were statistically significant transgressive segregants (at standard deviation 0.31; $P < 0.0001$). Eighteen lines had a mean lower than Hartwig and nine were significantly transgressive segregating lines.

Trait Heritability and Correlations

The heritability of mean seed yield on a line mean basis among the recombinant inbred lines was 57% across four environments. Among locations heritability ranged from 72 to 89%. All of the SCN resistant progeny had higher yields than the Hartwig, SCN resistant parent. However, none yielded better and only two as well as the SCN susceptible parent Flyer (Figure 1). There were a significant correlations between mean yield and resistance to SCN race 3 ($P = 0.005$, $R^2 = 31\%$); SCN race 14 ($P = 0.005$, $R^2 = 24\%$); mean SDS in other field locations infested with *F. virguliforme* ($P = 0.001$, $R^2 = 33\%$ for DX) and ($P = 0.006$, $R^2 = 19\%$ for IS at R8). There was no significant association between yield and growth habit ($P = 0.06$, $R^2 = 4.0\%$) or yield and lodging ($P = 0.055$, $R^2 = 9\%$) in the locations tested for yield. Yield means were 2.92 ± 0.03 and $2.79 \pm 0.07 \text{ Mg ha}^{-1}$ for indeterminate ($n = 72$) and determinate ($n = 22$) lines, respectively.

Significant Genomic Regions for Yield

There were two regions detected on linkage group K (Figure 3) associated with mean seed yield across four environments and two consecutive years by CIM (Table 1; Figure 3). The first region on linkage group K, identified by the microsatellite marker Satt539-Satt242, was significantly associated ($P = 0.0005$, $R^2 = 20.5\%$). The interval identified by IM had a peak LOD score of 2.9 and explained about 13% of total variation of soybean yield. However, the interval was not associated with seed yield at every location. The interval was significantly associated ($P = 0.0001$, $R^2 = 34.5\%$) with yield at Harrisburg, IL in 1999 (H99) and at Nashville in 1999 ($P = 0.01$, $R^2 = 6\%$). It was not associated with yield at Ridgway or at Nashville in 1998 ($P = 0.056$, $R^2 = 2.4\%$). Neither Satt539 nor Satt242 was associated with soybean yield at Ridgway in 1998 by IM or ANOVA (not shown). The region derived beneficial allele for seed yield from Hartwig (the Flyer allelic mean was $2.76 \pm 0.06 \text{ Mg ha}^{-1}$; the Hartwig allelic mean was $2.98 \pm 0.03 \text{ Mg ha}^{-1}$).

The second region of linkage group K associated with mean seed yield (Figure 3) was the interval between Satt337 and Satt326. The interval contained a mean seed yield QTL that spanned 1.4cM had a peak LOD of 5.31 and explained about 25.2% (0.4 Mg ha^{-1}) of total variation in soybean yield (Figure 3). The region derived the beneficial allele from Flyer (Table 1). However, at individual locations the interval was no associated with mean seed yield. In addition, by ANOVA the associated locus was only significantly associated with seed yield in N98 and the mean of all locations (Table 1).

Table 1. Examples of markers associated with seed yield at harvest among recombinant inbred lines in two soybean populations. Essex × Forrest was evaluated in Carbondale, IL (C96) and Ridgway, IL (R96) in 1996 and in Desoto, IL (D97) in 1997. Flyer × Hartwig was evaluated in Nashville, IL (N98 and N99) in 1998 and 1999; Ridgway, IL (R98) in 1998 and Harrisburg, IL (H99) in 1999. Clustered yield QTL were found on LG K (Gm9) and G (Gm18) that could be fine mapped in NILs. Loci on C1 (Gm4) and I(Gm20) proved difficult to isolate in NILs and so might be inaccurately mapped or blends of conditional QTL.

Population, marker and map position	Location	P value	R ²	LOD	QTL var	Yield mean±SEM (Mg ha ⁻¹) (alleles from)	
						Flyer	Hartwig
Satt337 LG K (Gm9) 47.4 cM	N98	0.0042	10	5.9	27	3.20 ± 0.06	2.69 ± 0.08
	N99	0.0001	27	2.2	26	2.53 ± 0.05	2.36 ± 0.04
	Mean	0.0006	14	2.7	14	2.98 ± 0.03	2.77 ± 0.05
Satt326 LG K (Gm9) 49.5 cM	N98	0.0001	26	5.4	26	3.20 ± 0.06	2.69 ± 0.08
	N99	0.0082	9	1.6	8	2.52 ± 0.06	2.33 ± 0.04
	Mean	0.0004	15	3	15	2.98 ± 0.04	2.76 ± 0.05
Satt539 LG K (Gm9) 2.0 cM	R98	0.049	5	1.8	5	3.39 ± 0.09	3.62 ± 0.07
	H99	0.0006	15	2.6	13	2.45 ± 0.07	2.74 ± 0.04
	N99	0.012	8	1.5	8	2.35 ± 0.05	2.54 ± 0.05
	Mean	0.0008	14	2.5	13	2.77 ± 0.06	2.99 ± 0.03
TMD1 LG G (Gm18) ‡	N98	0.0047	11	2.5	10	2.88± 0.07	3.38± 0.05
	R98	0.9552	3	0.5	2	3.50± 0.04	3.70± 0.07
	H99	0.0029	10	2.8	11	2.54 ± 0.07	2.99± 0.06
	N99	0.0001	16	4.1	17	2.33± 0.06	2.79± 0.07
	Mean	0.0007	9.7	3.1	27.4	2.32 ± 0.03	2.64 ± 0.05
Satt514_LGD2 (Gm17) 85.7 cM	Mean	0.0006	8.0	2.52	12.1	2.77±0.048	3.0±0.041

†Mean involves all location used for evaluation within each population.

‡ Also significantly associated with SCN resistance (within the intron of the receptor like kinase at *rhg1*).

Table 2. Regions detected by CIM (a.) and ANOVA or IM (b.) that were associated with mean yield (Mg ha⁻¹) across various locations in the Flyer by Hartwig (RIL) population on linkage group K.

Marker	Location	P value	R ² (%)	LOD ^a	QTL var. ^b	Yield Means ± SEM (Mg ha ⁻¹) with Alleles from	
						Flyer	Hartwig
a.		ANOVA	ANOVA	IM	IM	IM	IM
Satt337	Nashville 98	0.0001	12.4	2.06	25.2	3.19±0.05	2.69± 0.07
(K)	Mean yield	0.002	4.6	2.2	20.2	2.98±0.03	2.77±0.04
Satt326	Nashville 98	0.0001	15.4	2.14	14.5	3.14±0.07	3.35±0.05
(K)	YieldMN98	0.006	10	2.6	7.7	3.38±0.05	3.08±0.07
	Mean yield	0.001	6.2	2.50	5.5	2.97±0.04	2.75±0.04
b.				CIM	CIM	CIM	CIM
	Nashville 98	-	-	5.31	25.2	-	-
Satt337 -	Ridgway 98	-	-	2.48	13.6	-	-
Satt326	Mean yield	-	-	2.55	13.8	2.98±0.03	2.77±0.04
(K)							
Satt539 -	Nashville 98	0.056	2.4	1.04	5.6	3.01±0.08	2.94± 0.07
Satt242	Ridgway 98	-	-	-	-	-	-
(K)	Mean 98	0.0001	7.12	3.76	13.1	3.13±0.08	3.34±0.05
	Harrisburg99	0.0001	34.5	2.81	14.5	2.44±0.03	2.747±0.05
	Nashville 99	0.01	6.0	1.3	9.2	2.34±0.05	2.52±0.52
	Mean 99	0.001	5.1	2.65	13.6	2.41±0.05	2.61±0.03
	Mean yield	0.0005	6.5	2.9	13.0	2.76±0.05	2.98±0.03

a. LOD was the probability of the presence of a locus.

b. var. was the amount of variability in seed yield explained by the marker loci.

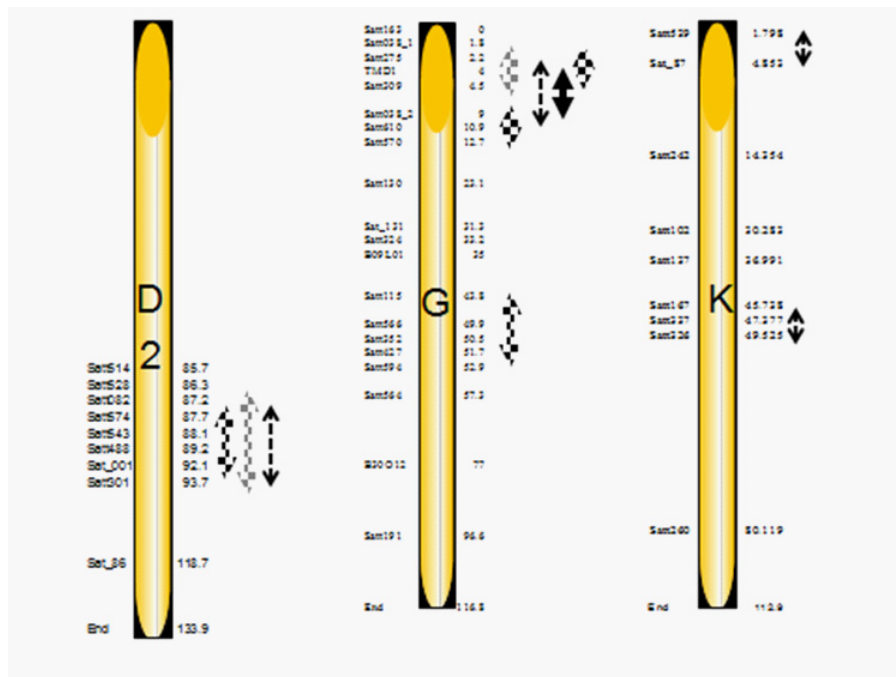


Figure 3. Genetic map of seed yield QTL on linkage group D2, G and K (chromosomes 17, 18 and 12 respectively) in the F × H RIL population. Estimated map distances are shown (cM). QTL are presented to the right of the linkage group as arrows. Discontinuous black arrows are QTL underlying seed yield. The QTL underlying resistance to SDS are black stippled arrows. The QTL underlying resistance to SCN race 14 are grey stippled arrows. The QTL underlying resistance to SCN race 3 are black solid arrows.

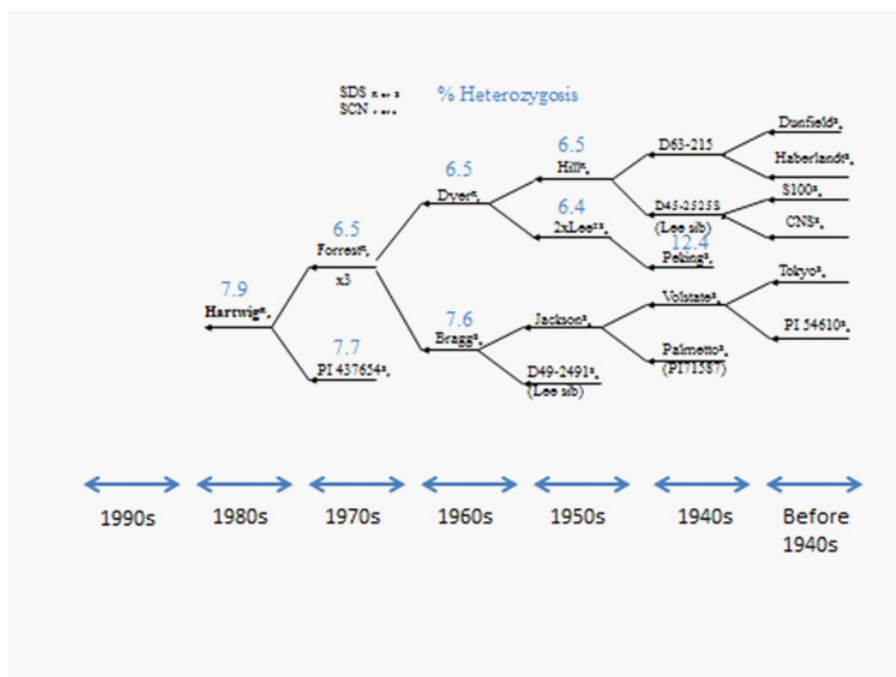


Figure 4. Parents and ancestors of Hartwig and Flyer. SDS and SCN resistance is indicated along with approximate year of release and percent of heterozygosity in a single plant estimated by GBS where known (blue number; Li et al., 2014).

Table 3. Regions on linkage groups G (18) and D2 (17) detected by ANOVA and IM (a.) or CIM (b.) that were associated with mean yield (Mg ha^{-1}) and disease resistance in the Flyer by Hartwig (RIL) population.

Marker or Interval	Location	P value	R ² (%)	LOD ^a	QTL var. ^b (%)	Yield Means \pm SEM (Mg ha^{-1}) With Alleles from	
						Flyer	Hartwig
a.		ANOVA	ANOVA	IM	IM	IM	IM
TMD1 (G/18)	Mean yield	0.0007	9.7	.07	27.4	2.32 \pm 0.03	.64 \pm 0.05
Satt610 (G/18)	Mean yield	0.01	9.0	2.57	11.8	2.72 \pm 0.09	.96 \pm 0.052
Satt514 (D2/17)	Mean yield	0.0006	8.0	2.52	12.1	2.77 \pm 0.048	3.0 \pm 0.041
b.				CIM	CIM	CIM	CIM
TMD1-Satt610 (15.5cM) (G/18)	Mean yield	-	-	3.05	15.8	2.37 \pm 0.035	2.91 \pm 0.058
Satt514-Satt488 (32.6cM) (D2/17)	Mean yield	-	-	2.57	13.3	2.79 \pm 0.049	3.1 \pm 0.043

a. LOD was the probability of the presence of a locus.

b. var. was the amount of variability in seed yield explained by the marker loci.

Two additional QTL were identified in genetic linkage groups D2 and G of the F \times H population. One region on linkage group G (chromosomes 17 and 18: Figure 3) identified by the microsatellite marker TMD1 was significantly ($P=0.0005$, $R^2=9.7\%$) associated with mean yield over two years (1998, 1999) in the population. The adjacent marker was Satt610. The interval between these two markers spanned a genetic distance of about 15.5 cM contained a yield QTL (Figure 3) and had a peak Log-likelihood (LOD) of 3.05 and explained about 15.8 % of total variation in mean yield (Table 2). This region was also significantly associated with resistance to SCN HG Types 0 and 1.3.5- (Kazi et al., 2007, 2009) and is the likely location for rhg1. The region on linkage group D2 (Figure 3) identified by the microsatellite marker Satt514 was significantly ($P=0.0006$, $R^2=7.4\%$) associated with seed yield (across environment Grand mean yield; 1998-1999) in the population at four locations. The adjacent marker was Satt488. The interval between these two markers spanned a genetic distance of about 32.6 cM. The yield QTL had peak Log-likelihood (LOD) of 2.57 and explained about 13.3% of total variation (Table 2).

Discussion

The F \times H population was very important for the analysis of seed yield QTL and other agronomic traits because it segregates for many traits (Kazi et al., 2007). A QTL for DAP (days after planting) maturity was found to be located on Linkage group C2 by Satt277 that gave a highly significant Peak-LOD of 6.45 and underlies segregation across maturity groups 4 and 5. Previously, this locus has been found significantly associated with lodging and plant height in Minsoy \times Archer (Orf et al., 1999a), seed weight and seed yield in Noir1 \times Archer (Orf et al., 1999a) and plant height/maturity date in Noir1 \times Archer (Orf et al., 1999b).

For quantitative traits, genotype and environment interactions (G \times E) are commonly observed (Kearsey and Farquar 1998). Specific environmental conditions have effects on quantitative trait values like yield and lodging (Brim 1973). Environment seems to be an important factor in the number of QTL detected. For instance, in F \times H, there was no yield QTL that associates with all environments, suggesting that these QTL might be environment specific (Palomeque et al., 2009). The QTL that was identified based on average data from two years may be the most significant QTL for breeders, because these may represent the genes that will lead to yield stability irrespective of the environment.

QTL detected in one population could be detected at similar positions in different populations (Lin et al., 1995). Most populations had a reasonably good distribution of markers across the major linkage groups, like in E \times F (Meksem et al., 1999) but inevitably gaps were present that prevented detection of QTL across populations. Also, different QTL could be identified in different populations due to genetic background effects. This is probably the reason why a large effect is detected in one population but not in others (Brummer et al., 1997). As all populations were not mapped with the same markers, it is difficult to provide reliable estimates of the number of QTL that are common across various populations. However, common QTL are detected.

In this study, F \times H RIL verified putative QTL for yield on linkage group K. The genetic marker Satt337 was significantly associated with yield in both F \times H and E \times F population (Yuan et al., 2002; Table 3). This indicates that the QTL associated with this marker is consistent. Here the susceptible parent, Flyer, in F \times H provided the beneficial allele. As shown in Figure 4, these parents share a significant amount of their genome with Essex and Forrest i.e. Flyer originated from the cross Essex \times L24. On the other hand, Hartwig was derived from a cross between Forrest3 \times PI437654. Two new QTL for yield were also detected

in linkage group G (TMD1) and D2 (Satt514) but they were not responsible for yield drag associated with SCN resistance. It is possible low SCN pressures in the environment used were sufficient to reduce yield in susceptible lines. Cultivars that have shown a history of high yield should be crossed with cultivars with alleles or genes for SDS and SCN. Recombination events should be identified within the loci reported and used for breeding purposes to provide a good source for high yielding and resistant genes.

Acknowledgements

This research was funded by grants from the United Soybean Board to DAL; and the family of YCL. The authors thank Dr. P. Gibson, O. Myers Jr. and M. Schmidt for assistance with germplasm development and maintenance from 1991-2000. We thank J.H. Klein III for assistance with germplasm maintenance from 1991-2011. We thank Ted Ballard for assistance with the population from 2012 to 2015.

References

- Anand SC (1992) Registration of 'Hartwig' soybean. *Crop Sci* 32: 1060-1070.
- Basten CJ, BS Weir, and Z Zeng (2001) QTL Cartographer Version 2.0. Raleigh, NC: Department of Statistics, North Carolina State University, USA.
- Concibido VC, BL Valley, P McLaird, N Pineda, J Meyer, L Hummel, J Yang, K Wu, and X Delannay (2003) Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. *Theor Appl Genet* 106: 575-582.
- Chung J, HL Babka, GL Graef, PE Staswick, DJ Lee, PB Cregan, RC Shoemaker, and JE Specht (2003) The seed protein, oil and yield QTL on soybean linkage group I. *Crop Sci* 43: 1053-1067.
- Fehr RW (1987) Principals of cultivar development. Vol 1: Theory and techniques. McMillan, New York.
- Ferris R, TR Wheeler, RH Ellis, and P Hadley (1999) Seed yield after environmental stress in soybean grown under elevated CO₂. *Crop Sci* 39: 710-718.
- Frederick JR, CR Camp, and PJ Bauer (2001) Drought-stress effects on branch and main stem seed yield and yield components of determinate soybean. *Crop Sci* 41: 759-763.
- Gizlice Z, T Carter Jr, TM Gerig, and JW Burton (1996) Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. *Crop Sci* 36: 753-765.
- Hamilton JG, O Dermody, M Aldea, AR Zangerl, A Rogers, MR Berenbaum, and EH DeLucia (2005) Anthropogenic changes in tropospheric composition increase susceptibility of soybean to insect herbivory. *Environ Entomol* 34: 479-485.
- Han Y, X Zhao, D Liu, Y Li, DA Lightfoot, Z Yang, L Zhao, G Zhou, Z Wang, L Huang, and Z Zhang (2016) Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytologist* 209 (2): 871-884.
- Holsinger KE, Lewis PO, Dey DK (2002) A Bayesian approach to inferring population structure from dominant markers. *Mol Ecol* 11: 1157-1164.
- Hnetkovsky N, SJC Chang, TW Doubler, PT Gibson, and DA Lightfoot (1996) Genetic Mapping of Loci Underlying Field Resistance to Soybean Sudden Death Syndrome (SDS). *Crop Sci* 36 (2): 392-400.
- Kabelka EA, SR Carlson, and BW Diers (2006) *Glycine soja* PI 468916 SCN resistance loci associated effects on soybean seed yield and other agronomic traits. *Crop Sci* 46: 622-629.
- Karangula UB, MA Kassem, L Gupta, HA El-Shemy, and DA Lightfoot (2009) Locus interactions underlie seed yield in soybeans resistant to *Heterodera glycines*. *Curr Issues Mol Biol* 11 (Suppl. 1): i73-84.
- Kassem MA, JL Shultz, K Meksem, AJ Wood, MJ Iqbal, and DA-Lightfoot (2006) An updated Essex by Forrest linkage map and first composite interval map of QTL underlying six soybean traits. *Theor Appl Genet* 113: 1015-1026.
- Kazi S, VN Njiti, TW Doubler, J Yuan, MJ Iqbal, S Cianzio, and DA Lightfoot (2007) Registration of the Flyer by Hartwig Recombinant Inbred Line Mapping Population. *J Plant Reg.* 1: 175-178.
- Kazi S, J Shultz, R Bashir, J Afzal, VN Njiti, and DA Lightfoot (2008) Separate loci underlie resistance to soybean sudden death syndrome in 'Hartwig' by 'Flyer'. *Theor Appl Genet* 116: 967-977.
- Kazi S, J Shultz, J Afzal, R Hashmi, M Jasim, J Bond, PR Arelli, and DA Lightfoot (2010) Iso-lines and inbred-lines confirmed loci that underlie resistance from cultivar 'Hartwig' to three soybean cyst nematode populations. *Theor Appl Genet* 120: 633-640.
- Lark KG, K Chase, F Adler, LM Mansur, and JH Orf (1995) Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. *Proc Natl Acad Sci USA* 92: 4656-4660.
- Luckew A, L Leandro, M Bhattacharyya, D Nordman, DA Lightfoot, and S Cianzio (2013) Usefulness of ten genomic regions in soybean associated with sudden death syndrome resistance. *Theor Appl Genet* 126: 2391-2403.
- Mansur LM, JH Orf, and KG Lark (1993) Determining the linkage of qualitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (*Glycine max* L. Merr.). *Theor Appl Genet* 86: 914-918.
- Mansur LM, JH Orf, K Chase, T Jarvik, PB Cregan, and KG Lark (1995) Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Sci* 36: 1327-1336.
- Nakagawa S (2004) A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology* 15: 1044-1045.
- Njiti VN, RJ Suttner, LE Gray, PT Gibson, and DA Lightfoot (1997) Rate reducing resistance to *Fusarium solani* underlies field resistance to soybean sudden death syndrome (SDS). *Crop Sci* 37 (1): 132-138.
- Njiti VN, TW Doubler, RJ Suttner, L Gray, PT Gibson, and DA Lightfoot (1998) Loci underlying resistance to *Fusarium solani* and soybean Sudden Death Syndrome (SDS) mapped in near isogenic lines. *Crop Sci* 38 (3): 472-477.
- Njiti VN, K Meksem, MJ Iqbal, JE Johnson, KF Zobrist, VY Kilo, and DA Lightfoot (2002) Common loci underlie field resis-

- tance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. *Theor Appl Genet* 104: 294–300.
- Orf JH, K Chase, T Jarvik, LM Mansur, PB Cregan, FR Adler, and KG Lark (1999) Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. *Crop Sci* 39: 1642–1651.
- Palomeque L, L Li-Jun, W Li, B Hedges, ER Cober, and I Rajcan (2009) QTL in mega-environments: I. Universal and specific seed yield QTL detected in a population derived from a cross of high-yielding adapted \times high-yielding exotic soybean lines. *Theor and Appl Genet* 119 (3): 417–427.
- Prabhu RR, VN Njiti, JE Johnson, ME Schmidt, RJ Klein, and DA Lightfoot (1999) Selecting soybean cultivars for dual resistance to cyst nematode Sudden Death Syndrome with two DNA markers. *Crop Sci* 39 (4): 982–987.
- Rector BG, JN All, WA Parrott, and HR Boerma (2000) Quantitative trait loci for antibiosis resistance to corn earworm in soybean. *Crop Sci* 40: 233–238.
- Reyna N and CH Sneller (2001) Evaluation of markers-assisted introgression of yield QTL alleles into adapted soybean. *Crop Sci* 41: 1317–1321.
- Specht JE, DJ Hume, and SV Kumudini (1999) Soybean yield potential—a genetic and physiological perspective. *Crop Sci* 39 (6): 1560–1570.
- Sun D, W Li, Z Zhang, Q Chen, H Ning, L Qiu, and G Sun (2006) Quantitative trait loci analysis for the developmental behavior of soybean (*Glycine max* L. Merr.). *Theor and Appl Genet* 112 (4): 665–673.
- Wang D, GL Graef, AM Procopiuk, and BW Diers (2003) Identification of putative QTL that underlie yield in interspecific soybean backcross populations. *Theor Appl Genet* 108: 458–467.
- Yuan J, VN Njiti, K Meksem, MJ Iqbal, K Triwitayakorn, MA Kassem, GT Davis, ME Schmidt, and DA Lightfoot (2002) Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. *Crop Sci* 42:271–277.