Multivariate selection of nutritional aspects of soybean in an F_5 segregating family

S.M. Fachi[1], I.R. Carvalho[2], J.A.G. da Silva[2], C.D. Ferreira[3],
M.H. Barbosa[3], D.A. Magano[2], F.S. Sanes[2], L.L. Ferreira[5],
V.J. Szareski[3], N.B. Moura[2], F. Lautenchleger[4] and V.Q. de Souza[6]

[1] Universidade Federal de Santa Maria, Santa Maria, RS, Brasil
[2] Universidade Regional Noroeste do Estado do Rio Grande do Sul, Ijuí, RS, Brasil
[3] Universidade Federal de Pelotas, Capão do Leão, RS, Brasil
[4] Universidade Estadual do Centro-oeste, Guarapuava, PR, Brasil
[5] Centro Universitário de Mineiros, Mineiros, GO, Brasil
[6] Universidade Federal do Pampa (PQ 312329/2019-9), São Gabriel, RS, Brasil

Corresponding author: I.R. Carvalho
E-mail: irc.carvalho@gmail.com

ABSTRACT. The soybean cultivars currently available for cultivation are used mainly due to their productive and phytosanitary characteristics, without defining specific centesimal compositions as an ideotype. Our study aimed to contribute to this aspect and to identify profiles and establish similarities between the genotypes, taking into account components such as starch, lipids and mineral materials in 54 soybean genotypes. The experiment was conducted in the 2017/2018 agricultural year in Campos Borges, RS. The treatments consisted of four cultivars, considered pure lineages, and 50 F5 segregating families (theoretical level of inbreeding of 94% with approximately 6% heterozygosis). The experimental units were composed of a sowing line five meters in length, spaced 0.45 m. The characters of agronomic interest measured in the grains collected from 10 random plants were: protein, humidity, lipids, total fibers, mineral material and starch. The percentages of the characters of agronomic interest: proteins, lipids, fibers, mineral matter, starch and

oil were determined in intact soybeans by the technique of reflectance in the near infrared. There was genetic dissimilarity among the 54 soybean genotypes. The protein character was the most decisive factor for genetic variability, based on Singh's relative contribution. Using patterns such as standardized Euclidean mean distance, Tocher's optimized method and the UPGMA clustering method, it was possible to show patterns in the genotype clusters. Based on these results, we can define crosses between accessions with genetic potential for the formation of genotypes with superior centesimal characteristics.

Key words: Agricultural genetics; Competition test; *Glycine max*; Seed composition

INTRODUCTION

The soybean (*Glycine max*) is a species of annual cycle belonging to the Fabaceae family, originating in East Asia, a region between China and Manchuria (Panwar et al., 2018). It is the most widely produced legume in the world; in the 2017/2018 harvest the world production was 337 million tons (USDA, 2019); in this same period, Brazilian production was 118 million tons, which ranked Brazil as the world's second largest soybean producer, only after the United States (CONAB, 2018).

Versatility is the main factor that makes soybean one of the most cultivated crops in the world. The grains are used in human and animal food, composing the food diets of most countries. The soybean grain has, in general, 20% oil and 40% proteins, which constitute about 60% of the dry mass (Bezerra, 2015). The rest of the dry matter of the seed is composed mainly of carbohydrates, about 35%, and 5% of mineral material (Liu, 2012).

The main use of soybeans is in the crushing industry, which produces crude oil, degummed and bran (Finoto, 2008). Of all soybeans produced in the Brazilian territory, 42.1% is destined for national industry, of these, 8.33% is processed into oil and 31.9% in bran (Abiove, 2016). Soybean oil is used as a base for food products (refined oil, hydrogenated fats, margarines, mayonnaises, etc.) and bran is one of the main raw materials for the manufacture of ration.

The composition of soybean grains varies according to several factors, among which the genotype, cropping site, environmental conditions and sowing season stand out. Alves (2011) evaluating the centesimal composition of soybean grains from eight different cultivars, observed that the contents of these components varied among the cultivars. In the same study, the cultivar BRS 258 had a higher protein content (44.37%), being possible the production of Hipro soybean bran (48% protein) without the need to use a system of separation of soybean hulls, before processing.

In a study that analyzed the oil content and the genetic diversity of 22 soybean genotypes for the choice of progenitors aiming to increase the oil content in soybean, even with reasonable genetic diversity it was possible to verify a wide variation in the oil content in the grains (Rodrigues, 2016). In another study, Pípolo (2002), evaluating five genotypes in two agricultural years (1998/1999 and 1999/2000) and 10 different locations, verified differences between the genotypes at the mean concentrations of protein and oil.

Studies carried out by Hwang (2014), detecting regions of the genome and controlling the quantitative characteristics, managed to analyze where they were located, so the chromosomal regions can be used to identify genes or DNA markers that are used in the selection to alter the protein and the oil in the soybean seed in a predictable way.

The current cultivars available for cultivation are selected mainly due to their productive and phytosanitary characteristics, not defining specific centesimal compositions as an ideotype. The increasing consumption and utilization of the centesimal components of soybean (especially lipids and protein) has been problematic for the industry since the concentrations that are needed, for example, the protein concentration in the bran, are not always reached (Rangel, 2007).

Likewise, there is a growing market for the development of new soybean products that require high values of certain centesimal components. Thus, the selection of soybean genotypes for their nutritional constituents (protein, lipids, mineral materials and others) allows the grain to be destined for specific uses with greater efficiency, which adds qualitative value and makes its participation in the market more effective. The present study was designed to contribute to this aspect and to identify profiles and determine differences among soybean genotypes, taking into account constituents such as starch, lipids and minerals.

MATERIAL AND METHODS

The experiment was conducted in the agricultural crop of 2017/2018 in Campos Borges - RS, located at coordinates 28°52'31' S and 53°00'55" W, with humid subtropical climate of the Cfa type according to the classification of Köppen (Kottek et al., 2006) and soil classified as dark red Latosol (Oxisol) (Streck et al., 2008). The treatments consisted of four cultivars, considered as pure lineages and 50 F5 segregating families (theoretical endogamy level of 94% with approximately 6% heterozygous) (Table 1), obtained through directed crossings.

Table 1. Genotypes that compose the F5 segregating soybean generation and the pure lineages analyzed in this study.

Segregating generation	IRC1_194; IRC1_196; IRC1_136; IRC2_111; IRC2_191; IRC3_120; IRC3_159; IRC7_29; IRC8_130; IRC11_135; IRC12_155; IRC12_134; IRC12_132; IRC13_117; IRC13_119; IRC13_164; IRC16_82; IRC16_153; IRC16_97; IRC17_193; IRC17_46; IRC19_62; IRC19_124; IRC19_114; IRC21_160; IRC21_113; IRC22_49; IRC25_154; IRC26_9; IRC28_110; IRC28_123; IRC29_122; IRC30_182; IRC30_104; IRC31_102; IRC31_137; IRC32_103; IRC34_174; IRC32_129; IRC32_189; IRC33_175; IRC33_192; IRC35_118; IRC36_152; IRC36_121; IRC37_116; IRC39_133; IRC39_106; IRC39_105 IRC40_58
Pure lineages	M6210IPRO; NA5909RG; TMG7069IPRO; NS6601IPRO

The experimental units were composed of a sowing line 5 m in length, spaced by 0.45 m. Direct and manual sowing was done in the second half of November 2017, using a density of 14 seeds per linear meter, base fertilization of 250 kg.ha^{-1} of N-P-K using the formulation 10-20-20.

Preventive control of weeds, pest insects and diseases was as recommended in order to minimize the biotic effects in the result of the experiment. The characters of agronomic interest measured in the grains collected in 10 random plants were: protein (PTN); moisture (MO); lipids (LIP); total fibers (FB); mineral material (MM) and starch (ST). The percentages of the characters of agronomic interest: protein, lipids, fiber, mineral matter, starch and oil were determined in intact soybean grains by the technique of Near Infrared Reflectance (NIR).

The data were submitted to the assumptions of the analysis of variance where the normality and homogeneity of the residual variances were verified, after which the descriptive analysis was carried out by frequency distribution and definition of the phenotypic classes of interest. Then the Euclidian algorithm was used to make the genetic distances matrix of the studied genotypes using all the measured characters. Through the distance matrix the Tocher optimized grouping method was applied with intra-group and inter-group stratification, we used the methodology of Singh's (1981) relative contribution in order to define which characters are most important. Statistical analyzes were performed using Genes software (Cruz, 2013) and R (Team RC, 2018).

RESULTS

Descriptive analysis

The descriptive analysis revealed the formation of different quantitative components of the centesimal components for the studied genotypes (Table 2). Segregant families had a higher proportion of lipids (18.52%), compared to pure lines with 17.36% respectively. In contrast, pure lines obtained a higher proportion of protein in comparison to the segregating generation, a fact that can be explained by the inverse relationship between the lipid and protein contents in soybean genotypes (Akond et al., 2012; Popović et al., 2012) In addition to this factor, it is interesting to note that the F_5 segregating generation still presents close to 6% of loci in heterozygosis, and therefore, there may be non-additive effects on the contribution of centesimal component expression.

Table 2. Descriptive analysis of the nutritional components of stratified genotypes for pure lines and F5 segregating families of soybean, conducted in the municipality of Campos Borges in the agricultural crop year 2017/2018.

Genotypes	PTN*	MO	LIP	FB	MM	ST
			------------------------------(%)------------------------------			
Pure lines	36.42	10.78	17.36	5.75	4.66	3.23
F_5 segregating families	35.26	10.60	18.52	5.55	4.77	3.31

* PTN, protein; MO, moisture; LIP, lipids; FB, fiber; MM, mineral material; ST, starch.

The PTN character revealed the formation of eight classes with amplitude of 32.4 to 38.0% (Figure 1). The class with 35.6% content was the one that composed the highest frequency of genotypes, around 25% respectively. However, in order to obtain higher concentrations of protein in the grain, two genotype frequencies were better able to meet these conditions, the 37.2% content that composed a frequency of 10% of the genotypes and

the 38% content that composed one frequency of 3% of the genotypes. The selection of genotypes to increase the protein content adds value to the grain and can guarantee the competitiveness of Brazilian soybeans in the world market (Rodrigues et al., 2010).

The ST character presented eight classes with amplitude ranging from 1.25% to 4.75%. The highest frequency of genotypes (32%) was allocated to the class 3.75% of starch, a value commonly found in pure lines. About 13 and 2% of the genotypes were allocated to the classes 4.25 and 4.75% of starch, respectively, allowing selection to increase this component in the grains. Increase in the starch content can improve the taste and texture of soybeans that are consumed unprocessed (Masuda, 2004; Oliveira et al., 2010). The physiological processes of germination and vigor are influenced by protein, lipid, starch and sugar contents. In this way, the increase in the starch content can positively influence the germination process and seedling vigor by providing soluble carbohydrates available and absorbable by the embryo (Moraes et al., 2006).

The FB character revealed nine classes that presented amplitude of 5.22 to 6.18%. The highest frequency of the studied genotypes (25%) presented 5.58% of FB, values similar to those found in the pure lines available in the market. Soybean fiber is used in nutrition products and in some bakery products (Slavin, 1991). The increase in fiber content in soybeans is important because it allows it to be classified as functional foods, ensuring that its use within a diet can reduce the risk of some diseases, such as the heart and some types of cancer (Poutanen et al., 2014)

The highest frequency of the genotypes shows 10.5% moisture. Nine classes were formed for this character, and the amplitude ranged from 9.7 to 11.3%. Soybeans are usually harvested at high moisture levels, from 16 to 25%, which makes them unsuitable for storage because they make the grains more susceptible to fungal infections or other microorganisms, reducing their quality (Coradi et al., 2016). Damage caused by seed moisture occurs because there is an unequal maturation in the plant population (Hamer, 1999). For storage and sale, the moisture content of soybeans should be around 12% guaranteeing the quality of storage (Coradi et al., 2016). Thus, the frequency of genotypes that presented 9.7% moisture has important characteristics regarding the conservation of the grains.

The MM character revealed eight classes with a magnitude of 4.05 to 5.10%. The highest frequency was obtained in the 4.8% MM content group, similar to those found in the pure lines of this study. As a large part of the soybean grains are used in the production of feed for the animal production chain, it is interesting that the total MM content is not high, because when MM is higher, the energy levels of the feed produced with these grains will be higher.

The LIP character formed eight classes with magnitude from 14.5% to 21.5%. The content of 18.5% was the one that composed the highest frequency of genotypes, close to 35%. The value of 18.5% is the average value of oil content in commercial cultivars. Two frequencies of genotypes around 6% and 4%, were promising for selection for increases in lipid content, with a content of 20.5% and 21.5%, respectively. Cavalcante et al. (2010), selecting soybean lineages for the oil content for biodiesel production, found average oil content in the grains of 16.75% and the highest value in only one lineage with 21.59%. Therefore, the genotypes present in two classes of the frequency distribution obtained in this study are shown to be promising for selection.
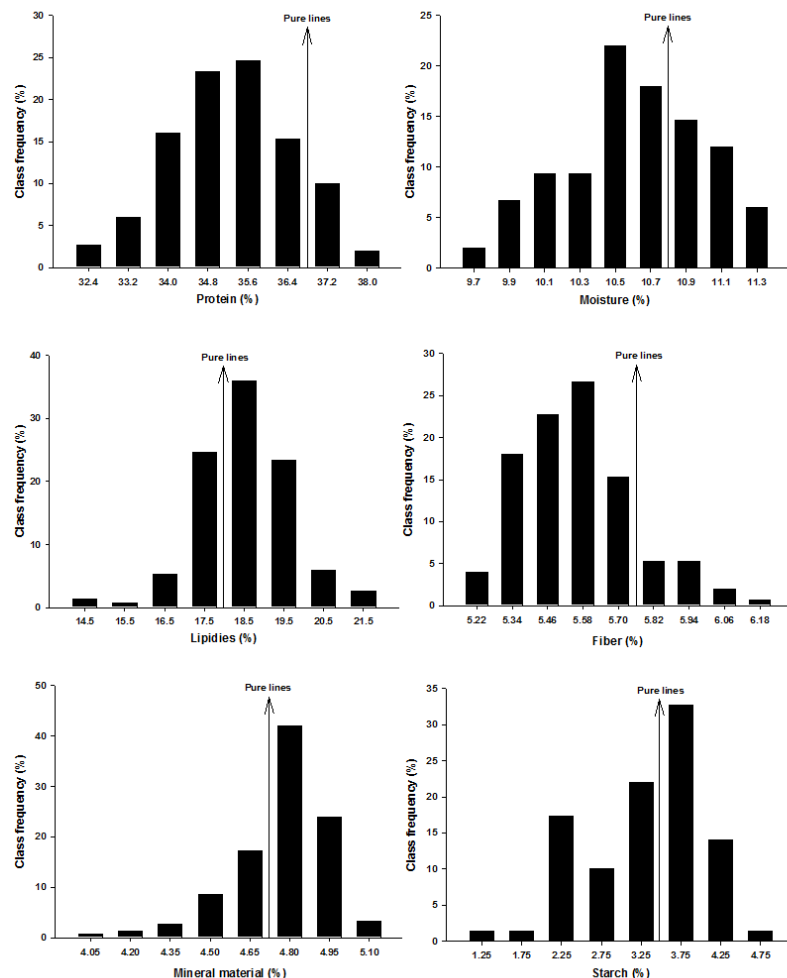
Figure 1. Class frequency distributions for the traits: PTN, protein; MO, moisture; LIP, lipids; FB, fiber; MM, mineral material; ST, starch.

The analysis of variance revealed significance at 5% probability for the interaction between maize soybean genotypes for characters (PTN) protein; (ST) starch; (FB) fiber; (MO) moisture; (MM) mineral matter and (LIP) lipids (Table 3).

Regarding the coefficients of variation obtained for the characters evaluated, it was observed a range of 1.05% to 6.51% (Table 3). According to the classification proposed by Pimentel Gomes (2000), the coefficients of variation (CV) are considered low (CV <10%) with high precision, medium (10 to 20%) being of good precision, high (20 to 30 %) that reveal low accuracy and very high (CV > 30%) reflecting low accuracy. Thus, the coefficients of variation obtained for the characters under analysis are legitimate and of high precision.

The F5 segregating generation composed of the genotypes IRC1_194 IRC1_196 IRC1_136; IRC2_111; IRC2_191; IRC3_120; IRC3_159; IRC7_29; IRC8_130; IRC11_135; IRC12_155; IRC12_134; IRC12_132; IRC13_117; IRC13_119; IRC13_164; IRC16_82;

IRC16_153; IRC16_97; IRC17_193; IRC17_46; IRC19_62; IRC19_124; IRC19_114; IRC21_160; IRC21_113; IRC22_49; IRC25_154; IRC26_9; IRC28_110; IRC28_123; IRC29_122; IRC30_182; IRC30_104; IRC31_102; IRC31_137; IRC32_103; IRC34_174; IRC32_129; IRC32_189; IRC33_175; IRC33_192; IRC35_118; IRC36_152; IRC36_121; IRC37_116; IRC39_133; IRC39_106; IRC39_105; IRC40_58 and pure lines M6210IPRO; NA5909RG; TMG7069IPRO and NS6601IPRO presented significant differences regarding the contents of the centesimal components (Table 4). The IRC17_193 genotype was the one with the highest PTN (37.63%), followed by the IRC25_154 (37.43%), IRC1_196 (37.25%), IRC19_62 (37.24%) and IRC8_130 (37.21%) genotypes, showing themselves promising for the increase of protein in the grain, with emphasis on the IRC17_193 genotype. For the chemical quality characteristics of the grains, the industries have internal standards regarding the content of oil and protein in the soybean grain, and these must meet the minimum for the production, as is the case of soybean meal with high protein content (Hipro $\geq$ 48% protein). To achieve such protein requirements in the bran, the grain must have, on average, 37% protein. The constitutional values described by the families are within the desired according to industry standardization

Table 3. Summary of analysis of variance for 54 soybean genotypes analyzed in Campos Borges - RS in the 2017/2018 crop season.

Variation factor		Mean square					
	GL	PTN	MO	LIP	FB	MM	ST
Genotypes	53	4.145*	0.393*	3.875*	0.109*	0.0813*	1.415*
Blocks	2	0.578	0.034	0.06	0.007	0.0461	0.0179
Residue	161	0.139	0.028	0.16	0.009	0.0075	0.046
CV (%)		1.057	1.578	2.175	1.709	1.828	6.512

[ns] For non-significant values; * For significant values at 0.05 by the *F* test. PTN, protein; MO, moisture; LIP, lipids; FB, fiber; MM, mineral material; ST, starch.

On the other hand, the genotypes that obtained the highest protein content were those that reached the lowest percentage of lipids. In order to assert the correlation between these characters, which corroborates the results of Helms and Orf (1999), which describe the direct selection for the protein content to promote the reduction in oil content.

The cultivar IRC28_110 was the most promising for selection of high lipid contents in the grains with 21.17%. The amplitude obtained by the genotypes was from 14.99% to 21.17% with an average of 18.44%. The results of this study are superior to those obtained by Finoto et al. (2008) that, by evaluating the phenotypic variation of oil and protein content in the grains of 15 soybean genotypes, obtained a mean oil content of 16.01%, ranging from 13.01% to 20.12%. The conducting of these families to compose possible cultivars becomes pertinent, so that superiority to what is described as the required average has been found.

The dendogram shows the genetic dissimilarity between the accessions (Figure 2). The analysis of the Euclidean Mean Distance provided the formation of eight groups. By means of these results, it is possible to identify that the grouping based on the arithmetic mean (UPGMA) shows results similar to those obtained by the Tocher optimization method. Through the formation of these groups with similar genetic behaviors, it is possible to group agronomic interest characteristics that show uniqueness as to the portion of the character variation that depends on the genetic factor.

Table 4. Centesimal constitution of soybean genotypes tested in Campos Borges during the 2017/2018 crop season.

GENOTYPE	PTN		MO		LIP		FB (%)		MM		ST	
IRC1_194	35.82	d	10.74	b	18.32	e	5.30	e	4.85	b	3.36	d
IRC1_196	37.25	a	10.84	a	17.34	f	5.70	c	4.92	a	2.30	f
IRC1_136	34.66	f	10.98	a	19.32	c	5.44	d	4.81	b	3.41	d
IRC2_111	34.82	e	10.68	b	19.12	c	5.52	d	4.77	b	3.67	c
IRC2_191	35.26	e	10.53	b	17.75	f	5.34	e	4.81	b	3.51	d
IRC3_120	36.15	c	10.31	c	16.76	g	5.63	d	4.81	b	3.68	C
IRC3_159	34.07	f	10.68	b	20.79	a	5.55	d	4.57	d	4.26	a
IRC7_29	34.30	f	9.91	d	18.35	e	5.36	e	4.83	b	3.48	d
IRC8_130	37.21	a	10.38	c	14.99	h	5.90	b	5.00	a	2.29	f
IRC11_135	33.73	g	10.77	b	19.70	c	5.34	e	4.73	c	3.84	c
IRC12_155	35.67	d	10.27	c	18.95	d	5.62	d	4.84	b	3.18	d
IRC12_134	34.01	f	9.93	d	18.77	d	5.31	e	4.90	a	2.39	f
IRC12_132	34.84	e	10.27	c	18.62	d	5.57	d	4.90	a	3.14	d
IRC13_117	35.60	d	10.93	a	18.53	d	5.63	d	4.79	b	3.29	d
IRC13_119	32.97	g	10.74	b	19.64	c	5.51	d	4.52	d	4.09	b
IRC13_164	35.91	d	10.91	a	18.31	e	5.52	d	4.28	e	4.41	a
IRC16_82	35.77	d	10.00	d	18.23	e	5.51	d	4.97	a	2.54	e
IRC16_153	34.58	f	9.93	d	17.41	f	5.38	e	4.79	b	2.67	e
IRC16_97	33.28	g	9.99	d	19.26	c	5.48	d	5.05	a	2.30	f
IRC17_193	37.64	a	11.08	a	17.91	e	5.50	d	4.70	c	2.84	e
IRC17_46	36.82	b	10.99	a	17.72	f	5.68	c	4.66	c	3.90	b
IRC19_62	37.24	a	10.52	b	17.10	g	5.82	c	4.93	a	2.44	f
IRC19_124	35.47	d	10.50	b	17.39	f	5.58	d	4.96	a	2.30	f
IRC19_114	36.70	b	10.59	b	17.01	g	5.60	d	4.83	b	2.97	e
IRC21_160	33.16	g	10.57	b	19.96	b	5.24	e	4.87	b	2.39	f
IRC21_113	34.44	f	10.52	b	19.58	c	5.34	e	4.88	b	2.22	f
IRC22_49	35.50	d	11.05	a	18.62	d	5.53	d	4.66	c	3.68	c
IRC25_154	37.43	a	10.64	b	17.51	f	5.51	d	4.78	b	3.51	d
IRC26_9	34.50	f	10.52	b	19.96	b	5.55	d	4.60	d	4.01	b
IRC28_110	34.21	f	9.83	d	21.17	a	5.37	e	4.91	a	2.14	f
IRC28_123	35.17	e	11.27	a	19.00	d	5.71	c	4.74	c	3.75	c
IRC29_122	35.67	d	10.72	b	17.69	f	6.06	a	4.92	a	2.74	e
IRC30_182	36.07	c	10.56	b	18.87	d	5.72	c	4.85	b	2.82	e
IRC30_104	34.57	f	11.11	a	18.75	d	5.54	d	4.25	e	4.56	a
IRC31_102	35.10	e	10.60	b	18.62	d	5.52	d	4.54	d	4.29	a
IRC31_137	33.12	g	11.26	a	19.19	c	5.57	d	4.76	b	3.47	d
IRC32_103	35.71	d	10.79	b	18.01	e	5.49	d	4.78	b	3.51	d
IRC34_174	34.66	f	10.70	b	19.14	c	5.70	c	4.60	d	3.80	c
IRC32_129	34.65	f	11.09	a	19.02	d	5.50	d	4.73	c	3.69	c
IRC32_189	35.83	d	10.56	b	17.58	f	5.52	d	4.69	c	3.46	d
IRC33_175	36.08	c	10.59	b	18.31	e	5.46	d	4.74	c	3.59	c
IRC33_192	34.88	e	10.51	b	17.96	e	5.36	e	4.71	c	3.63	c
IRC35_118	36.28	c	10.19	c	17.39	f	5.83	c	4.57	d	4.22	a
IRC36_152	34.21	f	11.15	a	18.62	d	5.44	d	4.70	c	3.39	d
IRC36_121	36.70	b	10.05	d	18.26	e	5.45	d	4.97	a	1.44	g
IRC37_116	34.97	e	10.94	a	18.50	d	5.77	c	4.64	d	3.66	c
IRC39_133	34.53	f	10.29	c	19.34	c	5.58	d	4.87	b	3.38	d
IRC39_106	34.09	f	10.63	b	21.04	a	5.33	e	5.06	a	3.78	c
IRC39_105	36.38	c	10.45	b	18.83	d	5.82	c	4.62	d	3.94	b
IRC40_58	35.31	e	10.91	a	17.99	e	6.00	b	4.65	c	4.25	a
M6210IPRO	37.10	a	10.76	b	16.26	g	5.67	c	4.59	d	2.94	e
NA5909RG	35.92	d	10.71	b	18.36	e	5.53	d	4.74	c	3.31	d
TMG7069IPRO	36.23	c	11.05	a	17.84	f	6.15	a	4.70	c	3.04	e
NS6601IPRO	36.42	c	10.58	b	16.98	g	5.64	d	4.60	d	3.62	c
Average	35.35		10.61		18.44		5.57		4.76		3.31	
CV (%)	1.05		1.57		2.17		1.70		1.82		6.51	

Means followed by the same lowercase letter in the column belong to the same group, according to Scott-Knott's grouping criterion (1974), at 5% probability. PTN, protein; MO, moisture; LIP, lipids; FB, fiber; MM, mineral material; ST, starch.
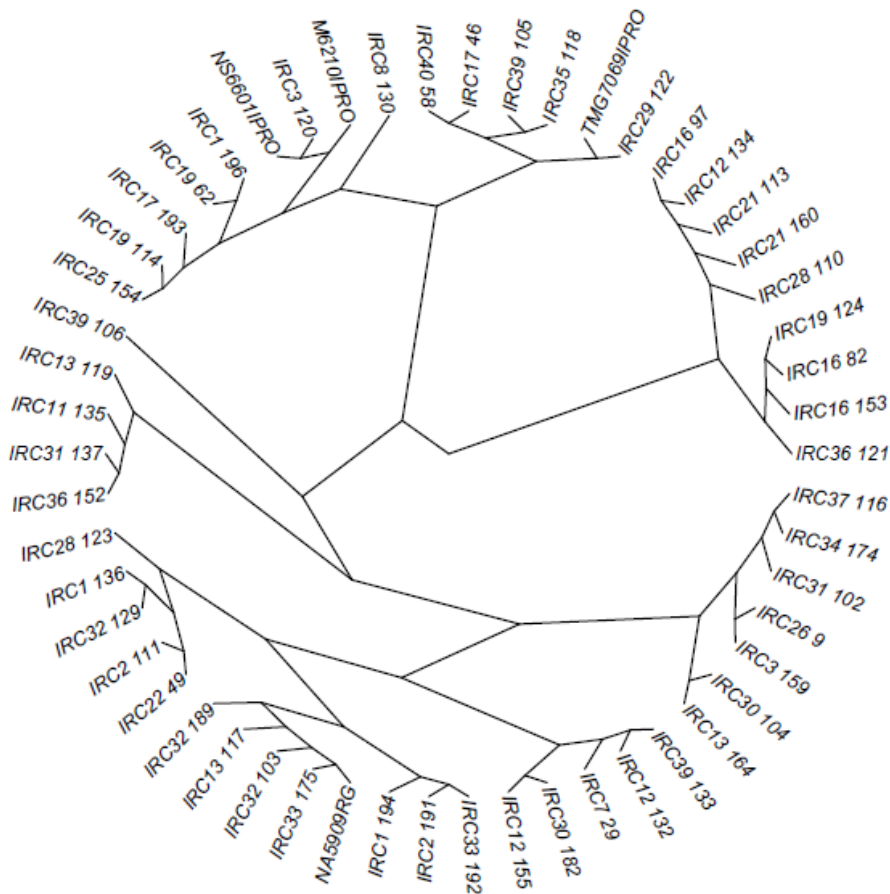
Figure 2. Dendogram based on Euclidean Mean Distance - UPGMA, composing 50 F5 segregating families and four pure lines of soybean. Characters used: PTN, protein; MO, moisture; LIP, lipids; FB, fiber; MM, mineral material; ST, starch.

The analysis to estimate the relative contribution of each character to the expression of the genetic divergence showed that three of the seven analyzed variables contributed with 68.47% of the divergence found (Table 5). Among the studied variables, PTN (29.76%), starch content (22.57%) and fiber (16.14%) were the most efficient in explaining dissimilarity among the 45 soybean genotypes evaluated.

The fact that PTN was the main contributor to genotype differentiation, only emphasized that this is an important characteristic that should be prioritized in breeding programs in order to produce gene complementarity and new recombinants to the grain quality characteristics for the industry (Rodrigues et al., 2017).

The Tocher optimized clustering method allowed to fragment the 45 soybean genotypes evaluated in eight groups, with similar behavior within the group and divergent between groups (Table 6). Of these, 3 groups were constituted by 4 accessions, 1 group by 3 accessions and 1 a group by only two accessions. Group I had the highest number of accessions (19) followed by group II with 9 accessions,

corresponding to 42.22% and 20% of the total, respectively. The accessions belonging to the same heterocyclic group are genetically similar to each other, thus their combinations provide low variability, so it is necessary that the combinations be carried out among genotypes of different groups because they are more promising due to greater dissimilarity. However, what has been evidenced is the extensive use of genotypes that are very closely related to the development of the new cultivars, resulting in the narrowing of the genetic basis of the crop (Bertini et al., 2006; Miranda et al., 2007). Therefore, it is possible to increase the genetic variability of the breeding program using the genotypes of genetically more distant groups. One option is to direct the crossings between the accessions of groups I and VII.

Table 5. Relative contribution of characters to soybean genetic diversity according to the criterion of Singh (1981).

Variable	S.j	Value in %
PTN*	41757.7	29.7672
MO	15586.6	11.111
LIP	19457.4	13.8703
FB	22648.9	16.1454
MM	9169.04	6.5362
ST	31661.4	22.57
Total	140281	

* PTN, protein; MO, moisture; LIP, lipids; FB, fiber; MM, mineral material; ST, starch. PTN, protein; MO, moisture; LIP, lipids; FB, fiber; MM, mineral material; ST, starch.

Table 6. Tocher optimized grouping method for the definition of groups through the dissimilarity of the soybean genotypes.

Groups	Accessions
I	IRC33_175 NA5909RG IRC32_103 IRC13_117 IRC32_189 IRC22_49 IRC2_111 IRC1_194 IRC2_191 IRC33_192 IRC12_132 IRC39_133 IRC37_116 IRC32_129 IRC1_136 IRC28_123 IRC34_174 IRC31_102 IRC26_9
II	IRC12_155 IRC30_182 IRC29_122 IRC16_82 IRC19_114 IRC1_196 IRC19_62 IRC17_193 IRC25_154
III	IRC3_120 NS6601IPRO M6210IPRO IRC17_46
IV	IRC12_134 IRC16_97 IRC21_113 IRC21_160
V	IRC31_137 IRC36_152 IRC11_135 IRC13_119
VI	IRC35_118 IRC39_105 IRC40_58
VII	IRC13_164 IRC30_104
VII	IRC7_29 IRC16_153 IRC19_124

Ten linear correlations were found with a level of 5% by the t-test (Table 7). The coefficients of linear correlation followed the classification proposed by Carvalho et al. (2004). The protein content presented intermediate positive correlation with the fiber content r = (0.38) and negative of high magnitude with lipids (-0.68) and low with starch (r = -0.16). The high negative correlation between protein and lipid content was already expected due to the competition of these two variables by the carbon skeletons (Hanson, 1992; Pípolo, 2002), showing the degree of genetic association among these

characters and explaining the great difficulty in selection to achieve simultaneous gains for oil and protein. Similar results from several studies demonstrate this same correlation (Bonato et al., 2000; Cober and Voldeng, 2000; Feng et al., 2004; Haghi et al., 2012; Qin et al., 2014; Dhungana et al., 2017). In relation to the negative association between protein and starch, this was also observed in studies conducted by Wee et al. (2018).

The moisture character positively correlated with intermediate associations with fiber content (0.26) and starch (0.39) and negatively with intermediary associations with MM (-0.42), probably these associations are related to water affinity of each component. The lipid content obtained a low positive correlation with starch (0.17) and negative mean with the fiber content (-0.42). Studies by Dhungana et al. (2017) and Wee et al. (2018) also found a positive correlation between starch and lipids.

Table 7. Pearson's Linear Correlation for protein, moisture, lipid, total fiber, mineral material and starch traits in soybean genotypes tested in Campos Borges in the 2017/2018 harvest.

	MO	LIP	FB	MM	ST
PTN	0.09^{ns}	-0.68^{*}	0.38^{*}	-0.02^{ns}	-0.16^{*}
MO	.	0.002^{ns}	0.26^{*}	-0.41^{*}	0.39^{*}
LIP		.	-0.42^{*}	-0.05^{ns}	0.17^{*}
FB			.	-0.06^{ns}	0.05^{ns}
MM				.	-0.70^{*}

[ns] For non-significant values; * For significant values at 5% by the t test. PTN, protein; MO, moisture; LIP, lipids; FB, fiber; MM, mineral material; ST, starch.

The analysis of main components allows the identification of the variables of greater and lesser contribution to accumulated variation (Cruz et al., 2004). Two main components were found, the first component explains 36.53% of the variation and the second component 35.53%, together the two components explain 72.06% of the total variance (Figure 3). Through the relative contribution of Singh (Table 3), we can infer the characters that most contribute to differentiate the genotypes, being: PTN, MO and LIP. These characters contribute significantly to the variability of the populations and therefore, they would be the variables more responsive to the selection for the genotypes.

Results obtained by Mello Filho et al. (2004) analyzing the means of eight populations correlations of protein with yield, seedling emergence and accelerated aging, had positive, but low and not significant correlations with yield, thus the higher protein content tends to have lower productivity.

In relation to the behavior of the genotypes within the graph, it is observed that these are well distributed, being possible to observe the formation of some groups, such as the one that compose the genotype NS6601IPRO and another group that compose the genotype NA5909RG. Some genotypes are more distant, e.g., 8_130 and 36_121. These genotypes have the highest values for the two main components, so they would be more likely to be selected for these characters.
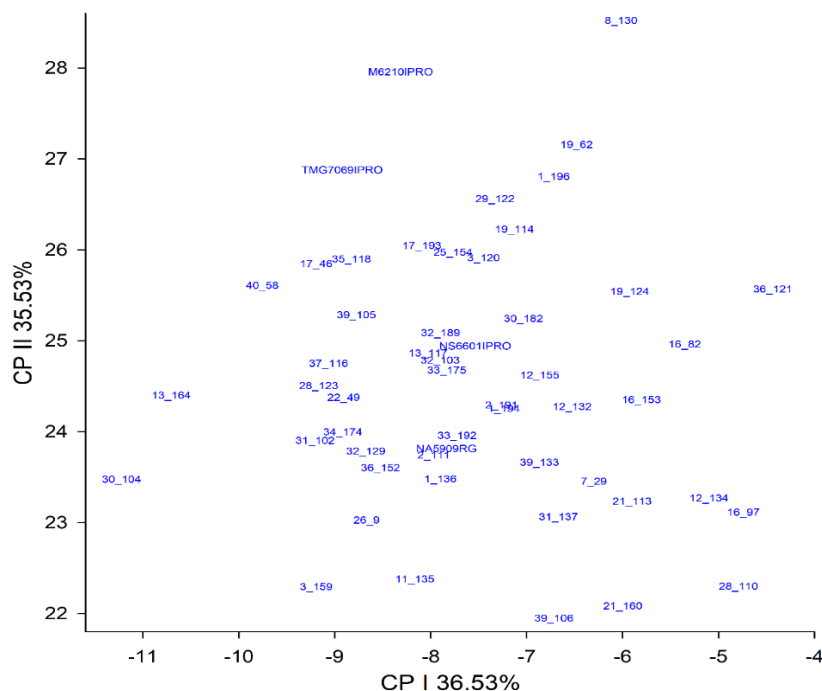
Figure 3 - Analysis of the main components referring to the heritabilities of the six characters of interest for the soybean crop.

CONCLUSIONS

We found genetic dissimilarity among the 54 soybean genotypes studied. The protein character was the most decisive factor for the genetic variability of this study, according to the results of Singh's (1981) relative contribution.

Through the use of standards such as the standardized Euclidean mean distance, Tocher optimized method and UPGMA grouping method, it was possible to show patterns in relation to genotype groupings.

With the results obtained it is possible to plan crosses between accessions with genetic potential for the formation of genotypes with superior characteristics.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Abiove. Estatística. 2016. Available at: http://www.abiove.org.br/site/index.php?page=estatistica&area=NC0yLTE.

Akond AM, Ragin B, Bazzelle R, Kantar S, et al. (2012). Quantitative trait loci associated with moisture, protein, and oil content in soybean [*Glycine max* (L.) Merr.]. *J. Agric. Sci.* 4(11): 16. http://dx.doi.org/ doi:10.5539/jas.v4n11p16.

Alves FP, Oliveira MA, Mandarino JMG, Benassi VT, et al. (2011). Composição centesimal de grãos de soja de oito diferentes cultivares. In: *Embrapa Soja-Resumo em anais de congresso (ALICE)*. In: Jornada Acadêmica Da Embrapa Soja, Londrina. Resumos expandidos. Londrina: Embrapa Soja, 7-9. (Embrapa Soja. Documentos 328).

Bertini CHC, Schuster I, Sediyma T, Barros EG, et al. (2006). Characterization and genetic diversity analysis of cotton cultivars using microsatellites. *Genet. Mol. Biol.* 29(2): 321-329. https://doi.org/10.1590/S1415-47572006000200021.

Bezerra ARG, Sedyma T, Borém A, Soares MM, et al. (2015). Importância Econômica. In: Sediyama, Tuneo; Silva, Felipe; Borém, Aluízio. *Soja: do plantio à colheita*. Viçosa: Editora UFV, 1(13).

Bonato ER, Bertangoli PF, Lang CE and Rubin SAL (2000). Teor de óleo e de proteína em genótipos de soja desenvolvidos após 1990. *Pesq. Agropec. Bras.* 35(12): 2391-2398. https://doi.org/10.1590/S0100-204X2000001200009.

Carvalho FIF, Lorencerti C and Benin G (2004). Estimates and Implications of Correlation.Pelotas. Ed. Universitária da UFPEL, 142p.

Cavalcante AK, Espindola SG, Hamawaki OT, Bisinotto FF, et al. (2010). Avaliação e seleção de linhagens de soja quanto ao teor de óleo para a produção de biodiesel. *FAZU Rev.* (6): 11-52. https://www.fazu.br/ojs/index.php/fazuemrevista/article/view/11/5.

Cober Erd and Voldeng H (2000). Developing high-protein, high-yield soybean populations and lines. *Crop Sci.* 40(1): 39-42. https://doi:10.2135/cropsci2000.40139x.

CONAB (2018). Companhia Nacional de Abastecimento. *Acompanhamento da Safra Brasileira de Grãos,* Brasília, 5(9). Available at: file:///C:/Users/Daiani/Downloads/Boletim_Graos_junho_2018.pdf.

Coradi PC, Fernandes CH, Helmich JC and Goneli AL (2016). Effects of drying air temperature and grain initial moisture content on soybean quality (*Glycine max* (L.) Merrill). *Eng. Agric.* 36(5): 866-876. https://doi.org/10.1590/1809-4430-Eng.Agric.v36n5p866-876/2016.

Cruz CD (2013). GENES – a software package for analysis in experimental statistics and quantitative genetics. *Acta Sci-Agron.* 35: 271-276. https://doi.org/10.4025/actasciagron.v35i3.21251.

Cruz CD, Regazzi AJ and Carneiro PCS (2004). Modelos biométricos aplicados ao melhoramento genético. (volume 1. Viçosa, Editora UFV.

Dhungana SK, Kulkarni KP, Kim M, Ha BK, et al. (2017). Environmental Stability and Correlation of Soybean Seed Starch with Protein and Oil Contents. *Plant Breed. Biotech.* 5: 293-303. https://doi.org/10.9787/PBB.2017.5.4.293.

Feng L, Burton JW, Carter TE and Pantalone VR (2004). Recurrent half-sib selection with testcross evaluation for increased oil content in soybean**.** *Crop Sci*. 44(1): 63-69. https://doi:10.2135/cropsci2004.6300.

Finoto E, Sediyama T, Reis M and Cruz C (2008). Teores de óleo e de proteína de genótipos de soja em duas épocas de plantio e dois tipos de solo em frutal-MG. *XXX Reunião de Pesquisa de Soja da Região Central do Brasil*, 53.

Finoto EL (2008). Variabilidade fenotípica dos teores de óleo e proteína de cultivares de soja em diferentes ambientes**.** 116p. Doctoral Thesis. Thesis (Doctoral Degree in Phytotechnology) - Universidade Federal de Viçosa, Viçosa.

Haghi Y, Boroomandan P, Moradin M and Hassankhali M (2012). Correlation and path analysis for yield, oil and protein content of Soybean (*Glycine max* L.) genotypes under different levels of nitrogen starter and plant density**.** *Biharean Biol.* 6: 32-37. http://biozoojournals.ro/bihbiol/cont/v6n1/bb.121101.Haghi.pdf.

Hamer E (1999). Maturação de sementes de soja no trópico úmido. 58f. Thesis (Doctoral Degree in Seed Science and Technology) - Universidade Federal de Pelotas, Pelotas, RS, Brazil.

Hanson WD (1992). Modified seed maturation and seed yield potentials in soybean. *Crop Sci.* 32: 972-976. http://doi:10.2135/cropsci1992.0011183X003200040028x.

Helms TC and Orf JH (1999). Protein, oil and yield of soybean lines selected for increased protein. *Crop Sci.* 38: 707-711. https://doi.org/10.2135/cropsci1998.0011183X003800030015x.

Hwang EY, Song Q, Jia G, Specht JE, et al. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics.* 15: 1-12. http://www.biomedcentral.com/1471-2164/15/1.

Kottek M, Grieser J, Beck C and Rudolf B (2006). World map of the Köppen-Geiger climate classification updated. *Meteorol. Z.* 15: 259-263. https://doi.org/10.1127/0941-2948/2006/0130.

Liu K (2012). Soybeans: Chemistry, Technology, and Utilization. Ed. Springer. 532.

Masuda R (2004). The strategy for sweetness increase of vegetable soybeans: maltose, another sugar in boiled seeds. In: Proceedings of 7th World Soybean Research Conference/6th International Soybean Processing and Utilization Conference/3rd Brazilian Soybean Congress, Foz do Iguacu, Brazil. 839-44.

Mello Filho OLD, Sediyama CS, Moreira MA and Reis MS (2004). Grain yield and seed quality of soybean selected for high protein content. *Pesq. Agropec. Bras.* 39(5): 445-450. https://doi.org/10.1590/S0100-204X2004000500006.

Miranda ZDFS, Arias CAA, Prete CEC, Kiihl RADS, et al. (2007). Genetic characterization of ninety elite soybean cultivars using coefficient of parentage. *Pesq. Agropec. Bras*. 42(3): 363-369. https://doi.org/10.1590/S0100-204X2007000300009.

Moraes RMAD, José IC, Ramos FG, Barros EGD, et al. (2006). Caracterização bioquímica de linhagens de soja com alto teor de proteína. *Pesq. Agropec. Bras*. 41(5): 725-729. https://doi.org/10.1590/S0100-204X2006000500002.

Oliveira MA, Carrão-Panizzi MC, Leite R, Campos Filho PJ, et al. (2010). Quantificação dos teores de açúcares, oligossacarídeos e amido em genótipos/cultivares de soja (Glycine max (L.) Merrill) especiais utilizados para alimentação humana. *Braz. J. Food Technol*. 13: 23-29. http://www.alice.cnptia.embrapa.br/alice/handle/doc/856227.

Panwar A, Chauhan P and Srivasta VAJ (2018). The Oxidative and Antioxidative Response in *Glycine max* (L.) Merr. Exposed to Various Abiotic Stress. *Int. J. Sci Res*. 7: 72-80. http://DOI: 10.21275/ART20175794.

Pípolo AE (2002). Influência da temperatura sobre as concentrações de proteína e óleo em sementes de soja (Glycine max (L.) Merrill). Thesis (Doctorate) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

Popović V, Vidić M, Jocković Đ, Ikanović J, et al. (2012). Variability and correlations between yield components of soybean [*Glycine max* (L.) Merr.]. *Genetika*. 44: 33-45. https://doi.org/10.2298/GENSR1201033P.

Poutanen K, Sozer N and Della Valle G (2014). How can technology help to deliver more of grain in cereal foods for a healthy diet?. *J. Cereal Sci*. 59: 327-336. https://doi.org/10.1016/j.jcs.2014.01.009.

Qin P, Song W, Yang X, Sun S, et al. (2014). Regional distribution of protein and oil compositions of soybean cultivars in China. *Crop Sci*. 54: 1139-1146. https://doi:10.2135/cropsci2013.05.0314.

Rangel MAS, Minuzzi A, Lucca A, Scapim CA, et al. (2007). Efeitos da interação genótipos x ambientes no rendimento de grãos e nos teores de proteína de cultivares de soja. *Acta Sci. Agron*. 29(3). https://DOI: 10.4025/actasciagron.v29i3.280.

Rodrigues JIDS, Arruda KMA, Cruz CD and Barros EGD (2017). Genetic divergence of soybean genotypes in relation to grain components. *Cienc. Rural*. 47(2). https://doi.org/10.1590/0103-8478cr20151258.

Rodrigues JIDS, Arruda KMA, Piovesan ND, Barros EGD, et al. (2016). Selection of progenitors for increase in oil content in soybean. *Rev. Ceres*. 63: 661-667. https://doi.org/10.1590/0034-737x201663050010.

Rodrigues JIDS, Miranda FDD, Ferreira A, Borges LL, et al. (2010). Mapeamento de QTL para conteúdos de proteína e óleo em soja. *Pesq. Agropec. Bras*. 45(5): 472-480. https://doi.org/10.1590/S0100-204X2010000500006.

Singh D (1981). The relative importance of characters affecting genetic divergence. *Indian J. Genet. Pl. Br*. 41: 237-245.

Slavin J (1991). Nutritional benefits of soy protein and soy fiber. *J. Am. Diet Assoc*. 91: 816-819.

Streck EV, Kämpf N, Dalmolin RSD, Klamt E, et al. (2008). *Solos do Rio Grande do Sul*. Porto Alegre: UFRGS: EMATER/RS-ASCAR.

Team RC (2018). R Foundation for Statistical Computing; Vienna, Austria: 2015. *R*: A language and environment for statistical computing, 2013.

USDA – United States Department of Agriculture Economic Research Service. *Crop Production 2018 Summary*. 2019. ISSN: 1057-7823. https://www.nass.usda.gov/Publications/Todays_Reports/reports/cropan19.pdf.

Wee CD, Hashiguchi M, Ishigaki G, Muguerza M, et al. (2018). Evaluation of seed components of wild soybean (*Glycine soja*) collected in Japan using near-infrared reflectance spectroscopy. *Plant Genet. Resour. C*. 16: 94-102. DOI: https://doi.org/10.1017/S1479262116000472.