Genetic and phenotypic parameters associated with soybean progenies in a recurrent selection program

Igor Oliveri Soares¹, Mariane Cristina Bianchi¹, Adriano Teodoro Bruzi¹, Gabriel de Siqueira Gesteira², Karina Barroso Silva¹, Scheila Roberta Guilherme¹ and Silvia R. Cianzio²

Abstract: The aim of this study was to estimate genetic and phenotypic parameters associated with progenies from the recurrent selection program for grain yield in soybean. The evaluation of $S_{o1}$ progenies was in Lavras in the 2015/2016 crop year. The $S_{o2}$ progenies were evaluated in Lavras, Nazareno, and Itutinga. The $S_{o3}$ progenies were evaluated in Lavras, Ijaci, and Itutinga in the following crop years. The following traits were evaluated: days to flowering, days to full maturity, bottom pod height, plant height, lodging, and grain yield. The expected gain and gain achieved from selection, genetic correlation, and correlated response were estimated. The variance components show variability among the progenies. It was possible to obtain gains from selection for grain yield for all selection intensities (1, 5, 10, 15, 20, 25, and 30%). The variability and the high performance of the progenies indicate that implementation of the recurrent selection program can be successful.

Keywords: Glycine max (L.) Merrill, plant breeding, correlated response, heritability, gain from selection.

INTRODUCTION

Brazil is the second largest soybean producer in the world and is one of the largest exporters of soybean products and by-products (CONAB 2019). The hybridization method is the strategy most used in soybean breeding programs in Brazil. In these programs, various biparental crosses are carried out annually for the purpose of obtaining segregating populations. After that, they are usually conducted by the bulk method for two or three generations of inbreeding. In later steps, the progenies/lines are evaluated aiming at selection of superior genotypes. The cultivars recommended through Value for Cultivation and Use (VCU) trials are incorporated in crossing blocks to obtain new populations, and the steps already mentioned are carried out once more. Even though this is a cyclical process, it is clear that recombination of superior genotypes occurs only after 6 or 8 generations, which attenuates an increase in the frequency of favorable alleles.

With the aim of maximizing resources, as well as efficient use of the time necessary for identification of superior genotypes for recombination, recurrent selection is a useful strategy. Although this strategy has been promoted for breeding of allogamous plants, its use is reported in the literature in autogamous
plants like common bean (Pereira et al. 2017a, Anjos et al. 2018, Melo et al. 2019) and soybean (Posadas et al. 2014, Vello and Nazato 2017, Nie et al. 2019). The use of this technique for the purpose of increasing grain yield in Brazil has not been reported in the literature.

Other agronomic traits are desirable in addition to yield. Plant height, bottom pod height, and lodging directly affect the efficiency of mechanized harvest (Soares 2015). One of the prominent factors of Brazilian agribusiness is the possibility of two harvests in the same crop year. To make this possible, early-maturity soybean cultivars must be used, to achieve a shorter time to harvest. In this context, days to maturity and days to flowering are essential traits.

Obtaining estimates of genetic and phenotypic parameters, such as genetic and phenotypic variances, heritability, genetic and phenotypic correlations, and expected gains from selection are essential for the success of plant breeding programs because the breeder makes decisions based on them. Authors highlight the importance of obtaining parameters from experimental data coming from more than one environment (year/location) because the parameters estimated from only one environment are considerably influenced by the interaction effect, and may lead to mistakes in making decisions related to genotype selection.

Thus, the aim of this study was to estimate genetic and phenotypic parameters associated with the progenies of the recurrent selection program for grain yield in soybean, and to select genetically superior progenies with good agronomic attributes.

MATERIAL AND METHODS

The S₀₁ progenies were evaluated in the 2015/2016 summer crop season in the municipality of Lavras at the Centro de Desenvolvimento Científico e Tecnológico em Agropecuária - Fazenda Muquém (lat 21º 14’ 45” S, long 44º 59’ 59”W, and 918 m asl). A plot consisted of 1 row of 2.0-meter length, with two replications, using a 12 × 12 simple lattice design (131 progenies + 13 parent lines).

The S₀₂ progenies were evaluated in the 2016/2017 summer crop season in the municipality of Lavras at the Fazenda Muquém; in the municipality of Itutinga at Fazenda Milanez (lat 21º 17’ 52” S, long 44º 39’ 28” W, and 969 m asl); and in the municipality of Nazareno at Fazenda Grupo G7 (lat 21º 12’ 59” S, long 44º 36’ 41” W, and 935 m asl). For the evaluations, an 8 × 8 triple lattice design was used (51 progenies + 13 parental lines). One row of 3-meter length was used as the experimental plot.

The S₀₃ progenies were evaluated in the 2017/2018 summer crop season in the municipality of Lavras at the Fazenda Muquém; in the municipality of Itutinga at Fazenda Milanez; and in the municipality of Ijaci at the Centro de Desenvolvimento Científico e Tecnológico em Agropecuária - Fazenda Palmital (lat 21º 09’ 24” S, long 44º 54’ 34” W, 920 m asl). For the evaluations, a 5 × 5 triple lattice design was used (12 progenies + 13 parental lines). Two rows of 3-meter length were used as the experimental plot.

In all the environments evaluated, a no-tillage system was used, with plant rows spaced at 0.50 m. Inoculation with Bradyrhizobium japonicum was performed in the row at the time of planting in a liquid inoculant form at a proportion of 1.200.000 bacteria per seed. Seeds were sown manually and thinned at 25 days after germination, ensuring the stand desired in each one of the experiments. Other crop treatments were carried out according to the procedure presented by Soares et al (2015). The following traits were evaluated:

a) Days to flowering: 50% of the plants of the plot in full bloom, R2 stage (Fehr and Caviness 1977);

b) Days to full maturity: 90% of the plants of the plot in the R8 stage (full maturity) (Fehr and Caviness 1977);

c) Bottom pod height and plant height: in centimeters, of 5 plants taken at random;

d) Lodging: evaluated according to Bernard et al. (1965), with the following scores: 1, for all upright plants; 2, for some leaning or slightly lodged plants; 3, for all the plants moderately leaning or 25-50% lodged; 4, for all the plants severely leaning or 50-80% lodged; and 5, for more than 80% of the plants lodged;

e) Grain yield: value in bag ha⁻¹ after conversion to 13% moisture content.
First, individual analysis was performed for each one of the environments. After that, combined analysis involving the four evaluation environments was carried out. Data were analyzed with the assistance of the R Development Core Team (2018) software through the mixed-model approach (Walsh and Lynch 2018). The variance components were estimated through the restricted maximum likelihood (REML) method. Estimates of the coefficient of experimental variation (CV) and selective accuracy ($r_{g^2}$) were used to assess experimental quality (Resende and Duarte 2007).

To obtain estimates of association between the traits, genotypic correlation analyses were carried out through use of R Development Core Team (2018) software. Significance was checked by the Mantel test. The heritabilities ($h^2$) on the progeny level were obtained in accordance with the estimator proposed by Piepho and Mohring (2007). The expected genetic gain from selection (EG) was estimated for all the traits based on deviation from the genotypic values of the progenies at seven selection intensities (1, 5, 10, 15, 20, 25, and 30%) (Ramalho et al. 2012). For plant height and grain yield, the highest values were selected. For days to flowering, days to full maturity, bottom pod height, and lodging, the lowest values were selected. For the estimates of correlated response ($CR\%$), the following estimator was used according to the modified expressions of Falconer and Mackay (1996):

$$CR_{y/y} (\%) = \frac{BLUP_{y/y}}{\overline{y}} \times 100$$

Where:

$BLUP_{y/y}$ = means of BLUP of the genotypes for trait $y$, by selection carried out for trait $y'$; and $\overline{y}$ = overall mean of the genotypes for trait $y$.

**RESULTS**

Upon obtaining estimates of genetic and phenotypic parameters considering only one environment, it is not possible to isolate the component of the interaction. Thus, the estimates of genetic variance may be over- or underestimated. To isolate the component of the interaction, at least two genotypes and two environments are necessary.

In the present study, the combined estimates of the genetic and phenotypic parameters for all the traits were obtained from the $S_{0:1}$, $S_{0:2}$, and $S_{0:3}$ progenies in all the environments evaluated. They are shown in Table 1. The accuracy estimates for combined analysis ranged from 25.79% (bottom pod height) to 97.45% (days to flowering). The coefficient of variation, for its part, ranged from 2.97% (days to maturity) to 40.21% (lodging) (Table 1).

Interaction variance was significant for all the traits evaluated (Table 1). From this, it can be inferred that progeny performance did not coincide in the environments evaluated for the different traits. It can also be inferred that there is a significant difference among the progenies, since the estimates of genetic variance were other than zero for all the traits in combined analysis (Table 1). Variability can be confirmed by the heritability estimates. This parameter ranged from 0.68 (bottom pod height) to 0.98 (days to flowering) (Table 1).

The frequency distributions of the mean values of combined BLUP for all the traits and environments evaluated are shown in Figure 1. Just as the estimates of the variance components, the frequency distributions of the mean BLUP values

<table>
<thead>
<tr>
<th>Parameters*</th>
<th>Yield (bag ha⁻¹)</th>
<th>DFM (days)</th>
<th>DTF (days)</th>
<th>PH (cm)</th>
<th>BPH (cm)</th>
<th>Lodg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_{gp}$</td>
<td>73.05*</td>
<td>39.51*</td>
<td>27.84*</td>
<td>64.57*</td>
<td>1.68*</td>
<td>0.08*</td>
</tr>
<tr>
<td>$\sigma^2_{gxA}$</td>
<td>69.34*</td>
<td>6.80*</td>
<td>2.524*</td>
<td>28.56*</td>
<td>2.73*</td>
<td>0.12*</td>
</tr>
<tr>
<td>$\sigma^2_{ep}$</td>
<td>179.44</td>
<td>15.17</td>
<td>6.01</td>
<td>42.061</td>
<td>8.06</td>
<td>0.42</td>
</tr>
<tr>
<td>$h^2$</td>
<td>0.79</td>
<td>0.96</td>
<td>0.98</td>
<td>0.91</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td>$r_{g^2}$</td>
<td>66.92</td>
<td>95.24</td>
<td>97.45</td>
<td>89.15</td>
<td>25.79</td>
<td>33.66</td>
</tr>
<tr>
<td>CV%</td>
<td>18.71</td>
<td>2.97</td>
<td>4.44</td>
<td>7.00</td>
<td>16.97</td>
<td>40.21</td>
</tr>
</tbody>
</table>

$\sigma^2_{gp}$ - genetic variance; $\sigma^2_{gxA}$ - variance of genotype × environment interaction; $\sigma^2_{ep}$ - environmental variance; $h^2$ - heritability; $r_{g^2}$ - accuracy; CV% - coefficient of variation; * Significant at 5% probability by the maximum likelihood ratio test. * Data refer to $S_{0:1}$, $S_{0:2}$, and $S_{0:3}$ progenies in the 2015/2016, 2016/2017, and 2017/2018 crop years.
confirm variability among the progenies and also denote the quantitative nature of the traits due to continuous distribution (Figure 1). Almost all progenies evaluated had performance higher than the Brazilian average of 55.55 bag ha\(^{-1}\) (CONAB 2019) (Figure 1). It is also noteworthy that 68% of the progenies had yield higher than 70 bag ha\(^{-1}\), exhibiting their high yield potential (Figure 1).

The estimates of genotypic correlation for all the traits evaluated are shown in Table 2. Analysis of genotypic correlation shows significant estimates between the traits of days to maturity and all other traits, between yield and plant height, and between days to flowering and the bottom pod height and lodging traits. The significant values ranged from 0.4360 for plant height and days to maturity to 0.9106 for correlation between days to flowering and days to maturity (Table 2).

When progenies/lines are evaluated, it is fitting to obtain gain from selection of the superior genotypes. In this study, estimates were obtained considering selection at different intensities. The combined estimates of expected gain from selection, for the three generations, are shown in Table 3.

The estimates varied according to the selection intensities used. As expected, with greater intensity of selection, i.e., when fewer individuals were selected, greater gain was achieved, though with reduced variability (Table 3). It is also important to emphasize that the expected gain from selection for the traits of days to full maturity, days to flowering, and bottom pod height have negative values, since selection for these traits is carried out to reduce them (Table 3).

The estimates of correlated response, shown in Table 4, allow progeny performance to be evaluated for different traits in regard to selection made for grain yield. The results show that once the highest yielding progenies were selected for all the selection intensities, there was also an increase in the values of plant height and bottom pod height, as well as an increase in the number of days to flowering and days to full maturity (Table 4).

**Table 2.** Estimates of genotypic correlation between progenies for the traits of yield (Yield), days to maturity (DFM), days to flowering (DTF), plant height (PH), bottom pod height (BPH), and lodging (Lodg.)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Yield (bag ha(^{-1}))</th>
<th>DFM (days)</th>
<th>DTF (days)</th>
<th>PH (cm)</th>
<th>BPH (cm)</th>
<th>Lodg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield</td>
<td>-</td>
<td>-0.2903</td>
<td>-0.1880</td>
<td>-0.5047*</td>
<td>0.0134</td>
<td>0.2427</td>
</tr>
<tr>
<td>DFM</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DTF</td>
<td>-</td>
<td>0.9106*</td>
<td></td>
<td>0.4360*</td>
<td>0.5425*</td>
<td>0.5276*</td>
</tr>
<tr>
<td>PH</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.3465</td>
<td>0.6660*</td>
<td>0.6626*</td>
</tr>
<tr>
<td>BPH</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.3404</td>
<td></td>
<td>0.0748</td>
</tr>
</tbody>
</table>

* Data refer to \(S_{0:1}, S_{0:2},\) and \(S_{0:3}\) progenies in the 2015/2016, 2016/2017, and 2017/2018 crop years. * Significant at 5% probability by the Mantel test.
In plant breeding programs, comparing genotypes in field experiments for the purpose of identifying differences and selecting superior individuals is a common practice. Therefore, for success in selection, it is essential that these experiments have high experimental accuracy (Matei et al. 2017). Accurate experiments ensure more accurate estimates and, consequently, more reliable recommendations. In this study, accuracy was assessed through the coefficient of experimental variation and of selective accuracy. Experimental accuracy assessed by the coefficient of variation is considered excellent when less than 10%, good from 10% to 20%, average from 20% to 30%, and bad when greater than 30% (Pimentel-Gomes 2009). Accuracy estimates are considered of high magnitude when higher than 70%, of average accuracy with values from 30% to 70%, and of low accuracy with values less than 30% (Resende and Duarte 2007).

In all the environments evaluated, differences could be observed in the magnitude of the estimates of CV and accuracy. However, the values were within the satisfactory interval for both the estimators, confirming the experimental accuracy obtained in this study. High estimates of the coefficient of variation were observed only for lodging. High values of CV for this trait were also observed by Soares et al. (2015). It is expected that higher values for attributes with lower mean values will be observed in CV estimates. When the mean value is near zero, the coefficient of variation will approximate infinity, and this is a possible explanation for the consistently low experimental accuracy observed for lodging (Soares et al. 2015). In contrast, when experimental accuracy is assessed by the accuracy estimate, the effect of the mean is suppressed, and this is one of the main advantages of adopting this tool (Resende and Duarte 2007).

In all the environments evaluated, differences could be observed in the magnitude of the estimates of CV and accuracy. However, the values were within the satisfactory interval for both the estimators, confirming the experimental accuracy obtained in this study. High estimates of the coefficient of variation were observed only for lodging. High values of CV for this trait were also observed by Soares et al. (2015). It is expected that higher values for attributes with lower mean values will be observed in CV estimates. When the mean value is near zero, the coefficient of variation will approximate infinity, and this is a possible explanation for the consistently low experimental accuracy observed for lodging (Soares et al. 2015). In contrast, when experimental accuracy is assessed by the accuracy estimate, the effect of the mean is suppressed, and this is one of the main advantages of adopting this tool (Resende and Duarte 2007).

Genetic variability is fundamental for success in plant breeding programs (Walsh and Lynch 2018, Ramalho et al. 2012). This study showed that the components of genetic variance among progenies were significant for all traits evaluated. Thus, genetic variability among the progenies can be inferred. The amplitude of variation found among the BLUP means, for all the traits evaluated, also denotes variability. The significance of the genetic variance component and the variability found are a result of the choice of parents. In this study, they were used to obtain an elite genotype base population with excellent agronomic performance in the region (Gesteira et al. 2015) and with separate genetic
background coming from different companies (Coodetec, Tropical Melhoramento Genético, Monsoy, Embrapa, Nidera, Brasmax, and Dow Agroscience). The use of elite and contrasting genotypes as parents provided for variability and good agronomic performance among the progenies, and this led to success in selection of superior individuals (Walsh and Lynch 2018, Ramalho et al. 2012).

In the present study, progenies were evaluated in different locations and in different crop years. Under these conditions, the influence of predictable and unpredictable environmental factors is expected (Allard and Bradshaw 1964). It is also important to highlight that the environmental variation found in this study is not only due to the crop year/location effect, but also to the generation effect. The environmental variation associated with the variability among the progenies arose in the significance of the genotype × environment interaction for all the traits evaluated (Table 1). From this result, it can be inferred that in the soybean breeding programs in Brazil, quantification of the effect of the interaction is of crucial importance aiming at selection of superior genotypes. The genotype × environment interaction has been widely reported for the soybean crop in Brazil (Soares et al. 2015, Silva et al. 2016, Volpato et al. 2019) and confirms the results of the present study.

The correlations make it possible to measure the degree of association between two variables (Morais Júnior et al. 2017, Walsh and Lynch 2018). Correlations of high magnitude for two traits allow selection for a trait of interest, especially when it has complex inheritance through another correlated trait that is more easily measured. This strategy is an alternative to direct selection and aims at optimizing gains in plant breeding programs (De Lima et al. 2017). Correlation between two traits may be of a phenotypic, genotypic, or environmental nature. Only genotypic correlations involve an association of an inheritable nature, and these are of greater interest for breeding and are worked with in this study (Falconer and Mackay 1996).

Analysis of genotypic correlation showed significant estimates between the traits of days to maturity and all other traits, between yield and plant height, and between days to flowering and the bottom pod height and lodging traits (Table 2). The significant values ranged from 0.4360 for plant height and days to maturity to 0.9106 for correlation between days to flowering and days to maturity (Table 2). Correlations from 0.3 to 0.6 are considered of medium magnitude (De Lima et al. 2017). The significant correlation between days to full maturity and days to flowering was expected since the traits are directly related. The significant correlation between bottom pod height and days to flowering can be explained by the fact that genotypes that have a longer vegetative period tend to flower at a greater height and, consequently, their bottom pods are more distant from the ground. The positive correlations found between the lodging trait and the days to full maturity and days to flowering traits were also expected because later maturing plants have a greater tendency to lodging. The results here corroborate the findings of Gesteira et al. (2018) in the soybean crop in Minas Gerais.

Quantification of expected gain from selection is fundamental in selection of superior genotypes. Considering greater intensity of selection, i.e., when fewer individuals were selected, greater gain was obtained, since the individuals with greater frequency of favorable alleles are selected. Nevertheless, upon using greater selection intensity, there is reduction in variability (Ramalho et al. 2012).

When diverse traits are considered in evaluation of progenies and selection of superior genotypes, as in the present study, it is fitting to estimate the correlated response; that is, the gain obtained in the secondary trait due to selection in the primary trait. Correlated response was carried out selecting the progenies with higher grain yield and the response for the other traits. The results show that once the highest yielding progenies were selected for all the selection intensities, there was also an increase in the values of plant height, bottom pod height, number of days to flowering, and number of days to full maturity (Table 4). Higher yielding genotypes tend to have higher plant height and later maturity. Therefore, upon selecting the highest yielding progenies, an increase was observed for all the other traits. Similar results for the correlated response for grain yield and days to full maturity were reported by Pereira et al. (2017b).

Grain yield is a quantitative trait, controlled by many genes and strongly influenced by the environment. Thus, recurrent selection arises as an excellent alternative for obtaining consistent gains in long-term selection and maintaining variability in the population. When working with traits controlled by various genes, such as yield, an exorbitant number of plants would be necessary to select an individual with all the favorable alleles in only one generation, which would make the selection program unfeasible. Through recurrent selection, it is possible to undertake breeding in steps by means of successive selection cycles.
Recurrent selection is a cyclical breeding process that involves obtaining progenies, evaluating them, and recombining superior individuals. Thus, an increase in the frequency of favorable alleles is expected, making gains possible in the long term for quantitative traits (Ramalho et al. 2012). It is a tool initially developed for allogamous crops; however, application in autogamous plants, such as wheat, oats, rice, and common bean have increased in recent years. In the case of soybean, there are reports of the use of recurrent selection outside of Brazil for grain protein content (Patil et al. 2017, Rialch and Sharma 2019) and grain yield (Sumarno and Fehr 1982, Posadas et al. 2014, Vello and Nazato 2017, Nie et al. 2019, Hegstad et al. 2019). However, it should be highlighted that for the purpose of expanding the genetic variability of the base population of the recurrent selection program, the authors used exotic germplasm, i.e., germplasm with low adaptability and, likewise, lower frequency of favorable alleles. Reports were not found of the use of recurrent selection in Brazil with the aim of increasing grain yield, i.e., this strategy has not been used before in Brazilian soybean breeding programs.

Selection followed by recombination of the best individuals allows the frequency of favorable alleles in the population to increase, and this increases the chances of selecting superior individuals and obtaining good lines. Geraldi (1997) presents the following argumentation to justify selection followed by recombination and recurrent selection. The author considers the probability of obtaining an individual with at least 80% of the favorable alleles for a trait controlled by 20 genes and 5 populations that vary in regard to the frequency of the favorable allele, and he clearly shows that through selection followed by recombination, the chances of obtaining good individuals are much greater.

In our study, the 20 best $S_{0,2}$ progenies were selected to be crossed using the top cross method to obtain recurrent selection cycle I. The program will use elite genotypes, with high frequency of favorable alleles. According to our results, the variability and high performance of the progenies indicate the possibility of success in implementation of a recurrent selection program. With selection of the 20 best $S_{0,2}$ progenies, we expected gain from selection for all traits, that is, yield, days to maturity, days to flowering, plant height, and bottom pod height (BPH). For example, the gain is 7.3% for grain yield, which is more than the genetic progress obtained for soybean in Brazil per year (Todeschini et al. 2019). Given the various reports of success obtained from recurrent selection in different autogamous crops, this tool constitutes a promising strategy, not only through the capacity of generating superior cultivars, but also through reducing the complexity of conventional breeding programs.

ACKNOWLEDGMENTS

This work was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Financiagal Code 001. This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq/MCTIC, through grant of productivity to the author Adriano Teodoro Bruzi.

REFERENCES


Bernard RL, Chamberlain DW and Lawrence RE (1965) *Results of the cooperative uniform soybean tests Part I. Uniform Soybean Tests, Northern Region*, 134p.


Melo LC, Silva ACF, Pereira HS, Melo PGS and Souza TLPO (2019) Genetic diversity as an additional selection criterion in a carioca common bean recurrent selection program for resistance to common bacterial blight. Genetics and Molecular Research 18: gmr18368


This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.