

# Protein, Oil, and Yield of Soybean Lines Selected for Increased Protein

T. C. Helms\* and J. H. Orf

## ABSTRACT

Soybean breeders need to evaluate alternative strategies for increasing yield, protein, and oil content. The correlated response for yield when selection is conducted for increased protein among unreplicated plant rows is unknown. Our objective was to evaluate selection for increased protein content at an intensity of 50% using unreplicated plots to determine the correlated response for yield, gross value per hectare, oil content, and the direct response for protein content. Ten populations that included 30 lines each were evaluated for yield, protein, and oil content at each of seven environments. Lines were selected solely for high protein content with one replicate at each of three selection environments and based on the average of two replicates in three selection environments. Additionally, selection was conducted on the basis of the population mean in the selection environment(s). Selected and unselected lines were evaluated in four validation environments that were not the same as the selection environments. Selection among individual lines solely for increased protein content decreased yield 70 kg ha<sup>-1</sup> and increased protein content 7 g kg<sup>-1</sup>, when lines were selected on the basis of the mean of two replicates and three environments. Selection among populations for increased protein content, decreased yield 110 kg ha<sup>-1</sup> and increased protein content 5 g kg when selection was based on the population mean averaged across two replicates and three environments. On average, selection for increased protein content resulted in decreased yield.

SOYBEAN [*Glycine max* (L). Merr.] lines are evaluated in unreplicated plant rows the year proceeding yield trial evaluation. At this stage of testing, limited seed is available per genotype with large numbers of genotypes evaluated. Johnson et al. (1955b) suggested that independent culling could be conducted at the plant-row stage for traits such as protein and oil content and the selected lines could be evaluated for yield the following year. Typically, selection is conducted on the basis of the phenotype of an individual line. An alternative to selecting among phenotypes would be to select and discard whole populations based on the population mean. The population mean would be estimated with greater precision than the phenotypic value of individual lines.

Falconer (1960) discussed selection based on individual phenotypic values, the family mean, or within family selection. He stated that selection based on the mean of an individual genotype gives equal weight to selection among and within families. When selection is conducted among families, all the individuals within the selected families are kept and all the individuals in the nonselected families are discarded. Family selection is favored when the heritability is low and families are large. However, selection among population means is not the same as among family selection. All families are developed from a single population and among family selection is conducted within a single population.

T.C. Helms, Dep. of Plant Sciences, North Dakota State Univ., Fargo, ND 58105; J.H. Orf, Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, St. Paul, MN 55108. Received 5 June 1997. \*Corresponding author (helms@badlands.nodak.edu).

After selection has been conducted among plant rows, on the basis of either the population mean or individual phenotypic values, the selected lines are subjected to replicated testing in validation environments. Validation environments that are not the same as the selection environments can be used to evaluate the response to selection. This approach mimics the strategy commonly used to develop improved cultivars. Helms et al. (1995) used validation environments to compare selection based on nearest neighbor with selection based on unadjusted means. The purpose of the validation environments was to evaluate the actual response of two selection methods after selection had been conducted in one environment.

Shannon et al. (1972) reported that the heritability of protein content was higher than the heritability of yield for each of six soybean populations. This suggests that selection for increased protein content among unreplicated plant rows could be an effective strategy. Genotypes low in protein could be discarded prior to the expensive procedure of yield evaluation, which is typically replicated across and within locations. Negative genetic correlations between yield and protein content have been reported (Simpson and Wilcox, 1983; Shannon et al., 1972). The negative genetic correlations between yield and protein content suggest that independent culling for protein content would result in a correlated response of lower yield. However, selecting 50% of the lines with the highest protein content might be effective in increasing protein content with only a negligible decrease in mean yield.

Plant breeders typically have an opportunity to select among unreplicated plant rows for traits such as protein content, maturity, and visual appearance. Selection conducted in several environments would be useful to determine whether the correlated and direct response to selection is similar across environments. An alternative to selection based on individual phenotypes would be to select among populations. Our objective was to evaluate selection for increased protein content at an intensity of 50% using unreplicated plots to determine the correlated response for yield, gross value per hectare, oil content, and the direct response for protein content.

## MATERIALS AND METHODS

Experimental lines were developed from 10 different populations. The pedigrees of each population included 'Merit' × 'Ozzie' (Population 1), M84-93 × M81-18 (Population 2), {Ozzie × 'Sprite'} × 'Sigco KG20' (Population 3), {Ozzie × 'Sprite'} × Ozzie (Population 4), 'Kato' × 'Maple Ridge' (Population 5), 'Agassiz' × 'Northrup King S07-80' (Population 6), 'Pioneer 9061' × 'Evans' (Population 7), 'Bert' × M84-93 (Population 8), LMA-82/1 × Maple Ridge (Population 9), and LS-352 × 'McCall' (Population 10). M84-93 and M81-18 are experimental lines derived from the University of Minne-

sota soybean breeding project. The pedigree of M84-93 is {'Clay'  $\times$  Evans}  $\times$  Ozzie. The pedigree of M81-18 is Evans  $\times$  {'Anoka'  $\times$  'Amsoy'}. LMA-82/1 and LS-352 are experimental lines obtained from Poland by the University of Minnesota.

In the Pioneer 9061  $\times$  Evans and Bert  $\times$  M84-93 populations, lines were derived from  $F_3$  plants that were individually threshed and  $F_{3,4}$  plant-rows were evaluated. In the remaining eight populations,  $F_4$  plants were individually threshed and  $F_{4,5}$  plant rows were evaluated. Thirty plant rows within each of the 10 populations were selected for early physiological maturity and then visual appearance, but not for protein or oil content. After plant rows were selected for early maturity and visual appearance, seed within each plant-row was bulked for evaluation in multiple-row plots.

The 30 lines from each of the 10 populations were then evaluated in replicated yield trials at each of seven environments. Selection for increased protein content was conducted in three of these seven environments. Genotype  $\times$  environment interactions exist for protein content (Schutz and Bernard, 1967). Selection for protein content in more than one environment is useful to evaluate whether the response to selection is repeatable across environments.

Two-row plots were planted at Casselton and Mantador, ND, in 1995 and 1996. Plots were 6.1 m long with 76-cm spacing between rows. The center 4.3 m of the two-row plot was harvested. Planting rate was 41 seeds  $m^{-1}$  of row. Four-row plots were planted at Morris and Rosemount, MN, in 1995 and at Rosemount, MN in 1996. Plots were planted 3.7 m long with 25-cm spacing between rows. The center 2.4 m of all four rows in each plot were harvested. Planting rate was 34 seeds  $m^{-1}$  of row. Protein and oil content was reported on a dry weight basis as determined by near infrared reflectances at the National Center for Agricultural Utilization Research, Peoria, IL.

The 10 populations were evaluated by a sets within replicates arrangement. Six lines from each of the 10 populations were assigned to one of the five sets. Each set consisted of these 60 experimental lines plus four commercially available check cultivars for a total of 64 entries per set. The same experimental lines were assigned to the same set in all seven environments. There were five sets at each environment and each set was arranged in a separate eight-by-eight lattice design with two replicates. Each set within each environment was adjusted for lattice effects before conducting a combined analysis.

Lines were selected for high protein content at the 1995 Rosemount, MN, 1996 Casselton, ND, and the 1996 Mantador, ND environments. Selection for increased protein content among individual lines was conducted by choosing 30 lines from each of five sets for a total of 150 selected lines. Selection was conducted on the basis of data from the selection environment(s). Lines were selected without regard to the population from which they originated. The selection intensity was 50%. When selection was conducted in one environment, protein content of lines was determined with data from

one replicate of these multiple-row plots, which mimics selection among unreplicated plant rows. Also, selection was conducted on the basis of protein content of each line, averaged across three selection environments and the two replicates within each environment.

Alternatively, selection was conducted on the basis of the population means of data from the selection environment(s). Five of the 10 populations were selected, for a selection intensity of 50%. All 30 lines within each selected population were retained for a total of 150 lines.

All lines were then evaluated at four validation environments. The validation environments were Morris, MN, in 1995; Casselton, ND, in 1995; Mantador, ND, in 1995; and Rosemount, MN, in 1996. Validation environments were chosen randomly and the remaining three environments were used as selection environments.

Gross value per hectare was defined as price per kilogram multiplied by yield per hectare. Price per kilogram was determined by Updaw and Nichols (1980) formula with protein valued at \$0.4001  $kg^{-1}$  and oil valued at \$0.4741  $kg^{-1}$  (Orf and Helms, 1994). These protein and oil prices represented a market with roughly equal value for both constituents. Gross value per hectare was reported after yield and protein and oil contents were adjusted to 130 g  $kg^{-1}$  grain moisture.

The selection differential was the difference between the mean of the selected lines and the unselected population mean, averaged across the four validation environments. This procedure is comparable to that of Eta-ndu and Openshaw (1992). Procedures of Hallauer and Miranda (1981) were used to estimate heritability on an entry-mean basis and standard errors of variance components. Mean squares for genotypes pooled across sets were tested with the genotype  $\times$  environment within sets mean square as the denominator for the  $F$ -ratio.

## RESULTS AND DISCUSSION

The mean square among genotypes within sets was significant for protein and oil content ( $P < 0.01$ ) for all 10 populations. The mean square for genotypes within sets was significant for yield and gross value per hectare ( $P < 0.10$ ) for all 10 populations. The significant genotypes within sets mean squares for protein content are evidence that genetic variability existed for protein content. Mean protein content varied from 437 for population 5 to 412 g  $kg^{-1}$  for Population 7 (Table 1). Genetic variance estimates were greater than twice the standard error for all populations except Population 6. Heritability estimates on an entry-mean basis were large and suggest that the lines with the highest protein content could be identified when selection was based on the mean of all seven environments.

Correlations between data from one replicate of a selection environment with the mean, averaged across two replicates and four validation environments, were high at Rosemount, MN, in 1995 (Table 2). These correlations were lower at Casselton and Mantador, ND, 1996 locations. These correlations are evidence that improve-

**Table 1.** Mean, genetic variance estimate, standard error of genetic variance, and heritability on an entry mean basis for protein content of ten soybean populations, evaluated across seven environments.

Population†	Mean	Genetic variance	Standard error	Heritability
	g kg <sup>-1</sup>			
1	425	44.9	14.4	0.89
2	421	40.8	13.4	0.86
3	432	51.0	16.6	0.87
4	426	51.9	16.4	0.90
5	437	81.7	25.0	0.92
6	421	6.0	3.5	0.50
7	412	14.3	5.8	0.71
8	413	17.5	6.0	0.82
9	421	9.6	4.4	0.63
10	418	21.5	7.4	0.82
LSD (0.05)	2			

† Populations: 1 = Merit × Ozzie; 2 = M84-93 × M81-18; 3 = (Ozzie × Sprite) × Sigco KG20; 4 = Ozzie<sup>2</sup> × Sprite; 5 = Kato × Maple Ridge; 6 = Agassiz × Northrup King S07-80; 7 = Pioneer 9061 × Evans; 8 = Bert × M84-93; 9 = (LMA-82/1) × Maple Ridge; 10 = (LS-352) × McCall.

ments in protein content would be expected when lines are selected at a single location, on the basis of data from a single replicate.

The mean yield of 150 lines selected solely for high protein content from one replicate of data at Rosemount, MN, in 1995 was 2940 kg ha<sup>-1</sup>, when these lines were evaluated at four validation environments (Table 3). The overall mean of the 300 lines was 2930 kg ha<sup>-1</sup>, when these lines were averaged across four validation environments. Selection for protein content did not reduce yield when conducted at Rosemount, MN, in 1995, but did reduce yield when conducted at Casselton and Mantador, ND, in 1996. Selection solely for protein content decreased yield when selection was conducted on the basis of the mean of two replicates, averaged across three environments. The response to selection based on the mean of three environments suggests that in most environments, selection for increased protein content will reduce yield. Genotype × environment interactions effects are reduced when selection is based on the mean averaged across environments.

Protein content of selected lines was greater than the overall mean, when selection was conducted at each of the three environments (Table 3). However, the increase in protein content due to selection was small. Oil content decreased slightly when selection was conducted solely for protein content. The gross value per hectare increased when selection was conducted at Rosemount, MN, in 1995. Gross value per hectare due to the lower yield of selected lines decreased from selection solely for protein content at Casselton and Mantador, ND, in 1996. The higher protein content of lines selected at Casselton and Mantador, ND, in 1996 did not compensate for the decreased yield. Gross value per hectare decreased when selection of individual lines for increased protein content was based on the mean of three environments.

Hazel and Lush (1942) discussed independent culling. They stated that the advantage of independent culling is that selection can be conducted for one trait without waiting until all the traits can be measured. The disad-

**Table 2.** Correlations between selection and validation environments for yield, protein, and oil content among 300 soybean lines derived from ten populations. Data from one replicate in the selection environment was correlated with the mean of validation environments, averaged across two replicates and four environments.

Selection environment	Correlation coefficient		
	Yield	Protein	Oil
1995 Rosemount	0.78**	0.83**	0.66**
1996 Casselton	0.43**	0.43**	0.55**
1996 Mantador	0.32**	0.40**	0.56**
All three environments	0.81**	0.76**	0.73**

\*\* Significant at the 0.01 probability level.

vantage is that "increasing the intensity of culling for one characteristic automatically lowers the intensity of selection possible for other characteristics." In our experiment independent culling was conducted first for maturity, then visual score, and finally for protein content. Independent culling can first be conducted for maturity and visual score without measuring protein content on all plant rows. Only the lines selected for protein content would be evaluated for yield. Our results demonstrate that averaged across selection environments, culling for increased protein content decreased yield. These results have implications for all breeders who conduct independent culling for traits that have negative genetic correlations with yield.

We simulated plant-row selection using multiple-row plots. Pazdernik et al. (1996) reported that selection for protein and oil content was equally effective for short one-row plots as it was for long four-row plots. Correlations for protein content conducted between locations within the same year were significant. However, correlations for protein content between single locations in different years were not significant. They concluded that selection for high protein in short, one-row plots would be likely to increase protein content. They did not evaluate the correlated response of selection for protein content on yield.

**Table 3.** Selection differential for yield, physiological maturity (days from 1 Aug.), protein, oil content, and gross value per hectare (GVH) of 150 soybean lines selected from 300 lines tested in three selection environments and evaluated averaged across four validation environments. Lines were selected solely for protein content based on individual line data.

Selection environment†	Selection differential averaged across four validation environments				
	Yield	Physiological maturity	Protein‡	Oil	GVH§
	kg ha <sup>-1</sup>	days	g kg <sup>-1</sup>		\$ ha <sup>-1</sup>
1995 Rosemount	10	0	7	-3	7
1996 Casselton	-120	-2	4	-2	-22
1996 Mantador	-50	-1	5	-2	-7
All three environments	-70	-1	7	-3	-11
300 line mean	2930	50	416	204	591
LSD (0.05)¶	20	0.4	1	1	5

† Lines were selected based on data from one replicate at each individual selection environment and selected based on the mean of two replicates, averaged across three selection environments.

‡ Protein and oil content are reported on a dry matter basis.

§ GVH is reported at 130 g kg<sup>-1</sup> grain moisture.

¶ Least significant difference between 150 selected lines and the overall mean.

Our results are consistent with those of Orf and Helms (1994). They found that selection for protein content decreased yield, and selection solely for yield would identify the lines with the highest gross value per hectare. However, they did not examine the influence of independent culling at a low selection intensity. Also, they only evaluated three populations with mean yields of 1430, 1340, and 1700 kg ha<sup>-1</sup>. The low yield of the populations in their report was due to unfavorable environmental conditions. Our results confirm that, in high yielding environments, selection solely for protein content decreased gross value per hectare in two of the three selection environments. Our results mimic selection among plant rows because selection was conducted on the basis of data from a single replicate.

Johnson et al. (1955a) suggested that early generation testing at one location should be effective in increasing protein content of soybean. Their strategy was to select for traits other than yield in the early generations and then select for yield in a later generation after the quantity of seed had been increased. They also stated that effective selection for traits other than yield "would result in discarding some of the better genotypes for yield, if selection pressures for the character in question was very strong." Johnson et al. (1955b) predicted that selection solely for protein content would decrease the mean yield of the selected lines in two soybean populations. They stated that the negative genetic correlation between protein and yield "may prove to be serious limitations in breeding for high protein." However, they based these considerations on predicted correlated response for yield when selection was for protein content and did not evaluate actual correlated response. Our results, determined by measuring actual correlated response for yield when selection was conducted for protein content, were consistent with the predicted correlated response estimated by Johnson et al. (1955b).

When selection was conducted solely for protein content, on the basis of the population mean, the results

were similar to selection based on individual lines (Table 4). At Rosemount, MN, selection among populations for increased protein content increased yield, protein, and gross value per hectare. However, selection among populations for high protein at Casselton, ND, in 1996 identified lines that yielded 2740 kg ha<sup>-1</sup>, which has a lower yield than selection of individual lines at this location with a mean yield of 2810 kg ha<sup>-1</sup>. Also, when selection was conducted among populations at Mantador, ND, in 1996, the mean yield of selected lines in validation environments was 2820 kg ha<sup>-1</sup>, which was lower than the 2880 kg ha<sup>-1</sup> mean yield when selection was conducted at this location on the basis of selection among individual lines. Protein content of lines selected on the basis of the population means (Table 4) was equal to selection among individual lines (Table 3). These results show that selection for protein content among individual lines resulted in higher yielding lines than selection among populations in two of the three selection environments.

Falconer (1960) suggested that selection among families would be superior to selection of individual phenotypes when the heritability of a trait is low. Use of a single replicate when selecting among plant-rows for protein content would reduce the heritability compared with selection among replicated plots. However, selection among family means is different from selection among population means. Families are derived from a common population. Measured in the four validation environments, the genetic variance for protein content among populations was 14 and the pooled genetic variance for lines within populations was 42. The lack of genetic variance among population means is the reason that selection among populations was no more effective in increasing protein content than selection of individual phenotypes.

Wilcox et al. (1984) evaluated selection among and within families derived from three-parent crosses. Twenty families, derived from 20 F<sub>1</sub> plants, were evaluated and four lines were developed from each of these 20 families. They found that the genetic variance within families was 1.4 times larger than the variance among families. Lines with superior agronomic performance tended to come from the families that had a high family mean. Our results are different from those of Wilcox et al. (1984) and this could be due to the fact that our populations were each derived from a different combination of parents, while they evaluated family means within a cross.

Orf and Helms (1994) reported that combined selection for yield, protein, and oil content in replicated plots did not identify lines that had a higher gross value per acre than lines selected solely for yield. Our results show that selection among unreplicated plots for protein content increased protein content slightly and decreased yield and gross value per hectare in two of the three selection environments. Selection in one environment for increased protein content increased protein, and gross value per hectare. However, when selection was based on the mean of three environments, selection for increased protein content decreased yield and gross

**Table 4. Selection differential for yield, physiological maturity (days from 1 Aug.), protein, oil content, and gross value per hectare (GVH) of 150 soybean lines selected from 300 lines tested in three selection environments and evaluated averaged across four validation environments. Lines were selected solely for protein content, based on the population mean.**

Selection environment†	Selection differential averaged across four validation environments				
	Yield	Physiological maturity	Protein‡	Oil	GVH§
	kg ha <sup>-1</sup>	days	g kg <sup>-1</sup>		\$ ha <sup>-1</sup>
1995 Rosemount	50	0	6	-3	14
1996 Casselton	-190	-3	2	-2	-38
1996 Mantador	-110	-1	5	-3	-19
All three environments	-110	-1	5	-3	-19
300 line mean	2930	50	416	204	591
LSD (0.05)¶	20	0.4	1	1	5

† Lines were selected based on data from one replicate at each individual selection environment and selected based on the mean of two replicates, averaged across three selection environments.

‡ Protein and oil content are reported on a dry matter basis.

§ GVH is reported at 130 g kg<sup>-1</sup> grain moisture.

¶ Least significant difference between 150 selected lines and the overall mean.

value per hectare. On average, selection for increased protein content resulted in a correlated response of decreased yield.

#### ACKNOWLEDGMENTS

This research was supported in part by grants from the Minnesota Soybean Research and Promotion Council. Appreciation is expressed to the National Center for Agricultural Utilization Research, Peoria, IL.

#### REFERENCES

- Eta-Ndu, J.T., and S.J. Openshaw. 1992. Selection criteria for grain yield and moisture in maize yield trials. *Crop Sci.* 32:332-335.
- Falconer, D.S. 1960. *Introduction to Quantitative Genetics*. Ronald Press Co., New York.
- Hallauer, A.R., and J.B. Miranda, F.O. 1981. *Quantitative genetics in maize breeding*. The Iowa State Univ. Press, Ames, IA.
- Hazel, L.N., and J.L. Lush. 1942. The efficiency of three methods of selection. *J. Hered.* 33:393-399.
- Helms, T.C., J.H. Orf, and R.A. Scott. 1995. Nearest-neighbor-adjusted means as a selection criterion within two soybean populations. *Can. J. Plant Sci.* 75:857-863.
- Johnson, H.W., H.F. Robinson, and R.E. Comstock. 1955a. Estimates of genetic and environmental variability in soybeans. *Agron. J.* 47:314-318.
- Johnson, H.W., H.F. Robinson, and R.E. Comstock. 1955b. Genotypic and phenotypic correlations in soybeans and their implications in selection. *Agron. J.* 47:477-483.
- Orf, J.H., and T.C. Helms. 1994. Selection to maximize gross value per hectare within three soybean populations. *Crop Sci.* 34:1163-1167.
- Pazdernik, D.L., L.L. Hardman, J.H. Orf, and F. Clotaire. 1996. Comparison of field methods for selection of protein and oil content in soybean. *Can. J. Plant Sci.* 76:721-725.
- Schutz, W.M., and R.L. Bernard. 1967. Genotype  $\times$  environment interactions in the regional testing of soybean strains. *Crop Sci.* 7:125-130.
- Shannon J.G., J.R. Wilcox, and A.H. Probst. 1972. Estimated gains from selection for protein and yield in the  $F_4$  generation of six soybean populations. *Crop Sci.* 12:824-826.
- Simpson, A.M., Jr., and J.R. Wilcox. 1983. Genetic and phenotypic associations of agronomic characteristics in four high protein soybean populations. *Crop Sci.* 23:1077-1081.
- Updew, N.J., and T.E. Nichols, Jr. 1980. Pricing soybeans on the basis of chemical constituents. p. 781-799. *In* F.T. Corbin (ed.) *Proc. 2nd World Soybean Res. Conf.*, Raleigh, NC 26-29 Mar. 1979. Westview Press, Boulder, CO.
- Wilcox, J.R., A.G.M. Khalaf, and G.D. Brossman. 1984. Variability within and among  $F_1$  families from diverse three-parent soybean crosses. *Crop Sci.* 24:1055-1058.