

HOSTED BY



Contents lists available at ScienceDirect

Electronic Journal of Biotechnology



Research article

SSR genetic diversity assessment of popular pigeonpea varieties in Malawi reveals unique fingerprints



Vincent Njung'e^{a,*}, Santosh Deshpande^b, Moses Siambi^a, Richard Jones^{a,c}, Said Silim^{a,e}, Santie De Villiers^{a,d}

^a International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)—Nairobi, PO Box 39063-00623, Nairobi, Kenya

^b International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)—India, Patancheru 502 324, Andhra Pradesh, India

^c Alliance for a Green Revolution in Africa (AGRA), P.O. Box 66773, Westlands, 00800, Kenya

^d Pwani University, PO Box 195-80108, Kilifi, Kenya

^e International Center For Agricultural Research In The Dry Areas (ICARDA), P.O. Box 5689, Addis Ababa, Ethiopia

ARTICLE INFO

Article history:

Received 30 May 2015

Accepted 15 March 2016

Available online 3 April 2016

Keywords:

Genetic variation

Genotyping

Microsatellites

Seed purity

ABSTRACT

Background: Pigeonpea (*Cajanus cajan* (L.) Millsp.) is a drought tolerant legume of the Fabaceae family and the only cultivated species in the genus *Cajanus*. It is mainly cultivated in the semi-arid tropics of Asia and Oceania, Africa and America. In Malawi, it is grown as a source of food and income and for soil improvement in intercropping systems. However, varietal contamination due to natural outcrossing causes significant quality reduction and yield losses. In this study, 48 polymorphic SSR markers were used to assess the diversity among all pigeonpea varieties cultivated in Malawi to determine if a genetic fingerprint could be identified to distinguish the popular varieties.

Results: A total of 212 alleles were observed with an average of 5.58 alleles per marker and a maximum of 14 alleles produced by CCTtc019 (Marker 40). Polymorphic information content (PIC), ranged from 0.03 to 0.89 with an average of 0.30. A neighbor-joining tree produced 4 clusters. The most commonly cultivated varieties, which include released varieties and cultivated land races, were well-spread across all the clusters observed, indicating that they generally represented the genetic diversity available in Malawi, although substantial variation was evident that can still be exploited through further breeding.

Conclusion: Screening of the allelic data associated with the five most popular cultivated varieties, revealed 6 markers – CCB1, CCB7, Ccac035, CCTtc003, Ccac026 and CCTtc019 – which displayed unique allelic profiles for each of the five varieties. This genetic fingerprint can potentially be applied for seed certification to confirm the genetic purity of seeds that are delivered to Malawi farmers.

© 2016 Pontificia Universidad Católica de Valparaíso. Production and hosting by Elsevier B.V. All rights reserved.

1. Introduction

Pigeonpea, *Cajanus cajan* [L.] Millsp. is a drought tolerant crop and one of the most important legumes grown in the tropics and sub tropics. As a rich source of protein for humans it is largely used to supplement cereal staples [1,2] and is also a good source of fodder. In the southern Great Plains of the United States of America, pigeonpea provides high quality forage when other feeds are less productive [3] and poultry performs particularly well when pigeonpea is included in their diet [4,5]. The extensive and deep root system of pigeonpea fixes atmospheric nitrogen and improves the quality and structure of soils [6] and perennial pigeonpea types provide fuel wood and material for basket weaving and roofing in African villages [7]. Due to this

versatility, pigeonpea is an established and valued crop among small-scale farmers in Malawi.

Analyzing genetic relationships in species is important for revealing diversity. In addition to displaying the existing variability among cultivars [8] genetic diversity provides valuable information on target trait availability and diversity for successful breeding programs [9,10]. Molecular markers are useful tools for genetic diversity assessment of various crops and among them simple sequence repeats (SSRs) are popular since they reveal more variation e.g. in pea [11], rice [12], maize [13] and wheat [14]. This is also the case for pigeonpea [15,16,17,18,19,20]. In fact SSRs in pigeonpea are bound to be more informative as most have now been mapped in the pigeonpea genome [21,22].

Pigeonpea production in Malawi has increased from 64 to 193 kt between 2005 and 2010 making Malawi Africa's top pigeonpea producer in 2010 [23]. It is an economically important crop for small-scale farmers especially in southern Malawi as it provides food security, is highly nutritious, improves soils and serves as a valuable

* Corresponding author.

E-mail address: vnjung@cgiar.org (V. Njung'e).

Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

cash crop [24]. Traditional varieties are predominantly cultivated, leaving potential for increased production if farmers had access to improved high yielding varieties [25]. In addition, production suffers greatly from low quality seeds that result from mixing and/or contamination with pathogen propagules. Natural outcrossing, which can be as high as 45% in pigeonpea, is the major source of varietal contamination [26] and causes significant yield losses for farmers in Malawi. This is further exacerbated by the lack of effective channels to avail sufficient high quality seeds to farmers by various stakeholders [27]. Besides, since genetic purity directly affects pigeonpea yields, access to pure seeds is essential [28]. It is therefore important to determine the general level of purity of each pigeonpea variety available in Malawi and to what extent varieties become mixed. This will help determine how to maintain seed purity. This study aimed to assess the diversity of all known pigeonpea varieties cultivated in Malawi using SSR markers and to determine if the released varieties could be distinguished to establish a basis for future tracking of dissemination and adoption of improved and released varieties.

2. Materials and methods

2.1. DNA extraction

Seventy nine varieties, (Table 1) representing all accessions held in the Malawi gene bank as well as released varieties of pigeonpea in Malawi, and one variety (ICP 2309) used as a control in allele scoring, were obtained with the assistance of ICRISAT-Lilongwe and planted in Nairobi, Kenya in a screen house. Two weeks after germination, DNA

was extracted from leaves of 5 individual seedlings of each genotype according to the protocol described by Mace et al. [29], omitting the phenol: chloroform extraction step. However, for accessions that failed to germinate at all or produced fewer than 5 seedlings DNA was extracted from the seeds using the protocol described by Sharma et al. [30]. The homogenization solution was modified to contain 5 M NaCl, 2% (w/v) Sarcosyl, 100 mM Tris and 20 mM EDTA. For all samples, DNA quality was determined by agarose gel electrophoresis (0.8% (w/v)) stained with 5 µL/100 mL Gel Red® (Biotium Inc., USA) while the quantity was determined by spectrophotometry (Nanodrop® 1000, Thermo Scientific, USA). All the DNA samples were then diluted to 10 ng/µL and used for PCR.

2.2. SSR amplification and optimization

PCR was done using 48 publicly available polymorphic markers (Table S1.). All the forward primers contained a 5'-M13 tag (CACGAC GTTGTAACGAC) to allow incorporation of a fluorochrome during the PCR process [31]. The fluorochromes used were 6-Carboxyfluorescein (FAM®), NED®, VIC® and PET®; Life Technologies Corporation, USA). Each PCR reaction contained 1 × PCR buffer (20 mM Tris-HCl (pH 7.6); 100 mM KCl; 0.1 mM EDTA; 1 mM DTT; 0.5% (v/v) Triton X-100; 50% (v/v) glycerol), 2 mM MgCl₂, 0.16 mM dNTPs, 0.16 µM of a labeled M13-primer, 0.04 µM M13-forward primer, 0.2 µM reverse primer, 0.2 units of *Taq* DNA polymerase (SibEnzyme Ltd., Russia) and 30 ng of template DNA. The volume for each PCR was topped up with sterile water to a final volume of 10 µL. Reactions were performed on a thermocycler (GeneAmp PCR system 9700®, Applied Biosystems, USA)

Table 1
List of genotypes used in this study.

Unit	Sample ID	Type	Maturity	Unit	Sample ID	Type
1	ICEAP00040	Released Variety	Long	41	MW2291	Malawi Genebank
2	ICP9145	Landrace	Long	42	MW2292	Malawi Genebank
3	ICEAP00557	Released Variety	Medium	43	MW2295	Malawi Genebank
4	ICPV87105	Landrace	Short	44	MW2296	Malawi Genebank
5	Mtawanjuni	Landrace	Medium	45	MW2298	Malawi Genebank
6	ICEAP00020	Released Variety	Long	46	MW2299	Malawi Genebank
7	ICEAP00068	Released Variety	Medium	47	MW2300	Malawi Genebank
8	ICP13076	Landrace	Long	48	MW2302	Malawi Genebank
9	Kat60/8	Released Variety	Short	49	MW2303	Malawi Genebank
10	ICP2309	Reference Variety		50	MW2305	Malawi Genebank
11	MW2047	Malawi Genebank		51	MW2306	Malawi Genebank
12	MW2097	Malawi Genebank		52	MW2308	Malawi Genebank
13	MW2238	Malawi Genebank		53	MW2309	Malawi Genebank
14	MW2243	Malawi Genebank		54	MW2311	Malawi Genebank
15	MW2244	Malawi Genebank		55	MW2313	Malawi Genebank
16	MW2245	Malawi Genebank		56	MW2317	Malawi Genebank
17	MW2251	Malawi Genebank		57	MW2321	Malawi Genebank
18	MW2256	Malawi Genebank		58	MW2323	Malawi Genebank
19	MW2258	Malawi Genebank		59	MW2240	Malawi Genebank
20	MW2261	Malawi Genebank		60	MW2324	Malawi Genebank
21	MW2263	Malawi Genebank		61	MW2325	Malawi Genebank
22	MW2264	Malawi Genebank		62	MW2326	Malawi Genebank
23	MW2265	Malawi Genebank		63	MW2327	Malawi Genebank
24	MW2266	Malawi Genebank		64	MW2328	Malawi Genebank
25	MW2267	Malawi Genebank		65	MW2331	Malawi Genebank
26	MW2268	Malawi Genebank		66	MW2332	Malawi Genebank
27	MW2269	Malawi Genebank		67	MW2333	Malawi Genebank
28	MW2270	Malawi Genebank		68	MW2336	Malawi Genebank
29	MW2271	Malawi Genebank		69	MW2355	Malawi Genebank
30	MW2276	Malawi Genebank		70	MW326	Malawi Genebank
31	MW2279	Malawi Genebank		71	MW454	Malawi Genebank
32	MW2281	Malawi Genebank		72	MW470	Malawi Genebank
33	MW2282	Malawi Genebank		73	MW480	Malawi Genebank
34	MW2283	Malawi Genebank		74	MW587	Malawi Genebank
35	MW2284	Malawi Genebank		75	MW648	Malawi Genebank
36	MW2285	Malawi Genebank		76	MW690	Malawi Genebank
37	MW2286	Malawi Genebank		77	MW765	Malawi Genebank
38	MW2287	Malawi Genebank		78	MW786	Malawi Genebank
39	MW2288	Malawi Genebank		79	MW793	Malawi Genebank
40	MW2289	Malawi Genebank				

with initial denaturation of 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, 59°C for 1 min and 72°C for 2 min followed by final elongation at 72°C for 20 min. For markers that did not amplify with this PCR protocol, changes were made only to the annealing temperature using the published annealing temperatures for the respective markers, followed by testing annealing temperatures calculated using the SSR primer sequences in the first step of BioMath Calculators (<http://www.promega.com/a/apps/biomath/index.html?calc=tm>). A group of eleven SSRs primers that failed to amplify after these annealing temperature adjustments, were submitted to gradient PCR using a Techne TC-5000 Thermo cycler®, (Bibby Scientific Group, United Kingdom), which allocated different annealing temperatures to each column in a 96-well PCR plate. The temperatures used for this study were between 48.8°C in column 1 to 61.1°C in column 12, with intervals of 1.05°C increasing for each of the 12 columns.

Amplification was confirmed by electrophoresis using 2% (w/v) agarose gel stained with GelRed® (Biotium, USA) and visualized under UV light. Depending on the efficiency of amplification, 2.5 µL–3.5 µL of 3 to 4 different amplification products were co-loaded along with the size standard, GeneScan™-500 LIZ® (Applied Biosystems, USA) and Hi-Di™ Formamide (Applied Biosystems, USA) and separated by capillary electrophoresis using an ABI Prism® 3730 Genetic analyzer (Applied Biosystems, USA) [32].

2.3. Allele calling and fingerprinting

Fragment analysis was performed with Gene Mapper 4.0 (Applied Biosystems, USA) and allelic data for each marker analyzed with PowerMarker V3.25 [33] and DARwinV.5.0.158 software [34]. A table of summary statistics including major allele frequencies, number of alleles, heterozygosity and polymorphic information content (PIC) of each marker was produced using PowerMarker software. DARwin software was used to produce a pair-wise dissimilarity matrix calculated using the formula described in [Equation 1]:

$$d_{ij} = 1 - \frac{1}{L} \sum_{l=1}^L \frac{m_l}{\pi} \quad [\text{Equation 1}]$$

where d_{ij} was the dissimilarity between units i and j , L was the number of loci, π was the ploidy and m_l was the number of matching alleles for locus l . This matrix was then used in displaying the unweighted neighbor-joining dendrogram and the principle coordinate analysis with the same software.

Allelic results were investigated to identify markers with the potential to provide a DNA fingerprint for cultivated and released pigeonpea varieties from Malawi. The ideal fingerprinting markers were considered to be those that can unambiguously discern all the varieties from one another. It was highly unlikely that a single marker would fit these criterion and more likely that a set of markers would have to be considered for this purpose. In order to identify such a set of markers, the following steps were followed. First, the allelic data for the target varieties were selected from the complete dataset and considered in isolation from the gene bank and reference data. Second, the data were screened to eliminate all the markers that had low success in PCR amplification (presented $\geq 40\%$ missing data), were monomorphic or heterogeneous (provided multiple different alleles within an accession). If a marker presented a different allele for two individuals within an accession, it was considered heterogeneous and not included. If a marker presented a different allele for only a single individual, it was considered homogeneous and included, provided that it was polymorphic across all the accessions under consideration.

3. Results

3.1. DNA quality and quantity

All 328 samples extracted from fresh leaves provided good quality, high molecular weight DNA. However, the 67 DNA samples extracted from seeds showed some degradation. The average concentration of DNA from leaf samples was 0.574 µg/µL, ranging from 0.066 µg/µL to 1.342 µg/µL. DNA extracted from seeds contained more contaminants and lower concentrations ranging from 0.096 µg/µL to 0.689 µg/µL. The mean $A_{260/280}$ for DNA obtained from fresh leaves was 1.9 ranging from 1.7 to 2.1 while that of DNA obtained from seeds was 1.6 ranging from 1.6 to 2.0.

3.2. PCR amplification

Amplification was successful for 37 of the 48 SSRs tested. To ensure amplification for the remaining 11 SSRs, PCR conditions were optimized by testing various annealing temperatures between 48.8°C and 61.1°C. This was successful for a further eight SSRs, with the annealing temperatures shown in Table 2. Of these markers, three (CCttc001, Cccta003 and CCttc007) still did not amplify and were excluded from this study. Another three markers (CCttc006, CCttc012 and CCttc020) still amplified in fewer than 50% of the samples. Further optimization, through reduction of the fluorescent dye for one of these markers (CCttc006) resulted in amplification. However, the fluorescent signals from these amplification products were too low to be detected during capillary electrophoresis to allow allele scoring on GeneMapper® software and this marker was also excluded.

3.3. Allele scoring and analysis

Allelic data obtained from GeneMapper V4.0 analysis after capillary electrophoresis, was analyzed using PowerMarker and showed that one marker, CCttc008, appeared to amplify two different loci, since each sample produced two distinct alleles of different sizes, one 251 bp and the second ranging from 251 bp to 255 bp. This marker was therefore not useful for discerning genetic diversity in this germplasm. Similarly, marker Ccat006 was highly heterozygous, *i.e.* each sample produced two different alleles and these often differed among the individuals within an accession, which complicated the interpretation of the allelic data for this marker within the germplasm. This marker was therefore also excluded from further data analysis. DNA samples from genotypes MW 2243_3, MW 2243_4 and MW 2355_7, which were extracted from seeds, did not work well as only 33% of allelic data was available and these individuals were also excluded from the subsequent analysis.

Following this data curation, the data matrix obtained was for 38 markers and 392 genotypes. This data set was analyzed by PowerMarker to produce a table of summary statistics (Table 3). Two hundred and twelve alleles were revealed with an average of

Table 2
Annealing temperatures used during gradient PCR.

Marker Name	T _m (°C)
CCttc006	49.2
CCttc012	53.0
CCttc020	54.3
CCB8	57.1
CCttc013	51.8
CCttc019	51.8
CCac006	53.0
CCttc031	53.0
CCcttc001	53.0
Cccta003	54.3
CCttc007	53.0

Table 3
PowerMarker® Summary statistics output for the selected 38 markers across 392 individual DNA samples.

Marker Names	Major Allele Frequency	No of Alleles	Heterozygosity	PIC
CCB1	0.92	8.00	0.06	0.14
CCB7	0.58	6.00	0.03	0.39
CCB8	0.94	6.00	0.02	0.12
CCB9	0.85	5.00	0.06	0.25
CCB10	0.97	2.00	0.01	0.05
CCttc008	0.96	2.00	0.00	0.07
CCac035	0.94	2.00	0.00	0.11
CCac036	0.63	7.00	0.24	0.40
CCttc033 (CCttc020)	0.94	3.00	0.03	0.10
CCac021	0.99	2.00	0.02	0.03
CCtc007	0.91	5.00	0.00	0.17
CCttc005 (CCttc004)	0.50	8.00	0.93	0.55
CCttc003 (CCttc002)	0.86	5.00	0.03	0.23
CCtta001	0.91	4.00	0.01	0.16
CCtta007	0.90	6.00	0.03	0.17
CCggt001	0.98	2.00	0.00	0.04
CCtc002	0.97	3.00	0.04	0.06
PKS30	0.84	3.00	0.00	0.23
CCac003	0.39	7.00	0.67	0.65
CCttc008	0.91	9.00	0.00	0.17
CCtta011 (CCtta006)	0.42	8.00	0.11	0.67
CCac006	0.13	11.00	0.03	0.89
CCggt002	0.47	3.00	0.31	0.51
CCggt003	0.53	7.00	0.28	0.51
CCtc013 (CCtc007)	0.93	3.00	0.00	0.13
CCac012 (CCac010)	0.90	3.00	0.05	0.18
CCac013 (CCac011)	0.57	7.00	0.67	0.60
Ccat011 (Ccat006)	0.45	14.00	0.03	0.73
CCtc009	0.72	4.00	0.14	0.41
CCac018	0.92	3.00	0.00	0.15
CCac026	0.86	7.00	0.01	0.23
CCttc018	0.82	10.00	0.05	0.30
CCttc019	0.59	14.00	0.82	0.58
CCac029	0.80	3.00	0.00	0.31
CCB4	0.47	7.00	0.07	0.56
CCcct004	0.94	7.00	0.01	0.11
ICPM1E04	0.73	4.00	0.05	0.40
ICPM1G04	0.93	2.00	0.01	0.13
Mean	0.76	5.58	0.13	0.30

5.58 alleles per marker and a maximum number of 14 alleles produced by marker CCttc019. PIC, an indicator of how well a marker is able to distinguish the samples tested due to the diversity of alleles detected across the samples, ranged from 0.03 to 0.89 with an average of 0.30.

DARwin software was used to produce a dissimilarity matrix which was displayed in a neighbor-joining tree, presented in Fig. 1, and also used for principle component analysis (PCA), presented in Fig. 2. Four major clusters were evident. Clusters I and IV comprised some of the gene bank materials as well as all five released varieties. All of the gene bank materials grouped together in Clusters II, III and IV. The landraces also grouped in Cluster III, apart from ICP 13076, which was in Cluster I.

Screening of the allelic data associated with the selected released and improved varieties (ICEAP 00040, ICEAP 00020, KAT60/8, ICEAP 00068 and ICEAP 00557 and Mtawanjuni a popular traditional cultivar) for which a DNA fingerprint was to be developed, revealed that 6 markers – CCB1, CCB7, Ccac035, CCttc003, Ccac026 and CCttc019 – could unambiguously discern these six varieties from one another. These markers were also highly homozygous and the amplified fragments were easy to score. The fingerprint developed with the 6 markers listed above, are presented in Table 4.

4. Discussion

4.1. DNA extraction and amplification

High quality DNA was obtained in this study, even without using the prescribed phenol: chloroform extraction step described by Mace et al.

[29]. The total average amount of DNA obtained from leaves was 55 µg, which is higher than the 7.5 µg reported for pigeonpea by Mace et al. [29]. The mean $A_{260/280}$ of DNA extracted from fresh leaves was 1.9 while that of DNA extracted from seeds was 1.6. This made the extraction both safer and cheaper by eliminating the use of phenol, which is hazardous and expensive to dispose of [35]. DNA extracted from the seeds was degraded and of lower quality than that obtained from leaf material, likely due to the polysaccharides and polyphenols present in pigeonpea seeds [1]. These compounds co-precipitated with the DNA after the addition of isopropanol/ethanol: sodium acetate [30] and inhibited *Taq* DNA polymerase activity in the subsequent SSR genotyping [36], which explained the recalcitrance to PCR amplification of the 3 DNA samples that were obtained from seeds.

PCR optimization is an important step to ensure the successful amplification of the target DNA fragment. All aspects of a PCR protocol can be considered in optimization [37,38]. However, this study focused only on the annealing temperature and primer concentration. Amplification for 37 of the 48 primer pairs was successful using a fixed annealing temperature of 59°C, the standard protocol that worked well in our hands. Eight of the remaining 11 primer pairs successfully amplified the target SSR loci when the annealing temperature was adjusted, as indicated in Table 2. For marker CCttc006, CCttc012, and CCtc020, it was necessary to increase the amount of forward primer and reduce the fluorescently labeled M13 tag concentrations in the PCR reaction mixture. However, with the reduced fluorescent label, the resultant fragments did not incorporate enough fluorescence to be detected by the laser during capillary electrophoresis. This has been experienced before in other studies that used labeled M13 sequences according to the method described by Schuelke [31]. All in all, 45/48 or 94% of the markers tested did amplify by PCR and this was considered sufficient for this study. However, not all markers amplified equally well and another 7 (CCttc031, CCac009, CCttc001, Cccta003, CCac036, CCtta015, CCttc007) had to be excluded from analysis. Although this represented a substantial amount of data that was excluded from the analysis, the final number of 38 good markers compared well with other published studies on genetic diversity analysis where 30 to 40 SSR markers were typically considered adequate e.g. in pigeonpea [19], sorghum [39], groundnut [40], wheat [41] and rice [42].

4.2. Genetic diversity

Allelic data analysis showed an average of 5.58 alleles per marker. This was higher than other pigeonpea diversity studies published to date, which used similar markers on cultivated varieties [15,16]. Diversity in cultivated pigeonpea is generally reported to be low [1, 43]. This was observed even when other types of markers were used e.g. diversity arrays technology (DArTs) and amplified fragment length polymorphisms (AFLPs) [44,45]. Consequently, studies that included wild species reported higher PIC and allele number averages [17]. Despite the relatively low polymorphism, the markers used in this study grouped the genotypes clearly into four major groups. After ten thousand iterations the highest bootstrap value was observed in Cluster I. Other clusters showed lower confidence levels and these could be due to low polymorphism/genome coverage of the SSRs used [46].

Most of the released varieties (ICEAP 00040, ICEAP 00020, KAT60/8, ICEAP 00068 and ICEAP 00557) were developed from Kenyan and Tanzanian varieties and subsequently introduced to Malawi [47,48]. ICEAP 00068 and ICEAP 00557 are released varieties originating from Tanzania, which grouped in different clusters (I and IV respectively). Released varieties that were developed in Kenya (ICEAP 00040, ICEAP 00020, and KAT60/8) all grouped together in Cluster I except for ICEAP 00040, which was in Cluster IV. All these released varieties were selected and improved for traits such as disease resistance, high

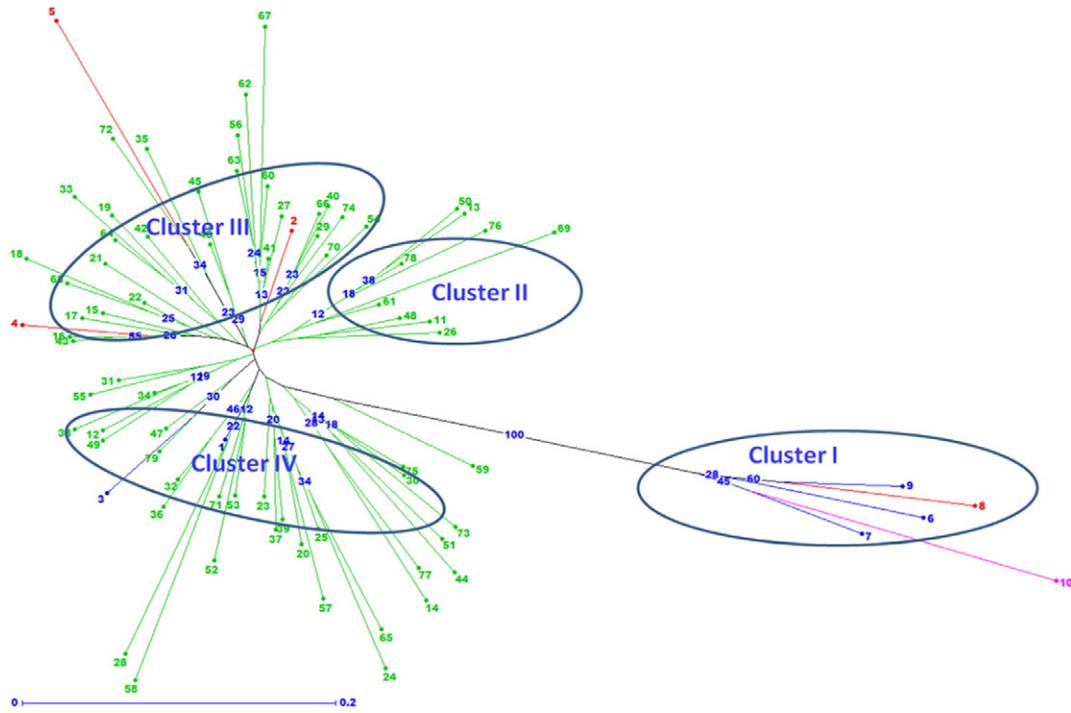


Fig. 1. Unweighted neighbor-joining dendrogram of Malawi pigeonpea varieties allelic data showing the four major clusters with two sub-groups that appeared in Cluster I (A and B) with a bootstrap value of 10,000. Color key: Genebank materials (green), Released Varieties (blue), Landraces (red) and Reference variety (violet).

yields or drought tolerance and have different maturity durations [49]. ICEAP 00040 and ICEAP 00020 are medium and long duration maturity genotypes, respectively, which are resistant to *Fusarium* wilt while ICEAP 00068, of medium duration, is susceptible to wilt but is popular with farmers as it yields large grains [48]. ICPV 9145 and ICP 13076 were ICRISAT-India accessions collected from Kenya, although they grouped in different clusters (III and I, respectively). Both

genotypes and ICPV 87105 have moderate resistance to *Fusarium* wilt [50]. The obvious genetic differences observed between ICPV 9145 and ICP 13076 in this study could indicate possible different sources or mechanisms of *Fusarium* wilt resistance inherent in these two varieties. This should be further investigated in association mapping studies to confirm if this is the case so that this diversity can be exploited in future in breeding programs.

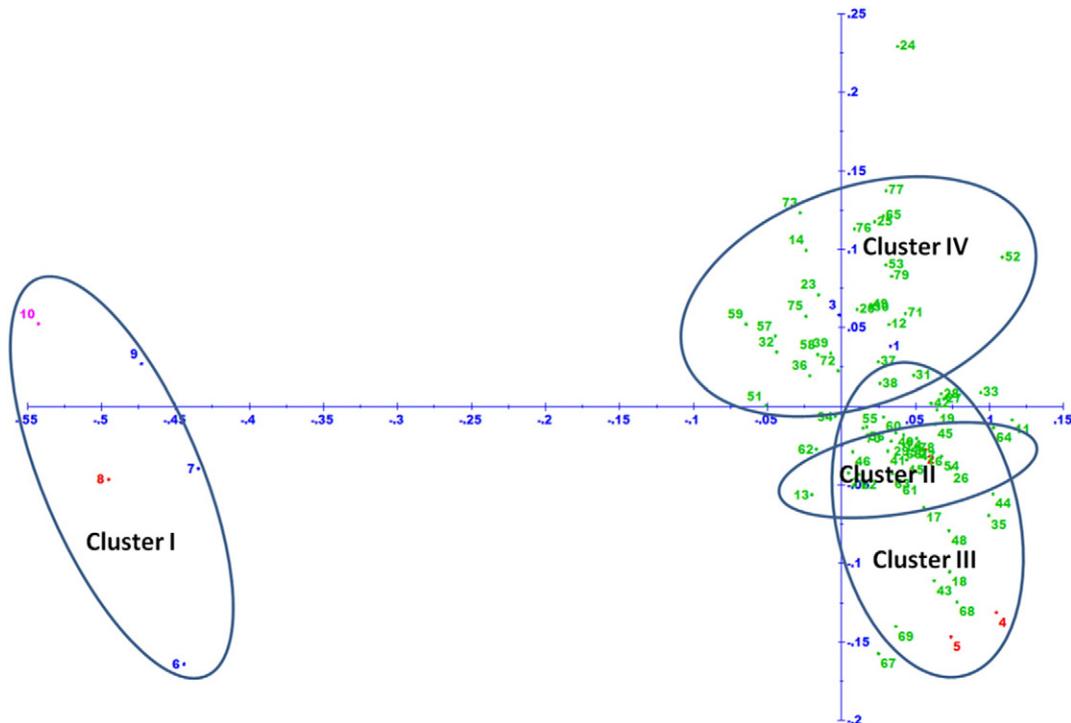


Fig. 2. Graphical display of the principle coordinate analysis on Malawi pigeonpea genotypes on axes 1/2 with a bootstrap value of 10,000. Color key: Genebank materials (green), Released varieties (blue), Landraces (red), Reference variety (violet).

Table 4

Allelic data from 6 selected SSR markers that provides a genetic fingerprint for the five released and improved pigeonpea varieties cultivated in Malawi.

Marker name	Variety name					
	ICEAP 00040	ICP9145	ICEAP 00557	ICPV 87105	Mtawanjuni	KAT60/8
CCB1	222	222	222	222	220	214/222
CCB7	172	174	172	174	174	172/174
CCac035	267	267	267	265	265	267
CCttc003	196	196	196	193	193	190
CCac026	268	268	268	266	266	266/268
CCttc019	222	219	225/234	219	219/225/228	196

Although individuals of the same genotype grouped together for the most part, some were spread out among different clusters, such as ICP 9145 and ICEAP 00040. This was probably due to contamination or mixture of the seeds. Two landraces, Mtawanjuni and ICP 9145 grouped with gene bank materials. Mtawanjuni is a popular traditional cultivar in Malawi. It is a high yielding medium duration variety, which farmers prefer due to its relatively good insect resistance. ICP 9145 is a Kenyan landrace and one of the first varieties to be introduced to Malawi in 1987. It is high yielding and has resistance to *Fusarium* wilt [24].

From the neighbor-joining tree (Fig. 1) the most commonly cultivated varieties in Malawi, which include the four released varieties and four landraces from the region, were spread across three of the clusters observed, indicating that they generally represented the genetic diversity available in Malawi. However, cluster III and cluster IV showed only two released varieties each and none in cluster II; thus there is substantial variation that can still be exploited through further breeding. The markers used in this study were not known to be linked to any traits of interest and this should be the next step in pigeonpea genomics to allow visualization of which varieties harbor important traits such as the differing maturity duration, number and duration of flowering times during a season, high yields, large, cream colored seeds, insect resistance (especially to pod borers) and *Fusarium* wilt resistance [22]. Markers linked to these traits will allow scientists to determine sources and mechanisms controlling these traits. In addition, germplasm containing these traits can be identified and the traits transferred to the best yielding and most popular varieties [10, 51]. Markers linked to these traits will also allow pyramiding the traits into a select few varieties. The recent sequencing of the pigeonpea genome is a major step in this direction [52].

Natural outcrossing, due to insect pollination, is high in pigeonpea and is difficult and expensive to control in the field since plants have to be isolated under insect-proof nets if outcrossing is to be avoided [26]. In Malawi, this causes contamination of seeds in farmers' fields since many farmers plant more than one variety on their farms or have neighbors who plant different varieties whose flowering times overlap. For example, after obtaining pure Mtawanjuni seeds used in this study from breeders, other seeds of this variety were obtained randomly from different Malawi farmers. The seeds obtained from the farmers had five different seed coat colors and none was similar to seeds obtained from breeders. Such contamination can cause yield losses due to loss or dilution of insect or *Fusarium* wilt resistance and often closes market opportunities when mixtures give rise to different seed colors or seed size [28].

4.3. Genetic fingerprint

To our knowledge, there is no available software that can screen allelic data and identify markers suited for a DNA fingerprint. Therefore, this study attempted a logical approach to identify markers that will provide such a fingerprint and the criteria were developed. The six markers identified for the DNA fingerprint, generally had low heterozygosity and intermediate to high PIC scores according to the

PowerMarker results of the entire dataset (Table 3). Since the resulting number of markers and genotypes were both small, the fingerprint could be determined visually and is presented in Table 4. In all cases, at least four out of the five individuals always presented the same alleles, except for individual ICEAP 00557/3 and Marker CCac026 where missing data reduced this number to 3/5. CCttc019 was a heterozygous marker, which presented a monomorphic allele of 196 bp for all individuals across all the released varieties. This allele was excluded for the fingerprint and only the second, polymorphic alleles from all varieties were included. When the combination of alleles for each variety across the six markers are considered, this preliminary DNA fingerprint for pigeonpea can discern each variety with confidence. In a similar way, advantage of SSR marker assays was evidenced in pigeonpea hybrid breeding through ensuring the genetic purity of hybrids and their parents [53,54]. However, this fingerprint needs to be further tested for robustness, repeatability and ability to discern admixtures due to cross pollination.

5. Conclusion

This study set out to investigate the level of genetic diversity in all cultivated Malawi pigeonpea varieties with SSR markers. While this was successful, it was observed that the level of diversity is low and further studies should exploit more new SSR markers, such as those identified from resequenced pigeonpea genomes. It is also recommended that such studies include wild pigeonpea genotypes as they could reveal a new genetic resource. It was however noted that the released varieties are representative of the genetic base available in Malawi pigeonpea.

With a small number of markers it was possible to create a genetic fingerprint of six important pigeonpea varieties in Malawi. Although this needs to be tested further, it indicates the potential of using SSR markers to discern pigeonpea varieties. Moreover, use of more polymorphic markers will increase the number of genotypes in the fingerprint. This can be used to detect seed contamination, which is a major cause of low yields and ensure availability of high quality seeds for Malawi farmers.

Adequate high quality DNA was obtained from leaves despite omitting the phenol: chloroform extraction step. This and the advent of new methods that eliminate use of hazardous substances during DNA extraction show clearly that DNA extraction is becoming safer and cheaper.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ejbt.2016.02.004>.

Conflict of interest

The authors declare that they have no conflict of interest.

Financial support

This project was funded by Irish Aid under the "Malawi Seed Industry Development Project" (2010/2011).

Acknowledgments

The authors acknowledge technical assistance from Wilson Kimani and Geoffrey Meru.

References

- [1] Saxena KB, Kumar RV, Rao PV. Pigeonpea nutrition and its improvement. *J Crop Prod* 2002;5:227–60. http://dx.doi.org/10.1300/J144v05n01_10.
- [2] Saxena KB, Kumar RV, Sultana R. Quality nutrition through pigeonpea—A review. *Health* 2010;02:1335–44. <http://dx.doi.org/10.4236/health.2010.211199>.

- [3] Rao MR, Mathuva MN. Legumes for improving maize yields and income in semi-arid Kenya. *Agric Ecosyst Environ* 2000;78:123–37. [http://dx.doi.org/10.1016/S0167-8809\(99\)00125-5](http://dx.doi.org/10.1016/S0167-8809(99)00125-5).
- [4] Amaefule KU, Obioha FC. Performance of pullet chicks fed raw or processed pigeon pea (*Cajanus cajan*) seed meal diets. *Livest Res Rural Dev* 2005;17.
- [5] Abdelati KA, Mohammed HAR, Ahmed ME. Influence of feeding processed pigeon pea (*Cajanus cajan*) seeds on broiler chick performance. *Int J Poult Sci* 2009;8: 971–5. <http://dx.doi.org/10.3923/ijps.2009.971.975>.
- [6] Kumar Rao JV, Dart PJ, Sastry PV. Residual effect of pigeonpea (*Cajanus cajan*) on yield and nitrogen response of maize. *Exp Agric* 1983;19:131–41. <http://dx.doi.org/10.1017/S0014479700022572>.
- [7] Odeny DA. The potential of pigeonpea (*Cajanus cajan* (L.) millsp.) in Africa. *Nat Resour Forum* 2007;31:297–305. <http://dx.doi.org/10.1111/j.1477-8947.2007.00157.x>.
- [8] Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants – Sialient statistical tools and considerations. *Crop Sci* 2003;43:1235–48. <http://dx.doi.org/10.2135/cropsci2003.1235>.
- [9] Sneller CH, Nelson RL, Carter TE, Cui Z. Genetic diversity in crop improvement. *J Crop Improv* 2005;14:103–44. http://dx.doi.org/10.1300/J411v14n01_06.
- [10] Varshney RK, Graner A, Sorrells ME. Genomics-assisted breeding for crop improvement. *Trends Plant Sci* 2005;10:621–30. <http://dx.doi.org/10.1016/j.tplants.2005.10.004>.
- [11] Loridon K, McPhee K, Morin J, Dubreuil P, Pilet-Nayel ML, Aubert G, et al. Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.). *Theor Appl Genet* 2005;111:1022–31. <http://dx.doi.org/10.1007/s00122-005-0014-3>.
- [12] Jin L, Lu Y, Xiao P, Sun M, Corke H, Bao J. Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theor Appl Genet* 2010; 121:475–87. <http://dx.doi.org/10.1007/s00122-010-1324-7>.
- [13] Chakraborti M, Prasanna BM, Hossain F, Mazumdar S, Singh AM, Guleria S, et al. Identification of kernel iron- and zinc-rich maize inbreds and analysis of genetic diversity using microsatellite markers. *J Plant Biochem Biotechnol* 2011;20: 224–33. <http://dx.doi.org/10.1007/s13562-011-0050-9>.
- [14] Emon RM, Gustafson JP, Nguyen H, Musket T, Jahiruddin M, Islam MA, et al. Molecular marker-based characterization and genetic diversity of wheat genotypes in relation to Boron use efficiency. *Indian J Genet Plant Breed* 2010; 70:339–48.
- [15] Burns MJ, Edwards KJ, Newbury HJ, Ford-Lloyd BV, Baggott CD. Development of simple sequence repeat (SSR) markers for the assessment of gene flow and genetic diversity in pigeonpea (*Cajanus cajan*). *Mol Ecol Notes* 2001;1:283–5. <http://dx.doi.org/10.1046/j.1471-8278.2001.00109.x>.
- [16] Odeny DA, Jayashree B, Ferguson M, Hoisington D, Crouch J, Gebhardt C. Development, characterization and utilization of microsatellite markers in pigeonpea. *Plant Breed* 2007;126:130–6. <http://dx.doi.org/10.1111/j.1439-0523.2007.01324.x>.
- [17] Odeny DA, Jayashree B, Gebhardt C, Crouch J. New microsatellite markers for pigeonpea (*Cajanus cajan* (L.) millsp.). *BMC Res Notes* 2009;2:35. <http://dx.doi.org/10.1186/1756-0500-2-35>.
- [18] Saxena RK, Prathima C, Saxena KB, Hoisington DA, Singh NK, Varshney RK. Novel SSR markers for polymorphism detection in pigeonpea (*Cajanus spp.*). *Plant Breed* 2010; 129:142–8. <http://dx.doi.org/10.1111/j.1439-0523.2009.01680.x>.
- [19] Saxena RK, Saxena KB, Kumar RV, Hoisington DA, Varshney RK. Simple sequence repeat-based diversity in elite pigeonpea genotypes for developing mapping populations to map resistance to *Fusarium* wilt and sterility mosaic disease. *Plant Breed* 2010;129:135–41. <http://dx.doi.org/10.1111/j.1439-0523.2009.01698.x>.
- [20] Varshney RK, Penmetsa RV, Dutta S, Kulwal PL, Saxena RK, Datta S, et al. Pigeonpea genomics initiative (PGI): An international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). *Mol Breed* 2010;26:393–408. <http://dx.doi.org/10.1007/s11032-009-9327-2>.
- [21] Bohra A, Dubey A, Saxena RK, Penmetsa RV, Poornima K, Kumar N, et al. Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea (*Cajanus spp.*). *BMC Plant Biol* 2011;11:56. <http://dx.doi.org/10.1186/1471-2229-11-56>.
- [22] Bohra A, Saxena RK, Gnanesh BN, Saxena K, Byregowda M, Rathore A, et al. An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. *Theor Appl Genet* 2012;125:1325–38. <http://dx.doi.org/10.1007/s00122-012-1916-5>.
- [23] FAOSTAT. Food and Agriculture Organisation of the United Nations statistical databases. cited June 24, 2013. 2010 Available from Internet <http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor>.
- [24] Soko H. Pigeonpea Research, Technology Exchange, and Seed Production in Malawi. In: Silim S, Mergeai G, Kimani P, editors. Status potential pigeonpea East. South. Africa Proc. a Reg. Work. 12–15 Sep 2000, Nairobi, Kenya. B-5030 Gembloux, Nairobi; 2001.
- [25] Simtowe F, Shiferaw B, Kassie M, Siambi M, Silim S, Kananji G, et al. Assessment of the current situation and future outlooks for the pigeonpea sub-sector in Malawi. *ICRISAT* 2009;1–48.
- [26] Saxena KB, Singh L, Gupta MD. Variation for natural out-crossing in pigeonpea. *Euphytica* 1990;46:143–8. <http://dx.doi.org/10.1007/BF00022307>.
- [27] Simtowe F, Asfaw S, Shiferaw B, Siambi M, Monyo E, Muricho G, et al. Socioeconomic assessment of pigeonpea and groundnut production conditions-farmer technology choice, market linkages, institutions and poverty in rural Malawi. Patancheru: ICRISAT; 2010.
- [28] Jones R, Freeman HA, Lo Monaco G. Improving the access of small farmers in eastern and southern Africa to global pigeonpea markets. *Agricultural Research and Extension Network Paper No. 120*; 2002.
- [29] Mace ES, Buhariwalla KK, Buhariwalla HK, Crouch JH. A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant Mol Biol Report* 2003;21:459–60. <http://dx.doi.org/10.1007/BF02772596>.
- [30] Sharma AD, Gill PK, Singh P. DNA isolation from dry and fresh samples of polysaccharide-rich plants. *Plant Mol Biol Report* 2002;20:415a–415f. <http://dx.doi.org/10.1007/BF02772129>.
- [31] Schuelke M. An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol* 2000;18:233–4. <http://dx.doi.org/10.1038/72708>.
- [32] Koumi P, Green HE, Hartley S, Jordan D, Laheç S, Livett RJ, et al. Evaluation and validation of the ABI 3700, ABI 3100, and the MegaBACE 1000 capillary array electrophoresis instruments for use with short tandem repeat microsatellite typing in a forensic environment. *Electrophoresis* 2004;25:2227–41. <http://dx.doi.org/10.1002/elps.200305976>.
- [33] Liu K, Muse SV. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* 2005;21:2128–9. <http://dx.doi.org/10.1093/bioinformatics/bti282>.
- [34] Perrier X, Jacquemoud-Collet J. DARwin software; 2006[cited November 15, 2010]. Available from Internet <http://darwin.cirad.fr/>.
- [35] Maréchal-Drouard L, Guillemot P. A powerful but simple technique to prepare polysaccharide-free DNA quickly and without phenol extraction. *Plant Mol Biol Rep* 1995;13:26–30. <http://dx.doi.org/10.1007/BF02668390>.
- [36] Pandey RN, Adams RP, Flournoy LE. Inhibition of random amplified polymorphic DNAs (RAPDs) by plant polysaccharides. *Plant Mol Biol Rep* 1996;14:17–22. <http://dx.doi.org/10.1007/BF02671898>.
- [37] Caetano-Anollés G. DAF optimization using Taguchi methods and the effect of thermal cycling parameters on DNA amplification. *Biotechniques* 1998;25:472–80.
- [38] Roux KH. Optimization and troubleshooting in PCR. *Cold Spring Harb Protoc* 2009;4. <http://dx.doi.org/10.1101/pdb.ip66>.
- [39] Billot C, Ramu P, Bouchet S, Chantreau J, Deu M, Gardes L, et al. Massive sorghum collection genotyped with SSR markers to enhance use of global genetic resources. *PLoS One* 2013;8. <http://dx.doi.org/10.1371/journal.pone.0059714>.
- [40] Tang R, Gao G, He L, Han Z, Shan S, Zhong R, et al. Genetic diversity in cultivated groundnut based on SSR markers. *J Genet Genomics* 2007;34:449–59. [http://dx.doi.org/10.1016/S1673-8527\(07\)60049-6](http://dx.doi.org/10.1016/S1673-8527(07)60049-6).
- [41] Roussel V, Leisova L, Exbrayat F, Stehno Z, Balfourier F. SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. *Theor Appl Genet* 2005;111:162–70. <http://dx.doi.org/10.1007/s00122-005-2014-8>.
- [42] Chakravarthy BK, Narayanan R. SSR marker based DNA fingerprinting and diversity study in rice (*Oryza sativa* L.). *Afr J Biotechnol* 2006;5:684–8.
- [43] Varshney RK, Kudapa H, Roorkival M, Thudi M, Pandey MK, Saxena RK, et al. Advances in genetics and molecular breeding of three legume crops of semi-arid tropics using next-generation sequencing and high-throughput genotyping technologies. *J Biosci* 2012;37:811–20. <http://dx.doi.org/10.1007/s12038-012-9228-0>.
- [44] Yang S, Pang W, Ash G, Harper J, Carling J, Wenzl P, et al. Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology. *Theor Appl Genet* 2006;113:585–95. <http://dx.doi.org/10.1007/s00122-006-0317-z>.
- [45] Panguluri SK, Janaiah K, Govil JN, Kumar PA, Sharma PC. AFLP fingerprinting in pigeonpea (*Cajanus cajan* (L.) Millsp.) and its wild relatives. *Genet Resour Crop Evol* 2006;53:523–31. <http://dx.doi.org/10.1007/s10722-004-2031-5>.
- [46] Berry V, Gascuel O. On the interpretation of bootstrap trees: Appropriate threshold of clade selection and induced gain. *Mol Biol Evol* 1996;13:999–1011. <http://dx.doi.org/10.1093/molbev/13.7.999>.
- [47] Silim SN, Gwata ET, Mligo JK, Siambi M, Karuru O, King SB, et al. Registration of pigeonpea cultivar “ICEAP 00040”. *Crop Sci* 2005;45:2647. <http://dx.doi.org/10.2135/cropsci2004.0732>.
- [48] Gwata E, Mligo J, Silim S. Registration of pigeonpea cultivar Tumia. *Crop Sci* 2007;47: 436. <http://dx.doi.org/10.2135/cropsci2006.03.0141>.
- [49] Silim SN, Mergeai G, Kimani PM. Status and potential of pigeonpea in Eastern and Southern Africa: Proceedings of a regional workshop, 12–15 Sep 2000, Nairobi, Kenya. B-5030 Gembloux, Belgium: Gembloux Agricultural University; and Patancheru 502324. Andhra Pradesh, India: International; 2001.
- [50] Silim S, Tuwafe S, Singh L. Pigeonpea Improvement in eastern and southern Africa-Annual Research Planning Meeting 1993. 25–27 Oct 1993–1994 [Cited June 15, 2013]. Available from Internet http://oar.icrisat.org/447/1/CPE_089.pdf.
- [51] Odeny D, Gebhardt C. Towards molecular breeding of pigeonpea (*Cajanus cajan* (L.) Millsp.): A case for increased production in Africa. *Acta Hort* 2008:151–4. <http://dx.doi.org/10.17660/ActaHortic.2009.806.16>.
- [52] Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, et al. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnol* 2011;30:83–9. <http://dx.doi.org/10.1038/nbt.2022>.
- [53] Bohra A, Singh IP, Yadav AK, Pathak A, Soren KR, Chaturvedi SK, et al. Utility of informative SSR markers in the molecular characterization of cytoplasmic genetic male sterility-based hybrid and its parents in pigeonpea. *Natl Acad Sci Lett* 2015;38: 13–9. <http://dx.doi.org/10.1007/s40009-014-0288-6>.
- [54] Saxena RK, Saxena K, Varshney RK. Application of SSR markers for molecular characterization of hybrid parents and purity assessment of ICPH 2438 hybrid of pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *Mol Breed* 2010;26:371–80. <http://dx.doi.org/10.1007/s11032-010-9459-4>.