Plant Breeding 126, 130—136 (2007)
Journal compilation © 2007 Blackwell Verlag, Berlin
No claim to original US government works
doi:10.1111/j.1439-0523.2007.01324.x

Development, characterization and utilization of microsatellite markers in pigeonpea

D. A. Odeny[1], Jayashree B.[2], M. Ferguson[3], D. Hoisington[2], J. Crouch[4] and C. Gebhardt[5]

[1]Centre for Development Research (ZEFc), University of Bonn, Walter-Flex Str.3, D-53113 Bonn, Germany, E-mail: odeny@mpiz-koeln.mpg.de; [2]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502 324, Hyderabad, India; [3]International Institute of Tropical Agriculture (IITA), PO Box 30709, Nairobi, Kenya; [4]International Maize and Wheat Improvement Centre (CIMMYT), CP 56130, Mexico; [5]Max-Planck Institute for Plant Breeding Research (MPIZ), D-50829 Köln, Germany

With 1 figure and 5 tables

Received May 5, 2006/Accepted August 15, 2006
Communicated by C. Jung

Abstract

Pigeonpea is a major legume of the semi-arid tropics that has been neglected in terms of molecular breeding. The objectives of this study were to develop microsatellite markers and evaluate their potential for use in pigeonpea genetics and breeding. Two hundred and eight microsatellite loci were isolated by screening a non-enriched partial genomic library. Primers were designed for 39 microsatellite loci, 20 of which amplified polymerase chain reaction products of the expected size. Nineteen of the primer pairs were polymorphic amongst 15 cultivated and nine wild pigeonpea accessions providing evidence for cross-species transferability within the genus *Cajanus*. A total of 98 alleles were detected at the 19 polymorphic loci with an average of 4.9 alleles per locus. The observed heterozygosity ranged from 0.17 to 0.80 with a mean of 0.42 per locus. Less allelic variation (31 alleles) was observed within the cultivated species than across the wild species (92 alleles). The diversity analysis readily distinguished all wild relatives from each other and from the cultivated germplasm. Development of more microsatellites is recommended for future genomic studies in pigeonpea.

Key words: *Cajanus cajan* — microsatellites — Simple Sequence Repeat — molecular markers

Pigeonpea [*Cajanus cajan* (L.) Millsp.] is a grain legume belonging to the Cajaninae sub-tribe of the economically important leguminous tribe Phaseoleae. The tribe Phaseoleae also contains soybean (*Glycine max* L.), common bean (*Phaseolus vulgaris* L.) and mungbean (*Vigna radiata* L. Wilczek; Young et al. 2003). Pigeonpea is the only cultivated food crop of the Cajaninae sub-tribe and has a diploid genome (2n = 22) estimated at about 0.853 pg (Greilhuber and Obermayer 1998). India is the largest pigeonpea producer followed by Myanmar and Kenya (FAOSTAT data 2006). Pigeonpea has become increasingly important in recent years because of its inherent ability to perform well under marginal input systems and to withstand a range of environmental stresses including drought.

Despite the existence of substantial variability among pigeonpea landraces and varieties for various traits, no effective molecular breeding programme has been developed to facilitate its improvement. The genetic basis of most important traits is not known and, to date, no mapping strategies have been reported. With the current interest in the genetic potential of wild relatives (Sharma et al. 2003) and the recent introduction of cytoplasmic male sterility (CMS) technology into cultivated genotypes (Saxena and Kumar 2003), there is an urgent need to develop a robust set of molecular markers in pigeonpea.

Microsatellites (Tautz and Rentz 1984), also known as Simple Sequence Repeat (SSR) markers, are DNA-based molecular markers that offer several advantages because they are reproducible, polymorphic, polymerase chain reaction (PCR)-based and readily portable within a species (Edwards et al. 1996). Only 20 SSRs have been previously reported in pigeonpea, of which only half were polymorphic in cultivated pigeonpea germplasm (Burns et al. 2001). In contrast, more than 1000 microsatellites have been mapped in soybean [*Glycine max* (L.)] (Song et al. 2004), and several hundreds are available in chickpea (*Cicer arietinum* L.; Lichtenzveig et al. 2005), common bean (*Phaseolus vulgaris* L.; Blair et al. 2003) and groundnut (*Arachis hypogaea* L.; Ferguson et al. 2004). In this report, the isolation and characterization of additional 20 SSR markers from pigeonpea and their utilization in diversity analysis of *Cajanus* germplasm is described.

Materials and Methods

Plant material and DNA extraction: Genomic DNA from accession ICP 2376 of pigeonpea, *Cajanus cajan* (L.) Millsp., was extracted and purified as described by Oberhagemann et al. (1999) for the development of a small insert library. All genotypes used in the diversity analysis are described in Table 1. Seeds of these accessions were obtained from the ICRISAT GenBank, India. The cultivated genotypes were selected on the basis of one representative accession from each of the major clusters (based on phenotypic analysis) across the entire diversity of pigeonpea core collection (Reddy et al. 2005). DNA from all cultivated pigeonpea accessions was extracted using a cetyltrimethyl-ammonium-bromide (CTAB) method (Murray and Thompson 1980) with modifications as described by Mace et al. (2003). Genomic DNA from the wild species was extracted as described by Sivaramakrishnan et al. (1997).

Genomic DNA library preparation: About 5 μg genomic DNA were partially digested with *Tsp*5091 (▼AATT) and size-fractioned on a 1.5% agarose gel. Fragments between 700 and 1000 bp were recovered using the Qiagen Gel Extraction Kit (Qiagen, Hilden, Germany). The vector pBluescript (Stratagene, La Jolla, CA, USA) was isolated from *Escherichia coli* as described by Sambrook et al. (1989). Plasmid DNA was digested with *Eco*R1 and dephosphorylated using Shrimp Alkaline Phosphatase (Roche, Mannheim, Germany). Recovered pigeonpea DNA fragments were ligated into the vector, precipitated and transformed into DH-10B cells (Invitrogen, Carlsbad, CA, USA) by

Table 1: Accessions used to characterize polymorphism

Accession	Species	Gene pool[1]/description	Country of origin
ICPW 1	*Cajanus acutifolia*	Secondary	Australia
ICPW 13	*C. albicans*	Secondary	India (Karnataka)
ICPW 28	*C. cajanifolia*	Secondary	India (Madhya Pradesh)
ICPW 39	*C. latisepala*	Secondary	Australia
ICPW 41	*C. lineata*	Secondary	India (Tamil Nadu)
ICPW 68	*C. platycarpa*	Tertiary	India (Uttar Pradesh)
ICPW 74	*C. reticulata*	Secondary	Australia
ICPW 119	*C. carabaeoides*	Secondary	Philippines
ICPW 162	*C. sericea*	Secondary	Australia
ICP 9267	*C. cajan*	Landrace	Guyana
ICP 11181	*C. cajan*	Landrace	India (Andhra Pradesh)
ICP 14144	*C. cajan*	Landrace	Jamaica
ICP 13575	*C. cajan*	Landrace	Sierra Leone
ICP 15145	*C. cajan*	Landrace	Zaire
ICP 9266	*C. cajan*	Landrace	Guyana
ICP 4167	*C. cajan*	Landrace	India (Uttar Pradesh)
ICP 14576	*C. cajan*	Landrace	Thailand
ICP 12058	*C. cajan*	Landrace	Tanzania
ICP 14352	*C. cajan*	Landrace	Venezuela
ICP 1514	*C. cajan*	Landrace	India (Andhra Pradesh)
ICP 7543	*C. cajan*	Landrace	India (Madhya Pradesh)
ICP 7852	*C. cajan*	Landrace	India (Karnataka)
ICPL 87119	*C. cajan*	Advanced cultivar	ICRISAT Breeding Line
ICP 2376	*C. cajan*	Landrace	India (Andhra Pradesh)

[1]Secondary: wild species from secondary gene pool, easily hybridize with pigeonpea; tertiary: wild species from tertiary gene pool, more distant from cultivated pigeonpea and do not easily hybridize.

electroporation at 1.8 KV/cm. Transformed cells were incubated at 37°C for 1 h with vigorous shaking at 250 rpm before culturing on X-Gal/IPTG/ampicillin LB-agar plates for blue/white selection.

White clones were carefully scraped with sterile toothpicks and used as templates for colony PCR using T7 and T3 primers in a total volume of 25 µl. Each PCR contained 5 pmol of dNTP, 0.4 µm of each primer, 1.5 units of *Taq* polymerase, 1.6 mm $MgCl_2$ and 1X buffer (Invitrogen). Clone insert lengths were confirmed through 1.2% agarose gel electrophoresis using 5 µl of PCR product. Selected colonies were PCR amplified and subjected to 1.2% agarose gel electrophoresis at 100 V for 1 h. The DNA was denatured by incubating the gel for 30 min in 0.4 m NaOH and then transferred overnight on to Hybond N^+ filters (Boehringer-Mannheim, Mannheim, Germany). The filters were air dried and DNA covalently cross-linked using Stratalinker (Stratagene, Heidelberg, Germany) at 1200 µJ × 100. The filters were further dried at 80°C for 1 h and stored aseptically at room temperature.

Library Screening and sequencing: Synthetic oligonucleotide repeats $(CA)_{15}$, $(CT)_{15}$, $(AAT)_{10}$, $(GCC)_{10}$, $(CAA)_{10}$ (Qiagen) were end-labelled with γ-(^{32}P) dATP using T4 polynucleotide kinase (New England Biolabs, GmbH, Frankfurt, Germany) (Sambrook et al. 1989). The membranes were prehybridized for 2 h in 10X SSPE (0.18 m NaCl, 0.01 m NaH_2PO_4, 0.001 m Na_2EDTA), 100X Denhardt's (2% bovine serum albumin, 2% Ficol, 2% polyvinylpyrrolidone) solution, 20% sodium dodecyl sulphate (SDS) and 100 µg/ml of herring sperm DNA to minimize unspecific binding. Hybridization was carried out for 16 h using a similar buffer to that in prehybridization but this time incorporating the labelled oligonucleotides. The temperatures were maintained at 63°C for $(CA)_{15}$ and $(CT)_{15}$, 65°C for $(GCC)_{10}$, 55°C for $(CAA)_{10}$ and 43°C for $(ATT)_{10}$. Posthybridization washes were carried out in two steps at 40°C with stringent Saline Sodium Citrate (SSC) buffer washes. The filters were exposed to X-ray films (Eastman Kodak Company, Rochester, NY, USA) with intensifying screens overnight and thereafter processed using a Kodak M35A X-OMAT automated processor.

Positive clones were purified for sequencing using EXOSAP (Amersham Biosciences, Freiburg, Germany) according to the manufacturer's instructions. Approximately 25–50 ng of each amplicon was used as a template for DNA sequencing. DNA sequences were determined by the Max-Planck Institute for Plant Breeding Research DNA core facility (ADIS) on Applied Biosystems (Weiterstadt, Germany) ABI Prism 377, 3100 and 3730 sequencers using Big Dye-terminator v3.1 chemistry. Applied Biosystems supplied premixed reagents. A total of 208 genomic DNA clone inserts were sequenced in both directions using T7 and T3 (Metabion Martinsried, Germany) primers.

Sequence analysis and primer design: All raw sequences were trimmed using SEQUENCHER v4 (Gene Codes Corp, Ann Arbor, MI, USA) software and redundant sequences determined using CAP3 (Huang and Madan 1999). Simple Sequence Repeat Identification Tool (SSRIT; Temnykh et al. 2001) was used to identify SSRs present in all non-redundant sequences. Primers were designed for each SSR locus using PRIMER 3 (Rozen and Skaletsky 1998) and confirmed using NETPRIMER (http://www.premierbiosoft.com/netprimer/index.html). Thirty-nine primer pairs were designed and synthesized by GENOMECHANIX (Florida, USA).

PCR optimization and amplification: Polymerase chain reaction conditions were optimized for each primer pair using a modified Taguchi method (Cobb and Clarkson 1994) as described by Buhariwalla et al. (2005). Three concentrations each of primer (0.2, 0.3 and 0.5 pm), Mg^{++} (1.0, 1.5 and 2.0 pm) and enzyme (0.2, 0.3 and 0.5 U), and two concentrations each of DNA (5 and 10 ng) and dNTP (0.1 and 0.15 mm) were varied across five different

Table 2: Temperature ranges and reactions that led to amplification of products

Reaction[1]	Primer (pmol)	DNA (ng)	Mg^{++} (mm)	Enzyme (U)	dNTP (mm)	Number of primers amplified with 'Touchdown' programmes (°C)		
						55–45	60–55	65–60
1	0.2	5	1	0.2	0.1	0	7	1
2	0.2	10	2	0.2	0.15	1	6	0
3	0.3	5	1.5	0.5	0.15	0	0	0
4	0.3	10	2	0.3	0.1	2	0	0
5	0.5	10	1	0.5	0.15	0	2	1

[1]Total reaction volume = 10 µl.

protocols (Table 2). Amplifications were performed in a final volume of 10 μl with 1X reaction buffer (BioLine, London, UK). Three different programmes of 'Touchdown' PCR (Don et al. 1991) with base annealing temperature ranging from 55 to 45°C, 60 to 55°C, and 65 to 60°C and optimum annealing temperatures of 48°C, 56°C and 59°C, respectively, were used to reduce spurious amplification. PCR products were separated on 6% non-denaturing polyacrylamide gels for 3 h at 600 V and visualized by silver staining. All amplifying SSR markers were named following the nomenclature by Yu et al. (2000) and screened across all genotypes listed in Table 1.

Statistical analysis: The presence or absence of each band was determined and designated '1' if present and '0' if absent for each genotype. Markers that produced the expected size (100–500 bp) of amplification product were recorded and the polymorphism information content (PIC) calculated as described by Botstein et al. (1980) using the formula:

$$\mathrm{PIC} = 1 - \left[\sum_{i=1}^{n} p_i^2\right] - \left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2p_i^2 p_j^2\right],$$

where p_i equals the frequency of the ith allele and p_j the frequency of the $(I + 1)$th allele. Only data from polymorphic SSR loci were used for diversity analysis. Genetic similarities between any two genotypes were estimated according to Nei and Li (1979). All 24 genotypes were clustered with the Unweighted Pair Group Method using arithmetic average (UPGMA) in the SAHN procedure of the NTSYS-PC v2.10t (Rohlf 1994).

Results

Sequence characterization

A total of 753 colonies were screened, of which 208 positive clones were detected after hybridization and sequenced in both directions. The average length of these sequences was 963 bp. Forty-eight unique SSRs were identified including dinucleotide, trinucleotide, tetranucleotide and hexanucleotide repeats (Table 3). Twelve of the 48 sequences had insufficient flanking regions for primer design, reducing the number of sequences to which primers could be designed to 36. Thirty-nine primer pairs were designed with three sequences containing two different SSR motifs. Five of the motifs identified were tetranucleotides and hexanucleotides even though these were not probed for. The longest repeat motif was an AT with 17 uninterrupted repeats. The average number of repeats was 10 for dinucleotide, five for trinucleotide, and four each for tetranucleotide and hexanucleotide repeats. Most of the microsatellite motifs were perfect repeats except three dinucleotide repeats.

Most amplifying SSR markers in this study required very low primer concentrations providing optimum amplification through either reaction 1 or 2 (Table 2). Both reactions 1 and 2 made use of only 0.2 pmol (the lowest primer concentration) of primer while reaction 2 used double concentrations of DNA, Mg^{++} and dNTP. The 'Touchdown' programmes at 55–45°C and 65–60°C were not useful as they only amplified three and two primers, respectively. Increasing DNA concentrations without varying primer and enzyme concentrations was enough to improve amplification, provided Mg^{++} and dNTP concentrations were proportionately increased. All the primer sequences that amplified expected products and the respective optimized PCR conditions are shown in Table 4.

SSR marker analysis of diverse pigeonpea germplasm

Twenty primer pairs, all amplifying no more than two bands when tested on accession ICP 2376, were selected to assess polymorphism in 15 cultivated and nine wild pigeonpea relatives (Table 1). All markers detected at least one allele in all genotypes tested suggesting transferability for all markers across the genus. Nearly all amplifying SSR markers detected polymorphisms amongst the 24 diverse accessions (except CCtta001; Table 5). For the 20 SSR loci, a total of 98 alleles were amplified with an average number of 4.9 alleles per locus. The PIC value ranged from 0.17 to 0.80 with five SSR markers having a PIC value more than 0.5 (Table 5). Most of the polymorphic SSR loci contained dinucleotide (35%) and trinucleotide (50%) repeats and only three (15%) had tetranucleotide and hexanucleotide repeats.

Less genetic variation (Fig. 1) was detected within the cultivated species, with only nine markers detecting polymorphism and a total of 31 alleles. The average number of alleles was 3.4 with an average PIC of 0.39 per locus. Loci CCat002, CCtta002, CCtacccg001 and CCgtt001 contained five unique alleles out of the 31 alleles detected in cultivated species. Two of the unique alleles were from locus CCtta002 and were only found in the landrace accession ICP 13575. Expectedly, there was substantially more variation within the wild species (Fig. 1) with 19 polymorphic loci and 92 different alleles. The average number of alleles was 4.8 with an average PIC value of 0.60. A total of 56 alleles were unique to the wild species.

Data from all the polymorphic SSRs (19) were used for diversity analysis of the 24 *Cajanus* genotypes. Within the landraces studied, ICP 7543 and ICP 14144 showed the highest genetic similarity (98%) even though the two landraces were collected from India and Jamaica, respectively. This level of similarity could be indicative of germplasm exchange between

Table 3: Characterization of microsatellite clones

Repeat group	Repeat family	SSRs identified	Primers designed	Primers amplifying ICP 2376	Polymorphic SSRs[1]
Dinucleotides	AT	10	8	3	3
	TC/TG	11	6	3	3
	TA + TG	1	1	1	1
Trinucleotides	TTA	11	10	5	4
	CAT/CAA	4	4	3	3
	GGT/GGA	3	3	1	1
	GGC/GCC	3	2	1	1
Tetranucleotides	CCTC	1	1	–	–
	GAAA	1	1	–	–
	TTAT	1	1	1	1
Hexanucleotides	TACCCG	2	2	2	2
	Total	48	39	20	19

[1]Simple Sequence Repeats (SSRs) polymorphic in both cultivated and wild accessions.

Table 4: Pigeonpea microsatellite motifs, primer sequences and optimized PCR conditions

GenBank accession number	SSR name	Motif	Primer sequences (5′–3′)	Temperature (°C)	Reaction
CZ445530	CCttc001	(ttc)5	F: Cgggcttccttttcttctct R: Aaaaccccgaaaacaccatt	60–55	1
CZ445525	CCtta001	(tta)10	F: Ttctggatccctttcatttttc R: Tgacacccttctaccccataa	60–55	1
CZ445522	CCat001	(ta)8	F: Cttcccccaactaagatcca R: Gttcgttctctttaattgacttgc	65–60	5
CZ445523	CCat002	(ta)10(tg)9	F: Tttcctgagccatcagtcg R: Aagcatcaacgtaccagcaa	55–45	4
CZ445531	CCat003	(ta)11	F: Tgaattgctgagaggacgttt R: Ctgttccaattccacggttt	60–55	1
CZ445520	CCtta002	(tat)9	F: Cccatttagtgagggttaat R: Gactactccaggtcaaacacg	55–45	4
CZ445535	CCtacccg001	(tacccg)4	F: Gtcggggcgtgtaagtcata R: Ccgaaataaggatggcaaat	55–45	2
CZ445536	CCtacccg002	(cgggta)4	F: Gtctttgagggacggaacc R: Ggggcggggaaagtacata	60–55	5
CZ445538	CCtta003	(tta)4	F: Ccaagaaaaggtgctccaagt R: Ttgcttcttttctcgcttgc	60–55	2
CZ445539	CCcat001	(cat)4	F: Tgatagggaccacaacgaca R: Agcgttgactcctccctctt	60–55	2
CZ445540	CCggt001	(ggt)4	F: Acgcttctgatgctgtgttg R: Catcagcatcatcgttaccc	60–55	2
CZ445519	CCtc001	(ct)6tt(ct)2	F: Gactcttcacctcacactcatcac R: Acctcatacaacaaccctaagcac	60–55	2
CZ445530	CCggc001	(ggc)4	F: Ccattgtgcgtctttgtgtt R: Gcttttcctcttcctttctcg	60–55	1
CZ445544	CCttat001	(ttat)4	F: Tacagcagccacatcaaagc R: Tgaaccgtgaaagtgggatt	65–60	1
CZ445553	CCtta004	(tta)4	F: Acccattattgatttgggta R: Ccaaatttcacccaagaaa	60–55	2
CZ445545	CCtta005	(aat)4	F: Tcttccattgcatggtgtt R: Gcatgatatgagatgatgacga	60–55	2
CZ445524	CCac001	(tg)6	F: Ctgggcctctagcatagcaa R: Aaacttctggacgcaaaatga	60–55	1
CZ445555	CCtc002	(ga)12	F: Ggaaaaccccgagacaaaag R: Gggcaacccataaaccctaa	60–55	5
CZ445554	CCgtt001	(aac)4	F: Ataggcccatctccaggttc R: Ttaatgcccagccaattctt	60–55	1
CZ445521	CCat004	(ta)4(gatag)(at)4	F: Ctacaatcccagggaaaagg R: Aacaaacgtaatctgtgttgatctc	60–55	1

SSR, Simple Sequence Repeat; PCR, polymerase chain reaction.

Table 5: Properties of pigeonpea microsatellite loci, including range of amplification products sizes, number of alleles and PIC values

Repeat group	SSR name	Allele size (bp)	N_ac	N_aw	$N_ac + w$	PIC_c	PIC_w	PIC_{c+w}
Dinucleotide (perfect)	CCat001	280–340	5	7	7	0.64	0.78	0.69
	CCac001	250–290	1	3	3	0	0.45	0.17
	CCtc002	185–200	3	3	3	0.47	0.50	0.49
	CCat003	265–300	1	5	5	0	0.74	0.48
Dinucleotide (compound)	CCtc001	160–225	1	6	6	0	0.73	0.49
	CCat002	355–500	6	9	10	0.47	0.82	0.76
	CCat004	210–220	3	3	3	0.21	0.31	0.26
Trinucleotide	CCttc001	185–220	1	9	9	0	0.85	0.57
	CCtta001	220	1	1	1	0	0	0.00
	CCtta002	240–320	5	3	5	0.56	0.55	0.58
	CCgtt001	170–190	2	3	4	0.19	0.41	0.31
	CCtta003	180–190	1	3	3	0	0.53	0.28
	CCcat001	155–185	1	3	3	0	0.44	0.21
	CCggt001	205–210	1	2	2	0	0.16	0.07
	CCggc001	200–260	2	5	5	0.12	0.7	0.49
	CCtta004	250–260	1	4	4	0	0.66	0.44
	CCtta005	245–290	1	6	6	0	0.73	0.39
Tetranucleotide	CCttat001	210–250	1	6	6	0	0.75	0.43
Hexanucleotide	CCtacccg001	170–230	2	9	10	0.37	0.83	0.80
	CCtacccg002	305–320	3	3	3	0.46	0.46	0.47

N_ac, number of alleles in cultivated genotypes; N_aw, number of alleles in wild relatives; $N_ac + w$, total number of alleles in both wild and cultivated; PIC, polymorphism information content; PIC_c, PIC for cultivated genotypes; PIC_w, PIC for wild relatives; PIC_{c+w}, total PIC for both wild and cultivated; SSR, Simple Sequence Repeat.

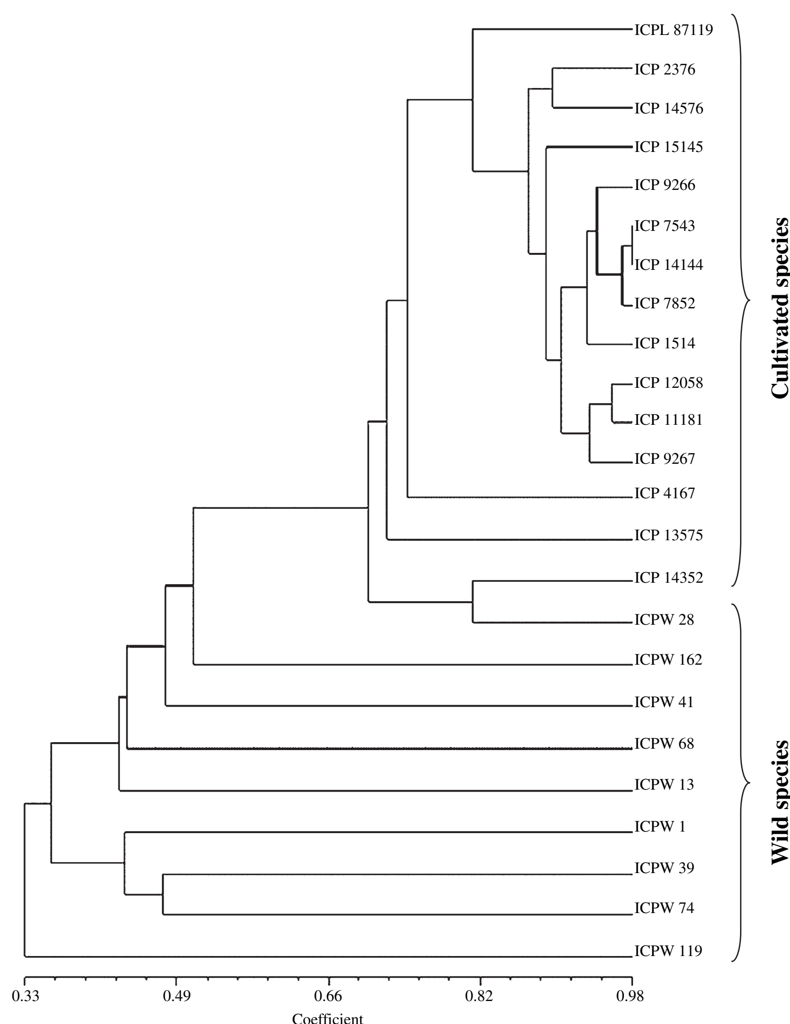

Fig. 1: Unweighted Pair Group Method using arithmetic average dendrogram showing relatedness among the 24 *Cajanus* genotypes. The scale at the bottom of the dendrogram indicates the level of similarity between the genotypes

the two geographical regions. The smallest genetic distance and hence the largest similarity coefficient among the various species of *Cajanus* studied was between *C. cajanifolius* and cultivated pigeonpea. This revealed a very close relationship between the two species consistent with the results of Sivaramakrishnan et al. (2002) using mtDNA restriction fragment length polymorphic analysis. *Cajanus cajanifolius* has been reported as the nearest wild relative of pigeonpea (Van der Maesen 1990). Morphologically, *C. cajanifolius* resembles pigeonpea in all traits except the presence of a prominent strophiole (Van der Maesen 1990).

Discussion

The present study doubles the amount of pigeonpea SSR markers available which, together with those reported earlier (Burns et al. 2001), will be useful in genetic diversity studies and ultimately for marker-assisted breeding. The identification of 6.4% positive clones compares well with results from similar studies (Squirrell et al. 2003) even though the high percentage of false positives (77%) could have been further eliminated by using longer oligonucleotide probes. Longer probes have been reported to favour the isolation of relatively long arrays and also to eliminate mismatches (Armour et al. 1994).

Employing an enrichment process is a recommended alternative (Edwards et al. 1996, Panaud et al. 1996, Billotte et al. 1999) towards enhanced microsatellite isolation efficiency. However, this approach suffers redundancy problems especially when PCR amplification of selected fragments is done prior to cloning (Rallo et al. 2000, Mba et al. 2001). Indeed, previous efforts to develop microsatellite markers in pigeonpea (Burns et al. 2001) did employ an enrichment process but experienced a similar efficiency level. Several other enrichment techniques are now available (Zane et al. 2002) and if redundancy could be eliminated, these procedures are likely to yield more microsatellites in pigeonpea.

The most abundant motifs in this study were AT based (AT, AAT, TTAT) followed by TC class of repeats. AT-based motifs have been reported to be the most abundant in plants (Morgante and Olivieri 1993, Cardle et al. 2000, Morgante et al. 2002) even though previous studies have excluded them due to the problems with self-complementarity and difficulties in amplification (Su et al. 1996). The amplification of these motifs in the current study was greatly enhanced by optimization making them equally good sources of microsatellites. TC class of repeats have been reported to be equally abundant in other legumes (Ferguson et al. 2004, Lichtenzveig et al. 2005) and plants in general (Wang et al. 1994).

This study also investigated a number of factors that affect SSR amplification in order to devise an optimal PCR procedure for each primer designed. Some studies have reported that modifications in PCR conditions led to insignificant improvement in band interpretability (Ashworth et al. 2004). Here, it was observed that varying 'Touchdown' programmes greatly

improved band clarity and hence ease of scoring. However, there does not appear to be any additional value in increasing primer and enzyme concentrations in as much as it is in increasing DNA concentration. The advantage of being able to optimize the enzyme and primer concentrations has great cost reduction implications for similar projects in the future.

The average PIC among dinucleotide repeats was 0.48 while that of trinucleotide repeats was 0.33. There are similar reports of higher levels of polymorphism for dinucleotide repeats than trinucleotide repeats for tomato (*Lycopersicon esculentum*; He et al. 2003) and in avocado (*Persea americana* Mill.; Ashworth et al. 2004). Dinucleotide repeats have been reported to reside outside coding regions of genes (Temnykh et al. 2001) and are characterized with more repeat numbers (Li et al. 2004) making them the best source of highly polymorphic SSR markers. In contrast, trinucleotide repeats are more abundant in protein coding regions (Tóth et al. 2000) with relatively small repeat numbers and total length (Thiel et al. 2003). Despite their frequent polymorphism, the use of dinucleotide repeats may be limited by the occurrences of stutter bands in the amplification products that may lead to incorrect scoring of alleles (Ashworth et al. 2004, Wang et al. 2005).

Although longer repeats have been linked to higher degrees of polymorphism in the past, this new study did not find any clear relationship in this respect consistent with reports in common bean (Yu et al. 2000). Theoretically, the number of repeats is correlated with the mutation rate (Li et al. 2002) meaning more recently evolved microsatellites would have fewer polymorphisms because of fewer occasions for mutations even if they have longer repeats (Budak et al. 2003). While other studies (Budak et al. 2003) have discarded SSRs with < 10 repeats, some of the shortest SSRs (four repeats) in the current study were highly polymorphic detecting up to 10 alleles per locus.

Transferability of all markers tested across all species screened was 100%. The amplification of SSR loci across species of the target genus appears to be widespread in plants (Moretzsohn et al. 2004, Wang et al. 2004, Zou et al. 2004). The possibility of using the same microsatellite flanking primers in more than one species is encouraging in view of the efforts and expense involved in designing them. Results also demonstrate that pigeonpea SSRs will be a valuable source of molecular markers for analysing the relatively unknown *Cajanus* gene pool and for studying gene introgression between the various species.

The allelic variation observed at microsatellite loci varies greatly between different plant species. Previous diversity analysis of cultivated pigeonpea species reported a range of 2–5 alleles per locus for 10 polymorphic loci (Burns et al. 2001), which is comparable with the present study. This is relatively lower than in other legume crops such as soybean (11–26 alleles per locus; Rongwen et al. 1994). Ninety-eight different alleles have been observed for the 20 loci with over 50% of alleles (56) unique to wild species. The low genetic variability amongst cultivars when compared with the wild suggests that natural and artificial selection has contributed to the selection of specific alleles and to changes of allelic frequencies at specific loci. Similar findings have been reported in groundnut (Moretzsohn et al. 2004) and mulberry (*Morus* L.; Zhao et al. 2005).

The UPGMA dendrogram (Fig. 1) obtained by the analysis of these SSR data suggests their potential use in diversity analysis. The current results clearly demonstrate that pigeonpea SSRs constitute efficient sources of molecular markers for other *Cajanus* species. With the ongoing breeding emphasis on development of hybrid pigeonpea using a source of CMS derived from wild species (Saxena and Kumar 2003), the markers developed here and previously by Burns et al. (2001) will be useful for confirmation of hybridity at an early stage and preliminary mapping of fertility restoration genes. Nevertheless, future studies should incorporate the use of several accessions within each species.

Acknowledgements

The authors thank Dr Agim Ballvora and Ms Birgit Walkemeier (MPIZ) for their assistance in genomic library preparation. Many thanks to ICRISAT's Ms Seetha Kannan for providing DNA of the wild species, Dr H. Upadhyaya for selecting and providing seeds of diverse genotypes and Mr Prasanth V.P. for statistical analysis. The authors gratefully acknowledge the leading role of Prof. Dorothea Bartels (University of Bonn) in co-ordinating this work and Prof. Paul Vlek (Centre for Development Research, Bonn, Germany) for organizing funds for this project.

References

Armour, J. A., R. Neumann, S. Gobert, and A. J. Jeffreys, 1994: Isolation of human simple repeat loci by hybridization selection. Hum. Mol. Genet. **3,** 599—605.

Ashworth, V. E. T. M., M. C. Kobayashi, M. De La Cruz, and M. T. Clegg, 2004: Microsatellite markers in avocado (*Persea americana* Mill.): development of dinucleotide and trinucleotide markers. Sci. Hortic. **101,** 255—267.

Billotte, N., P. J. L. Lagoda, A.-M. Risterucci, and F.-C. Baurens, 1999: Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. Fruits **54,** 277—288.

Blair, M. W., F. Pedraza, H. F. Buendia, E. Gaitán-Solís, S. E. Beebe, P. Gepts, and J. Tohme, 2003: Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. **107,** 1362—1374.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis, 1980: Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. **32,** 314—331.

Budak, H., F. Pedraza, P. B. Cregan, P. S. Baenziger, and I. Dweikat, 2003: Development and utilization of SSRs to estimate the degree of genetic relationships in a collection of pearl millet germplasm. Crop Sci. **43,** 2284—2290.

Buhariwalla, H. K., B. Jayashree, K. Eshwar, and J. H. Crouch, 2005: Development of ESTs from chickpea roots and their use in diversity analysis of the *Cicer* genus. BMC Plant Biol. **5,** 16.

Burns, M. J., K. J. Edwards, H. J. Newbury, B. V. Ford-Lloyd, and C. D. Baggott, 2001: Development of Simple Sequence Repeat (SSR) markers for the assessment of gene flow and genetic diversity in pigeonpea (*Cajanus cajan*). Mol. Ecol. Notes **1,** 283—285.

Cardle, L., L. Ramsay, D. Milbourne, S. M. Macaulay, D. Marshall, and R. Waugh, 2000: Computational and experimental characterisation of physically clustered Simple Sequence Repeats in plants. Genetics **156,** 847—854.

Cobb, B. D., and J. M. Clarkson, 1994: A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods. Nucleic Acids Res. **22,** 3801—3805.

Don, R. H., P. T. Cox, B. J. Wainwright, K. Baker, and J. S. Mattick, 1991: 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. **19,** 4008.

Edwards, K. J., J. H. A. Barker, A. Daly, C. Jones, and A. Karp, 1996: Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques **20,** 758—759.

FAOSTAT data, 2006: Available at: http://faostat.fao.org/faostat/collections?version=ext&hasbulk=0&subset=agriculture. Accessed 6 July 2006.

Ferguson, M. E., M. D. Burow, S. R. Schulze, P. J. Bramel, A. H. Paterson, S. Kresovich, and S. Mitchell, 2004: Microsatellite identification in peanut (*A. hypogaea* L.). Theor. Appl. Genet. **108,** 1064—1070.

Greilhuber, J., and R. Obermayer, 1998: Genome size variation in *Cajanus cajan* (Fabaceae): a reconsideration. Plant Syst. Evol. **212,** 135—141.

He, G., R. Meng, M. Newman, G. Gao, R. N. Pittman, and C. S. Prakash, 2003: Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biol. **3,** 3.

Huang, X., and A. Madan, 1999: CAP3: a DNA sequence assembly program. Genome Res. **9,** 868—877.

Li, Y.-C., A. B. Korol, T. Fahima, A. Beiles, and E. Nevo, 2002: Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol. Ecol. **11,** 2453—2465.

Li, Y.-C., A. B. Korol, T. Fahima, and E. Nevo, 2004: Microsatellites within genes: structure, function, and evolution. Mol. Biol. Evol. **21,** 991—1007.

Lichtenzveig, J., C. Scheuring, J. Dodge, S. Abbo, and H.-B. Zhang, 2005: Construction of BAC and BIBAC libraries and their applications for generation of SSR markers for genome analysis of chickpea, *Cicer arietinum* L. Theor. Appl. Genet. **110,** 492—510.

Mace, E. S., H. K. Buhariwalla, and J. H. Crouch, 2003: A high throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol. Biol. Rep. **21,** 459a—459h.

Mba, R. E. C., P. Stephenson, K. Edwards, S. Melzer, J. Nkumbira, U. Gullberg, K. Apel, M. Gale, J. Tohme, and M. Fregene, 2001: Simple Sequence Repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theor. Appl. Genet. **102,** 21—31.

Moretzsohn, M. C., M. S. Hopkins, S. E. Mitchell, S. Kresovich, J. F. M. Valls, and M. E. Ferreira, 2004: Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. BMC Plant Biol. **4,** 11.

Morgante, M., and A. Olivieri, 1993: PCR-amplified microsatellites as markers in plant genetics. Plant J. **3,** 175—182.

Morgante, M., M. Hanafey, and W. Powell, 2002: Microsatellites are preferentially associated with non-repetitive DNA in plant genomes. Nat. Genet. **30,** 194—200.

Murray, M. G., and W. F. Thompson, 1980: Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. **8,** 4321—4325.

Nei, M., and W. H. Li, 1979: Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl Acad. Sci. USA **76,** 5269—5273.

Oberhagemann, P., C. Chatot-Balandras, R. Schäfer-Preg, D. Wegener, C. Palomini, F. Salamini, E. Bonnel, and C. Gebhardt, 1999: A genetic analysis of quantitative resistance to late blight in potato: towards marker-assisted selection. Mol. Breed. **5,** 399—415.

Panaud, O., X. Chen, and S. R. McCouch, 1996: Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol. Gen. Genet. **252,** 597—607.

Rallo, P., G. Dorado, and A. Martin, 2000: Development of Simple Sequence Repeats (SSRs) in olive tree (*Olea europaea* L.). Theor. Appl. Genet. **101,** 984—989.

Reddy, L. J., H. D. Upadhyaya, C. L. L. Gowda, and S. Sube, 2005: Development of a core collection in pigeonpea [*Cajanus cajan* (L.) Millspaugh] using geographic and qualitative morphological descriptors. Genet. Res. Crop Evol. **52,** 1049—1056.

Rohlf, F. J., 1994: NTSYS-PC: Numerical Taxonomy and Multivariate Analysis System. Exeter Software, New York.

Rongwen, J., M. S. Akkaya, A. A. Bhagwat, U. Lavi, and P. B. Cregan, 1994: The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. **90,** 43—48.

Rozen, S., and H. J. Skaletsky, 1998: Primer3. code. Available at: http://www-genome.wi.mit.edu/genome_software/other/primer3.html. Accessed 1 May 2005.

Sambrook, J., E. F. Fritsch, and T. Maniatis, 1989: Molecular Cloning: a Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Saxena, K. B., and R. V. Kumar, 2003: Development of a cytoplasmic nuclear male-sterility system in pigeonpea using *C. scarabaeoides* (L.) Thouars. Indian J. Genet. Plant Breed. **63,** 225—229.

Sharma, H. C., G. Pampapathy, and L. J. Reddy, 2003: Wild relatives of pigeonpea as a source of resistance to the pod fly (*Melanagromyza obtuse* Malloch) and pod wasp (*Tanaostigmodes cajaninae* La Salle). Genet. Res. Crop Evol. **50,** 817—824.

Sivaramakrishnan, S., K. Seetha, A. Nageshwar Rao, and L. Singh, 1997: RFLP analysis of cytoplasmic male-sterile lines of pigeonpea [*Cajanus cajan* (L.) Millsp.] developed by interspecific crosses. Euphytica **93,** 307—312.

Sivaramakrishnan, S., K. Seetha, and L. J. Reddy, 2002: Diversity in selected wild and cultivated species of pigeonpea using RFLP of mtDNA. Euphytica **125,** 21—28.

Song, Q. J., L. F. Marek, R. C. Shoemaker, K. G. Lark, V. C. Concibido, X. Delannay, J. E. Specht, and P. B. Cregan, 2004: A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. **109,** 122—128.

Squirrell, J., P. M. Hollingsworth, M. Woodhead, J. Russell, A. J. Lowe, M. Gibby, and W. Powell, 2003: How much effort is required to isolate nuclear microsatellites from plants? Mol. Ecol. **12,** 1339—1348.

Su, X.-Z., Y. Wu, C. D. Sifri, and T. E. Wellems, 1996: Reduced extension temperatures required for PCR amplification of extremely A + T-rich DNA. Nucleic Acids Res. **24,** 1574—1575.

Tautz, D., and P. Rentz, 1984: Simple sequences are ubiquitous repetitive components of eukaryotic genome. Nucleic Acids Res. **12,** 4127—4138.

Temnykh, S., G. DeClerck, A. Lukashova, L. Lipovich, S. Cartinhour, and S. McCouch, 2001: Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genet. Res. **11,** 1441—1452.

Thiel, T., W. Michalek, R. K. Varshney, and A. Graner, 2003: Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. **106,** 411—422.

Tóth, G., Z. Gáspári, and J. Jurka, 2000: Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. **10,** 967—981.

Van der Maesen, L. J. G., 1990: Pigeonpea: origin, history, evolution and taxonomy. In: Y. L. Nene, S. D. Hall, and V. K. Sheila (eds), Pigeonpea, 15—46. CAB International, Wallingford, UK.

Wang, Z., J. Weber, G. Zhong, and S. Tanksley, 1994: Survey of short tandem DNA repeats. Theor. Appl. Genet. **88,** 1—6.

Wang, M. L., A. G. Gillaspie, M. L. Newman, R. E. Dean, R. N. Pittman, J. B. Morris, and G. A. Pederson, 2004: Transfer of Simple Sequence Repeat (SSR) markers across the legume family for germplasm characterization and evaluation. Plant Genet. Resour. **2,** 107—119.

Wang, H., F. Li, and J. Xiang, 2005: Polymorphic EST–SSR markers and their mode of inheritance in *Fenneropenaeus chinensis*. Aquaculture **249,** 107—114.

Young, N. D., J. Mudge, and T. N. Ellis, 2003: Legume genomes: more than peas in a pod. Curr. Opin. Plant Biol. **6,** 199—204.

Yu, K., S. J. Park, V. Poysa, and P. Gepts, 2000: Integration of Simple Sequence Repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). Am. Genet. Assoc. **91,** 429—434.

Zane, L., L. Bargelloni, and T. Patarnello, 2002: Strategies for microsatellite isolation: a review. Mol. Ecol. **11,** 1—16.

Zhao, W., X. Miao, S. Jia, Y. Pan, and Y. Huang, 2005: Isolation and characterization of microsatellite loci from the mulberry, *Morus* L. Plant Sci. **168,** 519—525.

Zou, J. J., R. J. Singh, and T. Hymowitz, 2004: SSR marker and ITS cleaved amplified polymorphic sequence analysis of soybean X *Glycine tomentella* subgeneric derived lines. Theor. Appl. Genet. **109,** 769—774.