

# Low level of genetic diversity in cultivated Pigeonpea compared to its wild relatives is revealed by diversity arrays technology

Shiying Yang · Wen Pang · Gavin Ash ·  
John Harper · Jason Carling · Peter Wenzl ·  
Eric Huttner · Xuxiao Zong · Andrzej Kilian

Received: 7 November 2005 / Accepted: 9 May 2006 / Published online: 15 July 2006  
© Springer-Verlag 2006

**Abstract** Understanding the distribution of genetic diversity among individuals, populations and gene pools is crucial for the efficient management of germplasm collections and breeding programs. Diversity analysis is routinely carried out using sequencing of selected gene(s) or molecular marker technologies. Here we report on the development of Diversity Arrays Technology (DArT) for pigeonpea (*Cajanus cajan*) and its wild relatives. DArT tests thousands of genomic loci for polymorphism and provides the binary scores for hundreds of markers in a single hybridization-based

assay. We tested eight complexity reduction methods using various combinations of restriction enzymes and selected *PstI/HaeIII* genomic representation with the largest frequency of polymorphic clones (19.8%) to produce genotyping arrays. The performance of the *PstI/HaeIII* array was evaluated by typing 96 accessions representing nearly 20 species of *Cajanus*. A total of nearly 700 markers were identified with the average call rate of 96.0% and the scoring reproducibility of 99.7%. DArT markers revealed genetic relationships among the accessions consistent with the available information and systematic classification. Most of the diversity was among the wild relatives of pigeonpea or between the wild species and the cultivated *C. cajan*. Only 64 markers were polymorphic among the cultivated accessions. Such narrow genetic base is likely to represent a serious impediment to breeding progress in pigeonpea. Our study shows that DArT can be effectively applied in molecular systematics and biodiversity studies.

Communicated by D. A. Hoisington

**Electronic Supplementary Material** Supplementary material is available to authorised users in the online version of this article at <http://dx.doi.org/10.1007/s00122-006-0317-z>.

S. Yang · J. Carling · P. Wenzl · E. Huttner · A. Kilian (✉)  
DArT P/L, PO Box 7141, Yarralumla, ACT 2600, Australia  
e-mail: a.kilian@diversityarrays.com

W. Pang · S. Yang  
Guangxi Academy of Agricultural Sciences, Nanning,  
Guangxi 530007, People's Republic of China

G. Ash · J. Harper · S. Yang  
EH Graham Centre for Agricultural Innovation, School of  
Agricultural and Veterinary Sciences, Charles Sturt  
University, PO Box 588, Wagga Wagga, NSW 2678,  
Australia

S. Yang  
Center for the Application of Molecular Biology to  
International Agriculture (CAMBIA), GPO Box 3200,  
Canberra, ACT 2601, Australia

X. Zong  
Chinese Academy of Agricultural Sciences, Beijing 100081,  
People's Republic of China

## Introduction

Molecular marker technologies are increasingly important tools for genetic and genomics studies, breeding and biodiversity research. Currently, several DNA-based molecular marker technologies are available for genetic diversity analysis, e.g. Restriction Fragment Length Polymorphisms (RFLP) (Botstein et al. 1980; Miller 1990), Random Amplified Polymorphic DNA (RAPD) (Williams et al. 1990; Gonzalez 1993), Simple Sequence Repeats (SSR) (Weber and May 1989; Jain et al. 1994), Amplified Fragment

Length Polymorphisms (AFLP) (Mackill et al. 1995; Vos et al. 1995), Single Nucleotide Polymorphisms (SNPs) (Wang et al. 1998) and Diversity Arrays Technology (DArT) (Jaccoud et al. 2001). Most technologies suffer from a combination of impediments with low throughput and the high cost of marker discovery and/or of routine marker assays being most critical limitations. SNP markers are the most abundant class of DNA markers and numerous platforms were recently developed to efficiently analyse SNP polymorphism in humans (Wang et al. 1998) and in some crop species (reviewed by Rafalski 2002). However, the high cost of SNP marker discovery and assay development limits their applicability for many crops, especially for the ‘orphan’ crops like pigeonpea (*Cajanus cajan* (L.) Millsp).

Diversity Arrays Technology (DArT) can overcome these limitations. DArT is a novel genotyping method developed originally using the rice genome (Jaccoud et al. 2001) and recently applied to barley, cassava and *Arabidopsis thaliana* (Wenzl et al. 2004; Xia et al. 2005; Wittenberg et al. 2005). The technology offers low cost, high throughput and sequence-independent genotyping. DArT technology is a solid-state, open-platform method for DNA polymorphism analysis providing comprehensive genome coverage in a single assay. DArT can be rapidly developed for practically any genome, as both marker discovery and routine analysis are carried out using the same, hybridization-based, assay.

Legumes are an integral part of subsistence agriculture since they benefit both humans, by providing protein-rich food, and soils, by improving their nutrition and structure. Pigeonpea is a grain legume crop of the tropics and subtropics, grown in the fields or backyards in more than 90 countries and representing 5% of the total world production of pulses. Pigeonpea has a total cultivated area of approximately 4.2 million hectares and a yield of three million tonnes worldwide, predominately in Asia, Africa, Latin America and Australia. Its protein content averages 24%, and can be as high as 31%. Besides protein-rich food, this environment-friendly crop also provides much needed fodder and fuel wood in the dry areas. Its cultivation over a period of time helps to improve soil fertility and structure (Saxena et al. 2000). It is used as a food and fodder crop and for water conservation and ecological restoration in south China (Yang et al. 2001). The pigeonpea growing area has continued to increase in the last several decades because of increasing recognition of the value of this crop and its multiple uses (Saxena et al. 2000). Unfortunately, due to the small investment in breeding and research of this crop, ge-

netic progress remains slow. Insufficient level of genetic diversity within breeding materials, often a consequence of limited breeding activity, could also significantly impede genetic improvement. However, there is practically no genetic diversity information currently available for pigeonpea. A quick and efficient technology for genotyping and precise germplasm diversity assessment is therefore of great importance for the continued development of this important crop.

Here we report on the development and application of DArT for pigeonpea. The aims of this study were (i) to identify effective complexity reduction method(s) for pigeonpea DArT; (ii) to quantify the level of genetic diversity in cultivated materials and pigeonpea wild relatives; and (iii) to establish genetic relationships among *Cajanus* species based on DArT whole-genome profiles.

## Materials and methods

### Plant material

A total number of 232 accessions of pigeonpea were obtained from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Guangxi Academy of Agricultural Sciences, China Academy of Agricultural Sciences and Australian Tropical Crops and Forages Germplasm Collection (ATCFGC) (Supplementary Table S1). The pigeonpea samples from ICRISAT belong mostly to *C. cajan* species and include accessions from the following countries: India, Australia, Dominican Republic, Grenada, Guyana, Kenya, Tanzania, Malawi. Also included were ICRISAT breeding lines. Samples from ATCFGC included accessions of *C. cajan* and 16 wild species (Table 1) Several leaf samples were collected from plants of a distant relative of pigeonpea, *Cytisus* sp. in Canberra on the campus of the Australian National University. The *Cytisus* sp. samples were used as an outgroup for the phylogenetic analysis.

### DNA extraction and purification

The pigeonpea seeds from ICRISAT and China were planted in the fields in Nanning and in a glasshouse in Beijing, China. The youngest 2–3 leaves from pigeonpea plants of about 30–40 cm in height were snap-frozen in liquid nitrogen, and stored at  $-70^{\circ}\text{C}$  prior to extraction. The samples from ATCFGC were young leaves which were dried at  $50^{\circ}\text{C}$  for 24 h and stored with silica gel until DNA was extracted. All DNA samples were extracted using a modified DNA

**Table 1** Pigeonpea accessions analyzed with *PstI/HaeIII* DArT array

Accession No.	Variety	Taxonomic position	Source country
P002	YunnanW	<i>Cajanus cajan</i>	China
P003	Yunnansu	<i>Cajanus cajan</i>	China
P004	UPAS 120	<i>Cajanus cajan</i>	India
P005	ICPL 88039	<i>Cajanus cajan</i>	India
P009	cms 85010 B line	<i>Cajanus cajan</i>	India
P010	cms 88039-1 A line	<i>Cajanus cajan</i>	India
P014	ICPL 129 R line	<i>Cajanus cajan</i>	India
P016	HPL 24	<i>Cajanus cajan</i>	India
P035	ICP 12044	<i>Cajanus cajan</i>	Tanzania
P041	ICP 13110	<i>Cajanus cajan</i>	Kenya
P056	ICP8863	<i>Cajanus cajan</i>	India
P061	ICP13389	<i>Cajanus cajan</i>	Malawi
P062	ICP13828	<i>Cajanus cajan</i>	Grenada
P063	ICP13961	<i>Cajanus cajan</i>	Dominican Republic
P068	Msprabhatd7	<i>Cajanus cajan</i>	India
P091	ICP12031	<i>Cajanus cajan</i>	Tanzania
P095	ICP12058	<i>Cajanus cajan</i>	Tanzania
P134	ICP13555	<i>Cajanus cajan</i>	Grenada
P156	ICEAP00068	<i>Cajanus cajan</i>	Africa
P158	ICEAP557	<i>Cajanus cajan</i>	Africa
P243	ICPL87119	<i>Cajanus cajan</i>	India
P354	ICPL 93092	<i>Cajanus cajan</i>	India
P371	14Kamya	<i>Cajanus cajan</i>	Kenya
P373	Guangxi Local line 3	<i>Cajanus cajan</i>	China
P374	ICPL 93097	<i>Cajanus cajan</i>	India
P382	ICP 11762	<i>Cajanus cajan</i>	India
P388	ICP 13555	<i>Cajanus cajan</i>	Grenada
P390	ICP 13952	<i>Cajanus cajan</i>	Grenada
P395	ICP 14085	<i>Cajanus cajan</i>	Venezuela
P397	ICP 14163	<i>Cajanus cajan</i>	Indonesia
P398	ICPL 151	<i>Cajanus cajan</i>	India
P403	ANU1	<i>Cytisus species</i>	Australia
P406	ANU4	<i>Cytisus species</i>	Australia
P408	ANU7	<i>Cytisus species</i>	Australia
P410	ANU9	<i>Cytisus species</i>	Australia
P414	Guangxi wild 5	<i>Cajanus sp.</i>	China
P501	AusTRCF310389	<i>Cajanus cajan</i>	India
P502	AusTRCF310438	<i>Cajanus cajan</i>	India
P503	AusTRCF310439	<i>Cajanus cajan</i>	India
P504	AusTRCF310459	<i>Cajanus cajan</i>	Australia
P505	AusTRCF310460	<i>Cajanus cajan</i>	Australia
P506	AusTRCF310461	<i>Cajanus cajan</i>	Australia
P507	AusTRCF316760	<i>Cajanus cajan</i>	India
P508	AusTRCF316763	<i>Cajanus cajan</i>	India
P509	AusTRCF316766	<i>Cajanus cajan</i>	India
P510	AusTRCF317713	<i>Cajanus cajan</i>	Australia
P511	AusTRCF316925	<i>Cajanus acutifolius</i>	Australia
P512	AusTRCF106492	<i>Cajanus albicans</i>	Australia
P515	AusTRCF316931	<i>Cajanus lanuginosus</i>	Australia
P516	AusTRCF316992	<i>Cajanus lanuginosus</i>	Australia
P517	AusTRCF309207	<i>Cajanus latisepalus</i>	Australia
P518	AusTRCF309208	<i>Cajanus latisepalus</i>	Australia
P519	AusTRCF300160	<i>Cajanus marmoratus</i>	Australia
P520	AusTRCF309209	<i>Cajanus marmoratus</i>	Australia
P521	AusTRCF309206	<i>Cajanus pubescens</i>	Australia
P522	AusTRCF300150	<i>Cajanus reticulatus</i>	Australia
P525	AusTRCF100105	<i>Cajanus scarabaeoides</i>	Australia
P526	AusTRCF106053	<i>Cajanus scarabaeoides</i>	Australia
P527	AusTRCF316919	<i>Cajanus scarabaeoides</i>	Australia
P528	AusTRCF55796	<i>Rhynchosia edulis</i>	Australia
P529	AusTRCF320724	<i>Rhynchosia minima</i>	Australia
P530	AusTRCF34133	<i>Rhynchosia verdourtii</i>	Australia

**Table 1** continued

Accession No.	Variety	Taxonomic position	Source country
P531	AusTRCF309210	<i>Cajanus acutifolius</i>	Australia
P532	AusTRCF321876	<i>Cajanus albicans</i>	Australia
P533	AusTRCF309202	<i>Cajanus aromaticus</i>	Australia
P534	AusTRCF300207	<i>Cajanus Confertiflorus</i>	Australia
P535	AusTRCF321881	<i>Cajanus crassus</i>	Australia
P537	AusTRCF321878	<i>Cajanus lanceolatus</i>	Australia
P538	AusTRCF309205	<i>Cajanus latisepalus</i>	Australia
P540	AusTRCF318236	<i>Cajanus platycarpus</i>	Australia
P541	AusTRCF310308	<i>Cajanus cajan</i>	Australia
P542	AusTRCF310443	<i>Cajanus cajan</i>	Australia
P543	AusTRCF310458	<i>Cajanus cajan</i>	Australia
P544	AusTRCF310460	<i>Cajanus cajan</i>	Australia
P545	AusTRCF310477	<i>Cajanus cajan</i>	Australia
P546	AusTRCF321879	<i>Cajanus</i> sp.	Australia
P547	AusTRCF321880	<i>Cajanus</i> sp.	Australia
P548	AusTRCF316916-1	<i>Cajanus acutifolius</i>	Australia
P549	AusTRCF316916-2	<i>Cajanus acutifolius</i>	Australia
P550	AusTRCF316916-3	<i>Cajanus acutifolius</i>	Australia
P551	AusTRCF321877	<i>Cajanus albicans</i>	Australia
P552	AusTRCF321883	<i>Cajanus albicans</i>	Australia
P553	AusTRCF318282	<i>Cajanus crassus</i>	Australia
P554	AusTRCF321882	<i>Cajanus crassus</i>	Australia
P555	AusTRCF318279	<i>Cajanus crassus</i>	Australia
P556	AusTRCF316926-1	<i>Cajanus lanuginosus</i>	Australia
P557	AusTRCF316926-2	<i>Cajanus lanuginosus</i>	Australia
P558	AusTRCF316926-3	<i>Cajanus lanuginosus</i>	Australia
P559	AusTRCF318147-1	<i>Cajanus latisepalus</i>	Australia
P560	AusTRCF318147-2	<i>Cajanus latisepalus</i>	Australia
P561	AusTRCF318147-3	<i>Cajanus latisepalus</i>	Australia
P562	AusTRCF300129	<i>Cajanus lanceolatus</i>	Australia
P563	AusTRCF106127	<i>Cajanus platycarpus</i>	Australia
P564	AusTRCF300150	<i>Cajanus reticulatus</i>	Australia
P565	AusTRCF53911	<i>Flemingia macrophylla</i>	Australia
P566	AusTRCF320724C	<i>Rhynchosia minima</i> var. <i>nuda</i>	Australia

extraction method used for rice (Ronald 1999; Stein et al. 2001). The DNA was further purified by using 2% of Polyvinylpyrrolidone (PVP), (Povidone, K29-32) and sarcosyl to remove the polysaccharides and polyphenolic contaminants. DNA was dissolved in 0.1× TE and adjusted to 100 ng l<sup>-1</sup>.

## DArT procedure

### Preparation of genomic representations

The representations, which are the product of a genome complexity reduction method, were generated using a combination of restriction enzyme (RE) digestion and adapter ligation, followed by amplification (Wenzl et al. 2004). For the *PstI/HaeIII* representation approximately 100 ng of DNA was digested with 2 units of *PstI* and the frequent cutter *HaeIII* (New England Biolabs; NEB, USA) in a buffer containing 10 mM Tris-OAc, 50 mM KOAc, 10 mM Mg(OAc)<sub>2</sub> and 5 mM DTT. A *PstI* adapter (5'-GTT

CAG TCA AGT TAG ATG GTG CA-3' annealed with 5'-CCA TCT AAC TTG ACT G-3') was simultaneously ligated to the complementary overhangs with T4 DNA ligase (NEB). The ligation reaction was diluted 20 fold, and a 1 µl aliquot was used as a template in 50 µl amplification reactions using a primer complementary to the adapter sequence (5'-CAG TCA AGT TAG ATG GTG CAG-3'). A thermal cycling program applicable to all plant species tested so far was used: 94°C for 1 min, followed by 30 cycles of 94°C for 20 s, 58°C for 40 s, 72°C for 1 min, and a final extension of 72°C for 7 min.

The procedures for all other representations were the same as for *PstI/HaeIII* representation, except that the appropriate adaptors and primers were used (Table 2).

### Microarray preparation

Diversity Arrays Technology genomic libraries were prepared and inserts amplified from bacterial clones as described in Wenzl et al. (2004). After the insert

**Table 2** Sequences of primers and adapters used in this study

Oligo name	Adapter sequence	Oligo name	Primer sequence
PstI adapter	5' CACGATGGATCCAGTGCA 3' 5' CTGGATCCATCGTGCA 3'	PstI+0 primer	5' GATGGATCCAGTGCAG 3'
PstI2 adapter	5' GTTCAGTCAAGTTAGATGGTGCA 3' 5' CCATCTAACTTGACTG 3'	PstI2+0 primer	5' CAGTCAAGTTAGATGGTGCA 3'
NdeI adapter	5' CCTCGTAGACTGCGTATCCG 3' 5'TACGGATACGCAGTCTACG 3'	NdeI primer	5'CCTCGTAGACTGCGTATCCG 3'
Bsp1286I adapter	5'CTGAGTAGTGCCAGAACGGTCTGCA 3' 5' GACCGTTCTGGCA3'	Bsp1286I primer	5'CTGAGTAGTGCCAGAACGGTCTGCA 3'
		<sup>a</sup> M 13f primer	5'GTTTTCCCAGTCACGACGTTG3'
		<sup>a</sup> M13r Primer	5'TGAGCGGATAACAATTTTCACACAG3'

Note: <sup>a</sup>M13f primer + M13r primer used for colonies insert amplification

amplification, the amplified inserts were precipitated and the DNA was resuspended in spotting buffer (1 M sucrose + 50% DMSO). The products were then arrayed onto polylysine-coated slides using a microarrayer, either a GMS 417 arrayer (Affymetrix) or a MicroGrid II arrayer (Genomics Solutions). After printing, slides were baked at 80°C for 2 h, incubated in hot water (95°C) for 2 min and dried by centrifugation.

#### Preparation of genomic representations and hybridization to genotyping arrays

Genomic representations were generated from the selected varieties and wild accessions, using the same complexity reduction method used for library construction (*PstI/HaeIII*). Representations were precipitated with one volume of isopropanol, denatured at 95°C for 3 min and labelled with fluorescent dye (1.5 µl of 500 µM Cy3- or Cy5-labelled random decamers synthesized by Sigma, Australia), using the exo-Klenow fragment of *E.coli* DNA polymerase I (NEB). Labelled representations, called “targets”, were added to 50 µl of a 50:5:1 mixture of ExpressHyb buffer (Clontech), 10 g l<sup>-1</sup> herring sperm DNA, and the cy5-labeled or FAM-labelled polylinker fragment of the plasmid used for library preparation as a reference (Jaccoud et al. 2001). After denaturing, labelled targets were hybridized onto microarray surface, covered with a glass coverslip. The slides were placed into a humid chamber at 65°C and incubated overnight.

After 14–16 h of incubation, the coverslips were removed, slides were placed into slide-racks and washed in 1× SSC + 0.1% SDS for 5 min; in 1× SSC for 5 min; in 0.2× SSC for 2 min and in 0.02× SSC for 1 min. Slides were spun-dried immediately at 200× g at room temperature for 7 min.

#### Slide scanning and data extraction

Slides were scanned using a fluorescent microarray scanner (GMS 418 scanner or Tecan LS300 scanner)

and images were generated for each of the fluorescent dyes using the appropriate laser/filter combination.

DARtsoft, a software package developed in-house, was used to automatically analyze each batch of TIF image pairs generated in an experiment, typically comprising a few dozen slides. Two versions of the software were used in this report. The earlier version of DARtsoft, exclusively based on a variance partitioning algorithm (ANOVA), was used in the initial tests of complexity reduction methods. A more advanced version was used in the validation experiments with full-sized arrays. The improved software version localized spots, rejected those with a weak reference signal, computed and normalized the relative hybridization intensities [ $= \log(\text{cy3target}/\text{cy5reference})$  or  $\log(\text{cy3target}/\text{FAMreference})$ ], calculated the median value for replicate spots, identified polymorphic clones by using a combination of ANOVA and fuzzy *K*-means clustering at a fuzziness level of 1.5 and classified polymorphic clones as being present (“1”) or absent (“0”) in the representation hybridized to a slide. The clustering algorithm also provided a probability estimate for each individual genotype call (Cayla et al. in preparation).

#### Evaluation of complexity reduction methods and selection of genotyping method

Two 1536-clone *PstI/HaeIII* libraries (each developed with a different adapter sequence) were produced from a mixture of the DNA of 48 selected pigeonpea accessions. Five 768-clone libraries of NdeI + Bsp1286I fragments, each recut with different frequent cutting enzyme, were generated using the same 48 DNA samples (Supplementary Table 1). The percentage of polymorphic clones in each of these libraries was calculated using DARtsoft with the same parameter settings.

The *PstI/HaeIII* library was expanded by 5,376 clones to produce a first-generation genotyping array using *PstI/HaeIII* with PstI2 + 0 adapter/primer. This

array was used for genotyping 96 pigeonpea accessions (including cultivated and wild materials).

#### Analysis of DArT data

The DArTsoft-generated 0–1 scores were used as input for the RESTDIST and NEIGHBOR programs of the PHYLIP 3.6 software package to construct an Unweighted Pair Group Method with Algorithmic Mean (UPGMA) dendrogram based on Felsenstein's modification of the Nei/Li restriction fragment distance (Felsenstein 1989, 2004). The DArTsoft-generated Hamming distance matrices were used as input for the MEGA3 (Kumar 2004) software package to build a UPGMA dendrogram. The Hamming distance matrices were also used as input into software package for Principal Coordinate Analysis (PCO) (Anderson 2003).

The Polymorphism Information Content (PIC), a measure of informativeness of a genetic marker, was calculated according to Anderson et al. (1993):

$$PIC = 1 - \sum_{i=1}^n p_i^2$$

where  $p_i$  is the population frequency of the  $i$ th allele and  $n$  the total number of allelic states.

## Results

### Optimization of complexity reduction methods and library construction

Producing suitable genomic representations is the key step for DArT technology development. It was shown for small genome rice (430 Mbp) and large genome barley (5,000 Mbp) that digestion with *Pst*I RE together with a frequently cutting RE, combined with adapter ligation-based amplification of intact *Pst*I fragments is an efficient method of DArT (Jaccoud et al. 2001; Wenzl et al. 2004). Similar methods of complexity reduction were also effective in discovering DArT polymorphism in cassava with genome size in-between rice and barley and similar to that of pigeonpea, which has been estimated to be around 1 pg per haploid genome (Ohri et al. 1994; Ohri and Singh 2002).

There were significant differences in polymorphism detection efficiency observed among the complexity reduction methods using different frequently cutting RE both in barley and in cassava (Wenzl et al. 2004; Xia et al. 2005). We therefore started developing DArT for pigeonpea by testing a large number of combinations of *Pst*I and one of the following frequent cutting RE: *Apo*I, *Alu*I, *Bst*NI, *Ban*II, *Taq*I, *Mse*I,

*Rsa*I, *Bso*BI, *Bsp*1286I and *Hae*III. We analysed the representation by electrophoresis on a 2% agarose gel and excluded all representations showing one or more strong bands, since highly abundant amplicons would result in highly redundant clones in the library created from such representations. Only the *Pst*I/*Hae*III representation was free from observable bands (data not presented). Therefore, we selected this representation to construct the first library.

Nearly all DArT complexity reduction methods reported so far have used *Pst*I RE to generate adaptor-compatible overhangs, either alone (Wenzl et al. 2004; Xia et al. 2005) or in combination with *Eco*RI (Wittenberg et al. 2005). We tested a new enzymes combination, *Nde*I + *Bsp*1286I, as adaptor-compatible REs, and used them in combination with either a single frequent cutting RE (*Taq*I, *Bst*NI, *Apo*I, *Alu*I, *Ban*II, *Bso*BI, *Hae*III, *Mse*I, *Rsa*I and *Dpn*II) or a pair of REs (*Alu*I + *Bst*NI, *Alu*I + *Bso*BI, *Alu*I + *Dpn*II, *Bso*BI + *Bst*NI, *Bso*BI + *Dpn*II and *Bst*NI + *Dpn*II). Based on the agarose gel analysis of these representations we built another five libraries: *Nde*I + *Bsp*1286I/*Dpn*II, *Nde*I + *Bsp*1286I/*Bst*NI, *Nde*I + *Bsp*1286I/*Alu*I + *Bso*BI, *Nde*I + *Bsp*1286I/*Alu*I + *Dpn*II and *Nde*I + *Bsp*1286I/*Bst*NI + *Taq*I.

A set of 48 accessions (Supplementary Table I) was used to test the frequency of polymorphic clones on the arrays containing either 768 or 1,536 clones derived from relevant representations. Approximately eight fold difference in polymorphism frequency was observed among the complexity reduction methods (Table 3). The highest levels of polymorphism were observed for

**Table 3** Polymorphism level obtained for the complexity reduction methods tested

Complexity reduction method	Number of polymorphic clones <sup>a</sup>	Total number of clones	Frequency of polymorphic clones (%)
<i>Pst</i> I/ <i>Hae</i> III (F) <sup>b</sup>	86	1,536	5.9
<i>Pst</i> I/ <i>Hae</i> III (G) <sup>b</sup>	90	1,536	5.6
<i>Nde</i> I + <i>Bsp</i> 1286I/ <i>Dpn</i> II	6	768	0.8
<i>Nde</i> I + <i>Bsp</i> 1286I/ <i>Bst</i> NI	21	768	2.7
<i>Nde</i> I + <i>Bsp</i> 1286I/ <i>Alu</i> I + <i>Bso</i> BI	41	768	5.2
<i>Nde</i> I + <i>Bsp</i> 1286I/ <i>Alu</i> I + <i>Dpn</i> II	19	768	2.5
<i>Nde</i> I + <i>Bsp</i> 1286I/ <i>Bst</i> NI + <i>Taq</i> I	31	1,536	2.0

<sup>a</sup>Polymorphic clones were detected by DArTsoft program using standard settings

<sup>b</sup>*Pst*I/*Hae*III representations were prepared using either *Pst*I + 0 (F) or *Pst*I2 + 0 (G) adaptors/primers, respectively

the two *PstI/HaeIII* representations generated with either *PstI* + 0 or *PstI2* + 0 primers (5.9 and 5.6%, respectively). The small difference observed between the two *PstI/HaeIII* representations could be attributed to between experimental variation or differences in performance of respective adapters/primers. The *NdeI* + *Bsp1286I/AluI* + *BsoBI* complexity reductions method was only slightly less polymorphic (5.2%). The lowest frequency of polymorphism was detected for the *NdeI* + *Bsp1286I/DpnII* complexity reduction method (0.8%). Based on these results, we developed a full-size array using the *PstI/HaeIII* method.

#### Development of a full-size array

Firstly, we developed a *PstI/HaeIII* array containing 3,456 random clones from cultivated materials listed in Table 1. Next, we expanded this array by using additional *C. cajan* samples from Australia, India and China, one wild species from China and five wild species from Australia. We identified 760 potential polymorphic clones in a total of 3,072 clones (24.7% of polymorphic clones) in a genotyping experiment involving the samples used for array development. The UPGMA dendrogram based on the *Nei/Li* restriction fragment distance showed clear separation between the *C. cajan* accessions and the wild accessions. The wild species from China and Australia clustered in one group, while all the cultivars clustered into two close subgroups (data not shown).

Finally, we expanded this library array to 5,376 random clones by using 96 additional samples (48 samples from 14 wild relatives from Australia and 48 cultivated accessions). The 48 cultivated accessions were selected based on their highest genetic dissimilarity in preliminary DArT assays involving 300 accessions (data not presented). The selection of initial set of 300 lines was done to maximize diversity in qualitative and quantitative characters in combination with passport information (K.B. Saxena, personal communication). With the first level of selection based on phenotypic and geographic diversity and the second level based on DNA polymorphism we were confident that the 48 lines used for final array development and polymorphism detection (below) represent well genetic diversity of cultivated pigeonpea germplasm.

#### Evaluation of the performance of full-size array

In order to evaluate the performance of the expanded *PstI/HaeIII* array we analyzed the genetic diversity of 96 accessions used for array development. Thirty two accessions were analysed in duplicate, and the

remaining 64 accessions were analyzed on a single microarray. An improved version of DArTsoft (version 7.4) was used for data analysis and the application of highly stringent criteria resulted in the identification of 696 polymorphic clones on the array. The average call rate was 96.04% and the average scoring reproducibility was 99.7%. DArT markers displayed high polymorphism information content (PIC) values, ranging from 0.02 to 0.50 with an average of  $0.34 \pm 0.15$  and a median of 0.42.

#### Genetic diversity of cultivated accessions

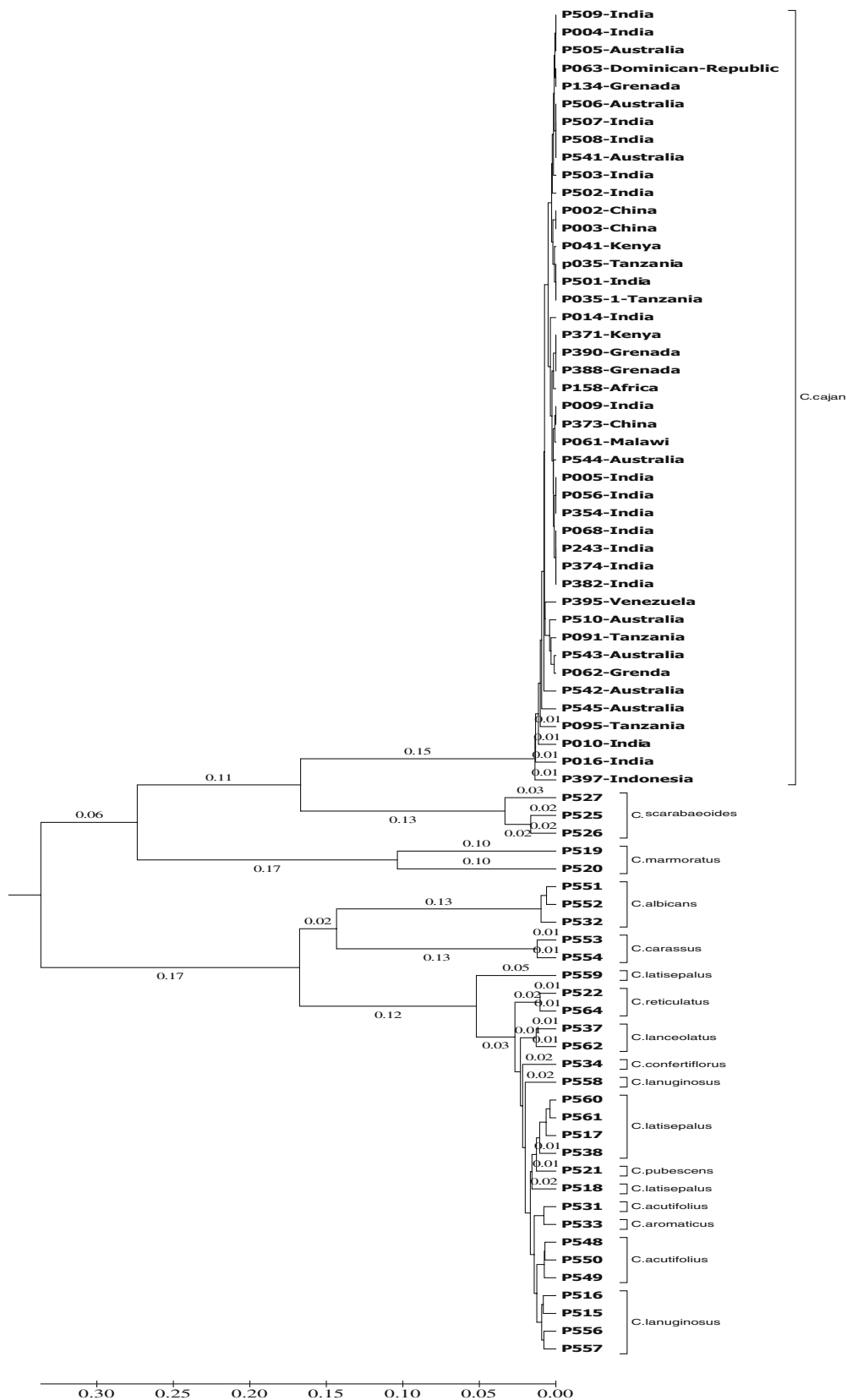
The cultivated accessions (all belonging to *C. cajan*), selected to represent a significant variation for a number of morphological traits, showed very little diversity in DArT analysis. Out of nearly 700 markers identified with the full size genotyping array only 64 markers detected variation among the 48 *C. cajan* samples. Importantly, 50% of these markers (30) had one of the two alleles present at low frequency (below 5%) among the cultivated accessions. Such low PIC markers contributed to the low level of genetic differentiation among the pigeonpea cultivars.

There was no clear differentiation among the cultivars from various geographic areas of pigeonpea cultivation, with Africa (represented by eight accessions) being the most diverse region (Supplementary Figure 1).

#### DArT marker reflect the relationship among pigeonpea and related species

A dendrogram based on the UPGMA method was constructed based on DArT markers for 76 accessions of pigeonpea (Genus *Cajanus*, Fig. 1). This cluster analysis discriminated well between 14 species of pigeonpea. At an evolutionary distance threshold of 0.11, all genotypes were divided into six distinct clusters. The first cluster contained 44 lines of cultivated species *C. cajan* which were collected from tropical and subtropical countries of Asia, Africa, Australia and South America. Consistently with the results reported above, there was no clear separation of the accessions based on the country/continent of origin, although some grouping of samples with similar provenience could be observed. The second group included three lines of the species *Cajanus scarabaeoides*, the sexually compatible, closest relative of cultivated *C. cajan*. The third group included two lines of the species *C. marmoratus*, the fourth included three lines of *C. albicans* and the fifth two lines of *C. crassus*. The sixth cluster consisted of lines from all remaining eight species: *C. latiseptalis*,

**Fig. 1** An UPGMA dendrogram representing 76 accessions of pigeonpea constructed with MEGA version 3.0 software (Kumar and Nei 2004). The clustering was performed on the dissimilarity matrix 696 DArT markers from the full size *PstI/HaeIII* array. At a genetic distance index of 0.11 and above, genotypes were divided into six distinct clusters. The information about the accessions analysed is provided in Table 1





*C. reticulates*, *C. lanceolatus*, *C. confertiflorus*, *C. lanuginosus*, *C. pubescens*, *C. acutifolius* and *C. aromaticus*.

The principal coordinate analysis (Anderson 2003) plot (Fig. 2) shows the picture of diversity within the genus *Cajanus* and several related genera. The picture for genus *Cajanus* is very consistent with the dendrogram described above. The low level of diversity among the 44 cultivated genotypes is clearly visible, with just a few genotypes (ICEAP00068, UPAS120, HPL24 and Quest) slightly separated from the tight cluster of remaining accessions. All cultivated accessions are well separated from all 18 wild relatives. Only five wild *Cajanus* species (*C. scarabaeoides*, *C. marmoratus*, *C. platycarpus*, *C. albicans* and *C. crassus*) form independent groups. However, discrimination of these groups can be influenced by the small number of accessions per species. Similarly to the results of the clustering analysis, all remaining wild *Cajanus* species (*C. acutifolius*, *C. aromaticus*, *C. confertiflorus*, *C. latisepalis*, *C. lanceolatus*, *C. lanuginosus*, *C. pubescens* and *C. reticulates*) formed one tight group. Interestingly, our observations of the glasshouse and field grown plants from a number of these species (e.g. *C. latisepalis*, *C. lanceolatus*, *C. lanuginosus* and *C. pubescens*) identified strong similarities of their leaf shape, leaf and flower color and the growth habit. All

these species are endemic to Australia. The *Cytisus sp.* samples are clearly the farthest away from *C. cajan* on the PCO plot. Their spread on the plot indicates substantial level of genetic diversity in the morphologically very similar samples (Fig. 2). This is in contrast to situation in *C. cajan*, for which morphological variation was apparently much higher than the level of DNA variation revealed by DArT.

## Discussion

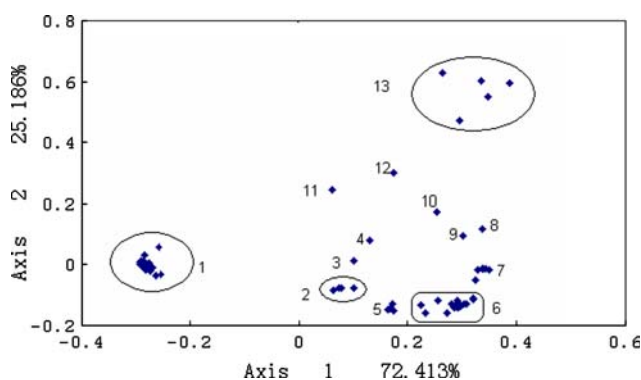
This is the first report of the use of DArT technology in pigeonpea and the first comprehensive systematics study of the *Cajanus* genus using molecular markers. Our results demonstrate that DArT markers are of good quality, as measured by their high PIC value, call rate and scoring reproducibility. All the marker quality parameters reported here are very similar to those obtained for barley (Wenzl et al. 2004) and cassava (Xia et al. 2005).

DArT markers performed well in the analysis of 17 species of pigeonpea, grouping samples with similar morphology and/or systematic position. Genetic distance analysis shows that pigeonpea genetic differentiation mainly occurred between *C. cajan* and its different wild relatives. Both the UPGMA dendrogram and the principal coordinate analysis plot showed a similar picture of genetic differentiation among the *Cajanus* samples, grouping them into six major groups. The classification into six groups is also consistent with the taxonomic classification of these species by van der Maesen (1990). Two species grouped together in the PCO plot, *C. acutifolius* and *C. latisepalis*, were also very similar in the RFLP analysis of mitochondrial DNA (Sivaramakrishnan et al. 2002).

Our molecular marker analysis in combination with our morphological observations of plants in controlled environment (data not presented) suggest the need for further molecular and phenotypic studies to assess the systematic position of several species in the genus *Cajanus*. Such explorations should include several related genera, since the accessions from *Flemingia macrophylla*, *Rhynchosia minima* var. *nuda*, *Rhynchosia edulis* and *Rhynchosia verdcourtii* are not clearly separated from several *Cajanus* species on the PCO plot.

Our data enabled the classification of several samples for which the systematic position could not be asserted based on morphological evaluation by the staff of the Australian Tropical Crops and Forages Germplasm Collection (S. Dillon, personal communication).

Based on their position in the PCO plot and the dendrogram the sample P546 (Aus TRCF 321879)



**Fig. 2** A Principal Coordinate Analysis (PCO) plot in representing the position of samples analysed in the two dimensional space of Principal Coordinate 1 (PC1) (72.4% of total variance) and the PC2 (25.2% of total variance). The PCO analysis was performed using 278 highest quality DArT markers for the 96 accessions of pigeonpea using the software by Anderson (2003). The samples (complete list in Table 1) were grouped according to their position on the PCO plot into the following 13 groups: 1 *Cajanus cajan*; 2 *Cajanus scarabaeoides*; 3 *Cajanus marmoratus*; 4 *Cajanus platycarpus*; 5 *Cajanus albicans*; 6 *Cajanus latisepalis*, *Cajanus lanceolatus*, *Cajanus lanuginosus*, *Cajanus pubescens*, *Cajanus reticulates*, *Cajanus acutifolius*, *Cajanus aromaticus*, *Cajanus confertiflorus* 7 *Cajanus crassus*; 8 *Rhynchosia minima*; 9 *Rhynchosia edulis*; 10 *Rhynchosia minima* var. *nuda*; 11 *Rhynchosia verdcourtii*; 12 *Flemingia macrophylla*; 13 *Cytisus species*

could be classified as *C. latiseptalis* and the sample P548 (Aus TRCF 321880) as *C. scarabaeoides*.

Our study revealed a very low level of DNA polymorphism in cultivated *Cajanus*, a result reminiscent of the recent findings in groundnut (Moretzsohn et al. 2004) and chickpea (Berger et al. 2003). Low level of genetic diversity within *C. cajan* is likely to represent a significant impediment in any genetic improvement program for this crop. We therefore strongly encourage any attempt to broaden the genetic base of cultivated *C. cajan*, especially through genetic crossing with wild relatives. Advanced backcross QTL (AB QTL) technology has been used productively for such purpose in species like tomato (Frery et al. 2004), rice (Septiningsih et al. 2003) and wheat (Huang et al. 2003), but such approaches are constrained by the high cost of the whole-genome genotyping required. The availability of an inexpensive whole genome DNA profiling technology like the one reported here is therefore likely to contribute significantly to the effective utilization of genetic diversity of wild relatives of cultivated pigeonpea.

**Acknowledgments** The authors would like to thank Dr. K. B. Saxena of ICRISAT, Patancheru, India, and Sally Dillon of ATGCRC, for providing seeds; Dr. H. D. Upadhyaya of ICRISAT, for providing pedigree information; Dr. Rongbai Li, Ms Yuanhua Chen of Guangxi Academy of Agricultural Sciences, Nanning, China and Ms Jianping Guan and Mr. Long Yuan of Chinese Academy of Agricultural Sciences, Beijing, China for helping with DNA extraction; Ms Vanessa Caig and Ms Margaret Evers, for technical help; Mr. Cyril Cayla, Mr. Grzegorz Uszynski, Dr. Yuantu Huang and Damian Jaccoud for help with data analysis. We thank all the other colleagues at DArT P/L for helpful discussions. Special thanks to Center for the Application of Molecular Biology to International Agriculture and The Finkel Family Trust for financial support for Shiyang Yang.

## References

- Anderson M (2003) PCO: a FORTRAN computer program for principal coordinate analysis. University of Auckland, New Zealand
- Anderson JA, Churchill G.A, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. *Genome* 36:181–186
- Berger J, Abbo S, Turner NC (2003) Ecogeography of annual wild *Cicer* species: the poor state of the world collection. *Crop Sci* 43:1076–1090
- Botstein D, Skolnick M, Davis R (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32(3):314–331
- Felsenstein J (1989) PHYLIP—phylogeny inference package (Version 3.2). *Cladistics* 164–166
- Felsenstein J (2004) PHYLIP (Phylogeny Inference Package) version 3.6. University of Washington, Seattle
- Frery A, Fulton TM, Zamir D, Tanksley SD (2004) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *L. pennellii* cross and identification of possible orthologs in the Solanaceae. *Theor Appl Genet* 108:485–96
- Gonzalez J (1993) Random amplified polymorphic DNA analysis in *Hordeum* species. *Genome* 36:1029–1031
- Huang XQ, Coster H, Ganai MW, Roder MS (2003) Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). *Theor Appl Genet* 106:1379–89
- Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* 29(4):e25
- Jain A, Bhatia S, Banga S, Lakshmikumaran M (1994) Potential use of random amplified polymorphic DNA (RAPD) technique to study the genetic diversity in Indian mustard (*Brassica juncea*) and its relationship to heterosis. *Theor Appl Genet* 88:116–122
- Kumar S, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163
- van der Maesen LJG (1990) Pigeonpea: origin, history, evolution, and taxonomy. In: Nene YL, Hill SH, Sheila VK (eds) *The pigeonpea*. CABInternational, Wellingford, UK, P15–46
- Mackill D (1995) Classifying *Japonica* rice cultivars with RAPD markers. *Crop Sci* 35:889–894
- Miller J (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theor Appl Genet* 80:437–448
- Moretzsohn M de C, Hopkins MS, Mitchell SE, Kresovich S, Valls JF, Ferreira ME (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* 14(4):11
- Ohri D, Singh S (2002) Karyotypic and genome size variation in *Cajanus cajan* (L.) Millsp (pigeonpea) and some wild relatives. *Genet Resour Crop Evol* 49:1–10
- Ohri D, Jha S, Kumar S (1994) Variability in nuclear-DNA Content within Pigeonpea, *Cajanus cajan* (Fabaceae). *Plant Syst Evol* 189:211–2
- Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* 5:94–100
- Ronald D (1999) A rapid DNA minipreparation method suitable for AFLP and other PCR applications. *Plant Mol Biol Rep* 17:53–57
- Saxena K, Yang S (2000) ICRISAT Pigeonpea Jumps over Himalayas. *Int Chickpea and Pigeonpea Newslett* 2–3
- Saxena K, Zong X, Yang S, Li Z, Zhou C (2000) Potential of Pigeonpea in China and its Genetic Improvement at ICRISAT, ICETS 2000, Beijing
- Septiningsih EM, Trijatmiko KR, Moeljopawiro S, McCouch SR (2003) Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theor Appl Genet* 107:1433–1441
- Sivaramakrishnan, Reddy L (2002) Diversity in selected wild and cultivated species of pigeonpea using RFLP of mtDNA. *Euphytica* 125:121–128
- Stein N (2001) A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum*. *Plant Breed* 120:354–356
- Vos P, Hoger R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Rijters A, Pot J, Peleman J, Kuiper M et al (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23:4407–4414

- Wang D, Fan J, Siao C, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J et al (1998) Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280(5366):1077–1082
- Weber J, May P (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44(3):388–396
- Wenzl P, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity arrays technology (DArT) for whole-genome profiling of barley. *Proc Natl Acad Sci USA* 101(26):9915–9920
- Williams J, Kubelik A, Livak K, Rafalski J, Tingey S (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18(22):6531–6535
- Wittenberg AH, Van der Lee T, Cayla C, Kilian A, Visser RG, Schouten HJ (2005) Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. *Mol Gen Genomics*, 274:30–39
- Xia L, Peng K, Yang S, Wenzl P, Carmen de Vicente M, Fregene M, Kilian A (2005) DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. *Theor Appl Genet* 110:1092–1098 DOI:101007/s100122-101005-101937-101004
- Yang S, Pang W, Zong X, Li Z, Zhou C, Saxena K, Liang H (2001) A potential fodder crop for Guangxi province of China. *Int Chickpea Pigeonpea Newslett* 54