Australian Journal of Crop Science AJCS
AJCS 9(11):1089-1094 (2015) ISSN:1835-2707

Clustering fastigiata peanut accessions for selection of early-mature types suitable for the food market

Jean Pierre Cordeiro Ramos[1], Lucas Nunes da Luz[2], José Jaime Vasconcelos Cavalcanti[3], Liziane Maria de Lima[3], Rosa Maria Mendes Freire[4], Péricles de Albuquerque Melo Filho[5], Roseane Cavalcanti dos Santos[3*]

[1]Department of Postgraduate in Agricultural Sciences, State University of Paraiba, Rua Baraúnas, 351, Bairro Universitário, CEP 58429-500 Campina Grande-PB, Brazil
[2]Department of Agronomy, Federal University of Cariri, Rua Vereador Sebastião Maciel Lopes, s/n, Muriti, CEP: 63133-610, Crato-CE, Brazil
[3]Biotechnology Laboratory, Embrapa Cotton, Rua Oswaldo Cruz, nº 1143, Centenário, CEP 58428-095, Campina Grande-PB, Brazil
[4]Laboratory of Soil Science and Plant Nutrition, Embrapa Cotton, Rua Oswaldo Cruz, nº 1143, Centenário, CEP 58428-095, Campina Grande-PB, Brazil
[5]Department of Agronomy, Rural Federal University of Pernambuco, Rua Dom Manoel de Medeiros, s/n, Dois Irmãos, CEP: 52171-900, Recife-PE, Brazil

***Corresponding author: roseane.santos@embrapa.br**

Abstract

Germplasm collections are the main raw material for breeding programs. Peanut is an important oilseed that serves to oil and food markets. The identification of early-mature genotypes and earliness adds benefits to crop production as it minimizes the costs of management and the losses of pod production due to occurrence of dry periods. In this paper, we estimate the genetic divergence in 77 peanut accessions (subsp. f*astigiata*), using three clustering methods, in order to identify early-mature genotypes and earliness for food market. The accessions were grown in field and phenotyped for 19 qualitative and quantitative traits. The Tocher methodology was employed for discriminating Valencia accessions. Both UPGMA and Principal Components Analysis (PCA) were used to identify and cluster genotypes. The results identified at least, two groups including high yield-earliness (*fastigiata*) and drought tolerance (*vulgaris*) accessions, promising for food market. These methodologies provided interesting combinations for further using in breeding program aiming to generate early-mature lines for food market with adaptability to semi-arid environments.

Keywords: *Arachis hypogaea*, genetic divergence, Tocher, UPGMA, Principal Components.
Abbreviation: PCA_Principal Component Analysis; UPGMA_Unweighted Pair Group Method with Arithmetic Mean.

Introduction

Germoplasm collections are the input for any genetic improvement program. In autogamous species, the natural variability is often limited due to broad phenotype similarities among genotypes, especially those with cleistogamic flowers. Consequently, the genetic gains in selection procedures are not expressive. The knowledge of genetic variation among and within populations is key to planning improvement strategies, which may meet the several demands of a program (Granja et al., 2009; Santos et al., 2013). The use of multivariate methods provides broad contribution to classification and identification of genotypes, which may be potentially useful to genetic improvement proposals. The dissimilarity analysis, obtained by Euclidean distance, distance of Mahalanobis, Principal Components (PC), and canonical variables are widely reliable and employed by plant breeders (Pereira et al., 1992; Santos et al., 2000a; Cruz et al., 2011, 2012).

Peanut (*Arachis hypogaea* L.) is a tetraploid species that reproduces through autogamy. The specie is benefited by cleistogamic flowers, which ensure a high uniformity to commercial cultivars (Santos et al., 2013). The species is subdivided into two subspecies, *fastigiata*, with accessions that belong to Valencia (variety *fastigiata*) and Spanish (variety *vulgaris*) groups; and *hypogaea*, with accessions belonging to Virginia (variety *hypogaea*) group. All of them have phenotypical peculiarities, which contribute to distinction of accessions (Valls, 2013).

The *fastigiata* subspecies is characterized by upright and short cycle accessions with flowers on main axis. Several commercial cultivars belong to this subspecies. They are valorous genetic resources to genetic improvement focused on environments with irregular or short rainy season (Santos et al., 2013; Melo Filho and Santos, 2010). The *fastigiata* and *vulgaris* varieties have several similar traits whose distinction of accessions based on minimal classes of descriptors are not recommended, because it does not favors reliably the selection procedures. The use of multivariate methods have contributed widely to separate autogamous genotypes and identify divergent groups that may be further used in hybridization procedures, in order to broaden the genetic basis of commercial cultivars.

The Brazilian Company of Agricultural Research (EMBRAPA) coordinates a robust program of peanut improvement focused on semiarid environment. This region

is characterized by soils of low fertility, erratic rainfall and moderate veranicos that often occur in rainy season. Periodically, several intra and interspecific lines are generated by crossings for further use in selection procedures in order to identify high yield and drought tolerant materials, recommended to food market (Santos et al., 2013).

In this study, we will estimate the genetic divergence of peanut accessions (subsp. *fastigiata*) based on different clustering methods, in order to select genotypes for further use in breeding program to food market, focusing on semiarid environment.

Results and Discussion

Clustering of accessions

The analysis of genetic divergence in 77 accessions through Tocher method resulted in five groups, in which two of them had only two accessions (groups 4 and 5), more phenotypically isolated from others (Table 1). The Group 1 clustered 52% of all accessions, with 40 genotypes belonging to Valencia type (*fastigiata*). The Group 2 clustered 28 accessions belonging to *fastigiata* and *vulgaris* var., including genotypes with high tolerance to drought, inherited genetically from 55 437 (43), an African cultivar with wide adaptation to semiarid environment. The accessions derived from 55 437, such as 76AMPoitara (48), 76AM51AM (49), L7 bege (50) and BRS 151 L7 (51) show earliness and drought tolerance (Santos et al., 2010). The Group 3 clustered only five accessions, all Spanish-earliness types, generated by *International Crops Research Institute for the Semi-Arid-Tropics* **(**ICRISAT), with broad tolerance to drought (Santos et al., 2000b; Santos et al., 2010; Pereira et al., 2012).

The D^2 analysis was carried out using all the nineteen traits and generalized distance (D^2) was calculated for each pair of genotypes. The relative contribution of traits to genetic divergence is found in Table 2. The traits with the highest loading, based on D^2, were those directly involved with production, such as pod yield, seed yield and oil content, agreeing with others reports found in literature (Santos et al., 2000a; Ajay et al., 2012). Pod width showed low contribution to genetic divergence based on data collected from 77 accessions. The clustering analysis by UPGMA showed high sensibility to differentiation of accessions, revealing interesting arrangements with possibility for further genetic gains during the selection procedures (Fig 1). The Groups 1 and 4 clustered low oil-accessions, developed by ICRISAT and EMBRAPA, respectively, most of them with high yield and earliness, such as BRS Havana (42), BRS 151 L7 (51), and BR 1 (39), all short cycles (90 d) and indicated to semiarid environment. Gomes et al. (2007) estimated the yield stability and adaptation of several peanut genotypes grown in semiarid and non-semiarid environments and found that these genotypes showed broad adaptation and stability in all environment studied**.** They may represent valuable genetic resources for use in breeding programs aiming tolerance to dry conditions**.** The Groups 2 grouped only three accessions from EUA and indicated to tropical climates.

The Group 3 clustered 53% of whole accessions, including genotypes from *fastigiata* and *vulgaris* var. Five subgroups were established based on peculiarity of accessions: (a) subgroup 3.1: Spanish types with small-tan seeds and low oil; (b) subgroup 3.2: earliness and short cycle accessions (below 90 d), with large seeds, assigned to food market. Two valorous genetic resources are in this group: 283 AM (63), a high yield, obtained by crossing between Manfredi 407 and Florunner, and a drought tolerant 76AM51AM (49), obtained by crossing between 55 437 and IAC Oirã, an African and a Brazilian cultivars, respectively. Santos et al. (2010) evaluated these genotypes in semiarid and tropical climates located at Brazilian Northeast during three years in rainy season and found that both genotypes were more responsive in semiarid than in tropical environments. The authors report that the adequate water availability associated with high temperatures benefited the performance of these genotypes: (c) subgroup 3.3: contained only two Spanish accessions, both from India and very similar in all traits, which must probably be related; (d) subgroup 3.4: showed the same pattern seen in Group 5, ie, earliness-Valencia types, with large seeds and suitable to food market; and (e) subgroup 3.5: intermediate to late cycle accessions adapted to tropical climates, such as cv. Botutatu, a upright and mid-cycle genotype, developed to environment with rainfall up to 1000 mm (Zanotto, 1993). Finally, Group 5 clustered only red seeds-Valencia accessions, with adequate pattern for food market.

The settings of UPGMA-clusters presented here with 77 *fastigiata* accessions allowed selecting parents to attend a peanut breeding program in several segments, focusing on semiarid environments or even in tropical climates. Based on composition of groups, UPGMA was more responsive to genetic divergence than Tocher method.

Graphic dispersal of accessions by PCA

The PCA was performed to identify the major components that could explain the total variations observed from 19 traits. The adjustment of the model was not able to account for total variation in the first three main components, which mostly occupied 54% (data not shown). The analysis was again performed adopting only quantitative traits (Pod yield, Seed yield, Number of pods per plant, Main axis height, Pod lenght, Oil content, Harvest index, Blooming and Full pod maturation), and the two first PC accounted for 70.41% of the total variation observed among 77 *fastigiata* peanut genotypes (Table 3). This percentage attends the recommendations outlined in Rencher (2002) and Härdle and Simar (2003). The dispersion of points corresponding to 77 accessions is found in Fig 2. Seven groups were clustered, whose composition of most accessions showed coherence with those found in Tocher (Table 1) and UPGMA (Fig 1) methods. Genotypes clustered in Group I, all earliness and drought tolerant genotypes, matched almost 100% in all three methodologies. They might be employed as robust parents in breeding program for semiarid environments.

Taking into account, the improvement to food market, lines clustered in Group IV were more promising, which showed coherence with those showed in Groups 4 and sub-group 3.2, (Fig 1). The best materials in these groups are cvs. BRS Havana (42), BRS 151 L7 (51), and BR 1 (39), all have earliness developed by EMBRAPA, and IAC 8112, developed by Agronomic Institute of Campinas, Brazil, and indicated to tropical environment (Santos et al., 2013).

Several authors have adopted clustering methodologies to estimate the genetic similarity in vegetal species. The efficiency of the method is dependent on reproductive system, discriminant traits and level of genetic variability of

Table 1. Clustering of 77 peanut genotypes based on Tocher method.

Group	Genotype
1	23 25 17 24 20 26 27 61 47 73 9 16 44 29 66 45 28 76 1 60 11 34 22 74 64 21 77 30 39 31 69 37 42 35 75 59 65 6 36 41
2	19 49 5 63 15 50 51 12 10 13 7 32 48 67 52 55 53 62 68 54 72 58 38 70 71 33 2 14
3	40 43 46 56 57
4	4 8
5	3 18

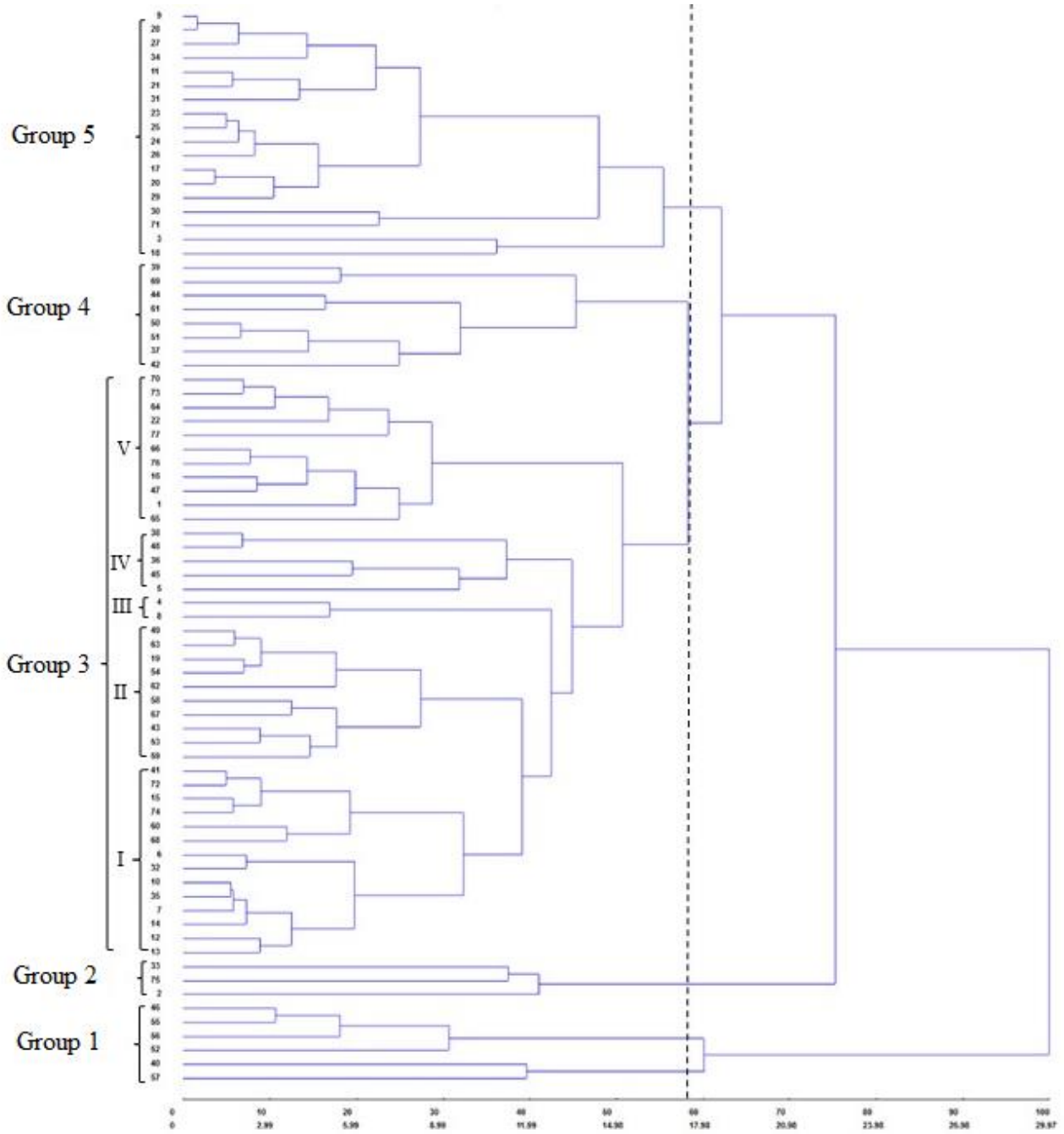

Fig 1. Dendrogram illustrating genetic relationship among 77 fastigiata peanut genotypes generated by UPGMA cluster. Cophenetic correlation coefficient 0.829. Dotted line is a screening adopted to the genetic similarity index of 60%.

Table 2. Relative contribution of traits to genetic divergence of 77 *fastigiata* peanut genotypes.

Trait	Sj	Sj (%)	Ranks
Oil content	277.677	15.56	3
Main axis height	155.668	8.72	5
Blooming	129.418	7.25	8
Harvest index	164.886	9.23	4
Seed yield	301.770	16.91	2
Pod/plant number	140.965	7.90	6
Pod yield	410.976	23.02	1
Pod width	63.98	3.58	9
Pod maturation	139.442	7.81	7

Sj- relative importance of traits to genetic divergence (Singh, 1981).

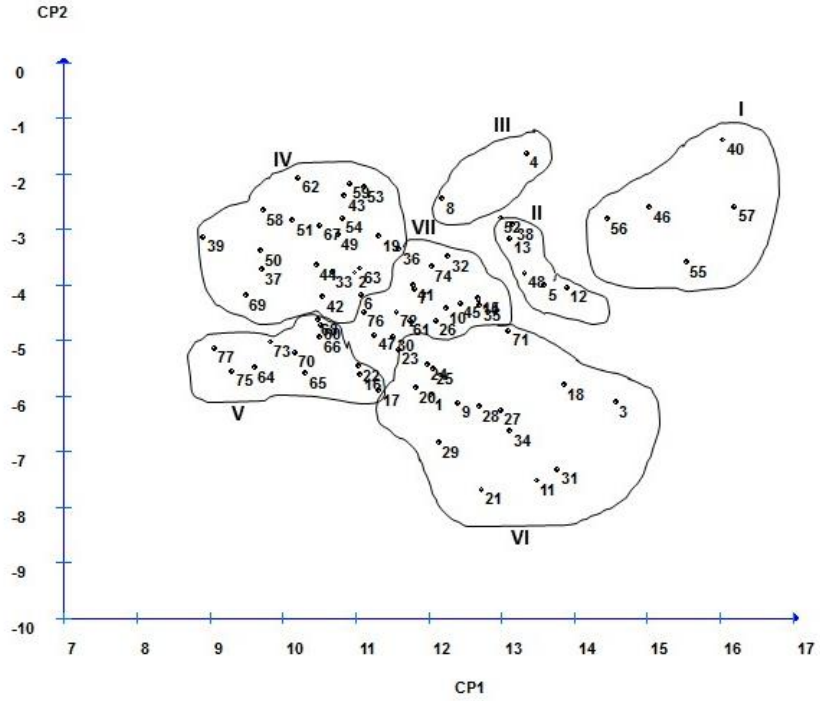

Fig 2. Two-dimension principal component analysis (PCA) established by linear combination of quantitative traits from 77 *fastigiata* preanut genotypes. Numbers of genotypes are listed in Supplementary Table 1.

Table 3. Estimates of eigenvalues, individual and cumulative variance explained by Principal components based on 19 traits of *fastigiata* peanut genotypes.

Components	Individual value	Variance (%)	CPV
CP1	6.55	38.42	38.42
CP2	4.98	31.99	70.41
CP3	2.19	16.28	86.70
CP4	1.05	10.71	95.41
CP5	0.88	3.79	99.20
CP6	0.66	0.70	99.90
CP7	0.33	0.19	100.00
CP8	0.32	0.60	100.00
CP9	0.01	0.20	100.00

PC: Principal componente; CPV- Cumulative percentage of variance.

population studied. In this study, the genetic divergence of *fastigiata*-peanut bred-lines was estimated based on Tocher, UPGMA and PC methods. Based on our results, the UPGMA and PC methods showed coherence in results of *fastigiata* and *vulgaris* var. clustering, and; therefore, were more contributive to assist the selection procedures of *fastigiata* subsp. Esquivel et al. (1993) adopted the PCA to estimate the genetic divergence in geographically closed *fastigiata* and *hypogaea* subsp., based on agronomic, biochemical, morphological, and phytopathological traits, and did not find results to discriminate the intraspecific accessions, even using agronomic traits that are more robust for characterization of germoplasm. Mehndiratta and Phul (1970) stated that when population is originated from the same geographical region, or when it undergoes selection pressures to fix common traits, the tendency to detect diversity among accessions is smaller.

Ajay et al. (2012) used UPGMA and PC methods in order to identify parents with complementary traits for further use in food-breeding program. According to the authors, both

multivariate methods were very adequate and represent reliable tools to assist the breeders in improvement programs.

Materials and Methods

Genetic resources and conduction of the experiment

Seventy-seven peanut bred-lines (subsp. *fastigiata*) were used in this study (Supplementary Table 1). The experiment was carried out at experimental field of EMBRAPA, in Campina Grande, PB, Brazil (7º13'50"S, 35º52'52"W, 551 m), in rainy seasons (May-August, 2014). The total volume of rainfall during growing season was 536 mm. The soil was classified as Vertisoil, previously limed and fertilized (NPK, 20:60:30, ammonium sulfate, single superphosphate and potassium chloride).

Each genotype was sown in three rows (5 m length), spaced in 70 cm each. The population density was 10 plants/meter. Four seeds were sown per hill and after emergence. They were thinned to only two seedlings/hole. A randomized complete block design was adopted with five replications. The crop was grown by adopting recommended package of practices, described in Santos et al. (2006). Harvest took place between 87 and 115 d after emergence, when the pods reached 70% of maturity (Santos et al., 2013).

Nineteen traits were collected from each genotype located at central rows. They were: Pod yield, Seed yield, Number of pods per plant, Number of seeds per pod, 100 pods-weight, 100 seeds-weight, Oil content, Pod length, Harvest index, Hairiness, Color of main stem, Growth habit, Main axis height, Seed color, Leaf color, Seed size, Emergence, Blooming and Full pod maturation.

Analysis of genetic divergence and clustering techniques

The genetic distances of accessions were estimated by using Gower's algorithm (Gower, 1971). The clustering analysis were carried out through Tocher's optimization (Rao, 1952) based on generalized Mahalanobis D^2 statistics, and UPGMA (*Unweighted Pair Group Method with Arithmetic Mean*) methods. The UPGMA based on genetic distances has been considered an efficient estimator of phylogenetic linkages (Nei et al., 1983). In order to eliminate the non-hierarchical effects, the cophenetic correlation coefficient was estimated (Sneath and Sokal, 1973), which indicates the distortion produced by clustering in the original genetic distances. This coefficient is a matrix correlation between original genetic distances and a new distance matrix (the cophenetic matrix), derived directly from the UPGMA dendrogram.

The PCs were estimated through transformation of original data into a set with equivalent dimension of uncorrelated data (Cruz et al., 2012). The first PC often explains the maximal amount of variance in the data set and its direction. The variance along the vector is known as the eigenvalue, which was used for the determination of variances of the major PCs. The scores corresponding to the PCs were calculated from the correlation matrix. The first two PC scores were used to group the genotypes in dispersion graphic. Cluster analysis was performed using the software GENES, version 2013.5.1 (Cruz, 2013).

Conclusion

The genetic divergence of seventy-seven *fastigiata* peanut accessions was estimated using clustering analysis. The UPGMA and PC methods significantly assisted to the selection procedures of *fastigiata* subsp. The clustering revealed interesting combinations for further use in breeding program, focusing on food market demand in semiarid environments.

Acknowledgments

To acknowledge the National Network of Agricultural Biodiversity and Sustainability (*Rede Nacional de Pesquisa em Agrobiodiversidade e Sustentabilidade Agropecuária* - REPENSA [National Network for Research on Agricultural Biodiversity and Sustainability] (MCT/ CNPq/ MEC/ CAPES/ CT AGRO/ CT IDRO/ FAPS/ EMBRAPA), for the financial support and grants.

References

Ajay BC, Gowda MVC, Rathnakumar AL, Kusuma VP, Fiyaz RA, Holajjer P, Ramya KT, Govindaraj G, Babu HP (2012) Improving genetic attributes of confectionary traits in peanut (*Arachis hypogaea L.*) using multivariate analytical tools. J Agric Sci. 4:247-258.

Cruz CD (2013) GENES: A software package for analysis in experimental statistics and quantitative genetics. Acta Sci Agron. 35:271-276.

Cruz CD, Regazzi AJ, Carneiro PCS (2012) Modelos biométricos aplicados ao melhoramento genético. Editora UFV, Viçosa, p 514.

Cruz CD, Ferreira FM, Pessoni LA (2011) Biometria aplicada ao estudo da diversidade genética. Editora UFV, Viçosa, p 620.

Esquivel M, Fundora Z, Hammer K (1993) Peanut (*Arachis hypogaea* L.) genetic resources in Cuba. II. Preliminary germoplasm evaluation. Plant Genet Resour Newsl. 91/92: 17-20.

Gomes LR, Santos RC, Anunciação Filho CJ, Melo Filho PA (2007) Adaptabilidade e estabilidade fenotípica de genótipos de amendoim de porte ereto. Pesqui Agropecu Brasil. 42:985-989.

Gower JC (1971) A general coefficient of similarity and some of its properties. Biometrics. 27:857-874.

Granja MMC, Melo Filho PA, Santos RC (2009) Análise genética em uma população intraespecífica de amendoim baseada em descritores fenotípicos. Rev Bras Ciênc Agr. 4:257-260.

Härdle W, Simar L (2003) Applied multivariate statistical analysis. MD Tech, Berlin, p 488.

Mehndiratta PD, Phul PS (1970) Genetic diversity in relation to fodder field and its components in Shorgum. Indian J Genet Plant Breed. 31:300-306.

Melo Filho PA, Santos RC (2010) A cultura do amendoim no Nordeste: situação atual e perspectivas. An Acad Pernamb Ciênc Agron. 7:192-208.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol. 19:153-170.

Pereira JWLP, Melo Filho PA, Albuquerque MB, Nogueira RJMC, Santos RC (2012) Mudanças bioquímicas em genótipos de amendoim submetidos a déficit hídrico moderado. Rev Ciênc Agron. 43:766-773.

Pereira AV, Vencovsky R, Cruz CD (1992) Selection of botanical and agronomical descriptors for the characterization of cassava (*Manihot esculenta* Crantz) germoplasm. Rev Bras Genét. 15:115-124.

Rao CR (1952) Advanced statistical methods in biometric research. New York: John Wiley & Sons. p 390.

Rencher AC (2002) Methods of multivariate analysis. Wiley-Interscience, New York, p 740.

Santos RC, Moreira JAN, Farias RH, Duarte JM (2000) Classificação de genótipos de amendoim baseada nos descritores agromorfológicos e isoenzimáticos. Ciênc Rural. 30:55-59.

Santos RC, Moreira JAN, Duarte JM (2000a) Variabilidade isoenzimática entre linhagens de amendoim resistentes à seca. Ciênc Rural. 30:269-274.

Santos RC, Rêgo MG, Silva APG, Vasconcelos JOL, Coutinho JLB, Melo Filho PA (2010) Produtividade de linhagens avançadas de amendoim em condições de sequeiro no Nordeste brasileiro. Rev Bras Eng Agríc Ambient. 14:589–593.

Santos RC, Godoy IJ, Fávero AP (2013) Melhoramento do amendoim e cultivares comerciais. In: Santos RC, Freire RMM, Lima LM (eds.). O agronegócio do amendoim no Brasil. Editora Embrapa, Campina Grande

Santos RC, Freire RMM, Suassuna TMF, Rego GM (2006) BRS Havana: nova cultivar de amendoim de pele clara. Pesq Agropec Bras. 41:1337-1339

Sneath PH, Sokal RR (1973) Numerical taxonomy: The principles and practice of numerical classification. San Francisco: W.H. Freeman, p 573

Valls JFM (2013) Recursos genéticos do gênero *Arachis*. In: Santos RC, Freire RMM, Lima LM (eds.). O agronegócio do amendoim no Brasil. Embrapa, Campina Grande

Zanotto MD (1993) Botutatu: novo cultivar de amendoim (*Arachis hypogaea* L.). Pesq Agropec Bras. 9:1101-1102.