

## Assessment of genetic variability in sesame accessions using SSR markers and morpho-agronomic traits

Eveline de Sousa Araújo<sup>1</sup>, Nair Helena Castro Arriel<sup>2</sup>, Roseane Cavalcanti dos Santos<sup>2</sup>, Liziane Maria de Lima<sup>\*2</sup>

<sup>1</sup>Universidade Estadual da Paraíba/ Embrapa Algodão, Pós-Graduação em Ciências Agrárias, Campina Grande, PB, Brazil

<sup>2</sup>Embrapa Algodão, Campina Grande, PB, Brazil

\*Corresponding author: [liziane.lima@embrapa.com](mailto:liziane.lima@embrapa.com)

### Abstract

The genetic variability of thirty-six sesame accessions were evaluated using molecular and morpho-agronomic data, aiming to identify divergent genotypes for further use in breeding program. Ten SSR markers and twenty seven morpho-agronomic traits were used to estimate genetic divergence by means of multivariate Tocher and UPGMA methods. The GENES program was used for statistical analysis of data. We observed that molecular and morpho-agronomic data were efficient to estimate divergence among the accessions. In molecular aspect, the ZM\_22, ZM\_45 and ZM\_34 markers showed broad contribution by presenting polymorphism rates of 0.53, 0.44 and 0.39, respectively. The groups formed in the Tocher model were not similar to those formed by the UPGMA method. Among the groups formed, the study of genetic diversity allowed identification of characteristics of interest such as precocity in the genotypes ABG 591 and ABG 649. The degree of similarity assembling revealed the most similar (ABG 591, ABG 616, ABG 649, ABG 141 and ABG 688) and identifying the most divergent (ABG 648 and ABG 200) genotypes. These results revealed significant genetic variability among the investigated accessions of the sesame ABG that can be applied in the breeding programs.

**Keywords:** Genetic Diversity; Multivariate Methods; *Sesamum indicum* L.; Tocher Grouping and UPGMA.

**Abbreviations:** SSR\_simple sequence repeat; ABG\_active germplasm bank; PH\_plant height; HIFF\_height of insertion of the first fruit; MMS\_mass of 1000 seeds; NFA\_number of fruits per leaf axils; NFP\_number of fruits per plant; FS\_size of the fruit.

### Introduction

The sesame (*Sesamum indicum* L.) is an oleaginous species of ancient use. The seeds have high nutritional value, due to protein quality and oil, and are also an important source of anti-oxidants compounds, like sesamol and sesamolina, both with anti-hypertensive and anti-cancerous properties (Dar et al., 2015).

The sesame plant is easy to handle and has wide environmental adaptation; however, biotic and abiotic stresses may affect the crop yield, depending on the intensity and duration (Gepstein and Glik, 2013; Cooper et al., 2014; Wang et al., 2014). Nowadays, the main challenge of agriculture is to develop technologies to increase food production and allow sustainability, especially under limited resources of the environment. Therefore, the efficient exploitation of germplasm based on molecular genetics and phenotyping generates important information to the breeding programs and the characterization of the species. The studies of genetic divergence have been widely adopted for accession classification, aiming to estimate the similarity

genetic degree so that it comes to help in the choice of divergent parental in breeding program (Kaur et al., 2014). Several descriptors may be used for the classification of germplasm. The morphologic and agronomic traits are the most adopted due to the availability of valuation of variables that contributes the phenotyping. Besides, several researchers have adopted the molecular markers by acting directly on the individual genome, presenting the advantage of generating a stronger classification. This efficiently helps in the genetic discrimination of the accessions (Wei et al., 2011; Sant'ana et al., 2012; Luo et al., 2015).

The analysis of database is based on unique and multivariate models. The latter have offered large contribution in the discrimination of the accessions and indication of possible genitors, which may be used in breeding programs. The literature provides several methodologies to estimate the genetic diversity in collections of germplasm, based on qualitative and quantitative descriptors. The grouping methods are the most adopted as long as they can join accessions based in some criteria that presents similarity in

the behavior pattern in relation to the character set (Cruz et al., 2013). In this work, we adopted multivariate method to estimate genetic similarity in sesame accessions, based in morphological, agronomic, and molecular descriptors.

## Results and discussion

### *Morpho-agronomic featuring*

The classification of accessions was proceeded based on the morphologic evaluation of the sesame collection (Table 1), in which they grouped for each morpho-agronomic descriptor (Supplementary Fig 1) (IPGRI and NBPGR, 2004). The importance of this phenotype classification is that it provides direct information to the breeder about the potential of each descriptor for using in the selection procedures, as well as the accessions that may be used for further compound populations to be improved. Considering the number of classes for each descriptor, the most contributive traits for accession classification were: stem, leaf, and capsule hairiness; shape of leaf, symmetry of leaf basis, angle of leaf on the main stem, extrafloral nectary, color of corolla and seeds. The least contributive were: growth habit, stem shape, and stem and leaf color. The Table 2 contains a synthesis of variance analysis of agronomic characters (PH, HIFF, NFP, NFA, FS and MMS; see abbreviations list) of evaluated sesame collection. We verified a meaningful statistic difference among the accessions for all characters, indicating that variability exists among the materials. Therefore, they may be adopted for characterizing collection material or as source of diversity for genetic improvement of the species. The accessions presented variation of 0.74 m (ABG 648) to 2.26 m (ABG 616) for plant height (PH), 12.0 cm (ABG 648) to 74.6 cm (ABG 668) for height of insertion of the first fruit (HIFF). For number of fruits per leaf axils (NFP), the average values of the accessions varied from 67 (ABG 608) to 322 (ABG 663), whereas for fruit size (FS) the variation was 2 to 3.66 cm in (ABG 668) and (ABG 649), respectively. For mass of 1000 seeds (MMS) the 36 accessions from the collection presented values ranging from 2.25 g (ABG 2135) to 4.25 g (ABG 648). Despite of the observed meaningful difference to all variables, Manivannan and Nadarajan (1996) and Arriel et al. (2007b) reported that NFP is one of the most contributive traits for the selection of genotypes of this species. Among the variables, we could remark MMS as differentiating in the accessions, in which genotypes with bigger seeds present bigger mass. The observed value on ABG 648 is similar to MMS of the BRS Anahi sesame cultivar, recently launched by Embrapa Algodão, presenting an MMS value of approximately 4.22 g (Arriel et al., 2015).

### *Diversity assessment through molecular tools*

Nine pairs of micro-satellite markers generated profiles of polymorphic bands on the 36 sesame accessions. The ten SSR loci revealed 22 alleles, the number of alleles per locus varied from 1 to 3, with average of 2, with amplified bands from 100 to 200 pb. The oligonucleotides ZM\_22, ZM\_34 and ZM\_45 were the ones with most contributions to characterization of accessions, for they presented three

alleles each (Table 3; Supplementary Fig 2). It is important to highlight that the markers SSR used in the present work were already selected by Wei et al. (2011). They used 40 SSR polymorphic markers for characterization of 24 accessions of sesame and obtained an average number of 6.55 alleles per loci, with a variation from 4 to 18, much superior quantity than the obtained in the present work. The frequency of alleles in molecular markers, as amplified by SSRs, depends on various factors such as the structure or composition of the genome (Toth et al., 2000), statistic method for detecting SSR, as well as the parameters for exploration of microsatellites (Wei et al., 2011). Some researches on determination of genetic diversity on sesame accessions have reported that SSR markers have presented low polymorphism. This has probably occurred due to the narrow genetic basis of the evaluated materials on those studies. However, this low polymorphism on sesame is also similar to other species like wheat and peanut, as reported by Zhang et al. (2012b). Several works on genetic variability in sesame have used different methodologies of evaluation and generated different results, depending on the type of marker, composition and size of germplasm. For example, Badri et al. (2014), evaluated 16 sesame accessions (among cultivated and wild species) and observed that, from 120 SSR markers, 92 were polymorphic and the number of alleles per microsatellite loci varied from 2 to 5, with average of 3.11 alleles. The size of the bands varied from 100 – 510 bp among the pairs of indicators. Dossa et al. (2016) used 33 SSRs in 96 accessions of sesame originated from different countries of Africa and Asia, and obtained 4.15 alleles in average, with variation from 2 to 10 alleles per loci from a total of 137 alleles. The content of polymorphic information (PIC) for each marker presented variation from 0.00 (ZM\_21) to 0.53 (ZM\_22), with average of 0.27 (Table 3), indicating that there is heterogeneity of polymorphism among the used markers. Such results are similar to the ones obtained by Zhang et al. (2012b) for evaluation of 32 SSR markers in 24 accessions of sesame, where the content of polymorphism varied from 0.08 to 0.67. Dossa et al. (2016) also obtained PIC values varying from 0.28 to 0.54 for the determination of 96 accessions of sesame coming from six African and Asian regions. The PIC was firstly described by Boststein et al. (1980), expressing the quality of the markers in genetic studies. This analysis takes the total number of fragments detected for each loci into consideration for an established marker and the frequency of those alleles in the set of genotypes investigated (Varshney et al., 2009). In theory, the PIC values may vary from 0 to 1. In a zero PIC value, the marker has just one allele. In a one PIC value, the marker could have an infinite numbers of alleles. A PIC >0.7 value is considered highly informative (Hildebrand et al., 1992) and presents existence of variability. Thus, SSR markers presenting this characteristic are the most indicated for the study of genetic variability (Ribeiro, 2011). In the present study, the molecular markers present low PIC values, even when there was polymorphism among the investigated accessions (Table 3). Barbosa et al. (2003) observed in the corn that the SSR markers presented heterogeneous distribution for individual PIC values, and

**Table 1.** The data obtained from morphologic descriptors.

Stem				
Plant growth type	indeterminate	75%	determinate	25%
Plant growth habit	erect	97%	semi-erect	3%
Stem shape in cross section	squared	97%	round	3%
Main stem colour	green	94%	yellow	6%
Stem trichome	glabrous (trichome absent)	6%	weak or sparse	78%
	medium	13%	strong or profuse	3%
Leaves				
Stem branching	mixed	81%	alternating	19%
Leaf shape	lanceolate	86%	ovate	11%
	linear	3%	-	
Basal leaf profile	flat	70%	reverse cup shaped (convex)	11%
	cup shaped (concave)	19%	-	
Basal leaf margin	entire	87%	serrate	13%
Lobe incision of basal leaf	absent	91%	weak	6%
	medium	3%	-	
Leaf angle to main stem	horizontal =90°	80%	acute <90°	6%
	drooping >90°	14%		
Leaf hairiness	glabrous (trichome absence)	16%	weak or sparse	78%
	medium	6%	-	
Petiole colour	green	97%	yellowish green	3%
Flower				
Corolla interior pigmentation	pigmentation along the lip region of corolla tube	98%	pigmented throughout	2%
Exterior corolla colour	white with pink shading	64%	light violet	22%
	white	11%	pink	3%
Interior corolla color	pink	94%	dark violet	6%
Extra-floral nectary development	medium	47%	rudimentary	36%
	small	17%	-	
Seed coat color	white	55%	mixture of white and beige	31%
	cream	8%	mixture of dull black and cream	3%
	dull black	3%	-	
Capsule				
Bicarpellate capsule shape	tapered at apex	61%	narrow oblong	33%
	square	5%	-	
Capsule hairiness	strong or profuse	11%	medium	22%
	weak or sparse	64%	Glabrous (trichome absent)	3%
Capsule arrangement	multi-capsular	83%	monocapsular	17%
	white	55%	mixture of white and beige	31%
Seed coat color	cream	8%	mixture of dull black and cream	3%
	dull black	3%	-	

**Table 2.** Summary of the analysis of variance for the three agronomic characters of sesame evaluated in Patos, PB, 2015.

Source of variation	GL	Mean square					
		PH	HIFF	NFP	NFA	FS	MMS
Blocks	4	0.03	0.18	11.55	0.54	0.07	0.00
Genotypes	35	0.89**	11.16**	35.38**	6.13**	0.51**	1.48**
Error	140	0.03	0.83	10.37	0.23	0.07	0.00
Mean		1.43	37.44	206.35	2.86	3.11	3.02
CV (%)		12.63	15.37	23.26	14.37	8.78	0.02

\*\*Meaningful for the test F ( $p < 0.01$ ). plant height (PH); height of insertion of the first fruit (HIFF); number of fruits per plant (NFP); number of fruits per leaf axils (NFA); size of fruit (FS); mass of 1000 seeds (MMS).

**Table 3.** Parameters of SSR oligonucleotides used and estimated in the sesame accessions. Campina Grande, PB, 2015.

Primer	Nº of allele	He	Ho	F	PIC
ZM_10	2	0.42	0.00	1.0	0.33
ZM_11	2	0.47	0.22	0.53	0.36
ZM_13	2	0.19	0.00	1.00	0.17
ZM_20	2	0.47	0.00	1.00	0.36
ZM_21	1	0.00	0.00	1.00	0.00
ZM_22	3	0.60	0.44	0.26	0.53
ZM_34	3	0.50	0.11	0.78	0.39
ZM_40	2	0.02	0.02	-0.01	0.02
ZM_45	3	0.55	0.47	0.14	0.44
ZM_47	2	0.15	0.00	1.00	0.14
Média		0.34	0.12	0.62	0.27

He – Expected heterozygosity, Ho – observed heterozygosity, F- endogamy coefficient, PIC – content of polymorphism information.

**Table 4.** Grouping of the 36 accessions of sesame by Tocher's optimization method, from the morphoagronomic and molecular data. Campina Grande, PB, 2015.

Groups	Genotypes
1	ABG-177, ABG-257, ABG-606, ABG-586, ABG-579, ABG-619, ABG-698, ABG-615, ABG-221, ABG-591, ABG-688, ABG-141, ABG-G6, ABG-232, ABG-675, ABG-587, ABG-2135, ABG-649
2	ABG-536, ABG-666, ABG-608, ABG-235, ABG-172
3	ABG-668, ABG-663, ABG-673, ABG-G5
4	ABG-612, ABG-664
5	ABG-616, ABG-G3
6	ABG-648, ABG-200
7	ABG-511
8	ABG-687
9	ABG-G4

**Table 5.** Correlation among pairs of matrices of dissimilarities with qualitative and quantitative data obtained from evaluation of 36 sesame accessions. Campina Grande, PB, 2015.

Correlation between matrices	Agronomic	Morphological	Molecular
Agronomic	-	0,54 <sup>ns</sup>	0,82*
Morphological	0,54 <sup>ns</sup>	-	0,55 <sup>ns</sup>
Molecular	0,82*	0,55 <sup>ns</sup>	-

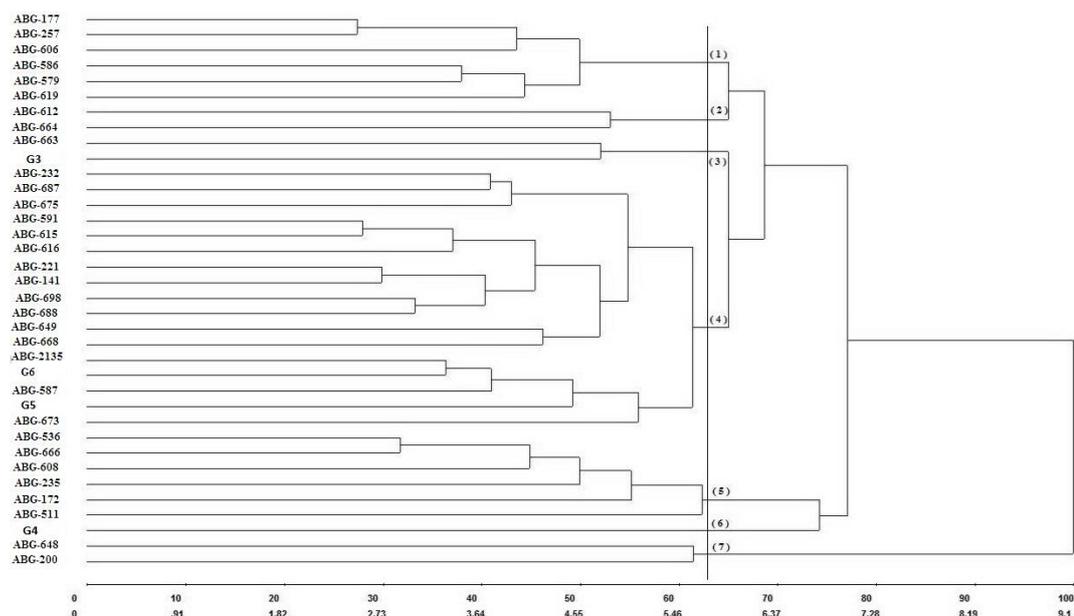
<sup>ns</sup>, \*: non-significant, significant to 5% of probability by t-test.

**Table 6.** Identification of sesame accessions used for genetic similarity study. Campina Grande, PB, 2015.

Nº	Identification	Nº	Identification	Nº	Identification	Nº	Identification
1	ABG 172	10	ABG 221	19	ABG 235	28	ABG 666
2	ABG 586	11	ABG 177	20	ABG 511	29	ABG 608
3	ABG 591	12	ABG 257	21	ABG G4	30	ABG 141
4	ABG 615	13	ABG 579	22	ABG 232	31	ABG 663
5	ABG 616	14	ABG 698	23	ABG 2135	32	ABG G3
6	ABG 612	15	ABG 649	24	ABG G6	33	ABG 688
7	ABG 648	16	ABG 664	25	ABG 536	34	ABG 606
8	ABG 200	17	ABG 668	26	ABG G5	35	ABG 687
9	ABG 619	18	ABG 587	27	ABG 673	36	ABG 675

**Table 7.** List of primers SSRs with respective names and sequences of nitrogen basis (Wei et al., 2011).

Primers		Sequence (5' → 3')
ZM_10	Sense	ATGCCCATCTCCATATACTCT
	Antisense	AATTCTTGCTGACTCTACG
ZM_11	Sense	GGATTCTCTAGACATGGCTTT
	Antisense	AACGCAGAATTCTCTCCTACT
ZM_13	Sense	GCAGAAGGCAATAAAGTCAT
	Antisense	GCGTCAGAAGAAAAACTGG
ZM_20	Sense	GGGATGTTGATAGAGATGTTG
	Antisense	TCTTCACTCTCACACACACA
ZM_21	Sense	CTCTCTCTCTGCTGTTTCA
	Antisense	GCCATACGATCTCAAATCAC
ZM_22	Sense	ACCACCGATCTACTACTTTT
	Antisense	CCACTGCACACTACAGTTTTT
ZM_34	Sense	AAGTCCCTTTTCAAGCAATC
	Antisense	GAGAGAGGAAAATGCAGAGAG
ZM_40	Sense	CGAAAAGGGAGTTACAGTTAAG
	Antisense	CTTCCTCTCTATCATCCTGT
ZM_45	Sense	GCAAAATCTCTGTTGTCTCAG
	Antisense	GCAAAATCTCTGTTGTCTCAG
ZM_47	Sense	GTTTCCAGGTCTATTCCTTTG
	Antisense	AGGTAGAGCTAATCCTTACCG



**Fig 3.** Dendrogram obtained by the UPGMA hierarchical clustering method from the morpho-agronomic and molecular descriptors, representing the thirty six genotypes of *S. indicum* L. investigated.

that this has probably occurred due to low number of polymorphic loci evaluated for those markers.

The expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_o$ ) and endogamy coefficient ( $F$ ) values were 0.34, 0.12 and 0.62, respectively. This indicates considerable level of endogamy among the accessions, which is corroborated from the difference among  $H_e$  and  $H_o$  values in most of loci, and may be result of a high rate of self-pollination of this population under study (Wei et al., 2011). This suggests that such accessions are not very contributory to generate genetic variability in pre-breeding works. This suggests crosses between more divergent accessions to increase the probability of increasing the percentage of variability in the segregating population to be obtained.

#### **Analysis of genetic diversity by multivariate techniques**

From the sum of the matrixes of dissimilarities from morphologic and agronomic descriptors, and the band profile generated by SSR markers, a multivariate analysis was performed, in which the dissimilarities measures varied from 2.49 to 9.10, suggesting a narrow genetic basis among the 36 accessions evaluated. The most similar genotypes were ABG 177 and ABG 257, and the most diverging ones were ABG 172 and ABG 648. By the grouping methods of optimization of Tocher, and hierarchical UPGMA, it was possible to find out the formations of different agglomerates.

Nine groups were formed by Tocher's optimization method (Table 4). The first grouping consists of approximately 53% of the accessions, in which the most frequent ones are precocious and plants with erect growth habit, green leaf color, absence of incision in the leaf. In the second group the genotypes with sparse hairiness on the stalk, shape of the blade leaf with entire margin and absence of global incision in the leaf, extrafloral nectary of middle size, shape of wide rectangular capsule were assembled. In group 3 there were plants with undeterminate stalk shape, erect growth habit, green leaf with mixed arrangement, angle of the flower on stem = 90°, sparse hairiness of capsule and multi-capsular arrangements.

In group 4, the shape of stalks of the plants was determinate, the habit of erect growth, lanceolate shape green colored leaf, with smooth basal profile and entire margin, average extrafloral nectary, white with shadowed pink color of the corolla of inside the flower, wide rectangular capsule shape with sparse hairiness. In group 5, genotypes were found out with erect growth habit, green leaves with sparse hairiness, mixed arrangement and lanceolate shape, full leaf margin with absence of global incision, rudimental extrafloral nectary, capsule with wide rectangular shape. Group 6 kept the accessions ABG 648 and ABG 200 with particular characteristics, as violet flower, lower plants, with higher seed production.

The groups 7, 8, and 9 were formed by only one accession each, being the most distant ones (Table 4). By the Tocher's agglomerative method of optimization, one single individual was included into a determined forming group, if its average distance related to this group does not surpass certain established value. This value is generally taken as the higher

amplitude of the set of the lower distance estimative, involving each of the genotypes in agglomerates (Arriel, 2004). So, we can confirm the ABG-511, ABG-687 and ABG-G4 accession divergence related to the others evaluated genotypes.

The formation of the clusters by the UPGMA method (Figure 3) was not similar to that of Tocher's analysis. In the hierarchical groupings, delimitation is done in a subjective way, observing the points of high level change of the dendrogram; thus, different patterns of grouping can be adopted in function of the genetic relation of the genotypes (Arriel, 2004). After construction of the dendrogram, a Cophenetic Correlation Coefficient of 0.71 was generated, which translates the degree of fit between the data matrix of original dissimilarities and the resultant matrix of the groups for the dendrogram formation, indicating the higher the coefficient value, the lower is the distortion caused in the genotypes grouping. Values over 0.70 provide an excellent representation of distances in the formation of the dendrograms (Cruz, 2008).

The formation of seven groupings was observed adopting a percentage of genetic divergence of about 63%. In group 1, the accessions were assembled with particular characteristics as erect growth habit, green main stem with sparse hairiness, stem squared shaped, green colored leaf with lanceolate shape, full leaf margin with absence of incision. Group 2 included the accessions with development of determined stalk, leaves with lanceolate shape and symmetry of smooth foliar basis.

Group 3 encompassed plants with sparse leaf hairiness, mixed leaf arrangement, with angle of leaves on the main stem of 90° and rudimental extrafloral nectary. Group 4 kept the most part of accessions (47%), among them the CNPA G2 (ABG 698) cultivar, with medium cycle and tolerance to angular spot (Arriel et al., 2006) and is part of the work collection of Embrapa Algodão together with ABG G5 and ABG G6 accessions. They belong to exotic genotypes from Pakistan and Mexico, respectively. We also see in group 3 the most premature accessions (ABG 591 and ABG 649), with characteristics like: undeterminate stalk shape, arrangement of multi-capsular capsules, higher plants, consequently, the height of insertion of the first fruit is also bigger, in which the biggest NFP (ABG 663, with 322 fruits) and biggest FS (ABG 649, with 3.66 cm) were obtained. The NFP is one of the components of production that can determine higher productivity with possibility to select superior genotypes, associated with other agronomic characteristics important to the genetic improvement program of the species.

In group 5, the genotypes have the green colored main stems, with squared stem shape, green colored leaf, lanceolate shaped leaf, full leaf margin, absence of incision on the leaf and sparse hairiness. Group 6 held only one accession, showing most divergent nature from the others. In group 7, the accessions presented an all violet pigmented color on the flower, and the smallest plants; however with higher MMS values.

By comparing the formation of the groupings by the hierarchical methods (UPGMA), and optimization (Tocher), despite the formation and composition of the different

conglomerates, some accessions have kept their position in function of the criteria that each method imposes. In this context, the same identification of accessions of higher similarity was observed, e.g. for the ABG-177, ABG-257, ABG-606, ABG-586, ABG-579 and ABG-619 accessions, and higher divergence accessions, like ABG-511 and ABG-G4.

Another point that should be taken into consideration in the composition of the agglomerates is that, besides the pressure of selecting improvement works with focus on genetic divergence, there is still the geographic origination as synonym of dissimilarity. The relation of geographic origination as indicator of genetic diversity usually does not result in a greater genetic distance (Cruz and Carneiro, 2003). However, in both groupings (Tocher and UPGMA) the presence of genotypes of different origins in different groups can be observed. Dossa et al. (2016) evaluated 96 sesame accessions, originated from 22 countries in six Asian and African regions, and reported that 44.66% of genetic variability was due to diversity among geographic regions and 55.34% inside regions. However, Arriel et al. (2006), through RAPD markers, and Bandilla et al. (2011) from agromorphologic descriptors, reported distribution patterns of genotypes in different groups. However, they indicated that divergence was not related to geographic differentiation. Moreover, the markers that they used allowed identifying variability among the accessions, placing them in function of the methods of grouping adopted in the respective criteria. Therefore, in the present work, the morphologic, agronomic, and molecular markers were also determiners in the differentiation of the 36 accessions analyzed.

The results of the correlations among the matrices of the three marker types (morphological, agronomic and molecular) used in the evaluation of the accessions are given in Table 5. A significant and high correlation was observed between the agronomical and molecular matrices, which will probably facilitate the identification of phenotypically different individuals in reference to the agronomic descriptors. It is emphasized that, although most of the correlation indexes presented values above the average, these were not significant. In the evaluation of diversity determination of 12 sesame accessions, Zhang et al. (2012a) did not find any significance among the phenotype and molecular markers.

The non-significance of the correlations among the evaluated markers has probably occurred due to assignment of different values by each marker to dissimilarities among the accessions (Arriel, 2004). By the analysis, correlations involving the morphological markers were not significant. However, in the morphological classification (descriptive analysis), we found that more than 3 classes occurred for the identification of the accessions, reflecting variability inside the population under study for this kind of marker. Vieira et al. (2013) mentioned that such results should not be considered as a limitation of different markers used to access the genetic variability, but as an indication of complementarity among them.

It is also emphasized that in application of multivariate techniques to determinate the diversity of a group of individuals, the basic aims are to reduce the number of variables that explain the total available variation and the

consequent simplification in obtaining distance among individuals that are being compared. In this context, the information in an individual level is lost, and the remainders are referred to the group average. So, the morphologic descriptors used in the classification of sesame accessions were capable to differentiate the accessions and reflected phenotypic variability among the accessions. Probably, what remained was the information of the means of the groups, when analyzing the three types of markers used together.

The different markers used in this work were helpful to collaborate in the classification of sesame accessions. The germplasm evaluated presented higher variation through agronomical and molecular descriptors and the correlation among them was significant (Table 5). Furthermore this germplasm highlighted formation of grouping among the most premature accessions (ABG 591 and ABG 649) and the ones which have the higher number of capsules per plant (ABG 616, ABG 221, and ABG 141) in the same group by both agronomic and molecular evaluation. The primers ZM\_22 and ZM\_45 generated greater polymorphism and were able to identify the most divergent accessions (ABG 648 and ABG 200). These accessions were differentiated by the plants with lower height and the highest MMS values as important agronomic characteristics for selection of progenitors. A positive and significant correlation between measures of genetic variation of quantitative characteristics and molecular markers has been verified in other studies and even in different species such as *Cagaiteira* (*Eugenia dysenterica* DC., Myrtaceae) (Aguiar et al., 2011; Vieira et al., 2013). In spite of lack of significant correlation among morphologic descriptors and the molecular markers in this study, it is emphasized that the stem hairiness, leaf shape, foliar basis symmetry, leaf incision, angle of the leaf on the main stem, leaf hairiness, extrafloral nectary, color of the flower corolla, capsule hairiness, and color of the seeds were contributive to discrimination of sesame accessions, because they presented more than one class. However, taking into consideration the number of descriptors used, the relative contribution of these ten descriptors for the association was influenced by the multivariate analysis criteria. In the genetic diversity studies using multivariate analysis, where various features are evaluated simultaneously, the use of this technique allows to unify the information. In this way, each genotype is represented by a single value related to its analyzed characteristics (Val et al., 2014).

The data generated in this study will provide support for the breeding program of sesame in order to support the various segments. The molecular tools, such as SSR markers may be used to identify accessions that have better characteristics of agronomic interest, widening possibilities of crossing for the obtainment of segregating populations.

## Materials and Methods

### Genetic resources and experiment conduction

The experiment was conducted in the Experimental Field of Embrapa Algodão (Cotton), in the city of Patos-PB (07°00'11" S and 37°18'53" W, 250 m). According to Köppen's classification, the climate of the region is semiarid

Bsh, with a hot and dry season from June to December, and another season with scarce rain fall from January to May, with irregularly distributed annual average of 700 mm. The soil of the experimental area is classified as Chromic Luvisol (Brasil, 1978; Burgos et al., 2002). The area of cultivation was previously corrected and fertilized according to the recommendation suggested to the crop, as 90 kg ha<sup>-1</sup> monoammonium phosphate (MAP) and 20 kg ha<sup>-1</sup> urea. After the final roughing, at nearly 40 days, coverage of 90 kg ha<sup>-1</sup> of urea and 20 kg ha<sup>-1</sup> of potassium chloride was applied and at 60 days a second coating was applied with 90 kg ha<sup>-1</sup> of urea and 15 Kg ha<sup>-1</sup> of potassium chloride. The crop management was performed according to Arriel et al. (2007a). During the conduction of experiment (from March to June, 2015), the average temperature records varied from 22.2 °C to 34.3 °C and the relative humidity of 61%. The total rainfall registered in the growing period was 214 mm (AESA, 2015).

Thirty six sesame accessions belong to the Embrapa's Germplasm Bank were used for the analysis (Table 6). The experimental unit consisted of three rows of 3 m in length, spacing of 1.0 x 0.2 m and one plant per plot, after roughing. The experimental design adopted was randomized block with five repetitions. The harvesting, batting, sifting, and ventilation were performed manually. The evaluations were done at approximately 60 days in five plants selected in the central row (morphological variables) and 90 days (agronomic variables) after germination.

#### **Morphologic and agronomic descriptors**

The phenotyping of the accessions was performed using morphological, agronomic and molecular descriptors. Morphologic descriptors were adapted based on IPGRI and NBPGR (2004): stem development, growth pattern, stem shape, stem color, stem hairiness, leaf arrangement, leaf shape, leaf basis symmetry, leaf margin, leaf incision, leaf angle of main stem, hairiness of leaf, color of leaf, extra-floral nectar, corolla color, flower pigment, color of the flower lobules, capsule arrangement, capsule shape, hairiness of capsule, and color of the seeds. Besides these, other two descriptors were analyzed: absence and presence of honeycombing (present in all flowers) and opening of the capsule to maturation (all accessions presented characteristics of dehiscence fruits, meaning that fruits were opened asynchronously during the plant ripening, providing possible loss of seeds, even due to the wind action). However, these were irrelevant because they did not show divergence among the accessions; therefore, they were not included in the statistical evaluations.

The agronomic descriptors adopted were: plant height (PH), height for insertion of the of the first fruit (HIFF), number of fruits per plant (NFP), number of fruits per leaf axils (NFA), size of the fruit (SF), and mass of 1000 seeds (MMS).

#### **Molecular analysis**

For a better featuring of molecular characterization of the samples, Simple Sequence Repeats (SSRs) like marker were adopted, previously isolated for sesame (Wei et al., 2011),

and used in PCR (Polymerase Chain Reaction) experiments. The genomic DNA of sesame seeds was extracted following the method of McDonald et al. (1994). The reactions of PCR were made by the use of 10 µg genomic DNA; reaction buffer (1X); MgCl<sub>2</sub>, (2.5 mMol.L<sup>-1</sup>); dNTPs (0.25 µMol.L<sup>-1</sup>), primers sense and anti-sense (5 mMol.L<sup>-1</sup>) (Table 7); and Taq DNA polymerase (2.5 U). the reactions were conducted in thermo-cycler (*Amplitherm Thermal Cyclers*) with the following program: 94 °C/4 min, 35 cycles of 94 °C/40s, 55 °C/40 s and 72 °C/1 min, finally 72 °C/10 min. The products of the reactions were colored with SYBR® *Gold* (Invitrogen) and separated in electrophoresis in agarose gel (3.5%). The band pattern was photo documented (*Imaging System, model El Logic 200*).

#### **Statistical analysis**

The proceeding of multi-categorical analysis with all agronomic and morphological database were applied to the subsequent grouping analysis. For the molecular markers, the descriptive statistic was made for the following parameters: effective and total number of alleles per loci, endogamy coefficients (F), expected heterozygosis (He), and observed (Ho) and information content of polymorphism (PIC). The analysis involving all markers was made through sum and correlation among the matrixes of dissimilarities with qualitative data (molecular and morphologic), and quantitative (agronomic) and value significance obtained by the t-tests and Mantel (Cruz, 2008). For the grouping analysis, two methods were adopted: Optimization of Tocher and UPGMA. All statistical procedures were performed using the GENES program, 2015.5.0 version (Cruz, 2013).

#### **Conclusion**

The molecular and morpho-agronomic markers were effective to estimate the variability among the sesame accessions and allowed to identify the most similar (ABG 591, ABG 616, ABG 649, ABG 141 and ABG 688) and the most divergent (ABG 648 and ABG 200) genotypes. Moreover, it was possible to identify the genotypes with important agronomic characteristics for the genetic improvement of the species such as: most premature accessions ABG 591 and ABG 649, the latter with longer fruits; and the ABG 648 and ABG 663 accessions, with higher NFP and higher MMS, respectively.

#### **Acknowledgements**

To Embrapa Algodão for the financial support and to CAPES for the grants.

#### **Conflict of interest**

The authors declare no conflict of interest between the partners with the dissemination of results.

#### **References**

AESA. Agência executiva de gestão das águas da Paraíba. Available in: <<http://www.aesa.pb.gov.br/>> Access in 04 August 2015.

- Aguiar AV, Moura NF, Moura MF, Zucchi MI, Vencovsky R, Chaves LJ (2011) Relação entre a variação genética de caracteres quantitativos e marcadores moleculares em subpopulações de cagaiteira (*Eugenia dysenterica* DC). Rev Bras Frutic. 33:157-169.
- Arriel NHC (2004) Diversidade genética em gergelim (*Sesamum indicum* L.) a partir de marcadores moleculares (RAPD) e caracteres morfológicos e agrônômicos. 114 p. These – Universidade Estadual Paulista, Jaboticabal-SP.
- Arriel NHC, Di Mauro AO, Di Mauro SMZ, Bakke OA, Unêda-Trevisoli SH, Costa MM, Capeloto A, Corrado AR (2006) Técnicas multivariadas na determinação da diversidade genética em gergelim usando marcadores RAPD. Pesq Agrop Bras. 41:801-809.
- Arriel NHC, Di Mauro AO, Arriel EF, Unêda-Trevisoli SH, Costa MM, Bárbaro IM, Muniz FRS (2007a) Genetic divergence in sesame based on morphological and agronomic traits. Crop Breed Appl Biotechnol. 7:253-261.
- Arriel NHC, Firmino PT, Beltrão NEM, Soares JJ, Araújo AE, Silva AC, Ferreira GB (2007b) A cultura do gergelim – Brasília, DF: Embrapa Informação Tecnológica. Coleção Plantar. 50:9-72.
- Arriel NHC, Sousa SL, Heuert J, Medeiros AA, Gondim TMS, Firmino PT, Vasconcelos RA, Dantas ESB (2015) Gergelim BRS Anahí, EMBRAPA-CNPQ, Campina Grande, (Folder).
- Bandilla S, Ghanta A, Natarajan S, Subramoniam S (2011) Determination of genetic variation in Indian sesame (*Sesamum indicum*) genotypes for agro-morphological traits. J Res Agricult Sci. 7(2):88-99.
- Badri J, Yepuri V, Ghanta A, Siva S, Siddiq EA (2014) Development of microsatellite markers in sesame (*Sesamum indicum* L.). Turk J Agric For. 38:603-614.
- Barbosa AMM, Geraldi IO, Benchimol LL, Garcia AAF, Souza JCL, Souza AP (2003) Relationship of intra- and interpopulation tropical maize single cross hybrid performance and genetic distances computed from AFLP and SSR markers. Euphytica. 130(1):87-99.
- Brasil/Ministério da Agricultura (1978) Estudos básicos para o levantamento agrícola: Aptidão agrícola das terras da Paraíba. Brasília: Binagri. 3:1-23.
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. 32:314-331.
- Burgos N, Cavalcanti AC, Silva FHBB, Oliveira Neto MB (2002) Solos do campo experimental da Embrapa Algodão, em Patos-PB. Levantamento detalhado e potencial edáfico. Rio de Janeiro: Embrapa Solos. 1-112.
- Cooper M, Messina CD, Podlich D, Totir LR, Baumgarten A, Hausmann NJ, Wright D, Graham G (2014) Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci. 65:311-336.
- Cruz CD, Carneiro PCS (2003) Modelos biométricos aplicados ao melhoramento genético. Editora UFV 2, p.585. Viçosa.
- Cruz CD (2008) Programa Genes - Diversidade Genética. Editora UFV. p.278. Viçosa.
- Cruz RN, Azevedo CAV, Fernandes JD, Monteiro Filho AF, Wanderley JAC (2013) Adubação orgânica residual no crescimento e produção do gergelim irrigado com água residuária. Rev Verde. 8:257-263.
- Dar AA, Verma NK, Arumugam N (2015) An updated method for isolation, purification and characterization of clinically important antioxidant lignans - Sesamin and sesamol, from sesame oil. Ind Crops Prod. 64:201-208.
- Dossa K, Wei X, Zhang Y, Fonceka D, Yang W, Diouf D, Liao B, Cissé N, Zhang X (2016) Analysis of genetic diversity and population structure of sesame accessions from Africa and Asia as major Centers of its cultivation. Genes. 7(4):14.
- Gespstein S, Glick BR (2013) Strategies to ameliorate abiotic stress-induced plant senescence. Plant Mol Biol. 82:623-633.
- Hildebrand CE, Torney DC, Wagner RP (1992) Informativeness of polymorphic DNA markers. Los Alamos Sci. 20:100-102.
- IPGRI, NBPGR (2004) Descriptors for sesame (*Sesamum* spp.). International Plant Genetic Resources Institute, Rome, Italy; and National Bureau of Plant Genetic Resources, New Delhi, India 76p.
- Kaur S, OI CN, Forster JW, Paull JG (2014) Assessment of genetic diversity in faba bean based on single nucleotide polymorphism. Diversity. 6:88-101.
- Luo H, Wang X, Zhan G, Wei G, Zhou X, Zhao J, Huang L, Kang Z (2015) Genome-wide analysis of simple sequence repeats and efficient development of polymorphic SSR markers based on whole genome re-sequencing of multiple isolates of the wheat stripe rust fungus. PLoS One. 10:1-18.
- Manivannan N, Nadarajan N (1996) Genetic divergence in sesame. Madras Agric J. 83:789-790.
- McDonald MB, Elliot LJ, Sweeney PM (1994) DNA extraction dry seeds for RAPD analyses in varietal identification studies. Seed Sci Technol. 22:171-176.
- Ribeiro ICNS (2011) Análise da divergência genética em acessos de *Mangifera indica* com base em descritores agro-morfológicos e marcadores microssatélites. 114 p. Dissertation, Universidade Estadual de Feira de Santana, Feira de Santana-BA.
- Sant'ana GC, Ferreira JL, Rocha HS, Borém A, Pasqual M, Cançado GMA (2012) Comparison of a retrotransposon-based marker with microsatellite markers for discriminating accessions of *Vitis vinifera*. Genet Mol Res. 11:1507-1525.
- Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 10:967-981.
- Val BHP, Ferreira Júnior JA, Bizari EH, Di Mauro AO, Trevisoli SHU (2014) Diversidade genética de genótipos de soja por meio de caracteres agromorfológicos. Ciênc Tecnol Fatec-JB. 6(1):72-83.
- Varshney RK, Thiel T, Sretenovic-Rajcic T, Baum M, Valkoun J, Guo P, Grando S, Ceccarelli S, Graner A (2009) Identification and validation of a core set of informative genetic SSR and SNP markers for assaying functional diversity in barley. Mol Breed. 22:1-13.
- Vieira EA, Fialho JF, Faleiro FG, Bellon G, Fonseca KG, Silva MS, Paula-Moraes SV, Carvalho LJC (2013) Caracterização fenotípica e molecular de acessos de mandioca de indústria com potencial de adaptação às condições do

- Cerrado do Brasil Central. *Semina: Ciência Agrár.* 34:567-582.
- Wang L, Han X, Zhang Y, Li D, Wei X, Ding X, Zhang X (2014) Deep resequencing reveals allelic variation in *Sesamum indicum*. *BMC Plant Biol.* 14:1-225.
- Wei W, Qi X, Wang L, Zhang Y, Hua W, Li D, Lv H, Zhang X (2011) Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genom.* 19:12-451.
- Zhang Y, Zhang X, Che Z, Wang, L, Wei W, Li D (2012a) Genetic diversity assessment of sesame core collection in China by phenotype and molecular markers and extraction of a mini-core collection. *BMC Genet.* 13:102.
- Zhang H, Wei L, Miao H, Zhang T, Wang C (2012b) Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genom.* 13-316.