



# Genetic divergence in a soybean (*Glycine max*) diversity panel based on agro-morphological traits

M.B. Marconato<sup>1</sup>, E.M. Pereira<sup>1</sup>, F.M. Silva<sup>1</sup>, E.H. Bizari<sup>1</sup>, J.B. Pinheiro<sup>2</sup>,  
A.O. Mauro<sup>1</sup> and S.H. Unêda-Trevisoli<sup>1</sup>

<sup>1</sup>Departamento de Produção Vegetal,  
Faculdade de Ciências Agrárias e Veterinárias,  
Universidade Estadual de São Paulo “Júlio de Mesquita Filho”, Jaboticabal,  
SP, Brasil

<sup>2</sup>Laboratório de Diversidade Genética e Melhoramento,  
Departamento de Genética, Escola Superior de Agricultura “Luiz de Queiroz”,  
Universidade de São Paulo, Piracicaba, SP, Brasil

Corresponding author: M.B. Marconato  
E-mail: marcela\_bmarconato@yahoo.com.br

Genet. Mol. Res. 15 (4): gmr15048980  
Received July 20, 2016  
Accepted September 30, 2016  
Published November 21, 2016  
DOI <http://dx.doi.org/10.4238/gmr15048980>

Copyright © 2016 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

**ABSTRACT.** Owing to the narrow genetic basis of soybean (*Glycine max*), the incorporation of new sources of germplasm is indispensable when searching for alleles that contribute to a greater diversity of varieties. The alternative is plant introduction, which may increase genetic variability within breeding programs. Multivariate techniques are important tools to study genetic diversity and allow the precise elucidation of variability in a set of genotypes of interest. The agro-morphological traits of 93 soybean accessions from various continents were analyzed in order to assess the genetic diversity present, and to

highlight important traits. The experimental design was incomplete blocks (Alpha lattice, 8 x 12) with three replicates. Nine agro-morphological traits were analyzed, and principal component analysis and cluster analysis were performed, the latter by Ward's method. The dendrogram obtained contained eight subgroups, confirming the genetic diversity among the accessions and revealing similarities between 11 national genotypes. The geographical origin of the accessions was not always related to the clusters. The traits evaluated, and the methods used, facilitated the distinction and characterization of genotypes between and within groups, and could be used in Brazilian soybean breeding programs.

**Key words:** *Glycine max*; Germplasm; Multivariate analysis; Plant introduction

## INTRODUCTION

A major goal of plant breeding is the introduction of superior cultivars through the study and manipulation of germplasm (Bueno et al., 2006). There is a narrow genetic basis of soybean (*Glycine max*) cultivars in Brazil, because of the poor insertion of ancestral lineages and the existing relatedness between them (Wysmierski and Vello, 2013). Such narrowing causes less variability, lower levels of productivity, and cultivars that are less resistant to diseases and pests (Kisha and Diers, 1997; Manjarrez-Sandoval et al., 1997). To increase the genetic variability of cultures within breeding programs, a viable alternative is the incorporation of new sources of germplasm, such as genotypes known as plant introductions (PIs). The use of exotic germplasm is an important alternative in the process, and can contribute to the introduction of specific alleles of interest (Sneller et al., 1997).

The study of genetic diversity is of fundamental importance in understanding the genetic variability of populations and germplasm banks. Various multivariate analysis techniques may be used for this, such as principal component analysis (PCA) and cluster analysis (Cruz and Carneiro, 2003), which optimize genotype evaluation. Cluster analysis is the allocation of individuals or objects to groups, such that those that are in the same group are more similar to each other than those that are in other groups. The goal of this analysis is to maximize the homogeneity within groups while maximizing the heterogeneity between groups (Hair et al., 2005). PCA aims to simplify the description of a set of interrelated variables, and reduce the space of variables forming orthogonal axes that are linear combinations of the original variables, which are called principal components. This method transforms the original variables into new, uncorrelated variables. The principal component is a linear combination of the original variables, and the variance of each component is the amount of information explained by each one (Ferraudo, 2012).

The aims of this study were to evaluate a set of soybean accessions from various regions of the world based on agro-morphological traits of importance, study their genetic diversity using multivariate methods, and highlight traits of importance.

## MATERIAL AND METHODS

The experiment was conducted at an experimental station located at Faculdade de Ciências Agrárias e Veterinárias of Universidade Estadual de São Paulo, Jaboticabal, São

Paulo, Brazil, at 21°15'22" S and 48°18'58" W and an average of 595 m above mean sea level in altitude. The climate, according to the Köppen (1948) classification, is Aw and humid tropical, with a rainy season in the summer and dry season in the winter. The predominant soil type is Red Eutrophic Latossoil.

Sowing was conducted manually, after the planting area had been harrowed twice and ploughed deeply. The culture and management practices were conducted according to the technical guidelines for soybean provided by EMBRAPA (2012).

A total of 93 soybean genotypes (Table 1) were evaluated, which were provided by the EMBRAPA germplasm bank. The experimental design was incomplete blocks (Alpha lattice 8 x 12), totaling 93 treatments with three replicates. The genotypes were sown in November 2012 for cultivation in the agricultural year 2012/2013. Each plot consisted of four 5-m rows that were spaced 0.5 m apart, with a total area of 4 m<sup>2</sup>

**Table 1.** Characteristics of the accessions used in the study.

FN	PI	Origin	FN	PI	Origin
1	36906	Manchuria (China)	49	341254	Sudan
2	79861	China	50	341264	Liberia
3	84910	North Korea	51	360851	Japan
4	90251	South Korea	52	377573	China
5	133226	Indonesia	53	381660	Uganda
6	145079	Zimbabwe	54	381680	Uganda
7	148259	Indonesia	56	407744	China
8	148260	South Africa	57	407764	China
9	153681	El Salvador	58	416828	Japan
10	159097	South Africa	59	417563	Vietnam
11	159927	Peru	60	417581	EUA
12	164885	Guatemala	61	417582	EUA
13	165524	India	62	427276	China
14	166141	Nepal	63	438301	North Korea
15	170889	South Korea	64	90577	China
16	171437	China	65	159922	Peru
17	172902	Turkey	66	209839	Nepal
18	189402	Guatemala	67	222546	Argentina
19	200832	Burma Myanmar	68	240665	Philippines
20	203400	Brazil	69	281898	Malaysia
21	203404	Brazil	70	281911	Philippines
22	204333	Suriname	71	284816	Malaysia
23	204340	Suriname	72	306712	Tanzania
24	205384	Pakistan	74	281907	Malaysia
25	205912	Thailand	75	IAC 100	Brazil
26	210178	Taiwan	76	Paranagoiania	Brazil
27	210352	Mozambique	77	A7002	Brazil
29	215692	Israel	78	CD 215	Brazil
30	222397	Pakistan	79	Conquista	Brazil (TMG)
31	222550	Argentina	80	Pintado	Brazil (TMG)
32	229358	Japan	81	Sambaiba	Brazil (EMBRAPA)
33	239237	Thailand	82	Dowling	EUA
34	253664	China	83	Shira Nuhi (200526)	Japan
35	259540	Nigeria	84	Kinoshita (200487)	Japan
36	265491	Peru	85	Orba (471904)	Indonesia
37	265497	Colombia	86	Bignam	EUA
38	274454-A	Japan	87	227687	Japan
39	274454-B	Japan	88	171451	Japan
40	274507	China	89	VMáx	Brazil
41	283327	Taiwan	90	Potência	Brazil
42	285095	Venezuela	91	Sandra 1	Brazil
43	297550	Russia	92	Sandra 2	Brazil
44	306702	Tanzania	93	LQ 1050	Brazil
45	315701	EUA	94	LQ 1505	Brazil
46	322695	Angola	95	LQ 1421	Brazil
47	331793	Vietnam	96	LQ 1413	Brazil
48	331795	Vietnam			

FN, field number; PI, plant introduction.

Evaluations of agro-morphological traits, which required plant manipulation, were conducted on a sample of six plants from each plot at the R<sub>7</sub> state of maturity (Fehr and

Caviness, 1977). The remainder of the plants was manually harvested. To characterize the genotypes, nine agronomically important traits were considered. Grain yield (GY) was obtained after harvest. The corrected total weight was obtained based on 13% moisture and converted to kilograms per hectare (kg/ha). The number of pods (NP) and number of branches (NB) were obtained by counting the pods and branches, respectively, on each plant evaluated. The weight of 100 seeds (WHS), given in grams, was evaluated by weighing 100 seeds harvested from each plot. The oil content (OC) was analyzed using a near-infrared spectrometer (Model Tango Bruker, Bruker Optics Inc., Billerica, MA, USA), and the results were expressed as a percentage that was obtained from the average of three readings. The number of days to maturity (NDM) was calculated from the day of emergence to the day when at least 50% of the plants exhibited 95% maturation of the pods. The grain-filling period (GFP) was calculated as the number of days between the  $R_5$  and  $R_7$  stages, according to the Fehr and Caviness (1977) scale. Plant height at maturity (PHM) was the distance (cm) between the neck of the plant and the insertion point at the last productive pod. The height of insertion of the first pod (HIP) was the distance (cm) between the neck of the plant and the insertion of the first pod.

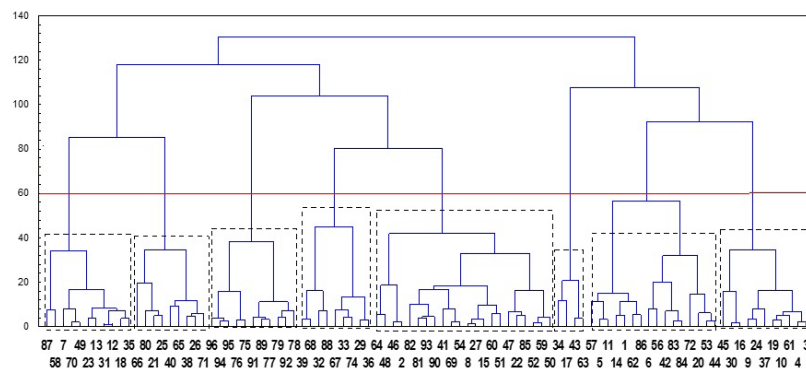
To estimate genetic divergence among the 93 accessions, we conducted multivariate analysis. Two exploratory approaches were used, PCA and cluster analysis by Ward's method, which depends on the existence of a dependency structure in the original set of variables. The data were standardized, so that all of the variables had zero mean and unit variances. The statistical software used was Statistica version 10 ([www.statsoft.com](http://www.statsoft.com)).

The goal of PCA is to evaluate the importance of each variable in relation to the total available variation among genotypes. Using this method, it is possible to exclude less important traits in the group studied (Cruz and Carneiro, 2003), and simultaneously determine which traits are the most important. After calculating the means for each replicate, the data obtained were processed by PCA with the covariance matrix giving eigenvalues that generated eigenvectors, which are linear combinations of the original variables. Only eigenvalues greater than one were considered, because these are components with a significant amount of information from the original variables (Kaiser, 1958).

Subsequently, the centroids of the genotypes, which were specific to each quadrant, were calculated based on the results of the PCA. With the data obtained, a two-dimensional graphic of the groups was produced, which displays the standardized values of the averages of the original variables. The similarity between genotypes was measured by the Mahalanobis distance (Mahalanobis, 1936), and connections between groups were obtained by Ward's method, whereby the distance between two groups is defined as the sum of squares of the two groups obtained from all of the variables. At each stage of the clustering procedure, the internal sum of squares is minimized on all partitions that are obtained by combining two earlier-stage groups (Ferraudo, 2012).

## RESULTS AND DISCUSSION

The dendrogram produced by the Ward method shows two groups separated by the maximum distance. At a shorter distance (60), eight subgroups were identified (Figure 1). The first subgroup had 11 genotypes (87, 58, 7, 70, 49, 23, 13, 31, 12, 18, and 35); 18% were African, 36% were Hispanic, and 45% were Asian (18% East Asian, 18% Southeast Asian, and 9% Southern Asian).



**Figure 1.** Dendrogram derived from a hierarchical cluster analysis using the Mahalanobis generalized distance and Ward's method for connecting groups based on agro-morphological traits. Eight subgroups (indicated with dashed lines) are below the solid red line.

The second group was composed of nine genotypes (66, 80, 21, 25, 40, 65, 38, 26, and 71); 33% were Latino and 66% were Asian (33% East Asian, 22% Southeast Asian, and 11% Southern Asian). The third group consisted entirely of Brazilian genotypes (96, 94, 95, 76, 75, 91, 89, 77, 79, 92, and 78). The fourth group contained nine genotypes (39, 68, 32, 88, 67, 33, 74, 29, and 36); 22% were Latino and 77% were Asian (33% East Asian, 33% Southeast Asian, and 11% West Asian). The fifth group was composed of 22 genotypes (64, 48, 46, 2, 82, 81, 93, 90, 41, 69, 54, 8, 27, 15, 60, 51, 47, 22, 85, 52, 59, and 50); 26% were African, 44% were Asian (22% East Asian and 22% Southeast Asian), and 30% were American (18% Brazilian). The sixth group had only four genotypes (34, 17, 43, and 63), all of which were Asian; 50% were East Asian, 25% were West Asian, and 25% were Northern Eurasian. The seventh group contained 16 genotypes; 25% were African, 25% were American, and 50% were Asian (37.5% East Asian, and the remainder equally from Southeast and Southern Asia). The final group consisted of 11 genotypes (45, 30, 16, 9, 24, 37, 19, 10, 61, 4, and 3); 36% were from the USA, 9% were African, and 54% were Asian (27% East Asian, 18% Southern Asian, and 9% Southeast Asian).

Overall, there was a moderate association between the genotypes and their geographical distributions. Perry and McIntosh (1991) reported an association between New World accessions, including Brazilian PIs, and Chinese accessions, with striking morphological similarities between the two groups. This association could be seen between groups 2, 5, 6, and 8. Griffin and Palmer (1995) stated that the long history of soybean domestication and trade in Asia has contributed to the spread of its alleles across regions, thereby reducing the influence of geography on patterns of variation among Asian soybean accessions. Similarly, Brown-Guedira et al. (2000) did not detect any geographical variation in a genetic diversity study using random amplification of polymorphic DNA and simple sequence repeat (SSR) markers, conducted with a group of 105 genotypes that consisted of American ancestors and PIs.

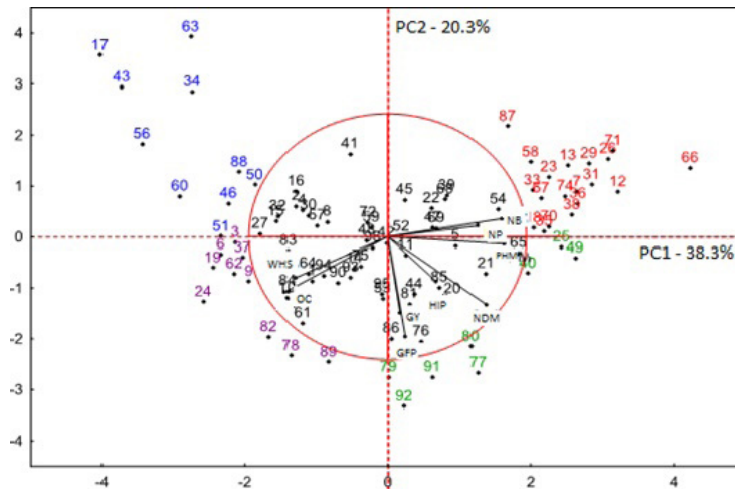
Five of the eight subgroups had Chinese accessions. This was expected, because China is where the soybean originated, and similar results were obtained in a study that included 79 soybean accessions using genomic (SSR) and functional (expressed sequence tag-SSR) microsatellite markers (Mulato et al., 2010). Indeed, 73 of the 79 genotypes were also used in the present study.

Although the sample size was not very large, we did find some associations or groupings based on the traits evaluated. Li and Nelson (2001) reported that the number of

accessions of each region was not representative of the diversity found in each country, and the data allowed the identification of genetic patterns. Another method of identifying genotypes individually is molecular characterization (Oliveira et al., 2010).

In the PCA, the first three components accounted for 71.07% of the total variance. According to Kaiser (1958), only eigenvalues greater than 1.0 should be considered, but those with values above 0.6 in each major component were considered relevant. The first principal component (PC1) accounted for 38.28% of the total variance explained by PHM, NB, OC, NDM, WHS, and NP. The second principal component (PC2) accounted for 20.30% of the total variance explained by GFP and GY, and the third principal component (PC3) accounted for 12.50% of the total variance explained by HIP. However, PC3 did not discriminate between genotypes, which supports the results obtained by Muniz et al. (2002), who reported no significant phenotypic correlation between GY and HIP. Alcantara Neto et al. (2011) investigated correlations between PHM, HIP, NP, and WHS and GY, and found that HIP did not have a cause-and-effect relationship with the other variables, and, therefore, did not directly affect productivity.

Considering the first two principal components, PC1 (38.28%) and PC2 (20.30%), the data were analyzed on a two-dimensional plane, in which the accessions were broken down by quadrants (Figure 2). As can be seen in Figure 2, the genotypes 87, 58, 23, 13, 29, 26, 71, 33, 67, 74, 7, 31, 12, 36, 38, 70, 35, 18, and 66 are highlighted in relation to the variables NB and NP, whose vectors are contained in the first quadrant. The genotypes 51, 46, 50, 88, 60, 56, 34, 43, 17, and 63 are in the second quadrant despite diverging from the others, and had no outstanding variable that grouped them. The third quadrant was characterized by the variables OC and WHS, and contained the genotypes 89, 78, 82, 24, 9, 62, 19, 37, 6, and 3. The fourth quadrant contained the genotypes 25, 49, 40, 80, 77, 91, 92, and 79 with respect to PHM, NDM, GY, and GFP.



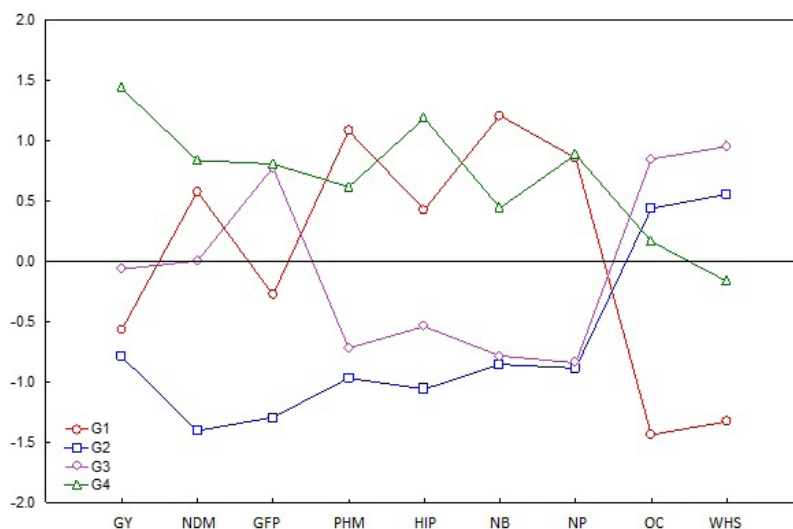
**Figure 2.** Principal component analysis of 93 soybean accessions for agro-morphological traits. The first quadrant shows accessions highlighted in red, the second in blue, the third in purple, and the fourth in green.

Groups were formed based on the genotypes being broken down into each quadrant of Figure 2. The means of the genotypes, and of the groups formed in the quadrants, are shown in Table 2, and the group centroid profiles for each variable are shown in Figure 3.

**Table 2.** Means of agro-morphological traits and four genotype groups that were broken down by principal component analysis.

Group 1	GY	NDM	GFP	PHM	HIP	NB	NP	OC	WHS
18 Guatemala	2004	139	45	151	13	6	51	17	10
35 Nigeria	2885	139	45	128	10	6	72	15	9
70 Philippines	3039	135	41	118	15	6	120	17	10
38 Japan	2997	136	42	116	10	7	169	16	11
36 Peru	1792	139	40	124	19	6	102	15	9
74 Malaysia	1098	137	37	134	27	6	55	15	9
7 Indonesia	1846	134	43	129	16	7	94	15	8
31 Argentina	1801	137	43	129	16	7	94	15	9
12 Guatemala	1700	140	45	159	10	7	109	14	9
33 Thailand	1188	138	38	128	20	6	71	16	9
67 Argentina	1468	138	39	97	19	7	98	17	9
23 Suriname	1955	133	37	133	17	6	78	15	9
58 Japan	2310	132	36	161	7	6	84	15	10
87 Japan	1334	126	34	184	10	4	73	16	8
13 India	1518	138	38	151	12	6	74	14	8
29 Israel	1599	135	35	137	18	6	133	15	10
26 Taiwan	1912	135	37	144	11	7	147	16	8
71 Malaysia	1397	137	37	122	12	7	180	15	10
66 Nepal	1757	139	42	113	15	9	187	14	8
Average	1874	136	40	136	14	6	106	15	9
Group 2									
51 Japan	2844	119	36	58	10	4	59	22	20
46 Angola	1747	114	37	78	14	3	51	22	15
50 Liberia	1751	120	31	85	13	3	34	21	17
88 Japan	750	134	39	33	5	1	36	19	12
60 USA	2414	112	35	61	9	2	50	21	19
56 China	460	112	37	38	5	3	43	20	23
34 China	1958	101	26	62	10	2	27	18	13
43 Russia	1378	100	25	46	4	3	81	21	19
17 Turkey	2420	88	13	35	12	3	47	19	22
63 North Korea	1100	97	22	45	5	3	100	19	17
Average	1682	110	30	54	9	2	53	20	18
Group 3									
3 North Korea	1732	119	44	69	12	2	71	22	18
6 Zimbabwe	2193	122	43	60	10	3	82	20	27
9 El Salvador	1445	128	51	71	12	2	44	21	20
19 Burma Myanmar	1865	123	48	57	8	3	57	23	19
37 Colombia	2001	127	50	39	5	4	50	20	19
62 China	1605	123	48	70	17	2	28	19	23
24 Pakistan	1012	134	57	46	7	2	30	22	20
82 USA	3435	138	51	54	6	3	69	23	17
78 Brazil	4260	130	49	80	16	2	43	22	16
89 Brazil	3515	138	51	95	15	2	67	22	18
Average	2307	128	49	64	11	3	54	21	20
Group 4									
79 Brazil	4178	138	48	106	22	3	57	20	18
91 Brazil	3592	140	53	133	20	3	89	21	16
92 Brazil	4092	139	52	113	26	3	64	22	15
77 Brazil	3983	142	52	133	19	5	89	19	17
80 Brazil	4190	137	49	107	17	7	109	21	18
40 China	2333	141	51	123	10	4	168	19	11
49 Sudan	3822	138	45	123	12	6	140	16	9
25 Thailand	2586	139	47	99	12	8	143	18	10
Average	3597	140	50	117	17	5	106	19	14

GY = grain yield (kg/ha); NDM = number of days to maturity; GFP = grain-filling period (days); PHM = plant height at maturity (cm); HIP = height of insertion of the first pod (cm); NB = number of branches; NP = number of pods; OC = oil content (%); WHS = weight of 100 seeds (g).



**Figure 3.** Centroid profiles of each group (G) broken down by principal component analysis for agro-morphological traits in 93 soybean genotypes. GY = grain yield; NDM = number of days to maturity; GFP = grain-filling period; PHM = plant height at maturity; HIP = height of insertion of the first pod; NB = number of branches; NP = number of pods; OC = oil content; WHS = weight of 100 seeds.

Group 1 (Table 2 and Figure 3) had the second-highest value for NDM (136), indicating that plants of this group have a late cycle. This group also had above-average values for the related traits PHM (136 cm) and HIP (14 cm). However, the GFP (40 days) was below average. This short GFP was associated with the lowest WHS among all of the groups (9 g), which negatively affected the grain yield (1874 kg/ha), even though NB and NP were above the average (6 and 105, respectively). The OC was the lowest among all of the groups, with 15%.

Most of the variables in Group 2 (Table 2 and Figure 3) were below average, with the exception of OC (20%) and WHS (18 g). Despite having the second-largest WHS, the GY was the lowest among all of the groups, with an average of 1682 kg/ha. The accessions in this group were the earliest, with an average NDM of 110 days. In general, the plants were shorter (PHM = 54 cm), which is commonly observed in early plants, in addition to having few pods (NP = 53). The specific genotypes that formed Group 2 were not characterized by any outstanding variable.

Group 3 (Table 2 and Figure 3) had a mean NDM value of 128 days, and 50 days were used for grain filling, which is why the grains were bigger, which, in turn, contributed to the higher WHS (20 g). The NB and NP values were below average (3 and 54, respectively), and negatively influenced the GY (2307 kg/ha), which was below average (Table 2). PHM had a relatively low value (64 cm), and the OC was 20%.

Group 4 (Table 2 and Figure 3) had above-average values for all of the variables, except WHS. These were the latest accessions, with a NDM of 140 days and a long GFP (50 days). The plants were tall (PHM = 117 cm) and had a high HIP (17 cm), which contributed to the low NP value (106). They were the most productive accessions, with an average of 3597 kg/ha; this was expected, because five of the eight genotypes in this group were Brazilian.

Rigon et al. (2012) found a positive, linear relationship between WHS and GY, indicating that indirect selection for this characteristic can increase productivity. Among the



genotypes broken down by PCA, the highest values for WHS were obtained in Groups 2 and 3, the GY values of which were below average. Group 1 had the highest value for GY and a below-average WHS value. These results suggest that other characteristics affect GY. Indeed, Alcantara Neto et al. (2011) found that NP affects productivity; in the present study, the NP values in Groups 2 and 3 were low, which may have contributed to the inconsistency between WHS and GY. In contrast, the opposite was observed in Group 1, which had the highest NP and GY values, demonstrating the close relationship that exists between NP and GY.

According to Muniz et al. (2002), there is a strong correlation between productivity and PHM, indicating that tall plants are more productive. Selection for NP can increase the GY. A congruent result was seen in Group 4, which had above-average values for these variables. The converse was the case for Groups 2 and 3, which had negative standardized values for these variables. However, Group 1 had positive standardized values for PHM and NP and negative standardized values for GY. The low productivity of this group can be explained by the WHS, which was below average; seed size can influence the final GY (Pádua et al., 2010). Silva et al. (2016) reported a negative phenotypic correlation between GY and OC, i.e., high productivity was associated with low OC. The opposite was observed in Groups 2 and 3, which had below-average GY values and above-average OC values.

## CONCLUSIONS

Our analyses revealed the presence of several groups, indicating genetic variability in the soybean accessions studied. However, the geographical origins of the accessions were not always related to the groups in which they were in. The multivariate analyses characterized the genotypes between and within groups, which can be used in Brazilian soybean breeding programs. GY was high in a group that contained five Brazilian genotypes and three PIs (China, Sudan, and Thailand), indicating that these genotypes should be studied further.

## Conflicts of interest

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

The authors wish to thank CAPES for a fellowship, which was essential for the conclusion of the research, granted to M.B. Marconato.

## REFERENCES

- Alcantara Neto F, Gravina GA, Monteiro MMS, Morais FB, et al. (2011). Análise de trilha do rendimento de grãos de soja na microrregião do Alto Médio Gurguéia. *Commun. Sci.* 2: 107-112.
- Brown-Guedira GL, Thompson JA, Nelson RL and Warburton ML (2000). Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. *Crop Sci.* 40: 815-823. <http://dx.doi.org/10.2135/cropsci2000.403815x>
- Bueno LCS, Mendes ANG and Carvalho SP (2006). Natureza e objetivo no melhoramento de plantas. *Melhor. Genet. Plantas- Princ. Proced.* 2: 1724.
- Cruz CD and Carneiro PCS (2003). Modelos biométricos aplicados ao melhoramento genético 2. Editora UFV, Viçosa, 356-434.
- EMBRAPA (2012). Empresa Brasileira de Pesquisa Agropecuária. Tecnologias de Produção de Soja- Região Central do Brasil. Sistemas de Produção. <https://www.embrapa.br/soja/cultivos/soja1> Accessed June 2, 2014.

- Fehr WR and Caviness CE (1977). Stages of soybean development. Cooperative Extension Service, Agriculture and Home Economics Experiment Station, Iowa State University, Ames.
- Ferraudo AS (2012). Técnicas de Análise Multivariada: uma introdução. StatSoft South América, São Caetano.
- Griffin JD and Palmer RG (1995). Variability of thirteen isozyme loci in the USDA soybean germplasm collections. *Crop Sci.* 35: 897-904. <http://dx.doi.org/10.2135/cropsci1995.0011183X003500030045x>
- Hair JF, Anderson RE, Tatham RL and Black WC (2005). Análise multivariada de dados 5th edn. Bookman, Porto Alegre.
- Kaiser HF (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23: 187-200. <http://dx.doi.org/10.1007/BF02289233>
- Kisha TJ and Diers BW (1997). Allele contribution of parents to selected progeny from two-way crosses. *Soybean Genet. Newsletter* 24: 190-193.
- Köppen W (1948). Climatologia: con un estudio de los climas de la tierra. Fondo de Cultura Económica, Mexico, 478.
- Li Z and Nelson RL (2001). Genetic diversity among soybean accessions from three countries measured by RAPDs. *Crop Sci.* 41: 1337-1347. <http://dx.doi.org/10.2135/cropsci2001.4141337x>
- Mahalanobis PO (1936). On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India* 2: 49-55.
- Manjarrez-Sandoval P, Carter TE, Jr., Webb DM and Burton JW (1997). RFLP genetic similarity estimates and coefficient of parentage as genetic variance predictors for soybean yield. *Crop Sci.* 37: 698-703. <http://dx.doi.org/10.2135/cropsci1997.0011183X003700030002x>
- Mulato BM, Möller M, Zucchi MI, Quecini V, et al. (2010). Genetic diversity in soybean germplasm identified by SSR and EST-SSR markers. *Pesqui. Agropecu. Bras.* 45: 276-283.
- Muniz FRS, Mauro AO, Unêda-Trevisoli SH, Oliveira JA, et al. (2002). Parâmetros genéticos e fenotípicos em populações segregantes de soja. *Rev. Bras. Oleag. Fib.* 6: 609-616.
- Oliveira MB, Vieira ESN and Schuster I (2010). Construction of a molecular database for soybean cultivar identification in Brazil. *Genet. Mol. Res.* 9: 705-720. <http://dx.doi.org/10.4238/vol9-2gmr706>
- Pádua GP, Zito RK, Arantes NE and França-Neto JB (2010). Influência do tamanho da semente na qualidade fisiológica e na produtividade da cultura da soja. *Rev. Bras. Sem.* 32: 9-16. <http://dx.doi.org/10.1590/S0101-31222010000300001>
- Perry MC and McIntosh MS (1991). Plant genetic resources. *Crop Sci.* 31: 1350-1355. <http://dx.doi.org/10.2135/cropsci1991.0011183X003100050054x>
- Rigon JPG, Capuani S, Brito Neto JF, Rosa GM, et al. (2012). Dissimilaridade genética e análise de trilha de cultivares de soja avaliada por meio de descritores quantitativos. *Rev. Ceres* 59: 233-240. <http://dx.doi.org/10.1590/S0034-737X2012000200012>
- Silva KB, Bruzi AT, Zuffo AM, Zambiazzi EV, et al. (2016). Adaptability and phenotypic stability of soybean cultivars for grain yield and oil content. *Genet. Mol. Res.* 15: <http://dx.doi.org/10.4238/gmr.15026756>. PubMed
- Sneller RC, Miles J and Hoyt JM (1997). Agronomic performance of soybean plant introduction and their genetic similarity to elite lines. *Crop Sci.* 37: 1595-1600. <http://dx.doi.org/10.2135/cropsci1997.0011183X003700050032x>
- Wysmierski PT and Vello NA (2013). The genetic base of Brazilian soybean cultivars: evolution over time and breeding implications. *Genet. Mol. Biol.* 36: 547-555. <http://dx.doi.org/10.1590/S1415-47572013005000041>