Pure line selection in a heterogeneous soybean cultivar

Lígia de Oliveira Amaral[1]*, Adriano Teodoro Bruzi[1], Pedro Milanez de Resende[1] and Karina Barroso Silva[1]

Abstract: *Although cultivars of autogamous plants are homogeneous genotypes, they may show natural variability due to mechanical mixing, natural hybridization, and mutation. The aim of the present study was to estimate genetic and phenotypic parameters and to identify and select superior genotypes that associate good performance in traits of interest from a heterogeneous population derived from the cultivar BRS Favorita RR. The evaluation experiments were carried out in two crop years and five cities of the state of Minas Gerais, Brazil, in a lattice design, using progenies and the control 'BRS Favorita RR'. Plant height, first pod height, days to full maturity, lodging score, and grain yield were evaluated. For estimation of the variance components, analysis of deviance was performed by the restricted maximum likelihood (REML) method. The results show that there is difference among treatments and that it is possible to select progenies that outperform the control for all traits evaluated.*

Keywords: *Glycine max, selection method, superior genotypes, genetic and phenotypic parameters, Blup.*

INTRODUCTION

Although the cultivars in autogamous plants are obtained from homogenous genotypes, they become heterogeneous over time after successive cultivations. This is because plants from a genotype can undergo mechanical mixing, cross-pollination, and/or mutation (Ramalho et al. 2012). Prolonged use of the same cultivar by farmers over successive generations and reuse of the grain as "seeds" can provide natural variability derived from the aforementioned mechanisms, which can be useful in breeding programs. Variability in the cultivar is essential for successful selection of superior genotypes (Ramalho et al. 2012).

The concept of pure lines arose from a series of experiments conducted by the Danish botanist W. L. Johannsen in 1903 with the bean cultivar Princess. He investigated the effect of selection on the trait of bean seed weight.

The pure line selection strategy has been used in some breeding programs, such as in coffee (Carvalho et al. 1952), beans (Ramalho et al. 1982, Yokoyama et al. 1999, Santos et al. 2002), and rice (Rangel et al. 1998). In the soybean crop, the importance of the pure line selection method is also shown in the study of Priolli et al. (2013), in which the genetic structure of soybean cultivars in Brazil was systematically evaluated. For that purpose, the cultivars were genotyped with the use of microsatellite markers, which showed that of the 435 cultivars evaluated, 42 were from selection programs with an emphasis

[1] Universidade Federal de Lavras, Campus Universitário, 37.200-000, Lavras, MG, Brazil

on photoperiod insensitivity (Andrews, BRS 65, Missões, Emgopa 306, UFV 1, FT 440, Cristalina, IAS 1, Monarca, IAC 8-2, Ocepar 7, Embrapa 64). Sebastian et al. (2010) and Tokatlidis (2015) have shown the existence of variability and obtained success in increasing grain yield through selection. However, in Brazil there are no reports on the use of this strategy to select superior lines for grain yield and days to full maturity.

Thus, the aim of this study was to estimate genetic and phenotypic parameters and to identify and select superior genotypes that associate good performance in traits of interest from a heterogeneous population of the cultivar BRS Favorita RR.

MATERIAL AND METHODS

The experiment was carried out in two crop seasons: first, in 2014/2015 in the municipalities of Lavras, MG, at the Center for Scientific and Technological Development in Agriculture of UFLA - Fazenda Muquém (lat 21° 14' 43" S, long 44° 59' 59" W, alt 919 m asl), and in Patos de Minas, MG, at the Sertãozinho experimental farm of EPAMIG (lat 18° 44' 34" S, long 46° 31' 04" W, alt 832 m asl); and in 2015/2016, once more in Lavras, in Ijaci, MG, at the Center of Technology Transfer of UFLA - Fazenda Palmital (lat 21° 9' 24" S, long 44° 55' 34" W, alt 833 m asl), and in Nazareno, MG, on private property, Fazenda G7 (lat 21° 12' 46" S, long 44° 35' 54 W", 926 m alt asl).

In March 2014, 400 plants from the cultivar BRS Favorita RR were collected at Fazenda Milanez (lat 21° 17' 52" S, long 44° 39' 28" W, alt 969 m asl) in the municipality of Itutinga, MG, in an area of about 10 hectares. The plants were chosen by phenotypic selection, i.e., visibly superior plants were selected. In the present study, the occurrence of variability may be based on natural hybridization, but variability may also arise from mechanical mixing of seed or mutations. From this selection, individual trials of plants were performed. A total of 255 lines/progenies were obtained, which were evaluated in a 16 × 16 simple lattice design together with a control ('BRS Favorita RR'), adopting a 3-m row as a plot in Lavras and Patos de Minas in the 2014/2015 crop season.

A no-tillage system was adopted and seeds were inoculated with *Bradyrhizobium japonicum* using a liquid inoculant in the rate of 1,200,000 bacteria per seed. Seeds were sown manually and thinning was performed 25 days after germination, ensuring the desired final stand of 10-12 plants per linear meter. Other crop practices were performed according to the procedures described by Soares et al. (2015).

The following traits were evaluated. Grain yield (in 60 kg bags per ha) was determined from the harvest of each plot, corrected to 13% moisture, and extrapolation of the value obtained to one hectare. First pod height (cm) and plant height (cm) were obtained from five plants per plot taken at random - with direct measurement from the base of the plant to the point of attachment of the first pod to the stem for first pod height, and from the base of the plant to its tip for plant height, using gauge. Days to full maturity consisted of the number of days from planting to time of collection, the time at which 95% of the plants had mature pods. Finally, the lodging index was evaluated according to Bernard et al. (1965), with a score of 1 for all plants upright, 2 for some leaning or slightly lodged plants, 3 for all plants moderately leaning or 25%-50% lodged plants, 4 for all plants leaning severely or 50%-80% lodged plants, and 5 for more than 80% lodged plants.

From the 255 lines/progenies evaluated, 80 were selected for evaluation with the control in the 2015/2016 crop season in the municipalities of Lavras, Ijaci, and Nazareno. A 9 × 9 triple lattice design was used and two 3-m rows for each. Soil tillage and the crop practices adopted were as previously described. Just as in the 2014/2015crop season, grain yield (in bags ha^{-1}), plant height (cm), first pod height (cm), days to full maturity, and lodging index (scores from 1 to 5) were recorded.

The data obtained were analyzed using R software from the R Development Core Team (2016 or 2017?) using a mixed model approach (Bernardo 2010). First, individual analysis was performed for each of the five experiments. Subsequently, joint analyses were performed for the sites in the same crop season, and, finally, a general combined analysis containing all the evaluation environments, i.e., sites and crop seasons, was performed. The models used in the analyses of variance are shown in Table 1.

For estimation of variance components, analysis of deviance was performed by the restricted maximum likelihood (REML) method. Fixed effects were verified by the F-factor significance test, and the significance of variances associated

Table 1. Statistical models adopted to perform each analysis of variance

Analyses	Statistical model
Individual/Location	$Y_{ijk} = \mu + p_i + b_{(j(k))} + r_k + E_{ijk}$
Joint/Crop season	$Y_{ijkl} = \mu + p_i + l_i + b_{jkl} + r_{(k(l))} + (pl)_{il} + E_{(ij(kl))}$
Combined	$Y_{ijkm} = \mu + p_i + a_i + b_{(j(km))} + r_{(k(m))} + (pa)im + E_{(ij(km))}$

$Y_{ijk} = Y_{ijkl} = Y_{ijkm}$ = observed value; μ = constant associated with all observations; p_i = random effect of progeny i; l_i = random effect of site; a_i = random effect of environment; $b_{j(k)}$ = random effect of block j within replication k; r_k = random effect of replication k; E_{ijk} = experimental error associated with observation Y_{ijk}; $b_{j(kl)}$ = random effect of block j within site k and replication l; $r_{k(l)}$ = random effect of replication k within site l; $(pl)_{ij}$ = effect of progeny × site interaction; $E_{ij(kl)}$ = error associated with observation Y_{ijkl}; $b_{j(km)}$ = random effect of block j within replication k and environment m; $r_{k(m)}$ = random effect of replication k within environment m; $(pa)_{im}$ = effect of progeny × environment interaction; $E_{ij(km)}$ = Error associated with observation Y_{ijkm}.

with the random effects was verified by the Likelihood Ratio Test (Bernardo 2010). The packages used in the software routines were lme4 (Bates et al. 2014), lsmeans (Lenth 2016), and car (Fox and Weisberg 2011). For calculation of heritability, the following estimator was used:

$$h^2 = \frac{\sigma^2_G}{\sigma^2_G + \frac{\sigma^2_{GE}}{n} + \frac{\sigma^2_E}{nr}}$$

Where σ^2_G is the genotypic variance, σ^2_{GE} is the variance of the genotype × environment interaction, σ^2_E is the environment variance, n is the number of environments, and r is the harmonic mean number of replications. Narrow sense heritability was used, which uses only the additive genetic variance, and associated with the reproductive value, which is fixed by selection. To evaluate experimental quality, the experimental coefficient of variation (CV) and selective accuracy (rgg') were determined by the following estimators:

$$CV = \frac{\sqrt{\hat{\sigma}^2_e}}{\overline{x}} \qquad\qquad rgg' = \sqrt{h^2}$$

Where $\hat{\sigma}^2_e$ is the error variance, $\overline{x}$ is the average, and h^2 is heritability at the progeny level.

The efficiency of the mixed model analysis procedures was evaluated by the genetic gain from selection (GS), calculated from the BLUP averages of the progenies (GS), as well as the realized gain (RG) at seven selection intensities (1, 5, 10, 15, 20, 25, and 30%) (Ramalho et al. 2012). In order to estimate the realized gain (RG %), the following estimator was used:

$$GR(\%) = \frac{\overline{BLUP'S}_{j/i}}{\overline{y}_j} * 100$$

Where $\overline{BLUP'S}_{j/i}$ is the BLUP average of the genotypes in generation j, by the selection made in generation i, and $\overline{y}_j$ is the general average of genotypes in generation j.

RESULTS AND DISCUSSION

In the present study, two estimates were used to verify experimental accuracy: the coefficient of experimental variation (CV), classified according to Pimentel Gomes (2009), and accuracy, classified according to Resende and Duarte (2007) (Table 2). It is clear that experimental accuracy varied according to environments and traits. The trait that showed the highest CV values was lodging. In estimation of the coefficient of variation, higher values are expected for attributes with lower average values (Soares et al. 2015), which may explain the low accuracy associated with the lodging trait. When we consider accuracy by its estimator, the average effect is suppressed, which is the main advantage in adopting this tool to measure experimental accuracy (Resende and Duarte 2007). It was also clear that the experimental accuracy measured by the CV and by accuracy is dependent on the number of sites/environments contemplated in the analysis of variance. Individual analyses denote less experimental accuracy. This was expected because there are a larger number of replications and hence greater accuracy when joint analysis is performed, i.e., greater reliability in recommendations and even in identification of superior genotypes (Ramalho et al. 2012). When estimates of genetic and phenotypic parameters are obtained considering only one site, it is not possible to isolate the component due to the interaction, and estimates of genetic variance are over- or underestimated. In this context, joint analysis of variance was performed (Table 2).

Table 2. Estimation of genetic and phenotypic parameters for plant height (cm), first pod height (FPH) (cm), yield (bags ha^{-1}), days to full maturity (FM), and lodging of soybean progenies based on combined joint analysis

Parameter	Height	FPH	Yield	FM	Lodging
σ^2_P	35.89**	1.40**	12.15**	5.02**	0.18**
σ^2_{PxA}	8.49**	0.96**	12.53**	0.26**	0.03**
σ^2_E	47.66	15.30	72.46	4.98	0.68
h^2	0.39	0.08	0.13	0.49	0.20
Accuracy (%)	62.45	28.21	35.37	69.94	45.24
CV (%)	8.01	18.09	17.07	1.48	21.70
Overall mean	86.13	21.62	49.84	150.77	3.80

** Significant at 1% probability by the maximum likelihood ratio test; σ^2_P: variance among progenies; σ^2_{PxA}: progeny × environment interaction variance; σ^2_E: error variance; h^2: heritability among progenies.

The components of genetic variance among progenies and the interaction were significant for all traits evaluated. In the case of an autogamous plant, when the farmer does not buy seeds, i.e., the grain is reused as "seed" for the next crop season, variability is common in the cultivars (Ramalho et al. 2012). Variability may arise due to mechanical mixing, mutation, and natural hybridization. Variability in heterogeneous cultivars is reported in the studies of Priolli et al. (2013) for insensitivity to photoperiod and in Sebastian et al. (2010) for grain yield in the USA. The selection efficiency for grain yield in soybean adopting heterogeneous cultivars, such as coffee, dry edible bean, and rice in other crops, has been reported (Carvalho et al. 1952, Ramalho et al. 1982, Rangel et al. 1998, Yokoyama et al. 1999, Santos et al. 2002). However, it has not yet been reported in Brazil. Therefore, there may be genetic variability among the progenies and the relative response of progenies varied in the different environments (Table 2 and Figures 1 and 2).

Heritability estimates indicate the ratio of genetic and non-environmental components in the variation observed. These estimates ranged from 0.08 to 0.49 of selection, corroborating the results found in this study (Rocha and Vello 1999).

In this study, progenies were evaluated at different sites and in different crop seasons. Under such conditions, influence from predictable and unpredictable environmental factors is expected (Allard and Bradshaw 1964). The environmental variation associated with the variability among lines/progenies allowed the occurrence of the progeny × environment interaction. This suggests that the progeny response did not coincide for the different traits in the environments evaluated. The variance component associated with the interaction for grain yield was of greatest magnitude, compared to the other traits (Table 2). For days to full maturity, first pod height, and lodging score, although the effect of the genotype × environment interaction was significant, it was of small magnitude compared to grain yield estimates (bags ha^{-1}) (Table 2). This can be explained from the fact that grain yield is a quantitative trait controlled by a large number of genes and, hence, it is highly affected by the environment. However, other traits, such as first pod height, plant height, days to full maturity, and lodging index, show less influence from environmental factors due to their simpler inheritance (Ramalho et al. 2012, Pereira et al. 2017, Zambiazzi et al. 2017, Gesteira et al. 2018).

In any selection strategy, success is achieved by quantifying the expected and realized gain with selection. Estimates of the expected gain from selection were obtained considering selection at different intensities. In Table 3, the gain for grain yield was 5.95%, considering an intensity of 30%. For plant and first pod height, the estimates are negative, because selection is made with the aim of selecting progenies with lower plant and first pod height. The same comments are relevant for Table 4. For days to full maturity and the lodging score, the gain is also negative, since the objective is to select progenies with a shorter cycle and that do not exhibit lodging (Table 4). Estimates varied according to the selection intensities used. As expected, with greater selection intensity, i.e., when fewer individuals were selected, greater gain was obtained, however, with reduction in variability (Tables 3 and 4). In Table 4, for example, the expected gain ranged from 9.45 (1% SI) to 4.96 (30% SI) for grain yield. The estimates for realized gain were obtained for the progenies selected in the 2014/2015 crop season and evaluated in the 2015/2016 crop season (Table 3). It was expected that the gain would decrease as the number of selected individuals increased, but that did not occur in some situations (Table 3). This can be explained from the occurrence of the genotype × environment interaction throughout the experiments/generations evaluated (Zambiazzi et al. 2017). Furthermore, the estimates

of realized gain were smaller than those of expected gain. For grain yield, for example, when a selection intensity of 1% was considered, the gain was only 2.69% (Table 3), and the effect of the genotype × environment interaction is also a likely cause.
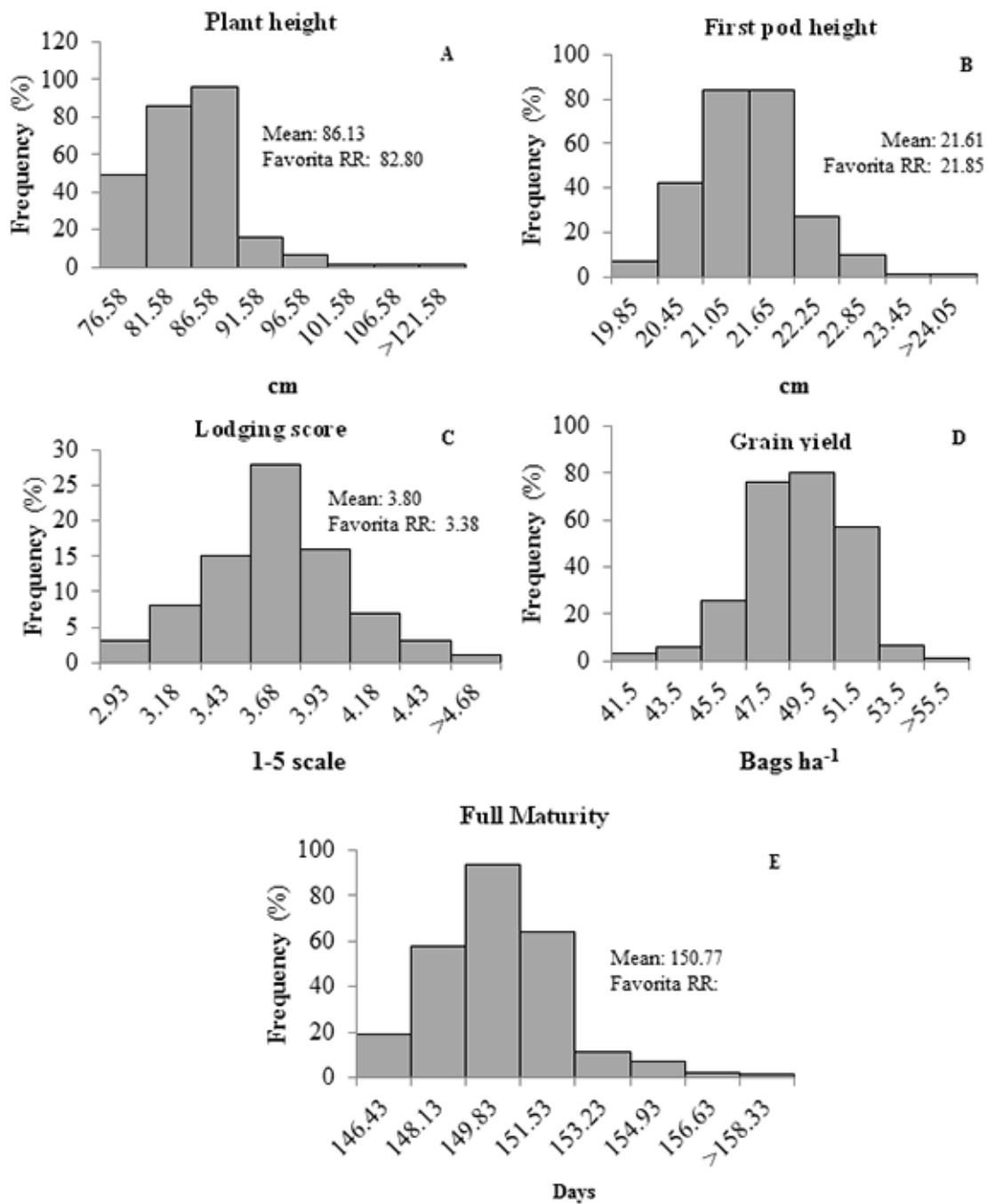

Figure 1. Frequency distribution of the adjusted averages for: A. Plant height (cm); B. First pod height (cm); C. Lodging (1-5); D. Grain yield (bags ha⁻¹); E. Days to full maturity (days).
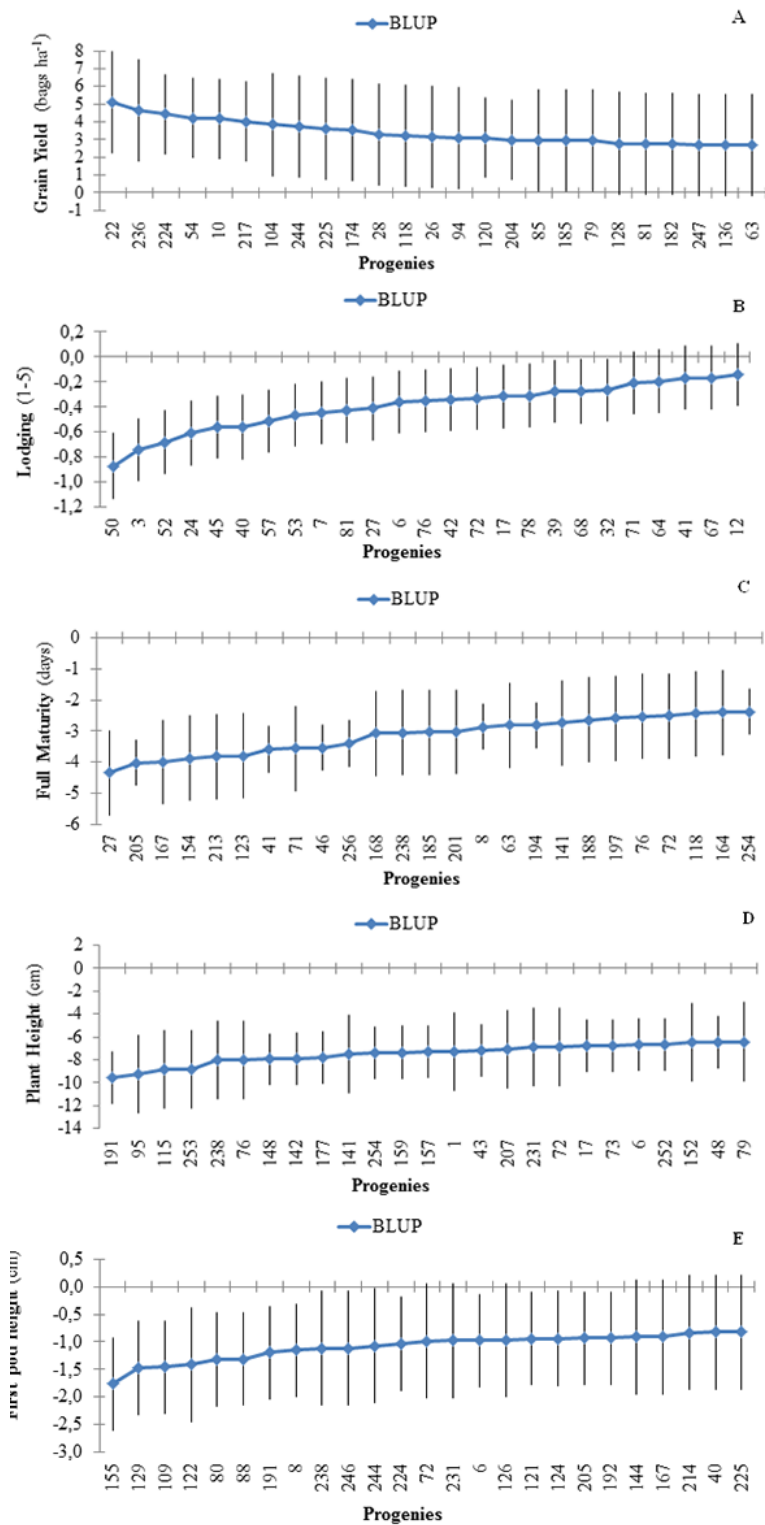
Figure 2. Genotypic values (BLUPs) and associated errors from the 25 best progenies for grain yield (bags ha[-1]), lodging (1-5), full maturity (days), plant height (cm), and first pod height (cm).

Table 3. Expected gain (GS) and realized gain (RG) from selection of soybean progenies at different selection intensities in the 2014/2015 crop season

SI (%)	Height		FPH		Yield	
	GS	GR	GS	GR	GS	GR
1	-11.49	-4.27	-4.62	-2.68	12.40	2.69
5	-9.82	-3.86	-3.71	-2.12	9.90	1.17
10	-8.73	-2.56	-3.23	-0.38	8.30	0.54
15	-8.12	-2.63	-2.97	0.43	7.55	0.34
20	-7.51	-2.14	-2.70	-0.01	6.93	0.39
25	-6.97	-1.21	-2.49	0.16	6.42	0.82
30	-6.48	-0.44	-2.31	0.01	5.95	0.09

SI - Selection intensity; GS - Expected gain from selection; GR - Realized gain. Height of plants (Height) (cm); First pod height (FPH) (cm); Grain yield (Yield) (bags ha^{-1}).

Table 4. Expected gains from selection of soybean progenies at different selection intensities in the 2014/2015 and 2015/2016 crop seasons

SI (%)	Height	FPH	Yield	FM	Lodging
1	-10.68	-7.23	9.45	-2.73	-22.82
5	-9.44	-5.83	7.84	-2.40	-19.02
10	-8.65	-5.00	6.76	-2.07	-16.39
15	-8.07	-4.49	6.18	-1.85	-14.52
20	-7.46	-4.07	5.72	-1.67	-13.08
25	-6.91	-3.69	5.33	-1.51	-11.93
30	-6.38	-3.38	4.96	-1.37	-10.74

SI - Selection intensity. Height of plants (Height) (cm); First pod height (FPH) (cm); Grain yield (Yield) (bags ha^{-1}); Days to full maturity (FM); lodging index (Lodging) evaluated according to Bernard et al. (1965).

Variability among the progenies for all the traits can also be observed in the frequency distribution graphs (Figure 1). It should be noted that some progenies outperformed 'BRS Favorita RR'. For grain yield (Figure 1D), in which the average of the control was 48 bags ha^{-1}, progenies with grain yield above 55 bags ha^{-1} could be selected. For the other traits (Figure 1A, 1B, 1C, and 1E), in which the objective is to reduce average expression of the traits, there is also the possibility of selecting progenies that achieve this objective. Considering days to full maturity, lines with a cycle shorter than the control and associated with good agronomic performance can be obtained (Figure 1E).

The scatter plots (Figure 2) show that there is genotypic value that differed from zero considering the 25 best genotypes and a selection intensity of 1% for all the traits. In other words, there is success from selection, and these progenies will contribute to improvement in expression of the trait in subsequent generations.

Among the traits evaluated, grain yield is the most important. Furthermore Santos et al. (2002), Sebastian et al. (2010), and Tokatlidis (2015) confirm the existence of variability and success in selection. With this as reference, progeny 22 was the most prominent (Figure 2A). In addition, the control (treatment 81) was among the genotypes with superior grain yield. However, it should be noted that the genotypic value associated with this treatment did not differ from zero, i.e., it was null (Pereira et al. 2017).

One of the major objectives of soybean breeding is to increase yield potential and reduce days to full maturity. Figure 2C shows the genotypic values associated with the 25 best progenies, i.e., those with fewer days to full maturity. All the estimates differ from zero. Therefore, it can be inferred that gain can be obtained from selection of progenies that show fewer days to full maturity.

This study showed that there is variability in the heterogeneous cultivar BRS Favorita RR, allowing selection of lines that combine grain yield, plant height, first pod height, and resistance to lodging which are superior to 'BRS Favorita RR'.

REFERENCES

Allard RW and Bradshaw AD (1964) Implications of genotype-environmental interactions in applied plant breeding. **Crop Science 5**: 503-508.

Bates D, Mächler M, Bolker B and Walker S (2014) Fitting linear mixed-effects models using lme4. **Journal of Statistical Software 67**: 1-48.

Bernard RL, Chamberlain DW and Lawrence RD (1965) **Results of the cooperative uniform soybean tests.** USDA, Washington, 134p.

Bernardo R (2010) **Breeding for quantitative traits in plants**. Stemma Press, Woodbury, 400p.

Carvalho A (1952) Melhoramento do cafeeiro: IV-Café Mundo Novo. **Bragantia 4**: 97-130.

Fox J and Weisberg S (2011) An {R} companion to applied regression. **Thousand oaks CA**. Available at: <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>. Accessed on Nov 22, 2017.

Gesteira GS, Bruzi AT, Zito RK, Fronza V and Arantes NE (2018) Selection of early soybean inbred lines using multiple index. **Crop Science 58**: 2494-2502.

Lenth RV (2016) Least-squares means: The R package lsmeans. **Journal of Statistical Software 69**: 1-33.

Pereira FC, Bruzi AT, Matos JW, Rezende BA, Prado LC and Nunes JAR (2017) Implications of the population effect in the selection of soybean progeny**. Plant Breeding 136**: 679-687.

Pimentel Gomes F (2009) **Curso de estatística experimental**. FEALQ, Piracicaba, 451p.

Priolli RHG, Wysmierski PT, Cunha CPD, Pinheiro J B and Vello NA (2013) Genetic structure and a selected core set of Brazilian soybean cultivars. **Genetics and Molecular Biology 36**: 382-390.

R Core Team (2017) R: A Language and environment for statistical computing. R Foundation for statistical Computing, Vienna. Available at: <http://www.R-project.org/>. Accessed on Nov 22, 2016.

Ramalho MAP, Abreu AFB, Santos JB and Nunes JAR (2012) **Aplicações da genética quantitativa no melhoramento de plantas autógamas**. UFLA, Lavras, 522 p.

Ramalho MAP, Pinto CABP and Santa Cecília FC (1982**)** Avaliação de amostras de cultivares de feijão roxo e seleção de progênies. In **Reunião nacional de pesquisa de feijão 1**. Embrapa, Brasília, p. 20-21.

Rangel PHN, Zimmermann FJP and Neves PCF (1998) Estimativas de parâmetros genéticos e resposta à seleção nas populações de arroz irrigado CNA-IRAT 4PR e CNA-IRAT 4ME. **Pesquisa Agropecuária Brasileira 33**: 905-912.

Resende MDV and Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical 3**: 182-194.

Rocha MDM and Vello NA (1999) Interação genótipos e locais para rendimento de grãos de linhagens de soja com diferentes ciclos de maturação. **Bragantia 1**: 69-81.

Santos PSJ, Abreu AFB and Ramalho MAP (2002) Seleção de linhas puras no feijão 'carioca'. **Ciência e Agrotecnologia 20**: 1492-1498.

Sebastian SA, Streit LG, Stephens PA, Thompson JA, Hedges BR, Fabrizius MA and Feng L (2010) Context-specific marker-assisted selection for improved grain yield in elite soybean populations. **Crop Science 50**: 1196.

Soares IO, Rezende PM, Bruzi AT, Zuffo AM, Zambiazzi EV, Fronza V and Teixeira CM (2015) Interaction between soybean cultivars and seed density. **American Journal of Plant Sciences 6**: 1425-1434.

Tokatlidis IS (2015) Conservation breeding of elite cultivars. **Crop Science 55**: 2417-2434.

Yokoyama LP, Del Peloso MJ, Di Stefano JG and Yokoyama M (1999) Nível de aceitabilidade da cultivar de feijão "Pérola": avaliação preliminar. Embrapa Arroz e Feijão, Santo Antônio de Goiás, 20p. (Documento 98).

Zambiazzi EV, Bruzi AT, Guilherme SR, Pereira DR, Lima JG, Zuffo AM and Carvalho MLM (2017) Estimates of genetics and phenotypics parameters for the yield and quality of soybean seeds. **Genetics and Molecular Research 16**: 1-11.