Genetic divergence of soybean genotypes in relation to grain components

Josiane Isabela da Silva Rodrigues1* Klever Márcio Antunes Arruda2 Cosme Damião Cruz1 Everaldo Gonçalves de Barros1 Newton Deniz Piovesan1 Maurilio Alves Moreira4

1Institute of Biotecnology Applied to Agronomy, Universidade Federal de Viçosa (UFV), Campus Universitário, Avenida PH Rolfs, 36570-400, Viçosa, MG, Brasil. E-mail: josianeisabela@gmail.com. 2Corresponding author.

ABSTRACT: The objective of this paper was to estimate the genetic divergence among 49 soybean (Glycine max L. Merrill.) genotypes to assist grain quality-focused breeding programs in the choice of progenitors. The genetic divergence was estimated using the Mahalanobis generalized distance from the percentages of protein, oil, and fatty acids oleic, linoleic and linolenic after cultivation of genotypes in different environments. Genotypes were grouped by agglomerative methods and the two and three-dimensional projections of the distance matrix were obtained. The average protein and oil contents in the four environments ranged from 34.25 to 45.18% and from 16.48 to 23.01%, respectively. The average contents of the fatty acids oleic, linoleic and linolenic ranged from 20.2 to 42.41%, from 44.17 to 63.18%, and from 5.89 to 10.39%, respectively. The genetic distances ranged from 0.11 to 251.02 and indicated genetic variability among the accessions. The most divergent pair of accessions was P417360/CD01RR8384, followed by P417360/B3PTA213-3-4 and P417360/BRARC-8. The most similar pair of accessions was CS3032PTA190-5-1, followed by UFV18/M-SOY9914 and BRSMG Garantia/CD983321RR. In this study we indicated as promising in terms of genetic variability the hybridizations involving BARC-8, CD2013PTA, CD01RR8384, CS303TNKCA, P1181544, and P417360. Among these genotypes we can stand out BARC-8 and CD2013PTA, with protein contents above 43%, and CD01RR8384 and CS303TNKCA, with oil contents above 20%. The use of these genetically divergent genotypes and with high phenotypic means in future crosses should produce desirable recombinants for grain quality.

Key words: Glycine max L. Merrill., protein, oil, fatty acids.

INTRODUCTION

Soybean has prospect of improvement for oil and protein contents and for grain oil composition, because a small fraction of its germplasm constitutes the genetic basis of improved cultivars. Historically the use of elite materials maintained narrow the genetic basis of improved varieties.

In the 80 decade, 15% of the cultivated soybean in Brazil were introduced in the southern region of the United States and the remaining 85% came from hybridization of the North American introductions (VELLO et al., 1988). According to HIROMOTO & VELLO (1986), 79 cultivars recommended at that time descended from 26 ancestors of which 11 formed 89% of the gene pool. VELLO et al. (1988) at the same time observed an average relatedness value of 0.16 between 69 soybean cultivars, the equivalent to an effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006). Equal relatedness value and effective size equal to N=11-15. Other 100 cultivars released between 1984 and 1998, showing an average relatedness value equal to 0.21 (BONATO et al., 2006).
studies with molecular markers it is clear the similarity among the Brazilian cultivars (vieira et al., 2009; priolli et al., 2010). Thus, a relatively low level of diversity is maintained in different breeding programs, although there is some heterogeneity in a minority of the programs, according to priolli et al. (2004).

Due to the limited genetic diversity in soybean elite germplasm and to the relative lack of variability in breeding programs, the use of new sources of germplasm can be useful to broaden the genetic basis of breeding programs and to enable the new cultivars additional gains in the protein and oil contents or improvement in oil fraction. The objective of this research was to estimate the genetic divergence among 49 soybean genotypes, based on the protein, oil and fatty acids contents in order to help breeding programs focused on quality in the choice of progenitors.

Several multivariate methods can be used to estimate genetic divergence. The methods of principal components and canonical variables seek to evaluate the similarity through a graphic dispersion, being usually two Cartesian axes. The agglomerative methods are based on the use of dissimilarity measures as the Euclidean distance and Mahalanobis generalized distance, etc. Among the most used ones, the mean standardized Euclidean distance circumvents problems due to the number and scale of the variables, but with the inconvenience of not taking into account the residual correlations. The Mahalanobis distance takes into account the correlation between the characters which is advantageous because most of the time there is a certain degree of correlation between the agronomic traits. The advantage of these methods in relation to the dialelic analyses is that they do not require hybrid combinations. In hybridization, manual pollination can be costly, difficult to perform with low probability of success in obtaining the hybrid seed, or even the hybrid seed that can be obtained in a very small quantity for proper evaluations. Then the predictive methods are faster and simpler to apply.

MATERIALS AND METHODS

Forty-nine soybean genotypes for which we expected a wide variation in oil and protein contents, according to preliminary analyses, were cultivated in Viçosa, MG (12/2009, 20° 45' S, 42° 52' W), Visconde do Rio Branco, MG (02/2010, 21°00' S, 42°50' W) and São Gotardo, MG (02/2010, 10/2011, 19°18' S, 46°02' W) using the randomized block design with three replicates. In each trial conducted in the field we sown 15 seeds per 1m length row with spacing between rows of 0.5m. The harvest was done manually and after beneficiation the grains were ground in an industrial mill (model MA020, Marconi, Brazil). The bran was analyzed for protein, oil and fatty acids oleic, linoleic and linolenic contents by the infrared spectrometry method using a FT-NIR spectrometer (model II Antaris, Thermo Scientific). For analysis of phenotypic data we performed analysis of variance based on the factorial design using the statistical model below:

$Y_{ijk} = \mu + G_i + E_j + G\times E_{ij} + B/E + \varepsilon_{ijk}$

In which: $Y_{ijk}$: observation of the k-th block evaluated in the i-th genotype and j-th environment; m: general mean; B/E: effect of the block k within the environment j; Gi: effect of the genotype i; Ej: effect of the environment j; GEij: effect of the interaction between genotype i and environment j; and $\varepsilon_{ijk}$: random error associated with ijk observation. The effect of genotypes (Gi) was considered fixed and the effect of environments (Ej) was considered random. The variance components were estimated as follows:

$\varphi_\sigma^2 = \frac{QMG - QMGE}{e r}$

$\sigma_\varepsilon^2 = \frac{(QMG - QMRE)}{r}$

In which: MSG is the mean square of genotypes; MSGE is the mean square of GE interaction; MSR is the mean square of residue; e is the number of environments; r is the number of replicates; and in which g is the number of genotypes.

We used the Mahalanobis gener, GMA, Tocher, and modified Tocher. The Mojema method (mojema, 1977) and value of k = 1.25 were used as criteria to identify the cut off values from the dendrogram. The consistency of the grouping pattern was evaluated by bootstrap using 5,000 simulations. The importance of variables in determining the distance values among the pairs of accessions was evaluated by the Singh method (sINGH, 1981). All analyses were performed using the GENES program <http://www.ufv.br/dbg/genes/>.

RESULTS AND DISCUSSION

Analyses of variance revealed significant differences among the 49 soybean genotypes in relation to protein, oil and fatty acids oleic, linoleic and linolenic contents. Coefficients of variation were below 5% and demonstrated precision in controlling the causes of experimental variation. Based on the ratio between the highest and lowest residual mean square there was also homogeneity in residual variances. Effects of genotype, environment and GE interaction were
Genetic divergence of soybean genotypes in relation to grain components.

Significant (P<0.01) for the five traits, indicating genetic variability among the accessions, difference among the environments and differential behavior of the accessions in the environments. The effects of environments had the highest contributions to variation in the characters in relation to the other main effects. Results of the analyses of variance are presented in table 1.

The average protein and oil contents in the four environments ranged from 34.25 to 45.18% and from 16.48 to 23.01%, respectively. The average contents of the fatty acids oleic, linoleic and linolenic ranged from 20.2 to 42.41%, from 44.17 to 63.18% and from 5.89 to 10.39%, respectively. Among the environments, the variation in the average percentages of protein and oil was from 32.07 to 48.57% and from 14.40 to 24.20%, respectively. The variation in the average percentages of the acids oleic, linoleic and linolenic was from 16.38 to 57.35%, from 29.56 to 69.51%, and from 4.27 to 11.54%, respectively. These values confirmed a wide variation among the accessions in relation to each component of the grain, as it was expected for the protein and oil contents (Table 2).

The Mahalanobis distances from the variation in the components of the grains ranged from 0.11 to 251.02 and indicated genetic variability among the accessions. The most divergent pair of accessions was PI417360/CD01RR8384, followed by PI417360/B3PTA213-3-4 and PI417360/BARC-8. The most similar pair of accessions was CS3032PTA276-1-2/CS3032PTA190-5-1, followed by UFV18/M-SOY8914 and BRSMG Garantia/CD983321RR. Regarding the importance of the environmental variation.

The grouping by the Tocher method differentiated PI417360 in relation to all other 48 accessions, which were gathered in a single group. The modified method produced an identical grouping, which highlights the genetic distance of this accession in relation to the others. In the grouping by the UPGMA method, PI417360 proved to be the most divergent accession again, followed by PI181544. The accessions with the initials CS3032PTA proved to be close to each other as well as the accessions with the initials B3PTA. All the cultivars with the initials CS were close to each other in the projection, except CS3033TNKCA. The accessions BARC-8 and BR8014887 proved to be very close to the previous genotypes, indicating elevated genetic similarity between all of them. Other genotypes showed to be genetically close to each other, as PI371610 and PI3571611, most cultivars with the initials CD, except CD2013PTA and CD01RR8384, and the cultivars with the initials UFV and M-SOY. Cultivars A7002 and Suprema also showed a high degree of genetic similarity. The cultivar CD01RR8384, conversely, showed certain genetic divergence in relation to the previous ones. Most of the time there was greater proximity among cultivars of breeding programs of the same institution, which was expected. The values of cophenetic correlation, distortion and stress in the projection were equal to 0.95, 6.89% and 26.25%, respectively, which indicated good adjustment of the original and graphic values and small distortion of the distances in the dendrogram (Figure 1).

The statistical criterion of Mojema (MOJEMA, 1977), based on the relative size of the fusion levels in the dendrogram, provides the possibility of making cuts in different stages of the grouping for groups definition. A cut performed at 18.56% of the dissimilarity establishes the same relationship observed in the grouping by the Tocher method. while a cut at 7.31% of the dissimilarity

Table 1 - Analysis of variance of the protein, oil and fatty acids oleic, linoleic and linolenic contents of the 49 soybean genotypes cultivated in Viçosa-MG (12/2009), Visconde do Rio Branco-MG (02/2010) and São Gotardo-MG (02/2010, 10/2011).

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>DF</th>
<th>Mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Protein (%)</td>
</tr>
<tr>
<td>Blocks/Environments</td>
<td>8</td>
<td>11.73</td>
</tr>
<tr>
<td>Genotypes</td>
<td>48</td>
<td>89.22**</td>
</tr>
<tr>
<td>Environments</td>
<td>3</td>
<td>236.44</td>
</tr>
<tr>
<td>GE</td>
<td>144</td>
<td>6.56**</td>
</tr>
<tr>
<td>Error</td>
<td>384</td>
<td>2.00</td>
</tr>
<tr>
<td>Mean</td>
<td>-</td>
<td>39.67</td>
</tr>
<tr>
<td>CV(%)</td>
<td>-</td>
<td>3.56</td>
</tr>
<tr>
<td>Higher/LowerMSR</td>
<td>-</td>
<td>3.77</td>
</tr>
<tr>
<td>CVg/CVe</td>
<td>-</td>
<td>1.86</td>
</tr>
</tbody>
</table>

**Significant at 1% probability by the F test; DF, degree of freedom; MSR, residual mean square; CVg, coefficient of genetic variation; CVe, coefficient of environmental variation.
(another value indicated by the Mojema method) established two main groups, where six accessions are not included in any of the established groups (BARC-8, CD2013PTA, CD01RR8384, CS303TNKCA, PI181544 and PI417360). For this number of groups it is possible to observe the same relationships discussed earlier, which reflects the larger genetic distances of the ungrouped accessions in relation to each group of accessions.

At the bi and three-dimensional projections it was clear the divergence of the accessions PI417360 and PI181544 in relation to the majority. Values of cophenetic correlation, distortion and stress at the bi and three-dimensional projections were equal to 0.96 and 0.9, 39.72 and 45.68%, and 45.31 and 64.88%, respectively, indicating in both cases low efficacy in representing the distance matrix. Therefore, the projection that best represents the genetic distance relationships is the dendrogram obtained by the UPGMA method.

Other studies about genetic diversity using the Mahalanobis statistics have been made in

Rodrigues et al.

soybean. JAIN et al. (2016) using the Mahalanobis distance observed that 39 soybean genotypes formed 11 distinct groups. PAWAR et al. (2015) using the same measure of distance observed that 80 soybean genotypes formed 14 groups, where the level of genetic diversity was probably lower in relation to the previous study. The most recent studies with Brazilian cultivars, in turn, do not use multivariate analysis. The most recent studies with soybean germplasm related to genetic diversity use molecular markers (BONATO et al., 2006; VIEIRA et al., 2009; Prioli et al., 2010) and the Malecot’s coefficient of relatedness (MIRANDA, 2005; MIRANDA et al., 2007). In these studies the authors reported low to moderate diversity in the Brazilian cultivars.

The genetic diversity studies can identify among many accessions the most promising combinations in terms of variability and those less promising for which it is expected a low variability in the crosses, as between accessions of the same group or between accessions with low genetic distance values (ALMEIDA et al., 2011). In this study we indicated as promising in terms of genetic variability the hybridizations involving BARC-8, CD2013PTA, CD01RR8384, CS303TNKCA, PI181544 and PI417360, once such accessions showed larger genetic distances in relation to the majority. So, these genotypes when used in hybridizations in breeding programs may increase the number of desired recombinants and/or be a source of germplasm to create superior genotypic constitutions.

The importance of genetic diversity studies for plant breeding lies in the fact that crosses involving genetically divergent progenitors are most likely in producing genetic variability and heterotic effect. Genetically close progenitors, conversely tend to share alleles in common, causing little complementarity and low vigor in the crossing due to the low level of heterozygosity. In this way, genetic diversity can be considered in the prediction of the potential of
populations, in the phase of selection of progenitors, avoiding generate populations with low genetic variability (RIAZ et al., 2008; FILHO et al., 2010).

CONCLUSION

The most genetically divergent pair of genotypes based on the Mahalanobis distance is PI417360/CD01RR8384, followed by PI417360/B3PTA213-3-4 and PI417360/BARC-8. Based on the average divergence values, the genotypes BARC-8, CD2013PTA, CD01RR8384, CS303TNKCA, PI181544, and PI417360 were the most divergent Among these genotypes we can stand out BARC-8 and CD2013PTA, with protein contents above 43%, and CD01RR8384 and CS303TNKCA, with oil contents above 20%. The use of these genotypes in new crosses should produce genic complementarity and new recombinants to the respective grain traits.

ACKNOWLEDGEMENTS

We thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support (314899/2014-6).

REFERENCES


