Genet Resour Crop Evol (2020) 67:191–208
https://doi.org/10.1007/s10722-019-00864-9

RESEARCH ARTICLE

Genetic diversity and population structure in Beninese pigeon pea [*Cajanus cajan* (L.) Huth] landraces collection revealed by SSR and genome wide SNP markers

Fiacre Zavinon · Hubert Adoukonou-Sagbadja · Jens Keilwagen · Heike Lehnert · Frank Ordon · Dragan Perovic

Received: 23 January 2019 / Accepted: 27 November 2019 / Published online: 13 December 2019

Abstract Pigeon pea [*Cajanus cajan* (L.) Huth] is a valuable multipurpose crop locally used for household food security as well in traditional medicine in Benin. However, due to the neglected status of the crop, its genetic resources are not well evaluated and its agronomic potential remains undetermined. For breeding purpose, a total of 77 landraces from Benin and three breeding lines were simultaneously genotyped with 30 preselected SSRs and 794 GBS derived SNPs in order to estimate the genetic diversity and infer the population structure within the collection. Both marker types were found informative in polymorphism analysis revealing all high genetic variability. The 30 SSR markers led to a total of 209 alleles with an average of 6.97 alleles per locus, whereas only biallelic SNPs were extracted from GBS data according to specific filter criteria. The polymorphism information content value was 0.57 and 0.25 for SSR and SNP, respectively. The genetic diversity calculated as expected heterozygosity was higher for SSR (0.62) than for SNP (0.35). The inference of the genetic structure subdivided the entire collection into three major groups independent of the marker type. However, the resolving power in population structure analysis was higher for SNP than for SSR. AMOVA and PCoA analyses showed clearer population structure with SNP than SSR. The present study provides a clear insight on the genetic diversity in Beninese pigeon pea and represents the first report comparing the performance of SSR and SNP markers for population genetics analysis in cultivated pigeon pea. It provides useful information for further pigeon pea conservation and breeding in Benin.

Keywords Benin · Genetic diversity · Pigeon pea · Population structure · Single nucleotide polymorphisms (SNP) · Simple sequence repeat (SSR)

F. Zavinon · H. Adoukonou-Sagbadja (✉)
Laboratory of Genetic Resources and Molecular Breeding (LaREGAME), Department of Genetic and Biotechnology, Faculty of Sciences and Techniques, University of Abomey-Calavi, BP 1947, Abomey-Calavi, Republic of Benin
e-mail: hadoukas@yahoo.fr

F. Zavinon · F. Ordon · D. Perovic
Institute for Resistance Research and Stress Tolerance, Federal Research Centre for Cultivated Plants, Julius Kühn-Institut (JKI), Erwin-Baur-Str. 27, 06484 Quedlinburg, Saxony-Anhalt, Germany

J. Keilwagen · H. Lehnert
Institute for Biosafety in Plant Biotechnology, Federal Research Centre for Cultivated Plants, Julius Kühn-Institut (JKI), Erwin-Baur-Str. 27, 06484 Quedlinburg, Saxony-Anhalt, Germany

Introduction

Pigeon pea [*Cajanus cajan* (L.) Huth] is an important food legume crop that is predominantly cultivated in tropical and subtropical regions of the world

(Varshney et al. 2010). It is a diploid plant species (2n = 2x = 22) with genome size of 853Mbp and belongs to the subfamily of Papilionoideae of the plant family of Fabaceae. It belongs to *Cajanus* genus and represents the only one cultivated of the 32 species known under its sub-tribe *Cajaninae* (Van der Maesen 1990; Odeny et al. 2007; Bohra et al. 2017). The legume is also known to be often a cross-pollinated species with an outcrossing rating up to 25–30% (Saxena et al. 1990). In the semi-arid tropics where it is grown for both subsistence and commercial use, the crop is increasingly playing a critical role in the daily life of many farmers (Odeny et al. 2016). Pigeon pea is mainly cultivated for its grains rich in protein (about 20–22%) with appreciable amount of essential amino acid and minerals (Kumar et al. 2016; Saxena et al. 2015). It is also useful in restoring soil fertility by fixing atmospheric nitrogen and for its ability to solubilize fixed phosphorus (Ae et al. 1990; Sharma et al. 2016). It is adapted to widely differing climatic and soil conditions and considered as drought tolerant crop (Kumar et al. 2011). Indeed, pigeon pea can be grown under rain fed conditions with low inputs (Sharma et al. 2011) making it suitable for subsistence agriculture (Bohra et al. 2012).

In Benin, pigeon pea is a minor crop that substantially contributes to the food security of the rural communities. It is mainly grown by farmers for home consumption and commercial uses. According to Ayenan et al. (2017a) pigeon pea is used for making up the shortage of cowpea, maize and other staple foods during the lean season. The dry stems and leaves of pigeon pea are also used by the local population in Benin for several other multipurpose. Although pigeon pea is a source of income generation and plays an important role in food security for the rural households, its cultivation is unfortunately neglected in many countries of West Africa (Zavinon et al. 2018) and in Benin (Dansi et al. 2012). The national production of Benin averages yearly about 2800 tons (Ayenan et al. 2017b) and ensured by a minority of small farmers on relatively reduced areas lower than 0.5 ha (Zavinon et al. 2018). A number of factors are responsible for the poor pigeon pea production in Benin, including the neglected status of the species and the lack of improved cultivars (Dansi et al. 2012; Ayenan et al. 2017a, b). Hence, it is highly imperative to breed new cultivars with enhanced agronomic performances for increasing pigeon pea production in Benin.

Genomics-assisted breeding (GAB) provides strategies to develop superior varieties by using properly genome information in combination with the phenotypic data in conventional breeding (Varshney et al. 2012; Saxena et al. 2017). This implies a better understanding of the extent and distribution of genetic diversity within and between different genotype groups. According to Chen et al. (2017), studying the diversity is essential not only to assist plant breeders in parents' selection but also to provide a more rational basis for expanding the gene pool and for identifying materials that harbour alleles valuable for plant improvement. In pigeon pea, due to the recent availability of genomic resources, several molecular markers have been developed and employed mainly in germplasm characterization. Among the DNA markers developed for pigeon pea, SSR and SNP are seen as the most robust molecular markers widely used for plant genetic studies due to their abundance in genome and co-dominance inheritance. For instance, these two molecular marker types were extensively employed to study genetic diversity (Njung'e et al. 2016; Bohra et al. 2017; Kassa et al. 2012; Saxena et al. 2014), association mapping (Mir et al. 2014; Varshney et al. 2017) and QTL analysis (Kumawat et al. 2012; Saxena et al. 2017) in pigeon pea.

However, despite the recent availability of genomic resources in pigeon pea, this important genetic resource has not been adequately characterized because of the neglected status of the crop (Dansi et al. 2012). Indeed, the pigeon pea landraces' diversity used by farmers has been only recently documented (Ayenan et al. 2017a; Zavinon et al. 2018) and broadly evaluated for its phenotypic diversity using agro-morphological traits (Zavinon et al. 2019). Although these studies are important, they have limited advantages to pigeon pea breeding program (Adoukonou-Sagbadja et al. 2007). In the future, intensive genetic and genomic research must be used to mobilize various useful alleles, which could be channeled into breeding programs, to increase the productivity of pigeon peas and also to address major limitations, especially in the light of recent climatic changes. In the present investigation, an effort has been made to study the genetic diversity in a representative pigeon pea germplasm collection grown in Benin and to infer its population structure using both SSR and genome wide SNP markers. The study, additionally, aimed to compare the performance of these two marker types in population genetic studies in pigeon pea.

Material and methods

Plant material

A total of eighty (80) pigeon pea accessions including three breeding lines and seventy-seven (77) landraces' representative of pigeon pea cultivars in use in Beninese agriculture were used in this study. The landraces were sampled in four administrative departments (Collines, Couffo, Plateau and Zou) covering three different agro-ecological zones, especially the major tribal areas of pigeon pea cultivation in Benin. They differed from each other for many agronomic characteristics including maturity cycle and seed color. The three breeding lines known as TC8126 (T1), TC4 (T2) and ICPL87 (T3) were obtained from the International Institute of Tropical Agriculture (IITA/ Ibadan). The geographic origin and maturity cycle of all analyzed pigeon pea accessions is listed in Table 1.

DNA extraction and quantification

All pigeon pea accessions were grown under greenhouse conditions at the Julius Kuehn-Institute (JKI), Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance in Quedlinburg/ Germany. Five seeds of each accession were sown in 96-well pot plates. Three weeks after germination, leaves of the five seedlings of each accession were harvested and bulked (cf. Mohamed et al. 2019) for DNA extraction. Total genomic DNA was extracted from 300 mg fresh bulked leaves of each genotype following CTAB (Cetyl Trimethyl Ammonium Bromide) extraction protocol described in detail in Stein et al. (2001) and Adoukonou-Sagbadja et al. (2007). The quality of the extracted genomic DNA was checked by electrophoresis on 1% agarose gel and later quantified by using the NanoDrop ND-100 spectrophotometer (PeQLab, Erlangen, Germany). DNA samples showing absorbance ratio above 1.8 at 260/280 nm were used and diluted to a final concentration of 20 ng/µl working solution.

Genotyping by sequencing for SNP genotyping

Library construction and sequencing

A major step in the GBS approach is genomic library construction. In this analysis, the two-enzyme GBS method described by Poland et al. (2012) was used to digest genomic DNA. Indeed, all genomic DNA samples were normalized to a final concentration of 20 ng/µl first and genome complexity was reduced by digesting individual sample genomic DNA with *Pst*I and *Msp*I. The resulting fragments were later ligated to a pair of enzyme-specific adapters in the same tube/plate without purification as reported in Wendler et al. (2014). The ligated products were purified with Magna beads (Thermo Scientific, Inc, Waltham MA) and eluted in 20 µl in Elution Buffer (EB) (10 mM Tris–HCl pH 8.0). To increase the amount of fragment, 8 µl of eluate DNA fragment were amplified with primer that were complementary to the adapters. The PCR amplification was also conducted following Wendler et al. (2014). The amplicons corresponding to the target fragments were purified using solid phase reversible immobilization (SPRI) and pooled in equimolar ratios. Besides, 500 ng of pooled DNA was size fractionated by electrophoresis using 2% agarose gel. The average DNA fragment length for the final library was determined using the Agilent 2100 Bio-analyzer (Agilent Technology, Santa Clara). Working library was diluted according to Qubit quantification and quantitative real-time PCR was used to quantify the GBS library. The concentration was determined based on standard curve and the average size of the GBS library. According to the manufacturer's protocol, GBS library was sequenced on a single lane of Illumina MiSeq instrument.

Sequencing data analysis and SNP genotyping

The sequencing produced millions of reads. They were de-multiplexed according to the barcodes and the adapters/ barcodes using the CASAVA pipeline 1.8 (Illumina, Inc.). Trim Galore software from Babraham Bioinformatics (2012) was utilized for adapter and quality trimming of the amplified genomic sequences. After this first filtering, the trimmed sequences were then aligned to draft genome sequence of pigeon pea (Varshney et al. 2012) using the SOAP alignment program as described in Li et al. (2008). Aligned sequencing reads were used for SNP detection after quality check (Q score > 20). Multi-allelic SNPs, SNPs with minor allele frequency (MAF) $< 5\%$, missing values $\geq 5\%$ or heterozygosity $\geq 90\%$ were further excluded and a high-quality SNP genotyping dataset was compiled.

Table 1 List of genotypes used in the study and their geographic origin and maturity cycle

Sample ID	Type of accessions	Department of origin	Maturity cycle	Sample ID	Type of varieties	Department of origin	Maturity cycle
BCA01	Landraces	Couffo	Late	BCT01	Landraces	Couffo	Late
BCA02	Landraces	Couffo	Intermediate	BCT03	Landraces	Couffo	Late
BCA03	Landraces	Couffo	Intermediate	BCT04	Landraces	Couffo	Late
BCA04	Landraces	Couffo	Late	BCT05	Landraces	Couffo	Late
BCA05	Landraces	Couffo	Intermediate	BCT06	Landraces	Couffo	Late
BCA06	Landraces	Couffo	Late	BCT07	Landraces	Couffo	Late
BCA07	Landraces	Couffo	Late	BCT08	Landraces	Couffo	Late
BCA08	Landraces	Couffo	Late	BPK01	Landraces	Plateau	Late
BCA09	Landraces	Couffo	Late	BPK02	Landraces	Plateau	Late
BCK01	Landraces	Couffo	Late	BPK03	Landraces	Plateau	Late
BCK02	Landraces	Couffo	Late	BPK04	Landraces	Plateau	Late
BCK03	Landraces	Couffo	Late	BPK05	Landraces	Plateau	Late
BCK04	Landraces	Couffo	Late	BPK06	Landraces	Plateau	Late
BCK05	Landraces	Couffo	Late	BPK07	Landraces	Plateau	Late
BCK06	Landraces	Couffo	Late	BPP01	Landraces	Plateau	Late
BCK07	Landraces	Couffo	Late	BPP02	Landraces	Plateau	Late
BCK08	Landraces	Couffo	Late	BPP03	Landraces	Plateau	Late
BCoG01	Landraces	Collines	Late	BPP04	Landraces	Plateau	Late
BCoG02	Landraces	Collines	Late	BPP05	Landraces	Plateau	Late
BCoG03	Landraces	Collines	Late	BPP06	Landraces	Plateau	Late
BCoG04	Landraces	Collines	Late	BPP07	Landraces	Plateau	Intermediate
BCoG05	Landraces	Collines	Late	BPP08	Landraces	Plateau	Late
BCoG06	Landraces	Collines	Late	BPP09	Landraces	Plateau	Late
BCoG07	Landraces	Collines	Late	BPP10	Landraces	Plateau	Late
BCoO01	Landraces	Collines	Late	BPP11	Landraces	Plateau	Late
BCoO02	Landraces	Collines	Late	BPP12	Landraces	Plateau	Late
BCoO04	Landraces	Collines	Late	BPP13	Landraces	Plateau	Late
BCoO05	Landraces	Collines	Late	BPP14	Landraces	Plateau	Late
BCoO06	Landraces	Collines	Late	BZD02	Landraces	Zou	Late
BCoO07	Landraces	Collines	Late	BZD03	Landraces	Zou	Late
BCoO08	Landraces	Collines	Late	BZD04	Landraces	Zou	Late
BCoO09	Landraces	Collines	Late	BZD05	Landraces	Zou	Late
BCoO11	Landraces	Collines	Late	BZD07	Landraces	Zou	Late
BCoO12	Landraces	Collines	Late	BZD08	Landraces	Zou	Late
BCoS01	Landraces	Collines	Late	BZD09	Landraces	Zou	Late
BCoS02	Landraces	Collines	Late	BZD10	Landraces	Zou	Late
BCoS03	Landraces	Collines	Late	BZD11	Landraces	Zou	Late
BCoS05	Landraces	Collines	Late	T1	Advanced	IITA/Ibadan	Short
BCoS06	Landraces	Collines	Late	T2	Advanced	IITA/Ibadan	Short
BCoS08	Landraces	Collines	Late	T3	Advanced	IITA/Ibadan	Short

SSR genotyping

Thirty (30) SSR markers distributed throughout the *Cajanus* genome were used to genotype the 80 pigeon pea accessions. They were developed and used to construct a consensus genetic map in pigeon pea by Bohra et al. (2011). The list and detailed characteristics of these microsatellites' markers are presented in the Table 2.

PCR reactions were carried out in a total volume of 10 µl containing 1 µl of 10 × buffer, 1 µl of 25 mM $MgCl_2$, 0.2 µl of each 10 mMdNTPs, 0.25 µl of 1.0 pmol/µl forward primer, 0.25 µl of 10.0 pmol/µl reverse primer, 0.08 µl 5U Hot FIREPol®DNA polymerase (Solis BioDyne, Tartu, Estonia), 6.12 µl HPLC gradient grade water (Carl Roth, Karlsruhe, Germany) and 1 µl template DNA. For SSR amplification, M13 tailed forward primers were used according to Perovic et al. (2013), so that 0.1 µl of 'M13′ primer (10.0 pmol/µl) labeled with 5′ fluorescent dye was added to the reaction mix. DNA amplification was performed in a Gene Amp® PCR System 9700 (Applied Biosystems, Darmstadt, Germany) with initial denaturation of 94 °C for 5 min followed by a touchdown PCR with 12 cycles of 30 s at 94 °C, 30 s at 62 °C, 30 s at 72 °C; and then 35 cycles with 30 s at 94 °C, 30 s at 56 °C, 30 s at 72 °C, and a final extension at 72 °C for 10 min.

All PCR products (SSR fragments) were separated by capillary electrophoresis using the genetic analyzer ABI PRISM® 3100 (Applied Biosystems, Darmstadt, Germany). A single PCR product or a set of two PCR products were pooled with the HiDi-Rox-Mastermix and denatured at 94 °C for 5 min. Fragment analysis was performed with GeneMapper software to measure the molecular size of each SSR allele and allelic data for each marker was scored manually.

Statistical analysis

Genetic diversity analysis

Genetic parameters such as number of alleles (N_A), allele frequency, observed heterozygosity (H_O), expected heterozygosity (H_E) and the polymorphism information content (*PIC*) were calculated for both marker types. The software PowerMarker vs. 3.25 (Liu and Muse 2005) was used for the SNP data to calculate genetic parameters while GenAlEx program vs. 6.5 (Peakall and Smouse 2012) was used to generate the same genetic parameters for SSR data. The PIC values for each marker type were calculated using the following equation (Botstein et al. 1980):

$$PIC = 1 - \sum_{i=1}^{n} Pi^2 - \sum_{i}^{n-1} \sum_{j=i+1}^{n} 2Pi^2 Pj^2$$

where *Pi* and *Pj* are the frequencies of i^{th} and j^{th} alleles for the selected marker, respectively.

Analysis of population structure

The model-based approach implemented in the software package STRUCTURE (Pritchard et al. 2000) was used to infer population genetic structure. SSR and SNP data were analyzed separately and merged. For each analysis, the numbers of subpopulations (K) ranging from 1 to 10 and the admixture model with correlated allele frequency were adopted to estimate the ancestry fractions of each cluster attributed to each accession. Besides, for each K, 10 replications were tested. Each run was implemented with a burn-in period of 100,000 steps followed by 100,000 Monte Carlo Markov Chain replicates. The delta-K method of Evanno et al. (2005) available in the structure Harvester program (Earl and Vonholdt 2012) was used to determine the most probable K-value. Accessions with membership probabilities ≥ 0.70 were considered to belong to the same group.

Genetic distances between pairs of accessions were calculated based on each dataset to investigate the genetic relationship between accessions. The dissimilarity matrices generated from each dataset were used to construct phylogenetic trees using Adegenet package from R software (Version 3.3.2). Correlations between genetic distance matrices were assessed using the Mantel test as implemented in GenAlEx vs. 6.5 (Peakall and Smouse 2012). Finally, based on the generated distance matrices, Principal Coordinate Analysis (PCoA) and Analysis of Molecular Variance (AMOVA) were completed to investigate genetic differentiation among pigeon pea cultivar populations using GenAlEx software package.

At each level of above analyses, the genetic parameters generated for the two types of markers (SSR and SNP) were compared to infer their appropriateness and efficiency in diversity and population structure analyses in pigeon pea germplasm.

Table 2 List and characteristics of the 30 SSR markers used for the genotyping of the 80 pigeon pea accessions from Benin

Maker name	SSR motif	Forward Primer (5′–3′)	Reverse Primer (5′–3′)	Allele number (N_A)	Chromosome
CcM0121	(TA)17	AGAAATTGGAGGCTTGGTCA	GGTATAAGGCTCAAACCCGA	9	LG01
CcM0195	(AT)11	CAACAATAAAGCATAAACCACCA	TGACGTAGATTGGGTAGTTAGGA	10	LG06
CcM0248	(TA)8	CAAACTCAACCCTACCAATGC	CATTCCTTGTCATCAATGAAGTTT	90	LG10
CcM0374	(TA)11	GAACCGTCTTAAAATTTCTCATTT	CAATGGCACATTGTCAAAAA	3	LG06
CcM0443	(TA)17n(AT)5	TGACAAAATAATGCGGTCACA	CAAGCCAAAGTTTGTTTGAACT	6	LG01
CcM0444	(TA)7	TGTCATGAGTGGCTGATCCT	TCAACCAAAATCCAAACCAA	7	LG02
CcM0484	(T)12n(ATT)5n(AT)5	TGGAAATTAAACACCATGAAACA	TGCATGCTACCAAGGAATTG	5	LG07
CcM0494	(AT)21	ACGTGAAAAATCCGCAACTT	GCTTGTGTTTCAAAATCCAACTT	8	LG02
CcM0594	(GA)9n(TC)9	GGCTTGGTTCTTTCTTGGTG	AAGTCCCTGACTTTCCCCAT	4	LG08
CcM0673	(AT)6(AG)9	TGACCACCAACCATTACCAA	CATGCACCAGACCAGAATCA	8	LG10
CcM0721	(AT)19	ATCCAACCACGTGTTTCACA	TTTGAAATGGTATCGATGATTAAA	10	LG05
CcM0785	(AT)9	GCATGTGTTTTTACTTGAGTCGTC	TGGAGGCGATCTCTTTCTTG	8	LG09
CcM0834	(AT)10	GTCCGGCTTGCCTATAAGGT	AAGGCAACCTCCCCAGTATT	4	LG11
CcM0956	(AT)16	AGCCCCAACTCAATTATCAAA	TTCCTTGCGGTTTGAGCTAT	11	LG08
CcM0974	(AT)13	CGTCTTACAGACGATCTGCATC	CAAAGAAACAGACATGATAAAGAGAGA	6	LG03
CcM1045	(AT)6	AACCTTAGTTGGTGATAGATTTCAGA	ACCGTCAAGTCCCAAATCAC	4	LG10
CcM1251	(CCA)9	CAAATGGCAGAACAGAGCAG	CGGAGATTGCATTGTTCCTT	7	LG02
CcM1357	(AT)15n(ATA)5	TCTAGCATCTCCATTAAACCATTT	ACACATATGACATTTAGCAAATAAAAA	5	LG08
CcM1982	(TC)17	TATCAAACCTGGCGATCACA	ATTCCGCAAACACATCACAA	10	LG09
CcM2004	(CT)7n(AG)12	AGGAATGCGACATTTTGGAG	TCCCCATCCCTTTCTTTCTT	3	LG08
CcM2044	(TAT)9	ATCACTCCAAGCACCCAAAC	TGCAAATGGAAGGGAATAGC	10	LG01
CcM2049	(TAT)9	GCGACCAGGTACTTTCAAGC	CGAAAAGCGATTTCAGAATTT	6	LG07
CcM2097	(CT)12	TGATAGGAATATTTCGGCGG	CCTTTGAAATTGAAGGCGAG	6	LG02
CcM2379	(TC)10	CCGGAAAAATTGCCTATTGA	TTCGATGACAGAATTTAGGTGC	8	LG03
CcM2394	(TC)12	TGGAAACGATTTCCTACCACA	ACAAGGGGAAAAGGGAAAGA	3	LG10
CcM2409	(TTA)6	TGAAGGTTGATCCAAGGAGG	CGTGCAAAATAATTGTCCAAAA	5	LG02
CcM2505	(GAA)8	CCTCGGAAGAGATTGCAGTT	TGATGAATTGGGAAGCAACA	6	LG02
CcM2697	(CT)9n(T)14	AGAGTTCGGTGACGGTTACG	GATCTGTCGAGGTTGAGGCT	2	LG11
CcM2704	(AT)10	AAAAATGTTCAATGTCGTAGTATTTGA	TGCCATATATCATGCCCTCA	11	LG09
CcM2895	(AT)24	AATGATAATTGGACACTTCTTTTTC	TGCGTTAATTAAACAAGCAAGC	15	LG3

Results

Sequence data and SNP discovery

A large number of sequencing data was generated from the 80 pigeon pea accessions on the MiSeq Illumina Instrument. The results showed that all the 80 accessions produced a total of 27.7 million of raw reads with an average of 346,030.3 raw reads per accessions. Figure 1 illustrates the distribution of the number of sequences reads in individuals DNA sample. From the 27.7 million of raw reads produced, 138,247 reads were discarded after a first filtering based on sequence quality. The number of trimmed reads obtained varied from 60,181 to 688,466 with an average of 344,302. Further, filtered sequencing reads were analyzed and as result, 1,730 high quality SNPs were identified. Among these, 936 SNPs passed the filter for bi-allelic SNPs, missing values, heterozygosity and minor allele frequency. The final markers available were composed of 794 high quality SNPs in total, among which 519 were mapped across the eleven chromosomes of pigeon pea while 275 could not be located on a specific chromosome. The distribution of the 519 mapped SNPs (Table 3) revealed that the highest number of SNPs (27.36%, 142 SNPs) was physically linked to chromosome 11. The average marker density was 45.76 kb. The lowest and highest marker densities were observed on chromosome 5 (17.83 kb) and chromosome 3 (72.96 kb), respectively. Transitions (462 allelic sites, 58.18%) were more frequent than transversions (322 allelic sites, 41.81%), with a ratio of 1.4. The C/T transitions and G/C transversions occurred at the highest and lowest frequencies, respectively. The frequencies of the two types of transitions were 31.36% and 26.82% for C/T and A/G, respectively. The frequencies of the four transversions types ranked as follows: A/T 13.22%, A/C 9.57%, G/T 10.45%, G/C 8.56% (Table 4).

Genetic diversity revealed by SNP and SSR markers

For all the 30 SSRs used, a total of 209 alleles were observed across the 80 pigeon pea accessions with an average of 6.96 alleles per locus (Table 5). The polymorphism information content (*PIC*) value ranged from 0.042 to 0.83 with an average of 0.57. The genetic diversity, calculated here as expected heterozygosity (H_e) ranged from 0.04 to 0.86 with mean value of 0.61. The observed heterozygosity (H_O) varied from 0.00 to 0.94 with a mean value of 0.41.

For SNP markers, the range and mean values of different genetic parameters were summarized in the Table 5. All the 794 detected GBS SNPs were of bi-allelic type, revealing a total of 1588 alleles. The major allele frequency (MAF) varied from 0.5 to 0.95 and averaged 0.79. The estimated mean *PIC* value for SNP was 0.25 which was lower than that calculated for the SSR. The expected heterozygosity (He) was higher for SSR than for SNP. It ranged from 0.1 to 0.5 with mean value of 0.3 for SNP. The observed heterozygosity (Ho) across the 80 pigeon pea accessions ranged from 0.0 to 0.9 and averaged 0.35 for the SNPs.

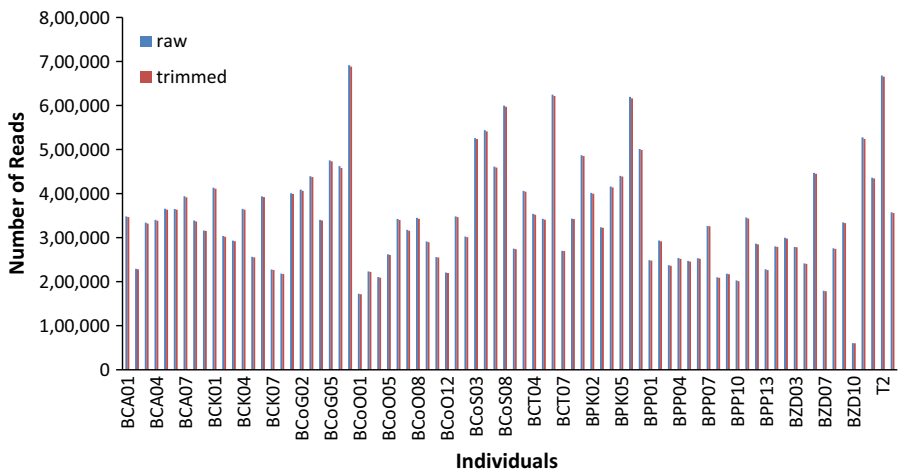

Fig. 1 Distribution of the number of sequences' reads in individuals DNA sample of the 80 *Cajanus cajan* accessions from Benin

Table 3 Genomic distribution of 519 SNPs physically mapped on 11 *C. cajan* chromosomes

Chromosomes	No. of SNPs	% SNP	Start position	End position	Length (Mb)	Density (Kb)
1	35	6.74	4,370,857	17,592,346	13.22	37.78
2	51	9.83	1,154,512	36,467,772	35.31	69.24
3	37	7.13	653,123	27,650,031	27.00	72.96
4	42	8.09	1,085,351	12,080,821	11.00	26.18
5	22	4.24	702,394	4,625,933	3.92	17.83
6	58	11.18	623,532	23,308,249	22.68	39.11
7	42	8.09	102,405	18,884,929	18.78	44.72
8	28	5.39	1,735,101	19,319,498	17.58	62.80
9	30	5.78	511,381	10,220,876	9.71	32.36
10	32	6.17	204,814	21,881,234	21.68	67.74
11	142	27.36	30,587	46,434,989	46.40	32.68

Table 4 Percentage of transition and transversion identified across the *C. cajan* genome

SNP type	Transitions		Transversions			
	A/G	C/T	A/T	A/C	G/T	G/C
Number of allelic sites	213	249	105	76	83	68
Frequencies (%)	26.82	31.36	13.22	9.57	10.45	8.56
Total (percentage)	462 (58.18%)		332 (41.81)			

Table 5 Summary of genetic parameters calculated from SSR and SNP data in the 80 pigeon pea accessions from Benin

Genetic parameters	SSR		SNP	
	Range	Mean	Range	Mean
MAF	0.23–0.98	0.51	0.5–0.95	0.79
N_A	2–15	6.97	–	2
H_O	0.0–0.94	0.41	0.0–0.9	0.35
H_e	0.04–0.86	0.62	0.1–0.5	0.3
PIC	0.04–0.84	0.57	0.09–0.38	0.25

MAF = major allele frequency; N_A = number of alleles; PIC = polymorphism information content; H_O = observed heterozygosity; H_e = expected heterozygosity

The allele frequency spectrum was also calculated to examine distribution of allele frequency for both data sets (Fig. 2a, b). Based on frequency of alleles generated with the SSR, the majority (61.24%) was rare with allele frequencies lower than 0.1 (Fig. 2a). The number of SSR alleles with high frequency (0.8 to 1) was very low and represents only 1.42% of total alleles detected. For the bi-allelic GBS SNPs, the majority of SNP markers (27.64%) showed a major allele frequency above 0.8, indicating that rare alleles are enriched (Fig. 2b).

Population structure within pigeon pea collection

The software package STRUCTURE was used to infer the population structure within the analyzed pigeon pea collection using the data generated by SSR and SNP genotyping. Based on ΔK method described by Evanno et al. (2005), the peak of ΔK was observed at K = 3 for both marker types (Fig. 3). However, there were also some minor peaks which indicate that a more fine-grained structure could possibly be observed if more markers or accessions could be utilized. Referring to these results, three main genetic subpopulations were identified in the Beninese pigeon pea germplasm collection and both SNP and SSR were efficient in detecting such subdivisions.

Pigeon pea accessions were assigned to subpopulations based on membership coefficients (membership probability higher than 0.70). However, the number of individuals assigned into different subpopulations varied substantially according to the marker type used (Fig. 4). Indeed, 16 accessions representing

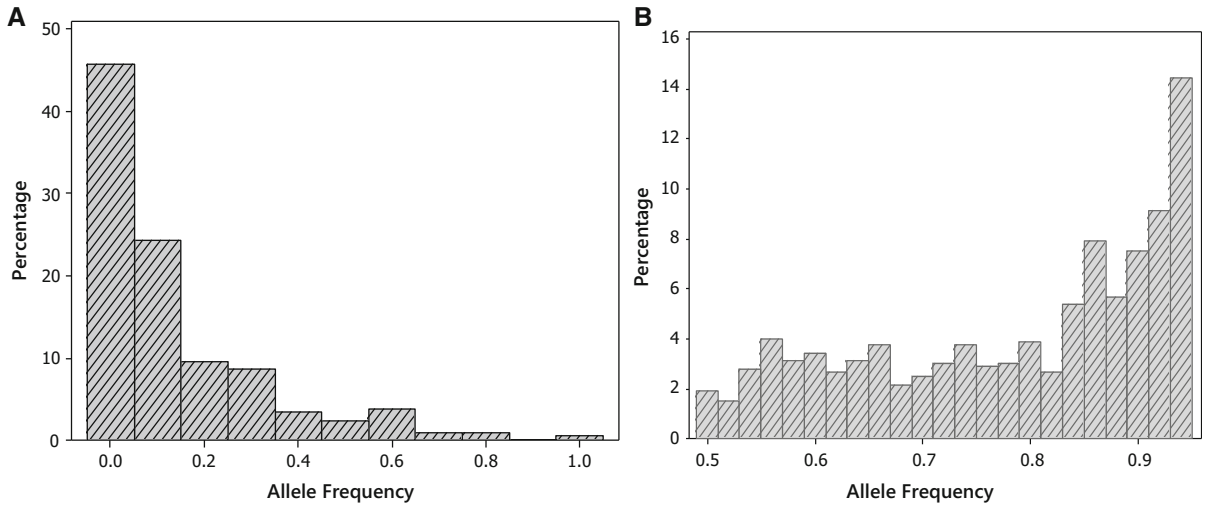

Fig. 2 Allele frequency distribution in the 80 analyzed *Cajanus cajan* accessions from Benin **a** for the 30 SSR and **b** for the 794 SNP markers. **a** The allele frequency spectrum of SSR markers and **b** the major allele frequencies of the SNP markers

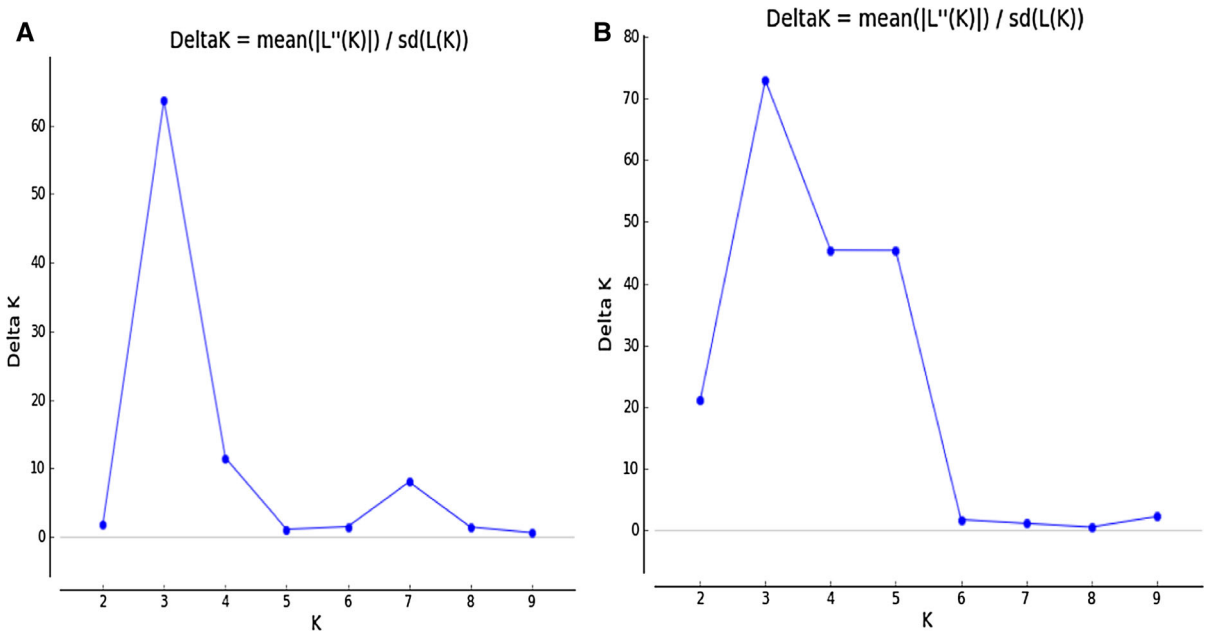

Fig. 3 Estimation of number of subpopulations in the 80 Beninese pigeon pea accessions using ΔK method **a** for 30 SSRs and **b** for 794 SNPs

20% of the total number of analyzed accessions were detected as admixture (membership probability lower than 0.70) for SNP markers while 19 accessions were found admixed in case of SSR markers. For the latter, the three subpopulations identified, i.e. $K1_{SSR}$ (red color), $K2_{SSR}$ (green color) and $K3_{SSR}$ (blue color), were respectively composed of 14, 21 and 26 accessions (Fig. 4a). In case of SNP markers, the three clusters $K1_{SNP}$ (red color), $K2_{SNP}$ (green color) and $K3_{SNP}$ (blue color) included 2, 46 and 16 accessions, respectively (Fig. 4b). The composition of these groups showed that the two individuals of $K1_{SNP}$ were also clustered together in $K1_{SSR}$. Besides, among the 21 individuals of $K2_{SSR,}$ the majority (14 accessions) were found together in $K2_{SNP}$. The $K3_{SNP}$ and $K3_{SSR}$ had in common 9 accessions. Among the

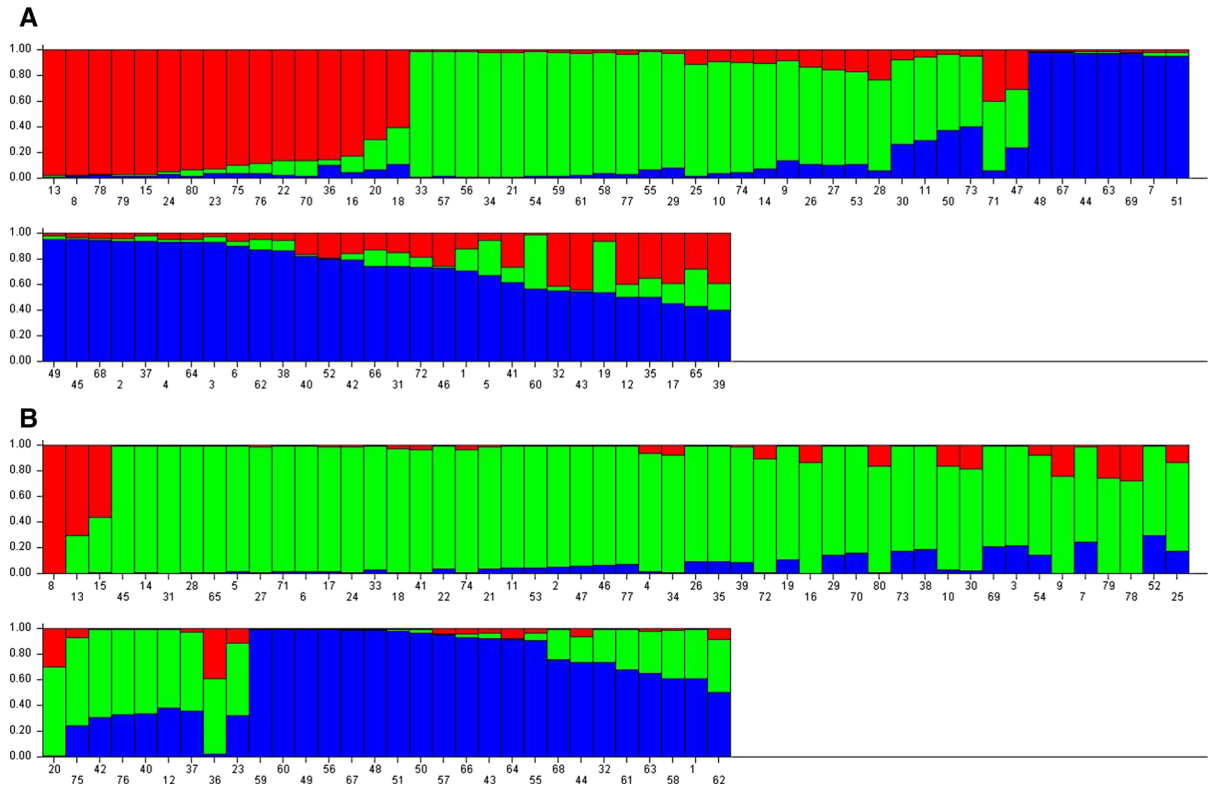

Fig. 4 Population structure of the 80 *Cajanus cajan* accessions from Benin (K = 3) inferred based on **a** 30 SSRs and **b** 794 SNPs

accessions identified as admixture, three were simultaneously detected as such by both marker types.

Genetic divergences of subpopulations inferred by AMOVA and PCoA analyses

To appreciate the degree of genetic divergence among the different subpopulations identified by STRUCTURE, an analysis of molecular variance (AMOVA) followed by a principal coordinate analysis (PCoA) was further conducted. For SSR markers, the AMOVA showed that only 1% (Table 6) of the total genotypic variation was explained by the difference among inferred subpopulations. The maximum variation has been observed within subpopulation (99%) across individuals (41%) and within individuals (58%). In case of SNP markers, no diversity existed across individuals within subpopulation (Table 6), whereas 12% of the total genotypic variation was caused by the difference among the subpopulations and 88% of variation was observed within individuals within subpopulation. Supporting the above AMOVA results, a low F_{ST} value (0.014) and high Nm (17.01) were observed for the subpopulations identified with SSR data, while moderate overall F_{ST} values (0.134) and low Nm (1.614) were obtained with SNPs for subpopulations (Table 6).

Principal coordinate analysis (PCoA) with SSR showed that the first two axes explained 14.96% of the overall genotypic variation (Table 7). The projection of the pigeon pea accessions into the biplot formed by the two first PCo axes revealed that no population structure was apparent. The distribution was uniform (Fig. 5a) and it was not possible to identify the three subpopulations delimited with STRUCTURE. Contrary to the spatial distribution observed in case of SSRs, highly structured population was observed in case of SNPs (Fig. 5b). The proportion of genotypic variance explained by the first two principal coordinates in case of SNPs was 20.57% (Table 7). The PCoA results for SNPs were consistent with STRUCTURE analysis. The three subpopulations identified by STRUCUTRE were clearly separated by the first and second principal coordinate (Fig. 5b). As observed in AMOVA results,

Table 6 Analysis of molecular variance (AMOVA) among and within subpopulations of 80 *Cajanus cajan* accessions using the 30 SSR and the 794 SNP

Sources of variations	AMOVA for SSR					AMOVA for SNP				
	df	SS	MS	Est. Var	%	df	SS	MS	Est. Var	%
Among subpopulations	2	38.98	19.49	0.14	1	2	1149.18	574.59	17.87	12
Across Individuals within subpopulation	58	801.77	13.82	4.04	41	61	5743.20	94.151	0.00	0
Within Individuals within subpopulation	61	349.50	5.73	5.73	58	64	8751.00	136.73	136.73	88
Total	121	1190.26		9.92	100	127	15,643.39		154.61	100
Fixation index (F_{ST})	F_{ST} = 0.014					F_{ST} = 0.134				
Nm (haploid)	Nm = 17.01					Nm = 1.614				

Table 7 Percentage of variation explained by the first three axes

Variation	SSR			SNP		
	Axis 1	Axis 2	Axis 3	Axis 1	Axis 2	Axis 3
%	8.97	5.99	5.68	11.15	9.42	6.63
Cum %	8.97	14.96	20.63	11.15	20.57	27.20

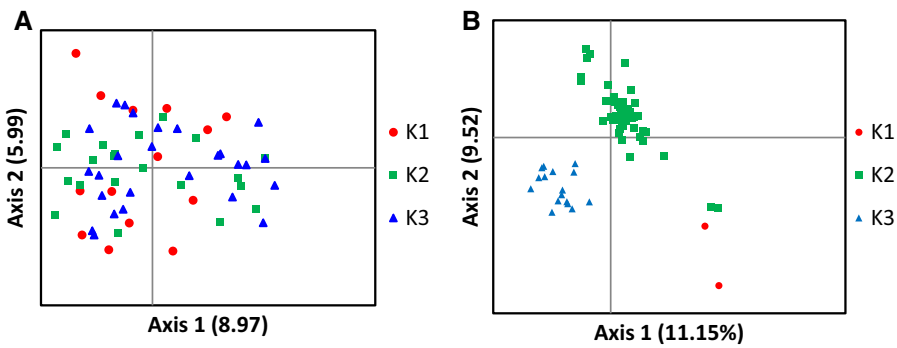

Fig. 5 Principal coordinates analysis showing the divergence within Beninese pigeon pea accessions based on SSR (**a**) and SNP (**b**) markers. Individuals with same color belong to the same genetic group

PCoA performed with SNPs showed high differentiation between the subpopulations while PCoA results from SSR markers revealed low genetic differentiation between subpopulations identified. Globally, by combining STRUCTURE analysis with AMOVA and PCoA results, it appeared that (1) a genetic structure exists in Beninese pigeon pea collection and (2) the SNP markers performed better in population genetic structure analysis than SSRs.

Genetic variability within pigeon pea subpopulations

The genetic diversity parameters estimated for each subpopulation identified with SSR and SNP markers are summarized in the Table 8. Among the three subpopulations identified with SSR markers, the highest diversity (He = 0.57) was observed within subpopulation 3 (K_{3SSR}) composed of 26 individuals. This subpopulation (K_{3SSR}) displayed also a high number of polymorphic loci (100%) and the mean values of Na, I and uHe within K_{3SSR} were 5.63, 1.228 and 0.625, respectively. For subpopulations identified by SNP markers, the highest diversity (H*e* = 0.30) was observed within K_{2SNP} that included 46 individuals and held the highest number of polymorphic loci (99.37%). The mean values in Na, I and uHe within K_{2SNP} were 1.994, 0.47 and 0.31 respectively.

Table 8 Genetic variability within each pigeon pea subpopulations identified with SSR and SNP data

Genetic parameters	Genetic groups SSR			Genetic groups SNP		
	K_{1SSR}	K_{2SSR}	K_{3SSR}	K_{1SNP}	K_{2SNP}	K_{3SNP}
N	14.000	21.000	26.000	2.000	46.000	16.000
N_A	4.600	5.233	5.633	1.364	1.994	1.788
N_E	2.857	3.081	3.139	1.261	1.510	1.348
I	1.126	1.187	1.228	0.219	0.470	0.334
H_O	0.374	0.443	0.376	0.229	0.391	0.233
H_E	0.579	0.600	0.612	0.150	0.306	0.214
uHe	0.601	0.615	0.625	0.201	0.310	0.221
F	0.363	0.288	0.434	− 0.487	− 0.200	− 0.042
Arp	0.300	0.333	0.800	0.001	0.126	0.005
PPL	100.00%	96.67%	100.00%	36.52%	99.37%	78.84%

N = number of samples; N_A = number of alleles; N_E = number of effective alleles, I = Shannon's information index, H_O = observed heterozygosity; H_e = expected heterozygosity, uHe = unbiased expected heterozygosity, F = fixation index, Arp = private allelic richness, PPL = percentage of polymorphic loci

Phylogenetic analysis with SSR and SNP

The genetic relationships between the 80 accessions were separately assessed with SSR and SNP data. Nei's genetic distances between pairs of accessions estimated for both marker types showed wide variations. For SSRs, the Nei's genetic distance ranged from 0.3 to 0.88, while it varied from 0.08 to 0.44 for SNPs. A Mantel test, performed based on the genetic distance matrices generated from the SNP and SSR markers, revealed a positive and significant correlation ($r = 0.39$, $P = 0.01$). The phylogenetic trees constructed from the two dissimilarity matrices illustrated the genetic relationships between accessions and their grouping (Fig. 6). A comparative analysis of these phylogenetic trees generated based on these genetic distances revealed clear discrepancies between the two marker types. In case of SNP markers, the 80 accessions were classified into two major groups (Fig. 6b). Almost all Beninese pigeon pea cultivars were clustered together in a same group except four accessions which were classified in a separate group with two breeding lines obtained from Ibadan. In case of SSRs, the dendrogram constructed grouped the 80 accessions into three major groups (Fig. 6a). The groups labeled in red, green and blue was composed of 37, 11 and 32 accessions respectively. The results observed in phylogenetic analysis were thus not consistent for SSR and SNP markers even though moderate correlation was found between the genetic distances generated with both marker types.

Discussion

Molecular characterization of crop germplasm is essential for its efficient exploitation (Adoukonou-Sagbadja et al. 2007). This study provides the first and detailed insight on the genetic diversity and population structure in a large and representative collection of pigeon pea landraces grown in Benin using both SSR and SNP markers. Furthermore, it was possible to compare the efficiency of these two molecular markers in diversity and population genetic analysis in pigeon pea. To conclude, results gathered from this study are highly useful not only for conservation management of pigeon pea genetic resources in Benin, but also for exploitation in breeding.

Efficiency of SSR and SNP markers in pigeon pea population genetic analysis

Although several molecular markers have been developed and now available for pigeon pea (Varshney et al. 2013), SSR and GBS SNP have been considered as markers of choice for genetic and breeding applications (Gupta and Varshney 2000; Mir and Varshney 2013). These two marker types have been previously

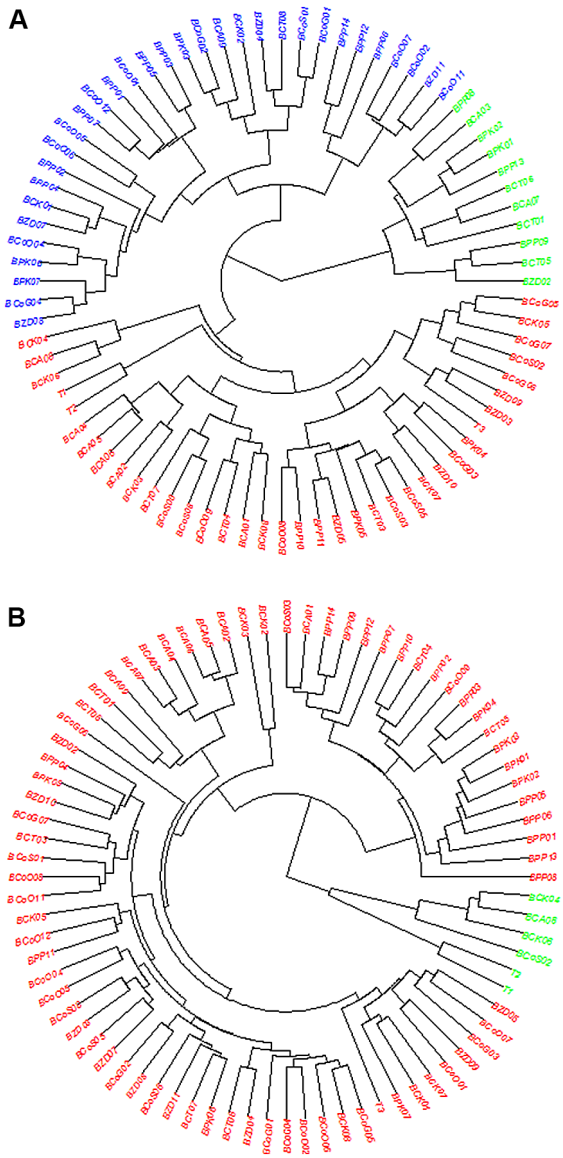

Fig. 6 Phylogenetic trees constructed from **a** SSR and **b** SNP data showing the relationships between the 80 Beninese pigeon pea accessions evaluated

well solely used to assess the genetic diversity in cultivated pigeon pea as well as in its wild relatives (Odeny et al. 2007; Kassa et al. 2012; Singh et al. 2013a, b; Saxena et al. 2014; Njung'e et al. 2016). The comparative study of these two markers was earlier reported for other crops such as spring barley (Bengtsson et al. 2017), sunflower (Filipi et al. 2015), winter wheat (Wurschum et al. 2013) or rice (Singh et al. 2013a, b). The results observed here by genotyping simultaneously the Beninese pigeon pea landraces' collection with SSR and SNP markers revealed that both markers were well informative for diversity analysis. For instance, based on theoretical expectation for bi-allelic markers such as SNPs, the PIC and gene diversity values range from 0 (fixation of one allele) to 0.5 (equal allele frequencies), whereas for multiallelic markers like SSRs the PIC and the gene diversity values can exceed 0.5 and approach 1 (Wurschum et al. 2013). The values of PIC and gene diversity observed here for SSR and SNP support such assertion and clearly indicate that both marker types are all powerful in genetic diversity estimation in pigeon pea.

However, as expected, the information gathered from the study differed significantly according to the marker system. The mean values of inferred genetic parameters were higher for SSRs than those of SNPs. For instance, the average PIC and genetic diversity values were twice higher for the SSRs compared to the SNPs. Although contrasting Singh et al. (2013a, b) where both marker systems (SNP and SSR) generated almost similar values of genetic parameters, this finding is consistent with many other previous studies in which the genetic parameters estimated with SSR were higher than those calculated of SNPs (Bengtsson et al. 2017; Chen et al. 2017; Desalegne et al. 2017; Emanuelli et al. 2013; Li et al. 2010). This could likely be due to the difference in the number of SSR and SNP markers used and their distribution. According to Hamblin et al. (2007), it could also be explained by the differences in the allele frequency spectrum of these two marker types. Inspection of the distribution of SNP allele frequencies showed a pattern different from that observed for SSR. A larger proportion of alleles were observed at intermediate frequency for SNPs, whereas there was a higher presence of SSR alleles at low frequencies (Fig. 2).

Similar results have been previously also reported by Filipi et al. (2015). Besides, a number of other differences were found between the two markers types in Bayesian population structure analysis as well in phylogenetic tree construction with the number of individuals assigned into different clusters varying substantially according to the marker systems (Fig. 4). Although partly contradicting those from Bengtsson et al. (2017) in Nordic spring barley, these findings are nonetheless in line with the fact that the marker type can significantly influence the quality of information

revealed in diversity and population structure analyses (Wurschum et al. 2013). The discrepancies in efficiency of these two marker types could be related to the different mutational mechanisms behind the two marker systems (replication slippage for SSRs against point mutation for SNPs), the high mutation rate of SSRs, the biallelic (SNPs) versus multi-allelic (SSRs) status of the two markers (Martinez-Arias et al. 2001; Bengtsson et al. 2017). Although these two marker systems were found all informative, it can be concluded from the present study that significant differences exist in their efficiency for diversity and population structure analyses in pigeon pea with a resolving power in population structure analysis higher for SNP than for SSR.

Genetic diversity in Beninese pigeon pea germplasm

The 30 SSR loci explored in the study showed a high polymorphism with 209 alleles revealed across the 80 accessions and an average of 6.97 alleles per locus. This polymorphism is higher compared to those reported by Sousa et al. (2011), Petchiammal et al. (2015), Bohra et al. (2017) and Odeny et al. (2007) although the latter additionally included wild species in their investigations. This result suggests that there is an important genetic variability within the analyzed Beninese pigeon pea collection. However, the high number of alleles observed here compared to previous studies could be attributed to the difference in the methods for revealing PCR products. Indeed, in this study, a capillary electrophoresis was used to detect and determine allele size. This method is more effective in allele detection compared to agarose and polyacrylamide gel electrophoreses which were used in earlier pigeon pea diversity studies and are known to underestimate the number of alleles (Gupta et al. 2010). This hypothesis is in congruence with the large number of alleles detected in Malawi pigeon pea collection using capillary electrophoresis for analyzing PCR products (Njung'e et al. 2016). Besides, the highest average *PIC* value (0.57) estimated in the present study gives hint to the high discriminatory power of the SSR markers used. The relatively high genetic diversity detected in the collection with SSR markers was confirmed by the high value of *PIC* and gene diversity calculated with SNP markers. Indeed, the overall *PIC* value (0.25) and gene diversity (0.30) detected within the pigeon pea collection using SNP markers were higher than that calculated with the same marker type by Saxena et al. (2014) across 21 pigeon pea landraces (0.2). This discrepancy could be attributed to the differences in the number of accessions analyzed and the composition of the genotype set. However, the values of these two genetic parameters (PIC and gene diversity) were similar to *PIC* value (0.24) and gene diversity (0.30) observed in 107 accessions from 18 pigeon pea wild species analyzed by Saxena et al. (2014). In addition, the 794 detected SNP markers detected here by GBS represent different mutation points revealing a total of 1,588 alleles across pigeon pea genome. All these results supported our previous conclusion that importantly large genetic diversity exists in Beninese pigeon pea collection (Zavinon et al. 2019) in contrast to the general trend of narrow genetic base of cultivated pigeon pea gene pool (Yang et al. 2006; Odeny et al. 2007; Kassa et al. 2012). This large genetic diversity in Beninese landraces collection could partly result from local gene flow between cultivated and wild pigeon pea populations as it has previously been reported by Kassa et al. (2012). According to these authors, gene flow is possible since many crops, particularly the minor/ neglected crops, are still grown alongside their wild relatives and the outcrossing in cultivated pigeon pea and its wild species may raise up to 70% and 17%, respectively. Among the 34 species of *Cajanus* genus (Lewis et al. 2005) one (*Cajanus kerstingii*) is endemic to West-Africa (Odeny 2006). This wild species through local gene flow could reinforce the diversity in the cultivated germplasm as outcrossing rate can reach 45% in pigeon pea (Saxena et al. 1990).

Genetic structure in Beninese pigeon pea germplasm

The occurrence of population structure in Beninese collection was separately assessed with the 794 SNPs as well the 30 SSRs markers using the Bayesian cluster analysis implemented in the software package STRUCTURE. The results revealed the existence of a genetic structure within Beninese pigeon pea collection. Indeed, whatever the marker type used, the ΔK method described by Evanno et al. (2005) suggested the subdivision of the collection into three main genetic sub-populations. Other value of K are possible since some small peaks were observed for ΔK

at K = 4; K = 7 (SSR; Fig. 3a) and at K = 4; K = 5 (SNP; Fig. 3b). This result indicates that more than 3 subpopulations are also possible but it would not qualitatively affect our conclusion (3 subpopulations). The presence of genetic structure in cultivated pigeon pea was previously detected by Sousa et al. (2011) in South American pigeon pea germplasm using SSR markers. Recently, several different genetic groups have been detected within the cultivated pigeon pea varieties collection from Malawi (Njung'e et al. 2016). In the present study, the consistency in the determination of number of subpopulations observed with the SSR and SNP confirmed the evidence of genetic substructure within the Beninese pigeon pea collection and showed that both marker types were equally powerful for identifying population subdivisions. This corroborates the findings reported earlier in other studies (Filipi et al. 2015; Chen et al. 2017). However, the composition and size of inferred subpopulations were not identical with for the two marker types even though some accessions were grouped together whatever the marker type. Indeed, the STRUCTURE analysis revealed high number of accessions as admixture varieties. With SSR and SNP markers, the number of admixed varieties detected by STRUCTURE analysis was important (16 and 19 respectively), comparably to those reported in earlier studies (Sousa et al. 2011; Bohra et al. 2017). The high number of admixed varieties in Beninese collection could be firstly explained by the cultural practices linked to seed management, which favored gene flow through seeds exchanges. Such cultural practices had been already reported by Zavinon et al. (2018). According to the authors, pigeon pea producers in Benin do not conduct any specific agricultural practices to ensure the varietal purity. The second most likely explanation of this result is related to the gene flow due to the reproduction system through pollens exchanges. Indeed, despite the fact that pigeon pea is considered as a self-pollinated species (Reddy 1990; Abrol and Shankar 2015), the out-crossing rate can reach 45% (Saxena et al. 1990) and represents a major source of varietal contamination. In a similar way, the result may be attributed to the spread of plant material among localities confirming the gene flow between populations. This is in line with the absence of correlation of the subpopulations identified with the geographic origins of the analyzed accessions even though they were collected from four major pigeon pea growing areas in Benin. Similar observations have been reported in many previous studies such that on fonio millet landraces in West-Africa (Adoukonou-Sagbadja et al. 2007).

Whatever the marker type used, the three advanced varieties included in the present study were clustered together with some landraces. This high similarity between varieties and some landraces indicates that these varieties possibly shared a common genetic background or ancestries with the landraces. On the other hand, the close relationships between varieties and some landraces could be also explained by gene flow or seed material contamination due to the introduction of new varieties or adoption of improved varieties by the farmers. This result was confirmed by the genetic relationship observed between individuals based on the two dendrograms constructed with SSR and SNP data, respectively (Fig. 6).

To further investigate the population structure analysis in Beninese pigeon pea germplasm, we conducted an analysis of molecular variance (AMOVA) followed by a principal coordinate analysis (PCoA) for each marker type. Although the differentiation was low (1%) between the subgroups identified with the SSR data, a relatively high differentiation (12%) was observed between clusters identified with SNP data.

In practical sense, the present study on pigeon pea genetic diversity and its population structure offers prospects for improving the crop characteristics. The diverse genetic parameters estimated, especially the high level of allelic diversity with significant amount of rare alleles detected (Table 5; Fig. 2) indicate that the analysed collection constitutes a valuable resource that can be exploited by breeders to improve pigeon pea breeding in Benin. In applied breeding, understanding the population structure in a germplasm collection is pivotal for the identification of genes and quantitative trait loci underlying phenotypic variation (Soto-Cerda et al. 2012). Therefore, the existence of genetic structure in Beninese pigeon pea collection which possesses an enormous variability in morphological traits (Zavinon et al. 2019) will facilitate hybrid breeding as distinct heterotic groups are expected to provide predictors for high heterosis effects (Adoukonou-Sagbadja et al. 2007). Besides, genetic subdivisions are more interesting if they correlate with the agro-morphological traits of accessions. Indeed, a clear separation was observed between

the three early maturing improved varieties and the intermediate maturing landraces. Moreover, these three improved lines known for their high grain yield potential clustered together with some landraces presenting also high grain yield (Zavinon et al. 2019). Certainly, these accessions may share in common some particular alleles which are likely associated with high grain yield. As it has been already stressed in some studies (e.g. Adoukonou-Sagbadja et al. 2007), additional experiments such as QTL mapping studies are needed to identify specific genes or genomic regions that have an influence on the phenotypic variation. These results globally offer a good opportunity not only for improving particular traits of interest, but also for defining conservation strategies to safeguard pigeon pea landraces in cultivation in Benin.

Conclusion

This study provides a detailed insight in the genetic diversity in pigeon pea landraces and cultivars grown in Benin using molecular markers. It has also compared the efficiency of SSR and SNP markers in population genetic analysis in cultivated pigeon pea. The results obtained with the two marker types revealed a great genetic diversity and highlighted a genetic structure within the Beninese pigeon pea collection. Though the two marker types were equally appropriate for estimating genetic diversity, at population level the resolving power of the SNP was higher than that of the SSR. The results of this study are an important contribution to pigeon pea breeding and conservation in Benin.

Acknowledgements Fiacre Zavinon thanks DAAD for financial support and Dr. Rajeev Varshney from ICRISAT for help in selecting SSR markers used in this study. Special thanks to Katy Niedung, Marlis Weilepp and Thomas Berner for technical assistance. The authors extend also special thanks to Ayenan Mathieu for supplying a part of analyzed plant material and the two anonymous reviewers for their critical comments.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abrol DP, Shankar U (2015) Role of pollination in pulses. Adv Pollen Spore Res 33:101–103

Adoukonou-Sagbadja H, Wagner C, Dansi A, Ahlemeyer J, Daïnou O, Akpagana K, Ordon F, Friedt W (2007) Genetic diversity and population differentiation of traditional fonio millet (*Digitaria* spp.) landraces from different agroecological zones of West Africa. Theor Appl Genet 115:917–931

Ae N, Arihara J, Okada K, YoshiharaT JC (1990) Phosphorous uptake by pigeon pea and its role in cropping systems of the Indian subcontinent. Science 248:477–480

Ayenan MAT, Danquah A, Ahoton LE, Ofori K (2017a) Utilization and farmers' knowledge on pigeon pea diversity in Benin, West Africa. J Ethnobiol Ethnomed 13:37. https://doi.org/10.1186/s13002-017-0164-9

Ayenan MAT, Ofori K, Ahoton LE, Danquah A (2017b) Pigeonpea [(*Cajanus cajan* (L.) Millsp.)] production system, farmers' preferred traits and implications for variety development and introduction in Benin. Agric Food Secur 6:48. https://doi.org/10.1186/s40066-017-0129-1

Bengtsson T, The PPP Barley Consortium, Manninen O, Jahoor A, Orabi J (2017) Genetic diversity, population structure and linkage disequilibrium in Nordic spring barley (*Hordeum vulgare* L. subsp. vulgare). Genet Resour Crop Evol 64(8):2021–2033

Bohra A, Dubey A, Saxena RK, Penmetsa RV, Poomima KN, Kumar N, Farmer AD, Srivani G, Upadhyaya HD, Gothalwal R, Ramesh S, Singh D, Saxena K, Kishor PBK, Singh KN, Town CD, May GD, Cook DR, Varshney RK (2011) Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeon pea (*Cajanus* spp.). BMC Plant Biol 11:56. https://doi.org/10.1186/1471-2229-11-56

Bohra A, Saxena RK, Gnanesh BN, Saxena K, Byregowda M, Rathore A, KaviKishor PB, Cook DR, Varshney RK (2012) An intra-specific consensus genetic map of pigeon pea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. Theor Appl Genet 125:1325–1338

Bohra A, Jha R, Pandey G, Patil PG, Saxena RK, Singh IP, Singh D, Mishra RK, Mishra A, Singh F, Varshney RK, Singh NP (2017) New hypervariable SSR markers for diversity analysis, Hybrid purity testing and trait mapping in pigeon pea [*Cajanus cajan* (L.) Millspaugh]. Front Plant Sci 8:377. https://doi.org/10.3389/fpls.2017.00377

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet 32(3):314–331

Chen W, Hou L, Zhang Z, Pang X, Li Y (2017) Genetic diversity, population structure, and linkage disequilibrium of a core collection of *Ziziphus jujuba* assessed with genome-wide SNPs developed by genotyping-by-sequencing and SSR markers. Front Plant Sci 8:575. https://doi.org/10.3389/fpls.2017.00575

Dansi A, Vodouhe R, Azokpota P, Yedomonhan H, Assogba P, Adjatin A, Dossou-Aminon I, Akpagana K (2012) Diversity of the neglected and underutilized crop species of

importance in Benin. Sci World J. https://doi.org/10.1100/2012/932947

Desalegne BA, Dagne K, Melaku G, Ousmane B, Fatokun CA (2017) Efficiency of SNP and SSR-based analysis of genetic diversity, population structure, and relationships among Cowpea (*Vigna unguiculata* (L.) Walp.) germplasm from East Africa and IITA Inbred Lines. J Crop Sci Biotechnol 20(2):107–128

Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361

Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, Myles S, Martinez-Zapater JM, Zyprian E, Moreira FM, Grando MS (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biol 13:39. https://doi.org/10.1186/1471-2229-13-39

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Filipi CV, Aguirre N, Rivas JG, Zubrzycki J, Puebla A, Cordes D, Moreno MV, Fusari CM, Avarez D, Heinz RA, Hopp HE, Paniego NB, Lia VV (2015) Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. BMC Plant Biol 15:52. https://doi.org/10.1186/s12870-014-0360-x

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 13:163–185

Gupta V, Dorsey G, Hubbard AE, Rosenthal PJ, Greenhouse B (2010) Gel versus capillary electrophoresis genotyping for categorizing treatment outcomes in two anti-malarial trials in Uganda. Malar J 9:19. https://doi.org/10.1186/1475-2875-9-19

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS ONE 2:e1367. https://doi.org/10.1371/journal.pone.0001367

Kassa MT, Penmetsa RV, Carraquille-Garde N, Sarma BK, Datta S, Upadhyaya HD, Varshney RK, von Wettberg EJB, Cook DR (2012) Genetic patterns of domestication in pigeon pea (*Cajanus cajan* (L.) Millsp) and wild *Cajanus* relatives. PLoS ONE 7(6):e39563. https://doi.org/10.1371/journal.pone.0039563

Kumar A, Priyanka VL, Lall V, Lal D (2011) A biotic factors and pigeon pea pod fly, *Melanagromyza Obtusa* (Malloch). Indian J Entomol 73:59–62

Kumar V, Khan AW, Saxena RK, Garg V, Varshney RK (2016) First-generation HapMap in *Cajanus* spp. reveals untapped variations in parental lines of mapping populations. Plant Biotechnol J 14(8):1673–1681

Kumawat G, Raje RS, Bhutani S, Pal JK, Mithra ASVCR, Gaikwad K, Sharma TR et al (2012) Molecular mapping of QTLs for plant type and earliness traits in pigeon pea (*Cajanus cajan* L. Millsp.). BMC Genet 13:84. https://doi.org/10.1186/1471-2156-13-84

Lewis G, Schrire B, MacKinder B, Lock M (2005) Legumes of the world. Royal Botanic Gardens, Kew Publishing, London

Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. Bioinformatics 24:713–714

Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS et al (2010) Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. New Phytol 188:242–253

Liu KJ, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Martinez-Arias R, Calafell F, Mateu E, Comas D, Andres A, Bertranpetit J (2001) Sequence variability of a human pseudogene. Genome Res 11(6):1071–1085

Mir RR, Varshney RK (2013) Future prospects of molecular markers in plants. In: Henry RJ (ed) Molecular markers in plants. Blackwell, Oxford, pp 170–190

Mir RR, Kudapa H, Srikanth S, Saxena RK, Sharma A, Azam S, Saxena K, Penmetsa RV, Varshney RK (2014) Candidate gene analysis for determinacy in pigeon pea (*Cajanus* spp.). Theor Appl Genet 127:2663–2678

Mohamed A, García-Martínez S, Loumerem M, Carbonell P, Ruiz JJ, Boubaker M (2019) Assessment of genetic diversity among local pea (*Pisum sativum* L.) accessions cultivated in the arid regions of Southern Tunisia using agro-morphological and SSR molecular markers. Genet Resour Crop Evol 66:1189–1203. https://doi.org/10.1007/s10722-019-00784-8

Njung'e V, Deshpande S, Siambi M, Jones R, Silim S, De Villiers S (2016) SSR genetic diversity assessment of popular pigeon pea varieties in Malawi reveals unique fingerprints. Electron J Biotechnol 21:65–71

Odeny AD (2006) Microsatellite development and application in Pigeon pea (*Cajanus cajan* (L.) Millsp). Ph.D. Dissertation, Bonn

Odeny DA, Jayashree B, Ferguson M, Hoisington D, Crouch J, Gebhardt C (2007) Development, characterization and utilization of microsatellite markers in pigeon pea. Plant Breed 126:130–136

Odeny DA, Shah T, Saxena RK (2016) Genomic resources for pigeon pea. Legumes Perspect 11:33–34

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537–2539

Perovic J, Silvar C, Koenig J, Stein N, Perovic D, Ordon F (2013) A versatile fluorescence-based multiplexing assay for CAPS genotyping on capillary electrophoresis systems. Mol Breed 32:61–69

Petchiammal KI, Muthiah AR, Jayamani P (2015) Molecular characterization of CGMS, maintainer and inbred lines and diversity analysis in pigeon pea [*Cajanus cajan* (L.) Millsp.]. Legume Res 38(2):734–741

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7(2):e32253. https://doi.org/10.1371/journal.pone.0032253

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170–181

Reddy LJ (1990) Pigeonpea: morphology. The pigeon pea. CAB International, Wallingford, Oxon, pp 47–88

Saxena KB, Singh L, Gupta MD (1990) Variation for natural out-crossing in pigeon pea. Euphytica 46:143–148

Saxena RK, von Wettberg E, Upadhyaya HD, Sanchez V, Songok S, Saxena K, Kimurto P, Varshney RK (2014) Genetic diversity and demographic history of *Cajanus* spp. illustrated from genome-wide SNPs. PLoS ONE 9(2):e88568. https://doi.org/10.1371/journal.pone.0088568

Saxena RK, Saxena KB, Pazhamala LT, Patel K, Parupalli S, Sameerkumar CV, Varshney RK (2015) Genomics for greater efficiency in pigeon pea hybrid breeding. Front Plant Sci. 6:793. https://doi.org/10.3389/fpls.2015.00793

Saxena RK, Kale SM, Kumar V, Parupali S, Joshi S, Singh V, Garg V, Das RR, Sharma M, Yamini KN, Ghanta A, Rathore A, Sameerkumar CV, Saxena KB, Varshney RK (2017) Genotyping-by-sequencing of three mapping populations for identification of candidate genomic regions for resistance to sterility mosaic disease in pigeon pea. Sci Rep 7(1):1813. https://doi.org/10.1038/s41598-017-01535-4

Sharma S, Agarwal N, Verma P (2011) Pigeon pea (*Cajanus cajan* L.): a hidden treasure of regime nutrition. J Funct Environ Bot 1(2):91–101

Sharma M, Ghosh R, Telangre R, Rathore A, Saifulla M, Mahalinga DM, Saxena DR, Jain YK (2016) Environmental influences on Pigeonpea–Fusarium udum interactions and stability of genotypes to Fusarium Wilt. Front Plant Sci 7:253. https://doi.org/10.3389/fpls.2016.00253

Singh N, Tyagi RK, Pandey C (2013) Genetic resources of pigeon pea: conservation for use. National Bureau of Plant Genetic Resources (NBPGR), New Delhi, pp 1–49

Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K et al (2013) Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. PLoS ONE 8(12):e84136. https://doi.org/10.1371/journal.pone.0084136

Soto-Cerda BJ, Maureira-Butler I, Munoz G, Rupayan A, Cloutier S (2012) SSR-based population structure, molecular diversity and linkage disequilibrium analysis of a collection of flax (*Linum usitatissimum* L.) varying for mucilage seed-coat content. Mol Breeding 30:875–888

Sousa ACB, Godoy R, Sforça DA, de Campos T, Zucchi MI, Jank L, de Sousa AP (2011) Genetic diversity analysis among pigeon pea genotypes adapted to South American regions based on microsatellites markers. Sci Agric 68(4):431–439

Stein N, Herren G, Beat K (2001) A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum saetivum*. Plant Breed 120(4):354–356

Van der Maesen LJG (1990) Pigeon pea: origin, history, evolution, and taxonomy. In: Nene YH, Hall SD, Sheila VK (eds) The pigeon pea. CAB International, Wallingford, pp 15–46

Varshney RK, Penmetsa RV, Dutta S, Kulwal PL, Saxena RK, Datta S, Sharma TR, Rosen B, Carrasquilla-Garcia N et al (2010) Pigeon pea genomics initiative (PGI): an international effort to improve crop productivity of pigeon pea (*Cajanus cajan* L.). Mol Breeding 26:393–408

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT et al (2012) Draft genome sequence of pigeon pea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83–89

Varshney RK, Mohan SM, Gaur PM, Gangarao NVPR, Pandey MK, Bohra A et al (2013) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Adv 31:1120–1134

Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, Kim C et al (2017) Whole-genome resequencing of 292 pigeon pea accessions identifies genomic regions associated with domestication and agronomic traits. Nat Genet 49(7):1082–1088

Wendler N, Mascher M, Noh C, Himmelbach A, Scholz U, Ruge-Wehling B, Stein N (2014) Unlocking the secondary gene-pool of barley with next-generation sequencing. Plant Biotechnol J 12:1122–1131

Wurschum T, Langer SM, Friedrich C, Longin H, Korzun V, Akhunov E, Ebmeyer E, Schachschneider R, Schacht J, Kazman E, Reif JC (2013) Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theor Appl Genet 126(6):1477–1486

Yang S, Pang W, Harper J, Carling J, Wenzl P, Huttner E et al (2006) Low level of genetic diversity in cultivated pigeon pea compared to its wild relatives is revealed by diversity arrays technology (DArT). Theor Appl Genet 113:585–595

Zavinon F, Adoukonou-Sagbadja H, Ahoton L, Vodouhê R, Ahanhanzo C (2018) Quantitative analysis, distribution and traditional management of pigeon pea [*Cajanus cajan* (L.) Millsp.] landraces' diversity in Southern Benin. Eur Sci J 14(9):184–211

Zavinon F, Adoukonou-Sagbadja H, Bossipkonnon A, Dossa H, Corneille A (2019) Phenotypic diversity for agro-morphological traits in pigeon pea landraces [(*Cajanus cajan* L.) Millsp.] cultivated in southern Benin. Open Agric 4:487–499