CBAB

CROP BREEDING AND
APPLIED BIOTECHNOLOGY

Selection of high protein content and high yield soybean families

Odilon Lemos de Mello Filho[1*], Carlos Sigueyuki Sediyama[2], Maurílio Alves Moreira[3], Múcio Silva Reis[2], Guilherme Augusto Massoni de Andrade[2], Rafael Rosalino Dias[2], and Newton Denis Piovesan[4]

ABSTRACT - *The present work aimed to show that it is possible to select soybean families bearing high seed protein contents without decreasing the grain yield. Four BC_1F_4 and four F_4 populations were used for this purpose. These populations were originated each from the crossing of a high protein material and a commercial variety. The population means, heritabilities, correlation coefficients, and gains by direct selection were estimated. The main conclusions were* i) *it was possible to select families presenting both high seed protein content and high yield;* ii) *population CD206-CR presented families with highest seed protein contents. Despite a high heritability of the protein content and an intermediate for grain yield, the negative genetic correlation between these traits did not permit a simultaneous selection.*

Key words: direct selection, heritability, correlations, *Glycine max*.

INTRODUCTION

Protein in soybean seed is a component of great economical value. On average, the protein content in commercial varieties is around 40%. However, there are accessions with percentages varying from 30 to 53%. In the last years, breeding programs have been focusing on the increase of the protein content and also its quality improvement (Sediyama et al. 1996, Sediyama et al. 1999).

The correlations among traits reflect the association degree among them. Breeding programs usually, besides focusing on the improvement of a target trait, are also directed towards maintaining or improving the expression of another one.

Thus, the determination of the correlation among traits becomes particularly important, since it provides knowledge about how the selection for a certain trait influences the expression of another (Piovesan 2000).

The progress in the development of materials with both high grain yield and desirable agronomic characters of the existent varieties, allied to the high protein content of exotic materials has been slow (Shannon et al. 1972). For 95 lines originated from three BC_2F_2 populations, Wehrmann et al. (1987) obtained genotypic correlation coefficients between protein content and grain yield of -0.54, -0.64, and -0.86 for lines descending from each one of the three populations.

[1]Departamento de Fitotecnia, Universidade Federal de Viçosa (UFV), 36570-000, Viçosa, MG, Brasil. *E-mail: odilonlemos@bol.com.br
[2]Departamento de Fitotecnia, UFV
[3]Departamento de Bioquímica e Biologia Molecular, UFV
[4]Bioagro, UFV

However, Thorne and Fehr (1970) were able to transfer the high protein content of exotic materials to the progenies, realizing the simultaneous selection of high protein content and high grain yield materials, in spite of the negative significant phenotypic correlations detected for some of the analyzed populations.

Shannon et al. (1972) conducted F_2 families of şix populations originated from the crossing between parents with good agronomic traits and high grain yield or high protein. They evaluated, in generation F_4, which types of crossings (high protein x high protein (P), high protein x high grain yield (G) or high grain yield x high grain yield) offered the best opportunity for the increase of grain yield, of protein, or of both traits. Populations P x G produced the largest number of families with high grain yield. Almost all families that presented a higher protein content were originated from the crossing of two families that bear this trait. The number of families that produced more protein ha^{-1} (combination of grain yield and protein content) was equally divided between P x P and P x G. The authors concluded, in relation to a population P x P, that even with a negative and significant correlation (-0.28) between protein and grain yield, this value could be considered sufficiently low, so that the selection of high yield and high protein content materials did not seem to be extremely difficult.

In this context, aim of the present study was to show the possibility of selecting lines that bear high seed protein content without setbacks for the grain yield.

MATERIAL AND METHODS

The experiment was carried out at the Federal University of Viçosa (UFV), Viçosa, State of Minas Gerais, in two parts. The first part of the experiment was conducted on the Experimental Field "Professor Diogo Alves de Mello", lat 20° 45' 20" S, long 42° 52' 53" W, and 650m asl of altitude. The second one was realized at the laboratories of the Plant Science Department and of the Institute of Biotechnology Applied to Agropecuary (BIOAGRO).

Eight populations, raised in the frame of the BIOAGRO/ UFV Soybean Breeding Program under an agreement with the Central Cooperative for Agropecuary Technological and Economic Development Ltd (COODETEC) were used.

The families that composed the populations were in the fourth generation of self-fertilization after artificial hybridization (F_4) or in the fourth generation of self-fertilization after the backcross (BC_1F_4). These families were originated, each one, from the cross between a line with high protein (donor

ancestor, for the BC families) with a commercial variety belonging to COODETEC (recurrent ancestor, for the BC families).

In the F_2 generation, plants were selected at 10% intensity (around 30 seeds) for high protein based on their seed protein content evaluated by the bicinconinic acid method. F_3 generation seeds were analyzed for protein with the Kjeldahl method. Amongst the plants with highest protein content in the seeds the four genetically closest the recurrent ancestor, in each family, were selected. The families were then, successively, self fertilized and harvested in bulks.

The F_4 and BC_1F_4 populations were planted January 12, 2001. They consisted in 17, 13, 20, and 11 BC_1F_4 families, identified as OC13BC, CD201BC, CD205BC, and CD206BC, and in 23, 22, 20, and 17 F_4 families, identified as OC13CR, CD201CR, CD205CR, and CD206CR, respectively. These families were originated from the four commercial varieties Ocepar 13, Coodetec 201, Coodetec 205, and Coodetec 206, identified respectively as OC13-P, CD201-P, CD205-P, and CD206-P and were planted in 6, 6, 5, and 3 replications. Additionally, three commercial varieties were planted as controls: CAC1, Coodetec 203, and Coodetec 205 (for this variety, only the central rows in the blocks were used as controls), identified as CAC1-T, CD203-T, and CD205-T, planted in 20 blocks. The high protein material (donor ancestor, for the BC families) for OC13 and CD205 varieties was the IAC12PTN line, for CD201, the BR 80 14887 line and for CD206, the CAC1PTN line. The mean seed protein contents of the three lines were respectively, 51.00, 47.80, and 54.20%.

Each experimental plot contained 20 plants, in 2 meter rows, spaced 0.5 m apart. The seeds were planted in the field, and were treated with Benomyl and Metalaxyl + Mancozeb fungicides and also with *Bradyrhizobium japonicum* inoculant. Weeds were controlled by 60g a.i. ha^{-1} of the pre-emergent herbicide Sulfentrazone. Four hundred kg ha^{-1} of 0-30-10 fertilizer were applied at sowing, besides the necessary cultural treatments and complementary sprinkling irrigation during the development of the plants.

The harvest was carried out ten days after 50% of the plants of a row had reached the R8 stadium (Fehr and Caviness 1977). Despite 20 plants were grown on each experimental plot, the grain yield data were collected just from the six most central plants of each row, due to operational questions. The grain yield data was obtained by weighing the seeds of each plant.

The total protein content of the seeds was evaluated by the modified Kjeldahl method according to Instituto Adolph Lutz (1985). Three seeds from each of the six plants from each plot were aggregated to form a bulk with 18 seeds, and analyzed in two replications. In cases where the differences between them

exceeded one percent, the analysis was repeated. A 6.25 factor was used to convert nitrogen into protein (AOAC 1984).

The applied experimental design was Families with Intercalated Controls (Cruz 2001), where only one replication of a family row was planted, due to the small number of seeds available. To test one population against the others, the completely randomized experimental design was used, so within each population, the family rows were considered replications. Every ten family rows, at the beginning and the end of each population, a block composed by the control varieties plus one progenitor of the respective population was inserted, totalizing 20 blocks all over the experiment. These 20 blocks were used to estimate the effect of blocks, and to evaluate the homogeneity of the environment. The variance analyses and estimations of means and phenotypic correlation coefficients were established using the software Genes (Cruz 2001).

RESULTS AND DISCUSSION

Initially, the variance analysis of the 20 replications (blocks) of control varieties used in the experiment was realized to avoid the influence of possible environmental variations on the results. There were no significant differences among the control blocks for the evaluated traits, a fact that indicates that the environmental variations were evenly distributed. Thus, it was not necessary to correct the population data for comparison. There were significant differences among control varieties, regarding the appraised traits, which suggests variability of the controls. Comparing the progenitors (P) among themselves, as well as with the controls (C) and populations (PP), significant differences of protein content in the seeds (PROT) may be observed (Table 1).

The precision levels of the experiment, evaluated by the experimental variation coefficients, were scored good to mean. According to Pimentel Gomes (1984), in agronomic experimentation, values below 10% indicate an experiment with high precision; between 10 and 20% mean precision; between 20 and 30% low, and over 30% a very low precision. As the variation coefficients presented appropriate levels, the data are also assumed to be reliable.

Significant variation was verified among the populations, indicating the importance of the use of several populations, in order to verify the persistence of the obtained results (Table 2). Besides, for the trait grain yield, variation was only detected within the populations CD206-BC and CD206-CR. However, regarding the protein seed content, the populations, except for OC13-BC and CD205-CR, presented significant difference (P < 0.05) within themselves.

As shown in Table 3, the populations – after family selection for high protein content - presented, as expected, a tendency of containing higher mean protein contents than those of their recurrent progenitors and the controls. This fact indicates the effectiveness of the selection for high protein content. While the available soybean seeds have a 40% protein mean (Sediyama et al. 1999), the mean protein content of the populations was 45.06%.

When the grain yield means of the recurrent progenitors, controls, and populations were compared (Table 3), it was observed that the populations presented lower grain yields than the recurrent progenitors and the controls, despite not having presented significant differences at 5% probability.

In Table 4, the heritability, the observed mean, the genetic variability, the range, the expected genetic gain and the expected mean - based on 20% selection intensity –

Table 1. Analysis of variance for the traits grain yield plant[-1] (YIELD) and seed protein contents (PROT)

Sources of variation[1]	df	Mean squares	
		YIELD	PROT
Blocks of C	19	16.29	1.25
Controls	2	63.72**	33.01**
Error	38	11.83	1.68
VC (%)		17.74	3.29
Progenitors	3	33.08	11.58**
Error	16	17.66	1.91
VC (%)		20.04	3.30
Among P, C and PP	2	17.46	39.14**
Error	12	20.56	1.94
VC (%)		23.74	3.23

[1]C = control varieties, P =progenitor varieties, and PP = populations; **significant at P< 0.01 probability, based on the F test

Table 2. Analysis of variance for the traits grain yield plant[-1] (YIELD) and seed protein contents (PROT) among and within the populations

Sources of variation	df	Mean squares		Variation coefficients (%)	
		YIELD	PROT	YIELD	PROT
Among populations	7	412.28**	29.40**	22.95	4.85
Error	135	15.00	4.82		
Within OC13-BC	16	8.17	1.96	20.21	3.04
OC13-CR	22	7.68	3.78**	20.39	3.00
CD201-BC	12	20.59	4.50**	18.47	3.10
CD201-CR	20	11.84	4.98**	20.08	3.02
CD205-BC	19	13.28	8.07**	20.40	3.04
CD205-CR	19	13.44	2.36	19.01	3.05
CD206-BC	10	33.17*	3.12*	17.40	3.10
CD206-CR	16	24.31**	9.04**	19.20	3.01

*, ** significant, respectively, at P<0.01 and P<0.05 probability, based on the F test

Table 3. Means of the recurrent progenitors, control varieties, and populations, regarding the traits grain yield plant[-1] (YIELD) and seed protein contents (PROT)

Genetic material	YIELD		PROT	
	—— g plant[-1] ——		—— % ——	
Progenitors	21.43	a[1]	41.85	b
Controls	19.39	a	39.39	b
Populations	17.82	a	45.06	a

[1]Means followed by the same letter, in the column, do not differ from each other, based on the Tukey test, at P < 0.05 probability

regarding the grain yield and seeds protein content of each population are presented.

The heritability ranged from 0 to 59.84% for yield and from 21.61 to 83% for protein content. The values found are close to the presented in the literature. Byth et al. (1969), working with F_6 and F_7 populations, in three environments, obtained values varying from 0 to 78%, for yield and between 71 and 90% for seed protein contents. These authors obtained smaller values of yield heritability when the families were grown under irrigation, which reduced the water stress. The low heritability estimates were ascribed to both the reduction of the genetic variance and increase of the environmental variance. Shannon et al. (1972) worked with F_4 families originated from crosses between materials with a high protein content and such of high yield. He obtained heritability values between 0 and 73% for yield and between 81 and 96% for protein. Brim and Burton (1979) found values between 20 and 37% for the realized heritability, regarding the protein seed content. The values found by these authors, lower than the comparable found in literature, are, as they explain, because these were calculated in realized sense. Santos et al.

(1995) found realized heritability of 57.68% for yield. Piovesan (2000), working with progenies F_2, obtained values between 17 and 56% for yield and between 22 and 81% for protein. Having obtained, for other traits, a broad-sense heritability zero, the author explains that this fact was due to the low variability of the progeny regarding to those other characters and/or, to the great influence of the environment on their expression, indicating also that the selection for them would be inefficient in that generation.

In general, the ranges of the families grain yield within each population were high, ascribed, however, not only to the genetic variability but also to the group of characteristics that influence the expression of this character, so that such a range is acceptable. The protein content range was reasonable and similar to the values presented in the literature (Shannon et al. 1972, Rebetzke et al. 1998), indicating the presence of considerable genetic variability within the populations. This variability may also be confirmed through the estimates of genetic variance, heritability, and selection gains. Smaller ranges for this character had been found by Wilcox (1998), who carried out recurrent selections with a selection intensity

of 20% for the cycles C_0 to C_3 and of 10% for cycles C_4 to C_7. During the eight cycles of recurrent selection, the mean protein content increased from 43.8% to 48.4%. From cycle C_5 to C_8, no increase in the minimum protein content was observed. The possible cause for the maintenance of these low limits was, according to the authors, the occurrence of casual pollination with plants from populations with low protein content.

The selection gains for yield ranged from 0 to 4.83 g plant[-1] and for protein content from 0.42 to 3.49%. For the two evaluated characters, the largest gains found are highly satisfactory. Population CD206-CR presented high and intermediate heritability, regarding the protein content and the grain yield, respectively (Table 4); though, it presented, as population CD206-BC, a high negative genotypic correlation between the cited traits (Table 5).

Population CD201-BC presented a high and positive genetic correlation for the two traits, which is extremely favorable for selection aiming at the simultaneous improvement of these characters. The other populations presented low genetic or, due to the absence of genetic variance, inexistent correlations. The phenotypic correlations were variable, however predominantly negative.

Table 6 shows the critical values for direct selection of families within each population as well as the percentage and the number of families selected in each one of them, regarding the traits protein content and grain yield. Regarding

the protein content, the selection criterion used was 0.84 times the standard deviation, added to the mean. This value, assuming a normal distribution, would result in a 20% selection of the families. In relation to grain yield, the selected families were those with yields equal or superior to their respective recurrent progenitors. Results indicate the possibility of selecting families for grain yield in all populations except CD201-CR. Simultaneous selection for protein and yield was possible in the populations OC13-BC, CD201-BC, CD205-CR, and CD206-BC. In spite of the low heritabilities, mainly for grain yield, their low magnitude indicates a limitation for the transfer of the phenotypic characters presented by the genitors to their progenies. In this context, the families selected in CD206-BC population are the ones that present a higher probability of transmitting these phenotypic characters to the offspring. Population CD206-CR Population, despite presenting good estimates of genetic parameters in relation to the separate two traits (Table 4), did not present families that meet, simultaneously, the two criteria used to select materials of high protein without losses in the grain yield (Table 6), due to the high and negative correlation between the two characters (Table 5). Due to the high mean of the protein content of this population, selection was not only tried with 20% (23.53%), but with all families that presented protein contents equal or superior to 48% (35.29% of the families); still, no line that would maintain at least the same yield as the recurrent progenitor met this new criterion.

Table 4. Means, genotypic variability (gv), heritability (H²), range, expected genetic gain (selection gain) - based on a selection intensity of 20% - and expected mean of the populations, regarding the traits grain yield plant[-1] (YIELD)[1] and seed protein contents (PROT)[1]

Population	Trait	mean	gv	H²	Range		Selection gain	Expected mean
OC13-BC	YIELD	13.01	0.00	0.00[2]	8.69	18.61	0.00	13.01
	PROT	45.70	0.42	21.61	43.63	48.71	0.42	46.47
OC13-CR	YIELD	14.01	0.00	0.00	8.21	19.37	0.00	14.01
	PROT	46.08	2.24	59.36	42.24	49.16	1.62	47.85
CD201-BC	YIELD	21.47	7.27	35.31	13.49	27.79	2.24	24.66
	PROT	43.03	2.97	65.87	39.11	47.43	1.96	45.07
CD201-CR	YIELD	14.68	0.00	0.00	8.31	20.22	0.00	14.68
	PROT	45.66	3.44	69.11	42.15	50.98	2.16	47.85
CD205-BC	YIELD	13.61	0.00	0.00	8.70	21.97	0.00	13.61
	PROT	44.81	6.53	80.95	39.03	49.56	3.22	47.83
CD205-CR	YIELD	18.61	0.12	0.86	11.06	23.49	0.04	19.01
	PROT	44.39	0.82	34.77	40.54	46.44	0.75	45.46
CD206-BC	YIELD	29.58	19.85	59.84	19.70	38.82	4.83	34.85
	PROT	43.50	1.58	50.76	40.61	46.51	1.26	44.99
CD206-CR	YIELD	17.63	10.99	45.20	8.93	28.20	3.12	21.55
	PROT	47.26	7.50	83.00	41.59	51.56	3.49	50.50
Means	YIELD	17.83	4.78	17.65	10.89	24.81	1.28	12.51
	PROT	45.05	3.19	58.18	41.11	48.79	1.86	47.00

[1]Yield in g plant[-1] and protein content in %.
[2]Heritability is zero since the estimate of the genetic variance was negative.

Table 5. Estimates of the coefficients of phenotypic and genotypic correlations between seed protein contents and grain yield in the analyzed soybean populations

Populations	Correlations	
	Phenotypic	Genotypic
OC13-BC	-0.27	_*
OC13-CR	-0.29	-
CD201-BC	0.35	0.83
CD201-CR	-0.41	-
CD205-BC	-0.49	
CD205-CR	-0.09	-0.14
CD206-BC	-0.21	-0.30
CD206-CR	-0.29	-0,42

*Families that presented a negative estimate of genetic variance for the character grain yield

Table 6. Number of families/population (f/p), percentage and number of selected families* within each population, regarding the traits grain yield (YIELD, in g plant[-1]) and seed protein contents (PROT, in %), simultaneously

Families	f/p	YIELD			PROT			PROT AND YIELD	
		Critical Values	%	Num fam selec.	Critical Values	%	Num fam selec.	%	Num fam selec.
OC13-BC	17	18.20	11.76	2	46.87	29.41	5	0.00	0
OC13-CR	23	18.20	8.70	2	47.71	17.39	4	4.35	1
CD201-BC	13	22.22	46.15	6	44.81	23.08	3	15.38	2
CD201-CR	21	22.22	0.00	0	47.49	18.18	4	0.00	0
CD205-BC	20	20.52	5.00	1	47.28	15.00	3	0.00	0
CD205-CR	20	20.52	40.00	8	45.68	15.00	3	5.00	1
CD206-BC	11	24.75	81.82	9	44.99	18.18	2	18.18	2
CD206-CR	17	24.75	5.88	1	49.79	23.53	4	0.00	0
Mean	17.8	21.42	27.63	4	46.83	19.46	3.43	5.36	0.75

*Direct selection of 20%, regarding the protein content, and equal or superior mean to the mean grain yield of the recurrent progenitor.

CONCLUSIONS

Analyzing the results, the following conclusions were drawn: i) the general means of the selected populations regarding the character seed protein contents were larger than the ones of the recurrent progenitors and of the control varieties; ii) the means of the lines tended to be lower than the ones of the recurrent progenitors and of the controls, regarding grain yield; iii) the CD206-CR family presented high and intermediate heritability, regarding the seed protein contents and grain yield, respectively; though, the high negative correlation between these characters made a simultaneous selection for both in this family impossible; and iv) in the OC13-CR, CD201-BC, CD205-CR, and CD206-BC populations the direct selection for protein content was possible, maintaining the grain yield means at least equal to the ones of the recurrent progenitors, even though a negative correlation was found between these characters in three of these populations.

ACKNOWLEDGEMENTS

Seleção de famílias de soja com alto teor protéico e elevada produtividade

RESUMO - *Objetivou-se mostrar a possibilidade de se selecionar famílias de soja que contenham alto teor de proteína nas sementes, sem prejuízos para a produção de grãos. Para tal, utilizaram-se quatro populações RC_1F_4 e quatro F_4 originárias, cada uma, do cruzamento de um material contendo alto teor protéico na semente e uma variedade comercial. Estimaram-se as médias, as herdabilidades, as correlações e os ganhos com seleção direta. As principais conclusões foram i) foi possível selecionar famílias que apresentassem ganhos de seleção positivos quanto ao teor protéico na semente, sem que houvesse redução na produção de grãos; ii) a população CD206-CR apresentou famílias com os maiores teores de proteína nas sementes. Embora tenha apresentado herdabilidade alta para esse caráter e intermediária para produção de grãos; a correlação negativa entre esses caracteres, presente nessa população, impossibilitou a seleção simultânea.*

Palavras-chave: seleção direta, herdabilidade, correlações, *Glycine max.*.

REFERENCES

Association of Official Analytical Chemistry - AOAC (1984) **Official methods of analysis**. 14th ed., Richmond Press, Byrd Gaithersburg, 1441p.

Brim CA and Burton JW (1979) Recurrent selection in soybeans II. Selection for increased percent protein in seeds. **Crop Science 19**: 494-498

Byth DE, Weber CR and Caldwell BE (1969) Correlated truncation selection for yield in soybeans. **Crop Science 9**: 699-702

Cruz CD (2001) **Programa Genes: versão Windows - aplicativo computacional em genética e estatística**. Editora UFV, Viçosa, CD ROM.

Fehr WR and Caviness CE (1977) **Stages of soybean development.** Iowa State University Press, Ames, 11p. (Co-operative Extension Service, Special Report 80).

Instituto Adolfo Lutz (1985) **Normas analíticas do Instituto Adolfo Lutz: métodos químicos e físicos para análise de alimentos.** 3rd ed., v1., Editora IAL, São Paulo, 533p.

Pimentel Gomes F (1984) **A estatística moderna na pesquisa agropecuária**. Editora POTAFOS, Piracicaba, 160p.

Piovesan ND (2000) **Aplicação de cruzamentos dialélicos no melhoramento genético do teor protéico em soja**. MSc. Thesis, Universidade Federal de Viçosa, Viçosa, 91p.

Rebetzke GJ, Burton JW, Carter Junior TE and Wilson RF (1998) Changes in agronomic and seed characteristics with selection for reduced palmitic acid content in soybean. **Crop Science 38**: 297-302.

Santos CAF, Reis MS, Sediyama CS, Cruz CD and Sediyama T (1995) Parâmetros genéticos e seleção indireta em progênies F_6 de um cruzamento de soja (*Glycine max* (L) Merril). **Revista Ceres 42**: 155-166.

Sediyama T, Pereira MG Sediyama CS and Gomes JLL (1996) **Cultura da soja-parte II**. Editora UFV, Viçosa, 75p.

Sediyama T, Teixeira RC and Reis MS (1999) Melhoramento da soja. In: Borém A (ed.) **Melhoramento de espécies cultivadas**. Editora UFV, Viçosa, p. 487-534.

Shannon JG, Wilcox JR and Probst AH (1972) Estimated gains from selection for protein and yield in the F_4 generation of six soybean populations. **Crop Science 12**: 824-830.

Thorne JC and Fehr WR (1970) Incorporation of high-protein exotic germplasm into soybean populations by 2- and 3-way crosses. **Crop Science 10**: 652-655.

Wehrmann VK, Fehr WR, Cianzio SR and Cavins JF (1987) Transfer of high seed protein to high-yielding soybean cultivars. **Crop Science 27**: 927-931.

Wilcox JR (1998) Increasing seed protein in soybean with eight cycles of recurrent selection. **Crop Science 38**: 1536-1540.