

Multivariate approach in the selection of superior soybean progeny which carry the RR gene¹

Abordagem multivariada na seleção de progênies de soja superiores e portadoras do gene RR

Anderson Dallastra^{2*}, Sandra Helena Unêda-Trevisoli², Antonio Sergio Ferraudo³ e Antonio Orlando Di Mauro²

ABSTRACT - Efficiency in the use of genetic variability, whether existing or created, increases when properly explored and analysed. Incorporation of biotechnology into breeding programs has been the general practice. The challenge for the researcher is the constant development of new and improved cultivars. The aim of this experiment was to select progenies with superior characteristics, whether or not carriers of the RR gene, derived from bi-parental crosses in the soybean, with the help of multivariate techniques. The experiment was carried out in a family-type experimental design, including controls, during the agricultural year 2010/2011 and 2011/2012 in Jaboticabal in the Brazilian State of São Paulo. From the F₃ generation, phenotypically superior plants were selected, which were evaluated for the following traits: number of days to flowering; number of days to maturity; height of first pod insertion; plant height at maturity; lodging; agronomic value; number of branches; number of pods per plant; 100-seed weight; number of seeds per plant; grain yield per plant. Given the results, it appears possible to select superior progeny by principal component analysis. Cluster analysis using the K-means method links progeny according to the most important characteristics in each group and identifies, by the Ward method and by means of a dendrogram, the structure of similarity and divergence between selected progeny. Both methods are effective in aiding progeny selection.

Key words: *Glycine max.* Genetic variability. Multivariate analysis.

RESUMO - A eficiência do aproveitamento da variabilidade genética, existente ou criada, aumenta quando esta é devidamente explorada e analisada. A incorporação de eventos biotecnológicos aos programas de melhoramento tem sido uma prática usual. O pesquisador tem como desafio, o desenvolvimento constante de novas e melhores cultivares. O objetivo deste experimento foi selecionar progênies com caracteres superiores e portadoras ou não do gene RR, oriundas de cruzamentos bi-parentais em soja, com o auxílio de técnicas multivariadas. O experimento foi conduzido em delineamento experimental do tipo famílias com testemunhas intercaladas no ano agrícola 2010/2011 e 2011/2012 em Jaboticabal-SP, Brasil. Na geração F₃, foram selecionadas plantas fenotipicamente superiores, sendo estas avaliadas para os caracteres: número de dias para o florescimento, número de dias para a maturidade, altura de inserção da primeira vagem, altura de planta na maturidade, acamamento, valor agrônomo, número de ramos, número de vagens por planta, peso de 100 sementes, número de sementes por planta e produção de grãos por planta. Diante dos resultados, verifica-se a possibilidade de selecionar progênies superiores através da análise de componentes principais. A análise de agrupamentos, através do método de K-means, une as progênies de acordo com os caracteres de maior importância em cada grupo e, através do método de Ward, identificou por meio do dendrograma a estrutura de similaridade e divergência entre as progênies selecionadas. Os dois métodos são eficientes no auxílio da seleção de progênies.

Palavras-chave: *Glycine max.* Variabilidade genética. Análise multivariada.

*Autor para correspondência

¹Recebido para publicação em 09/06/2013; aprovado em 23/03/2014

Parte da Dissertação de Mestrado do primeiro autor, apresentada ao programa de Pós-Graduação em Genética e Melhoramento de Plantas da UNESP/FCAV/Pesquisa financiada com Bolsa de mestrado CAPES e suporte financeiro da FAPESP

²Programa de Pós-Graduação em Genética e Melhoramento de Plantas, Departamento de Produção Vegetal, Faculdade de Ciências Agrárias e Veterinárias/UNESP, Campus de Jaboticabal, Jaboticabal-SP, Brasil, anderson_dallastra@hotmail.com, strevisoli@fcav.unesp.br, orlando@fcav.unesp.br

³Departamento de Ciências Exatas, Faculdade de Ciências Agrárias e Veterinárias/UNESP, Jaboticabal-SP, Brasil, ferraudo@fcav.unesp.br

INTRODUCTION

Progress in plant breeding is dependent on the skill of the breeder in identifying selection criteria that are able to promote the desired changes in characteristics which are of interest in a breeding program (REIS *et al.*, 2004).

The greater the degree of divergence between the parents, the greater the resulting variability in the segregating population and the greater the possibility of grouping the alleles into new and favourable combinations (BARBIERI *et al.*, 2005).

The genetic gains provided in the productive sector by new cultivars have been very significant - about 1.38% per year for the soybean crop (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA, 2011). However, the selection of superior progeny is no easy task, since characteristics of importance, the majority of which are quantitative, show complex behaviour, being influenced by the environment and interrelated in such a way that the selection or alteration of one causes a series of changes in another (RODRIGUES *et al.*, 2013; SILVA *et al.*, 2009; VASCONCELOS *et al.*, 2010).

Cultivation management directly influences the behaviour of genotypes. Chemical control of weeds is an excellent strategy due to its practicality, efficiency and speed (GAZZIERO; VARGAS; ROMAN, 2004). Glyphosate, a non-selective herbicide, is widely used around the world, being highly effective in controlling weeds (PLINE-SRNIC, 2006).

The soybean cultivar Roundup Ready-RR contains the enzyme EPSPs which comes from *Agrobacterium sp.* and makes the plants resistant to the action of glyphosate (ABREU; MATTA; MONTAGNER, 2008). According to Lima *et al.* (2011) the technology that has provided the greatest impact on weed management in the last 10 years has been the introduction of varieties which are resistant to herbicides.

As regards data analysis, multivariate techniques provide a way to run, in a single analysis, that which previously required multiple univariate analyses to achieve (HAIR *et al.*, 2009). Principal Component Analysis (PCA) aims to reduce the dimensionality of the data set while retaining as much information as possible in a smaller number of principal components (SILVA *et al.*, 2010). PCA can be used to cluster individuals with similar characteristics and study their correlation (VALLADARES *et al.*, 2008). In turn, Cluster Analysis (CA) can complement this information, whether by hierarchical methods or otherwise, classifying objects into groups (FERRAUDO, 2010) with high internal

consistency (within groups) and high external heterogeneity (between groups) (HAIR *et al.*, 2009; ALEIXO; SOUZA; FERRAUDO, 2007). These methodologies have been of great use in breeding, providing important information which aid in the selection process.

The objective of this experiment was to select superior soybean progeny from segregating populations carrying the glyphosate resistant gene (RR) and to identify more efficient crosses and parents, using multivariate approaches, and furthermore, to test the efficiency of the methods of principal component analysis and clustering in the process of selection of multiple characteristics of interest.

MATERIAL AND METHODS

In order to get segregating populations, 13 early strains belonging to the breeding program of the São Paulo State University - UNESP/FCAV, at Jaboticabal in the Brazilian State of São Paulo, were tested, being used as female parents, and 11 cultivars carrying the RR gene used as male parents, aiming at the introgression of the RR gene. The genealogy of the populations is shown in Table 1.

After performing the crosses, 37 F₁ populations were obtained and sown in pots to obtain the F₂ generation. The F₂ seeds from each cross were harvested separately and planted in the experimental area of the Farm for Teaching, Research and Extension - FEPE, of the São Paulo State University - UNESP/FCAV in Jaboticabal, during the agricultural year of 2010/11. From each cross, eight visually superior F₂ plants were selected, resulting in a group of plants that gave rise to the F₃ progenies. These F₂ plants were evaluated based on their agronomic traits: number of days to flowering (NDF), number of days to maturity (NDM), plant height at maturity (PHM) in cm, height of the first pod insertion (HPI) in cm; lodging (Lg), using a scale from 1 to 5 where 1 is the upright plant and 5 is totally flat; agronomic value (AV) using a scale from 1 to 5 in which 1 is undesirable and 5 desirable, number of branches per plant (NB), number of pods per plant (NP), number of seeds per plant (NS), 100-seed weight (HSW) in grams, and grain yield (GY) in grams per plant.

The progeny of the F₃ generation, which had been previously selected, were then sown in the field at FEPE/UNESP, Jaboticabal, during the agricultural year of 2011/12, using families with interspersed controls as the experimental design, and giving a total of 296 genotypes. For this purpose, the strains used as controls were: JAB.06-2/2C1D, JAB.01-21/4M1D (UNESP/FCAV) as well as the cultivars Conquista, V-Max, CD-207, CD-216 and CD-219. As the progeny reached maturity, a new evaluation began of six plants per genotype family, with individual progeny being selected based on the

agronomic characteristics already described. Statistical analyses were carried out using the STATISTICA v.7 computer software (STATSOFT, 2004).

The data for the traits Lg and AV were compared and used in subsequent analyses, as they had been assigned using pre-defined scales. The data were then standardised for analysis, according to recommendation (STATSOFT, 2004).

For analysis of the progeny by PCA, a method was used in which the scales of the graph were pre-determined in order to identify more easily progeny possessing characteristics which make it possible to differentiate them from the others within each cross. Thus, the first, less-stringent scale is made up of values on the X axis of +5 to -5 and on the Y axis of +2 to -2. On the second scale, which was more precise, values were determined for the X axis of +6 to -6, and for the Y axis of +2.5 to -2.5. These scales were pre-determined with the aim of achieving a selection of around ten percent of the total population under evaluation. Such scales can also be pre-determined with the aim of selecting a percentage of the maximum values (positive and negative). All progeny that were positioned near the centre do not show significant differences between themselves. Whereas those progeny that were positioned between the two scales can be thought of as differing from the others, with more superior characteristics depending on their positioning. Finally, those progeny positioned on the extremes of the more-precise scale are thought of as having a marked difference in one or more variables of expression, which distinguishes them from other progeny of the same cross. These progeny are the ones that should receive special attention and are likely to be selected.

The principal components were selected as proposed by Kaiser (1958), also cited by Hair *et al.* (2009), where only eigenvalues greater than one (1.0) were taken into consideration, as they generate components having a relevant amount of information of the original variables. These analyses were initially performed for each group of progeny (families) of each cross, so it is possible to identify which of the selected plants presented superior characteristics. This process was repeated for all crosses, i.e. one selection within each family and also the controls in the comparative study.

Cluster analysis by the K-means method is a procedure in which, given a predetermined number of groups, points are calculated that represent the "centres" of these groups. In this case, it was determined that division into six groups resulted in a positive response as to the distribution of results.

Cluster analysis was then performed using the Ward method to generate the dendrogram, with the aim of comparing the results of the formation of groups

between the different analyses. For a better understanding and comparison of the results of the division into groups made by the two analyses, each group generated by the K-means method was given a distinct colour when named.

RESULTS AND DISCUSSION

The F₃ selection of six plants from each of the 296 families of genotypes, taking into consideration losses due to the low germination of some of these, resulted in a total of 1382 selected progeny, in addition to the 110 plants evaluated as controls.

Initially, it was found that only three eigenvalues were greater than one. The eigenvalue for the first component was 4.062, corresponding to 45.14% of the total variance. The main variables that explained this retention were NB, NP, NS and GY. The eigenvalue for the second component was 1.578. The proportion of retained variance is approximately 17.54%, with the main variables being PHM and HSW. The third principle component has a value very close to one, making it open to analysis, however, in this study only the first two eigenvalues were used, as they had greater loadings and together retained a total of 62.68% of the original variance.

Due to the large amount of information generated by the PCA from each of the 37 crosses, an analysis of the mean of each cross was made, shown in Figure 1, where it is possible to verify their dispersion in accordance with the variables. It appears that the C5 and C6 crosses exhibit a strong correlation with the variables GY, NB, NS and NP however these crosses were not subsequently selected by the ACP (Table 1).

This fact is attributed to the low number of progeny in these crosses; four and five progeny respectively. The number of progeny was higher in cross C3, allowing the variability generated by the genetic divergence between the parents to be explored.

For purposes of explanation, data from cross C11 were used in the breakdown of the analysis, as they present data that allow detailed exploration of the results, shown in Figure 2.

In Figure 2 it is possible to see the dispersion of characteristics according to score and the correlation between them. The result of the PCA for the C11 cross, used as an example, shows that NS, NP and NB are positively and strongly correlated with GY; this is to be expected, since they are variable production components. Thus the higher the value of these components, the higher the value of GY for the progeny, with one thereby exerting a strong influence over the

Figure 1 - Biplot graph with dispersion of the 37 crosses and controls and vector projection for the traits: number of days to flowering (NDF); number of days to maturity (NDM); plant height at maturity (PHM); height of first pod insertion (HPI); number of branches per plant (NB); number of pods per plant (NP); number of seeds per plant (NS); 100-seed weight (HSW); grain yield per plant (GY)

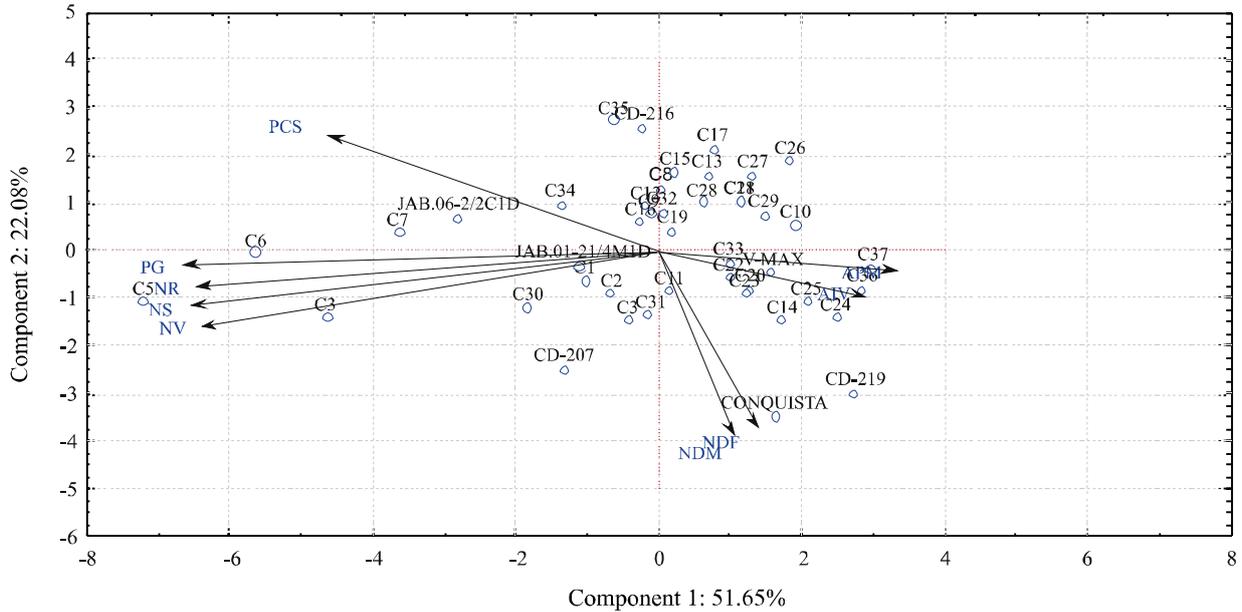


Table 1 - Genealogy of crosses made between conventional parents and carriers of the glyphosate-resistant gene (RR), with the respective numbers assigned to each cross, progeny per cross, progeny selected by PCA, and progeny selected by PCA and compared with the traits AV and Lg

Nbr of cross	Conventional parent		Monsoy RR parent	Nbr of progeny per cross	Nbr of progeny selected by PCA	Nbr of progeny selected by PCA + AV and Lg
	UNESP/FCAV	x				
C1	JAB.04-1/5A4D	x	M 7639 RR	63	8	3
C2	JAB.04-1/5A4D	x	M 8230 RR	36	3	3
C3	JAB.04-1/5A4D	x	M 8360 RR	44	6	5
C4	JAB.04-1/5A4D	x	M 8921 RR	24	2	2
C5	JAB.03-3/1C3D	x	M 7211 RR	4	0	0
C6	JAB.03-3/1C3D	x	M 7578 RR	5	0	0
C7	JAB.03-3/1H1D	x	M 7578 RR	24	2	2
C8	JAB.03-3/1H1D	x	M 7908 RR	39	5	1
C9	JAB.03-3/1H1D	x	M 8336 RR	29	3	3
C10	JAB.03-3/1H2D	x	M 7211 RR	40	3	1
C11	JAB.03-3/1H2D	x	M 7639 RR	30	2	2
C12	JAB.03-3/1H2D	x	M 7908 RR	28	3	2
C13	JAB.03-3/1H2D	x	M 8336 RR	36	4	1
C14	JAB.03-3/1H2D	x	M 8360 RR	41	3	1
C15	JAB.01-21/4M1D	x	M 7639 RR	20	3	2
C16	JAB.01-21/4M1D	x	M 8230 RR	44	9	2
C17	JAB.01-21/4M1D	x	M 8336 RR	37	5	3
C18	JAB.05-5/4A2D	x	M 7578 RR	64	7	4

Continuação Tabela 1

C19	JAB.05-5/4A2D	x	M 7908 RR	47	5	4
C20	JAB.05-5/4A2D	x	M 8221 RR	48	5	0
C21	JAB.05-1/5C3B	x	M 7211 RR	40	3	1
C22	JAB.05-1/5C3B	x	M 7639 RR	48	12	6
C23	JAB.05-1/5C3B	x	M 8211 RR	48	8	3
C24	JAB.05-1/5C3B	x	M 8336 RR	36	2	1
C25	JAB.05-1/5C3B	x	M 8360 RR	48	2	2
C26	JAB.05-8/2D3C	x	M 7211 RR	26	2	1
C27	JAB.05-8/2D3C	x	M 7636 RR	30	4	3
C28	JAB.05-8/2D3C	x	M 7908 RR	52	2	1
C29	JAB.05-8/2D3C	x	M 8336 RR	48	4	0
C30	JAB.06-2/3I3D	x	M 8230 RR	38	3	3
C31	JAB.06-2/3I3D	x	M 8236 RR	55	8	5
C32	JAB.06-2/2C1D	x	M 7211 RR	35	4	2
C33	JAB.02-26/1K1B	x	M 7211 RR	35	0	0
C34	JAB.02-30/1G4A	x	M 8360 RR	40	2	1
C35	JAB.02-3/6A4D	x	M 7211 RR	6	0	0
C36	JAB.02-3/6A4D	x	M 8360 RR	53	7	4
C37	JAB.02-3/6A4D1	x	M 8360 RR	41	5	3
TOTAL				1382	146	77

others. The HPI and PHM are also correlated, being positioned in the same quadrant, where it can be seen that the higher the HPI, the higher the PHM. For the variables NDF and NDM however, the results showed a negative correlation, i.e. progeny taking less time to start flowering, will not also always take less time to begin physiological maturity, which determines their cycle or maturity group. The type of soybean growth, classified as determinate, indeterminate or semi-determinate (SEDIYAMA; TEIXEIRA; REIS, 2005), may also influence the relationship between HPI/PHM and NDF/NDM. An interesting fact for evaluation in this analysis is that there is no direct correlation of the variable 100-seed weight (HSW) to the production components. This can be explained by the fact of this variable not directly influencing the GY alone, but only when combined with the other variables since generally the larger the NB and NP, the larger the NS and the smaller the HSW.

Analysis was then done of the dispersion of the progeny along the coordinate factor plane, also shown in Figure 1, where they were distributed according to their values and projection of their variables. In the result of this analysis, still for the C11 cross, it is possible to see a large group of progeny located in the centre of the plane, within

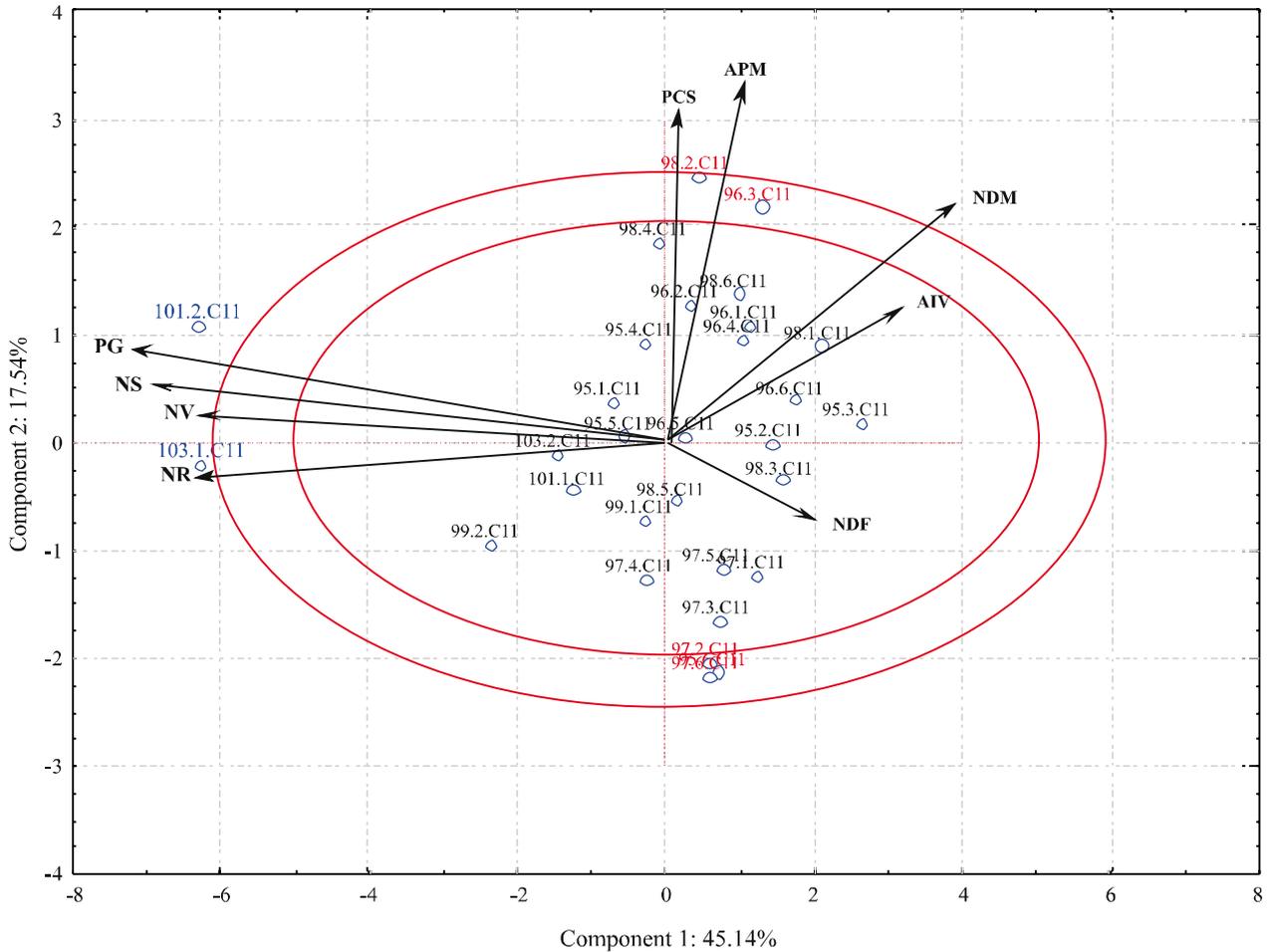
the boundaries of the predefined scales. This shows that in this group there are no differences between the progeny. For those progeny located between the scale boundaries (in red), it can be said that they have some differences to the progeny described above, and may be selected as superior, albeit using less stringent criteria. For those progeny at the edges, it can be seen that these differ from the others due to having specific characteristics with high values, that make them superior to the rest.

Thus, the further the progeny is from the centre of the coordinate system, the more specific is the pattern it presents. Progeny 101.2.C11 and 103.1.C11 therefore show a specific pattern due to having higher values of GY, NS, NP and NB; progeny 98.2.C11 and 96.3.C11, due to having higher values of PHM, HSW, NDM and HPI; while progeny 97.6.C11, 95.6.C11 and 97.2.C11 have low values of PHM, HSW, NDM and HPI.

Since in most breeding programs production is the factor of greatest importance, the progeny 103.1.C11 and 101.2.C11 were selected to represent their respective cross in the final evaluation.

This same process, described only for the C11 cross, is independently applied to all the other crosses being studied,

Figure 2 - Biplot graph with dispersion of the 30 soybean genotypes from crossing C11, as a function of Factor 1 x Factor 2, and vector projection for the traits: number of days to flowering (NDF); number of days to maturity (NDM); plant height at maturity (PHM); height of first pod insertion (HPI); number of branches per plant (NB), number of pods per plant (NP); number of seeds per plant (NS); 100-seed weight (HSW), grain yield per plant (GY). Genotypes selected on the less stringent scale are marked in red and those selected on the more precise scale in blue



thus allowing the selection only of superior progeny within each cross and the later identification of which obtained the greatest number of representatives in the final selection. This is important in order to determine whether the parents used in each cross were efficient in transmitting any marked superior characteristics to their offspring.

Analysis of the data related to the production components (NB, NP, NS and GY) is of great importance in the selection of productive genotypes. Costa *et al.* (2004), Alcantara Neto *et al.* (2011), Gomes *et al.* (2007) and Vianna (2013) showed satisfactory results in selection gains for these characteristics and in the relative contribution of these to grain yield in the soybean.

The number of progeny selected for each cross is variable, and thus not all the crosses in the study

were represented in the final selection. Another point to be discussed concerns the stringency of the selection, which in this study was set by two pre-defined scales. It is up to the researcher to decide which is the most appropriate criterion as regards accuracy to be imposed on the analysis and consequently on the number of selected individuals.

The final selection of progeny resulting from PCA, and later from the comparison with the data from the variables AV and Lg, resulted in the selection of 77 progeny from all the crosses used in the study, i.e. a selection of 5.6% of the total, plus one plant (samples) from the V-Max controls, one plant from CD-219, three plants from CD- 207 and three plants from CD- 216 which were also selected by the PCA. Although the

percentage of progeny selected is relatively low due to the stringency of the selection, it is in accordance with the selection index previously stipulated.

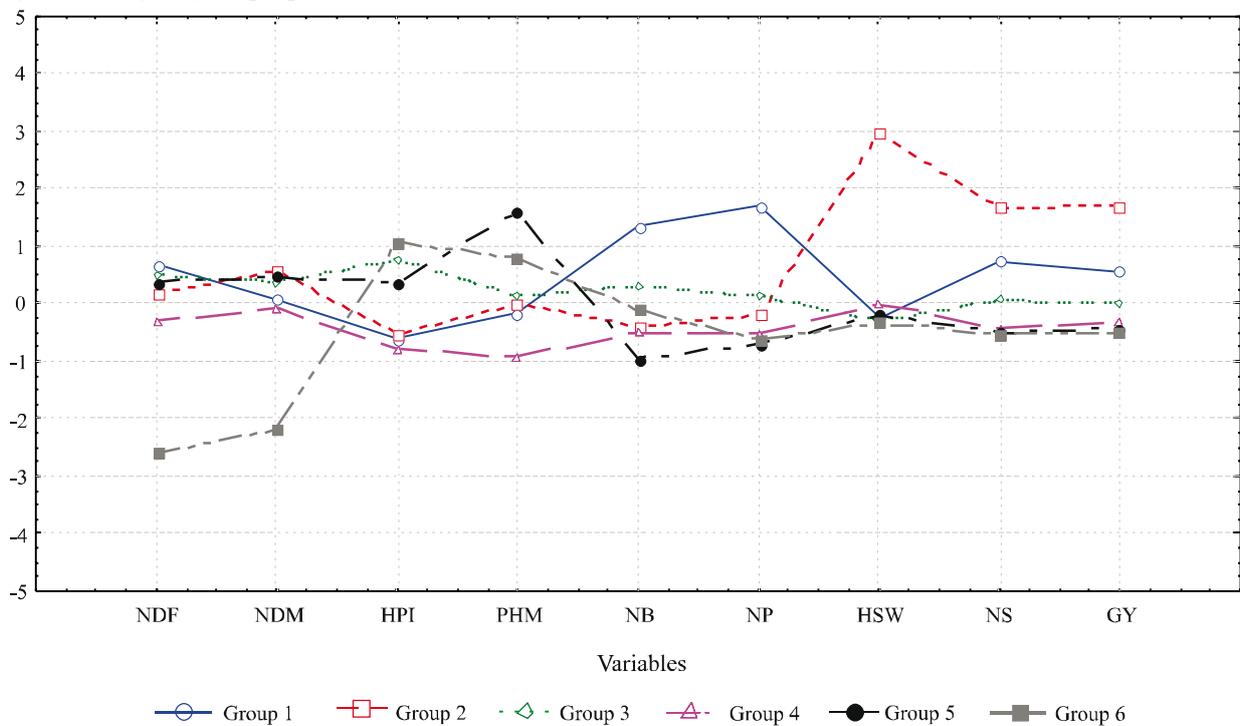
In relation to the parents involved in the crosses, the analysis found that the female parents JAB.04-1/5A4D and JAB.05-1/5C3B were the most efficient, with 13 and 12 progeny selected respectively. In turn, the male parent M 8360 RR was the most effective, with 16 progeny selected. When analysing crosses involving the combination of these more efficient parents, it can be seen that cross C3 has five progeny selected while cross C25 only two; results which are lower when compared to the C22 cross which presented the best combination with six progeny selected. This result is also shown in Table 1. This may happen due to the parents JAB.04-1/5A4D, JAB.05-1/5C3B and M 8360 RR having a high general combining ability (GCA) with a high frequency of favourable alleles, which does not mean they will consequently also have a high specific combining ability (SCA) and thus result in the best combinations, they may not possess the necessary complementarity between themselves, or have no genetic divergence which would allow further exploration of the genetic variability of such combinations.

Figure 3 shows the biplot graph generated by cluster analysis using the K-means method, containing the mean and distribution of each group according to the data of the traits, and based on the Euclidean distance between groups. Interpretation of this analysis is best if done independently, i.e. observing group by group, and then done generally.

Group one has the highest value for the trait NDF however, this value decreases on reaching maturity, getting close to the average (NDM). It also presents progeny of below average size, as can be seen for the traits HPI and PHM. For NR and NV the values are relatively high, whereas HSW was below average and, coupled with the fact of NS not being such a significant variable for this group, the trait GY was affected, although remaining above the average. This shows that the group has good productive potential, with the highest values for NR and NV which possibly, due to environmental factors such as water stress occurring during part of the cycle, led to low HSW values, however if the environmental conditions favour the crop this result could be positive.

Group two has a cycle which is slightly above average, but compared to the other groups, shows

Figure 3 - Biplot graph of the centroid profile of each formed group, using the K-means method and Euclidean distance, based on the traits: number of days to flowering (NDF); number of days to maturity (NDM); plant height at maturity (PHM); height of first pod insertion (HPI); number of branches per plant (NB); number of pods per plant (NP); number of seeds per plant (NS); 100-seed weight (HSW); grain yield per plant (GY)



the highest value, characterised by being formed of progeny with a longer cycle from germination to physiological maturity. It also presents a low HPI, even with a PHM very close to the average, and in this case, the lower values of HPI are not as significant. The traits NR and NV are below the average, but HSW has a significantly high value, which is characteristic of this progeny group, besides the high value for NS. Thus, the progeny of this group present the highest values found for GY among the groups formed. This shows that the combination of HSW and NS directly determines the productive potential of the progeny, strain or cultivar.

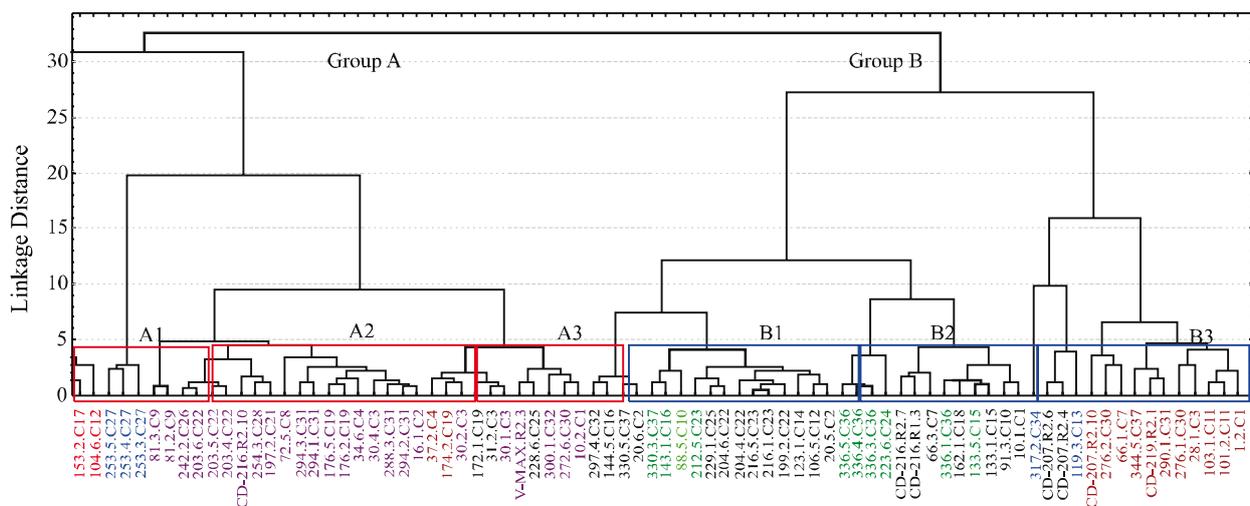
For group three, it can be seen that while some variables have values above and below the average, the group is generally characterised by progeny which maintain a regular overall average. The most significant variation occurs in the trait HPI, which in some way may be an undesirable characteristic, especially when presenting a low PHM, since this factor may reduce the productive capacity of the genotype (GY) mainly affecting the traits NB and NP. In group four, all variables have values below the average. For NDF and NDM this is a positive feature, this being the second group having more precocious progeny. It has the lowest values for both HPI and PHM, being the group with the smallest progeny. When it comes to production-component variables such as NB, NP, HSW, NS and GY, the below-average values found in this group are a very negative point, seeing as how they characterise less-productive progeny.

Group five has progeny with above average values for NDF, NDM, HPI and PHM, especially for PHM, however for that value to give better results, the traits NB, NP, HSW and NS should also follow this average, which did not happen, resulting in a low value for GY. Finally, group six is characterised by progeny having values significantly below average for NDF and NDM, with this being the group that has the earliest cycle among the groups under study. It is very common to find materials that are characterised by their precocity also being associated with low values for HPI and PHM, precisely because they have shorter reproductive and developmental stages. Therefore, the influence of the environment takes on great importance. In this case, the group did not display this characteristic, since the values are above the average. This factor is greatly influenced by the type of growth of the soybean (MARQUES; ROCHA; HAMAWAKI, 2008). When dealing with the production-component variables, these have below-average values for the traits NB, NP, HSW, NS and GY, resulting in the least productive progeny group among all the groups analysed.

This result assists the researcher in the use of a specific group of progeny according to a characteristic of interest, of a possible cultivar having differential characteristics or in targeted crossing and backcrossing.

Cluster analysis employing the Ward method resulted in the division of the progeny into two major groups, called groups A and B (Figure 4). With more

Figure 4 - Dendrogram of hierarchical cluster analysis using the Euclidean distance and the linkage between groups by the Ward method, showing the progeny and classification group to which they belong. The X axis represents the progeny and the Y axis, the distances for the agronomic traits characteristics: number of days to flowering (NDF); number of days to maturity (NDM); plant height at maturity (PHM); height of the first pod insertion (HPI); number of branches (NB); number of pods (NP); number of seeds (NS); 100-seed weight (HSW); grain yield per plant (GY)



criteria, using as a basis a value of five for the linkage distance, the formation of six subgroups can be seen, with three subgroups in group A, referred to as A1, A2 and A3, and three subgroups in group B referred to as B1, B2 and B3. The groups and subgroups can be divided based on the linkage distance indicated by the researcher, it being left to the researcher to determine the most appropriate distance according to genotype, stringency of selection and ultimate goal. Although the number of progeny per group was 40 and 45, and in the subgroups varied from 10 to 19, there was adequate distribution between the groups and subgroups, demonstrating the efficiency of the analysis in their determination and separation.

In the dendrogram generated by the Ward method, there was some disagreement with the K-means method. This fact can be explained by the two methods, which although different follow the same principle, therefore allowing the possibility of comparing them.

A similar result was found by José *et al.* (2012) when comparing hierarchical CA, non-hierarchical CA (K-means) and the PCA of the hydro-physical variables of soil, which were similar to the clusters and their interactions. Costa (2008) highlights that the results of K-means clustering were similar to those of the Ward method in the qualitative and quantitative study of the resistance of soybean genotypes to Asian soybean rust.

After the progeny reach a satisfactory level of homozygosity a selection, specifically for the RR gene, will be made by the application of glyphosate.

CONCLUSIONS

1. Principal Components Analysis aids the selection of 77 progeny with superior characteristics of agronomic importance, especially for those components related to grain production and the carriers of the glyphosate resistant (RR) gene. It also makes possible the identification of the parents JAB.04-1/5A4D and JAB.05-1/5C3B and M 8360 RR as being the most efficient in generating superior progeny;
2. Cluster analysis by the K-means method enables identification and characterisation of the different progeny groups based on the agronomic traits that most influenced them. Cluster analysis by the Ward method makes possible the generation of the dendrogram and characterisation of the similarity and divergence existing among selected groups of progeny;
3. Multivariate methods are effective in aiding selection of progeny with multiple characteristics of agronomic interest, thus confirming the potential of these methods for use in soybean breeding programs.

REFERENCES

- ABREU, A. B. G.; DA MATA, M. H. R.; MONTAGNER, E. Desenvolvimento e validação de método de análise de glifosato em grãos de soja. **Química Nova**, v. 31, n. 1, p. 5-9, 2008.
- ALCANTARANETO, F. *et al.* Análise de trilha do rendimento de grãos de soja na microrregião do Alto Médio Gurguéia. **Comunicata Scientiae**, v. 2, n. 2, p. 107-112, 2011.
- ALEIXO, S. S.; SOUZA, J. G. D.; FERRAUDO, A. S. Técnicas de análise multivariada na determinação de grupos homogêneos de produtores de leite. **Revista Brasileira de Zootecnia**, v. 36, n. 6, p. 2168-2175, 2007.
- BARBIERI, R. L. *et al.* Divergência genética entre populações de cebola com base em marcadores morfológicos. **Ciência Rural**, v. 35, n. 2, p. 303-308, 2005.
- COSTA, M. M. **Herança quali-quantitativa e marcadores moleculares para seleção assistida de genótipos de soja resistentes à ferrugem asiática**. 2008, 77 f. Tese (Doutorado em Genética e Melhoramento de Plantas) - Faculdade de Ciências Agrárias e Veterinárias - UNESP Campus Jaboticabal, 2008.
- COSTA, M. M. *et al.* Ganho genético por diferentes critérios de seleção em populações segregantes de soja. **Pesquisa Agropecuária Brasileira**, v. 39, n. 11, p. 1095-1105, 2004.
- EMPRESABRASILEIRADEPESQUISAAGROPECUÁRIA. **Tecnologias de produção de soja - Região Central do Brasil, 2012 e 2013**. Londrina: Embrapa Soja, 2011. 261 p. (Sistemas de Produção / Embrapa Soja, n.15).
- FERRAUDO, A. S. **Técnicas de Análise Multivariada: uma introdução**. São Caetano: StatSoft South América, 2010.
- GAZZIERO, D. L. P.; VARGAS, L.; ROMAN, E. S. Manejo e controle de plantas daninhas em soja. *In*: VARGAS, L.; ROMAN, E. S. **Manual de manejo e controle de plantas daninhas**. Bento Gonçalves: Embrapa Uva e Vinho, 2004. p. 595-635.
- GOMES, C. N. *et al.* Caracterização morfoagronômica e coeficientes de trilha de caracteres componentes da produção em mandioca. **Pesquisa Agropecuária Brasileira**, v. 42, n. 8, p. 1121-1130, 2007.
- HAIR, J. F. *et al.* **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009. 688 p.
- JOSÉ, J. V. *et al.* Análise multivariada de variáveis físico-hídricas de diferentes solos. *In*: WORKSHOP DE INOVAÇÕES TECNOLÓGICAS NA IRRIGAÇÃO, 4., 2012, Fortaleza. **Anais ...** Fortaleza: INOVAGRI, 2012. p. 1-5.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v. 23, n. 3, p. 187-200, 1958.
- LIMA, D. B. C. *et al.* Seleção de herbicidas para o controle de plantas voluntárias de soja resistente ao glyphosate. **Revista Brasileira de Herbicidas**, v. 10, n. 1, p. 01-12, 2011.

- MARQUES, M. C.; ROCHA, J. E. S.; HAMAWAKI, O. T. Avaliação do teor de óleo de cultivares de soja da UFU visando a produção de biodiesel. **Horizonte Científico**, v. 2, n. 1, p. 207-219, 2008.
- PLINE-SRNIC, W. Physiological mechanisms of glyphosate resistance. **Weed Technology**, v. 20, n. 2, p. 290-300, 2006.
- REIS, E. F. *et al.* Comparação de procedimentos de seleção para produção de grãos em populações de soja. **Ciência Rural**, v. 34, n. 3, p. 685-692, 2004.
- RODRIGUES, J. I. S. *et al.* Associação de marcadores microssatélites com teores de óleo e proteína em soja. **Pesquisa Agropecuária Brasileira**, v. 48, n. 3, p. 255-262, 2013.
- SEDIYAMA, T.; TEIXEIRA, R. C.; REIS, M. S. Melhoria da soja. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. 2 ed. Viçosa: Editora UFV, p.553-603. 2005.
- SILVA, M. A. *et al.* Análise de trilha para caracteres morfológicos do feijão-bravo (*Capparis flexuosa*) no cariri paraibano. **Archivos de Zootecnia**, Cordoba, v. 58, n. 221, p. 121-124, 2009.
- SILVA, M. *et al.* Discriminação de diferentes classes de solos irrigados com águas salinas, na região de Mossoró (RN), com o uso de análise multivariada. **Ambiência**, v. 6, n. 2 p. 261-270, 2010.
- STATSOFT. **Statistica**: data analysis software system, version 7. Tulsa, 2004. Disponível em: <<http://www.statsoft.com>>. Acesso em: 23 jun. 2013.
- VALLADARES, G. S. *et al.* Análise dos componentes principais e métodos multicritério ordinais no estudo de Organossolos e solos afins. **Revista Brasileira de Ciência do Solo**, v. 32, n. 1, p. 285-296, 2008.
- VASCONCELOS, E. S. *et al.* Estimativas de ganho genético por diferentes critérios de seleção em genótipos de alfafa. **Revista Ceres**, v. 57, n. 2, p. 205-210, 2010.
- VIANNA, V. F. **Estudo genético e influência de caracteres na seleção de genótipos superiores de soja**. 2013. 92 f. Tese (Doutorado em Agronomia) - Universidade Estadual Paulista "Júlio de Mesquita Filho" Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, 2013.