Research Article

Identification of common bean (*Phaseolus vulgaris*) duplicates using agromorphological and molecular data

Alisson Fernando Chiorato[1], Sérgio Augusto Morais Carbonell[1], Luiz Antonio dos Santos Dias[2], Rafaeli Ramos Moura[3], Marilia Barbosa Chiavegato[3] and Carlos Augusto Colombo[3]

[1]*Instituto Agronômico de Campinas, Centro de Análises e Pesquisa Tecnológica do Agronegócio dos Grãos e Fibras, Campinas, SP, Brazil.*
[2]*Universidade Federal de Viçosa, Departamento de Biologia Geral, Viçosa, MG, Brazil.*
[3]*Instituto Agronômico de Campinas, Pesquisa e Desenvolvimento de Recursos Genéticos Vegetais, Campinas, SP, Brazil.*

Abstract

We used agromorphological and random amplified polymorphic DNA (RAPD) molecular marker data to identify duplicate common bean (*Phaseolus vulgaris* L. Fabaceae) accessions in the Common Bean Germplasm Bank of the Agronomical Institute - IAC (Banco de Germoplasma de feijoeiro do Instituto Agronômico de Campinas (IAC), SP, Brazil). A total of 116 accessions with the same names and similar agromorphological traits was analyzed. The divergence between the accessions was initially evaluated by means of the agromorphological descriptors using single linkage clustering, from the Euclidean distance. Multivariate analysis identified four duplicate accessions (Carioca Lustroso, Bico de Ouro, Jamapa and Preto), with 17 other same-name accessions being suspected duplicates due to their low divergence levels. Accessions with low genetic distance values (indicating that they were duplicates) were further compared using RAPD markers which confirmed the results of the multivariate analyses in relation to the four duplicate accessions, although only two of the other 17 suspect accessions were confirmed to be duplicates, in this case of accessions IAPAR 57 and Sacavem. These results show that the combined use of agromorphological and molecular information allowed a better characterization of the acessions in the common bean Germplasm Bank.

Key words: Phaseolus vulgaris, multivariate analysis, germplasm banks, descriptors.
Received: December 23, 2004; Accepted: July 5, 2005.

Introduction

Germplasm banks are places where samples of genotypes, improved varieties, landraces, wild species and species related to a determined species of interest (generically denominated accessions) are stored under adequate conditions (Zimmerman *et al*., 1996). The characterization of accessions allows quantification and structuring of the genetic variability in the germplasm which is highly important for improvement programs and for the conservation and preservation of genetic diversity. To characterize a germplasm basically means to identify and describe differences between the accessions. Besides the information on the origin of the material (passport data) differences related to the agricultural performance of the accessions are normally also considered (*e.g.* yield, growth and flowering habit, pathogen and pest responses) as well as botanical differences related to normally taxon-specific descriptors (Vanderborght, 1988).

The Common Bean Germplasm Bank of the Agronomical Institute - IAC (IAC-BGF) (Banco de Germoplasma de Feijoeiro do Instituto Agronômico - IAC, Campinas, SP, Brazil) holds 1500 accessions representing the two principal centers of origin (Andean and Mesoamerican) of the species as well as ecotypes from different south-American countries and a large number of lines from both Brazilian and international genetic improvement programs, most of which were obtained by germplasm interchange. Although germplasm interchange is advantageous because it permits the amplification of the variability of traits of agricultural interest, almost inevitably, it leads to the appearance of duplicates (repeated accessions) and consequently increases the costs of germplasm maintenance and characterization. On the other hand, accessions with the same name do not

Send correspondence to Alisson Fernando Chiorato. Instituto Agronômico de Campinas, Centro de Análise e Pesquisa Tecnológica do Agronegócio dos Grãos e Fibras, Av. Barão de Itapura 1.481, Caixa Postal 28, 13001-970 Campinas, SP, Brazil. E-mail: afchiorato@iac.sp.gov.br.

always correspond to the same genotypes and in such cases the putative duplicate could be erroneously discarded.

Multivariate techniques such as Euclidean distance, Mahalanobis' generalized distance and clustering analyses techniques such as single linkage allow the identification of duplicates. These analyses are generally carried out using agromorphological (phenotypic) data to produce dissimilarity matrixes for the accessions, which are subsequently used for cluster analyses in order to identify and group the accessions by their similarity. Another multivariate technique is principal component analyses which uses aims at visualizing the proximity of accessions and any links between the analyzed descriptors. However, Tatieni *et al.* (1996) argue that the phenotypic traits traditionally used for the characterization and estimation of genetic divergence may be of limited importance because they are generally influenced by the environment and the developmental stage of the plant. Isoenzyme and DNA markers which are little influenced by the environment are therefore more adequate for germplasm characterizations. Singh *et al.* (1991a) suggest that agromorphological, biochemical and molecular data should be combined for diversity studies since this combination offers complementary results.

Based on these premises, the objective of the present study was to identify duplicate common bean accessions in the IAC-BGF and to verify if accessions with the same name correspond to the same genotype. Multivariate analysis was applied to agromorphological and random amplified polymorphic DNA (RAPD) data and principal component analyses used to quantify the variation and identify redundant agromorphological descriptors.

Material and Methods

Evaluated accessions

The IAC-BGF contains 1500 active accessions, 116 (Table 1) of which were selected for duplicate identification based on the fact that they had similar names and agromorphological descriptors (Table 2). The selected accessions were sown in 2002 and 2003 at the Regional Center of the University of Espírito Santo do Pinhal (Centro Regional Universitário de Espírito Santo do Pinhal - CREUPI), Espírito Santo of Pinhal, São Paulo state, Brazil. For seed propagation and descriptor evaluation the experimental field consisted of 3 meter rows spaced 0.5 meters apart without replications but with a row of *Phaseolus vulgaris* cultivar (cv) IAC-Carioca Eté sown between every 10 accessions as a local control.

Evaluation of the agromorphological descriptors

Seventeen agromorphological descriptors that refer to the plant, pod, and seed data and the reaction to the anthracnose pathogen (*Colletotrichum lindemuthianum*) were evaluated. The reaction to anthracnose was evaluated in the laboratory by inoculating the 116 accessions under

Table 1 - Common bean accessions evaluated to identify duplicates due to similarities in the names and agromorphological descriptors.

Accession name (number of accessions evaluated)	Accession code number
Alemão (2)	18, 92
Baetão (4)	25, 37, 79, 80
Bico de Ouro (8)	4, 6, 27, 29, 30, 39, 90, 93
Carioca (3)	91, 109, 114
Carioca Lustroso (2)	104, 108
Chumbinho (11)	28, 31, 34, 36, 38, 40, 41, 43, 44, 45, 111
Guatemala (6)	22, 32, 50, 54, 60, 78
Huasano (2)	76, 83
IAPAR-57 (2)	86, 113
Iguaçu (2)	85, 112
Jalo (3)	7, 84, 89, 97, 98, 99, 100, 101, 102, 103, 110, 115
Jamapa (3)	55, 82, 87
Phaseolus vulgaris (3)	17, 49, 69
Preto (6)	47, 51, 59, 71, 81, 88
Ret. Santa Rosa (3)	9, 13, 48
Rim de Porco (3)	15, 58, 72
Roxinho (12)	3, 11, 14, 19, 20, 26, 46, 65, 74, 75, 105, 106
Rubi (2)	107, 116
Sacavem (25)	1, 2, 5, 8, 10, 12, 16, 23, 24, 35, 42, 52, 53, 56, 61, 62, 63, 64, 66, 67, 68, 73, 77, 94, 95
Tupi (2)	33, 57
Turrialba (2)	21, 70

conditions of controlled temperature and humidity with the three principal *C. lindemuthianum* physiological races (race 31, 65 and 89) most frequently occurring in São Paulo state, these isolates being provided by the IAC Plant Health Research and Development Center (Centro de Pesquisa e Desenvolvimento de Fitossanidade do Instituto Agronômico - IAC).

Statistical analyses

Univariate analysis was performed to quantify the existing variability for each descriptor by means of the phenotypic variation coefficient. After univariate analysis we performed multivariate analysis on the data, including Euclidean distance, principal component analysis, and the clustering of the accessions by the single linkage method were used for the quantification of the genetic divergence among the accessions. Because different measurement scales were used for the various descriptors the data were standardized before the analyses. Thus, the original mean (X_{ij}) obtained for descriptor j of accession i was divided by the standard deviation of the corresponding descriptor j, leading to a reduced mean (Z_{ij}) of the unitary variance.

Principal component analysis was used to visualize the variation and for the identification of the less important descriptors, the criterion used to statistically identify the redundant descriptors being based on Jolliffe (1973) and completed using the procedure used by Dias *et al*. (1997). The descriptors associated with the largest loading coefficients in the eigenvectors of the last principal components (*i.e.* those responsible for the smallest part of the total variance) were taken into consideration for the disposal. The divergence between accessions was evaluated using a Euclidean distance dissimilarity matrix, the single-linkage clustering method being applied to this distance matrix to produce a dendrogram showing the relative position of the accession groups. Possible duplicates were located when the dissimilarity between accessions was equal to zero. All multivariate analyses were processed using the 'Genes' software (Cruz, 2001).

Some accessions with similar names and agromorphological descriptors presented low levels of genetic dissimilarity, casting doubts on the possibility that they were duplicates. Analyses with RAPD molecular markers were used to confirm whether-or-not they were duplicates.

Molecular analyses

Suspected duplicate accessions were sown in pots in a greenhouse. Young leaves were collected for the extraction of genomic DNA according to the protocol of Rafalski *et al.* (1996). Samples of DNA of each accession were RAPD amplified in a total volume of 15 µL containing 0.4 mM primer, 10 mM Tris-GHl (pH 8.3), 50 mM KCl, 2 mM $MgCL_2$, 4 ng DNA, 100 µM of each deoxyribonucleotide (dNTPs) and 1.0 unit of Taq polymerase (Invitrogen). The amplifications were conducted in a PTC-100 thermocycler (MJ Research) programmed for 95 °C during 4 min, followed by 45 cycles of 1 min at 95 °C, 1 min at 35 °C and 1.5 min at 72 °C, plus an extension of 72 °C during 7 min. Fifteen different *Operon Technologies* primers were used (AS13, A18, AZ20, AB03, F10, H18, H20, AL09, W13, C08, L4, B3, J01, I16 and G08) and triplicated to confirm the results. Fragments were separated by agarose gel (1.5%) electrophoresis, stained with ethidium bromide and visualized under ultra violet light.

Results

The most variable descriptors based on the coefficient of variation were those of the reaction to the anthracnose races 31, 65 and 89, pod profile and seed gloss (Table 2). In the same multivariate context these descriptors were responsible for 73.38% of the total variation in the accessions.

Principal component analysis showed that the mass of a thousand seeds was a redundant descriptor and that of the 17 principal components the first two components were responsible for 33% of the variation, the first three compo-

Table 2 - Relative contribution (%) of the descriptors to the multivariate divergence, loading coefficients (LC) of the eigenvectors of the last principal component, and coefficient of phenotypic variation (CPV) of each descriptor.

Descriptor	Relative contribution (%)	Loading coefficient (LC)	Coefficient of phenotypic variation (%) (CPV)
Pod characteristics			
Primary color	0.23	0.5	34.29
Secondary color	0.35	-0.04	35.27
Profile	1.73	0.09	48.88
Width	0.04	-0.35	18.75
Number of seeds	0.84	-0.14	13.60
Growth			
Growth habit	1.86	-0.10	36.80
Flower characteristics			
Flower color	1.4	0.23	43.60
Bracteole format	0.38	0.05	28.51
Anthracnose reaction			
C. lindemuthianum race 65	21.09	-0.03	52.35
C. lindemuthianum race 89	22.46	0.28	55.32
C. lindemuthianum race 31	18.32	-0.04	45.78
Seed characteristics			
Secondary color	10.22	0.01	28.98
Halo color	3.34	0.02	30.30
Shape	1.47	-0.39	37.84
Profile	0.80	-0.06	29.31
Mass of 1000 seeds	5.65	0.73*	28.81
Gloss	9.75	0.06	45.28

*Greatest loading coefficient in the last component, indicating a redundancy of the descriptor associated to this component.

nents for 42.75% and the first four components for 50.52%, the results showing that the first eight components accounted for 70% of the variation (Table 3). The traits responsible for separation along the first principal component were the number of seeds in a pod, seed profile and reaction to anthracnose race 65. Traits affecting separation along the second principal component were pod width, seed shape and reaction to anthracnose race 89. Traits affecting the third principal component were growth habit and seed gloss.

The single-linkage dendrogram based on agromorphological data only (Figure 1) revealed four duplicate accessions (accession numbers in parenthesis refer to Table 1 and Figure 1): Carioca Lustroso (104 and 108); Bico de Ouro (90 and 93); Jamapa (55 and 87); and Preto (81 and 88). Various homonymous accessions (*i.e.* those with the same accession name) exhibited weak dissimilarity, *e.g.* Turrialba (21 and 70), Huasano (76 and 83), Sacavem (52 and 53), Rubi (107 and 116), Preto (85 and 112), Sacavem

Table 3 - Eigenvalues associated to the principal components, their relative and accumulated (%) variances, referring to 17 agromorphological descriptors evaluated in 116 common bean accessions of the IAC Germplasm Bank.

Components	Eigenvalues	Total variance (%)	Accumulated variance (%)
1	3.5077811	20.6340060	20.6340063
2	2.1035333	12.3737250	33.0077315
3	1.6562113	9.7424193	42.7501507
4	1.3218365	7.7755088	50.5256596
5	1.2201000	7.1770586	57.7027182
6	1.0986939	6.4629051	64.1656233
7	0.9101293	5.3537018	69.5193252
8	0.8797037	5.1747275	74.6940527
9	0.7428281	4.3695769	79.0636296
10	0.6694736	3.9380801	83.0017097
11	0.5768662	3.3933306	86.3950403
12	0.5366015	3.1564795	89.5515198
13	0.4996943	2.9393781	92.4908980
14	0.4200861	2.4710949	94.9619929
15	0.3307639	1.9456702	96.9076631
16	0.3017643	1.7750842	98.6827473
17	0.2239330	1.3172527	100.0000000

(1 and 68), Tupi (33 and 57), and IAPAR - 57 (96 and 113). Although the Sacavem (67) and Huasano (76) accessions were very close in terms of genetic distance they were not considered duplicates because their seeds had different colors.

Homonymous accessions showing a high degree of similarity but not always identified as duplicates by agromorphological descriptor cluster analysis were subjected to RAPD analysis. Together with these accessions, another with the name Jamapa (82) was included which did not group with its homonymous accessions (Jamapa 55 and 87), identified in Figure 1. Cultivar Jamapa is known worldwide and is used a lot in genetic improvement programs which led to its inclusion in the study with markers. The RAPD analyses showed that only H18, F10, AZ20, W13, G08, AL09 and AS13 were polymorphic in relation to the 15 primers tested (Figure 2). The RAPD markers confirmed the results of the analysis based on agromorphological data in which four clusters were formed in which there was total similarity among the accessions, *i.e.* accessions within the same cluster were genetically identical (Figures 2.1, 2.2, 2.3 for accessions Jamapa 55 and 87 and 2.4). Genetic identity was confirmed in only two of the suspected duplicate accessions (IAPAR - 57 and Sacavem, Figures 2.5 and 2.6). Differences in RAPD banding patterns being found for the other accessions (Figures 2.3 for accession Jamapa 82, 2.7, 2.8, 2.9, 2.10, 2.11 and 2.12).

Discussion

In the IAC common bean genetic improvement program any cultivar recommended for planting must present resistance to the *C. lindemuthianum* anthracnose races 31, 65 and 89 due to the fact that these pathogen races predominate in São Paulo state (Carbonell *et al.*, 1999). The relative contribution of the three anthracnose race descriptors to accession divergence was 61.88%, indicating that the common bean accessions in the IAC active germplasm bank showed a high level of variability in their response to these three pathogen races. It is interesting to note that the accessions analyzed in our study were chosen because they had similar names and agromorphological descriptors and represented only 7.7% of the total of accessions of the active IAC common bean germplasm bank, indicating that the variability would be even higher if all the accessions in the germplasm bank had been included in the study.

The mass of a thousand seeds descriptor was discarded due to the fact that the accessions analyzed were fairly uniformity of in relation to this trait with the majority (75%) of accessions showing a mean mass of 200 to 250 g per thousand seeds. This descriptor also showed only moderate correlation ($r = 0.58**$) with seed profile and was considered redundant and disposed of. Other disposal simulations were performed but principal component analysis did not reveal any other redundant descriptors. Fonseca and Silva (1999) obtained different results during the identification of duplicates Common Bean Germplasm Bank of the Embrapa, were weight of 100 seeds descriptor was the main responsible for divergence between the common bean groups Amendoim, Bagagô and Chita Fina, where it was responsible for 47% of the total variation.

The percentage of variation obtained in the first principal components was similar to that obtained in the evaluation of 306 common bean accessions in the International Center for Tropical Agriculture (CIAT, Centro Internacional de Agricultura Tropical) Germplasm Bank by Singh *et al.* (1991b), who found that 43% of the variation was made up of the first three components. Castineiras (1990) evaluated 60 common bean cultivars using 34 quantitative and qualitative descriptors and found that the first three components were responsible for 37% of the total variation and that 23 of the 34 descriptors were disposable. Rodrigues *et al.* (2002) showed that after discarding 25 of the 40 descriptors evaluated in 37 common bean cultivars 69% of the total of variation was made up of the first four components. Chiorato *et al.* (2005) evaluated 993 common bean accessions belonging to the IAC Germplasm Bank and found that 33% of the variation occurred in the first two components and concluded that only 18 of the 23 agromorphological descriptors were needed for evaluation of the accessions. Results in literature suggest that 10 to 20 descriptors should be considered for the characterization of common bean germplasm banks because the use of higher numbers is costly and adds little the analysis. However, it

should be remembered that the most informative and variable descriptors are not always the same for all the accessions in a germplasm bank. Another important point is the lack of association between common bean descriptors,

Figure 1 - Single linkage dendrogram based on 17 agromorphological descriptors for 116 common bean accessions from the IAC Germplasm Bank. (●) Accessions considered duplicates (■). Accessions suspected of being duplicates
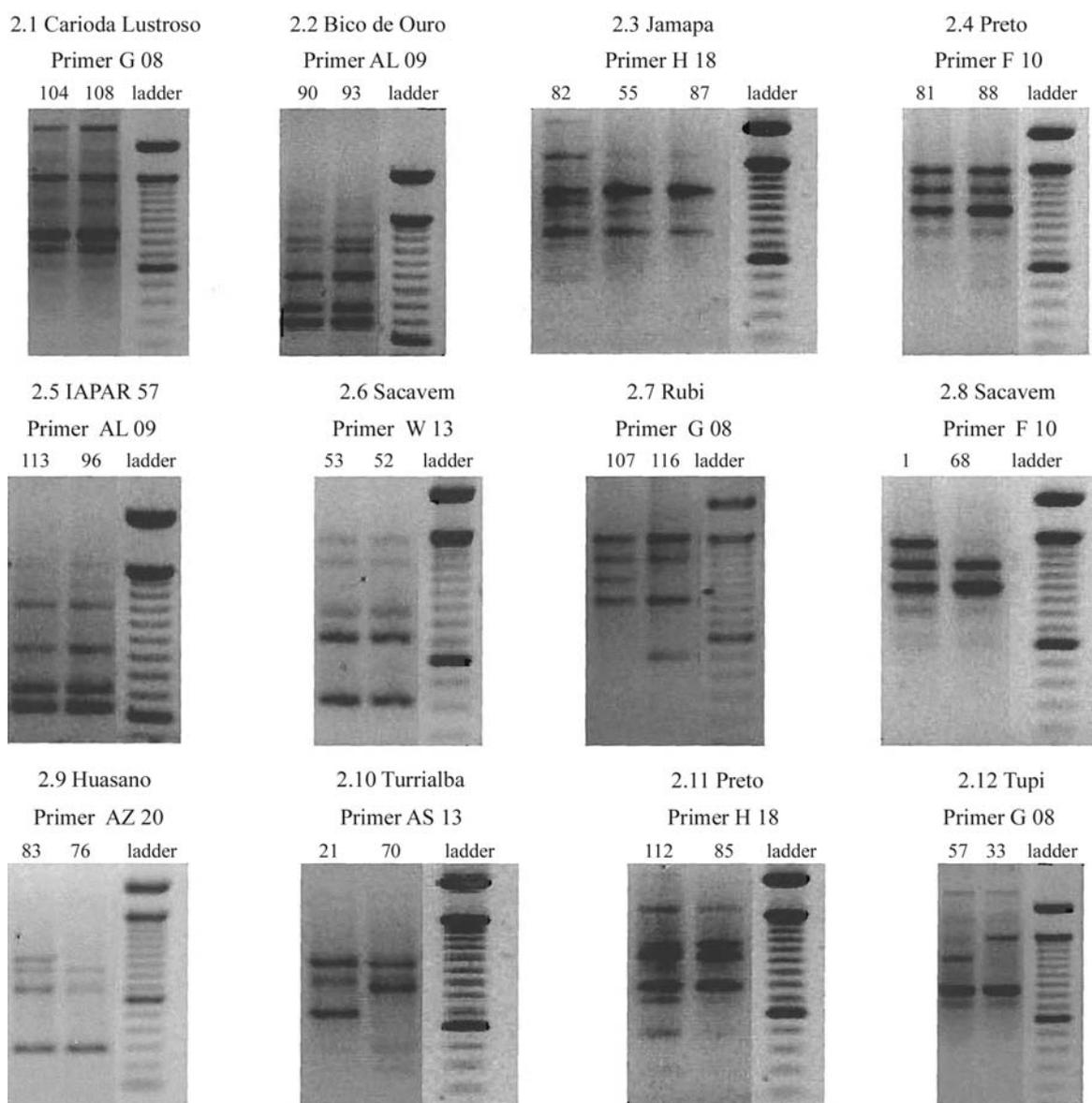
Figure 2 - Random amplified polymorphic DNA (RAPD) molecular marker agarose gel for the genetic differentiation of duplicate accessions in the active IAC Common Bean Germplasm Bank. The accession numbers correspond to those presented in Figure 1.

which means that disposal of the redundant descriptors does not result in a significant increase in the percentage of variation allocated to the first principal components.

Our multivariate analysis based on agromorphological descriptors detected four duplicate accessions but left the status of 17 other accessions in doubt, out of which RAPD analysis identified two duplicate accessions (Figures 2.5 and 2.6). Only in relation to the data of width and mean number of seeds per pod (data not shown) the dissimilarity observed among these accessions in the multivariate analyses was small (Figure 1). These quantitative traits being subject to environmental influences (including sowing accessions different cropping seasons) which may account for the small variations seen.

It is possible that the 15 phenotypically similar accessions which had similar names and which could not be separated either by agromorphological descriptors or RAPD analysis had been mistakenly indexed when stored, perhaps due to clerical errors when the accessions were received from other institutions during interchange. Another possibility is that recombination with other accessions occurred during seed propagation and that these segregation were not noticed.

Our data suggests that for the following accessions one of the homonymous genotypes could be discarded with no risks of the germplasm bank losing genetic variability: IAPAR-57 accessions 96 and 113; Carioca Lustroso 104 and 108; Bico de Ouro 90 and 93; Jamapa 55 and 87; Preto 81 and 88; and Sacavem 52 and 53. It also seems that

Jamapa accession 82 should be discarded due to the differences found in the banding-patterns in relation to Jamapa accessions 55 and 87.

For the homonymous accessions identified by molecular and multivariate analyses as not being duplicates (103 acessions which had 15 similar names) we propose the elaboration of new acronyms as a form of enhancing the identification of the genotypes (accessions) in the active IAC Common Bean Germplasm Bank.

Our results show that the combined use of agromorphological information and molecular data allowed the improved characterization of the accessions in the IAC Common Bean Germplasm Bank and that this approach could be applied to other plant germplasm collections.

Acknowledgments

References

Castineiras L (1990) Analisis de descriptores del frijol comum (*Phaseolus vulgaris*) mediante métodos multivariados. Ciência de la Agricultura 39:108-113.

Carbonell SAM, Ito MF, Pompeu AS, Francisco FG, Ravagnani S and Almeida ALL (1999) Raças fisiológicas de *Colletotrichum lindemuthianum* e reação de linhagens e cultivares de feijoeiro no Estado de São Paulo. Fitopatologia Brasileira 24:60-65.

Chiorato AF, Carbonell SAM, Colombo CA, Dias LAS and Ito MF (2005). Genetic diversity of common bens accessions in the germplasm bank of the Instituto Agronômico - IAC. Crop Breeding and Applied Biotechnology 5:1-9.

Cruz CD (2001) Programa Genes Versão Windows: Aplicativo Computacional em Genética e Estatística. Editora UFV, Viçosa, 648 pp.

Dias LAS, Kageyama PY and Castro GC (1997) Divergência fenética multivariada na preservação de germoplasma de cacau (*Theobroma cacao* L.). Agrotrópica 9:29-40.

Fonseca JR and Silva HT (1999) Identificação de duplicidades de acessos de feijão por meio de técnicas multivariadas. Pesquisa Agropecuária Brasileira 34:409-414.

Jolliffe IT (1973) Discarding variables in a principal component analysis. II. Real data. Applied Statistics 22:21-31.

Rafalski JA, Morgante M, Powell W, Vogel JM and Tingey SV (1996) Generating and using DNA markers in plants. In: Birren B and Lai E (eds) Analysis of Non-Mammalian Genomes - A Practical Guide. Academic Press, New York, pp 75-134.

Rodrigues LS, Antunes IF, Teixeira MG and Silva JB (2002) Divergência genética entre cultivares locais e cultivares melhoradas de feijão. Pesquisa Agropecuária Brasileira 37:1285-1294.

Singh SP, Gepts P and Debouck DG (1991a) Races of common bean (*Phaseolus vulgaris, Fabaceae*). Economic Botany 45:379-396.

Singh SP, Gutiérrez JA, Molina A, Urrea C and Gepts P (1991b) Genetic diversity in cultivated common beans: II. Markers based analysis of morphological and agronomic traits. Crop Science 31:23-29.

Tatieni V, Cantrell RG and Davis DD (1996) Genetic diversity in elite cotton germplasms determined by morphological characteristics and RAPD. Crop Science 36:186-192.

Vanderborght TA (1988) Centralized data base for common bean and its use in diversity analysis. In: Gepts, P (ed) Genetic Resources of *Phaseolus* Beans. Kluwer, Dordrecht, pp 51-65.

Zimmerman MJO and Teixeira MG (1996) Bancos de germoplasma. In: Araujo RS, Rava CA, Stone LF and Zimmermam MJO (eds) Cultura do Feijoeiro Comum no Brasil. Potafos, Piracicaba, pp 65-66.

Editor: Fábio de Melo Sene