

# Correlation estimates and assessment of selection strategies in five soybean populations

Rogério Luiz Backes\*<sup>1</sup>; Múcio Silva Reis<sup>2</sup>; Cosme Damião Cruz<sup>3</sup>; Tuneo Sedyama<sup>2</sup> and Carlos Sigueyuki Sedyama<sup>2</sup>

<sup>1</sup>Universidade Federal de Viçosa, Pós-Graduação em Genética e Melhoramento, CEP 36571-000, Viçosa, MG, Brazil (Bolsista do CNPq); <sup>2</sup>Departamento de Fitotecnia, UFV, CEP 36571-000 Viçosa, MG, Brazil; <sup>3</sup>Departamento de Biologia Geral, UFV, CEP 36571-000, Viçosa, MG, Brazil. (\* Corresponding Author. E-mail: rogerio@backes.com.br)

## ABSTRACT

The objectives of the study were to analyze correlation among some important traits for breeding and assess three family selection strategies in soybean populations. Five soybean populations were assessed, three in the  $F_6$  generation and two in the  $F_5$  generation in Viçosa, MG, in the 1997/98 growing season. The experiment involved non replicated families inserted with replicated controls to allow estimation of the environmental component associated to the phenotypic variance of the studied populations. This arrangement allows the estimation of genetic parameters. The estimates of genetic correlation showed that the number of pods per plant is a good indicator of grain yield. The estimates of the gains from simulation of the three selection strategies showed that the direct selection on yield strategy resulted in greatest gains for the same trait but indirect selection via number of pods per plant resulted in gains close to those of direct selection. Selection based on the Pesek and Baker index enabled grain yield gains and simultaneous reduction in the cycle and plant height, despite the positive correlation of these two traits with grain yield.

**KEY WORDS:** *Glycine max*, genetic breeding, gain prediction.

## INTRODUCTION

One of the most important properties of a trait is its heritability. This is due to its predictive capacity, expressing the confidence of the phenotypic value as a guide for the genetic value. Prediction of selection gain before selection, serving as a subsidy to define the selection strategy, is a direct use of the heritability value. Correlation is another important parameter in plant breeding because it indicates how selection on one trait can cause simultaneous effects on others. Plant breeding aims to improve not only one but rather a set of traits so that a balance of desirable attributes is obtained. Thus the indirect effect of selection must be assessed. Another important use of correlation is indirect selection, when the use of direct selection is hindered by inexpressive heritability magnitudes and, or, by measuring difficulties. In these cases, correlated selection can be chosen on a trait with high heritability that is easily measured (Falconer, 1987; Cruz and Regazzi, 1997; Reis et al., 2001).

When the indirect response to selection is estimated by the expression that considers the genotypic correlation among the traits, quicker progress would sometimes be possible by correlated selection compared to that expected by direct selection on the

desired trait. However, the correlation among the data sampled do not always reflect the correlation among the traits in the selected group, especially when the number of selected unities is small and there is also the imprecision associated with components of variance through which the correlations are estimated. Furthermore, correlation is not a measure of cause and effect. Thus the comparison among the predicted gain and the observed gain for correlated responses are generally less consistent than for responses to direct selection (Falconer, 1987). Alternatively, Cruz and Regazzi (1997) suggested estimating the correlated response based on heritability of the indirectly selected trait and on the differential of indirect selection obtained on the individuals whose superiority is shown based on the auxiliary trait. With this methodology, the indirect gain will be at most equal to the gain by direct selection but never superior to it.

Selection based on selection indexes is an interesting alternative because it permits the combination of information from various traits simultaneously enabling the selection materials with a series of favorable attributes and the distribution of gains among the traits considered in the construction of the index (Cruz and Regazzi, 1997; Martins, 1999). The index

based on the desired improvement was proposed by Pesek and Baker (1969). This index has the advantage of not requiring the previous establishment of economic weights for the traits involved because few breeders would be prepared to define them. Vieira (1988) and Barbosa and Pinto (1998) reported good results with the Pesek and Baker index when using the desired improvement equal to a genetic standard deviation for each trait. Oliveira et al. (1999) also obtained satisfactory improvement on selection on an  $F_2$  soybean population using this index and desired gains equal to a genetic standard deviation.

Quantitative genetics permits progress prediction by applying a certain selection strategy and then comparing among different selection strategies and thus technical bases can be used to choose the one that provides the best results, without needing to use all the strategies in practice. It is known that the estimates are not exact because the models on which the estimates are based do not always explain the totality of the phenomena involved (Vencovsky, 1987). Pesek and Baker (1971) compared the predicted gain with the realized gain in five wheat populations. All the families were cultivated to estimate the realized gain and not only the selected ones. In four populations the gain surpassed the predicted and in one population the contrary occurred, but there was no significant difference between the predicted gain and the realized gain. These authors suggested that the deviation between the predicted gain and the realized gain was due to the errors associated to the heritability estimate. Reis et al. (1999) worked with a soybean population in the  $F_6$  and  $F_7$  generations and concluded that the choice among different strategies based on the predicted gains was efficient because the order of superiority of the different strategies was identical for predicted gain and realized gain. The realized gains were inferior to the predicted, that, according to the authors, was due to the different environmental conditions among the years in which the original population and the selected population were conducted.

The objective of this study was to analyze the correlation of some agronomic traits with grain production and compare three selection strategies in five soybean populations by direct and indirect gain prediction.

## MATERIAL AND METHODS

Five experiments were set up in this study as part of the Soybean Breeding Program at the Crop Science

Department of the Federal University of Viçosa, in the 1997/98 growing season. Five soybean populations were assessed, structured in families, three in the  $F_6$  generation and two in the  $F_5$  generation, derived from the following crosses, with the respective number of families: I) CEPS 77-16 x Doko RC - 84 families ( $F_6$ ); II) CEPS 89-26 x IAC-8 - 81 families ( $F_6$ ); III) CEPS 89-26 x FT-Cristalina - 75 families ( $F_6$ ); IV) Coker 6738 x FT-Cristalina RC4F<sub>4</sub> - 92 families ( $F_5$ ); e V) Agratech 550 x FT-Cristalina RC4F<sub>4</sub> - 81 families ( $F_5$ ). The previous generations of these populations were assessed by Reis et al. (2002).

The experiments were sown on 05-12-97 at the Federal University of Viçosa, MG. The soil was prepared by plowing and two gradings, and the sowing drill was fertilized with 70 kg/ha of  $P_2O_5$  and 35 kg/ha of  $K_2O$ . Spraying for pest control and irrigation was performed as needed by the crop.

The material was distributed on the experimental field according to the families with intercalated controls design presented by Cruz (2001) and Backes et al. (2002). The system of families with intercalated controls is a derivation of the augmented block design, presented by Federer (1956) where the controls are equivalent to the common treatments and the families to regular treatments. The parents were used as controls, except CEPS 77-16 in population I and CEPS 89-26 in populations II and III that were substituted by the BR-16 cultivar. However, this presented a similar regional adaptation. The experimental design was the following: in the first row of each experiment (1<sup>st</sup> plot) a control was planted (father 1), families were planted (progeny from a selected plant) in the next five lines (2<sup>nd</sup> to 6<sup>th</sup> plot); the next row (7<sup>th</sup> plot) another control (father 2) and in sequence another five rows were sown with families (8<sup>th</sup> to 12<sup>th</sup> plot); and a row was again repeated (13<sup>th</sup> plot) with the first control (father 1) and so on until all the selected plants were sown.

The plot consisted of one 1m row with 0.7m spacing between rows. Thirty seeds were used per plot at planting and 15 to 20 days after emergence they were thinned to 15 plants per meter. For the assessment, six competitive plants were labeled within each plot.

The following traits were assessed: number of days to flowering (DF); plant height at flowering in cm (PHF); number of nodes at flowering (NNF); number of days to maturity (R8 stage) (DM); plant height at maturity (PHM); number of nodes at maturity (NNM); number of pods per plant (NPP); number of

seeds per pod (NSP); mean weight of 100 seeds (WHS) and grain yield of each plant in grams (PRO).

The analysis of variance of each experiment was performed considering the following statistical model, as presented by Cruz (2001) and Backes et al. (2002):

$$Y_{ij} = \mu + f_i + e_i + d_{ij} + \delta_{ij}$$

where:  $\mu$  = general mean (controls or families);  $f_i$  =

genetic effect attributed to the  $i$ -eth family (does not exist for controls);  $e_i$  = environmental effect among rows (on controls or families);  $d_{ij}$  = genetic effect attributed to the  $j$ -eth plant of the  $i$ -eth family (does not exist for the controls); and  $\delta_{ij}$  = environmental effect among plants within the row (of a controls or families).

The estimator of the phenotypic variance among families ( $\hat{\sigma}_{Fb}^2$ ) was the MSBf and the estimator of within phenotypic variance ( $\hat{\sigma}_{Fw}^2$ ) was the MSWf of the following analysis of variance design:

SV	Families		Parent 1		Parent 2	
	DF <sup>1/</sup>	MS <sup>3/</sup>	DF <sup>2/</sup>	MS <sup>3/</sup>	DF <sup>2/</sup>	MS <sup>3/</sup>
Between plots	$f - 1$	MSBf	$r_1 - 1$	MSBp1	$r_2 - 1$	MSBp2
Within plots	$fn - f$	MSWf	$r_1p - r_1$	MSWp1	$r_2p - r_2$	MSWp2
Total	$fn - 1$		$r_1p - 1$		$r_2p - 1$	

<sup>1/</sup>  $f$ : number of families in assessment; and  $n$ : number of plants per family; <sup>2/</sup>  $r_1$  and  $r_2$ : number of parent replications of parent 1 and 2, respectively, and  $p$ : number of plants per parent; <sup>3/</sup> MSB: mean square between families ( $f$ ) or among replications of the parents ( $p1$  and  $p2$ ); and MSW: mean square among plants within families ( $f$ ) or among plants within the replication of the parents ( $p1$  and  $p2$ ).

The environmental variance was estimated based on the phenotypic variance among the replication of the control intercalated in the families. Thus the environmental variance between ( $\hat{\sigma}_{Eb}^2$ ) and within ( $\hat{\sigma}_{Ew}^2$ ) families was estimated, respectively, by:

$$\hat{\sigma}_{Eb}^2 = \frac{(r_1 - 1)MSBp1 + (r_2 - 1)MSBp2}{r_1 + r_2 - 2} \quad \text{and}$$

$$\hat{\sigma}_{Ew}^2 = \frac{(r_1p - r_1)MSWp1 + (r_2p - r_2)MSWp2}{p(r_1 + r_2) - r_1 - r_2}$$

From difference, according to Ramalho and Vencovsky (1978) the genotypic variance was obtained between ( $\hat{\sigma}_{Gb}^2$ ) and within ( $\hat{\sigma}_{Gw}^2$ ) families, respectively, by:

$$\hat{\sigma}_{Gb}^2 = \hat{\sigma}_{Fb}^2 - \hat{\sigma}_{Eb}^2 \quad \text{and} \quad \hat{\sigma}_{Gw}^2 = \hat{\sigma}_{Fw}^2 - \hat{\sigma}_{Ew}^2$$

The genotypic variances were partitioned in additive variance ( $\hat{\sigma}_A^2$ ) and due to the dominance ( $\hat{\sigma}_D^2$ ) by the expression of genotypic variance distribution between and within selfed families quoted by Falconer (1987):

$$\hat{\sigma}_{Gb}^2 = 2F_n \hat{\sigma}_A^2 + F_n(1 - F_n) \hat{\sigma}_D^2 \quad \text{and}$$

$$\hat{\sigma}_{Gw}^2 = (1 - F_n) \hat{\sigma}_A^2 + (1 - F_n) \hat{\sigma}_D^2$$

considering the coefficient of endogamy  $F_n$  in the  $F_5$  and  $F_6$  generation equal to  $7/8$  and  $15/16$ , respectively, according to Ramalho and Vencovsky (1978).

The coefficient of genetic variation ( $CV_g$ ) and narrow sense heritability for family means ( $h^2$ ) were estimated, respectively by:

$$CV_g = \left( \sqrt{\hat{\sigma}_{Gb}^2 / \bar{X}} \right) * 100 \quad \text{and} \quad h^2 = (2F_n \hat{\sigma}_A^2 / \hat{\sigma}_{Fb}^2)$$

where  $\bar{X}$  is the mean of the families for the trait in question.

The phenotypic covariance among means of the same family for the traits X and Y was estimated by analysis of variance of each characteristic and their sum:

$$\text{cov}(\bar{F}_x, \bar{F}_y) = (MSBf_{(x+y)} - MSBf_{(x)} - MSBf_{(y)}) / 2$$

Based on each one of the two controls intercalated in each experiment a covariance was estimated ( $\text{cov}_{p_i}$ ), similarly the estimation of the phenotypic variance. The means weighted by the degrees of freedom ( $j1$  and  $j2$ ) of the covariance among the traits of each control was used as an estimate of the environmental covariance among family means for the X and Y traits, as follows:

$$\text{cov}(\bar{E}_x, \bar{E}_y) = \frac{j_1 \text{cov}_{p_1}(\bar{F}_x, \bar{F}_y) + j_2 \text{cov}_{p_2}(\bar{F}_x, \bar{F}_y)}{j_1 + j_2}$$

where:  $j_1 = r_1 - 1$  and  $j_2 = r_2 - 1$ .

The genotypic covariance for family means was obtained by the difference between the phenotypic and environmental covariance. The correlations were estimated conventionally by the quotient between the covariance (X,Y) and the product of the standard deviation of the traits X and Y.

Direct and indirect gains were estimated resulting from the simulation of three selection strategies, fixing the selection percentage at 20% of the families from each population:

- 1: selection of the families that presented greatest grain yield mean (PRO);
- 2: selection of families with greatest mean of number of pods per plant (NPP); and
- 3: selection among families based on the Pesek and Baker (1969) index considering the PRO, DM and PHM traits simultaneously. Because the gains by selection are limited by the genotypic constitution of

the population, equivalent desired gains and a genetic standard deviation for each trait were used in the index estimation, as recommended by Vieira (1988) and Cruz and Regazzi (1997).

By analysis of the correlation among the traits assessed, NPP was chosen as suitable for indirect selection on the most productive families. For the selection strategy based on the index, the DM and PHM traits and PRO were considered to assess the possibility of obtaining gains in yield and simultaneously, reduction in population height and cycle. For this, the desired improvements with negative sign were considered for PHM and DM.

Direct responses were estimated by:  $GS = DS * h^2$ , where DS is the selection differential. The indirect responses or correlated gains were estimated as suggested by Cruz and Regazzi (1997):  $GS_{x(y)} = DS_{x(y)} * h_{(x)}^2$ , where  $GS_{x(y)}$  is the response in X when selection is practiced on Y;  $DS_{x(y)}$  is the selection differential of the X trait when selection based on the Y trait, and  $h_{(x)}^2$  is the narrow sense heritability for family means for trait X.

The statistical analyses of this study were processed

**Table 1.** Estimates of the coefficients of phenotypic (P), genotypic (G) and environmental (E) correlation of grain yield with another nine traits assessed in five soybean populations.

Correlation	Traits <sup>1/</sup>								
	DF	PHF	NNF	DM	PHM	NNM	NPP	NSP	WHS
<b>Population I – ‘CEPS 77-16’ x ‘Doko RC’</b>									
P	-0.0369	0.1769	0.3180 <sup>3/</sup>	-0.2126	0.2395 <sup>3/</sup>	0.3258 <sup>3/</sup>	0.9440 <sup>3/</sup>	0.4016 <sup>3/</sup>	-0.0059
G	-0.0742	0.3009 <sup>3/</sup>	0.4110 <sup>3/</sup>	-0.2276 <sup>3/</sup>	0.3674 <sup>3/</sup>	0.4328 <sup>3/</sup>	0.9750 <sup>3/</sup>	0.2797 <sup>3/</sup>	-0.0188
E	0.0328	-0.1437	0.1329	-0.1965	-0.0977	0.1216	0.8987 <sup>3/</sup>	0.5331	0.0090
<b>Population II – ‘CEPS 89-26’ x ‘IAC -8’</b>									
P	-0.0419	0.1458	0.2484 <sup>3/</sup>	-0.0575	0.1212	0.1749	0.9202 <sup>3/</sup>	0.4674 <sup>3/</sup>	0.2430 <sup>3/</sup>
G	-0.0116	0.1739	0.4263 <sup>3/</sup>	0.0265	0.1632	0.3602 <sup>3/</sup>	0.9399 <sup>3/</sup>	0.7643 <sup>3/</sup>	0.3997 <sup>3/</sup>
E	-0.1647	0.0850	-0.2422	-0.2114	-0.0198	-0.1358	0.8867 <sup>3/</sup>	0.2345	0.2152
<b>Population III – ‘CEPS 89-26’ x ‘FT-Cristalina’</b>									
P	0.0112	0.0787	0.0266	-0.1646	0.0319	-0.0048	0.8950 <sup>3/</sup>	0.3941 <sup>3/</sup>	0.1498
G	0.0656	-0.0164	-0.1370	-0.0118	0.0208	-0.0140	0.8946 <sup>3/</sup>	0.3190 <sup>3/</sup>	-0.4465 <sup>3/</sup>
E	-0.0957	0.3543	0.2570	-0.2994	0.0681	0.0068	0.9097 <sup>3/</sup>	0.4620	0.1988
<b>Population IV – ‘Coker 6738’ x ‘FT-Cristalina RC4F<sub>4</sub>’</b>									
P	0.1149	0.2557 <sup>3/</sup>	0.2978 <sup>3/</sup>	0.0960	0.1882	0.1996	0.8823 <sup>3/</sup>	0.4267 <sup>3/</sup>	0.0103
G	0.0517	0.2304 <sup>3/</sup>	0.3011 <sup>3/</sup>	0.0581	0.1568	0.1474	0.9227 <sup>3/</sup>	1.4370 <sup>2/3/</sup>	-0.1411
E	0.4929	0.3932	0.2795	0.2328	0.3373	0.4509	0.5860	0.1244	0.2949
<b>Population V – ‘Agratech 550’ x ‘FT-Cristalina RC4F<sub>4</sub>’</b>									
P	0.1811	0.3425 <sup>3/</sup>	0.3088 <sup>3/</sup>	0.1793	0.3309 <sup>3/</sup>	0.2667 <sup>3/</sup>	0.7606 <sup>3/</sup>	0.3254 <sup>3/</sup>	0.0451
G	0.1890	0.4030 <sup>3/</sup>	0.1811	0.7437 <sup>3/</sup>	0.4808 <sup>3/</sup>	0.2438 <sup>3/</sup>	0.5887 <sup>3/</sup>	0.2005	0.0629
E	0.4496	0.4354	0.7412	-0.2546	0.2010	0.3521	0.8433 <sup>3/</sup>	0.4014	0.0308

<sup>1/</sup> DF: number of days to flowering; PHF: plant height at flowering; NNF: number of nodes at flowering; DM: number of days to maturity; PHM: plant height at maturity; NNM: number of nodes at maturity; NPP: number of pods per plant; NSP: number of seeds per pod; WHS: mean weight of 100 seeds; <sup>2/</sup> Estimate of real value 1.0; <sup>3/</sup> Coefficients of correlation significant at 5% probability by the t test.

with the GENES computer software, developed by Cruz (2001).

## RESULTS AND DISCUSSION

The traits DF, PHF, NNF, DM, PHM and NNM presented generally low phenotypic, genotypic and environmental correlations with grain yield, lower than 0.50 except for the genotypic correlation of DM in population V (Table 1).

Therefore, these are not helpful for indirect selection on the most productive families. However, they inform about alterations that may occur in the other traits in consequence of selection for yield. DF and

DM presented alteration in the sign of the coefficients of correlation with PRO in the different populations. However, the other coefficients were not significant except for the genotypic correlation of DM with PRO in populations I and V. The traits PHF, NNF and NNM presented positive and significant genotypic correlation with yield in population I. Only NNF and NNM in population II, PHF and NNF in population IV and PHF and NNM in population V were positively correlated with yield. The genotypic correlation between DM and PRO, negative in population I and nil in population III (Table 1), is a favorable situation for selection of earlier plants with good yield as shown by the predicted gain in the selection on yield (Table 2).

**Table 2.** Estimates of coefficients of genetic variance (CVg), narrow sense heritability for means of families ( $h^2$ ) and mean of eight traits, and the mean of the selected (MS) and gains by selection (GS) resulting from the simulation of three selection strategies in soybean  $F_6$  populations.

Trait <sup>1/</sup>	CVg (%)	$h^2$	Mean	Selection strategies								
				PRO based selection			NPP based selection			Pesek e Baker index		
				MS	GS	GS%	MS	GS	GS%	MS	GS	GS%
<b>Population I – ‘CEPS 77-16’ x ‘Doko RC’</b>												
DF	16.75	70.48	66.38	66.47	0.07	0.10	68.63	1.59	2.39	61.04	-3.76	-5.67
PHF	42.31	83.15	90.13	95.14	4.17	4.62	99.94	8.16	9.06	75.08	-12.51	-13.9
NNF	23.88	75.72	14.95	15.65	0.53	3.54	16.27	1.00	6.72	14.08	-0.66	-4.41
DM	9.13	77.73	142.67	140.59	-1.62	-1.13	142.98	0.24	0.17	137.11	-4.32	-3.03
PHM	52.94	85.21	113.11	123.33	8.71	7.70	128.77	13.35	11.80	94.36	-15.98	-14.10
NNM	32.24	73.09	17.59	18.83	0.91	5.18	19.69	1.53	8.72	16.32	-0.92	-5.25
NPP	63.25	51.86	76.04	110.73	17.99	23.66	112.80	19.07	25.07	94.88	9.77	12.85
PRO	64.24	48.55	18.64	28.27	4.68	25.09	27.76	4.43	23.76	24.86	3.02	16.20
<b>Population II – ‘CEPS 89-26’ x ‘IAC -8’</b>												
DF	22.91	88.99	63.74	64.60	0.77	1.21	64.73	0.88	1.38	57.35	-5.68	-8.92
PHF	56.90	95.97	84.50	91.77	6.98	8.26	90.19	5.46	6.46	63.85	-19.81	-23.45
NNF	17.18	76.72	12.67	13.20	0.41	3.20	13.23	0.43	3.39	12.50	-0.13	-1.02
DM	9.62	66.08	133.43	134.19	0.50	0.37	134.60	0.77	0.58	127.49	-3.93	-2.94
PHM	50.79	92.18	109.01	116.88	7.25	6.65	115.21	5.71	5.24	83.14	-23.85	-21.88
NNM	13.06	55.79	15.80	16.22	0.24	1.49	16.38	0.32	2.04	15.34	-0.25	-1.60
NPP	41.17	32.78	56.29	73.92	5.78	10.26	74.79	6.06	10.77	64.73	2.77	4.91
PRO	49.77	37.78	15.36	21.35	2.26	14.72	20.95	2.11	13.74	18.94	1.35	8.80
<b>Population III – ‘CEPS 89-26’ x ‘FT-Cristalina’</b>												
DF	12.61	82.59	62.52	62.58	0.05	0.07	63.36	0.69	1.10	60.60	-1.59	-2.54
PHF	46.71	87.61	78.48	83.21	4.15	5.29	84.13	4.96	6.32	71.51	-6.10	-7.78
NNF	21.68	66.76	14.76	14.96	0.13	0.88	15.14	0.26	1.74	14.23	-0.35	-2.38
DM	6.93	44.48	139.35	137.94	-0.63	-0.45	139.60	-0.34	-0.24	133.79	-2.48	-1.78
PHM	35.92	83.87	101.14	104.70	2.98	2.95	104.10	2.48	2.45	91.27	-8.28	-8.19
NNM	18.63	57.91	17.76	17.80	0.03	0.14	17.83	0.05	0.25	16.89	-0.50	-2.83
NPP	50.83	47.09	71.48	99.82	13.18	18.34	101.79	14.10	19.63	95.97	11.36	15.82
PRO	44.97	32.85	17.61	25.08	2.45	13.93	24.22	2.17	12.32	23.47	1.93	10.93

<sup>1/</sup>DF: number of days to flowering; PHF: plant height at flowering; NNF: number of nodes at flowering; DM: number of days to maturity; PHM: plant height at maturity; NNM: number of nodes at maturity; NPP: number of pods per plant; and PRO: grain yield of each plant.

Only NPP among the primary yield components presented positive environmental correlation with yield and these coefficients were significant except for population IV. This positive association indicates that the environmental variations favor or harm these traits simultaneously, that is, the environmental variation simultaneously intensifies or reduces the expression of these traits. The correlations of WHS with PRO were low magnitude and presented sign alteration among the different populations, mainly in genotypic correlation. The phenotypic and genotypic correlations of NPP and NSP with PRO were positive in all the populations and the other coefficients were significant except for the genotypic correlation of NPP with PRO in population V. The discrepancy in the phenotypic correlation of NPP with PRO was relatively small among the populations, ranging from 0.32 to 0.46. However, for the genotypic correlation between NSP and PRO, the discrepancy was greater between the different populations, ranging from 0.20 to 1.43 (Table 1). The 1.43 coefficient should be interpreted as an estimate of the real value 1.0. Estimates of genetic correlation above the unit are

attributed to errors associated to the components of variance used in the estimate. NPP was most consistently correlated genetically and phenotypically with PRO in all the populations although population V showed estimates a little lower than the others. NPP is therefore useful in selection for yield, because phenotypic selection of the families with greater NPP will have good correspondence with selection of the families also genotypically superior for the trait, as is expressed by the heritability (Tables 2 and 3). Whereas, because of the correlation between the genotypic value for NPP and PRO, in the same family, the selection of families genotypically superior for NPP will have good correspondence with the selection of families with superior genotypes for PRO, according to the magnitude of the genotypic correlation between these traits. However, this hypothesis cannot be applied to population V because of NPP low heritability (Table 3) and the lower magnitude of the genotypic correlation between NPP and PRO (Table 1). Nevertheless, when the indirect gains are estimated by the indirect selection differential, as suggested by Cruz and Regazzi (1997)

**Table 3.** Estimates of coefficients of genetic variation (CVg), narrow sense heritability for the means of families ( $h^2$ ) and mean of eight traits, besides the mean of the selected (MS) and gains by selection (GS) resulting from the simulation of three selection strategies in  $F_5$  soybean populations.

Trait <sup>1/</sup>	CVg (%)	$h^2$	Mean	Selection strategies								
				PRO based selection			NPP based selection			Pesek e Baker index		
				MS	GS	GS%	MS	GS	GS%	MS	GS	GS%
<b>Population IV – ‘Coker 6738’ x ‘FT-Cristalina RC4F<sub>4</sub>’</b>												
DF	26.63	83.07	62.64	63.65	0.84	1.34	64.64	1.66	2.65	55.74	-5.73	-9.15
PHF	44.08	73.64	70.26	74.41	3.06	4.35	75.91	4.16	5.92	59.69	-7.79	-11.08
NNF	36.37	84.63	14.74	15.71	0.82	5.59	15.81	0.90	6.12	1.94	-1.52	-10.30
DM	8.39	49.33	139.82	141.63	0.89	0.64	142.60	1.37	0.98	130.76	-4.47	-3.20
PHM	41.35	59.42	85.69	87.71	1.20	1.40	91.04	3.18	3.71	73.57	-7.20	-8.40
NNM	36.80	73.10	17.18	17.86	0.50	2.91	18.29	0.81	4.72	15.01	-1.59	-9.23
NPP	63.92	74.30	67.30	87.18	14.76	21.94	93.93	19.78	29.39	66.69	-0.45	-0.67
PRO	63.62	69.74	16.00	21.95	4.15	25.96	20.95	3.45	21.59	17.16	0.81	5.08
<b>Population V – ‘Agratech 550’ x ‘FT-Cristalina RC4F<sub>4</sub>’</b>												
DF	27.44	96.27	57.88	58.84	0.93	1.61	62.58	4.53	7.83	54.23	-3.51	-6.07
PHF	48.85	89.49	70.57	74.11	3.17	4.49	81.29	9.59	13.59	65.45	-4.59	-6.50
NNF	32.79	87.81	13.71	14.25	0.47	3.46	15.16	1.27	9.26	13.08	-0.55	-4.01
DM	9.12	50.67	136.41	138.24	0.93	0.68	140.65	2.14	1.57	130.09	-3.20	-2.35
PHM	51.36	81.00	91.90	96.42	3.66	3.98	105.03	10.64	11.58	83.95	-6.44	-7.01
NNM	34.00	75.55	16.69	17.31	0.47	2.82	18.36	1.27	7.58	15.64	-0.80	-4.77
NPP	27.22	19.10	65.68	81.80	3.08	4.69	87.71	4.21	6.41	70.19	0.86	1.31
PRO	35.71	16.43	16.60	22.45	0.96	5.78	21.32	0.78	4.67	20.51	0.64	3.87

<sup>1/</sup> DF: number of days to flowering; PHF: plant height at flowering; NNF: number of nodes at flowering; DM: number of days to maturity; PHM: plant height at maturity; NNM: number of nodes at maturity; NPP: number of pods per plant; and PRO: grain yield of each plant.

it is the phenotypic correlation that will determine the relative efficiency of the indirect selection. Thus the indirect response to PRO by selection via NPP should be satisfactory in all the populations because these traits are highly correlated phenotypically.

Indirect gain prediction indicates that selection with any one of the three strategies should cause alterations in all the traits, in all the populations, that reflects the correlation among the traits (Tables 2 and 3). The PRO trait presented greater predicted gain when selection was on the trait itself, in all the populations (Tables 2 and 3). However, it was observed that indirect selection for PRO via NPP presented gain estimates quite close to those of direct selection, with 81.25 to 94.66% relative efficiency. That is, when selection was made based on the NPP trait, the predicted gain in PRO by indirect selection was at least 81% of that obtained by direct selection indicating, as already expected based on the correlation among the two traits, that NPP is a good indicator of superior yield. However, considering PRO, direct selection is more efficient than correlated selection. Table 4 shows that many families were selected simultaneously by the strategy that considered PRO and that which considered NPP mainly in populations I, II and III. These are exactly the three populations that present the highest phenotypic correlations between PRO and NPP.

The ranking of populations based on the superiority of the percentage gains predicted for PRO by direct selection is IV, I, II III and V. The differences in the percentage gains are because of the differences in the genetic variability available in the population and the accuracy in identifying the genetically superior individuals, that is, heritability. In the among populations comparison, the possibility of obtaining gains by selection through the existence of genetic

variability should be considered in addition to the original mean. Starting from a population with a comparatively low mean, but with greater genetic variability, it can in a few selection cycles surpass others that initially presented better performance but lower genetic variability. Population I (Table 2), besides the greater initial mean for PRO (18.64g), presented the greatest coefficient of genetic variation, also giving the greatest predicted mean for the next generation (23.32g) among all the populations. Population IV (Table 3) had the fourth best PRO mean before selection (16.00g) but its high coefficient of genetic variation showed that it has a good potential for exploitation by selection. In this cycle it was the population that presented the highest predicted percentage gain, remembering that it also presented the greatest heritability coefficient for PRO. Thus, with only one selection cycle, population IV contained the second best predicted mean for the next generation (20.15g), and was only surpassed by population I.

The traits DF in population I, DF and DM in population II and DM and NNM in population III presented negative but not significant phenotypic correlations with PRO (Table 1) but the estimates of indirect gains for these traits were positive when for selection to increase PRO (Table 2). This finding may indicate that the estimated correlation in the population does not correspond to the correlation in the selected group. Thus the prediction of indirect gains based on the differentials of indirect selection as proposed by Cruz and Regazzi (1997) was appropriate because it did not use correlation to estimate the indirect gains, mainly for populations I, II and III. In spite of the non significance of the coefficients of correlation among PRO and DM, in populations I and III (Table 1) it was possible to select higher yielding and earlier families compared to the original population by direct selection for PRO even though the reduction in the cycle was very small (Table 2).

**Table 4.** Number of families that were simultaneously selected by different selection strategies simulated in five soybean populations.

Selection strategy <sup>1/</sup>	Population				
	I (17 families) <sup>2/</sup>	II (16 families)	III (15 families)	IV (18 families)	V (16 families)
S-PRO e S-NPP	14	13	12	10	10
S-PRO e S-IPB	8	7	9	4	8
S-NPP e S-IPB	6	8	9	2	6

<sup>1/</sup> S-PRO = direct selection on grain yield; S-NPP = selection on number of pods per plant; and S-IPB = selection via Pesek and Baker index; <sup>2/</sup> The values in parentheses indicate the total number of families selected in each population.

Considering the need to improve other traits besides PRO, selection based on the Pesek and Baker index was performed considering the traits DM, PHM and PRO in its construction and desired improvement equal to a standard genetic deviation for each trait. The objective was to reduce the cycle and height; for this, desired gains with negative sign were used for these two traits. Besides reducing PHM, increased resistance to lodging was expected, mainly in populations I, II and III where the predicted means for the next generation are above 100cm when of the selection on PRO and plants over 100cm tall tend to lodge more (Sediyama et al., 1999).

The predicted gains by selection based on the index (Tables 2 and 3) show that this strategy may be efficient in promoting PRO increase and simultaneous decrease in the PHM and DM traits, in the five populations, as required, in spite of the positive correlation between PRO and PHM especially in populations I and V and the positive correlation between DM and PRO in population V (Table 1). Selection was more efficient in reducing PHM than in reducing the cycle, because, in all the populations, PHM presented superior heritability estimates than DM. Therefore, obtaining gain by selection for PHM was facilitated. Traits DF, PHF, NNF and NNM also had negative gains with selection based on the index as shown in Tables 2 and 3.

Selection based on the index should give gain for grain yield (PRO) in all the populations but it should be less than the gain obtained with direct selection. The relative efficiencies of selection based on the index compared to direct selection for PRO were 78.8, 66.7, 64.5, 59.7 and 19.5 % for populations III, V, I, II and IV, respectively, as shown in Tables 2 and 3. Table 4 shows that the number of families simultaneously selected based on the index and on other strategies is much less than the number of individuals simultaneously selected by direct selection on PRO and via NPP. In population IV, the index presented the worst relative efficiency for PRO but it was exactly in this population that the least number of plants were simultaneously selected by direct selection on PRO and that based on the index. Direct or truncated selection on a single trait must always produce the best possible responses to selection on this same trait (Falconer, 1987).

The use of the selection index will be viable if the reduction in the height and cycle compensate the reduction in the yield gains in comparison with direct selection. As each population has different

characteristics, the decision will be specific for each population. In populations I and II where PHM was 113.1 and 109.01cm, respectively, and the predicted indirect gain with selection on yield was 8.71cm in population I and 7.25cm in population II thus the predicted PHM for the next generation is 121.82 and 116.26cm. Due to the tendency to lodging, it would be desirable to select shorter families, that was obtained in the strategy based on the Pesek and Baker (1969) index. Although selection based on the index restricted the PRO gains, this strategy is the most suitable for population I and II because it results in populations with a predicted PHM of 97.13 and 85.16cm, respectively. In the other populations, PHM is not restrictive and direct selection on PRO was the most viable strategy. Thus the index was efficient in the sense of permitting simultaneous gain in more than one trait even when the correlation presented sign contrary to the direction of the intended gains. This was the case, for example, of PHM and PRO among which the correlation was positive (Table 1) and even so it was possible to obtain gain for PRO and decrease in PHM. Obtaining simultaneous gains in various traits in selection, as required, with the use of indexes has also been reported by Oliveira et al. (1999), Martins (1999) and Barbosa and Pinto (1998).

## CONCLUSIONS

Direct selection for grain yield is the strategy that results in the best gains estimates for the same trait.

In the populations studied, the number of pods per plant trait (NPP) was the most consistently correlated with grain yield, showing that it can be used for indirect selection, as proved by the estimates of gain from selection via NPP.

Selection based on the Pesek and Baker index is efficient in promoting simultaneous gain in grain yield and reduction in the cycle and plant height, even where the correlation among these traits is positive.

## RESUMO

### Estimativas de correlações e avaliação de estratégias de seleção em cinco populações de soja

Os objetivos deste estudo foram analisar as correlações entre alguns caracteres de importância para o melhoramento e avaliar três estratégias de



seleção de famílias em populações de soja. Para tanto, avaliaram-se cinco populações de soja, três na geração  $F_6$  e duas na geração  $F_5$ , em Viçosa-MG, no ano agrícola 1997/98. Utilizou-se o esquema experimental de famílias sem repetições, intercaladas com testemunhas, com repetições, possibilitando a estimação do componente ambiental associado à variância fenotípica das populações em estudo e, conseqüentemente, dos parâmetros genéticos. As estimativas de correlações genéticas revelaram que o número de vagens por planta é um bom indicador da produção de grãos. As estimativas dos ganhos resultantes da simulação de três estratégias de seleção mostraram que a seleção direta na produção foi a estratégia que resultou em maiores ganhos para o mesmo caráter, mas a seleção indireta via número de vagens por planta resultou em ganhos próximos a seleção direta. A seleção baseada no índice de Pesek e Baker possibilitou ganhos na produção de grãos e, simultaneamente, redução no ciclo e na altura das plantas, apesar da correlação positiva desses dois caracteres com a produção de grãos.

## REFERENCES

- Backes, R.L.; Reis, M.S.; Sedyama, T.; Cruz, C.D. and Teixeira, R.C. 2002. Estimativas de parâmetros genéticos em populações  $F_5$  e  $F_6$  de soja. *Revista Ceres*. 49:201-216.
- Barbosa, M.H.P. and Pinto, C.A.B.P. 1998. Eficiência de índices de seleção na identificação de clones de batata. *Pesquisa Agropecuária Brasileira*. 33:149-156.
- Cruz, C.D. and Regazzi, A.J. 1997. Modelos biométricos aplicados ao melhoramento genético. Ed. UFV, Viçosa.
- Cruz, C.D. 2001. Programa GENES – Versão Windows - Aplicativo computacional em genética e estatística. Ed. UFV, Viçosa.
- Falconer, D.S. 1987. Introdução à genética quantitativa. Ed. UFV, Viçosa.
- Federer, W.T. 1956. Augmented (or Hoonuiaku) designs. *Hawaiian Planters Record*. 55:191-208.
- Martins, I.S. 1999. Comparação entre métodos uni e multivariados aplicados na seleção em *Eucalyptus grandis*. D.S. Thesis. Universidade Federal de Viçosa, Viçosa.
- Oliveira, A.C.B.; Sedyama, C.S. and Cruz, C.D. 1999. Selection for later flowering in soybean [*Glycine max* (L.) Merrill]  $F_2$  populations cultivated under short days conditions. *Genetics and Molecular Biology*. 22:243-247.
- Pesek, J. and Baker, R.J. 1969. Desired improvement in relation to selected indices. *Canadian Journal of Plant Science*. 49:803-804.
- Pesek, J. and Baker, R.J. 1971. Comparison of predicted and observed responses to selection for yield in wheat. *Canadian Journal of Plant Science*. 51:187-192.
- Ramalho, M.A.P. and Vencovsky, R. 1978. Estimação dos componentes da variância genética em plantas autógamas. *Ciência e Prática*. 2:117-140.
- Reis, E.F.; Reis, M.S.; Cruz, C.D.; Sedyama, T. and Sedyama, C.S. 1999. Seleção e eficiência da predição de ganhos, tendo em vista a produção de grãos, em uma população de soja (*Glycine max* (L.) Merrill). *Revista Ceres*. 46:411-422.
- Reis, E.F.; Reis, M.S.; Sedyama, T. and Cruz, C.D. 2002. Estimativas de variâncias e herdabilidades de algumas características primárias e secundárias da produção de grãos em soja (*Glycine max* (L.) Merrill). *Ciência e Agrotecnologia*. 26:749-761.
- Reis, E.F.; Reis, M.S.; Cruz, C.D.; Sedyama, T. and Sedyama, C.S. 2001. Resposta esperada à seleção correlacionada em uma população  $F_6$  de soja. *Revista Ceres*. 48:169-179.
- Sedyama, T.; Teixeira, R.C. and Reis, M.S. 1999. Melhoramento da soja. p.487-533. In: Borém, A. Melhoramento de espécies cultivadas. Ed. UFV, Viçosa.
- Vencovsky, R. 1987. Herança quantitativa. p.137-214. In: Paterniani, E. and Viegas, G.P. Melhoramento e produção do milho no Brasil. Fundação Cargill, Campinas.
- Vieira, J.V. 1988. Herdabilidades, correlações e índices de seleção em populações de cenoura (*Daucus carota* L.). M.S.Thesis. Universidade Federal de Viçosa, Viçosa.

Received: July 03, 2002;

Accepted: February 11, 2003.