CBAB

CROP BREEDING AND
APPLIED BIOTECHNOLOGY

Comparison of similarity coefficients in sesame cultivars clustering using RAPD markers

Nair Helena Castro Arriel[1*], Antônio Orlando Di Mauro[1], Cosme Damião Cruz[3], Sônia Marli Zingaretti Di Mauro[2], Marcelo Marchi Costa[2], Sandra Helena Unêda-Trevisoli[2], and Andréa Capeloto[2]

ABSTRACT - *Thirteen similarity coefficients were evaluated by the UPGMA method for clustering of nine sesame cultivars based on 613 RAPD markers. Distance matrices were estimated from the arithmetic complement of the Simple Matching, Jaccard, Kulczynski2, IndexIII, Ochiai, OchiaiII, Rogers and Tanimoto, Russel and Rao, Sorensen-Nei and Li, Baroni-Urbani and Buser, Sokal and Sneath, Haman and Phi coefficients. They all identified the same pairs of most divergent and most similar cultivars (2 and 7, 3 and 4, respectively). Except for the Russel and Rao's, the UPGMA dendrograms formed three identical groups. Based on the nature of the RAPD markers and the mathematical properties of Jaccard, Kulczynski2, Ochiai and Sorensen-Nei and Li coefficients, it is concluded that these coefficients are the most suitable for clustering sesame cultivars.*

Key words: *Sesamum indicum* L., genetic diversity, cluster analysis.

INTRODUCTION

Knowledge on a species' genetic variability is crucial to establish a sound program of genetic improvement. In a number of cultivars of commercially important crop plants grown in Brazil, little genetic diversity is found because the primary gene pool available for breeding is formed by few commercial cultivars and local varieties (Destro and Montalván 1999). This limited genetic base reduces the expected gains from any genetic improvement efforts. Thus, it is necessary to search for genetic variability and methods that maximize genotype discrimination and population diversity quantification.

Genetic diversity has basically been evaluated by quantitative techniques (diallels) or predictive processes. Predictive methods have been used thoroughly because no previous hybrid combinations, used in diallelic analysis, are necessary (Cruz and Carneiro 2003). This divergence can be evaluated dy agronomic, morphologic, and, more recently, molecular marker characteristics. Molecular markers provide an unlimited amount of data at any atmospheric condition and stage of plant development, making them a powerful tool in genetic studies of wild and domesticated populations. Among the different classes of molecular markers, the randomly amplified polymorphic DNA (RAPD) markers are of the most broadly used due to their easy and quick data collection process, low cost, and accessibility (Ferreira and Grattapaglia 1996).

[1]Embrapa Algodão; Departamento de Produção Vegetal, Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias (UNESP/FCAVJ), Via de Acesso Prof. Paulo Donato Castellane, s/n, 14884-900, Jaboticabal, SP, Brasil. *E-mail: narriel@fcav.unesp.br
[2]Departamento de Produção Vegetal, UNESP/FCAVJ
[3]Departamento de Biologia Geral, BIOAGRO, Universidade Federal de Viçosa, 36571-000, Viçosa, MG, Brasil

Different statistical methods can be applied to molecular marker data, such as cluster, factor, discriminant, and principal component and coordinate analyses. Cluster analysis demands no initial presupposition of data distribution and its results are easy to interpret. Clustering is based on a similarity coefficient to measure how similar the elements are (Bussab et al. 1990).

There are several coefficients proposed in literature related to cluster analysis. However, most of the published papers do not justify why a certain coefficient was employed, and different coefficients were used with the same purpose. Johns et al. (1997) used the Simple Matching coefficient in the classification of the gene pool of local bean cultivars in Chile; Sanz-Cortés et al. (2001) utilized the Jaccard and Nei coefficients to characterize 40 olive cultivars; and Magalhães et al. (2001) made use of the Sorensen-Nei and Li coefficient to classify tropical tree species.

Dias (1998) highlights that the choice of a coefficient should be based on research objectives, the variable type under analysis, and the coefficient properties. In genetic divergence studies, the determination of the degree of genetic similarity depends on the way the original data matrix is employed, as most of the coefficients are computed from 1-0 binary variables, representing, respectively, presence and absence of marks (bands). Thus, in the comparison of two individuals, there are four combinations of concordance and discordance of marks: a = 1-1; b = 1-0; c = 0-1 and d = 0-0. As each coefficient assigns different weights to (a), (b), (c) and (d) values, this can lead to different genotype groups.

Comparative studies related to the efficiency of similarity coefficients using RAPD molecular markers display different conclusions. Based on eight coefficients, Duarte et al. (1999) observed little alteration in clustering bean cultivars, Rocha (2002) spotted no alteration in the classification of potato genotypes, and Meyer (2002) verified that, except when using the Russel and Rao coefficient, the group discrimination of maize lineages using RAPD and AFLP markers was practically identical due to similar mathematical principles of the coefficients.

As the results of clustering methods can be influenced by the choice of the similarity coefficient, it is important to run comparative studies, in order to use the most appropriate measure. Objective of this work was to compare 13 similarity coefficients in the process of dendrogram generation and clustering of sesame cultivars, based on RAPD markers, in order to ultimately identify the most appropriate ones for the study of genetic divergence of these genotypes.

MATERIAL AND METHODS

Nine sesame cultivars (1-Guatemala, 2-Seridó 1, 3-Nicaragua, 4-Venezuela, 5-Paquistão, 6-Mexicana, 7-CNPA G2, 8-CNPA G3, and 9-CNPA G4) of different maturation cycle, height, seed color, and number of branches and fruits per leaf axilla were used.

Amplification reactions for RAPD marker were carried out according to Arriel et al. (2002). Marks of 55 Operon Technologies primers generated a binary matrix of 1's and 0's, respectively for presence and absence of the band. Pairwise genetic distance estimates were obtained from the arithmetic complement of the 13 similarity coefficients (Table 1).

Pearson and Spearman correlations were used to determine the level of linearity between similarity coefficients. Distance matrices and correlation values were calculated using software Genes (Cruz 2001). Clustering was based on the Unweighted Pair-Group Method using arithmetic Averages (UPGMA), run on software NTSYS-pc (Rohlf 1992).

The adequacy of the hierarchical grouping method was established by the coefficient of cophenetic correlation (CCC) that evaluates the correlation between the elements of the original and cophenetic matrices, obtained after dendrogram drawing.

The dendrograms were compared by the Consensus Index (CI) that estimates the level of similarity between two dendrograms. This index is based on the number of steps in common observed in the construction of a dendrogram. Its value is obtained by the division of the number of shared steps, excluding the last one, in which all cultivars are put together in a single group by the possible maximum number of grouping steps. This divisor equals n-2 (n=number of genotypes) for fully constructed dendrograms. CI values range from 0 (not a single consensus) to 1 (identical dendrograms). The coefficients of cophenetic correlation and consensus indexes were obtained by software NTSYS-pc (Rohlf 1992).

Six hundred and thirteen marks were bootstrapped with software Dboot (Coelho 2000). Dendrograms and errors associated to the genetic dissimilarity obtained from Jaccard, Sorensen-Dice, and Simple Matching coefficients were obtained based on 1000 permutations.

RESULTS AND DISCUSSION

The 55 primers used in RAPD reactions generated 613 amplification bands. Júnior Silva and Duarte (2003) emphasize that 91 primers are necessary to cover, with a probability of 90%, the genome of 18-to-24-chromosome species, which is fairly close to the 26 chromosomes present in sesame plants. However, the number of generated bands

Table 1. Similarity coefficients and respective estimation formulas and ranges used to estimate the pairwise genetic distance of nine sesame cultivars, based on RAPD markers

Similarity coefficients	Formulas	Ranges
Simple Matching	$a+d/a+b+c+d$	[0 to 1]
Jaccard	$a/a+b+c$	[0 to 1]
Kulczynski2	$1/2[(a/a+b)+(a/a+c)]$	[0 to 1]
IndexIII	$1/4[(a/a+b)+(a/a+c)+(d/d+b)+(d/d+c)]$	[0 to 1]
Ochiai	$a/\sqrt{(a+b)(a+c)}$	[0 to 1]
OchiaiII	$ad/\sqrt{(a+b)(a+c)(b+d)(c+d)}$	[0 to 1]
Roger and Tanimoto	$a+d/a+2(b+c)+d$	[0 to 1]
Russel and Rao	$a/a+b+c+d$	[0 to 1]
Sorensen-Nei and Li	$2a/2a+b+c$	[0 to 1]
Baroni-Urbani and Buser	$a+\sqrt{(ad)}/a+b+c+\sqrt{(ad)}$	[0 to 1]
Sokal and Sneath	$2(a+d)/2(a+d)+b+c$	[0 to 1]
Haman	$ad-bc/a+b+c+d$	[-1 to 1]
Phi	$ad-bc/\sqrt{(a+b)(a+c)(b+d)(c+d)}$	[-1 to 1]

a: (11) matching-type number; b: (10) matching type-number; c: (01) matching-type number; d: (0 0) matching-type number, for each pair of cultivars

is adequate as it is higher than 100. The bootstrapping technique demonstrated that this is a minimum band number that practically stabilizes the coefficient of variation of the genetic distance (CV_{GD}) between genotypes (Nienhuis et al. 1995). Benchimol et al. (2001) warn that, for a sound diversity characterization by RAPD markers, at least 400 loci or marks should be evaluated. In the present work, bootstrapping identified a CV_{GD} stabilization beginning at 215 bands. Also, random samples of the 613 bands generated the same grouping obtained with 1000 permutations, using Jaccard, Sorensen-Dice or Simple Matching coefficients.

Table 2 shows descriptive statistics of the distance matrices according to the coefficients of similarity. On average, the maximum and minimum distances were generated by the Phi (0.5772) and Sokal and Sneath (0.1669) coefficients, respectively. The latter presented the smallest standard deviation (0.0131). Based on CV values, higher precision is detected when negative co-occurrences of bands are considered, as is the case for the Russel and Rao (4.07%), OchiaiII (4.71%), Rogers and Tanimoto (4.97%), IndexIII and Phi (5.79%), Simple Matching and Haman (6.57%) coefficients.

Although different weights are assigned to the values of positive and negative co-occurrences, all the studied coefficients recognize the same pairs of most divergent (2 and 7) and most similar (3 and 4) cultivars (Table 2).

The Pearson and Spearman correlations show the existence of a high degree of linearity between the distance measures, in general close to unity (Table 3). Dias (1998) mentions that some coefficients can be considered as monotonic functions of each other, which means that they are proportional. Consequently, the correlation magnitude between most of the coefficients is close to 1. High rp (0.99) and rs (1.00) values are, in most of the cases, those involving Jaccard, Kulczynski2, Ochiai and Sorensen-Nei and Li coefficients. These coefficients do not consider the negative co-occurrences of bands as a factor of similarity. Equal magnitude results were observed for rp by Duarte (1999), and for both rp and rs by Meyer (2002), in most of the assessed similarity coefficients. Correlation values between the above cited and Haman ($rp = 0.96$ and $rs = 0.93$) and Phi ($rp = 0.90$ and $rs = 0.88$) coefficients show slightly depressed values. Note that these two last coefficients range from -1 to 1. For the other coefficients, differences are not so clear due to the criteria used to obtain distance data, and to their range limits, as can be seen by the high correlation between the Simple Matching, Haman, and Phi coefficients.

The lowest correlation estimates ($rp = 0.68$ and $rs = 0.52$) involve the Russel and Rao coefficient, although it can show a higher degree of linear correlation ($rp = 0.94$ and $rs = 0.78$) when it is correlated to Jaccard, Kulczynski2, Ochiai and

Table 2. Summary of the statistical of the arithmetic complement of the similarity coefficients obtained from RAPD marker data

Similarity Coefficients	Average Distance	Standard Deviation	CV[1] (%)	Maximum Distance	Divergent Cultivars	Minimum Distance	Similar Cultivars
Simple Matching	0.2803	0.0184	6.57	0.4263	2 and 7	0.2288	3 and 4
Jaccard	0.3743	0.0249	6.65	0.5660	2 and 7	0.3220	3 and 4
Kulczynski2	0.2368	0.0196	8.34	0.3871	2 and 7	0.1916	3 and 4
IndexIII	0.2886	0.0167	5.79	0.4281	2 and 7	0.2369	3 and 4
Ochiai	0.2375	0.0200	8.45	0.3881	2 and 7	0.1918	3 and 4
OchiaiII	0.4866	0.0229	4.71	0.6758	2 and 7	0.4201	3 and 4
Rogers and Tanimoto	0.4250	0.0211	4.97	0.5976	2 and 7	0.3722	3 and 4
Russel and Rao	0.5039	0.0205	4.07	0.6651	2 and 7	0.5181	3 and 4
Sorensen-Nei and Li	0.2382	0.0204	8.57	0.3891	2 and 7	0.1920	3 and 4
Baroni-Urbani and Buser	0.2604	0.0191	7.34	0.4084	2 and 7	0.2111	3 and 4
Sokal and Sneath	0.1669	0.0131	7.85	0.2709	2 and 7	0.1292	3 and 4
Haman	0.5606	0.0368	6.57	0.8525	2 and 7	0.4576	3 and 4
Phi	0.5772	0.0334	5.79	0.8563	2 and 7	0.4739	3 and 4

[1]CV: Coefficient of variation

Sorensen-Nei and Li coefficients. Duarte (1999) observed similar rs values when correlating Russel and Rao to Jaccard, Ochiai and Sorensen-Nei and Li ($rs = 0.95$) coefficients, and Russel and Rao to Simple Matching and Rogers and Tanimoto (0.87) coefficients.

Higher correlation values (Table 3) between RR and Jaccard group coefficients show that they do not fit with the coefficients as well as they presumably should, as the SM, OcII, RT, BUB, SS, H, and Phi do. These and RR coefficients take the negative co-occurrence of "d" bands in their respective formulas (Table 1) into account, while Jaccard, Kulczynski2, Ochiai and Sorensen-Nei and Li do not.

Figure 1 shows UPGMA-based dendrograms for all but the Russel and Rao similarity coefficients (Figure 2). Their structure was identical, and identified groups A (cultivars 3, 4, 5, 1, and 6), B (cultivars 7, 8, and 9), and C (cultivar 2). However, group A development was similarity coefficient dependent. For those coefficients that do not take the negative co-occurrence of bands into account (Jac, K2, Oc and SNLi), cultivars 3 and 4 are initially grouped together with cultivar 5, and then with cultivar 1 to form Group A, together with cultivar 6. For the other coefficients, cultivars 3 and 4 are initially pooled with cultivar 1, then with cultivar 5 to form Group A, together with cultivar 6.

The Russel and Rao coefficient reproduced the initial formation of the Jaccard-type coefficients, first combining cultivars 3, 4, and 5 (Figure 2). However, the sequence in which the cultivars were grouped was completely different from the others, except for cultivar 2, set alone in Group C in all dendrograms.

Cultivar 2 presents a longer vegetative cycle, higher growth, and richer branching than the other cultivars. This cultivar was originated from massal selection of local types grown in Jardim do Seridó, State of Rio Grande do Norte, while cultivars 3, 4, 5, and 6 were introduced from abroad, and cultivars 1, 7, 8, and 9 are the result of genetic selection from introduced material.

Table 4 shows estimated values of cophenetic correlation coefficients (CCC) that determine the degree of correspondence between the dissimilarity matrices and those resulting from clustering. Higher CCC values indicate less distortion provoked by genotype clustering (Bussab et al. 1990). CCC values are as high as 0.79 (IndexIII and Phi) and 0.90 (Russel and Rao), indicating a good representation of the dissimilarity matrices in the form of dendrograms. With the exception of the Russel and Rao coefficient, those that disregard negative co-occurrence of bands showed the best cluster adjustments.

Another comparison to these results refers to descriptive statistic data derived from the distance values based on the arithmetic complement of the 13 coefficients of similarities considered in the present study (Table 2), where higher precision (smaller CV%) was observed for RR, OcII, RT, InIII, Phi, H, and SS coefficients that take the negative

Table 3. Pearson-*rp* (diagonal below) and Spearman-*rs* correlations (diagonal above) between the distances computed from the complement of the similarity coefficients

Coefficients	SM[1]	Jac[2]	K2[3]	InIII[4]	Oc[5]	OcII[6]	RT[7]	RR[8]	SNLi[9]	BUB[10]	SS[11]	H[12]	Phi[13]
SM		0.93	0.93	0.98	0.93	0.98	1.00	0.62	0.93	0.95	1.00	1.00	0.98
Jac	0.97		1.00	0.88	1.00	0.88	0.93	0.78	1.00	0.98	0.93	0.93	0.88
K2	0.96	0.99		0.88	1.00	0.88	0.93	0.78	1.00	0.98	0.93	0.93	0.88
InIII	0.98	0.91	0.90		0.88	1.00	0.98	0.52	0.88	0.92	0.98	0.98	1.00
Oc	0.96	0.99	0.99	0.90		0.88	0.93	0.78	1.00	0.98	0.93	0.93	0.88
OcII	0.97	0.88	0.87	0.99	0.88		0.98	0.52	0.88	0.92	0.98	0.98	1.00
RT	0.99	0.96	0.96	0.98	0.96	0.97		0.62	0.93	0.95	1.00	1.00	0.98
RR	0.83	0.94	0.94	0.72	0.94	0.68	0.82		0.78	0.75	0.62	0.62	0.52
SNLi	0.96	0.99	0.99	0.90	0.99	0.88	0.96	0.95		0.98	0.93	0.93	0.88
BUB	0.99	0.99	0.98	0.95	0.98	0.93	0.99	0.89	0.98		0.92	0.95	0.95
SS	0.99	0.97	0.96	0.98	0.97	0.97	0.99	0.83	0.96	0.99		1.00	0.98
H	1.00	0.96	0.96	0.98	0.96	0.97	0.99	0.83	0.96	0.99	0.99		0.98
Phi	0.98	0.90	0.90	1.00	0.90	0.99	0.98	0.72	0.90	0.90	0.95	0.98	

[1]Simple Matching, [2]Jaccard, [3]Kulczynski2, [4]IndexIII, [5]Ochiai, [6]OchiaiII, [7]Rogers and Tanimoto, [8]Russel and Rao, [9]Sorensen-Nei and Li, [10]Baroni-Urbani and Buser, [11]Sokal and Sneath, [12]Haman, and [13]Phi
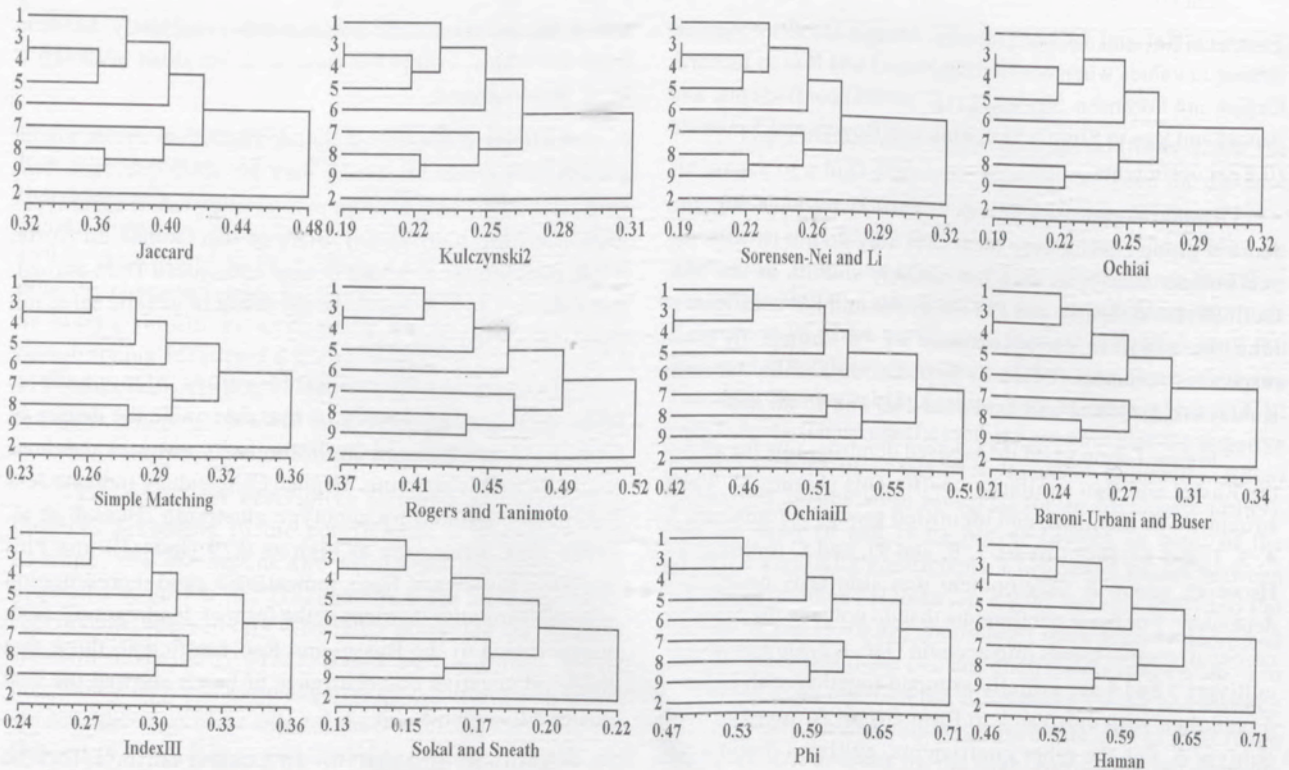

Figure 1. UPGMA dendrograms of nine sesame cultivars (1=Guatemala, 2=Seridó 1, 3=Nicarágua, 4=Venezuela, 5=Paquistão, 6=Mexicana, 7=G2, 8=G3 and 9=G4) constructed from matrices of genetic distance obtained from the complement of the similarity coefficients based on RAPD markers.
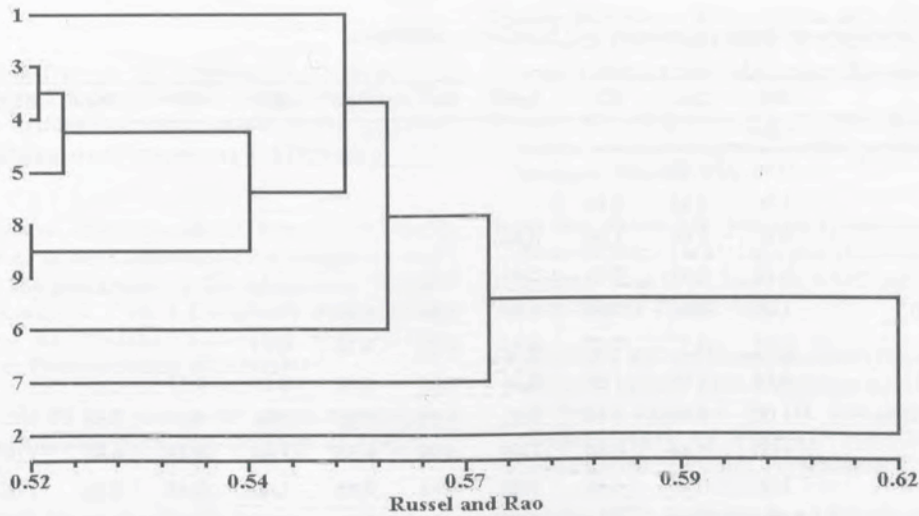
Figure 2. UPGMA dendrogram of nine sesame cultivars (1=Guatemala, 2=Seridó 1, 3=Nicarágua, 4=Venezuela, 5=Paquistão, 6=Mexicana, 7=G2, 8=G3 and 9=G4) constructed from the matrix of genetic distances obtained from the complement of the Russel and Rao similarity coefficient, based on RAPD markers

Table 4. Correlations between cophenetic and dissimilarity matrices for nine sesame cultivars, derived from RAPD-markers data

SM[1]	Jac[2]	K2[3]	InIII[4]	Oc[5]	OcII[6]	RT[7]	RR[8]	SNLi[9]	BUB[10]	SS[11]	H[12]	Phi[13]
0.83	0.86	0.84	0.79	0.85	0.81	0.83	0.90	0.85	0.83	0.81	0.82	0.79

[1]Simple Matching, [2]Jaccard, [3]Kulczynski2, [4]IndexIII, [5]Ochiai, [6]OchiaiII, [7]Rogers and Tanimoto, [8]Russel and Rao, [9]Sorensen-Nei and Li, [10]Baroni-Urbani and Buser, [11]Sokal and Sneath, [12]Haman, and [13]Phi

co-occurrence into account. This means that the distance measures tend to be less dissimilar because one more factor of similarity between individuals is considered. This reduces the variation of the analyzed data in relation to the average, as the coefficients of variation show. However, in the construction of the cluster matrix, the absence of "d" in the numerator decisively improved the degree of adjustment between the original matrices and the ones generated by clustering, as demonstrated by the high cophenetic correlation value (0.90) obtained by the Russel and Rao coefficient.

Consensus Index values (Table 5) show that Jac, K2, Och, and SNLi coefficients presented the same clustering path (CI=1.0). This was also observed for the SM, InIII, RT, BUB, SS, H, and Phi coefficients. CI values between OchiaiII and the group of Simple-Matching-type coefficients and OchiaiII and the group of Jaccard-type coefficients were, respectively, 0.86 and 0.71. These depressed CI values are due to the different clustering path of cultivars 7, 8, and 9.

Sesame cultivars clustered by the Russel and Rao coefficient followed a different path, especially when compared to how OcII, SM, RT, BUB, SS, H, and·Phi coefficients were used, as can be assumed by the CI values

ranging from 0.28 to 0.43 (Table 5). A slightly higher CI value (0.57), observed between the RR and Jaccard-type coefficients, characterized the similar path followed by these coefficients when clustering cultivars 3, 4, and 5. It also confirmed the previously described linear correlations (Table 3). The different RR-clustering path is probably due to the fact that RR coefficient considers the negative co-occurrence in the denominator, but not in the numerator.

Considering the results of the present study, it is noticed that, the thirteen analyzed coefficients did not alter the final structure of the clusters. Thus, coefficients should be chosen based on some data-related criteria, so that an appropriate number of bands is considered to cover the species genome as far as possible since the nature of the RAPD markers does not warrant that similarity is detected under the lack of amplification of a certain band of two genotypes. Coefficients that disregard the negative co-occurrence, such as Jaccard, Sorensen-Nei and Li, Ochiai, and Kulczynski2, are more appropriate for this type of marker.

Weight values attributed to the absence and presence of bands should also be considered when choosing a coefficient. As grouping analysis is an exploratory technique

Table 5. Consensus Index between dendrograms based on 13 similarity coefficients

Coefficients	SM[1]	Jac[2]	K2[3]	InIII[4]	Oc[5]	OcII[6]	RT[7]	RR[8]	SNLi[9]	BUB[10]	SS[11]	H[12]
Jaccard	0.86											
Kulczynski2	0.86	1.00										
IndexIII	1.00	0.86	0.86									
Ochiai	0.86	1.00	1.00	0.86								
OchiaiII	0.86	0.71	0.71	0.86	0.71							
Rogers and Tanimoto	1.00	0.86	0.86	1.00	0.86	0.86						
Russel and Rao	0.43	0.57	0.57	0.43	0.57	0.28	0.43					
Sorensen-Nei and Li	0.86	1.00	1.00	0.86	1.00	0.71	0.86	0.57				
Baroni-Urbani and Buser	1.00	0.86	0.86	1.00	0.86	0.86	1.00	0.43	0.86			
Sokal and Sneath	1.00	0.86	0.86	1.00	0.86	0.86	1.00	0.43	0.86	1.00		
Haman	1.00	0.86	0.86	1.00	0.86	0.86	1.00	0.43	0.86	1.00	1.00	
Phi[13]	1.00	0.86	0.86	1.00	0.86	0.86	1.00	0.43	0.86	1.00	1.00	1.00

[1]Simple Matching, [2]Jaccard, [3]Kulczynski2, [4]Index III, [5]Ochiai, [6]OchiaiII, [7]Rogers and Tanimoto, [8]Russel and Rao, [9]Sorensen-Nei and Li, [10]Baroni-Urbani and Buser, [11]Sokal and Sneath, [12]Haman, and [13]Phi

that seeks to formulate a hypothesis based on a data set, the most common procedure is to attribute the same weight to each variable (Bussab et al. 1990). In addition, population diversity should also be considered. For example, when dealing with top cultivars, as is the case in the present study, the occurrence of common band coincidences is an expected phenomenon. Therefore, the Jaccard coefficient would be the most appropriate, while in studies involving exotic materials or poorly related species, the Sorensen-Nei and Li coefficients would be the most suitable (Dias 1998).

Considering dendrograms and cophenetic correlation results, it can be concluded that Jaccard, Kulczynski2, Ochiai, and Sorensen-Nei and Li; Simple Matching, Rogers and Tanimoto, Sokal and Sneath, OchiaiII, Phi, IndexIII and Baroni-Urbani and Buser coefficients presented almost the same results in all situations of the cluster analysis, while the coefficient Russel and Rao is not recommended for sesame RADP data analysis as it generated discrepant results. Considering the nature of the molecular marker and the common properties of the coefficients, the use of a coefficient that disregards the negative co-occurrence of bands in the numerator of its formula is recommended, as for example Jaccard, Kulczynski2, Ochiai, and Sorensen-Nei and Li, which were equally capable to efficiently characterize the nine sesame cultivars.

Comparação de coeficientes de similaridade no agrupamento de cultivares de gergelim através de marcadores RAPD

RESUMO - *Avaliaram-se treze coeficientes de similaridade pelo método UPGMA no agrupamento de nove cultivares de gergelim, utilizando-se 613 bandas RAPD. As matrizes de distâncias foram obtidas pelo complemento aritmético dos coeficientes de Coincidência Simples, Jaccard, Kulczynski2, ÍndiceIII, Ochiai, OchiaiII, Rogers e Tanimoto, Russel and Rao, Sorensen-Nei e Li, Baroni-Urbani e Buser, Sokal e Sneath, Haman e Phi. Os coeficientes identificaram os mesmos pares de cultivares mais divergentes (2 and 7) e mais similares (3 and 4). Excetuando-se o de Russel e Rao, os dendrogramas formaram três grupos idênticos. Pela natureza do marcador molecular e as propriedades dos coeficientes, os de Jaccard, Kulczynski2, Ochiai e Sorensen-Nei e Li mostraram-se mais adequados ao agrupamento de cultivares de gergelim.*

Palavras chave: *Sesamum indicum* L., diversidade genética, análise de agrupamento.

REFERENCES

Arriel NHC, Unêda-Trevisoli SH, Capeloto A, Di Mauro AO and Di Mauro SMZ (2002) Análise comparativa de quatro protocolos de extração de DNA genômico em gergelim. **Revista Brasileira de Oleaginosas e Fibrosas 6**: 525-535.

Benchimol LL, Barbosa AMM, Geraldi IO, Souza Jr. CL, Garcia AAF and Souza AP (2001) Comparação entre marcadores RFLP, RAPD, AFLP e SSR para estudos de divergência entre linhagens de milho (*Zea mays* L.). In: **I Congresso Brasileiro de Melhoramento de Plantas**. Sociedade Brasileira de Melhoramento de Plantas, Goiânia (CD-ROM).

Bussab WO, Miazaki ES and Andrade DF (1990) **Introdução à análise de agrupamentos**. Associação Brasileira de Estatística, São Paulo, 105p.

Coelho ASG (2000) **Dboot: avaliação de erros associados a estimativas de distâncias/similaridade genéticas através do procedimento de bootstrap com número variável de marcadores (software)**. UFG, Instituto de Ciências Biológicas, Laboratório de Genética Vegetal, Goiânia, 10p.

Cruz CD (2001) **Programa Genes, versão 2003: aplicativo computacional em genética e estatística**. Imprensa Universitária, Viçosa, 648p.

Cruz CD and Carneiro PCS (2003) **Modelos biométricos aplicados ao melhoramento genético**. v 2, Imprensa Universitária, Viçosa, 585p.

Destro D and Montalván R (1999) **Melhoramento genético de plantas**. UEL, Londrina, 820p.

Dias LAS (1998) Análises multidimensionais. In: Alfenas AC (ed.) **Eletroforese de isoenzimas e proteínas afins: fundamentos e aplicações em plantas e microrganismos**. Editora Viçosa, UFV, p. 405-475.

Duarte JM, Santos JB and Melo LC (1999) Comparison of similarity coefficients based on RAPD markers in the common bean. **Genetics and Molecular Biology 22**: 427-432.

Ferreira ME and Grattapaglia D (1996) **Introdução ao uso de marcadores moleculares em análise genética**. 2nd ed., Embrapa-Cenargen, Brasília, 220p.

Johns MA, Skroch PW, Nienhuis J, Kinrichsen P, Bascur G and Muñoz-Schick C (1977) Gene pool classification of common bean landraces from Chile based on RAPD and morphological data. **Crop Science 37**: 605-613.

Júnior e Silva WC and Duarte JB (2003) Determinação do número mínimo de *primers* a serem utilizados por marcadores RAPD em diferentes espécies. In: **II Congresso Brasileiro de Melhoramento de Plantas**. Sociedade Brasileira de Melhoramento de Plantas, Porto Seguro (CD-ROM).

Magalhães MTQ, Suganuma E and Ciampi AY (2001) Análise da variabilidade genética em espécies arbóreas tropicais nativas com o uso de marcador molecular RAPD. In: **I Congresso Brasileiro de Melhoramento de Plantas**. Sociedade Brasileira de Melhoramento de Plantas, Goiânia (CD-ROM).

Meyer AS (2002) **Comparação de coeficientes de similaridade usados em análise de agrupamento com dados de marcadores moleculares dominantes**. MSc. Thesis, ESALQ, USP, Piracicaba, 106p.

Nienhuis J, Tivang J, Sckroch P and Santos JB (1995) Genetic relationships among cultivars and lines of lima bean (*Phaseolus lunatus* L.) as measured by RAPD markers. **Journal of the American Society for Horticultural Science 120**: 300-306.

Rocha BHG, Augustin E, Pereira AS, Fortes GRL and Peters JA (2002) Caracterização de cultivares de batata através de marcadores moleculares. **Revista Científica Rural 1**: 42-51.

Rohlf FJ (1992) **NTSYS-pc: numerical taxonomy and multivariate analysis system**. Exeter Software, New York, 270p.

Sanz-Cortés F, Badenes ML, Paz S, Íñiguez A and Llácer G (2001) Molecular characterization of olive cultivars using RAPD markers. **Journal of the American Society for Horticultural Science 126**: 7-12.