Genetic Variability for Soybean Seed Composition in Crosses Between High- and Low-Protein Parents[1, 2]

S. Rodríguez de Cianzio and W. R. Fehr[3]

ABSTRACT

The genetic variability for seed composition between high and low protein soybean (*Glycine max* (L.) Merr.) parents was evaluated. Wells and Woodworth were the low-protein cultivars, and PI 153.269 and Pando were plant introductions with high protein content. Wells was crossed to PI 153.269 and two backcrosses to Wells were obtained. Woodworth was crossed to Pando, and two backcrosses were made to Woodworth. F_2-, BC_1F_1-, and BC_2F_1-derived lines and the parents were evaluated in two environments. Protein and oil percentages were measured by infrared analysis.

No line from either set of crosses had protein concentrations as high as those of the high-protein donor parent. The F_2-derived line with the highest protein from the cross Wells × PI 153.269 had 48.0%, whereas PI 153.279 had 49.4%. In the Woodworth × Pando cross, the best F_2-derived line had 46.0%, whereas Pando had 49.2%. Mean protein percentage and genetic variance of the populations decreased with each backcross. Broad sense heritabilities on an entry mean basis ranged from 0.76 to 0.89. Discussed here are the implications of these findings in relationship to a breeding method for protein improvement.

INTRODUCTION

One procedure for augmenting protein production per hectare is to increase the protein percentage of high yielding cultivars. Widely grown soybean (*Glycine max* (L.) Merr.) cultivars have about 40% protein compared with about 50% for a few plant introductions.

Several studies on the inheritance of protein content in soybeans have been reported. Williams (10), in 15 crosses between the domestic soybean (*G. hispida* (Moench) Maxim) and the wild soybean (*G. ussuriensis* Regel and Maack) found high protein content to be dominant over low protein. The protein content of the F_1 was very close to that of the wild (high) parent, and some F_2 segregates had even higher protein content. No protein values were mentioned in that publication. Weber (9) crossed the *G. max* cultivar Dunfield (42.34% protein, 19.70% oil) with *G. ussuriensis*, strain P.I. 65.549 (52.04% protein, 7.68% oil) and studied the F_1, F_2, F_3 and backcrosses to both parents. He found that the highest protein segregate from the F_2 population exceeded the mean of the high-protein

parent, but the lowest protein segregate failed to reach the mean of the low-protein parent. The F_1 exhibited partial dominance for genes determining high protein content. The backcross data confirmed the findings reported for the F_1 generation. Viljoen (8) in South Africa, reported studies of protein and oil inheritance from three soybean crosses of the *G. max* species. He crossed P. 449 (37.8% protein, 21.4% oil) with Brownie (48.5% protein, 15.6% oil) and Dixie (44.4% protein, 17.1% oil), and P. 258 (44.6% protein, 17.9% oil) with Dixie. The oil and protein contents of the F_1 plants of the first two crosses fell midway between the values of the two respective parents. Their F_2's did not show evidence of transgressive segregation and he concluded that the inheritance of oil and protein for these crosses could be explained on the basis of incomplete dominance of quantitative factors. In the cross P. 258 × Dixie, however, the F_1 plants had a higher oil and a lower protein content than either parent. The F_2 generation of this cross showed evidence of transgressive segregation for both traits. He concluded that the genes responsible for high oil are dominant to the genes for low oil in the ratio of 9:7. Ishige (3) conducted a diallel analysis to study the inheritance of protein content in soybeans, *G. max* (L.) Merr. He used Hayman's model to analyze protein content and protein yield in F_1 and F_2 plants and found that the additive genetic effects for protein content were statistically significant in the generations considered.

Although the papers mentioned presented data on the variability for protein content for each of the generations under study, there is no available information concerning the segregation in backcross generations from crosses between high- and low-protein soybean (*G. max* (L.) Merr.) parents. The purpose of our research was to study the segregation for protein, oil, and protein + oil obtained from crosses between high- and low-protein soybean cultivars in the F_2 and the first and second backcross generations obtained without selection, with the low-protein cultivar as recurrent parent.

MATERIALS AND METHODS

Wells, a high-yielding soybean cultivar of Group II maturity, was crossed with PI 153.269, and two backcrosses were obtained with Wells as the recurrent parent. Woodworth, a high-yielding cultivar of Group III maturity, was crossed with Pando, and two backcrosses were obtained with Woodworth as the recurrent parent. The initial crosses were made at Ames in 1975, and the two backcrosses were obtained at the Isabela Substation of the Puerto Rico Agricultural Experiment Station in December 1975 and March 1976. The first backcross was obtained by crossing F_1 plants with the recurrent parent. For the second backcross, four BC_2 seeds were obtained from each of 64 BC_1 plants of the Wells cross and 52

BC_1 plants of the Woodworth cross. Each BC_2 seed was identified so that it could be traced to a known BC_1 plant. BC_2F_1 plants were grown at Ames in 1976.

Individual F_2 plants derived from each cross were obtained at the Isabela Substation in June 1976. In November 1976, seed from F_2, BC_1F_1, BC_2F_1, and parent plants were increased in Puerto Rico. One random BC_2 plant from each BC_1F_1 family was harvested.

Tests were conducted in 1977 at the Agronomy Research Center and the Burkey Farm near Ames, Iowa. Planting dates were 2 weeks apart at the two locations to obtain different environments during seed development. A separate randomized complete block design with two replications at each location was used for the two crosses. For the Wells × PI 153.269 cross, 73 F_2-derived lines in F_4, 64 Wells BC_1F_1-derived lines in F_3, 32 Wells BC_2F_1-derived lines in F_3, 10 entries of Wells, and eight entries of PI 153.269 were evaluated. For the Woodworth × Pando cross, 73 F_2-derived lines in F_4, 52 Woodworth BC_1F_1-derived lines in F_3, 27 Woodworth BC_2F_1-derived lines in F_3, 10 entries of Woodworth, and 10 entries of Pando were evaluated. A plot consisted of a single row 1.5 m long; rows were 68 cm apart. Each row was harvested in bulk, and protein and oil contents were measured from a 60-g sample by infrared analysis at the USDA Northern Regional Research Center, Peoria, Illinois. All percentages are reported on a dry-weight basis.

Entries and environments were considered random effects in the analysis of variance. Confidence intervals at the 5% probability level were computed for the parents, F_2, and backcrosses as described by Snedecor and Cochran (6). Variance components obtained by the method suggested by Fisher and described by Crump (1) from the analysis of variance were used to estimate heritabilities.

RESULTS AND DISCUSSION

No F_2-derived line had protein percentages as high as those of the high-protein donor parent (table 1). The F_2-derived line with the highest protein in the Wells × PI 153.269 cross had 48.0%, compared with 49.4% for PI 153.269. The highest F_2-derived line in the Woodworth × Pando cross had 46.0%, whereas Pando had 49.2%. Only 1% of the F_2-derived lines in each cross were within the confidence interval of Wells or Woodworth.

No F_2-derived lines had oil percentages as low as those for the donor parents (table 2). Only 1% of the F_2-derived lines had oil percentages within the confidence interval of the recurrent parent. The range among F_2-derived lines for oil was smaller than the range for protein because the parental range for oil was smaller.

The frequency distribution of F_2-derived lines for protein + oil per-

TABLE 1.—*Performance of parents and their progeny for protein content averaged across two environments*

Entry	Protein		σ_g^{2}[2]	H[3]	Protein		Within C.I.[1] of	
	$\bar{x}$	C.I.[1]			Lowest line	Highest line	Recurrent parent	Donor parent
	%				%		% of lines	
Wells	41.2	±0.25	—	—				
PI 153.269	49.4	±0.28	—	—				
F_2	44.5	±0.09	1.339	0.89	41.4	48.0	1	0
Wells BC₁	42.8	±0.10	0.572	0.82	41.0	44.6	6	0
Wells BC₂	42.0	±0.14	0.415	0.76	40.7	44.2	34	0
Woodworth	39.6	±0.16	—	—				
Pando	49.2	±0.24	—	—				
F_2	43.2	±0.09	0.921	0.84	39.4	46.0	1	0
Woodworth BC₁	41.4	±0.10	0.810	0.88	39.9	44.0	0	0
Woodworth BC₂	40.8	±0.14	0.455	0.81	39.5	42.2	11	0

[1] Confidence interval at the 5% probability level.
[2] Estimate of the genetic variance component.
[3] Broad-sense heritability on an entry mean basis.

TABLE 2.—*Performance of parents and their progeny for oil content averaged across two environments.*

Entry	Oil		σ_g^{2}[2]	H[3]	Oil		Within C.I.[1] of	
	$\bar{x}$	C.I.[1]			Lowest line	Highest line	Recurrent parent	Donor parent
	%				%		% of lines	
Wells	20.7	±0.23	—	—				
PI 153.269	14.2	±0.26	—	—				
F_2	18.2	±0.08	0.471	0.80	16.1	20.5	1	0
Wells BC₁	19.4	±0.09	0.257	0.69	18.2	21.1	2	0
Wells BC₂	20.2	±0.13	0.237	0.77	19.0	21.4	38	0
Woodworth	20.3	±0.14	—	—				
Pando	14.5	±0.20	—	—				
F_2	18.0	±0.07	0.519	0.82	16.2	20.3	1	0
Woodworth BC₁	19.0	±0.09	0.244	0.74	17.9	20.2	4	0
Woodworth BC₂	19.7	±0.12	0.135	0.39	18.8	20.4	30	0

[1] Confidence interval at the 5% probability level.
[2] Estimate of the genetic variance component.
[3] Broad-sense heritability on an entry mean basis.

centages was different for each set of parents (table 3). For the Wells × PI 153.269 cross, 1% of the F_2-derived lines were transgressive segregates beyond the confidence interval of either parent. For the remaining F_2-derived lines, 16% were within the confidence interval of Wells, and 12% were within the confidence interval for PI 153.269. For the Woodworth × Pando cross, 1% of the F_2-derived lines were transgressive segregates beyond the confidence interval of the recurrent parent, and only 7% of

the lines were within the confidence interval for Woodworth and 1% within the confidence interval for Pando.

Mean values for protein, oil, and protein + oil percentages of each generation changed in the direction of the recurrent parents (tables 1, 2, 3). For protein and oil percentages, frequency of lines within the confidence interval of the recurrent parents increased with every backcross (tables 1, 2), although no lines in the Woodworth BC_1 generation for protein percentage were within the confidence interval of Woodworth (table 1). For protein + oil percentages, the frequency of lines within the confidence interval of each recurrent parent increased at the BC_1 generation, but a decrease was found at the BC_2 generation.

TABLE 3.—*Performance of parents and their progeny for protein + oil content averaged across two environments*

Entry	Protein + Oil		σ_g^{2}[2]	H[3]	Protein + Oil		Within C.I.[1] of	
	$\bar{x}$	C.I.[1]			Lowest line	Highest line	Recurrent parent	Donor parent
	%				%		*% of lines*	
Wells	61.9	±0.24	—	—				
PI 153,269	63.6	±0.26	—	—				
F_2	62.7	±0.09	0.488	0.78	60.0	64.6	16	12
Wells BC_1	62.2	±0.09	0.138	0.56	61.2	63.2	39	0
Wells BC_2	62.2	±0.13	0.190	0.59	61.0	63.8	25	6
Woodworth	59.9	±0.16	—	—				
Pando	63.7	±0.24	—	—				
F_2	61.2	±0.09	0.345	0.68	59.6	63.2	7	1
Woodworth BC_1	60.4	±0.11	0.292	0.65	59.3	62.1	21	0
Woodworth BC_2	60.5	±0.15	0.108	0.24	59.6	61.6	11	0

[1] Confidence interval at the 5% probability level.
[2] Estimate of the genetic variance component.
[3] Broad-sense heritability on an entry mean basis.

For the two sets of parents, significant genetic variability was obtained in each generation for protein, oil, and protein + oil. Genetic variance of the F_2 was largest, and it decreased with every subsequent backcross. Genetic variation for all traits still was present at the BC_2 generation (tables 1, 2, 3).

Heritability estimates on an entry mean basis were within the range of those reported in the literature (4, 5, 7), (tables 1, 2, 3). The estimates suggest that selection for the traits in replicated tests should be effective in improving the populations.

The mean protein contents we found in the F_2 generation were lower than the mid-parent values calculated for both sets of parents. These findings are in agreement with those reported by Viljoen (8) for the three

sets of crosses he studied. The data reported by Williams (10) and Weber (9), however, showed mean values for the F_2 higher than the mid-parent value. The mean oil contents we observed in the F_2 generations were greater than the mid-parent values, as Viljoen (8) also found. In contrast, Williams (10) and Weber (9) reported that for their F_2 populations, mean oil values were equal to or very close to the mid-parent values. The difference between the results obtained for protein and oil by Williams (10) and Weber (9) and those of Viljoen (8) and ourselves may be associated to the fact that the former authors derived their populations from interspecific crosses, while our populations were obtained from intraspecific crosses.

Our results indicate that it will be difficult without selection to transfer genes for extremely high protein levels into cultivars with lower protein, by backcrossing to the low-protein parent. The feasibility of transferring genes for high-protein into low-protein soybean cultivars by backcrossing has not yet been evaluated. Hartwig and Hinson (2) used the backcross method to obtain closely related lines selected for differences in oil percentage. They crossed Bragg (2800 kg/ha, 39.6% protein) with a breeding line D60-7965 (2100 kg/ha, 44.4% protein) and minimized the contribution of D60-7965 by developing backcross populations, BC_1, and BC_2. The authors obtained 25 BC_1F_1 plants by backcrossing directly with the F_1, and each of the 25 BC_1F_1 plants was used as a pollen parent for making a second backcross to Bragg. They evaluated bulk F_2 and F_3 populations from each BC_1F_1 plant and F_2 populations from each BC_2F_1 plant, selected for low, intermediate, and high oil, and evaluated the selections in replicated experiments. They were able to isolate a BC_2 line with seed yield (2700 kg/ha) closely approximating the high-yielding parent and with protein content (43.5%), similar, but not equal to that of the donor parent.

The effectiveness of the backcross method may be improved by evaluating F_2-derived lines for seed composition and using the lines with the highest protein to cross with the recurrent parent. It may also be desirable to select among F_3 or F_4 plants within F_2-derived lines with high protein content to enhance the possibilities of identifying high-protein segregates. Research is in progress to evaluate this approach.

RESUMEN

Se evaluó la variabilidad genética de semillas de soja (*Glycine max* (L.) Mer.) obtenidas de cruces entre padres con alto contenido (PI 153.269, Pando) y bajo contenido en proteína (Wells, 'Woodworth'). Wells se cruzó con PI 153.269 y se lograron dos generaciones de retrocruces con Wells. Woodworth se cruzó con Pando y también se obtuvieron dos generaciones de retrocruces con Woodworth. Los padres

y las líneas derivadas de las generaciones F_2, BC_1F_1, y BC_2F_1, se evaluaron en dos ambientes. Se determinó el contenido de proteína y aceite mediante el uso de rayos infrarrojos.

En ninguno de los dos cruces se encontraron líneas con contenido proteínico tan alto como el de los padres donantes. En el cruce entre Wells × PI 153.269, la línea derivada F_2 más rica en proteína tuvo un contenido de 48.0%, mientras que el padre donante tuvo 49.4%. En el cruce entre Woodworth × Pando, la línea derivada F_2 con mayor riqueza de proteína tuvo 46.0%, mientras que Pando tuvo 49.2%. Las varianzas genéticas y las medias de contenido de proteína de las poblaciones fueron disminuyendo con cada generación de retrocruce. Las heredabilidades en sentido amplio de contenido de proteína, expresadas a base de las líneas incluidas en el experimento, tuvieron un rango de 0.76 a 0.89. Se discuten las implicaciones que estos resultados sugieren para la utilización de un método de hibridación encauzado a mejorar la riqueza en proteína de la soja.

LITERATURE CITED

1. Crump, S. L., 1946. The estimation of variance components in analysis of variance, Biom. Bull. 2: 7–11.
2. Hartwig, E. E. and Hinson, K., 1972. Association between chemical composition of seed and seed yield of soybeans, Crop Sci. 12: 829–30.
3. Ishige, T., 1978. Statistical studies on the inheritance of protein content in soybeans, *Glycine max* (L.) Merill, Annu. Rep. Div. Gen. Dep. Physiol. and Gen. Nat. Inst. Agric. Sci.
4. Shannon, J. G., Wilcox, J. R., and Probst, A. H., 1972. Estimated gains from selection for protein and yield in the F_4 generation of six soybean populations, Crop Sci. 12: 824–26.
5. Singh, L. and Hadley, H. H., 1972. Maternal and cytoplasmic effects on seed protein content in soybeans, *Glycine max* (L.) Merrill, Crop Sci. 12: 583–85.
6. Snedecor, G. W., and Cochran, W. G., 1967. Statistical methods, Iowa State Univ. Press, Ames, Iowa.
7. Thorne, J. C. and Fehr, W. R., 1970. Incorporation of high-protein, exotic germplasm into soybean populations by 2- and 3-way crosses, Crop Sci. 10: 652–55.
8. Viljoen, N. J., 1937. An investigation into the composition of the soybeans in South Africa, Union of South Africa Dep. Agric. Fore. Sci. Bull. 169.
9. Weber, C. R., 1950. Inheritance and interrelation of some agronomic and chemical characters in an interspecific cross in soybeans, *Glycine max* x *G. ussuriensis*, Iowa State Agric. Exp. Stn. Res. Bull. 374.
10. Williams, L. F., 1948. Inheritance in a species cross in the soybeans, Genetics 33: 131–32.