Molecular Characterization of Genotypes Selected from the Germplasm Bank of *Cajanus cajan* (L.) Millsp and Cross-species Amplification in Three Legume Species

Adna Cristina Barbosa de Sousa[1], Danilo Augusto Sforça[1], Tatiana de Campos[1], Rodolfo Godoy[2], Letícia Jungmann[3], Liana Jank[3] and Anete Pereira de Souza[4]

Introduction

Cajanus cajan (L.) Millsp (2n = 22) is one of the important grain legume crops in the tropics and subtropics. The centre of diversity is in India (SMARTT, 1990), which accounts for more than 80% of the world's *C. cajan* production. Its seed is primarily consumed as dhal (dried dehusked split cotyledons) and in Latin America the tender green seeds are used as canned peas. This specie has wide adaptability to diverse climates and soils. Because of its multiple uses as food, fodder, fuel wood, rearing lac insects, hedges, windbreaks, soil conservation, roof thatches and green manure, it plays an important role in subsistence agriculture. However, the average world productivity of *C. cajan* is rather low (709 kg ha $^{-1}$) (FAO, 2004), indicating an urgent need for improving the genetic potential of the crop.

Microsatellite markers are of high interest in *C. cajan* breeding programs. They are quite effective to estimate genetic diversity and genetic relationships, as well as to predict the genetic value of selected candidates derived from intraspecific crosses and the performance of their hybrid progenies. Microsatellite markers are short tandem repeat sequence motifs consisting of repeat units of 1-6 base pairs (bp). They are highly polymorphic DNA markers with discrete loci and co-dominant alleles (TAUTZ and SCHLOTTERER, 1994).

The present study reports the characterization of 67 microsatellite markers for investigated a genetic diversity de *C. cajan* and cross-species amplification in other legume species.

Materials and methods

A. Plant material

The set of 77 *C. cajan* genotypes was obtained from the Germplasm Bank of the Brazilian Agricultural Research Corporation (Embrapa) – Sudeste, São Carlos - São Paulo, Brazil. In addition, cross-amplification tests were applied to three legume species: *Phaseolus vulgaris* (CAL-143, IAC-UMA, BAT and JALO), *Phaseolus lunatus* (87-JP-12) and *Vigna* sp (Fradinho), from the Instituto Agronômico de Campinas (IAC) – Campinas, São Paulo, Brazil.

B.DNA extraction and amplification of microsatellite loci

Genomic DNA was extracted from freeze-dried leaf samples following the CTAB method of Doyle and Doyle (1990). DNA samples were quantified by comparison with known quantities of λ phage uncut DNA on 1% agarose gel. The sixty-seven microsatellite loci were analyzed. The Polymerase Chain Reaction (PCR) were performed in 25 µL final volume containing 10 ng of DNA template, 0.8 µM of each forward and reverse primers, 100 µM of each dNTP, 1.5 mM $MgCl_2$, 10 mM Tris-HCl, 50 mM KCl, and 0.5 U *Taq* DNA Polymerase (Invitrogen). PCR reactions were performed using the following conditions: 94°C for 1 min followed by 30 cycles of 94°C for 1 min, specific T*a* for 1 min, 72°C for 1 min, and a final extension of 72°C for 5 min. Amplification products were verified by electrophoresis on 3% agarose gel containing 0.1 mg ethidium bromide/ml in 1X TBE buffer (89 mM Tris-borate, 2 mM EDTA, pH 8.0) and 6% denaturing polyacrylamide gel in 1X TBE buffer, using a 10 bp ladder (Invitrogen) as a standard size. The amplified fragments were visualized in 6.0% polyacrylamide gels silver stained according to Creste et al. (2001).

C. Data analysis

Genotype data was used to calculate the number of alleles at each locus, observed (H_O) and expected (H_E) heterozygosity using the GDA program (LEWIS and ZAYKIN, 2002). The polymorphism information content (PIC) values were calculated to provide an estimate of marker informativeness (CORDEIRO et al., 2003). To compare the efficiency of the markers in varietal identification, we estimated the discriminating power (D) of each locus (TESSIER et al., 1999). The

[1]Adna Cristina Barbosa de Sousa, Danilo Augusto Sforça and Tatiana de Campos are undergraduate of Universidade Estadual de Campinas, CBMEG, Laboratório de Análise Genética e Molecular, Campinas, SP, CEP 13083-970. E-mail: adnacsousa@gmail.com, daniloalbi@gmail.com, tatyunic@gmail.com

[2]Rodolfo Godoy is researcher of Embrapa Pecuária Sudeste (CPPSE), São Carlos, SP. CEP 13060-970. E-mail: godoy@cppse.embrapa.br

[3]Liana Jank and Letícia Jungmann are researchers of Embrapa Gado de Corte, Campo Grande, MS, Caixa postal 154 – CEP 79002-970. E.mail: liana@cnpgc.embrapa.br, leticiajungmann@gmail.com

[4]Anete Pereira de Souza is teacher of Departamento de Genética e Evolução, Instituto de Biologia and researcher of Centro de Biologia Molecular e Engenharia Genética (CBMEG) and Biologia Vegetal of Universidade Estadual de Campinas - UNICAMP/Campinas, SP, CEP 13083-970. E-mail: anete@unicamp.br

Financial support: FAPESP, CNPq and EMBRAPA.

software TFPGA (Tools for Genetic Population Analysis) (MILLER, 1997) was used for calculating expected and observed heterozygosities, and to evaluate the Hardy–Weinberg equilibrium (HWE) using Bonferroni correction. All 77 genotypes were clustered with the Unweighted Pair Group Method using arithmetic average (UPGMA) in the SAHN procedure of the NTSYS-PC v2.1 (ROHLF, 2000). The Neighbor joining trees and the principal coordinates analysis plots were obtained with the software DARwin v. 5.0.157. The software STRUCTURE version 2.2 (PRITCHARD et al., 2000) was used to cluster individuals according to distinct allele frequencies sets.

Results and Discussion

Of the 67 microsatellite markers used, 16 (23,9%) were polymorphic. A total of 83 putative alleles were obtained with these 16 microsatellite polymorphic markers. The number of alleles per locus ranged from 2 to 12, with an average of 5.1 alleles per locus (Table 1). The PIC and D values ranged from 0.11 to 0.80 (average 0.44) and 0.23 to 0.91 (average 0.58), respectively. The highest PIC and D values ware found in locus CZ681938a whith presented 8 alleles. The observed (H_O) and expected heterozygosity (H_E) values ranged from 0.01 to 0.59 (average 0.26) and 0.01 to 0.82 (average 0.44), respectively. Three loci (CZ681930, CZ681938b and CZ681983) showed a significant deviation from Hardy-Weinberg equilibrium (HWE) after Bonferroni correction (P: (5%) $\leq$ 0.0039). No significant linkage disequilibrium was detected (P<0.001) using chi-squared test.

All polymorphic markers were tested for cross-amplification in *Phaseolus vulgaris* (CAL-143, IAC-UMA, BAT and JALO), *Phaseolus lunatus* (87-JP-12) and *Vigna* sp (Fradinho) (Table 2). Thirteen microsatellite loci successfully amplified all different species. Five loci (AJ306901, AJ312891, AJ312893, AJ312895 and CZ681930) were successfully amplified in all species. Non-specific amplification of CZ445531, CZ682017a and CZ682017b, was observed between species. The efficiency of heterologous amplification was 70%. However, this shows a considerable level of sequence, conservation within the primer regions flanking microsatellite loci. These results suggest that the microsatellites here reported the high potential for their use in comparative and phylogenetic studies.

Data from 16 polymorphic loci were used to construct a UPGMA dendrogram showing the genetic relationship among the 77 genotypes. This analysis revealed eleven distinct clusters (Figure 1). Based on the Rogers modified distance all genotypes were successfully differentiated. However, INPA and Fava-Larga were distinct. Genotypes of *C. cajan* fom the same locations in Africa and Brazil did not consistently cluster together (data not shown). The eleven clusters obtained comprised the three different legume species. The genotypes JALO and Fradinho showed 100% similarity to each other. The average genetic similarity value between *C. cajan* genotypes and legume species (*P. vulgaris*, *P. lunatus* and *Vigna* sp) was 0.34%. The cophenetic correlation observed for this clustering pattern was r = 0.71. The Bayesian algorithm used by STRUCTURE software allowed for the identification of four population clusters (K = 4). The phylogenetic (Neighbor-joining) tree, which was constructed based on the genetic distance matrix, was colored according to STRUCTURE results. Furthermore, a strong tendency of correspondence between the Bayesian clusters a neighbor-joining tree was observed (data not shown).

These markers satisfactorily assessed the genetic relationship among the 77 *C. cajan* genotypes. Knowledge of the genetic structure in this species is fundamental in elaborating of further breeding programs and germplasm conservation strategies.

Acknowledgments

We would like to thanks the Fundação de Amparo a Pesquisa de São Paulo (FAPESP) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial support. The authors also thank Dr. Rodolfo Godoy, from Embrapa Pecuária Sudeste, São Carlos - São Paulo, for providing plant tissue samples.

References

CORDEIRO, G. M.; TAYLOR, G. O.; HENRY, R. J. 2003: Characterization of microsatellite markers fromsugarcane (*Saccharum* sp.), a highly polyploidy species. Plant Science v. 155, p.161-168.

CRESTE, S.; TULMANN NETO, A.; FIGUEIRA, A. 2001. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Mol Biol, v. 19, p. 299-306.

DOYLE, J. J. & DOYLE, J. L.1990. Isolation of plant DNA from fresh tissue. Focus v.12, p.13-15.

FOOD and AGRICULTURE ORGANIZATION. 2004 http://www.FAO.ORG/FAOSTAT database.

LEWIS, P. & ZAYKIN, D. 2002. *Genetic Data Analysis (GDA): Computer Program for the Analysis of Allelic Data (Software)*, version 1.1 (d12). http://alleyn.eeb.uconn.edu/gda/.

SMARTT, J. 1990. Grain Legumes. Cambridge University Press, Cambridge, pp.278-293.

MILLER, M. P. 1997. Tools for population genetic analysis (TFPGA) 1.3. A windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author.

PRITCHARD, J. K.., STEPHENS, M., DONNELLY, P. 2000. Inference of population structure using multilocus genotype data. Genetics. v.155 (2), p.945-959.

TAUTZ, D. & SCHLOTTERER, C. 1994: Simple sequences. Curr. Opinion Genetics and Development , v.4, p.832-837.

TESSIER, C.; DAVID, J.; THIS, P.; BOURSIQUOT, J.M.; CHARRIER, A. 1999. Optimizations of the choice of molecular markers for varietal identification in Vitis viniferal. Theoretical and Applied Genetics, v. 98, p. 171-177.

ROHLF, F. J. 2000. NTSYS-pc Numerical Taxonomy and Multivariate Analysis System version 2.1.

Table 1. Characteristics of the sixteen microsatellite markers of *Cajanus cajan* (L.) Millsp.

GenBank Accessions no.	SSR Locos	Repeat motif	Primer sequences (5'–3')	Tm (ºC)	Size range (pb)	A	PIC	D	H_O	H_E	P-HWE
•CZ445531	CCat003	(TA)11	F: TGAATTGCTGAGAGGACGTTT R: CTGTTCCAATTCCACGGTTT	56	234-238	2	0.29	0.36	0.01	0.01	0.0096^{NS}
•CZ445540	CCggt001	(GGT)4	F: ACGCTTCTGATGCTGTGTTG R: CATCAGCATCATCGTTACCC	45	208-210	2	0.29	0.35	0.35	0.29	0.0749^{NS}
•CZ445530	CCggc001	(GGC)4	F: CCATTGTGCGTCTTTGTGTT R: GCTTTTCCTCTTCCTTTCTCG	56	206-208	2	0.11	0.25	0.01	0.11	0.0410^{NS}
*AJ306901	CCB1	(CA)10	F: AAGGGTTGTATCTCCGCGTG R: GCAAAGCAGCAATCATTTCG	56	186-202	4	0.32	0.41	0.19	0.30	0.0602^{NS}
*AJ312887	CCB2	(CA)21	F: CCATAATCCAATCCAAATCC R: AGAAGGCTTTCATGTAACGC	51	160-170	4	0.51	0.62	0.21	0.48	1.0000^{NS}
*AJ312891	CCB6	(CA)6	F: ACAATGCTAGGGAACACCGC R: TACCTTAACCCACAATGGCC	45,5	180-206	4	0.32	0.43	0.30	0.26	1.0000^{NS}
*AJ312892	CCB7	(CT)16	F: CAACATTTGGACTAAAAACTG R: AGGTATCCAATATCCAACTTG	56	150-158	5	0.49	0.56	0.27	0.51	0.0201^{NS}
*AJ312893	CCB8	(CT)30	F: TGCGTTTGTAAGCATTCTTCA R: ACTTGAGGCTGAATGGATTTG	50	126-150	12	0.61	0.72	0.43	0.54	0.0108^{NS}
*AJ312894	CCB9	(CT)22	F: CACTTGGTTGGCTCAAGAAC R: GCCAATGAACTCACATCCTTC	45	152-180	7	0.69	0.73	0.37	0.71	1.0000^{NS}
*AJ312895	CCB10	(CA)15	F: CCTTCTTAAGGTGAAATGCAAGC R:ATAACAATAAAAGACCTTGAATGC	45	228-242	5	0.60	0.75	0.32	0.62	0.6069^{NS}
+CZ681930	—	(TC)8	F: GCGCTAAGGGAAAACAAAAA R: AACTCCCTTGTTGTCATATGGTG	56	164-174	6	0.66	0.78	0.32	0.66	0.0008*
+CZ681938a	—	(ATT)21	F: TCAGGGGTAAATGCGGTATC R: GAATTGCTTTTTGCTTCCTCA	50	236-260	8	0.80	0.91	0.56	0.82	0.0204^{NS}
+CZ681938b	—	(ATT)21	F: TCAGGGGTAAATGCGGTATC R: GAATTGCTTTTTGCTTCCTCA	50	212-234	8	0.68	0.78	0.31	0.72	0.0035*
+CZ682017a	—	(AAG)13	F: TGAAATGAACAAACCTCAATGG R: TGTATTGCACATTGACTTGGCTA	45	200-222	8	0.74	0.82	0.21	0.77	0.0204^{NS}
+CZ682017b	—	(AAG)13	F: TGAAATGAACAAACCTCAATGG R: TGTATTGCACATTGACTTGGCTA	45	174-182	2	0.26	0.35	0.01	0.01	0.0096^{NS}
+CZ681983	—	(TGA)11	F: GAGGAGGAGGAAGAAGAAGAAGA R: TCGTCGCCGTATCACTACAA	45,5	73-79	4	0.48	0.57	0.25	0.54	0.0030*

*Burns et al. 2001; +Odney DA. 2006; •Odeny et al. 2007; F, forward sequence; R, reverse sequence; Tm, annealing temperatures; A, total number of alleles per locus; PIC, polymorphism information content; D, discriminating power; *Ho*, observed heterozygosity; *He*, expected heterozygosity; *P*-HWE, *P* values for the HWE test, significance threshold adjusted using Bonferroni correction: $P(P\%) \leq 0.0039$; ^{NS}, not significant, *, significant.

Table 2. Cross-species amplification of *Cajanus cajan* (L.) Millsp microsatellite markers in three common bean species.

	Phaseolus vulgaris				***Phaseolus lunatus***	***Vigna sp.***
Primers	**Cal-143**	**IAC-UNA**	**BAT**	**JALO**	**87-JP-12**	**Fradinho**
CZ445531	-	-	-	-	-	-
CZ445540	+	+	+	+	-	+
CZ445530	+	+	+	+	+	-
AJ306901	+	+	+	+	+	+
AJ312887	-	+	+	+	-	+
AJ312891	+	+	+	+	+	+
AJ312892	+	+	+	-	-	-
AJ312893	+	+	+	+	+	+
AJ312894	-	+	+	+	-	+
AJ312895	+	+	+	+	+	+
CZ681930	+	+	+	+	+	+
CZ681938a	+	+	+	-	+	+
CZ681938b	+	+	+	-	+	+
CZ682017a	-	-	-	-	-	-
CZ682017b	-	-	-	-	-	-
CZ681983	+	+	+	+	+	+

+ - Amplificação; – Não amplificação

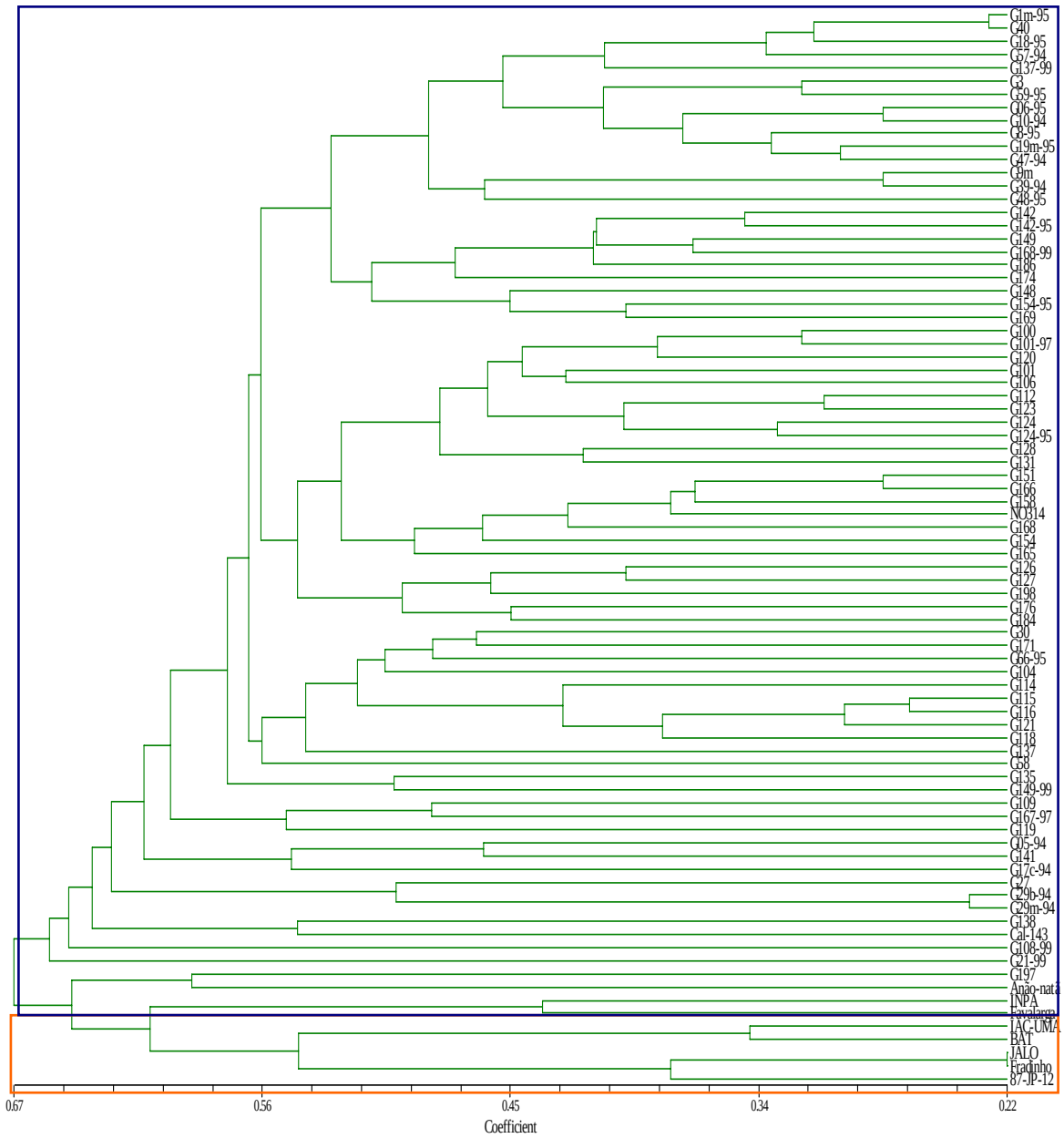

Figure 1. Unweighted pair group method analysis (UPGMA) dendrogram for 77 genotypes *of Cajanus cajan* and three legume species, using the Rogers modified distance.