



Discrimination of common bean cultivars using multiplexed microsatellite markers

P.C.B. Cardoso^{1,2}, C. Brondani¹, I.P.P. Menezes^{1,3}, P.A.M.R. Valdisser¹, T.C.O. Borba¹, M.J. Del Peloso¹ and R.P. Vianello¹

¹Laboratório de Biotecnologia, Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, Brasil

²Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, GO, Brasil

³Laboratório de Ciências Biológicas, Instituto Federal Goiano, Urutaí, GO, Brasil

Corresponding author: R.P. Vianello
E-mail: rosana.vianello@embrapa.br

Genet. Mol. Res. 13 (1): 1964-1978 (2014)

Received October 2, 2013

Accepted January 14, 2014

Published March 24, 2014

DOI <http://dx.doi.org/10.4238/2014.March.24.1>

ABSTRACT. Analysis of DNA polymorphisms allows for the genetic identification and precise discrimination of species with a narrow genetic base such as common bean. The primary objectives of the present study were to molecularly characterize commercial common bean varieties developed at various research institutions using microsatellite markers and to determine the degree of genetic diversity among the bean varieties analyzed. Fifty cultivars representing 12 grain classes and 64 genitors, i.e., accessions used to develop these cultivars, were characterized. Based on an analysis of 24 simple sequence repeats, the estimates for the average number of alleles and genetic diversity were 8.29 and 0.646, respectively. The combined probability of identity was estimated at 7.05×10^{-17} , indicating a high individual discriminatory power. Thirty-two percent of the cultivars exhibited heterogeneity for multiple loci that reflected either homozygosity for different alleles of a given locus in different individuals or heterozygosity for the locus. The average

genetic diversity for the groups of cultivars and genitors was 0.605 and 0.660, respectively, with no genetic differentiation (F_{ST}) between these groups. Although similar estimates of expected heterozygosity were observed when the cultivars were grouped by release date, a greater number of private alleles was observed in the most recent cultivars. The genetic differentiation among cultivars originating from different institutions was not different from zero ($F_{ST} = 0.01$). The molecular profile database derived from these analyses may increase the statistical power of genetic estimates and may be incorporated into breeding programs for common bean. Furthermore, the profiles obtained for the different cultivars may be used as molecular descriptors to complement traditional descriptors used in distinctiveness, uniformity and stability tests, thereby improving the traceability of samples and their derivatives and helping to protect the intellectual property rights of breeders.

Key words: *Phaseolus vulgaris*; Breeding program; Genetic diversity; SSR markers; Elite cultivars

INTRODUCTION

The domestication of plants is considered an ongoing and complex process in which plant species are increasingly adapted for cultivation and thus made more dependent on human intervention (Harlan, 1992). Multiple domestication processes have been reported for the *Phaseolus* genus (Gepts and Debouck, 1991), and the domestication of common bean (*Phaseolus vulgaris*) occurred independently in different gene pools approximately 8000 years ago (Mamidi et al., 2011; Serrano-Serrano et al., 2012). According to the Food and Agriculture Organization (FAO, 2012), dry bean is grown in 126 countries, and the estimated global production of this legume was 25 million tons in 2012. Brazil produces approximately 2.8 million tons of beans per year and is the third largest bean producer worldwide. Brazil also ranks third in terms of area under cultivation, with 2.7 million hectares used for bean cultivation.

Genetic breeding programs routinely release commercial cultivars that have advantages over existing cultivars. The production of new common bean cultivars typically requires 10 years of development and has a high aggregate cost, reflecting the numerous successive evaluation steps and selections conducted during this period. The first efforts to protect a developed variety date back to 1930, when legal mechanisms were established to protect plant varieties through asexual reproduction techniques. In 1961, the International Union for the Protection of New Varieties of Plants (UPOV) was created to establish and standardize requirements for the granting and revocation of rights for developed varieties. The 1970 United States statute known as the Plant Variety Protection Act sought to foster the development of new varieties using sexual reproduction techniques and to protect the rights of breeders. Although discussions regarding the regulation of intellectual property rights for cultivars were initiated in Brazil in 1945, plant variety protection in accordance with the 1978 UPOV Convention was not enacted into law until April 1997.

For a cultivar to be considered for registration by an authorized agency, UPOV requires that specific tests be performed for cultivar characterization and differentiation. These

tests are conducted by the official protection agency or the breeder of the cultivar. Tests of distinctiveness, uniformity and stability (DUS) are used to ensure that a new cultivar is distinguishable from other cultivars with known descriptors, that the characteristics of the new cultivar are homogeneous for each reproductive cycle and that the characteristics of the new cultivar are stably maintained through successive generations (MAPA, 2012). However, the instability of many of the evaluated characteristics and the limited number of polymorphisms associated with some descriptors restrict the discriminatory power of this evaluation, especially in species with a narrow genetic base. In this context, the development of systems based on DNA analysis has offered new possibilities for determining genetic identity and estimating genetic linkage. These systems may therefore complement DUS tests used for varietal protection. Moreover, the data generated from molecular analyses are conclusive, and the results are reproducible, consistent and not influenced by environmental fluctuations or variation reflecting the stage of plant development.

Microsatellite markers, or simple sequence repeats (SSRs), are internationally recognized as molecular descriptors for various genotypic characterization analyses, such as genetic identification tests and parentage testing in humans (FBI, 2012). Several collections of SSR markers are publicly available for molecularly characterized plant species, including common bean (Garcia et al., 2011). SSR-based molecular genotyping has rapidly evolved, and methods for the simultaneous amplification of multiple loci coupled to semi-automated detection systems have been developed (Masi et al., 2003; Borba et al., 2009; Cardoso et al., 2013). The numerous advantages of SSR-based genotyping include the increased speed and accuracy of analyzing the generated data, fewer errors in producing the genetic profiles and the potential for data standardization, all of which promote the transfer and consolidation of data among different laboratories or groups.

The primary objectives of the present study were as follows: to perform a microsatellite marker-based molecular characterization of commercial bean cultivars and experimental lines developed at various research institutions in Brazil and other countries and to assess the molecular profile database in terms of its discriminatory power, its ability to identify individuals and parentage and its utility for estimating genetic diversity.

MATERIAL AND METHODS

Plant material

A total of 114 common bean (*P. vulgaris*) genotypes, including 50 commercial cultivars released by research institutions in Brazil and other countries and 64 genitor lines used in breeding programs, were characterized (Tables 1 and 2). For each cultivar, up to four plants per sample were collected and analyzed in bulk, whereas individual plants were used for the analysis of the genitor lines. Genomic DNA was extracted from leaf tissue samples using the CTAB extraction method described by Ferreira and Grattapaglia (1998).

Molecular analysis

A set of 24 SSR markers assembled into a multilocus genotype (Cardoso et al., 2013) and a semi-automated detection system were used. PCRs combining more than one marker

were performed in a final reaction volume of 20 μ L containing 15 ng DNA, 1X buffer (10 mM Tris-HCl at pH 8.3, 50 mM KCl, 1.5 mM MgCl₂), 0.25 mM dNTPs, 5% DMSO, optimized concentrations of each primer pair (forward and reverse) and 1 U Taq DNA polymerase. The PCR program for amplification was as follows: an initial DNA denaturation at 94°C for 5 min; 40 cycles of denaturation at 94°C for 1 min, annealing at optimized temperatures for each primer

Table 1. Description of the common bean cultivars and their respective grain types, institution or place of origin and year of release.

Cultivar	Type of grain	Origin	Release date
Alba	Carioca	Has no official breeder	-
Aporé	Carioca	CIAT/EMBRAPA Arroz e Feijão	1992
BRS Campeiro	Black	EMBRAPA Arroz e Feijão	2003
BRS Cometa	Carioca	EMBRAPA Arroz e Feijão	2005
BRS Embaixador	Dark Red Kidney	CIAT/EMBRAPA Arroz e Feijão	2007
BRS Estilo	Carioca	EMBRAPA Arroz e Feijão	2010
BRS Executivo	Cranberry	CIAT/EMBRAPA Arroz e Feijão	2007
BRS Expedito	Black	EMBRAPA Clima Temperado	2005
BRS Explendor	Black	EMBRAPA Arroz e Feijão	2010
BRS Grafite	Black	EMBRAPA Arroz e Feijão	2003
BRS Horizonte	Carioca	CIAT/EMBRAPA Arroz e Feijão	2004
BRS MG Majestoso	Carioca	EMBRAPA Arroz e Feijão	2006
BRS MG Pioneiro	Carioca	BIOAGRO - UFV ^{2*}	2005
BRS Pitanga	Purple	EMBRAPA Arroz e Feijão	2004
BRS Pontal	Carioca	EMBRAPA Arroz e Feijão	2003
BRS Requite	Carioca	EMBRAPA Arroz e Feijão	2003
BRS Supremo	Black	EMBRAPA Arroz e Feijão	2004
BRS Timbó	Purple	CIAT/EMBRAPA Arroz e Feijão	2002
Campeão 1	Carioca	Has no official breeder	-
Campeão 2	Carioca	Has no official breeder	-
Corrente	Mulatinho	EMBRAPA Arroz e Feijão	1993
Diamante Negro	Black	EMBRAPA Arroz e Feijão	1991
EMGOPA 201-Ouro	Jalinho	CIAT	1984
EMGOPA 202-Rubi	Rosinha	EMBRAPA Arroz e Feijão	1988
FT-Bio Nobre	Black	FT - Pesquisa e Sementes S.A.	1999
FT-Nobre	Black	FT - Pesquisa e Semente S.A.	1996
Guará	Carioca	EPAGRI ³	2003
IAC Carioca Eté	Carioca	IAC ⁴	1999
IAC Carioca Tybatã	Carioca	IAC	2000
IAPAR 80	Carioca	IAPAR ⁵	1997
IAPAR 81	Carioca	IAPAR	1997
IPA 6	Mulato	IPA ⁶	1985
IPR Chopim	Carioca	IAPAR	2004
IPR Colibri	Carioca	IAPAR	2004
IPR Graúna	Black	IAPAR	2002
IPR Juriti	Carioca	IAPAR	2002
IPR Saracura	Carioca	IAPAR	2004
IPR-88 Uirapuru	Black	IAPAR	2000
Jalo Precoce	Manteigão	EMBRAPA Arroz e Feijão	1993
Magnifico	Carioca	FT - Pesquisa e Sementes S.A.	2001
Ônix	Black	EMBRAPA Arroz e Feijão	1992
Ouro Branco	White	CIAT	1993
Ouro Negro	Black	Honduras	1991
Ouro Vermelho	Red	EMBRAPA/UFLA ⁷ /UFV/EPAMIG	2005
Pitoco	Carioca	Has no official breeder	-
Princesa	Carioca	EMBRAPA Arroz e Feijão	1997
Rudá	Carioca	CIAT	1994
Soberano	Black	FT - Pesquisa e Sementes S.A.	2001
Vermelho 2157	Red	CIAT	1993
Xamego	Black	EMBRAPA Arroz e Feijão	1993

¹Centro Internacional de Agricultura Tropical; ²Universidade Federal de Viçosa; ³Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina; ⁴Instituto Agrônômico de Campinas; ⁵Instituto Agrônômico do Paraná; ⁶Instituto Agrônômico de Pernambuco; ⁷Universidade Federal de Lavras.

Table 2. Common bean cultivars characterized using SSRs and their respective genealogies.

Cultivar	Genealogy
Alba	Selection of Pérola cultivar *
Aporé	Carioca* / México 168* // Carioca* / BAT76*
BRS Campeiro	Radiação Gama na Cultivar Corrente* - A176* / CATU*
BRS Cometa	A769* /4/ EMP250* /// A429* / XAN252* // V8025* / Pinto UI 114* // WAF2 / A55* // GN31 / XAN170*
BRS Embaixador	XAN 42 / G 13922
BRS Estilo	EMP250* /4/ A769* /// A429* / XAN252* // V8025* / Pinto UI 114*
BRS Executivo	A 192 / BAT 1274
BRS Expedito	CNF5491 / FT Tarumã*
BRS Explendor	CB911863* / AN9123293*
BRS Grafite	AN512567* / México 168*
BRS Horizonte	EMP250* /4/ A769* /// A429* / XAN252* // V8025* / Pinto UI 114*
BRS MG Majestoso	Ouro Negro* / Pérola*
BRS MG Pioneiro	Rudá* / Ouro Negro*
BRS Pitanga	FEB163 / AN512879*
BRS Pontal	BZ3836* / FEB166* // AN910523*
BRS Requite	Carioca MG* / POT94* // AN910523*
BRS Supremo	W22-34* / VAN163*
BRS Timbó	A252* / XAN105 // A373* / A213* /// A445* / XAN112* // BAT477* / A213*
Campeão 1	-
Campeão 2	-
Corrente	A176* / CATU*
Diamante Negro	XAN87 / A367*
EMGOPA 201-Ouro	A30 / Acté 1 / 37*
EMGOPA 202-Rubi	Rio Iguaçu / Tayhú*
FT-Bio Nobre	-
FT-Nobre	FT120* / FT 84-1806 // FT84-424
Guará	Carioca* / FT 87-77
IAC Carioca Eté	L933 / LM30630 (Ônix)*
IAC Carioca Tybatã	L933 / LM30630 (Ônix)*
IAPAR 80	A176* / A259* /4/ BAT93* // carioca 99* / Great Northern Nebraska 1 sel. # 27* /// Sel. Aroana
IAPAR 81	BAT93* /// carioca 99* / Great Northern Nebraska 1 sel. # 27* // Sel. Aroana /4/ A248 / EMP117*
IPA 6	Rico 23* / Gordo*
IPR Chopim	F7 (BAT93* /2/ Carioca sel. 99* / Great Northern Nebraska 1 sel. # 27* / 3 / Sel. Aroana /4/ A176* / A259*) / NEP 2*
IPR Colibri	Seleção de Carioca Pitoco precoce
IPR Graúna	EP173 / 2 / Rio Iguaçu* / Great Northern Nebraska 1 sel. # 27* / 3 / Rio Tibagi* / Cornell49242* /4/ Iapar BAC25* /5/ IAPAR BAC26
IPR Juriti	BAT93 / 2 / Carioca Sel. 99* / Great Northern Nebraska 1 sel # 27* / 3 / Sel. Aroana /4/ A176* / A259* /5/ 11113 / XAN87
IPR Saracura	LPSP1 93-08 / IAPAR 31*
IPR-88 Uirapuru	IAPAR BAC29 / PR1711 / 3 / NEP 2* / 2 / Puebla 173 / ICA Pijão*
Jalo Precoce	Seleção de Goiano Precoce*
Magnífico	FT120* / FT84-179
Ônix	Porriño Sintético* / Turrialba 1* // ICA Pijão* / Negro Jamapa*
Ouro Branco	G10030 (Eibun) / G76 (Red Kloud)*
Ouro Negro	-
Ouro Vermelho	AN9022180* / Vermelhinho* // Vermelhinho*
Pitoco	-
Princesa	A252* / BAT1550*
Rudá	Carioca* / Rio Tibagi*
Soberano	FT90-109 / IAPAR 31* / FT Rio Preto
Vermelho 2157	BAT841* / BAT1230*
Xamego	LM20771* / BAT256* // LM20322* / BAT67*

*Components of the genealogy that were characterized. The symbols /, //, ///, /4/ and /5/ indicate the order of crossings.

pair (48° to 58°C) for 1 min and an extension at 72°C for 1 min; and a final extension at 72°C for 10 min. A 2-µL aliquot of the PCR product was combined with 7 µL Hi-Di formamide and 1 µL of a molecular weight marker, according to the methods previously described by Brondani and Grattapaglia (2001). The amplified fragments were separated by capillary electrophoresis on an ABI3100 Genetic Analyzer platform (Applied Biosystems), and the fragment analysis was performed using GeneScan 2.1 and GeneMapper 3.5 (Applied Biosystems).

Data analysis

The genotypic data were fitted into allelic classes based on a minimum difference of two bases. Genetic purity analysis was performed by testing for the presence of divergent alleles from a bulk analysis of the accessions. The observation of only one allele from the bulk analysis for a given microsatellite locus indicated genetic purity of the cultivar for the site. For instances when the bulk sample yielded more than one allele for at least two loci, the cultivar analyzed was considered heterogeneous. The samples for such cultivars were individually genotyped to determine whether the heterogeneity reflected heterozygosity for the locus or homozygosity for different alleles in different plants at the locus of interest (i.e., a mixture of pure lines).

The genetic diversity of the total sample was determined using estimates of allele frequency, the number of alleles per locus and the expected (H_E) and observed (H_O) heterozygosity values obtained using Cervus 3.0 (Kalinowski et al., 2007). The identification of private alleles and linkage disequilibrium analysis were performed using the GDA program (Lewis and Zaykin, 2001). The discriminatory power for each locus and multilocus was calculated based on the probability of genetic identity and the probability of paternity exclusion using the Identity program (Wagner and Sefc, 1999). The pairwise genetic distance (GD) was calculated using Wright's modified Roger's coefficient available in the NTSys program (Rohlf, 1989) and the matrix of allele frequencies generated by the Fstat 2.9.3.2 program (Goudet, 2001). The unweighted pair group method with arithmetic mean (UPGMA) and the Mantel test in the NTSys program (Rohlf, 1989) were used for the cluster analysis. Genetic differentiation was estimated by Wright's F_{ST} (Weir and Cockerham, 1984) and a 95% confidence interval for 10,000 bootstraps using Fstat 2.9.3.2 (Goudet, 2001). The factorial correspondence analysis (FCA) was generated using the Genetix 4.03 program (Belkhir et al., 2004). The Structure 2.2.3 program (Pritchard et al., 2000) was used to determine the structure of the germplasm analyzed, and the number of genotypic groups was determined as proposed by Evanno et al. (2005) using the Structure Harvester program (Earl and VonHoldt, 2012).

RESULTS

The 24 SSRs corroborated the assumption of independence according to the linkage disequilibrium test and were considered suitable for genotyping common bean varieties (Table 3). Altogether, 157 genotypes were analyzed, including cultivars genotyped by pooling four plants ($N = 35$), individual plants derived from the heterogeneous decomposition of bulk samples ($N = 58$) and the group of genitors ($N = 64$). An average of 8.29 alleles was identified, with the number of alleles ranging from 3 (PvBR11 and PvBR169) to 21 (BM154). The average genetic diversity (as measured by H_E) was 0.646, ranging from 0.296 for the BM212 locus to 0.880 for the BM154 locus. The estimated H_O value of 0.031, which was consistent with the high level of inbreeding indicated by a fixation index (F_{IS}) of 0.950, underscored the trend towards homozygosity among the loci analyzed (Table 3).

Of the 50 cultivars characterized in bulk, 29 (58%) exhibited a homozygous molecular profile for all 24 loci, and 5 cultivars (10%) had only one detectable heterogeneous locus. Accordingly, these 34 cultivars (68%) were characterized as pure. The remaining 16 cultivars (32%) were heterogeneous for at least 2 of the 24 SSRs analyzed. The heterogeneous loci

were decomposed by individually analyzing the sampled plants. For 3 of the 16 cultivars, the heterogeneity was attributable to homozygosity for different alleles in different individuals. In the remaining 13 cultivars, including IAC Carioca Tybatã, FT-Bio Nobre and Campeão 1, which had 4, 5 and 9 heterogeneous loci, respectively, heterozygosity for one or more loci was observed.

Table 3. Genetic parameters derived from the analysis of the 24 SSR markers in common bean.

Loci	No. of alleles	$F_{r_{max}}$	H_E	H_O	F_{is}	PI	PE
BM114	7	0.53	0.667	0.000	1.000	0.215	0.449
BM143	10	0.46	0.721	0.007	0.991	0.175	0.511
BM154	21	0.24	0.880	0.079	0.911	0.048	0.758
BM165	7	0.37	0.767	0.007	0.992	0.156	0.556
BM181	4	0.75	0.394	0.000	1.000	0.496	0.197
BM183	5	0.60	0.598	0.006	0.989	0.272	0.377
BM185	7	0.32	0.745	0.019	0.974	0.197	0.511
BM187	12	0.18	0.708	0.036	0.949	0.161	0.514
BM189	5	0.80	0.334	0.077	0.771	0.521	0.174
BM201	6	0.61	0.592	0.027	0.955	0.263	0.379
BM202	5	0.50	0.555	0.013	0.976	0.496	0.258
BM210	10	0.47	0.699	0.058	0.918	0.206	0.480
BM211	10	0.35	0.779	0.064	0.918	0.142	0.578
BM212	4	0.83	0.296	0.000	1.000	0.557	0.153
PV005	7	0.29	0.798	0.007	0.992	0.130	0.601
PV011	3	0.70	0.445	0.032	0.927	0.488	0.215
PV013	7	0.32	0.785	0.154	0.804	0.140	0.583
PV025	11	0.21	0.859	0.020	0.976	0.069	0.714
PV035	5	0.51	0.646	0.006	0.990	0.294	0.388
PV087	13	0.48	0.679	0.000	1.000	0.237	0.450
PV113	6	0.64	0.560	0.013	0.977	0.280	0.355
PV163	20	0.43	0.770	0.033	0.958	0.110	0.599
PV169	3	0.76	0.380	0.051	0.866	0.522	0.183
PV272	11	0.27	0.846	0.032	0.962	0.078	0.690
Average	8.29	-	0.646	0.031	0.950	-	-
Total	199	-	-	-	-	7.05×10^{-17}	0.999999*

$F_{r_{max}}$ = maximum allele frequency; H_E = expected heterozygosity; H_O = observed heterozygosity; F_{is} = fixation index; PI = probability of identity; PE = power of exclusion. *Multilocus value.

Overall, 35 private alleles were identified in 15 of the SSRs analyzed (17.6% of the total). Among the corresponding loci, PvBR163 and BM154 were clearly distinct from the others with 7 and 8 private alleles, respectively (Table 4). The G76 (Red Kloud), Gordo and Goiano Precoce genotypes, which were integrated into the genealogy of cultivars with White, Mulatto and Manteigão grains, respectively, and the BRS Executivo (Cranberry type grain) genotype were also distinct from the other cultivars insofar as each of these genotypes had at least three private alleles. The probability of identity (PI) for the groups of cultivars and genitors ranged from 0.048 (BM154) to 0.522 (PvBR169) for the loci with higher and lower powers of discrimination, respectively. The combined PI, which indicates the probability of two random samples having identical genotypes, was estimated to be 7.05×10^{-17} for the 24 loci (Table 3). Among the cultivars, the estimated PI was 2.2×10^{-15} . $H_E \geq 0.7$ was observed for only 11 SSRs (BM143, BM154, BM165, BM185, BM187, BM211, PvBR5, PvBR13, PvBR25, PvBR163, and PvBR272), which allows for the discrimination of all cultivated genotypes. These data yielded a combined PI of approximately 4.51×10^{-10} . In other words, the probability that two randomly selected cultivars would have the same genotype with this set of SSRs is less than 1

in 2 million. For the accessions used to generate the cultivars, the genitors NEP 2 and BAT256 were genetically indistinguishable. NEP 2 was used to develop the cultivars IPR Chopim and IPR 88 Uirapuru at the Agronomic Institute of Paraná (IAPAR), while BAT256 was used for the development of the Xamego cultivar at Embrapa Rice & Beans (Embrapa Arroz e Feijão).

Table 4. Private alleles (A_p) and the respective SSR markers identified in the groups of genitor accessions and common bean cultivars.

Groups	Accessions	A_p	Locus_bp
Genitor	1-México 168	1	BM201_100
	2-Goiano Precoce	2	PV163_301; BM165_182
	3-AN910523	1	BM211_234
	4-AN512879	1	BM143_168
	5-Pinto UI 114	1	BM212_206
	6-Cornell 49242(D)	1	PV087_168
	7-BAT67	1	PV087_150
	8-W2234	1	BM187_188
	9-Nebraska 1 sel. #27	2	BM187_190; PV163_238
	10-G76 (Red Kloud)	3	BM187_267; PV163_309; BM210_166
	11-XAN170	1	PV163_211
	12-Gordo	3	BM202_157; PV163_296; BM154_238
	13-CB911863	1	BM202_140
	14-Rio Tibagi	1	BM154_301
	15-Negro Jamapa	1	BM154_271
	16-Iapar BAC25	1	BM154_252
	17-BAT1230	1	BM154_288
	18-A445	1	BM154_245
	19-A213	1	BM154_242
	20-A55	1	PV272_93
Cultivars	1-Ouro Vermelho	1	BM143_149
	2-IPA6	1	BM143_147
	3-SUG33	3	BM114_237; PV163_317; BM189_99
	4-Jalo Precoce	2	PV087_187; BM154_240
	5-Ouro Negro	1	PV113_074
	6-Soberano	1	PV163_330

The average genetic distance observed among the 114 genitors and cultivars was 0.774, with 98.6% of the comparisons indicating a genetic distance greater than or equal to 0.5. The dendrogram generated by UPGMA clustering and the cophenetic correlation index (r) of 0.777 ($P < 0.001$) validated the grouping of the cultivars by grain type (Figure 1). Analysis for this grouping scheme (Figure 2) revealed an average genetic distance of 0.764 ($r = 0.931$, $P < 0.001$). The lowest genetic distance was observed between the Carioca and Black groups (0.30), and the highest genetic distance was observed when the groups with Cranberry, Manteigão, White, and Dark Red Kidney grain types were compared with the other groups (0.90). Thus, two main groups were formed, with the first representing small grain types (the Mesoamerican gene pool), such as Carioca, Black, Purple, Red, Rosinha, Jalinho, Mullato, and Mulatinho, and the second representing large grain types (Andean gene pool), such as Manteigão, Dark Red Kidney, White, and Cranberry (Figure 2). These results were corroborated by the FCA. Two groups of individuals were formed by the Bayesian model structure analysis including all of the genotypes, with the lower group formed by the seven accessions of Andean origin and the other group including the majority of the accessions of Mesoamerican origin.

When the average genetic diversity was estimated separately for the groups of cultivars and genitors, the values were 0.605 and 0.660, respectively. Separating the cultivar accessions by grain type revealed higher estimates within the Black group ($N = 14$, $A = 4.08$,

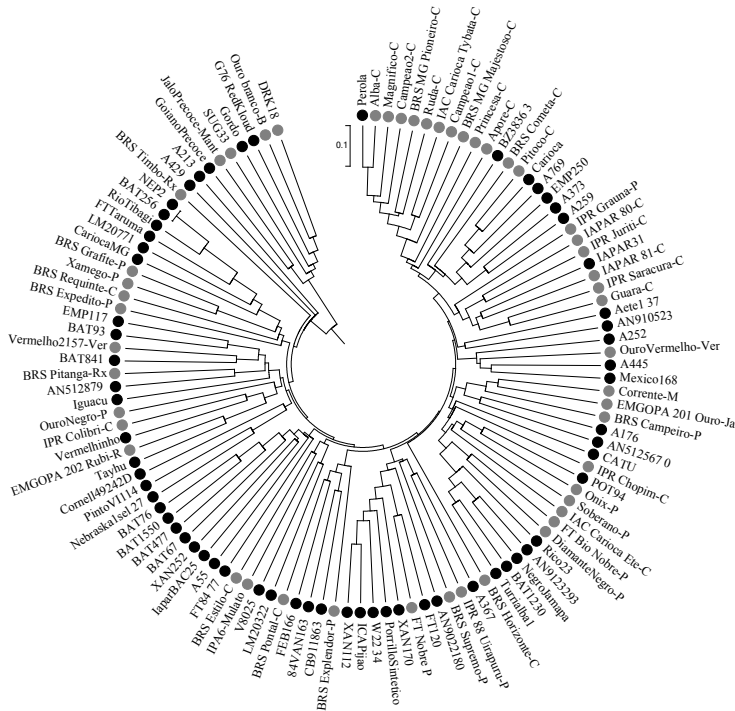


Figure 1. Dendrogram representing the genetic distances among all of the cultivars (gray circles) and genitors (black circles) analyzed.

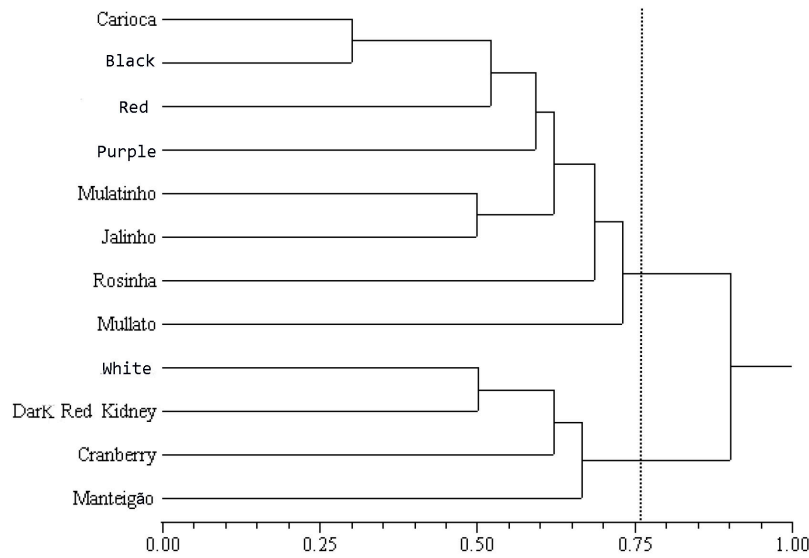


Figure 2. Dendrogram representing the genetic distances (Wright’s modified Roger’s genetic distance) among the cultivars (grouped by grain type).

$H_E = 0.544$, $GD = 0.734$) compared to the Carioca group ($N = 24$, $A = 4.33$, $H_E = 0.452$, $GD = 0.670$), except for the average number of alleles. The estimated genetic differentiation (F_{ST}) between the Carioca and Black groups was 0.134 and was significantly different from zero (95%CI = 0.082-0.196, $P < 0.05$). A smaller but nonetheless significant F_{ST} of 0.053 (95%CI = 0.015-0.102, $P < 0.05$) was observed for the cultivar and genitor accession groups.

When pairwise comparisons of the cultivars were made according to their release dates (1983 to 2000 classified as Older and 2001 to 2010 classified as Recent), the average genetic distances observed for the two groups were similar (Older = 0.77, Recent = 0.79; Table 5). The values for the average number of alleles were also similar (Older = 5.5, Recent = 6.1), but there was an increase in the number of private alleles in the Recent cultivars, with 17 private alleles in the Older group and 31 in the Recent group (Table 6). Consistent with the high average genetic dissimilarity of the cultivars within each period, genetic diversity was also high for both periods. H_O was low but higher for the Older period, indicating increased rigor in terms of purification in the Recent cultivars. The low F_{ST} estimate indicates the absence of a genetic structure for the common bean cultivars in both periods. The same results were observed for the Black and Carioca grain cultivars when analyzed separately. In general, the genetic parameters indicate that genetic variability has been maintained over the past 26 years. When the cultivars were grouped by their institutions of origin (CIAT, EMBRAPA, IAPAR, IAC, and FT), the estimated genetic distances varied from 0.76 to 0.83, with no differences observed among the institutions ($F_{ST} = 0.013$, $P > 0.05$).

Although a thorough analysis of genealogical relationships was not within the scope of the present study, we generally observed that genealogically similar cultivars were located close to each other in the dendrogram (Figure 1). Of the 64 accessions used to generate the cultivars, 22 were used for multiple cultivars. For example, the A176 and Great Northern Nebraska 1 selection 27 accessions were integrated into the genealogy of five different cultivars, and the BAT93 accession was a genitor for four cultivars. While some cultivars included only two accessions in their genealogy, other cultivars had large numbers of contributing accessions, such as BRS Cometa, which was derived from 10 genitors.

DISCUSSION

The 24 SSRs included in the genotyping system were in full compliance with the established requirements insofar as they allowed for a comprehensive resolution of alleles with type-specific amplification patterns and high consistency. Given that equally well-defined morphological and agronomic descriptors are required for the protection of developed genetic material (Plant Variety Protection Act, 1997), the molecular markers used in this study may potentially enhance varietal protection efforts. The standardization of accurate microsatellite genotyping systems that can be used in different laboratories and the effective use of the generated molecular profiles as descriptors for a particular crop of interest require validation steps to confirm that the same molecular patterns are obtained at different institutions. For instance, the establishment of a typing system in humans for forensic purposes was the result of an organized validation effort involving a number of laboratories worldwide (Kimpton et al., 1996), which yielded a set of 13 internationally recommended SSRs (Budowle et al., 1999). A similar study conducted for *Eucalyptus* involving an international network of laboratories in eight countries resulted in a set of 25 SSRs for the characterization of this genus, with six loci

Table 5. Average genetic distance (Wright's modified Roger's genetic distance) for each common bean cultivar in the 1983-2000 and 2001-2010 groups.

	Cultivars	1983-2000	2001-2010
1983-01	EMGOPA 201-Ouro	0.748	0.769
	IPA 6	0.802	0.777
	EMGOPA 202-Rubi	0.802	0.802
	Diamante Negro	0.753	0.760
	Ouro Negro	0.788	0.795
	Aporé	0.760	0.769
	Onix	0.741	0.762
	Ouro Branco	0.916	0.917
	Vermelho 2157	0.755	0.762
	Jalo Precocoe	0.918	0.914
	Corrente	0.780	0.781
	Xamego	0.821	0.802
	Rudá	0.730	0.723
	FT-Nobre	0.733	0.744
	Princesa	0.794	0.788
	IAPAR 80	0.704	0.716
	IAPAR 81	0.748	0.750
	FT-Bio Nobre	0.771	0.786
	IAC Carioca Eté	0.752	0.768
	IAC Carioca Tybatã	0.713	0.719
2001-10	IPR-88 Uirapuru	0.780	0.773
	Average	0.777	0.780
	Magnífico	0.738	0.745
	Soberano	0.735	0.753
	BRS Timbó	0.832	0.814
	IPR Juriti	0.718	0.754
	IPR Graúna	0.756	0.798
	BRS Pontal	0.766	0.753
	BRS Requite	0.816	0.812
	BRS Campeiro	0.769	0.790
	BRS Grafite	0.780	0.813
	Guará	0.724	0.744
	BRS Horizonte	0.788	0.779
	BRS Supremo	0.774	0.769
	BRS Pitanga	0.821	0.807
	IPR Chopim	0.752	0.780
	IPR Colibri	0.802	0.819
	IPR Saracura	0.727	0.726
	BRS MG Pioneiro	0.712	0.729
	OuroVermelho	0.776	0.772
BRS Cometa	0.718	0.739	
BRS Expedito	0.795	0.796	
BRS MG Majestoso	0.784	0.800	
BRS Executivo	0.916	0.935	
BRS Embaixador	0.886	0.910	
BRS Estilo	0.819	0.807	
BRS Explendor	0.789	0.789	
Average	0.780	0.789	

Table 6. Descriptions of the genetic variability of the common bean cultivars.

Period	A	P%	No. of alleles	A_p	H_E	H_O	GDW	F_{ST}
1983-00	21	100	132 (5.50)	17	0.650	0.042	0.777	0.01 ^{NS}
2001-10	25	100	146 (6.08)	31	0.656	0.017	0.789	

A = number of cultivars; P% = proportion of polymorphic loci; A_p = number of private alleles; H_E = expected heterozygosity; H_O = observed heterozygosity; GDW = genetic distance between cultivars in the same period; F_{ST} = estimated genetic differentiation.

established as the minimum set for the genetic discrimination of different *Eucalyptus* clones (Kirst et al., 2005). All of the markers used in the present study have been published and char-

acterized and have available mapping information (Gaitán-Solís et al., 2002; Blair et al., 2003; Buso et al., 2006; Grisi et al., 2007). There is strong evidence indicating the sharing of marker information among different laboratories, which represents the first step towards validating the markers with standardized amplification procedures.

The maximum allele frequency estimates exceeded 20% for all of the loci evaluated, and 50% of the loci contained alleles that occurred at rates exceeding 50%, indicating an uneven distribution of alleles and the predominance of certain alleles in establishing the different cultivars. The number of genitors shared by the different cultivars (approximately 35% of the genitors) certainly contributed to the reduced variability and increased frequency of certain alleles in the cultivars sampled. The high allelic frequency reduced the informative power of certain SSR loci, and the average H_E of 0.49 in the present study (with the H_E values ranging from 0.10 to 0.88) was used to select loci with the greatest informative potential for integrating the multilocus genotyping system. In this context, fewer but more informative loci would increase the discriminatory power of the analysis and make the process faster and less expensive. The findings that an H_E value greater than 0.70 was observed for only 11 loci (average of 0.79) and that the average number of alleles per locus was 11 are consistent with the high diversity of common bean. However, an average of 8.29 alleles per locus was observed for the set of 24 markers from which the genetic estimates were obtained. Similar estimates were reported by Gaitán-Solís et al. (2002) and Blair et al. (2006), who reported averages of 6 and 7.8 alleles per locus using 68 and 129 microsatellite markers, respectively. In these two studies, cultivated and wild *P. vulgaris* varieties of Andean and Mesoamerican origin were characterized. Although a smaller number of markers was used in the present study, the wide diversity of genotypes sampled, which included materials from various origins that had been used to develop the analyzed cultivars, contributed to the higher average number of alleles per locus observed. When the groups of genitors and cultivars were analyzed separately, allelic diversity was slightly higher among the genitors than in the cultivars, which had 7.42 and 6.83 alleles per locus. This finding is indicative of the allelic representativeness among genitors and their offspring.

The allele frequency database may prove critical for deriving more accurate estimates of discriminatory power in terms of determining the parentages of common bean genotypes. The database also ensures a more structured and organized means of storing the generated data. With wider ranges of genotyped samples and less population structuring between samples, there is a greater representation of alleles and their respective frequencies in the database, thereby increasing the power of the estimates derived from these samples. However, to ensure that the structure of the database is dynamic, it must be periodically expanded to incorporate data derived from new analyses, particularly for samples originating from germplasm collections, samples representing particular landraces and samples corresponding to introduced genotypes with greater genetic diversity. A far more conservative measure should be adopted for rare alleles, namely, the exclusion of alleles that occur at low frequencies, to ensure that the representative alleles have been sampled sufficiently and can be considered reliable in statistical tests.

The majority of the commercial cultivars analyzed (74%) exhibited a high degree of varietal purity. Moreover, the analysis of individual plants that were originally analyzed in bulk and exhibited allelic heterogeneity for more than one locus largely revealed mixtures of pure lines. The predominance of homozygous loci in genetically pure cultivars is expected for common bean because the species is normally preferentially autogamous. For those cultivars for which homozygosity for different alleles was observed at more than one gene locus, it

may be important to consider the commitment to varietal purity and the possible need to more closely monitor the propagation of these materials. A likely explanation for this mixture of homozygosity for different alleles in different plants at various loci is outcrossing or mechanical contamination during propagation followed by the fixation of exogenous alleles in some genotypes after successive generations of selfing. Another possible explanation for this heterogeneity is the method used to guide breeding populations during cultivar development. In two cultivars previously analyzed, the few heterogeneous loci identified were attributed to residual heterozygosity present in some individual progeny, which can persist even after several generations of selfing (Borba et al., 2009). However, nine heterogeneous loci were identified in one of the cultivars analyzed in the present study, strongly indicating outcrossing during seed production. In light of the fact that most studies analyze only one individual per cultivar, there have been few studies investigating the genetic purity of cultivars from cultivated autogamous species. Varietal purity tests of seed lots represent one possible approach in these situations, which can also occur in soybean (Schuster et al., 2004). However, once genetic variation is identified, it cannot be directly attributed to a lack of homogeneity in the original genetic material, particularly in light of all of the possible errors during the seed generation process.

The accession grouping dendrogram was consistent with the classification of the analyzed cultivars based on grain type. These findings clearly demonstrated the genetic differentiation among the analyzed cultivars, as the majority of the cultivars (96%) exhibited genotypic variation at three or more of the microsatellite loci evaluated; only the Alba and Magnifico cultivars (Carioca grain type) differed at only two loci. The estimates of genetic diversity for the common bean cultivars were similar to the previously reported estimates of 0.53 to 0.57 from an analysis of 20 SSRs in soybean cultivars; in terms of the ability to discriminate between cultivars with identical characteristics, these reported values were considered high (Diwan and Cregan, 1997). As reported by Priolli et al. (2002) in a study using 12 SSRs to analyze 186 soybean cultivars, a reduced number of SSRs may decrease discriminatory power and reduce diversity estimates. The absence of significant genetic differentiation between the Older and Recent groupings of the common bean cultivars in the present study indicates that more recent cultivars have been developed using the same genetic base in the respective breeding programs. Nevertheless, a clear increase in the number of private alleles was observed in the Recent cultivated germplasm, which may reflect the incorporation of new materials into the breeding programs over time in response to demands for cultivars with distinct characteristics. Classification by institution of origin also showed a lack of structure, i.e., homogeneity of the materials, which is likely attributable to the broad use of CIAT lines, the lines from which the bean cultivars from the various institutional breeding programs were derived.

While the use of multiplexed automated genotyping systems for common bean cultivars is not unprecedented (Masi et al., 2003), the current study derived a series of estimates by characterizing several cultivars and accessions, thereby demonstrating the efficiency of this genotyping system. Distinct molecular profiles for the 50 cultivars analyzed could be generated using 11 of the loci included in the analysis. The genetic diversity estimates obtained for five of these markers are consistent with the published literature (Blair et al., 2006; Buso et al., 2006; Garcia et al., 2011), validating these markers in terms of genetic content and supporting the recommendation to include these loci in an international genotyping system for common bean. The results for the remaining 19 loci, which were highly informative and demonstrated the analytical robustness of the present study, were consistent with previous reports (Garcia et al., 2011) and should also be considered for inclusion in the international genotyping set.

Considering the higher discriminatory power of molecular descriptors, the genotyping system of 24 microsatellite loci evaluated in the present study represents an efficient molecular tool that can be used to complement the morphological descriptors in DUS tests. This advantage is especially important because some cultivars exhibit highly similar morphological characteristics. Thus, the use of a stable and robust descriptor such as a microsatellite marker could significantly enhance the differentiation of morphologically indistinguishable cultivars. Although increasing numbers of SNP (single nucleotide polymorphism) markers are available for common bean, SSRs continue to be the most informative class of markers in genetic diversity studies (Yang et al., 2011). Studies comparing these markers have shown that at least 10 times more SNPs are necessary to generate the same informative estimates of genetic parameters obtained using SSRs (Yu et al., 2009; Van Inghelandt et al., 2010).

CONCLUSIONS

SSRs continue to represent the best method for the accurate estimation of genetic parameters, and the genotyping system based on 24 SSRs used in the present study was found to have high discriminatory and informative power in the genetic analyses. The molecular profile database generated from the analysis can be constantly updated by different laboratories using standardized analyses for the same SSRs and has much potential for use in common bean breeding programs, which would increase the statistical power of the genetic estimates. Additionally, the unique molecular profiles obtained for the accessions analyzed may potentially be used as molecular descriptors to complement the traditional descriptors used in DUS testing. In this way, these profiles may improve the traceability of samples and their derivatives and thereby help to protect the intellectual property rights of breeders.

ACKNOWLEDGMENTS

Research supported by the National Council for Scientific and Technological Development (CNPq; grants to C. Brondani, P.C.B. Cardoso, R.P. Vianello) and EMBRAPA/MP2. We thank Luiz Claudio de Faria for the common bean seed supply.

REFERENCES

- Belkhir K, Borsa P, Chikhi L, Raufaste N, et al (2004). Genetix. Version 4.05.2. Available at [<http://www.univ-montp2.fr/~genetix/genetix/genetix.htm>]. Accessed May, 5 2012.
- Blair MW, Pedraza F, Buendia HF, Gaitan-Solis E, et al. (2003). Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 107: 1362-1374.
- Blair MW, Iriarte G and Beebe S (2006). QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean x wild common bean (*Phaseolus vulgaris* L.) cross. *Theor. Appl. Genet.* 112: 1149-1163.
- Borba TCO, Brondani RP, Rangel PH and Brondani C (2009). Microsatellite marker-mediated analysis of the EMBRAPA Rice Core Collection genetic diversity. *Genetica* 137: 293-304.
- Brondani RP and Grattapaglia D (2001). Cost-effective method to synthesize a fluorescent internal DNA standard for automated fragment sizing. *Biotechniques* 31: 793-5, 798, 800.
- Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, et al. (1999). Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *J. Forensic. Sci.* 44: 1277-1286.
- Buso GSC, Amaral ZPS, Brondani RPV and Ferreira ME (2006). Microsatellite markers for the common bean *Phaseolus vulgaris*. *Mol. Ecol. Note* 6: 252-254.

- Cardoso PCB, Veiga MM, Menezes IPP, Valdisser PA, et al. (2013) Molecular characterization of high performance inbred lines of Brazilian common beans. *Genet. Mol. Res.* 12: 5467-5484.
- Diwan N and Cregan PB (1997). Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor. Appl. Genet.* 95: 723-233.
- Earl DA and vonHoldt BM (2012). Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conser. Genet. Resour.* 4: 359-361.
- Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.
- FAO (Food and Agriculture Organization of the United Nations) (2012). Statistical Database. Available at [<http://www.fao.org/faostat>]. Accessed January 10, 2014.
- FBI (Federal Bureau of Investigation) (2012). Available at [<http://www.fbi.gov/about-us/lab/codis>]. Accessed September 30, 2012.
- Ferreira ME and Grattapaglia D (1998). Introdução ao Uso de Marcadores Moleculares em Análise Genética. Embrapa, Brasília.
- Gaitán-Solís E, Duque MC, Edwards KJ and Tohme J (2002). Microsatellite repeats in common bean (*Phaseolus vulgaris*): Isolation, characterization and cross-species amplification in *Phaseolus* ssp. *Crop Sci.* 42: 2128-2136.
- Garcia RA, Rangel PN, Brondani C, Martins WS, et al. (2011). The characterization of a new set of EST-derived simple sequence repeat (SSR) markers as a resource for the genetic analysis of *Phaseolus vulgaris*. *BMC Genet.* 12: 41.
- Gepts P and Debouck DG (1991). Origin, Domestication, and Evolution of the Common Bean, *Phaseolus vulgaris*. In: Common Beans: Research for Crop Improvement (Voyses O and Van Schoonhoven A, eds.). CAB, Oxon, 7-53.
- Goudet J (2001). FSTAT: Program to Estimate and Test Gene Diversities and Fixation Indices (Software). Version 2.9.3. Available at [<http://www2.unil.ch/popgen/softwares/fstat.htm>]. Accessed May 8, 2012.
- Grisi MC, Blair MW, Gepts P, Brondani C, et al. (2007). Genetic mapping of a new set of microsatellite markers in a reference common bean (*Phaseolus vulgaris*) population BAT93 x Jalo EEP558. *Genet. Mol. Res.* 6: 691-706.
- Harlan JR (1992). Crops and Man. 2nd ed. American Society of Agronomy/Crop Science Society of America, Madison.
- Kalinowski ST, Taper ML and Marshall TC (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16: 1099-1106.
- Kimpton CP, Oldroyd NJ, Watson SK, Frazier RR, et al. (1996). Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification. *Electrophoresis* 17: 1283-1293.
- Kirst M, Cordeiro CM, Rezende GD and Grattapaglia D (2005). Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. *J. Hered.* 96: 161-166.
- Lewis P and Zaykin D (2001). Genetic Data Analysis: Computer Program for the Analyses of Allelic Data. Version 1.1. Available at [<http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php>]. Accessed May 5, 2012.
- Mamidi S, Rossi M, Annam D, Moghaddam S, et al. (2011). Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Funct. Plant Biol.* 38: 953-967.
- MAPA (Ministério da Agricultura, Pecuária e Abastecimento) (2012). Available at [<http://www.agricultura.gov.br/>]. Accessed September 30, 2012.
- Masi P, Spagnoletti Zeuli PL and Donini P (2003). Development and analysis of multiplex microsatellite markers sets in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 11: 303-313.
- Priolli RHG, Mendes-Junior CT, Arantes NE and Contel EPB (2002). Characterization of Brazilian soybean cultivars using microsatellite markers. *Genet. Mol. Biol.* 25: 185-193.
- Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Rohlf FJ (1989). NTSYS-Pc Numerical Taxonomy and Multivariate Analysis System: Version 2.02. Exeter Software, New York.
- Schuster I, Queiroz VT, Teixeira AI, Barros EG, et al. (2004). Determinação da pureza varietal de sementes de soja com o auxílio de marcadores moleculares microsatélites. *Pesq. Agropec. Bras.* 39: 247-253.
- Serrano-Serrano ML, Andueza-Noh RH, Martínez-Castillo J, Debouck DG, et al. (2012). Evolution and domestication of lima bean in Mexico: evidence from ribosomal DNA. *Crop Sci.* 52: 1698-1712.
- Van Inghelandt D, Melchinger AE, Lebreton C and Stich B (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor. Appl. Genet.* 120: 1289-1299.
- Wagner HW and Sefc KM (1999). Identity 1.0. Centre for Applied Genetics. University of Agricultural Sciences, Vienna.
- Weir BS and Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.
- Yang X, Xu Y, Shah T, Li H, et al. (2011). Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. *Genetica* 139: 1045-1054.
- Yu JM, Zhang ZW, Zhu CS, Tabanao DA, et al. (2009). Simulation appraisal of the adequacy of number of background markers for relationship estimation in association mapping. *Plant Genome* 2: 63-77.