Plant Breeding 129, 142—148 (2010)

doi:10.1111/j.1439-0523.2009.01680.x

Novel SSR Markers for Polymorphism Detection in Pigeonpea (*Cajanus* spp.)

R. K. Saxena[1]*, C. Prathima[1]*, K.B. Saxena[2], D. A. Hoisington[1], N. K. Singh[3] and R. K. Varshney[1,4,5]

[1]Centre of Excellence in Genomics (CEG), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Greater Hyderabad 502324, AP, India; [2]Global Theme – Crop Improvement, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Greater Hyderabad 502324, AP, India; [3]National Research Centre on Plant Biotechnology (NRCPB), Indian Agricultural Research Institute, New Delhi 110012, India; [4]Genomics towards Gene Discovery Sub-Programme, Generation Challenge Programme (GCP), c/o CIMMYT, Int APDO Postal 6-641, 06600 Mexico DF, Mexico; [5]Corresponding author, E-mail: r.k.varshney@cgiar.org

With 1 figure and 4 tables

Received December 9, 2008/Accepted April 30, 2009
Communicated by W. Swiecicki

Abstract

With an objective to expand the repertoire of molecular markers in pigeonpea (*Cajanus cajan*), 36 microsatellite or simple sequence repeat (SSR) loci were isolated from a SSR-enriched genomic library. Primer pairs were designed for 23 SSR loci, of which 16 yielded amplicons of expected size. Thirteen SSR markers were polymorphic amongst 32 cultivated and eight wild pigeonpea genotypes representing six *Cajanus* species. These markers amplified a total of 72 alleles ranging from two to eight alleles with an average of 5.5 alleles per locus. The polymorphic information content for these markers ranged from 0.05 to 0.55 with an average of 0.32 per marker. Phenetic analysis clearly distinguished all wild species genotypes from each other and from the cultivated pigeonpea genotypes. These markers should be useful for genome mapping, trait mapping, diversity studies and assessment of gene flow between populations in pigeonpea.

Keywords: molecular markers — microsatellites — simple sequence repeats — genetic diversity — polymorphism

Pigeonpea [*Cajanus cajan* (L.) Millspaugh] (2n = 22) is a short-lived perennial shrub that is traditionally cultivated for its grains as an annual crop in Asia, Africa, Caribbean region and Latin America. Considering the vast natural genetic variability available in pigeonpea and presence of its wild relatives in the region, it has been postulated that India is the primary center of origin of pigeonpea (van der Maesen 1980). It is hardy, widely adapted and drought tolerant crop with a large temporal variation (90–300 days) for maturity. These traits allow its cultivation in a vast range of agro climatic zones and cropping systems. Globally, pigeonpea is cultivated on a 4.68 m/ha of land with an annual production of 3.65-metric tonnes and productivity of 780 kg/ha (http://faostat.fao.org/site/381/default.aspx). Being a legume, pigeonpea crop fix 40 kg/ha atmospheric nitrogen (Kumar Rao et al. 1983) and add valuable organic matter to the soil through the fallen leaves. Its roots help in releasing soil-bound phosphorus to make it available for plant growth (Ae et al. 1990). With such advantages at low input, pigeonpea has become an ideal crop for sustainable cropping system in rainfed areas. However, because of exposure of the crop to a number of biotic (e.g. *Fusarium* wilt and sterility mosaic disease) and abiotic (e.g. salinity and water logging) stresses, the realized yield of the crop is very low in the marginal environment of Asia and Africa.

Use of molecular markers for improving the efficiency of plant breeding programmes has been successful especially for biotic/abiotic stress tolerance in several crops (Varshney et al. 2006). Because at the molecular level recognizing different alleles for a particular gene is independent of plant age or plant part. Also, in contrast to morphological traits, molecular markers are not influenced by various pleiotropic and epistatic interactions. Among different types of molecular markers, microsatellite or simple sequence repeat (SSR) markers have been proven the markers of choice for application in molecular breeding (Gupta and Varshney 2000). In case of pigeonpea, however, only 30 microsatellite markers are available so far (Burns et al. 2001, Odeny et al. 2007). Because of lower level of polymorphism in cultivated pigeonpea germplasm (Yang et al. 2006), there is an urgent need to develop the SSR markers in larger number so that germplasm characterization, molecular mapping and gene flow studies can be undertaken in pigeonpea.

In view of above, the present study was undertaken that deals with: (i) isolation and characterization of novel SSR from a SSR enriched genomic DNA library, (ii) validation and polymorphism assessment of new set of SSR markers and (iii) use of new SSR markers for understanding genetic relationships in selected genotypes of cultivated and wild species of pigeonpea.

Materials and Methods

Plant material and DNA extraction: In order to construct the genomic library, Asha (ICPL 87119) a released variety of pigeonpea was utilized. To measure level of microsatellite polymorphism, a set of 40 pigeonpea genotypes was used, consisting of cultivated varieties, inbred lines and wild pigeonpea species (Table 1). Total genomic DNA was isolated and purified following the protocol as mentioned in Cuc et al. (2008).

Construction of SSR enriched library: To isolate new SSR loci, a SSR enriched library was constructed from Asha genotype (ICPL 87119) using a bead capture enrichment protocol (Glenn and Schable 2005). Briefly, genomic DNA was digested with *Rsa*I and *Xmn*I (New England Biolabs, Inc) and fragments were ligated to double stranded SuperSNX24 (5′GTTTAAGGCCTAGCTAGCAGAATC) and super SNX24 + 4p (5′pGATTCTGCTAGCTAGGCCTTAAACAAAA). Fragments were then hybridized at their respective temperature to five

*These authors contributed equally to this work.

biotinylated oligonucleotides [$(CT)_{10}$ – 42°C, $(TCG)_{10}$ – 60°C, $(AAG)_8$ – 40°C and mixture of $(TG)_{12}$ and $(AG)_{14}$ – 45°C] using magnetic streptavidin beads (Dynabeads; DYNAL, Invitrogen, Carlsbad, CA, USA). Recovered enriched DNA fragments were amplified through polymerase chain reaction (PCR) using SuperSNX24 forward primer (30 cycles with 60°C annealing temperature). Subsequently, the enriched products were ligated into pCR2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA) and used to transform into competent *Escherichia coli* cells (TOP 10; Invitrogen, Carlsbad, USA), plated on LB agar containing ampicillin. To allow for blue-white selection, the plates were spread with 5- bromo-4-chloro-3-indolyl-β-D-galactopyranoside plus isopropyl β-D-thiogalactopyranoside (IPTG). White colonies were picked and plated in a grid on LB agar containing ampicillin, prior to making colony lifts with Nylon Membranes, positively charged. Hybridization was carried out at 55°C overnight using radioactive-labeled ($\gamma^{32}P$) probes containing the SSR motifs used for enrichment. Clones giving the strong signals in hybridization were selected as SSR positive clones.

Sequencing of SSR-positive clones: The SSR positive clones were grown overnight in 5 ml LB broth with 100 μg/ml ampicillin. Plasmid DNA was extracted using alkaline lysis method, the plasmid DNA was sequenced using M13 Forward 19-mer Sequencing Primer following the dideoxynucleotide chain termination method on ABI 3700 sequencer. Raw sequence data were trimmed using a sliding window of 50bp with a minimum average Phred score of 20 and filtered for a minimum length of 100bp.

SSR identification and primer designing: The sequencing data were analysed using CLUSTAL W software in order to determine redundancy in the library. Non-redundant sequences were analysed with *MI*cro*SA*tellite (MISA) perl script [http://pgrc.ipk-gatersleben.de/misa/, (Varshney et al. 2002)] to identify SSRs. Primer pairs flanking the SSR motifs were designed using program PRIMER3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi).

Amplification and visualization of SSR loci: Polymerase chain reaction was performed in a 5 μl reaction volume containing 1.0 μl (5 ng) of template DNA, 0.5 μl of 10 × PCR buffer, 0.3 μl of 25 mM $MgCl_2$, 0.5 μl of 2mM dNTPs, 0.15 μl of 10 pM primer, anchored with M13-tail and 0.3 U of *Taq* polymerase using GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA). A touch down PCR programme was used to amplify the DNA fragments with the initial denaturation for 15 min at 95°C, followed by 10 cycles of denaturation for 20 s at 94°C, annealing for 20 s at 55°C (reduction of 1°C per cycle) and extension for 30 s at 72°C. Subsequently, 35 cycles of denaturation at 94°C for 20 s, annealing for 20 s at 48°C and extension for 30 s at 72°C were used and followed by 20 min final extension at 72°C. Amplification products were separated on capillary electrophoresis using ABI3130 Genetic Analyzer. Fragment analysis was performed by using GENESCAN and GENOTYPER software version 3.7 (Applied Biosystems, Foster City, CA, USA).

Data analysis: Allelic data were recorded for each marker and subjected to ALLELOBIN software (http://www.icrisat.org/gt-bt/download_allelobin.htm) to get allele calls based on repeat motif. Allele number, observed heterozygosity, expected heterozygosity or gene diversity and polymorphic information content (PIC) was calculated by using POWERMARKER V3.25 (Liu and Muse 2005). Dissimilarity matrix was used to construct a dendrogram by using Neighbor-joining method with DARWIN V5.0.128 software (Perrier et al. 2003). Hardy–Weinberg equilibrium (HWE) test and linkage disequilibrium (LD) analysis were conducted using ARLEQUIN version 3.1 (Schneider et al. 2000).

Results

Isolation and characterization of microsatellites

The SSR enriched library was constructed from the genotype 'Asha' (ICPL 87119) following the method of Glenn and Schable (2005). This library was enriched for (CT), (TG), (AG), (AAG) and (TCG) SSR repeat motifs. In the first instance, a total of 3072 clones were picked from 32 96-well plates. After screening these clones for blue/white selection using X-gal/IPTG and hybridization with SSR probes, a total of 82 clones were considered as SSR positive clones. Sequencing of these clones indicated the insert size in the range of 100–800 bp. Sequence data of these clones have been submitted to GenBank (accession nos FI131355–FI131436).

To characterize the newly isolated sequences, commonly called Genome Survey Sequences (GSSs) of pigeonpea (FI131355–FI131436), these were analysed for sequence similarity with the available transcript (EST) sequence data of seven legume species (Table 2). As a result, almost all GSSs of pigeonpea got the hit with at least one legume species, however significant hit were found with 70 (85.3%) sequences. Sequence similarity of GSSs varied from 8.5% (cowpea or *Vigna unguiculata*) to 64.6% (groundnut or *Arachis hypogaea*). It is interesting to note that four (4.8%) GSSs showed significant similarity with all the legume species surveyed except Lotus (*Lotus japonicus*), 12 (14.6%) GSSs did not show a significant similarity with sequence data for any legume species analysed. Furthermore, to understand a putative function of newly isolated GSSs, BLAST X was performed for all newly isolated GSSs with SWISSPROT and a significant hit was considered at $<$1E-05 value. As a result, 53 (64.63%) GSSs showed a putative function, whereas 16 (19.51%) GSSs showed non-significant hit and 13 (15.85%) GSSs did not show any match with the protein sequences.

In order to identify the SSRs, sequence data for 82 GSSs were screened with *MI*cro*SA*tellite (MISA) search tool for the presence of SSR. A total of 23 GSSs representing 23 clones showed 36 SSRs. While 11 SSRs were present in compound formation and 10 clones contained more than one SSR. In terms of different classes of SSRs, di-nucleotide SSRs constituted the majority of the proportion (29, 80.5%) followed by mono-nucleotide SSRs (6, 16.7%). Other than mono- and di-nucleotide SSRs, only one tetra-nucleotide (2.8%) was present in the SSRs identified. The maximum repeat unit number for mono-nucleotide (A/T) and di-nucleotide SSRs were 21 and 20, respectively. The overall repeat unit numbers ranged from 5 to 21. Number of SSRs identified is low and therefore these results can not be used to speculate the frequency and abundance of SSRs in pigeonpea genome.

Marker development and validation

To develop the SSR markers, all the 23 GSSs that showed presence of SSRs, were used for primer designing. As a result, the primer pairs were developed for 23 (63%) SSRs, which have come from 22 clones.

Amplification conditions for all 23 primer pairs were optimized initially on two pigeonpea genotypes *viz*., ICPL 87119 and ICPA 2043. However only, 16 (69.5%) primer pairs (Table 3) showed amplification in both the genotypes. Subsequently, only these 16 primer pairs/markers were used for assessing the polymorphism in pigeonpea germplasm.

Polymorphism assessment in pigeonpea germplasm

A panel of 40 genotypes representing seven *Cajanus* species (Table 1) i.e. *Cajanus cajan, C. acutifolius*, *C. albicans*, *C. cajanifolius*, *C. platycarpus*, *C. scarabaeoides* and *C. sericeus*

Table 1: Details on germplasm used for diversity analysis

Pigeonpea genotypes	Species	Pedigree/origin	Maturity group	Habit	DM	DF	Seed colour
ICP 2376	*Cajanus cajan*	P 3888 Andhra Pradesh (India)	MD	NDT	176	118	C
ICP 7035	*C. cajan*	DSL R-55 Madhya Pradesh (India)	MD	NDT	192	130	P
ICP 8863	*C. cajan*	ICWR-6-7626-W1X-W16XB	MD	NDT	176	114	B
ICPB 2042	*C. cajan*	B-line	MD	NDT	165	90	W
ICPB 2043	*C. cajan*	B-line	MD	NDT	175	115	B
ICPB 2049	*C. cajan*	B-line	MD	NDT	160	118	B
ICPB 2051	*C. cajan*	B-line	MD	NDT	175	117	B
ICPL 20096	*C. cajan*	ICPL 87119 × ICP12746	MD	NDT	185	127	B
ICPL 20097	*C. cajan*	ICPL 87119 × ICP12746	MD	NDT	187	131	B
ICPL 20098	*C. cajan*	ICPL 87119 × ICP12746	MD	NDT	184	128	C
ICPL 20099	*C. cajan*	ICP 9150 × ICPL 87119	MD	NDT	184	127	B
ICPL 20108	*C. cajan*	ms 3783 × ICPL 87119	MD	NDT	181	125	C
ICPL 20110	*C. cajan*	ms 3783 × ICPL 87119	MD	NDT	186	130	C
ICPL 20112	*C. cajan*	ms 3783 × ICPL 87119	MD	NDT	182	127	C
ICPL 20113	*C. cajan*	ms 3783 × ICPL 87119	MD	NDT	185	129	LB
ICPL 20125	*C. cajan*	ms 3783 × GAUT 85–19	MD	NDT	190	137	B
ICPL 20127	*C. cajan*	ms 3783 × GAUT 85–19	MD	NDT	183	128	B
ICPL 20129	*C. cajan*	ms 3783 × GAUT 85–19	MD	NDT	185	131	B
ICPL 20135	*C. cajan*	ms 3783 × GAUT 85–19	MD	NDT	179	126	C
ICPL 332	*C. cajan*	ICP 1903-E1 (SPS)	MD	NDT	178	118	B
ICPL 84023	*C. cajan*	ICPX 78354	SD	DT	134	68	B
ICPL 85063	*C. cajan*	BDN1X 73054-55-5	MD	NDT	170	111	B
ICPL 86012	*C. cajan*	90 C–H6 (Florida)	SD	DT	130	65	C
ICPL 87051	*C. cajan*	ICP 7979 × C11	MD	NDT	179	131	W
ICPL 87091	*C. cajan*	(ICP 8504 × ICP 7220 × ((ICP 3783 × ICP 28)	SD	DT	121	74	C
ICPL 87119	*C. cajan*	C 11 × ICP-1-6 W 3X–W 1X	MD	NDT	180	122	B
ICPL 88034	*C. cajan*	ICPL 81 × ICPL 151	SD	NDT	137	88	B
ICPL 88039	*C. cajan*	((ICP 6 × ICP 6973)-46-BI-1-HINDT4-B-B*-B*)	SD	NDT	108	55	B
ICPL 96053	*C. cajan*	ICPL 87051 × ICPL 8357	MD	NDT	191	138	C
ICPL 96058	*C. cajan*	ICPL 88047 × ICPL 8357	MD	NDT	190	139	B
ICPL 99050	*C. cajan*	ICPL88039 × ICP8564	MD	NDT	175	123	B
ICPL 99052	*C. cajan*	ICPL88039 × ICP8564	MD	NDT	162	104	B
Wild relatives							
ICP 15602	*C. acutifolius*	Australia					
ICP 15614	*C. albicans*	Karnataka (India)					
ICP 15629	*C. cajanifolius*	Madhya Pradesh (India)					
ICP 15661	*C. platycarpus*	Himachal Pradesh (India)					
ICP 15695	*C. scarabaeoides*	Sri Lanka					
ICP 15706	*C. scarabaeoides*	Bihar (India)					
ICP 15758	*C. scarabaeoides*	Indonesia					
ICP 15761	*C. sericeus*	Maharashtra (India)					

MD, medium duration; SD, short duration; NDT, non determinate; DT, determinate; DM, days to maturation; DF, days to flowering; C, cream; P, purple; B, brown; LB, light brown; W, white.

Table 2: Analysis of pigeonpea Genome Survey Sequences (GSS) with selected legumes

	EST data surveyed			
Legume species	Number	Size (Mbp)	No. GSS showing homology	No. GSS showing significant homology (< 1E-05)
Soybean (*Glycine max*)	880 561	327.1	81	48
Cowpea (*Vigna unguiculata*)	183 757	108.6	81	7
Common bean (*Phaseolus vulgaris*)	83 448	49.2	82	45
Lotus (*Lotus japonicus*)	158 135	68.1	72	9
Chickpea (*Cicer arietinum*)	7097	3.3	82	44
Medicago (*Medicago truncatula*)	249 625	138.2	82	49
Groundnut (*Arachis hypogaea*)	41 489	22.5	82	53

was used for assessing the polymorphism of newly developed 16 SSR markers (Table 3). While 13 markers yielded a total of 72 alleles with an average of 5.5 alleles per marker in the germplasm analysed, three markers did not show any polymorphism (Table 4). Allele numbers detected by polymorphic SSR markers varied from 2 (ICPM1C11) to 8 (ICPM1G11 and ICPM2B08). The PIC value for these markers ranged from 0.05 (ICPM1C11) to 0.55 (ICPM1E04), with an average of 0.32 per marker. While major allele frequency at the polymorphic SSR loci ranged from 0.50 (ICPM1E04) to 0.97 (ICPM1C11), the expected heterozygosity (He) or gene diversity varied from 0.05 (ICPM1C11) to 0.62 (ICPM1E04) with an average of 0.34 per marker (Table 4). Furthermore, analysis of allelic data for 13 polymorphic markers for HWE test did not show any marker with significant deviation from HWE ($P < 0.05$). While analysing LD for all possible marker pairs, 14 marker pairs were found in significant LD (Table 4). Different genetic parameters/tests such as He, LD, HWE are important indicators of origin, evolution and distribution of diversity in the available gene pool. The heterozygosity

Table 3: Details on 16 novel pigeonpea simple sequence repeat (SSR) markers

SSR name	GenBank accession ID	Repeat motif	Forward primer sequence (5′-3′)	Reverse primer sequence (5′-3′)	*T*m (°c)	Putative function
ICPM1A08	F1131362	$(CA)_6$	AAGGCAAGATACTTGGTTCTGG	TCTCTCCCTGAGGTTTCCATT	58	ATP-dependent RNA helicase MRH4
ICPM1B04_a	F1131367	$(GA)_{13}$	GTGGGATACCATGTTCCAGG	CCGAATACATGCTAAAGGGG	57	Catalase
ICPM1B04_b	F1131367	$(T)_{10}$	CCCCTTTAGCATGTATTCGG	TTTAAACGAATTCGCCCTTG	57	Catalase
ICPM1C02	F1131371	$(GT)_6n(TG)_5$	GCATAGAGTTAGGAACATTCATTGC	AGCGTTCAACCCAAACAAAA	60	Catalase
ICPM1C11	F1131372	$(TA)_5$	TCCTCATGTAGCCTATTGGGTT	CATGTGAATATTCCTTCGATGC	58	–
ICPM1D01	F1131374	$(TA)_5$	AATTTTTAGCACAATGGCCG	AATTATCACAAATGGCACGC	53	–
ICPM1D10	F1131378	$(AT)_5(GT)_6$	GGATTAACCAATTGTGAGTGAACC	TGCACTTTATAAGCATTTACCAACA	59	Catalase
ICPM1E04	F1131382	$(A)_{10}$	TTTTTATGGAATATTTATGAGTTGGC	AAGAGTTTCCCAACCCTGCT	55	–
ICPM1E10	F1131384	$(CA)_7$	AGAACAACAAGGGAGCGAGA	CCATGACATCATTGCATGATAAA	57	–
ICPM1F11	F1131389	$(GT)_7$	ATCCCCACCCTTGTGTCATA	TCTTTCCATTTACACCCCGT	57	Transmembrane protein 9B precursor
ICPM1G01	F1131390	$(CA)_8$	ATCACCAACATCCCCATGAT	TCACCAACGATGAATTGTGAA	55	Catalase
ICPM1G04	F1131393	$(T)_{21}$	GCTCCAATTTTTCATTTCGG	ATCAAACAATGCACCCATGA	53	Protein in NOF-FB transposable element
ICPM1G11	F1131398	$(CA)_9$	GCAAGTGTTCCCTACGTTGC	CTCCAACGGCCATAGTAGGA	59	–
ICPM1H01	F1131399	$(A)_{10}$	CCCGAACTGCATTCAAAAAT	GCGTAGGTGGAAGAAGATCG	53	–
ICPMCT20	F1131410	$(GA)_{14}(AAGA)_5$	GAGGCTGAGGGGTGAAAAAT	CCTCTGGATTCCCTCTTTCC	57	–
ICPM2B08	F1131422	$(TG)_5n(TG)_5$	AGTTTGAAATTGCTTTTGGCT	GAATTGGGAGAGACCGCATA	52	–

*T*m, melting temperature; –, no hit was found.

measures for the new SSR markers indicated moderate heterozygosity in the tested germplasm. Indeed reflective of the genetic composition and mating behavior of the tested germplasm. The results thus suggest the suitability of the new markers for reliably ascertaining genetic diversity in the *Cajanus* gene pool.

As expected, higher polymorphism was observed in wild species gene pool as compared with cultivated species gene pool. For instance, out of 13 polymorphic markers, only nine markers were polymorphic in genotypes of cultivated species while all 13 were polymorphic in other *Cajanus* species. Similarly the average allele number and PIC values were higher in wild species genotypes (PIC value – 0.64, allele number – 5) as compared with cultivated species genotypes (PIC value – 0.15, allele number – 2.08).

Genetic relationships among cultivated and wild species

Allelic data obtained for all 13 polymorphic markers on 40 genotypes were used to prepare the genetic similarity matrix based on DICE similarity coefficient that was used to prepare the phenogram using DARWIN program. The phenogram classified the germplasm into four main clusters (Fig. 1). The cluster A contained 14 genotypes; the cluster B contained 13 genotypes while other two clusters namely C and D contained 9 and 4 genotypes, respectively. Under each of the main clusters, genotypes were grouped further into subclusters. For instance, the cluster A contained four subclusters (AI, AII, AIII and AIV), the cluster B contained three subclusters (BI, BII and BIII) and the cluster C (CI and CII) and D (DI and DII) contained two subclusters each.

It is interesting to note the grouping of the wild species genotypes and cultivated species genotypes in different clusters. For instance, all the wild species genotypes were grouped into main cluster A, while majority of genotypes belonging to the cultivated germplasm were grouped under the main clusters B, C and D. Furthermore, seven of the eight wild species genotypes were grouped into subcluster AI while the solitary genotype ICP 15629 (*C. cajanifolius*) was grouped with cultivated species genotypes under subcluster AIV indicating a closer relationship of *C. cajanifolius* species (at least the accession ICP 15629) with the cultivated species *C. cajan* (van der Maesen 1990). Data obtained by 13 polymorphic markers was utilized to understand the pedigree relationships among the 32 cultivated pigeonpea lines. Three lines (ICPL 20096, ICPL 20097 and ICPL 20098) developed by a single plant selection from the same parents (ICPL 87119 × ICP 12746) were grouped into the major cluster C. The same results were obtained for two more lines (ICPL 99050 and ICPL 99052) which were developed from cross ICPL 88039 × ICP 8564 and grouped into the major cluster A. In some other cases, where lines shared the same pedigree were not found grouped together in the phenogram. Hence no clear relationships were identified between phenetic analysis and pedigree data of the cultivated lines. Being an often cross pollinated crop, there can be a variation from plant to plant at the genome level in pigeonpea. As most of the cultivated lines used for the study were of a medium duration (MD) maturity (27), and only five lines were of a short duration (SD) maturity, four SD lines were found to be grouped into the major cluster B. However, a single SD line ICPL 87091 grouped with three MD lines was found in the major cluster D.

Table 4: Polymorphism features of 13 polymorphic SSR markers in 40 *Cajanus* genotypes

		No. alleles			Observed heterozygosity			Expected heterozygosity			Polymorphic information content (PIC)		
Marker ID*	Allele size (bp)	Cultivated	Wild	Across	Cultivated	Wild	Across	Cultivated	Wild	Across	Cultivated	Wild	Across
ICPM1A08	290–294	1	3	3	0	0.29	0.05	0	0.56	0.19	0	0.46	0.17
ICPM1B04_b	108–178	4	2	4	0	0	0	0.25	0.28	0.26	0.24	0.24	0.25
ICPM1C02	272–294	1	7	7	0	0.29	0.05	0	0.84	0.3	0	0.82	0.29
ICPM1C11	279–285	1	2	2	0	0	0	0	0.24	0.05	0	0.21	0.05
ICPM1D10	294–330	2	3	4	0	0	0	0.12	0.56	0.29	0.11	0.5	0.28
ICPM1E04	207–209	3	3	3	0	0	0	0.61	0.65	0.62	0.54	0.58	0.55
ICPM1E10	131–191	2	6	6	0	0	0	0.28	0.81	0.47	0.24	0.79	0.44
ICPM1F11	243–265	1	6	6	0	0	0	0	0.81	0.28	0	0.79	0.27
ICPM1G01	107–297	3	5	7	0	0.13	0.03	0.12	0.76	0.34	0.12	0.72	0.33
ICPM1G04	133–158	3	6	7	0	0	0	0.47	0.83	0.58	0.39	0.81	0.52
ICPM1G11	229–247	2	8	8	0	0.25	0.05	0.06	0.83	0.35	0.06	0.81	0.34
ICPMCT20	118–158	2	6	7	0	0	0	0.06	0.82	0.29	0.06	0.79	0.29
ICPM2B08	144–224	2	8	8	0.031	0.25	0.08	0.19	0.85	0.41	0.18	0.83	0.39
Mean		2.08	5	5.5	0.002	0.09	0.02	0.17	0.68	0.34	0.15	0.64	0.32

*Marker pairs showing significant LD: ICPM1D10/ICPM1B04_b, ICPM1B04_b/ICPM1G04, ICPM1E04/ICPM1D10, ICPM1G01/ICPM1D10, ICPM1E04/ICPM1E10, ICPM1E04/ICPM1G01, ICPM1E04/ICPM1G04, ICPM1E10/ICPM1G01, ICPM1E10/ICPM1G11, ICPM1E10/ICPMCT20, ICPM1G01/ICPM1G11, ICPM1G01/ICPMCT20, ICPM1G04/ICPM2B08 and ICPM1G11/ICPMCT20.

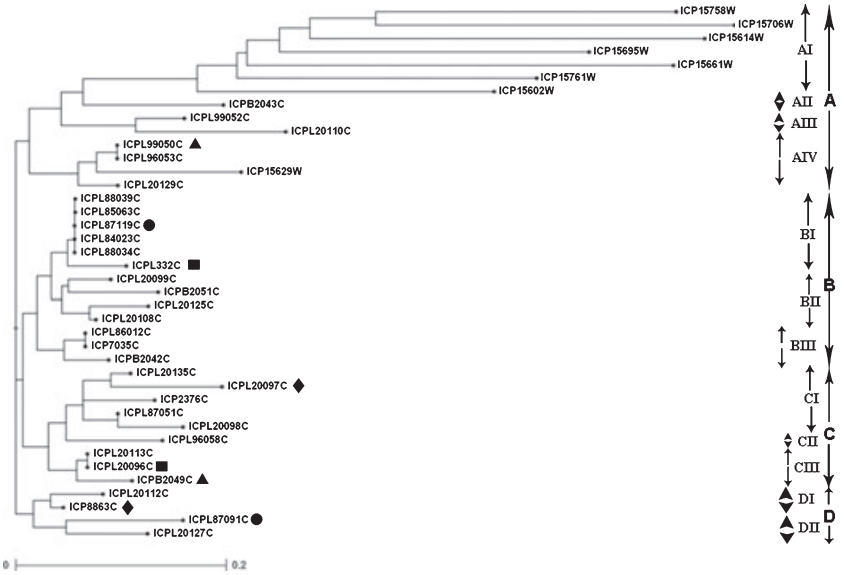

Fig. 1: Cluster analysis of 40 pigeonpea accessions based on genotyping data for 13 polymorphic simple sequence repeat (SSR) markers ▲, ●, ■, ◆: represent parental genotypes of four different mapping populations developed at ICRISAT

Discussion

A prerequisite for marker assisted breeding is a robust set of informative markers for the species of interest. Pigeonpea is an important pulse crop of tribe Phaseoleae with very limited genomic resources. As microsatellite or SSR markers are highly polymorphic, reproducible, co-dominant in nature and distributed throughout the genome, they have become the ideal marker system for genetic analysis and breeding applications (Gupta and Varshney 2000). In case of pigeonpea, however, only 30 SSR markers have been reported in literature (Burns et al. 2001, Odeny et al. 2007). The present study reports a new set of novel SSRs to the existing repertoire of molecular markers in pigeonpea.

For isolation of SSRs, construction and screening of partial genomic libraries and sequencing of SSR positive clones has been considered effective method (Rafalski et al. 1996). Enrichment of genomic DNA libraries for SSRs enhances the SSR isolation efficiency (Kijas et al. 1994, Edwards et al. 1996). Therefore SSRs-enriched genomic libraries were constructed in the present study. The SSR isolation efficiency in the present study is comparable to earlier SSR isolation studies in pigeonpea (Burns et al. 2001, Odeny et al. 2007). However as compared with SSR isolation studies in other species e.g., groundnut (Moretzsohn et al. 2004), common bean (Blair et al. 2003) etc., low frequency of SSRs was reported in the present study. By combining this study together with Burns et al. (2001) and Odeny et al. (2007), it seems that either frequency of SSRs in pigeonpea genome is lower than other legume genomes or the SSR isolation protocols deployed so far in pigeonpea are less effective.

In terms of examining the sequence similarity of newly isolated pigeonpea GSSs more than 50% GSSs showed homology with five legumes (soybean, common bean, chickpea, Medicago and groundnut) except cowpea and Lotus. Although the data surveyed for these seven legumes was not comparable, low sequence homology of pigeonpea GSSs with

cowpea and Lotus is a bit surprising. It indicates that homologous sequence for pigeonpea GSSs isolated in present study probably have not been sampled in transcript sequencing of cowpea and Lotus.

Majority of SSRs isolated in the present study comprised of di-nucleotide repeats (80.5%) followed by mono-nucleotide (16.6%) and tetra-nucleotide (2.7%) repeats. This can be attributed to (a) enrichment of library for three di-nucleotide and two tri-nucleotide repeats and (b) smaller sample size.

As number of SSRs isolated was low, mono-nucleotide SSRs were also considered for primer designing. The percentage of primer designed, in relation to the number of clones sequenced 28.04% is higher than in some studies like Cuc et al. (2008) (23.6%) while lower than in some other reports as Moretzsohn et al. (2004) (41.4%) and Wang et al. (2004) (43.7%). This may be due to the size range of insert, the restriction enzyme used for genomic DNA library construction and the method utilized for SSR enrichment (Gupta and Varshney 2000). The functionality of the primer pairs (69.5%) is comparatively higher than in the previous studies in pigeonpea [40.8% in Burns et al. (2001) and 47.6% in Odeny et al. (2007)].

As a result of screening of newly developed and validated 16 markers on 40 genotypes representing different *Cajanus* species, 81.3% markers showed polymorphism. The polymorphism rate observed in the present study is higher than in the study of Burns et al. (2001) (50%) as the present study employed screening of wild species genotypes while only cultivated genotypes were used by Burns et al. (2001). On the other hand, the observed polymorphism in the present study is lower than Odeny et al. (2007) (95%) and it can be attributed to the use of higher number of genotypes representing more wild species (nine) and tertiary gene pool as compared with the present study. Similar was the case for PIC value and average number of alleles detected by the marker. The average PIC value and number of alleles were 0.60 and 4.8 for the markers developed by Odeny et al. (2007) while 0.32 and 3.4, respectively, for the markers developed in the present study. As Burns et al. (2001) did not use any wild species; they could score only 3.1 alleles per marker.

The average PIC value of di-nucleotide repeats was found higher than that of other oligonucleotide repeats in the present study. Higher level of polymorphism for di-nucleotide repeats as compared with tri-nucleotide repeats has been reported in past as well (He et al. 2003, Ashworth et al. 2004). In general, di-nucleotide repeats have been reported to reside outside coding regions of genes (Temnykh et al. 2001) and are characterized with more repeat numbers making them the best source of highly polymorphic microsatellite markers. Furthermore, longer repeats have been shown to display higher degrees of polymorphism in several earlier studies (Li et al. 2002). In general, mono-nucleotide repeats are not considered for marker developed because of difficulty in resolving the polymorphism observed because of mutation in mono-nucleotide repeat motif. However, as we wanted to develop as many SSR markers as we could do, we considered mono-nucleotide SSRs as well as for marker developed. It is interesting to note that three (75%) of four SSR markers developed for mono-nucleotide repeats showed polymorphism. In fact, two mono-nucleotide repeat-based SSR markers (ICPM1E04 and ICPM1G04) showed higher PIC values as compared with other type of SSR markers. These observations indicate not to drop mono-nucleotide SSRs for converting them into markers especially in those species where only a limited number of markers are available.

As expected, less genetic variation was observed within cultivated species. Average number of alleles as well as the PIC value in cultivated genotypes is lower than half, than that observed in wild species. Although all the six wild species genotypes included in the present study represent only secondary gene pool, higher allele numbers and PIC values could have been observed if some genotypes representing the tertiary gene pool were included in the study. Indeed, similar kind of reports on observations of less genetic variation within cultivated gene pool as compared with inter-specific gene pool have been reported in past using SSR (Odeny et al. 2007) and diversity array technology (DArT) markers (Yang et al. 2006). These results further support the narrow genetic diversity available in cultivated gene pool. Therefore it is recommended to tap the diversity from wild species into cultivated gene pool for crop improvement programmes.

Genetic distance analysis based on SSR-allelic data showed differentiation between genotypes *C. cajan* and its wild relatives. The phenogram showed a distribution of genotypes into four major groups. Indeed such grouping of genotypes was observed in earlier diversity studies based on SSR (Odeny et al. 2007) and DArT (Yang et al. 2006) markers. Grouping of an accession (ICP 15629) representing *C. cajanifolius* together with genotypes of *C. cajan* indicates a close relationship of *C. cajan* with *C. cajanifolius*. The close relationships of *C. cajanifolius* with cultivated pigeonpea (*C. cajan*) has been shown earlier as well. For instance, similar observations were made based on mtDNA restriction fragment length polymorphism analysis (Sivaramakrishnan et al. 2002). Morphologically also, *C. cajanifolius* resembles *C. cajan* in all traits except the presence of a prominent strophiole. The species *C. cajanifolius* has been considered as the nearest wild relative of cultivated pigeonpea (*C. cajan*) (van der Maesen 1990).

In addition, to phylogenetic analysis and understanding the intra- and inter-specific relationships, the information obtained from the phenogram and phylogenetic analysis can be used for developing genetically diverse mapping populations. As at present not a single genetic map is available for pigeonpea because of a very low level of genetic variation between the parental genotypes of the mapping populations developed in the past and a limited number of molecular markers, efforts are underway at ICRISAT to develop genetically diverse mapping populations in addition to developing novel set of molecular markers. For instance, based on agronomic trait data and marker polymorphism data obtained using other set of 30 SSR markers (Burns et al. 2001, Odeny et al. 2007), the diverse parental genotypes were selected to develop the mapping populations segregating for *Fusarium* wilt and sterility mosaic diseases (ICPB 2049 × ICPL 99050; ICP 8863 × ICPL 20097; ICPL 332 × ICPL 20096 and ICPL 87091 × ICPL 87119; Fig. 1) (Saxena et al. 2009). While checking the grouping of these parental genotypes in the phenogram of the present study, these were found in different clusters/groups. It is interesting to note that these newly developed SSRs have capability to distinguish among these parents. For instance ICPB 2049 and ICPL 99050, the parental genotypes of a mapping population segregating for *Fusarium* wilt resistance were grouped in to CIII and AIV group, respectively. These analyses reconfirm the diverse nature of the parental genotypes of the mapping populations being developed at present. In addition to develop the mapping populations, selection and utilization of diverse genotypes is very important for breeding programmes to

enhance the diversity of breeding populations for selection gains in the future.

In summary, the present study adds a new set of 16 SSR markers to the repertoire of pigeonpea markers which are moderately polymorphic. The continued development of SSR markers in pigeonpea will facilitate construction of genetic map, trait mapping, diversity studies for selecting diverse lines for developing mapping populations and use in breeding programme, rapid assessment of gene flow between populations, monitoring the rates of genetic erosion and development of risk assessment strategies that are required before any transgenic pigeonpea can be grown in the field.

Acknowledgements

The study was funded by Pigeonpea Genomics Initiative of Indian Council of Agricultural Research (ICAR) under the framework of Indo-US Agricultural Knowledge Initiative (AKI).

References

Ae, N., J. Arihara, K. Okada, T. Yoshihara, and C. Johansen, 1990: Phosphorus uptake by pigeonpea and its role in cropping systems of the Indian sub-continent. Science **248,** 477—480.

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.

Blair, M. W., F. Pedraza, H. F. Buendia, E. Gaitan-Solıs, 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.

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.

Cuc, L. M., E. S. Mace, J. H. Crouch, V. D. Quang, T. D. Long, and R. K. Varshney, 2008: Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). BMC Plant Biol. **8,** 55.

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.

Glenn, T. C., and N. A. Schable, 2005: Isolating microsatellite DNA loci. Meth. Enzymol. **395,** 202—222.

Gupta, P. K., and R. K. Varshney, 2000: The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica **113,** 163—185.

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.

Kijas, J. M. H., J. C. S. Flower, C. A. Garbett, and M. R. Thomas, 1994: Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. BioTechniques **16,** 656—662.

Kumar Rao, J. V. D. K., P. J. Dart, and P. V. S. S. Sastry, 1983: Residual effect of pigeonpea (*Cajanus cajan* (L.) Millsp.) on yield and nitrogen response of maize. Exp. Agric. **19,** 131—141.

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.

Liu, K., and S. V. Muse, 2005: PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics **21,** 2128—2129.

van der Maesen, L. J. G., 1980: India is the native home of the pigeonpea. In: J. C. Arends, G. Boelema, C. T. de Groot, and A. J. M. Leeuwenberg (eds), Libergratulatorius in honerem HCD deWit, 257—262. Wageningen Agricultural University, Miscellaneous paper, The Netherlands.

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.

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.

Odeny, D. A., B. Jayashree, M. Ferguson, D. Hoisington, J. Crouch, and C. Gebhardt, 2007: Development characterization and utilization of microsatellite markers in pigeonpea [*Cajanus cajan* (L.) Millsp.]. Plant Breed. **126,** 130—137.

Perrier, X., A. Flori, and F. Bonnot, 2003: Data analysis methods. In: P. Hamon, M. Seguin, X. Perrier, and J. C. Glaszmann (eds), Genetic Diversity of Cultivated Tropical Plants, 43—76. Enfield, Science Publishers, Montpellier.

Rafalski, J. A., J. M. Vogel, M. Morgante, W. Powell, C. Andre, and S. V. Tingey, 1996: Generating and using DNA markers in plants. In: B. Birren, and E. Lai (eds), Non-Mammalian Genomic Analysis: A Practical Guide, 75—135. Academic, San Diego.

Saxena, R. K., K. B. Saxena, R. V. Kumar, D. A. Hoisington, and R. K. Varshney, 2009: SSR-based diversity in elite pigeonpea genotypes for developing mapping populations to map resistance to *Fusarium* wilt and sterility mosaic disease. Plant Breeding, in press.

Schneider, S., D. Roessli, and L. Excoffier, 2000: ARLEQUIN: A Software for Population Genetics Data Analysis. Ver 2.001. Genetics and Biometry Laboratory. Department of Anthropology, University of Geneva, Geneva.

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.

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.

Varshney, R. K., T. Thiel, N. Stein, P. Langridge, and A. Graner, 2002: In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol. Biol. Lett. **7,** 537—546.

Varshney, R. K., D. A. Hoisington, and A. K. Tyagi, 2006: Advances in cereal genomics and applications in crop breeding. Trends Biotechnol. **24,** 490—499.

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

Yang, S., W. Pang, G. Ash, J. Harper, J. Carling, P. Wenzl, E. Huttner, X. Zong, and A. Kilian, 2006: Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology. Theor. Appl. Genet. **113,** 585—595.