Published online May 3, 2018

RESEARCH

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

A. Akpertey, R. J. Singh, B. W. Diers, G. L. Graef, M. A. R. Mian, J. G. Shannon, A. M. Scaboo, M. E. Hudson, C. S. Thurber, P. J. Brown, and R. L. Nelson*

ABSTRACT

Soybean [*Glycine max* (L.) Merr.] breeding in the United States currently relies on a narrow genetic base. *Glycine tomentella* Hayata (2*n* = 78), native to Australia, is a perennial relative in the tertiary gene pool of soybean. No effort has been devoted to using this species to increase seed yield. The objectives of this research were (i) to identify high-yielding lines derived from backcrosses between the soybean cultivar Dwight (2*n* = 40) and *G. tomentella* PI 441001 (2*n* = 78) and to compare their agronomic performance with that of the recurrent parent, and (ii) to determine associations between *G. tomentella* introgressions and agronomic traits. PI 441001 was crossed to Dwight and immature seed rescue was used to produce a sterile F_1 plant. Amphidiploid plants (2*n* = 118) were produced by treating the F_1 hybrid with colchicine. Amphidiploid plants were backcrossed to Dwight and a series of backcrosses were made to obtain lines with 2*n* = 40 chromosomes, which were self-fertile and genetically stable. Preliminary yield testing was used to select 180 lines in maturity groups II, III, and IV that had yields greater than or less than that of the recurrent parent. Yield data collected in two replication tests at six to eight locations in 2013 and 2014 identified experimental lines in all three maturity groups that were higher yielding than the recurrent parent. All experimental and parental lines were genotyped, revealing an average of 1% *G. tomentella* introgression. Significant associations were detected between *G. tomentella* introgressions and time of flowering, height, lodging, and yield.

A. Akpertey, B.W. Diers, M.E. Hudson, P.J. Brown, and R.L. Nelson, Dep. of Crop Sciences, Univ. of Illinois, 1101 W. Peabody Dr., Urbana, IL 61801; A. Akpertey, current address, Cocoa Research Institute of Ghana, PO Box 8, New Tafo-Akim, Eastern Region, Ghana; R.J. Singh and R.L. Nelson, USDA-ARS, Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, 1101 W. Peabody Dr., Urbana, IL 61801; G.L. Graef, Dep. of Agronomy and Horticulture, Univ. of Nebraska, 202 Keim Hall, Lincoln, NE 68583; M.A.R. Mian, USDA-ARS, Soybean and Nitrogen Fixation Research Unit, 3127 Ligon St., Raleigh, NC, 27607; J.G. Shannon, Division of Plant Sciences, Univ. of Missouri-Delta Center, PO Box 160, Portageville, MO 63873; A.M. Scaboo, Division of Plant Sciences, Univ. of Missouri, 1-31 Agriculture Building, Columbia, MO 65211; C.S. Thurber, Dep. of Science and Math, Abraham Baldwin Agricultural College, ABAC 15, 2802 Moore Highway, Tifton, GA 31793; C.S. Thurber, current address, American Journal Experts, LLC, 601 West Main St., Suite 102, Durham, NC 27701; P.J. Brown, Institute for Genomic Biology, 1206 W. Gregory Ave., Urbana, IL 61801; P.J. Brown, current address, Univ. of California, Davis, Dep. of Plant Sciences, One Shields Ave., Davis, CA 95616;. Received 25 July 2017. Accepted 20 Nov. 2017. *Corresponding author (rlnelson@illinois.edu). Assigned to Associate Editor Stella Kantartzi.

Abbreviations: Chr, chromosome; GBS, genotyping-by-sequencing; MG, maturity group; QTL, quantitative trait loci; SNP, single-nucleotide polymorphism.

Commercial soybean [*Glycine max* (L.) Merr.] breeding in the United States currently relies on a narrow genetic base. Approximately half of the genetic contribution, calculated by pedigree analysis, comes from only five ancestral lines, and >80% of the North American gene pool was derived from fewer than a dozen introductions (Gizlice et al., 1994). For decades, but more intensely in recent years, efforts have been made to incorporate exotic soybean germplasm into the breeding pool.

Published in Crop Sci. 58:1277–1291 (2018).
doi: 10.2135/cropsci2017.07.0445

Species are separated by barriers preventing hybridization and gene flow. Harlan and de Wet (1971) proposed an informal three gene pool concept in which species are classified as being part of the primary (GP-1), secondary (GP-2), and tertiary (GP-3) gene pools according to plant breeders' experience with the rate of successful hybridization. Crosses within GP-1 species can be made easily and produce hybrids that are vigorous, exhibit normal meiotic chromosome pairing, and possess total fertility. GP-2 consists of species that can be crossed with GP-1 with some problems of F_1 infertility. GP-3 is the outer limit of potential genetic resources associated with cultivated species. Crosses with GP-3 species produce hybrids that are lethal or totally sterile, and gene transfer is not possible without extraordinary techniques including embryo culture, grafting, and chromosome doubling.

Most successful use of wild species to improve yield, a complex trait controlled by many genes, has involved species within the primary and secondary gene pools. For example, yield was improved in corn (*Zea mays* L.) (Prescott-Allen and Prescott-Allen, 1986), potato (*Solanum tuberosum* L.) (Gur and Zamir, 2004) and rice (*Oryza sativa* L.) (Brar, 2005). For legume crops, there is documentation for common bean (*Phaseolus vulgaris* L.) (Kelly, 2004) and chickpea (*Cicer arietinum* L.) (Jaiswal et al., 1987; Singh and Ocampo, 1993; Singh and Ocampo, 1997; Singh et al., 2005). In soybeans, an earlier attempt to broaden the genetic base using *Glycine soja* Sieb. & Zucc. was demonstrated by Hartwig (1973) who reported highly productive, high-protein lines derived from soybean by *G. soja* hybrids. Using the backcross breeding method, Lee et al. (2004) developed sprout soybeans with small seeds (100 mg seed^{-1}) and desirable agronomic traits from crosses involving *G. soja* accession KLG10084 and two *G. max* cultivars, Eunhakong and Sobaegnamulkong.

A number of quantitative trait loci (QTL) mapping studies have reported QTL for seed yield in soybean wild relatives. Using the advanced backcross method, Concibido et al. (2003) identified a *G. soja* allele from PI 407305 that increased seed yield, but that increase was expressed in only two of six genetic backgrounds. Wang et al. (2004) identified four significant ($P < 0.05$) yield QTL on linkage groups C2 (chromosome [Chr] 6), E (Chr 15), K (Chr 9), and M (Chr 7) from the *G. soja* PI 468916. One positive-yield QTL from the *G. soja* PI 245331 on LG A1 (Chr 5) was reported by Li et al. (2008). In more recent years, experimental lines have been derived from multiple *G. soja* accessions that are not significantly different from their recurrent soybean parents for seed yield (Akpertey et al., 2014). The percentage of *G. soja* single-nucleotide polymorphism (SNP) introgression ranged from 9 to 58% among BC_1 lines and from 1 to 32% among the BC_2 lines evaluated. There were both high- and low-yielding lines with <10% *G. soja* alleles from the *G. soja* donor parent PI 479767. A high-yielding line with 18% *G. soja* alleles was identified in lines derived from the *G. soja* donor parent PI 483461, but there were also lines with similar percentages of *G. soja* alleles that were significantly lower yielding, indicating that the percentage of *G. soja* alleles was not a good predictor of seed yield. Even with intense selection pressure to successfully recover good agronomic types, an average of 13% of alleles in the BC_2–derived lines came from the *G. soja* parents (Akpertey et al., 2014).

Despite the problems associated with crossing in the tertiary gene pool, there are benefits to making this diversity available for crop breeding. The use of tertiary gene pool species to transfer disease and pest resistance genes has been successful in the cereal crops of wheat (*Triticum aestivum* L.) (Henry et al., 1996; Mujeeb-Kazi et al., 1996; Singh et al., 2001) and rice (Jena and Khush, 1990; Amante-Bordeos et al., 1992; Brar et al., 1996), in the vegetable crop tomato (*Solanum lycopersicum* L.) (Rick and Chetelat, 1995; Chetelat et al., 2009), and in legume crops including common bean (*Phaseolus vulgaris* L.) (Honma, 1956; Coyne and Schuster, 1974), chickpea (Verma et al., 1995; Singh et al., 1999; Badami et al., 1997; Mallikarjuna 1999; Mallikarjuna and Jadhav 2008), and pigeonpea [*Cajanus cajan* (L.) Millsp.] (Mallikarjuna et al., 2005, 2006).

Wide hybridization experiments in soybean date back to 1979 (Ladizinsky et al., 1979), and hybrids have been produced between *G. max* and the perennial species *G. argyrea* Tindale (Grant et al., 1986), *G. canescens* F. J. Hermann (Broué et al., 1982), *G. clandestina* Wendl. (Singh et al., 1987), *G. falcata* Benth. (Newell et al., 1987), *G. tabacina* (Labill.) Benth. (Newell et al., 1987), and *G. tomentella* Hayata (Newell et al., 1987). Due to sterility or poor vigor, all but the hybrids with *G. tomentella* did not progress beyond the F_1 or amphidiploid stages.

Glycine tomentella is a species complex with plants having chromosome numbers of $2n = 38$, 40, 78, and 80. They can be further subdivided into five diploid and six tetraploid forms, each of which could be considered a species on the basis of reproductive isolation (Doyle et al., 2004). Hymowitz and Singh (1984) produced the first fertile *G. max* × *G. tomentella* ($2n = 118$) amphidiploids from an F_1 hybrid produced by Newell and Hymowitz (1982) and later repeated by Newell et al. (1987), both of whom doubled chromosome numbers of F_1 hybrids ($2n = 59$) grafted onto soybean using colchicine (Cheng and Hadley, 1983). Singh and Nelson (2015) successfully hybridized *G. max* to *G. tomentella* and produced fertile soybean plants, some of which were evaluated in this study.

Several useful traits have been identified in *G. tomentella*. Such reported traits include disease resistances to soybean rust (*Phakopsora pachyrhizi* Syd.) (Schoen et al., 1992; Hartman et al., 1992), soybean cyst nematode (*Heterodera glycines* Ichinohe) (Brucker 2004; Patzoldt et al., 2007), *Bean pod mottle virus* (Zheng et al., 2005), Sclerotinia

stem rot [*Sclerotinia sclerotiorum* (Lib.) de Bary] (Hartman et al., 2000), and sudden death syndrome [*Fusarium solani* (Mart.) Sacc. f. sp. *glycines*](Hartman et al., 2000). To date, no effort has been devoted to using this species to increase the seed yield of soybean.

The analysis of genomic variation is an essential part of crop improvement programs. One approach to identify SNPs among plants is genotyping-by-sequencing (GBS), which uses enzyme-based complexity reduction (using restriction endonucleases to target only a small portion of the genome) coupled with barcoded adapters to produce multiplex libraries of samples ready for next-generation sequencing. This approach has been demonstrated to be robust across a range of species and capable of producing tens of thousands to hundreds of thousands of molecular markers (Elshire et al., 2011; Poland et al., 2012). The flexibility of GBS in regards to species, populations, and research objectives makes this an ideal tool for plant genetics studies. It brings high-density genotyping to the vast majority of plant species that, until now, have had little to no investment in genomics resources. One of the most powerful applications of GBS is in the field of plant breeding. The objectives of this research were (i) to identify high-yielding lines derived from backcrosses between the soybean cultivar Dwight and *G. tomentella* PI 441001 ($2n = 78$) and to compare their agronomic performance with that of the recurrent parent, and (ii) to determine associations between *Glycine tomentella* introgressions and agronomic traits.

MATERIALS AND METHODS

Line Development

Methodology to produce fertile plants with $2n = 40$ chromosomes from soybean cultivar Dwight ($2n = 40$) (Nickell et al., 1998) × *G. tomentella* PI 441001 ($2n = 78$) is described by Singh (2010) and Singh and Nelson (2015), including procedures for extracting immature seeds and media composition for culturing these seeds. Briefly, young flower buds (prior to anthesis) of Dwight were emasculated and gynoecia were pollinated with pollen from newly opened flowers of PI 441001. Putative fertilized gynoecia were sprayed with a growth hormone mixture containing 100 mg gibberellic acid, 35 mg naphthalene acetic acid, and 5 mg kinetin per liter of distilled water 24 h post-pollination once a day for 19 to 21 d. Pods were removed and immature seeds were cultured in seed maturation medium to develop embryonic callus, allowing the production of multiple embryos per seed. Seeds were transferred to fresh medium every 2 to 3 wk and continued until callus developed, usually after ~3 mo. Embryonic calluses were transferred to embryo and shoot regeneration medium. Developing shoots were transferred to the rooting medium, and eventually developing seedlings were planted in pots with soil and moved to the greenhouse.

An amphidiploid plant ($2n = 118$) was produced by treating F_1 hybrid shoots with 0.1% colchicine in culture (Singh, 2010). Chromosome doubling was verified using Feulgen staining technique (Singh, 2003). Amphidiploid plants were backcrossed to Dwight, and BC_1 plants with $2n = 79$ chromosomes were produced through the immature seed rescue procedure. BC_1 plants were backcrossed to Dwight to obtain BC_2 plants, and a series of backcrossing was performed from 15 different BC_2 plants to BC_3 to BC_6, depending on when self-fertile plants with $2n = 40$ chromosomes were obtained (Singh and Nelson, 2015). Lines evaluated in this study were progeny from seven different BC_2 plants. The production of the BC_2 generation required hormone treatment and seed rescue. However, in later generations of backcrossing, culturing was not usually employed. Progeny from self-fertile plants were grown in the field and bulk harvested if they were homogeneous. Single plants were harvested from heterogeneous lines until progeny rows were phenotypically uniform. Among those included in this research, there were 15 pairs of lines where each pair was derived from the same initial, self-fertile plant. Inbred lines were visually selected for preliminary yield testing, and from those tests, 180 high, medium, and low-yielding lines in maturity groups (MG) II, III, and IV were selected for this research. Lines tested ranged from the F_2 to F_7 generation. The 180 experimental lines were derived from seven different BC_2 parents (06H1-1, 06H1-3, 07H1-7, 07H1-14, 07H1-16, 07H1-25, and 07H1-38). The number of lines from each BC_2 parent was <20 except for 06H1-1 and 06H1-3, and more than half came from 06H1-1 (Table 1).

Field Evaluation

Experimental lines in MG II were divided into two tests: IIA (48 lines) and IIB (47 lines). There was one MG III test (47 experimental lines) and one MG IV test (38 experimental lines). The experimental lines, the recurrent parent Dwight, and public check cultivars were evaluated in the field for 2 yr in 12 to 15 environments, depending on the test (Table 2). The lines in all tests were planted in a randomized complete block design with two replications. Plots at Lincoln, NE, were furrow irrigated every other week in both 2013 and 2014. Plots at Portageville, MO, were flood irrigated on an as-needed basis for both years. Plots at Columbia, MO, were irrigated once via lateral overhead irrigation in June 2013. Conventional tillage and weed management practices were followed at all locations.

The agronomic data collected included flowering date (R1), which was recorded when ~50% of the plants had at least one flower (Fehr et al., 1971); plant maturity date (R8), recorded when ~95% of the pods had reached mature pod color (Fehr et al., 1971); plant lodging, scored at maturity based on a 1-to-5 scale (1 = all plants are erect, 5 = all plants are prostrate); plant height (cm), measured from the soil surface to the top node of the main stem; and seed yield (kg ha^{-1}), recorded at 13% moisture.

In 2013, protein and oil concentrations were measured for all entries from all tests using seeds from Urbana, IL, and a DA 7200 near-infrared analyzer (Perten Instruments). Lines with significantly higher protein concentrations than the recurrent parent, Dwight, were identified in Tests IIA and IV. In 2014, protein and oil concentrations of the experimental lines and checks of only Tests IIA and IV grown at the Illinois locations were measured using near-infrared spectroscopy.

In 2013 and 2014 for Tests IIA and IIB, seed yield, plant height, maturity date, and lodging scores were collected from

Table 1. Total number of *G. tomentella*-derived lines from backcrosses to soybean cultivar Dwight evaluated in each test and the number of lines significantly higher yielding than the recurrent parent from each BC_2 parent.

	BC_2 parent							
Test	06H1-1	06H1-3	07H1-7	07H1-14	07H1-16	07H1-25	07H1-38	Total
IIA	35 (0)†	–‡	9 (0)	–	4 (0)	–	–	48 (0)
IIB	–	44 (2)	–	–	–	1 (0)	2 (0)	47 (2)
III	27 (2)	14 (4)	3 (1)	3 (0)	–	–	–	47 (7)
IV	32 (23)	6 (4)	–	–	–	–	–	38 (27)
Total	94 (25)	64 (10)	12 (1)	3 (0)	4 (0)	1 (0)	2 (0)	180 (36)

† The number of lines significantly higher yielding than the recurrent parent, Dwight, is in parenthesis.

‡ Lines from this BC_2 parent were not included in this test.

Table 2. Locations, planting dates, and planting information for Tests IIA, IIB, III, and IV grown in Nebraska, Missouri, Illinois, and Ohio in 2013 and 2014.

			Planting information			
Test	Date planted	Location	Geographical coordinates of location	No. of rows plot^{-1} planted at each location	Row spacing	Seeding rate
					m	seeds m^{-1}
IIA and IIB	8 May 2013	Wooster, OH	40°48′ N, 81°56′ W	8	0.20 × 6.4	10
	23 May 2014					
	14 May 2013	DeKalb	41°55′ N, 88°45′ W	2	0.76 × 3.6	30
	20 May 2014					
	18 May 2013	Urbana, IL	40°5′ N, 88°12′ W	4	0.76 × 3.1	30
	16 May 2014					
	14 May 2013	Pontiac, IL	40°52′ N, 88°37′ W	2	0.76 × 3.6	30
	7 May 2014					
	18 May 2013	Bellflower, IL	40°20′ N, 88°31′ W	4	0.76 × 3.1	30
	7 May 2014					
	22 May 2013	Lincoln, NE	40°48′ N, 96°40′ W	2	0.76 × 2.9	30
	16 May 2014					
IIA	16 May 2013	Ivesdale, IL	39°56′ N, 88°27′ W	4	0.76 × 3.1	30
	6 May 2014					
III and IV	8 May 2013	Wooster, OH	40°48′ N, 81°56′ W	8	0.20 × 6.4	10
	23 May 2014					
	13 May 2013	Portageville, MO	36°25′ N, 89°41′ W	4	0.76 × 3.7	40
	23 May 2014					
	23 May 2013	Arthur, IL	39°42′ N, 88°28′ W	2	0.76 × 3.6	30
	8 May 2014					
	23 May 2013	Brownstown, IL	38°59′ N, 88°57′ W	3	0.76 × 3.6	30
	7 June 2014					
	18 May 2013	Urbana, IL	40°5′ N, 88°12′ W	4	0.76 × 3.1	30
	6 June 2014					
	29 May 2014	Lincoln, NE	40°48′ N, 96°40′ W	2	0.76 × 2.9	30
III	7 June 2013	Columbia, MO	38°57′ N, 92°20′ W	4	0.76 × 3.7	40
	19 May 2014	Novelty, MO	40°0′ N, 92°12′ W	4	0.76 × 3.7	40
	18 May 2013	Bellflower, IL	40°20′ N, 88°31′ W	4	0.76 × 3.1	30
	7 May 2014					
	23 May 2014	Portageville, MO	36°25′ N, 89°41′ W	4	0.76 × 3.7	40
	6 June 2014	Urbana, IL	40°5′ N, 88°12′ W	4	0.76 × 3.1	30
IV	16 May 2013	Ivesdale, IL	39°56′ N, 88°27′ W	4	0.76 × 3.1	30
	5 June 2014					
	11 June 2013	Columbia, MO	38°57′ N, 92°20′ W	4	0.76 × 3.7	40
	20 May 2014	Novelty, MO	40°0′ N, 92°12′ W	4	0.76 × 3.7	40
	19 May 2014	Portageville, MO	36°25′ N, 89°41′ W	4	0.76 × 3.7	40
	16 June 2014	Urbana, IL	40°5′ N, 88°12′ W	4	0.76 × 3.1	30

Ivesdale (Test IIA only), Urbana, Bellflower, DeKalb, and Pontiac in Illinois, Wooster, OH, and Lincoln (2014 seed yield and maturity dates only). For Tests III and IV in 2013 and 2014, seed yield, plant height, lodging scores, and maturity dates were collected from Ivesdale (Test IV only), Bellflower (Test III only), Urbana, Arthur, Brownstown, IL, Wooster, Lincoln, (2014 only), and Columbia, Novelty, and Portageville, MO. Flowering dates were recorded for all tests grown at Urbana, Bellflower, and Ivesdale in Illinois in 2013 and 2014.

Genotyping-by-Sequencing

Genomic DNA was isolated from fresh trifoliolate leaf tissue of 10 greenhouse-grown plants for each entry and bulked within lines. It was frozen in a −80°C freezer prior to DNA extraction using a modified cetyltrimethylammonium bromide protocol (Mace et al., 2003) and quantified using PicoGreen (Invitrogen).

The GBS libraries were prepared using a protocol modified from Poland et al. (2012). Genomic DNA (~250 ng) from both parents and derived lines was double digested with *HindIII*-HF + *BfaI* and *HindIII*-HF + *HinP1I* at 37°C for 2 h with heat inactivation at 80°C for 20 min. Digested DNA was ligated to two separate adapters using T4 ligase with 1 mM ATP. The first adaptor contains the Illumina forward sequencing primer, one of 96 unique barcodes, and the *HindIII*-HF overhang. The second adaptor contains the Illumina reverse sequencing primer and the overhang for either *BfaI* or *HinP1I*. Ligation reactions were held at 25°C for 2 h, followed by heat inactivation at 65°C for 20 min. Pooled DNA from 96 barcoded libraries was cleaned using a 2:1 ratio of AmpureXP Beads (Beckman Coulter) to DNA solution using a Magnetic Particle Concentrator (Invitrogen) with two washes in 95% ethanol and resuspension in elution buffer (10 mM Tris). Cleaned DNA pools were amplified using Illumina primers in a 2× PhusionHF Master Mix (New England Biolabs) with cycler conditions as follows: 98°C for 30 s, 15 cycles (98°C for 10 s, 68°C for 30 s, and 72°C for 30 s), and 72°C for 5 min. Samples were run on agarose gels to confirm the presence of a genomic smear and cleaned a second time with AMPure beads. Amplified DNA sizes and relative concentrations were assessed using an Agilent Bioanalyzer 2100 and Agilent DNA1000 Kit (Agilent Technologies) and PicoGreen. The two separately digested samples were combined in equimolar concentrations and diluted to 10 nM in library buffer (elution buffer + 0.05% Tween-20) and submitted to the W.M. Keck Center at the University of Illinois for single-end sequencing on the Illumina HiSeq2500. The Keck Center performed an additional quantitative polymerase chain reaction assay on each library to adjust concentrations before sequencing.

Statistical Analysis

Agronomic Data

Flowering and maturity dates, plant height, lodging, and seed yield data were subjected to ANOVA using the Proc Mixed function in SAS (SAS Institute, 2013). Mean separation was conducted using Fisher's protected LSD test in the Proc GLM function in SAS. Normality of residuals was checked using the Proc Univariate function in SAS. The lodging data were log transformed (as the data were not normally distributed) before ANOVA and then back transformed on the original scale for reporting. Years by locations were defined as environments in the ANOVA. The ANOVA for each test were pooled among environments after homogeneity of variance has been checked with the Brown–Forsythe test in SAS. Environments and replications within environments were considered as random effects, whereas genotype was considered as a fixed effect. For each phenotypic trait, a least square mean was obtained across locations and years for each test after modeling out the effect of environments, using a mixed model (Proc mixed function) in SAS.

Marker-Trait Association

The SNPs were called from Illumina fastq files using the TASSEL3 GBS pipeline (http://www.maizegenetics.net/). Those with >95% missing data were discarded. The SNPs were not initially filtered by minor allele frequency, as we expected many *G. tomentella* introgressions to be rare. Imputation of missing data was performed using Beagle 4 (Browning and Browning, 2007) with a window size of 500 SNPs and an overlap of 100 SNPs. Because we could not get the exact Dwight plant that was used as the recurrent parent in the development of the breeding lines, four different sources of Dwight (USDA Soybean Germplasm Collection, Randall Nelson's USDA-ARS breeding program at the University of Illinois, Brian Diers's breeding program at the University of Illinois, and a Dwight DNA sample of unknown source that had been stored in −80°C freezer for 4 yr) were genotyped in this study.

For each SNP, homozygous reference allele, heterozygous, and homozygous alternate allele genotypes were coded as 0, 1, and 2, respectively. Least square means for each trait from PROC MIXED were used for marker-trait association in GAPIT (Lipka et al., 2012) using a mixed linear model with test (IIA, IIB, III, or IV) as a covariate. Markers included all SNPs discovered in this study that were present in at least two derived lines, represented fixed differences between Dwight and *G. tomentella*, and were homozygous or heterozygous for the *G. tomentella* allele at a locus.

RESULTS

Agronomic Data

There were significant differences among experimental lines for all traits measured in each test (Tables 3–6). There were also significant environmental effects for all the traits measured in all tests, as well as significant genotype × environment interaction for flowering date, maturity date, height, lodging score, and yield, except for yield in Test IIB (Tables 3–6). Protein and oil concentrations were only measured in Tests IIA (Table 3) and IV (Table 6) in six environments in Illinois over the 2 yr of the study. All major effects and interactions were significant except for the genotype × environment interactions for Test IIA (Table 3).

The agronomic data showed variations among derived lines for all traits measured for each test (Tables 7–10). The agronomic data presented for each test are for a subset of lines that shows the maximum variation for each trait and also represents derived lines from each BC_2 parent for each

Table 3. Combined ANOVA of agronomic traits from lines derived from backcrossing *G. tomentella* PI 441001 to Dwight in Test IIA across 14 environments in Nebraska, Illinois, and Ohio in 2013 and 2014.

Source of variation†	df	Mean square						
		R1‡	Lodging§	Plant height¶	R8#	Yield	Protein††	Oil‡‡
		d after 31 May			d after 31 May			
G	50	3.1***	0.3***	170.9***	42.1***	465,189***	3.5***	0.9***
E	13	7723.2***	6.7***	11,400.0***	5439.7***	32,861,010***	50.6**	40.0***
R(E)	14	2.9**	0.2***	374.4***	18.2***	2,105,696***	2.4***	0.3NS§§
E × G	650	1.2**	0.1***	28.5***	3.2***	166,029***	0.4NS	0.2NS
Residual	700	0.80	0.04	23.5	1.8	140,276	0.4	0.2

, * Significant at the 0.01 and 0.001 probability levels, respectively.

† G, genotype; E, environment; R, replication.

‡ Flowering date when 50% of plants had at least one flower measured in six environments over 2 yr.

§ Lodging score (1 = plant erect, 5 = prostrate) measured in 12 environments over 2 yr.

¶ Plant height was measured in 12 environments over 2 yr.

Maturity date when 95% of pods were at final color measured in twelve environments over 2 yr

†† Protein concentration on a 13% moisture basis measured in six environments over 2 yr

‡‡ Oil concentration on a 13% moisture basis measured in six environments over 2 yr

§§ NS, not significant.

Table 4. Combined ANOVA of agronomic traits from lines derived from backcrossing *G. tomentella* PI 441001 to Dwight in Test IIB across 12 environments in Nebraska, Illinois, and Ohio in 2013 and 2014.

Source of variation†	df	Mean square				
		R1‡	Lodging§	Plant height¶	R8#	Yield
		d after 31 May			d after 31 May	
G	49	1.8*	0.3***	122.2***	27.9***	417,230***
E	11	10,397.0***	9.2***	14,982.0***	5830.3***	42,919,393***
R(E)	12	2.2*	0.1**	182.6***	31.9***	975,084***
E × G	539	1.2***	0.1***	24.1**	3.4***	172,361NS††
Residual	588	0.7	0.03	18.1	2.2	152,076

*, **, *** Significant at the 0.05, 0.01, and 0.001 probability levels, respectively.

† G, genotype; E, environment; R, replication.

‡ Flowering date when 50% of plants had at least one flower measured in six environments over 2 yr.

§ Lodging score (1 = plant erect, 5 = prostrate) measured in 10 environments over 2 yr.

¶ Plant height was measured in 10 environments over 2 yr.

Maturity date when 95% of pods were at final color measured in 10 environments over 2 yr.

†† NS, not significant.

Table 5. Combined ANOVA of agronomic traits from lines derived from backcrossing *G. tomentella* PI 441001 to Dwight in Test III across 15 environments in Missouri, Illinois, and Ohio in 2013 and Nebraska, Missouri, Illinois, and Ohio in 2014.

Source of variation†	df	Mean square				
		R1‡	Lodging§	Plant height¶	R8#	Yield
		d after 31 May			d after 31 May	
G	49	8.1***	0.3***	887.7***	144.2***	1,435,022***
E	14	10,648.0***	7.9***	16,678.0***	7798.6***	58,378,507***
R(E)	15	1.3NS††	0.1***	193.7***	8.6***	690,497***
E × G	686	1.5*	0.1***	34.5NS	5.9***	208,868***
Residual	735	1.1	0.04	31.9	2.6	130,949

*, *** Significant at the 0.05 and 0.001 probability levels, respectively.

† G, genotype; E, environment; R, replication.

‡ Flowering date when 50% of plants had at least one flower measured in four environments over 2 yr.

§ Lodging score (1 = plant erect, 5 = prostrate).

¶ Plant height was measured in 15 environments over 2 yr.

Maturity date when 95% of pods were at final color measured in 13 environments over 2 yr.

†† NS, not significant.

Table 6. Combined ANOVA of agronomic traits from lines derived from backcrossing *G. tomentella* PI 441001 to Dwight in Test IV across 15 environments in Missouri, Illinois, and Ohio in 2013 and Nebraska, Missouri, Illinois, and Ohio in 2014.

Source of variation†	df	Mean square						
		R1‡	Lodging§	Plant height¶	R8#	Yield	Protein††	Oil‡‡
		d after 31 May			d after 31 May			
G	40	15.2***	1.3***	1,224.3***	134.9***	1,733,622***	4.2***	0.9***
E	14	10,161.0***	6.1***	12,849.0***	5,344.1***	37,841,871***	41.0**	24.7**
R(E)	15	3.6*	0.2***	220.4***	40.8***	789,079***	1.6***	0.5**
E × G	560	2.7***	0.1***	34.4**	5.2***	204,352***	0.4**	0.2*
Residual	600	1.1	0.04	26.7	2.2	109,255	0.3	0.1

*, **, *** Significant at the 0.05, 0.01, and 0.001 probability levels, respectively.

† G, genotype; E, environment; R, replication.

‡Flowering date when 50% of plants had at least one flower measured in four environments over 2 yr.

§ Lodging score (1 = plant erect, 5 = prostrate).

¶ Plant height was measured in 15 environments over 2 yr.

Maturity date when 95% of pods were at final color measured in 13 environments over 2 yr.

†† Protein concentration on a 13% moisture basis measured in five environments over 2 yr.

‡‡ Oil concentration on a 13% moisture basis measured in five environments over 2 yr.

Table 7. Selected *G. tomentella*-derived lines and checks showing variation in agronomic traits from Test IIA averaged over 14 environments in Nebraska, Illinois, and Ohio in 2013 and 2014.

Entry	Pedigree	BC_2 parent	R1†	Plant height‡	Lodging§	R8¶	Yield	Yield#	Protein††	Oil‡‡
			d after 31 May	cm		d after 31 May	kg ha^{-1}	%	—— g kg^{-1} ——	
Dwight	Check		32	86	2.0	108	4220	–	341	181
IA2102	Check		32	86	2.7	107	4543	108	342	181
LD02-4485	Check		31	83	2.0	106	4392	104	319	193
LG09-1013	F_3 Dwight (4) × PI 441001	06H1-1	31	90	2.1	109	4128	98	346	189
LG09-12980	F_3 Dwight (5) × PI 441001	06H1-1	32	83	2.0	109	4135	98	347	178
LG10-12128	F_4 Dwight (4) × PI 441001	06H1-1	32	83	1.6	109	4099	97	333	185
LG10-13322	F_3 Dwight (4) × PI 441001	06H1-1	33	85	1.9	108	4301	102	337	180
LG11-1582	F_6 Dwight (4) × PI 441001	06H1-1	33	82	2.1	113	3643	86	347	180
LG11-2575	F_5 Dwight (5) × PI 441001	06H1-1	32	82	1.6	110	4028	95	342	183
LG11-2725	F_6 Dwight (4) × PI 441001	06H1-1	32	77	1.7	109	4081	97	344	183
LG11-2860	F_6 Dwight (4) × PI 441001	06H1-1	32	84	2.2	111	4269	101	345	180
LG11-11433	F_2 Dwight (6) × PI 441001	06H1-1	31	80	1.7	106	4234	100	335	183
LG11-4428	F_2 Dwight (5) × PI 441001	07H1-7	32	86	2.2	109	4079	97	343	180
LG11-10940	F_2 Dwight (4) × PI 441001	07H1-7	32	84	1.6	109	4069	96	346	183
LG11-11279	F_2 Dwight (5) × PI 441001	07H1-7	32	87	2.6	109	4285	102	357	178
LG11-11291	F_2 Dwight (6) × PI 441001	07H1-7	32	82	1.8	108	4253	101	347	183
LG11-11362	F_2 Dwight (6) × PI 441001	07H1-7	33	80	1.8	110	4012	95	346	182
LG11-11120	F_2 Dwight (5) × PI 441001	07H1-16	32	81	1.8	108	4283	101	339	181
LG11-11387	F_2 Dwight (5) × PI 441001	07H1-16	33	82	1.8	108	4072	96	344	177
LG11-11395	F_2 Dwight (5) × PI 441001	07H1-16	32	81	2.0	108	4181	99	338	181
LSD ($P < 0.05$)			1	3	0.2	1	197		5	3

† Flowering date when 50% of plants had at least one flower measured in six environments over 2 yr.

‡ Plant height was measured in 12 environments over 2 yr.

§ Lodging score (1 = plant erect, 5 = prostrate) measured in 12 environments over 2 yr.

¶ Maturity date measured in 12 environments over 2 yr.

Yield as a percentage of the recurrent parent, Dwight.

†† Protein concentration measured in six environments over 2 yr and reported on a 13% moisture basis.

‡‡ Oil concentration measured in six environments over 2 yr and reported on a 13% moisture basis.

Table 8. Selected *G. tomentella*-derived lines and checks showing variation in agronomic traits from Test IIB averaged over 12 environments in Nebraska, Illinois, and Ohio in 2013 and 2014.

Entry	Pedigree	BC_2 parent	R1†	Plant height‡	Lodging§	R8¶	Yield	Yield#
			d after 31 May	cm		d after 31 May	kg ha^{-1}	%
Dwight	Check		35	85	1.9	110	4134	–
IA2102	Check		34	85	2.4	108	4427	107
LD02-4485	Check		33	85	2.0	108	4412	107
LG10-12179	F_2 Dwight (6) × PI 441001	06H1-3	35	86	2.5	112	4239	103
LG10-12211	F_3 Dwight (4) × PI 441001	06H1-3	34	83	1.8	112	3857	93
LG10-12212	F_3 Dwight (4) × PI 441001	06H1-3	34	82	1.8	111	3876	94
LG10-12298	F_3 Dwight (5) × PI 441001	06H1-3	35	86	2.5	113	4342	105
LG10-12313	F_3 Dwight (5) × PI 441001	06H1-3	36	83	2.2	112	4376	106
LG10-12468	F_3 Dwight (5) × PI 441001	06H1-3	35	86	1.8	111	4374	106
LG10-12579	F_4 Dwight (4) × PI 441001	06H1-3	35	79	1.5	110	4215	102
LG10-13076	F_2 Dwight (6) × PI 441001	06H1-3	35	89	2.6	112	4020	97
LG11-1241	F_4 Dwight (5) × PI 441001	06H1-3	34	78	1.6	109	4132	100
LG11-1266	F_4 Dwight (6) × PI 441001	06H1-3	34	80	1.5	108	4009	97
LG11-1432	F_4 Dwight (5) × PI 441001	07H1-25	35	84	1.9	109	4005	97
LG11-11017	F_2 Dwight (5) × PI 441001	07H1-38	34	80	1.9	111	4213	102
LG11-11018	F_2 Dwight (5) × PI 441001	07H1-38	34	81	1.9	110	4037	98
LSD ($P < 0.05$)			1	3	0.2	1	221	

† Flowering date when 50% of plants had at least one flower measured in six environments over 2 yr.

‡ Plant height was measured in 10 environments over 2 yr.

§ Lodging score (1 = plant erect, 5 = prostrate) measured in 10 environments over 2 yr.

¶ Maturity date measured in ten environments over 2 yr.

Yield as a percentage of the recurrent parent, Dwight.

Table 9. Selected *G. tomentella*-derived lines and checks showing variation in agronomic traits from Test III averaged across 15 environments in Missouri, Illinois, and Ohio in 2013 and Nebraska, Missouri, Illinois, and Ohio in 2014.

Entry	Pedigree	BC_2 parent	R1†	Plant height‡	Lodging§	R8¶	Yield	Yield#
			d after 31 May	cm		d after 31 May	kg ha^{-1}	%
Dwight	Check		35	79	1.8	107	3524	–
IA3023	Check		37	84	1.9	115	4094	116
IA3048	Check		35	81	2.0	113	3932	112
LG09-11994	F_3 Dwight (4) × PI 441001	06H1-1	35	75	1.5	108	3280	93
LG09-12041	F_2 Dwight (5) × PI 441001	06H1-1	36	92	2.2	115	3289	93
LG09-12315	F_3 Dwight (4) × PI 441001	06H1-1	37	79	1.8	110	3322	94
LG10-10821	F_2 Dwight (5) × PI 441001	06H1-1	36	78	1.9	110	3083	87
LG10-10833	F_2 Dwight (5) × PI 441001	06H1-1	36	76	1.7	112	3002	85
LG11-3444	F_6 Dwight (4) × PI 441001	06H1-1	37	87	1.9	115	3248	92
LG11-4311	F_6 Dwight (4) × PI 441001	06H1-1	36	82	1.6	115	3717	105
LG11-4330	F_6 Dwight (4) × PI 441001	06H1-1	36	86	2.3	113	3739	106
LG10-13149	F_3 Dwight (5) × PI 441001	06H1-3	36	81	2.1	115	3731	106
LG11-1499	F_6 Dwight (4) × PI 441001	06H1-3	40	101	2.4	114	3132	89
LG11-1502	F_6 Dwight (4) × PI 441001	06H1-3	38	92	2.2	117	3664	104
LG11-2051	F_7 Dwight (4) × PI 441001	06H1-3	38	84	2.1	115	3798	108
LG11-2052	F_7 Dwight (4) × PI 441001	06H1-3	38	82	2.1	116	3788	107
LG11-2963	F_6 Dwight (4) × PI 441001	06H1-3	36	85	2.3	114	3901	111
LG11-4421	F_2 Dwight (5) × PI 441001	07H1-7	36	79	1.7	110	3273	93
LG11-10931	F_2 Dwight (4) × PI 441001	07H1-7	37	81	2.0	116	3738	106
LG11-11326	F_2 Dwight (5) × PI 441001	07H1-14	37	74	1.6	113	3337	95
LSD ($P < 0.05$)			1	3	0.2	1	183	

† Flowering date when 50% of plants had at least one flower measured in six environments over 2 yr.

‡ Plant height was measured in 15 environments over 2 yr.

§ Lodging score (1 = plant erect, 5 = prostrate).

¶ Maturity date measured in 13 environments over 2 yr.

Yield as a percentage of the recurrent parent, Dwight.

Table 10. Selected *G. tomentella*-derived lines and checks showing variation in agronomic traits from Test IV averaged over 15 environments in Missouri, Illinois, and Ohio in 2013 and Nebraska, Missouri, Illinois, and Ohio in 2014.

Entry	Pedigree	BC_2 parent	R1†	Plant height‡	Lodging§	R8¶	Yield	Yield#	Protein††	Oil‡‡
			d after 31 May	cm		d after 31 May	kg ha^{-1}	%	—— g kg^{-1} ——	
Dwight	Check		35	80	1.9	108	3505	–	343	182
IA4005	Check		37	80	1.7	121	4229	121	335	188
LG04-6000	Check		39	97	2.4	121	4326	123	331	179
LG11-1389	F_6 Dwight (4) × PI 441001	06H1-1	37	84	2.0	116	3878	111	350	180
LG11-1402	F_6 Dwight (4) × PI 441001	06H1-1	36	80	1.4	115	3491	100	335	187
LG11-3187	F_6 Dwight (4) × PI 441001	06H1-1	36	85	2.0	114	3979	114	351	177
LG11-3191	F_6 Dwight (4) × PI 441001	06H1-1	36	81	1.6	116	3747	107	336	185
LG11-3224	F_6 Dwight (4) × PI 441001	06H1-1	38	97	3.2	119	3518	100	350	185
LG11-3233	F_6 Dwight (4) × PI 441001	06H1-1	36	89	2.2	116	3737	107	344	184
LG11-3341	F_6 Dwight (4) × PI 441001	06H1-1	42	105	3.3	119	3217	92	355	174
LG11-3370	F_6 Dwight (4) × PI 441001	06H1-1	37	85	2.0	117	3995	114	350	180
LG11-3375	F_6 Dwight (4) × PI 441001	06H1-1	36	83	1.9	117	3937	112	345	180
LG11-3441	F_6 Dwight (4) × PI 441001	06H1-1	36	86	1.9	117	3922	112	340	183
LG11-3463	F_6 Dwight (4) × PI 441001	06H1-1	36	92	1.8	118	3978	114	340	181
LG11-4258	F_6 Dwight (4) × PI 441001	06H1-1	39	90	2.1	119	3750	107	346	182
LG11-4265	F_6 Dwight (4) × PI 441001	06H1-1	40	89	2.0	120	3627	103	340	180
LG11-4301	F_6 Dwight (4) × PI 441001	06H1-1	38	77	1.5	115	3644	104	330	189
LG09-12682	F_4 Dwight (4) × PI 441001	06H1-3	37	81	1.5	116	3857	110	340	183
LG09-12732	F_4 Dwight (4) × PI 441001	06H1-3	36	78	1.7	115	3721	106	339	181
LG11-1210	F_4 Dwight (5) × PI 441001	06H1-3	37	91	2.4	117	3743	107	335	187
LG11-1463	F_6 Dwight (4) × PI 441001	06H1-3	37	93	3.0	118	3614	103	349	186
LG11-1464	F_6 Dwight (4) × PI 441001	06H1-3	36	97	2.3	117	3805	109	343	181
LG11-1496	F_6 Dwight (4) × PI 441001	06H1-3	39	98	3.5	113	2914	83	365	180
LSD ($P < 0.05$)			1	3	0.2	1	168		4	3

† Flowering date when 50% of plants had at least one flower measured in six environments over 2 yr.

‡ Plant height was measured in 15 environments over 2 yr.

§ Lodging score (1 = plant erect, 5 = prostrate).

¶ Maturity date measured in 13 environments over 2 yr.

Yield as a percentage of the recurrent parent, Dwight.

†† Protein concentration measured in five environments over 2 yr and reported on a 13% moisture basis.

‡‡ Oil concentration measured in five environments over 2 yr and reported on a 13% moisture basis.

test. Lines that yielded significantly more than the recurrent parent, Dwight, were identified in each test except for Test IIA (Table 7). For Tests IIB and III, all of the lines there were significantly higher yielding than Dwight were included in the Tables 8 and 9, respectively, but for Test IV, only 14 of the 27 lines that yielded more than Dwight were presented in Table 10. For all other traits, lines that were significantly different from the recurrent parent were identified in each test (Tables 7–10). There were 36 lines that yielded more than the recurrent parent, Dwight, in Tests IIB, III, and IV (Akpertey, 2015). These higher yielding lines were progeny from three of the seven different BC_2 parents included in this study. There were 127 lines that were equivalent in yield to Dwight and 17 that were lower yielding that Dwight. Lines that were lower yielding than Dwight were also identified in each of the four tests and also represented the three BC_2 plants that produced the highest yielding lines. Only three lines were tested from 07H1-14, and one of those (LG11-11326) was lower yielding than Dwight (3337 vs. 3524 kg ha^{-1}).

SNP Calling

Genotypic data were successfully obtained for 177 of the 180 lines included in this research (Table 1); restriction-ligation reactions likely failed for the remaining three lines due to low DNA quality. A total of 201,304,982 reads were obtained using GBS, out of which 193,225,777 matched perfectly to a barcode and the remnant *HindIII* restriction enzyme cut site, with a mean and median number of reads per entry of 530,796 and 504,729, respectively. We obtained 1,070,347 unique tags that were 64 bp in length and present at least five times in the entire dataset and aligned 998,634 (93%) of these unique tags to the Williams 82 reference genome (Gmax1.01) using Bowtie 2 (Langmead et al., 2009) with the "sensitive" option. These tags were used to identify 13,334 SNPs between *G.*

tomentella PI 441001 and the soybean cultivar Dwight. There were 3297 SNPs that were homozygous for the reference allele in all four sources of Dwight and homozygous or heterozygous for the alternate allele in *G. tomentella*; this dataset was used to calculate the introgression frequencies of *G. tomentella* alleles across different lines and SNPs. Of these, there were 871 SNPs (26%) for which the *G. tomentella* allele was present in at least one of the 177 genotyped derived lines, and 667 (20%) for which the *G. tomentella* allele was present in at least two derived lines. Only SNPs present in at least two derived lines were used for marker-trait association analysis. Overall, 9095 genotype calls (7.7% of the 667 SNP dataset) were heterozygous for the *G. tomentella* allele, and 1191 genotype calls (1.0% of the dataset) were homozygous for the *G. tomentella* allele.

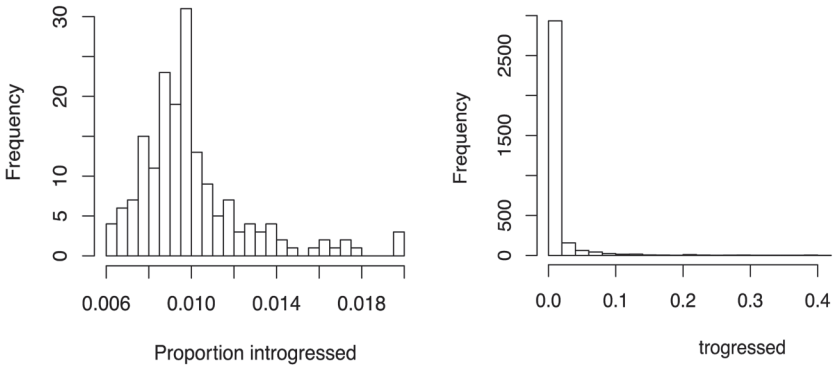

Fig. 1. (A) Proportion *G. tomentella* introgression for each of the 177 genotyped derived lines across 3297 single-nucleotide polymorphisms (SNPs). (B) Proportion *G. tomentella* introgression for each of the 3297 SNPs across the 177 derived lines.

Glycine tomentella Allele Introgression

The 177 genotyped derived lines contained a mean of 1% *G. tomentella* introgression (range 0.6–2.0%, Fig. 1A) across all 3297 SNPs. The amount of *G. tomentella* introgression showed little change with progressive backcrossing (Table 11, Supplemental Fig. S1) and was higher for lines included in later maturing tests (Supplemental Fig. S2). The female parents for the first generation of backcrossing were amphidiploid ($2n$ = 118) plants induced by colchicine treatment, so in all backcrossing generations, both parents had a full complement of *G. max* chromosomes. Because *G. max* chromosomes were always paired during the breeding process, it is not clear how or when introgression from *G. tomentella* occurred in the derived lines. Since the number of backcrosses did not change the amount *G. tomentella* introgression, atypical mechanisms are likely involved. Most of the 3297 *G. tomentella* alleles were present in none of the derived lines (n = 2426, 74%) or present in a single derived line (n = 204, 6%). However, there were 37 *G. tomentella* alleles present in >20% of the derived lines (Fig. 1B), which may represent repetitive sequences unique to *G. tomentella*. *Glycine tomentella*-specific alleles and introgressions are found on every chromosome arm.

Associations between *Glycine tomentella* Alleles and Phenotypic Traits

Significant associations were detected between *G. tomentella* alleles and each of the agronomic traits (q < 0.05) (Table 12). Nearly all *G. tomentella* alleles associated with agronomic traits had frequencies <0.05. For every significant association, the *G. tomentella* allele was associated with increases in height, lodging, and time of flowering, and with decreases in yield (Table 12). Time of flowering was associated with several higher frequency (minor allele frequency = 0.12–0.29) introgressions on chromosome 4. A genomic region on chromosome 12 was associated with lodging and yield, and a region on chromosome 13 was associated with height and yield (Table 12). Additional regions on chromosomes 2, 9, and 10 affected height and lodging individually.

DISCUSSION

Agronomic Data

Growth habit and morphology of all lines were similar to the recurrent parent, Dwight. None of the lines in Test IIA yielded more than the recurrent parent (Table 7), but the yields of 43 entries were not different from that of Dwight (Akpertey, 2015). The highest (LG10-13322) and lowest (LG11-1582) yielding lines were from the same BC_2 parent, 06H1-1. Both lines were similar in all traits measured except for yield (Table 7). The highest yielding lines derived from the other BC_2 plants (LG11-11279 and LG11-11120) were similar to LG10-13322, except LG11-11279 had the highest lodging score in the test (Table 7). There was little variation in time to flowering, with all entries within 1 d of Dwight. The earliest line matured a week before the latest maturing line, and both extremes were different from Dwight (Table 7). The latest maturing line was the lowest yielding line in the test. Only lines derived from 06H1-1 matured earlier than Dwight,

Table 11. Average percentage of *G. tomentella* introgressed into all lines for each backcrossing generation to Dwight based on 3297 single-nucleotide polymorphisms.

No. of backcrosses†	No. of entries	*G. tomentella* average	*G. tomentella* range
		%	
3	102	1.1	0.7–2.0
4	43	1.0	0.6–1.4
5	34	0.9	0.6–1.3
6	1	1.1	N/A

† Number of backcrosses to recurrent parent, Dwight.

Table 12. *Glycine tomentella* specific alleles associated with agronomic traits.

Trait	Chr.	Location	Minor allele frequency	*p*	*Q*	*a*†
		bp				
Height‡	2	45,368,057	0.011	3.55×10^{-5}	0.0066	5.31
Height	2	50,288,124	0.028	5.26×10^{-6}	0.0034	4.15
R1§	4	17,274,362	0.119	5.57×10^{-5}	0.0182	0.88
R1	4	38,901,619	0.285	1.25×10^{-5}	0.0082	0.68
Lodging¶	9	36,197,359	0.023	1.73×10^{-7}	0.0001	0.51
Lodging	10	6,468,493	0.011	0.000703514	0.0393	0.34
Lodging	10	44,277,173	0.006	0.000154125	0.0126	0.72
Lodging	12	400,071	0.017	5.78×10^{-5}	0.0054	0.35
Lodging	12	1,113,993	0.028	2.06×10^{-5}	0.0034	0.35
Seed yield	12	2,974,449	0.020	7.78×10^{-6}	0.0017	–216.18
Height	13	38,510,582	0.017	4.02×10^{-5}	0.0066	5.92
Seed yield	13	38,510,582	0.017	1.10×10^{-6}	0.0004	–292.22

† Additive effect of the alternate (*G. tomentella*) allele.

‡Plant height was measured in fifteen environments over 2 yr

§ Flowering date (d after May 31) when 50% of plants had at least one flower measured in six environments over 2 yr.

¶ Lodging score (1 = plant erect, 5 = prostrate).

and lines from both 06H1-1 and 07H1-7 matured later than Dwight. Only one line was significantly taller than Dwight (LG09-1013 from 06H1-1) (Table 7). The difference between the shortest and tallest lines was only 13 cm. Fourteen lines derived from all three BC_2 plants lodged significantly less than Dwight, and only LG11-11279 from 07H1-7 lodged more (Table 7).

Differences among the derived lines for protein and oil concentrations were observed for Test IIA (Table 7). LG11-11279 (from 07H1-7) had the highest concentration of protein (357 g kg^{-1}), which was greater than that of Dwight (341 g kg^{-1}). LG11-11279 was similar to Dwight for all other traits, except LG11-11279 lodged more. Three additional lines, including two derived from 06H1-1 (LG09-12980 and LG11-1582), had higher protein concentrations than Dwight. Five lines, all from 06H1-1, had lower concentrations than Dwight (Table 7). Seven lines, all from 06H1-1, had higher oil concentrations than Dwight (Akpertey, 2015), and only LG11-11387 from 07H1-16 had a lower oil concentration than Dwight (Table 7). The four lines with the highest protein concentrations all had oil concentrations that were similar to Dwight's (Table 7). Typically in soybean, an increase in protein results in a decrease in oil and seed yield (Wilcox and Cavins, 1995; Cober and Voldeng, 2000).

In Test IIB, LG10-12468 was the same as Dwight for all traits, except that it yielded 240 kg ha^{-1} more (Table 8). LG10-12313 (4376 kg ha^{-1}) had the same yield as LG10-12468 (4374 kg ha^{-1}) but was 1 d later in maturity and had slightly higher lodging score. Both of the lines were derived from 06H1-3 and were comparable with the public check cultivar IA2102 (4427 kg ha^{-1}) (Table 8). LG10-12211 and LG10-12212, also derived from 06H1-3, were lower yielding than Dwight but were similar to Dwight for most traits (Table 8). The range in height in this test was 10 cm, with LG10-13076 being taller and five lines shorter than Dwight (Table 8). Only one line was earlier than Dwight, and five lines were 2 or 3 d later than Dwight. The two highest and two lowest yielding lines were the same maturity and similar to Dwight in maturity (Table 8). Three lines lodged less than and three lines lodged more than Dwight, but none had greater lodging scores than IA2102. In addition to the lines from 06H1-3, one line (LG11-1432) from the BC_2 parent 07H1-25 and two lines (LG11-11017 and LG11-11018) from the BC_2 parent 07H1-38 were evaluated in Test IIB. All of these lines were similar to Dwight for all traits measured (Table 8).

As expected in Test III, all but LG09-11994 matured later than Dwight (Table 9), but only 14 lines flowered later than Dwight (Akpertey, 2015). There was a 6-d range in R1 and a 9-d range in R8. There were six lines that were higher yielding than Dwight and ranged from 6 to 9 d later in maturity. LG11-2963, derived from 06H1-3, had the highest yield (3901 kg ha^{-1}) (Table 9). This was lower yielding than the check cultivar IA3023 (4094 kg ha^{-1}) but was similar for other traits measured, except that it lodged more (Table 9). From the same BC_2 parent, LG11-2051, LG11-2052, and LG10-13149 were also higher yielding than Dwight (Table 9). The lowest yielding line from 06H1-3 was LG11-1499. It flowered the latest (July 10), was the tallest (101 cm), and lodged the most (2.4) of all entries and matured 7 d later than Dwight (Table 9). LG11-4330 and LG11-4311 from 06H1-1 and LG11-10931 from 07H1-7 were also higher yielding than Dwight (Table 9). In addition to LG11-1499 noted above, there were eight lines representing 06H1-1, 06H1-7, and 07H1-14 that were lower yielding than Dwight (Table 9).

All lines evaluated in Test IV were later maturing than Dwight (Akpertey, 2015); however, all but one were 2 to 8 d earlier than the MG IV checks (Table 10). Twenty lines flowered later than Dwight, but only one (LG11-3341) was later than the latest MG IV check. Twenty-seven lines yielded more than Dwight but matured 6 to 11 d later (Akpertey, 2015). These highest yielding lines came from both BC_2 plants represented in the test.

The highest yielding line, LG11-3370 from 06H1-1, yielded 490 kg ha^{-1} more and flowered and matured 2 and 9 d later than Dwight, respectively (Table 10). It was slightly taller and had a greater protein concentration but was similar in oil concentration to Dwight (Table 10). From the same BC_2 parent, four lines—LG11-3187, LG11-3375, LG11-3441, and LG11-3463—yielded >400 kg ha^{-1} more than Dwight. All four lines flowered the same day as Dwight but ranged from 6 to 10 d later in maturity (Table 10). Three of the lines were 5 to 12 cm taller but were similar in lodging to Dwight. The highest yielding line (LG09-12682) from the BC_2 parent 06H1-3 matured 8 d later and yielded 352 g kg^{-1} more than Dwight (Table 10). It flowered 2 d later and lodged less but had no differences in seed composition compared with Dwight.

There were only two lines that yielded less than Dwight from Test IV, one derived from each of the BC_2 plants represented. The lowest yielding line, LG11-1496, came from the BC_2 parent, 06H1-3. It flowered and matured 4 and 5 d later, respectively, and yielded nearly 600 kg ha^{-1} less than Dwight (Table 10). However, it had the highest protein concentration of 365 g kg^{-1} (22 g kg^{-1} greater than Dwight) but the same oil concentration as Dwight. It was 18 cm taller than Dwight and lodged the most (3.5) in the entire test (Table 10). The lowest yielding line from 06H1-1 was LG11-3341 (3217 kg ha^{-1}). It flowered 1 wk later and matured 11 d later than Dwight (Table 10). At 105 cm, it was the tallest line in the test and lodged more (3.3) than Dwight (1.9) (Table 10). It had a protein concentration of 355 g kg^{-1}, which was higher than that of Dwight (343 g kg^{-1}) and the other check cultivars evaluated in this test, and was 8 g kg^{-1} lower in oil than Dwight (Table 10).

Genotypic Data

Glycine tomentella Allele Introgression

We identified 3297 SNPs between Dwight and *G. tomentella* accession PI 441001, of which 26% showed evidence of *G. tomentella* introgression in at least one derived line. This has significant implications for soybean breeding, as it shows that it is possible to access large segments of *G. tomentella* genomes through hybridization with *G. max*.

There were several regions of 10 Mbp or larger with no *G. tomentella*-specific alleles that could be used to detect introgression (Fig. 2). These were mostly in centromeric or pericentromeric regions of the genome, in which marker density is typically much lower in intraspecific crosses. However, higher divergence between *G. max* and *G. tomentella* could also be preventing read alignment in some of these regions. Surprisingly, introgressed *G. tomentella*-specific alleles were closely linked with *G. tomentella*-specific alleles that were not introgressed (Fig. 2), in a pattern not entirely consistent with chromosomal crossover.

Association between SNPs and Phenotypic Traits

Significant associations were detected between *G. tomentella*-specific alleles and each phenotypic trait measured except time of maturity, but most trait-associated SNPs were present at very low frequencies (minor allele frequency < 0.05) and await confirmation in biparental populations. The use of test as a covariate in the GAPIT mixed linear model probably prevented the detection of maturity-associated SNPs. Even though many of the 177 derived lines yielded more than their recurrent parent, we did not detect any *G. tomentella* alleles associated with significant yield increase. There are several possible explanations for these observations. First, individual *G. tomentella* introgressions associated with yield increase might be very rare; we did not test any of the 204 SNPs that were present in a single derived line. Second, rare recombination events might be required to break linkage between yield-increasing and

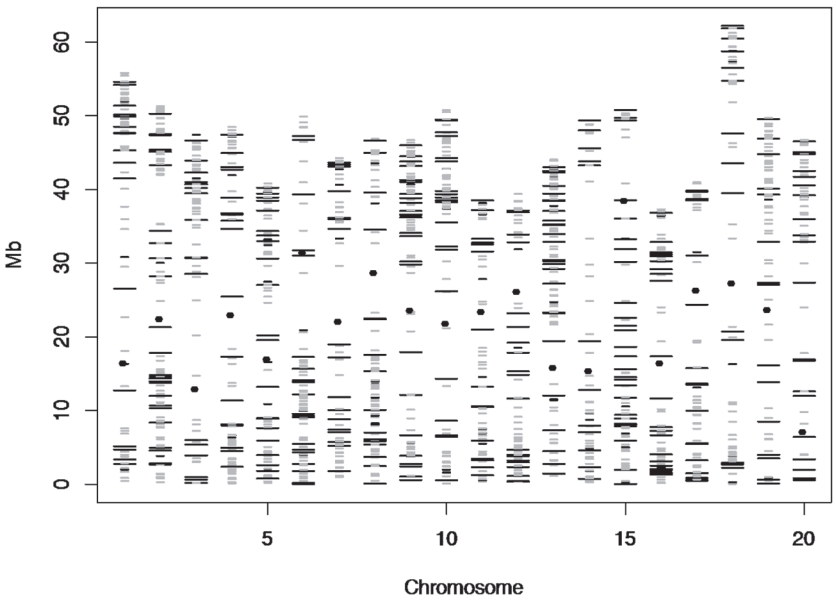

Fig. 2. Genome-wide distribution and introgression frequency of *G. tomentella*-specific alleles (n = 3297) across 177 derived lines. Grey lines (n = 2426) represent *G. tomentella*-specific alleles not present in any derived lines, and black lines (n = 871) represent *G. tomentella* alleles present in at least one derived line. Wide black lines (n = 677) represent the set of *G. tomentella* alleles present in at least two derived lines that were used for marker-trait association. Centromeres are indicated by black circles.

yield-decreasing *G. tomentella* alleles. Third, since we discarded GBS reads that did not align to the Williams 82 reference genome, we may be missing introgressions in regions of high divergence between *G. tomentella* and *G. max*. Construction of biparental populations using high-yielding *G. tomentella*-derived lines could help distinguish among these hypotheses.

This is the first report of field evaluation and GBS of experimental soybean lines derived from the tertiary gene pool species *G. tomentella*. A major achievement of this research is the phenomenal yield increases of nearly 500 kg ha^{-1} in the best line and between 200 and 350 kg ha^{-1} in progenies derived from two additional BC_2 plants. In producing the BC_2 generation, none of the *G. tomentella* chromosomes are paired so each BC_2 has one copy of approximately half of *G. tomentella* chromosomes. Because of the billions of possible combinations, each BC_2 plant is likely to have a different complement of *G. tomentella* chromosomes. High-yielding lines from different BC_2 plants could have a different genetic explanation for these large yield increases. To get such high yield increases from soybean × soybean crosses would be notable, but to obtain that from backcrosses with the perennial species *G. tomentella* is totally unexpected.

Conflict of Interest

The authors declare that there is no conflict of interest.

Supplemental Material Available

Supplemental Fig. S1. Box plots showing the range of *G. tomentella* introgression for lines in different backcross generations.

Supplemental Fig. S2. Box plots showing the range of range of *G. tomentella* introgression for lines in each of the four tests.

Acknowledgments

We gratefully acknowledge that partial funding for this research was provided by Project 1420532-5650 of the United Soybean Board and from the North Central Soybean Research Program.

References

Akpertey, A. 2015. Genetic introgression from *Glycine tomentella* to soybean to increase seed yield. Ph.D. diss., Univ. of Illinois, Urbana-Champaign.

Akpertey, A., M. Belaffif, G.L. Graef, M.A.R. Mian, J.G. Shannon, P.B. Cregan, et al. 2014. Effects of selective genetic introgression from wild soybean to soybean. Crop Sci. 54:2683–2695. doi:10.2135/cropsci2014.03.0189

Amante-Bordeos, A., L.A. Sitch, R. Nelson, R.D. Dalmacio, N.P. Oliva, and H. Aswidinnoor. 1992. Transfer of bacterial blight and blast resistance from the tetraploid wild rice *Oryza minuta* to cultivated rice, *Oryza sativa*. Theor. Appl. Genet. 84:345–354.

Badami, P.S., N. Mallikarjuna, and J.P. Moss. 1997. Interspecific hybridization between *Cicer arietinum* and *C. pinnatifidum*. Plant Breed. 116:393–395. doi:10.1111/j.1439-0523.1997.tb01019.x

Brar, D. 2005. Broadening the gene pool and exploiting heterosis in cultivated rice. In: K. Toriyama, K.L. Heong, and B. Hardy, editors, Rice is life: Scientific perspectives for the 21st century. Proceedings of the World Rice Research Conference, Tokyo and Tsukuba, Japan. 4–7 Nov. 2004. IRRI, Los Baños, the Philippines, and Japan Int. Res. Ctr. Agric. Sci. (JIRCAS), Tsukuba, Japan. p. 157–160.

Brar, D.S., R. Dalmacio, R. Elloran, R. Aggarwal, R. Angeles, and G.S. Khush. 1996. Gene transfer and molecular characterization of introgression from wild *Oryza* species into rice. In: G.S. Khush, editor, Rice genetics III. IRRI, Manila. p. 477–485.

Broué, P., J. Douglass, J.P. Grace, and D.R. Marshall. 1982. Interspecific hybridisation of soybeans and perennial *Glycine* species indigenous to Australia via embryo culture. Euphytica 31:715–724. doi:10.1007/BF00039210

Browning, S.R., and B.L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81:1084–1097. doi:10.1086/521987

Brucker, E.A. 2004. The effect of the soybean gene *rhg1* on reproduction of *Heterodera glycines* in the field and greenhouse and associated effects on agronomic traits. Ph.D. diss., Univ. of Illinois, Urbana-Champaign.

Cheng, S.H., and H.H. Hadley. 1983. Studies in polyploidy in soybeans: A simple and effective colchicine technique of chromosome doubling for soybean (*Glycine max* (L.) Merr.) and its wild relatives. Soybean Genet. Newsl. 10:23–24.

Chetelat, R.T., R.A. Pertuzé, L. Faúndez, E.B. Graham, and C.M. Jones. 2009. Distribution, ecology and reproductive biology of wild tomatoes and related nightshades from the Atacama Desert region of northern Chile. Euphytica 167:77–93. doi:10.1007/s10681-008-9863-6

Cober, E.R., and H.D. Voldeng. 2000. Developing high-protein, high-yield soybean populations and lines. Crop Sci. 40:39–42. doi:10.2135/cropsci2000.40139x

Concibido, V.C., B. La Vallee, P. Mclaird, N. Pineda, J. Meyer, L. Hummel, et al. 2003. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor. Appl. Genet. 106:575–582. doi:10.1007/s00122-002-1071-5

Coyne, D.P., and M.L. Schuster. 1974. Breeding and genetic studies of tolerance to several bean (*Phaseolus vulgaris* L.) bacterial pathogens. Euphytica 23:651–656. doi:10.1007/BF00022486

Doyle, J.J., J.L. Doyle, J.T. Rauscher, and A.H.D. Brown. 2004. Evolution of the perennial soybean polyploid complex (*Glycine* subgenus *Glycine*): A study of contrasts. Biol. J. Linn. Soc. Lond. 82:583–597. doi:10.1111/j.1095-8312.2004.00343.x

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. doi:10.1371/journal.pone.0019379

Fehr, W.R., C.E. Caviness, D.T. Burmood, and J.S. Pennington. 1971. Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. Crop Sci. 11:929–933. doi:10.2135/cropsci1971.0011183X001100060051x

Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. 34:1143–1151. doi:10.2135/cropsci1994.0011183X003400050001x

Grant, J.E., R. Pullen, A.H.D. Brown, J.P. Grace, and P.M. Gresshoff. 1986. Cytogenetic affinity between the new species *Glycine argyrea* and its congeners. J. Hered. 77:423–426. doi:10.1093/oxfordjournals.jhered.a110274

Gur, A., and D. Zamir. 2004. Unused natural variation can lift yield barriers in plant breeding. PLoS Biol. 2:e245. doi:10.1371/journal.pbio.0020245

Harlan, J.R., and J.M.J. de Wet. 1971. Toward a rational classification of cultivated plants. Taxon 20:509–517. doi:10.2307/1218252

Hartman, G.L., M.E. Gardner, T. Hymowitz, and G.C. Naidoo. 2000. Evaluation of perennial *Glycine* species for resistance to soybean fungal pathogens that cause Sclerotinia stem rot and sudden death syndrome. Crop Sci. 40:545–549. doi:10.2135/cropsci2000.402545x

Hartman, G.L., T.C. Wang, and T. Hymowitz. 1992. Sources of resistance to soybean rust in perennial *Glycine* species. Plant Dis. 76:396–399. doi:10.1094/PD-76-0396

Hartwig, E.E. 1973. Varietal development. In: B.E. Caldwell, editor, Soybeans: Improvement, production, and uses. ASA, Madison, WI. p. 187–210.

Henry, M., V. Rosas, and A. Mujeeb-Kazi. 1996. Utilization of alien *Triticeae* germplasm resistant to barley yellow dwarf virus for wheat improvement. In: 88th Annual Meeting of American Society of Agronomy Abstracts. ASA, Madison, WI. p. 92.

Honma, S. 1956. A bean interspecific hybrid. J. Hered. 47:217–220. doi:10.1093/oxfordjournals.jhered.a106642

Hymowitz, T., and R.J. Singh. 1984. A soybean × *Glycine tomentella* hybrid: Progress and problems. Soybean Genet. Newsl. 11:90.

Jaiswal, H.K., B.D. Singh, and R.M. Singh. 1987. Improvement of chickpea (*Cicer arietinum*) through introgression of genes from *Cicer reticulatum*. Indian J. Agric. Sci. 57:880–883.

Jena, K.K., and G.S. Khush. 1990. Introgression of genes from *Oryza officinalis* Well ex Watt to cultivated rice, *O. sativa* L. Theor. Appl. Genet. 80:737–745. doi:10.1007/BF00224186

Kelly, J.D. 2004. Advances in common bean improvement: Some case histories with broader applications. Acta Hortic. 637:99–122. doi:10.17660/ActaHortic.2004.637.11

Ladizinsky, G., C.A. Newell, and T. Hymowitz. 1979. Wide crosses in soybeans: Prospects and limitations. Euphytica 28:421–423. doi:10.1007/BF00056600

Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25. doi:10.1186/gb-2009-10-3-r25

Lee, J.-D., S.-K. Park, and Y.-H. Hwang. 2004. Evaluation of agronomic characteristics for development of small seed-size soybeans by interspecific crosses. Korean J. Breed. 36:207–213.

Li, D., T.W. Pfeiffer, and P.L. Cornelius. 2008. Soybean QTL for yield and yield components associated with *Glycine soja* alleles. Crop Sci. 48:571–581. doi:10.2135/cropsci2007.06.0361

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, et al. 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics 28:2397–2399. doi:10.1093/bioinformatics/bts444

Mace, E.S., K.K. Buhariwalla, H.K. Buhariwalla, and J.H. Crouch. 2003. A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol. Biol. Report. 21:459–460. doi:10.1007/BF02772596

Mallikarjuna, N. 1999. Ovule and embryo culture to obtain hybrids from interspecific incompatible pollinations in chickpea. Euphytica 110:1–6. doi:10.1023/A:1003621908663

Mallikarjuna, N., and D.R. Jadhav. 2008. Techniques to produce hybrid between *Cicer arietinum* L. × *C. pinnatifidum* Jaub. Indian J. Genet. Plant Breed. 68:398–405.

Mallikarjuna, N., D. Jadhav, and P. Reddy. 2006. Introgression of *Cajanus platycarpus* genome into cultivated pigeon pea, *C. cajan*. Euphytica 149:161–167. doi:10.1007/s10681-005-9063-6

Mallikarjuna, N., D. Jadhav, M.V. Reddy, and U. Dutta-Tawar. 2005. Introgression of phytophthora blight disease resistance from *Cajanus platycarpus* into short duration pigeonpea [*Cajanus cajan* (L.) Millsp.]. Indian J. Genet. Plant Breed. 65:261–263.

Mujeeb-Kazi, A., R.L. Villareal, L.A. Gilchrist, and S. Rajaram. 1996. Registration of five wheat germplasm lines resistant to Helminthosporium leaf blight. Crop Sci. 36:216–217. doi:10.2135/cropsci1996.0011183X003600010054x

Newell, C.A., X. Delannay, and M.E. Edge. 1987. Interspecific hybrids between the soybean and wild perennial relatives. J. Hered. 78:301–306. doi:10.1093/oxfordjournals.jhered.a110392

Newell, C.A., and T. Hymowitz. 1982. Successful wide hybridization between the soybean and a wild perennial relative, *G. tomentella* Hayata. Crop Sci. 22:1062–1065. doi:10.2135/cropsci1982.0011183X002200050041x

Nickell, C.D., G.R. Noel, T.R. Cary, and D.J. Thomas. 1998. Registration of 'Dwight' soybean. Crop Sci. 38:1398. doi:10.2135/cropsci1998.0011183X003800050046x

Patzoldt, M.E., R.K. Tyagi, T. Hymowitz, M.R. Miles, G.L. Hartman, and R.D. Frederick. 2007. Soybean rust resistance derived from *Glycine tomentella* in amphiploid hybrid lines. Crop Sci. 47:158–161. doi:10.2135/cropsci2006.05.0328

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253. doi:10.1371/journal.pone.0032253

Prescott-Allen, C., and R. Prescott-Allen. 1986. The first resource: Wild species in the North American economy. Yale Univ. Press, New Haven, CT.

Rick, C.M., and R.T. Chetelat. 1995. Utilization of related wild species for tomato improvement. Acta Hortic. 412:21–38. doi:10.17660/ActaHortic.1995.412.1

SAS Institute. 2013. The SAS system for Windows. Release 9.4. SAS Inst., Cary, NC.

Schoen, D.J., J.J. Burdon, and A.H.D. Brown. 1992. Resistance of *Glycine tomentella* to soybean leaf rust *Phakopsora pachyrhizi* in relation to ploidy level and geographical distribution. Theor. Appl. Genet. 83:827–832.

Singh, A., N.P. Singh, and A.N. Asthana. 1999. Genetic potential of wide crosses in chickpea. Legume Res. 22:19–25.

Singh, K.B., and B. Ocampo. 1993. Interspecific hybridization in annual *Cicer* species. J. Genet. Breed. 47:199–204.

Singh, K.B., and B. Ocampo. 1997. Exploitation of wild *Cicer* species for yield improvement in chickpea. Theor. Appl. Genet. 95:418–423. doi:10.1007/s001220050578

Singh, R.J. 2003. Plant cytogenetics. 2nd ed. CRC Press, Boca Raton, FL.

Singh, R.J. 2010. Methods for producing fertile crosses between wild and domestic soybean species. US Patent 7842850. Date issued: 30 November.

Singh, R.J., K.P. Kollipara, and T. Hymowitz. 1987. Intersubgeneric hybridization of soybeans with a wild perennial species, *Glycine clandestina* Wendl. Theor. Appl. Genet. 74:391–396. doi:10.1007/BF00274723

Singh, R.J., and R.L. Nelson. 2015. Intersubgeneric hybridization between *Glycine max* and *G. tomentella*: Production of F_1, amphidiploid, BC_1, BC_2, BC_3, and fertile soybean plants. Theor. Appl. Genet. 128:1117–1136.

Singh, R.P., M. Henry, J. Huerta-Espino, A. Mujeeb-Kazi, R.J. Pena, and M. Khairallah. 2001. Recombined Thinopyrum chromosome segments in wheat carrying genes *Lr19* and *Bdv2*. In: J. Reeves, A. McNab, and S. Rajaram, editors, Warren Kronstad Symposium, Obregon, Sonora, Mexico. 15–17 Mar. 2001. CIMMYT, El Batán, Mexico. p. 142–144

Singh, S., R.K. Gumber, N. Joshi, and K. Singh. 2005. Introgression from wild *Cicer reticulatum* to cultivated chickpea for productivity and disease resistance. Plant Breed. 124:477–480. doi:10.1111/j.1439-0523.2005.01146.x

Verma, M.M., Ravi, and J.S. Sandhu. 1995. Characterization of the interspecific cross *Cicer arietinum* LH. × *C. judaicum* (Boiss). Plant Breed. 114:549–551. doi:10.1111/j.1439-0523.1995.tb00856.x

Wang, D., G.L. Graef, A.M. Procopiuk, and B.W. Diers. 2004. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theor. Appl. Genet. 108:458–467. doi:10.1007/s00122-003-1449-z

Wilcox, J.R., and J.F. Cavins. 1995. Backcrossing high seed protein to a soybean cultivar. Crop Sci. 35:1036–1041. doi:10.2135/cropsci1995.0011183X003500040019x

Zheng, C., P. Chen, T. Hymowitz, S. Wickizer, and R. Gergerich. 2005. Evaluation of *Glycine* species for resistance to *Bean pod mottle virus*. Crop Prot. 24:49–56. doi:10.1016/j.cropro.2004.06.013