Genetic Introgression from *Glycine tomentella* to Soybean to Increase Seed Yield

Abraham Akpertey[1], Ram Singh[2], Carrie Thurber[3], Brian Diers[1], George Graef[4], Rouf Mian[5], Grover Shannon[6], Andrew Scaboo[6], Patrick Brown[7], and Randall Nelson[8]

[1]Dep. of Crop Sci., Univ. of Illinois, Urbana, IL; [2]USDA-ARS, Urbana, IL; [3]EBI, Univ. of Illinois, Urbana, IL; [4]Univ. of Nebraska, Lincoln, NE; [5]USDA-ARS/The Ohio State Univ., Wooster, OH; [6]Div. of Plant Sciences, Univ. of Missouri, MO; [7]IGB, Univ. of Illinois, Urbana, IL; [8]USDA-ARS and Univ. of Illinois, Urbana, IL

Objectives

- To identify high yielding lines derived from backcrosses between the cultivar Dwight and *G. tomentella* PI 441001.
- To determine the genetic contribution of PI 441001 to the derived lines.

Materials and Methods

Field of *G. max* × *G. tomentella* backcross lines

G. tomentella

Population development

The *G. max* × *G. tomentella* backcross lines were developed using the cultivar Dwight and the perennial *Glycine* species, *G. tomentella* PI 441001. PI 441001 was crossed to Dwight and immature seed rescue was used to produce a sterile F_1 plant. Amphidiploid plants (2n=118) were produced by treating the F_1 hybrid with colchicine. Amphidiploid plants were backcrossed to Dwight to obtain BC_1 plants. A series of backcrosses were made with eight different BC_2 plants to obtain BC_3, BC_4, BC_5 and BC_6 lines with 2n=40 chromosomes.

Field procedures

- Preliminary yield testing of inbred lines from these crosses was used to select 180 lines in maturity groups II, III, and IV. The derived lines and checks were evaluated in two replication tests in eleven to twelve environments in Illinois, Ohio, Missouri and Nebraska in 2013 and 2014.

Genotyping procedures

- All experimental and parental lines were genotyped using genotyping by sequencing (GBS). GBS libraries were prepared using *HindIII-Bfa1* and *HindIII-Hinp1* enzyme combinations. Sequencing was done with Illumina HiSeq 2500 at the W.M. Keck Center at the University of Illinois.

Statistical analysis

The field data were analyzed using the PROC Mixed and GLM functions in SAS (SAS 9.3, SAS Institute, Cary NC). GBS data was analyzed using the TASSEL GBS pipeline (Buckler lab for maize genetics and diversity, http://www.maizegenetics.net/). Association analysis was performed in R using a simple linear model.

Results and Discussion

Table 1. Data for high yielding lines, the recurrent parent Dwight, and the best check from Test II from 12 environments in Nebraska, Missouri, Illinois and Ohio in 2013 and 2014.

Entry	BC†	Yield (kg/ha)	Mat. (days)	Ldg. (1 to 5)	Hgt. (cm)	BC_2 Parent
LG11-1222	5	+311	0	0	-6	06H1-3
LG10-12313	4	+306	1	0.3	-1	06H1-3
IA2102		+294	-2	0.4	0	
Dwight		4194	Sept. 16	1.9	86	
LSD (0.05)		240	1.0	0.2	3.0	

†Backcrosses

Table 2. Data for high yielding lines from different BC_2 parents, the recurrent parent Dwight, and the best check from Test III from 11 environments in Nebraska, Missouri, Illinois and Ohio in 2013 and 2014.

Entry	BC†	Yield (kg/ha)	Mat. (days)	Ldg. (1 to 5)	Hgt. (cm)	BC_2 Parent
LG11-2963	3	+483	6	0.5	6	06H1-3
LG11-10931	3	+299	8	0.2	2	07H1-7
LG11-3223	3	+281	6	0.1	6	06H1-1
LG11-1340	4	+239	7	0.3	7	07H1-14
IA3023		+493	6	0.1	5	
Dwight		3529	Sept. 13	1.8	80	
LSD (0.05)		203	1.0	0.2	3.1	

†Backcrosses

Table 3. Data for high yielding lines from different BC_2 parents, the recurrent parent Dwight, and best check for Test IV from 11 environments in Nebraska, Missouri, Illinois and Ohio in 2013 and 2014.

Entry	BC†	Yield (kg/ha)	Mat. (days)	Ldg. (1 to 5)	Hgt. (cm)	BC_2 Parent
LG11-3187	3	+513	6	0.1	5	06H1-1
LG09-12682	3	+355	8	-0.4	1	06H1-3
LG11-4475	5	+269	7	0.2	3	07H6-17
IA4005		+724	13	-0.2	0	
Dwight		3624	Sept. 14	1.9	81	
LSd (0.05)		185	1	0.2	3.0	

†Backcrosses

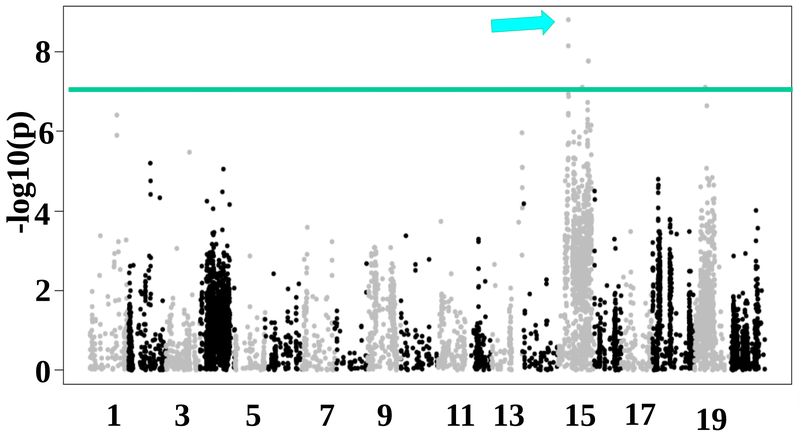

Figure 1. Manhattan plot depicting association between 6833 SNPs with seed yield, distributed across the 20 soybean chromosomes.

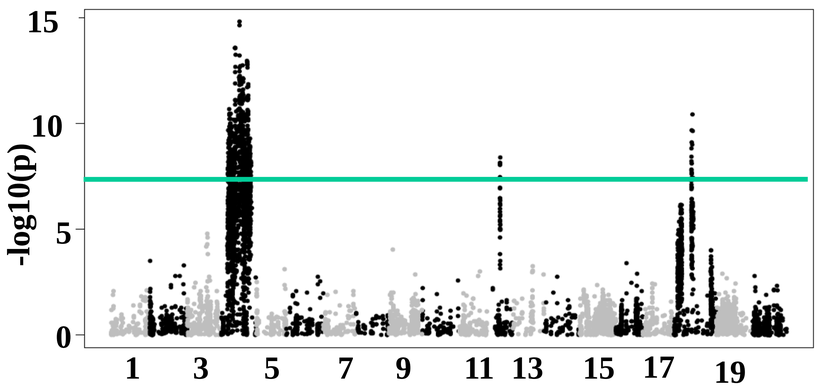

Figure 2. Manhattan plot depicting association between 6833 SNPs with maturity, distributed across the 20 soybean chromosomes.

From each test, we identified lines that were higher yielding than Dwight, the recurrent parent (Tables 1, 2 and 3). These superior lines were derived from 5 different BC_2 plants. Each BC_2 plant is likely to have a different complement of *G. tomentella* chromosomes. In Tests II and III, the highest yielding experimental lines were similar in maturity and yield to the best checks in each test. In Test IV the highest yielding experimental line was 7 d earlier than the best check.

A Manhattan plot (Fig. 1) of the association between seed yield and SNPs obtained from GBS showed 5 SNPs on chromosomes 15 and 19 that were significantly (-log 10 $P > 7$) associated with seed yield. The most significant SNP (arrowed) at 15.3 Mbp on chromosome 15 was associated with a 120 kg/ha, 119 kg/ha and 252 kg/ha increase in yield in Tests II, III and IV, respectively.

Fig. 2 shows a Manhattan plot of the association between the tested SNPs and date of maturity. A large number of SNPs (> 100) on chromosome 4 and a moderate number of SNPs (< 50) on chromosomes 12 and 18 that were significantly (-log10 $P > 7$) associated with maturity were identified.

The significant SNPs for seed yield and maturity are located on different chromosomes, indicating that the yield loci are separate from the maturity loci.

Conclusions

- Wide hybridization between *G. tomentella* and soybean can result in lines with increased seed yield.
- Because of the large genetic differences between soybean and *G. tomentella*, it is likely that the yield increases are a result of genes not in *G. max*.
- The highest yielding lines are good candidates for inclusion in breeding programs.

UNITED SOYBEAN BOARD

NCSRP NORTH CENTRAL SOYBEAN RESEARCH PROGRAM