

RESEARCH ARTICLE

Genetic diversity and population structure of pigeonpea (*Cajanus cajan* [L.] Millspaugh) landraces grown in Benin revealed by Genotyping-By-Sequencing

Géofroy Kinhoégbè^{1,2*}, Gustave Djèdatin¹, Rachit Kumar Saxena^{2^{aa}}, Anu Chitikineni^{2^{ab}}, Prasad Bajaj², Johiruddin Molla², Clément Agbangla³, Alexandre Dansi⁴, Rajeev Kumar Varshney^{2^{ac}}

1 Laboratory of Molecular Biology and Bioinformatics Applied to Genomics, National University of Sciences, Technologies Engineering and Mathematics of Abomey, Dassa-Zoumé, Benin, **2** Centre of Excellence in Genomics and Systems Biology, International Crop Research Institute for the Semi-Arid Tropics, Hyderabad, India, **3** Laboratory of Molecular Genetic and Genomes Analysis, University of Abomey-Calavi, Abomey-Calavi, Benin, **4** Laboratory of Biotechnology, Genetic Resources and Plant and Animal Breeding, National University of Sciences Technologies Engineering and Mathematics of Abomey, Dassa-Zoumé, Benin

^{aa} Current address: Department of Crop Improvement and Plant Biotechnology, Gujarat Biotechnology University, Gandhinagar, India

^{ab} Current address: The UWA Institute of Agriculture, The University of Western Australia, Perth, Western Australia, Australia

^{ac} Current address: Murdoch's Centre for Crop & Food Innovation, State Agricultural Biotechnology Centre, Food Futures Institute, Murdoch University, Murdoch, Western Australia, Australia

* dongefroy@gmail.com



OPEN ACCESS

Citation: Kinhoégbè G, Djèdatin G, Saxena RK, Chitikineni A, Bajaj P, Molla J, et al. (2022) Genetic diversity and population structure of pigeonpea (*Cajanus cajan* [L.] Millspaugh) landraces grown in Benin revealed by Genotyping-By-Sequencing. PLoS ONE 17(7): e0271565. <https://doi.org/10.1371/journal.pone.0271565>

Editor: Reyazul Rouf Mir, Faculty of Agriculture (FoA), Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir (SKUAST-K), Wadura Campus, INDIA

Received: October 20, 2021

Accepted: July 4, 2022

Published: July 20, 2022

Copyright: © 2022 Kinhoégbè et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The sequencing data have been deposited in NCBI Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>) with the BioProject ID PRJNA845980.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Genetic diversity studies provide important details on target trait availability and its variability, for the success of breeding programs. In this study, GBS approach was used to reveal a new structuration of genetic diversity and population structure of pigeonpea in Benin. We used a total of 688 high-quality Single Nucleotide Polymorphism markers for a total of 44 pigeonpea genotypes. The distribution of SNP markers on the 11 chromosomes ranged from 14 on chromosome 5 to 133 on chromosome 2. The Polymorphism Information Content and gene diversity values were 0.30 and 0.34 respectively. The analysis of population structure revealed four clear subpopulations. The Weighted Neighbor Joining tree agreed with structure analyses by grouping the 44 genotypes into four clusters. The PCoA revealed that genotypes from subpopulations 1, 2 and 3 intermixed among themselves. The Analysis of Molecular Variance showed 7% of the total variation among genotypes while the rest of variation (93%) was within genotypes from subpopulations indicating a high gene exchange ($N_m = 7.13$) and low genetic differentiation ($\Phi_{PT} = 0.07$) between subpopulations. Subpopulation 2 presented the highest mean values of number of different alleles ($N_a = 1.57$), number of loci with private alleles ($P_a = 0.11$) and the percentage of polymorphic loci ($P = 57.12\%$). We discuss our findings and demonstrate how the genetic diversity and the population structure of this specie can be used through the Genome Wide Association Studies and Marker-Assisted Selection to enhance genetic gain in pigeonpea breeding programs in Benin.

Introduction

Pigeonpea (*Cajanus cajan* [L.] Millspaugh) is an important food legume crop widely grown in tropical and subtropical climates [1]. It is a diploid ($2n = 22$) crop with a genome size of 833.07 Mbp [2]. Pigeonpea is counted among the five comestible legumes used for food, fodder and firewood [3]. It constitutes a rich source of protein and is used to supplement cereals [4, 5].

In Benin, pigeonpea is largely cultivated and consumed in Adja socio-linguistics area in the South-East of the country and in some others socio-linguistics areas where it improves household income [6, 7]. Despite the importance of this legume, till date, no research effort has been directed towards pigeonpea crop improvement in the country. In general, there are several factors that contribute to the current level situation of pigeonpea development in the country. This includes lack of sufficient characterization of genetic resources and therefore, lack of genetic data [8].

The success of breeding programs, in any crop, depends on the extent of genetic diversity through different traits available to breeders for breeding desirable varieties [9]. These parameters significantly affect parents' choice for the hybridization and the adopted selection procedures [10]. In addition, by showing the differences which exist among accessions [11], the genetic diversity studies provide important details on target trait availability and its variability for the success of breeding programs [12–14]. Thus, before establishing conservation and improvement programs of pigeonpea genetic resources in Benin, it is necessary to discover the level of genetic diversity and population structure of cultivated landraces.

Different methods are usually used to access genetic diversity. One of them is the agro-morphological traits approach, given the accessibility of their easily measurable characteristics is the most used approach to establish relationships between genotypes and provide information for varieties improvement programs [15, 16]. Hence, agro-morphological traits have been utilized in pigeonpea germplasm's characterization of both world reference collections [17] and national collections [18–23]. In Benin, agro-morphological traits have already been used in the evaluation of pigeonpea genetic diversity [24–26]. However, the environmental conditions affect agro-morphological traits. Hence, morphological characters do not only reflect the genetic constitution of the cultivars but also the available genetic diversity [27].

Molecular markers are one of the powerful tools used in the characterization of genetic resources [28]. In recent years, the advancement in pigeonpea genomic resources is a result of the development of molecular markers and genome sequence data which are all essentials in molecular breeding approach [29]. Thus, in several crops, a few approaches have been used for simultaneous SNP discovery and genotyping such as Restriction site Associated DNA sequencing (RADseq) [30] and Genotyping-By-Sequencing (GBS) [31], etc. Among these approaches, GBS is the simplest and most cost-effective approach [31] and has proven itself in trait mapping and Genome-Wide Association Studies (GWAS) as well as in diversity studies [32] in various crops like chickpea [33], common bean [34], wheat [35], pigeonpea [36], etc.

In Benin, SNPs markers have been used to assess genetic diversity and population structure of cultivated pigeonpea [37]. However recently, new morphotypes of cultivated pigeonpea have been identified in the country and corresponding accessions collected [7]. This assumes a new structure of the pigeonpea diversity other than that reported by the previous study [37].

Therefore, in the present study, GBS approach was used to reveal a new structure within the pigeonpea genetic resource and to provide important information for pigeonpea breeding programs. Consequently, the objective of the present study was to characterize the genetic diversity and population structure of 50 pigeonpea landraces collected in Benin. We hypothesized that beninese pigeonpea germplasm encompasses more genetic diversity than reported by Zavinon et al. [37] who stated that there are three genetic groups and three subpopulations using SNPs markers.

Materials and methods

Plant materials

The material included 50 pigeonpea landraces collected in Benin and held in the genebank of the Laboratory of Molecular Biology and Bioinformatics applied to Genomics of the National University of Sciences, Technologies Engineering and Mathematics of Abomey in Benin and at International Crop Research Institute for the Semi-Arid Tropics at Patancheru in India. These genotypes are originated from 39 villages ([S1 Table](#)).

Germplasm and DNA isolation

Total DNA was isolated from two to three young leaves using a high-throughput DNA extraction protocol. The quality of the DNA was checked on 0.8% agarose gels. DNA concentration was assessed with NanoDrop ND-1000 (Thermo Fisher Scientific Inc., Waltham, USA) and extract purity using the absorbance ratio 260nm to 230nm. According to the DNA concentration, different dilution rates were applied to each sample to normalize each sample at 10 ng/ μ l for further use.

Library construction and Genotyping by Sequencing (GBS)

GBS libraries were prepared following Elshire et al. [31] and in accordance with the genomic analysis platform of the Center of Excellence in Genomics and Systems Biology at International Crops Research Institute for the Semi-Arid Tropics in India. Thus, the restriction enzyme ApeKI was used to digest each sample. The digested product was ligated to adapters with cohesive ends (GWC) by addition of T4 DNA ligase enzyme [36] before being incubated at 22°C and heated at 65°C for 1h and 30 min respectively to make the T4 DNA ligase enzyme inactivate. The digested and ligated products from each sample were mixed in equal proportion to construct the GBS libraries. Each library was then amplified and purified. The libraries were sequenced on an Illumina sequencer HiSeq 2500 platform (Illumina Inc, San Diego, CA, USA).

SNP calling and genotyping

The raw sequence data generated was filtered using Tassel v5.2.63 [38] analysis pipeline according Saxena et al. [36]. In this manner, SNPs were identified and high-quality SNP genotyping dataset was compiled. The draft genome sequence of pigeonpea described by Varshney et al. [2] was used as a reference to align the compiled dataset using Bowtie 2 software [39]. Subsequently, the alignment file was processed through GBS analysis pipeline for SNP calling and genotyping. Eventually, quality control filters were applied to both SNPs and samples following Zhang et al. [40] and Hussain et al. [41] based on the following criteria. Firstly, variants should be bi-allelic SNPs. Secondly, SNPs with more than 20% of missing information were excluded and lastly the SNP markers with minor allele frequency greater than or equal to 0.05 were retained.

Statistical analysis

Major Allele Frequency (MAF), gene diversity, and Polymorphic Information Content (PIC) values for all SNP markers, were calculated using Power Marker v.3.25 software [42].

The Structure v2.3.4 Software [43] was used for population structure analysis. The membership of each accession was run for value of $K = 1$ to $K = 10$ with the admixture model and correlated allele frequency. For each K , it was replicated 3 times. Each run was implemented with a burn-in period of 50,000 steps followed by 100,000 Monte Carlo Markov Chain replicates.

“Structure harvester” (<http://taylor0.biology.ucla.edu/structureHarvester/>) available online was used to calculate the final population structure. The likelihood value Ln (K) (mean + SD) method [44] was used. This method translates the mean of estimated Ln (Prob of data) values as a function of the probable number of subpopulations (K). It assumes that when K approaches the true value, Ln (K) (mean + SD) reaches a plateau or has an inconsequential variance [44]. Genotypes with membership probabilities greater than or equal to 0.7 were grouped together genotypes with membership probability lower than 0.7 were assigned to the admixture group. Genetic variation within the subpopulations was assessed using the fixation index (Fst) and average distances between individuals in same cluster using the Expected Heterozygosity (HE).

To assess genetic diversity, genetic distances across genotypes were calculated using Tassel v5.2.63 [38]. The dissimilarity matrix generated was used to construct Neighbor Joining tree. In addition, Principle Coordinate Analysis (PCoA) and Analysis of Molecular Variance (AMOVA) was performed, based on the Structure results and the differentiation genetic index (PhiPT and Nm) allowed to describe the genetic diversity between and among subpopulations and their degrees of differentiation. Also, the genetic diversity index such as mean values of number of loci with private alleles (Pa), number of different alleles (Na); Shannon’s diversity index (H) and the Shannon’s Information Index (I) were calculated. PCoA and AMOVA performed using GenAEx 6.503 [45] and SNP data was numerically coded as suggested in GenAEx user manual.

Results

Sequence data and SNP discovery

Genotyping by sequencing (GBS) of the 50 genotypes provided 18.7 Gb data containing 3.16 million sequence reads. The sequencing data was mapped on to the reference pigeonpea genome and the alignment sequences provided 71,712 SNPs across 50 genotypes. Considering the filtering criteria applied both to samples and SNP, six genotypes (kk7, kk8, kk19, kk29, kk35 and kk36) were excluded from further analysis and a total of 688 high quality SNPs were retained for further analysis. The number of SNPs per chromosome ranged from 14 on chromosome 5 to 133 on chromosome 2 (Fig 1A and S2 Table). The rate of polymorphism was 100% across the other 44 genotypes. The PIC varied from 0.12 to 0.55 with a mean value of 0.30 (Fig 1B). The gene diversity ranged from 0.13 to 0.62 with an average of 0.34 (Fig 1C). A total of 1,376 alleles could be identified. The Major Allele Frequency for all 688 SNPs ranged from 0.48 to 0.93 with an average of 0.80 (Fig 1D). About 2% of SNPs were found to be least informative with PIC and gene diversity values equaling 0.12 and 0.13 respectively and the Major Allele Frequency of 0.93.

Population structure in pigeonpea landraces

The likelihood value method used showed a first clear peak at $K = 2$ (Fig 2A and S3 Table). However, others larger values of K corresponding to a continuous gradual increase in the likelihood values were observed but the largest was $K = 5$. This value of K indicated that 5 subpopulations could exist in the analyzed collection. Nevertheless, the best value of K which clearly defined the number of subpopulations was $K = 4$. As a result, four subpopulations (pop1, pop2, pop3 and pop4) were identified. The subpopulation 1 (in red) grouped 8 genotypes, the subpopulation 2 (in green) grouped 17 genotypes, the subpopulation 3 (in blue) grouped 15 genotypes and the subpopulation 4 (in yellow) grouped 4 genotypes (Fig 2B). 18.18% of admixture supported the 4 subpopulations (S4 Table).

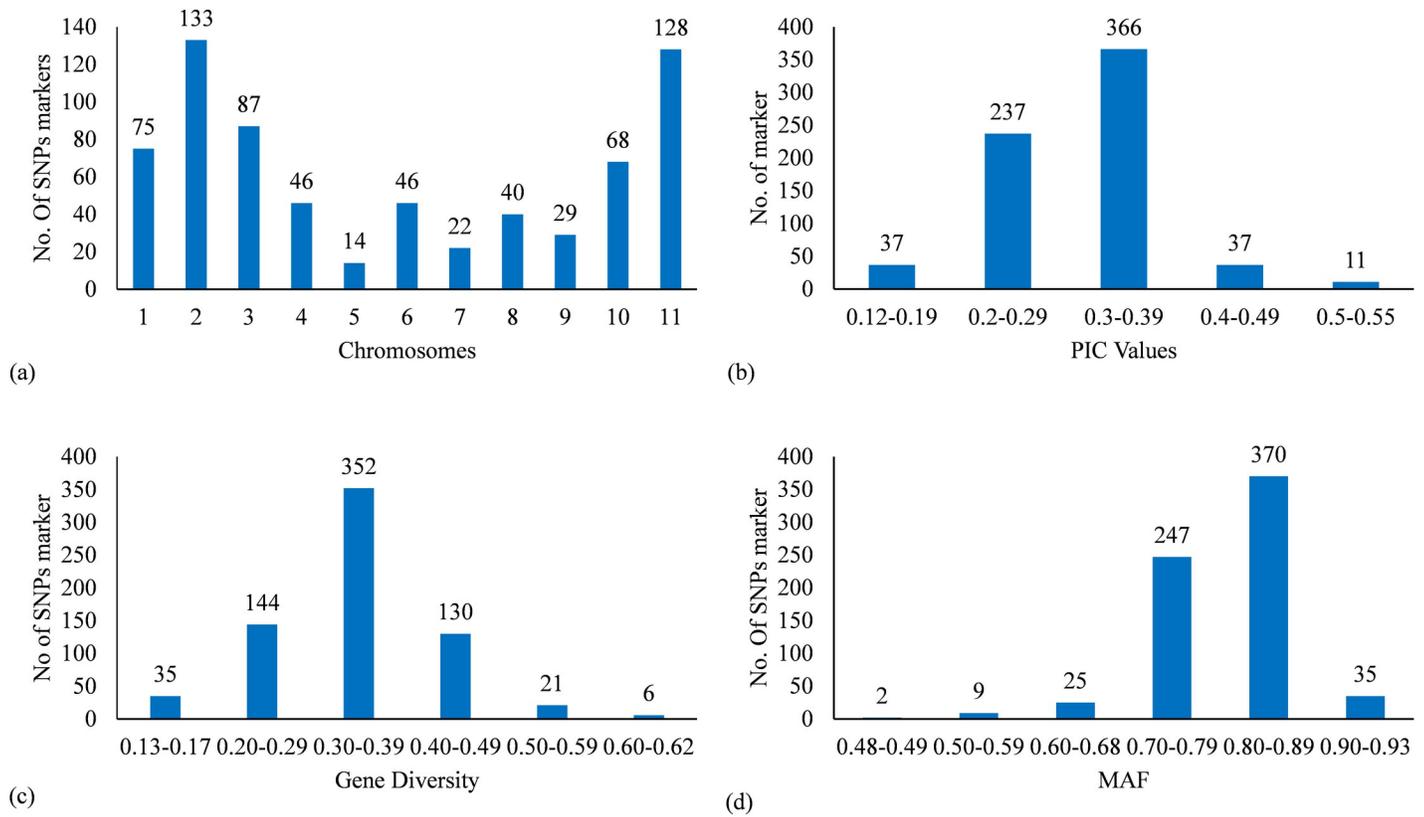


Fig 1. SNPs markers characteristics. (a) Distribution of SNPs across the 11 chromosomes. (b) Polymorphic Information Content (PIC). (c) Variation of Gene Diversity. (d) Frequency distribution of Major Allele Frequency.

<https://doi.org/10.1371/journal.pone.0271565.g001>

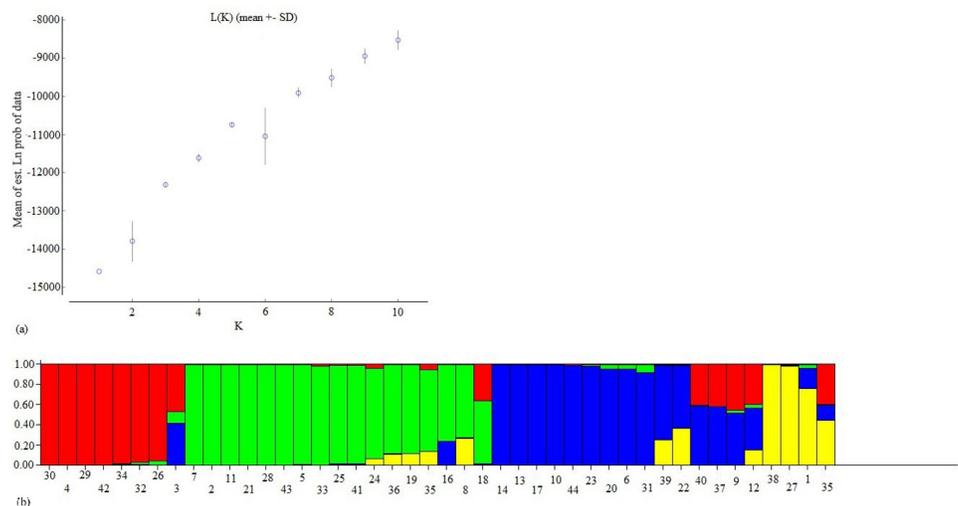


Fig 2. Estimated population structure of 44 pigeonpea genotypes. (a) Average of likelihood value $L(K)$ (mean + SD) against probable number of clusters (K). (b) Barplot of subpopulations at $K = 4$. Each genotype is represented by a vertical bar and partitioned into colored segments which represent the estimated probability membership in each of the K -inferred clusters on the y-axis. Bar colors (red, green, blue and yellow) indicate the groups identified through the structure program.

<https://doi.org/10.1371/journal.pone.0271565.g002>

Genetic diversity

The Neighbor Joining tree based on the genetic distances analysis showed four major clusters within all the 44 pigeonpea analyzed genotypes (S5 Table and Fig 3). Cluster 1 (in black) and cluster 2 (in blue) contained 4 and 18 genotypes respectively. Cluster 3 (in green) grouped 10 genotypes and Cluster 4 (in red) grouped 12 genotypes. PCoA was realized with all the 688 SNPs markers. Genotypes were labelled with four different colors according to structure based grouping of genotypes. The first, second and third PC accounted for 33.24% of the cumulative variation. The results did not agree with the structure results but showed an intermixing among genotypes from subpopulation 1, 2 and 3 (Fig 4).

Analysis of molecular variance and genetic diversity indices

The AMOVA results identified 7% of the total variation among subpopulation and 93% was observed within subpopulations. The differentiation genetic index (Φ_{PT} and N_m) were 0.07 and 7.13 respectively (Table 1). The mean value of different alleles (N_a) ranged from 1.38 in the subpopulation 1 to 1.57 in the subpopulation 2. The total number of alleles (N) ranged from 4 (in subpopulation 4) to 17 (in subpopulation 2). The percentage of polymorphic loci per population ranged from 37.97% in subpopulation 1 to 57.12% in subpopulation 2. Subpopulation 2 had the highest mean value of loci with private allele ($P_a = 0.11$) (Table 2). The number of private alleles ranged from 2 to 21 across the genotypes (Fig 5 and S6 Table). The highest Shannon Information index (I) and Shannon diversity index (H) were 0.28 and 0.20 respectively and hold by subpopulation 4 (Table 2). The mean values of F_{st} values for subpopulations pop1 pop2 pop3 and pop4 were 0.46; 0.48; 0.57 and 0.40 respectively. We observed a low average distance (H_E) between individuals in same cluster in subpopulation 3 (0.09) (Fig 6).

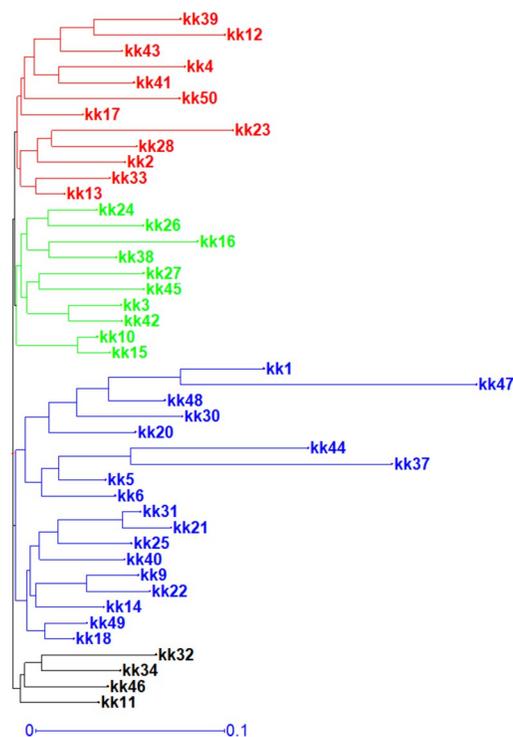


Fig 3. Neighbor—Joining tree of pairwise relatedness among 44 pigeonpea genotypes.

<https://doi.org/10.1371/journal.pone.0271565.g003>

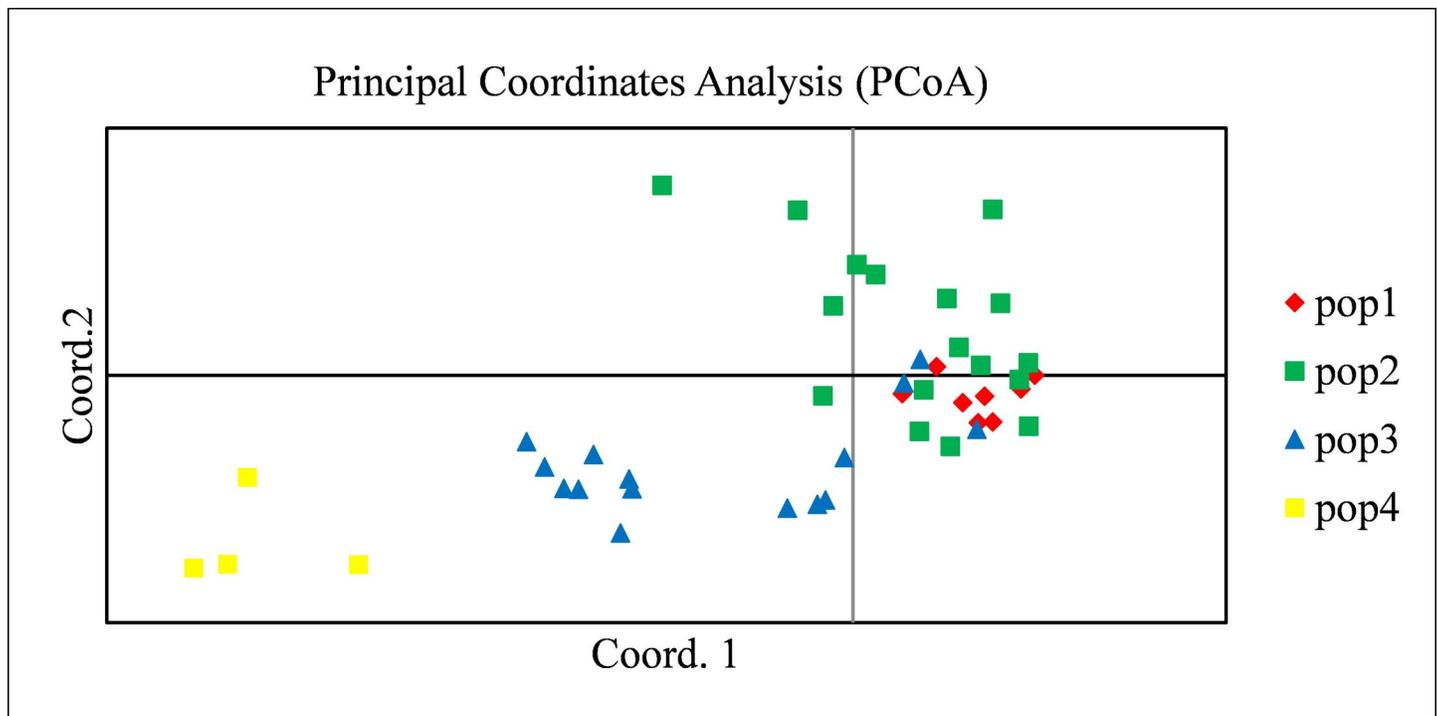


Fig 4. Principal Coordinate Analysis (PCoA) of 44 pigeonpea genotypes based on the population structure result.

<https://doi.org/10.1371/journal.pone.0271565.g004>

Discussion

The knowledge of genetic diversity is of paramount importance not only for the development of conservation strategies but also and more importantly the selection of parents to be used in genetic improvement [46–48]. In this study, we report the second genetic diversity analysis study of pigeonpea using Genotyping-by-sequencing approach in Benin. GBS approach provided 688 high-quality SNPs markers. The mean values of PIC and gene diversity were 0.30 and 0.34 respectively. These values were higher than those reported in Saxena et al. [49] which were 0.17 and 0.2 respectively for PIC and gene diversity across 21 pigeonpea landraces on the one hand and on the other hand 0.19, and 0.24 respectively across 56 pigeonpea elite cultivars. Previously in Benin, mean values of 0.25 and 0.30 for PIC and gene diversity respectively were reported on pigeonpea by using a set of 794 SNPs [37]. The difference between the present diversity index and those reported in Saxena et al. [49] could be explained through the radical certification of certain varieties of pigeonpea. Indeed, in the Indian context where there is selection followed by certification and commercialization of several varieties, the pigeonpea genetic diversity tends towards homogenization by an interesting alleles fixation which leads

Table 1. Analysis of molecular variance (AMOVA) of the genetic variation among and within the four subpopulations of *C. cajan* genotypes using the 688 SNPs.

Source	Df	SS	MS	Est.Var	%
Among subpopulations	3	578.89	192.96	7.90	7
Within subpopulations	40	4506.12	112.65	112.65	93
Total	43	5085.02		120.55	100

p value = 0.001. Df, Degrees of freedom; SS, Sum of Squares; MS, Mean Square; Est.Var, Estimated variance; Nm, Number of migrants.

<https://doi.org/10.1371/journal.pone.0271565.t001>

Table 2. Variation of different genetic parameters among the four identified subpopulations.

Indices de diversité	pop1	pop2	pop3	pop4
N	7.16 (0.03)	14.39 (0.06)	13.13 (0.05)	3.41 (0.03)
Na	1.38 (0.02)	1.57 (0.02)	1.50 (0.02)	1.45 (0.02)
I	0.19 (0.01)	0.20 (0.01)	0.18 (0.01)	0.28 (0.01)
Pa	0.06 (0.01)	0.11 (0.01)	0.08 (0.01)	0.04 (0.01)
H	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)	0.20 (0.01)
P	37.94%	57.12%	50.44%	44.91%

N, Total number of alleles; Na, Number of different alleles; H, Shannon's diversity index; I, Shannon's Information Index; P, Percentage of polymorphic loci; pop, Subpopulation. For each parameter, mean value is used followed by the Standard Error in brackets except the Percentage of polymorphic loci (P).

<https://doi.org/10.1371/journal.pone.0271565.t002>

to the reduction of its diversity [50]. Hence, compared with previous studies, the high level of index of diversity observed in the present study suggested that pigeonpea germplasm encompasses more genetic diversity than reported by previous studies in Benin [51]. Allelic diversity is known for its use as indicator of genetic variation [52]. For the set of 688 SNP markers, we could count 1 376 alleles. This non-negligible total allele's number is lower than that reported on pigeonpea by Zavinon et al. [37] who reported a total allele's number of 1 588 using a set of 794 SNP markers generated by GBS approach. For all genetic diversity index calculated, the observed difference with our results and those of previous studies could be explained by the difference between the genetic background of the accessions analyzed or between the number

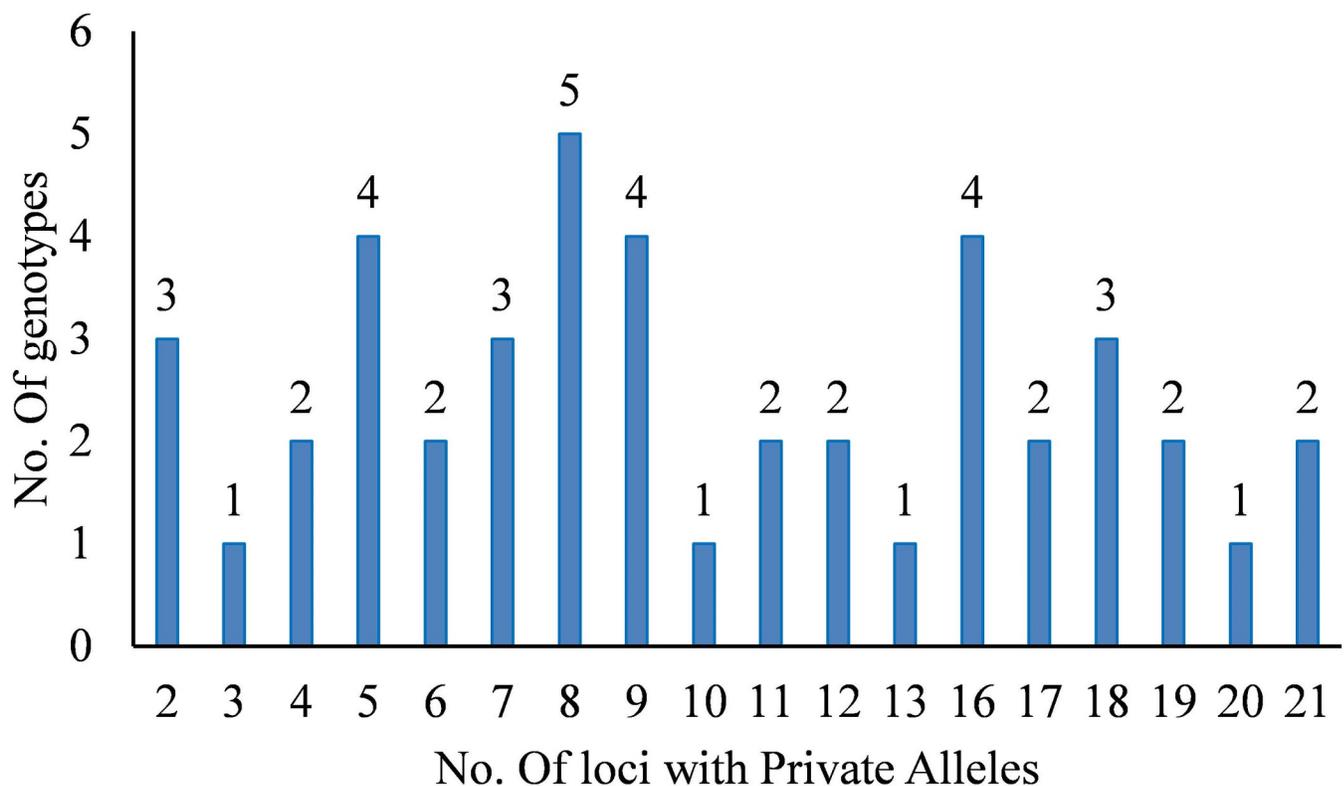


Fig 5. Variation of number of genotypes with private alleles within all the 44 pigeonpea analyzed genotypes.

<https://doi.org/10.1371/journal.pone.0271565.g005>

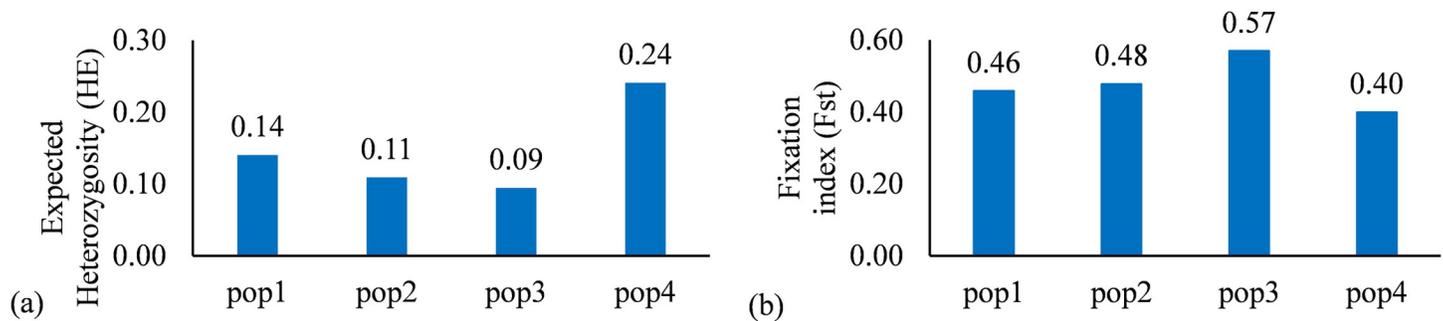


Fig 6. Genetic variation within the subpopulations. (a) Average distances (Expected Heterozygosity) between individuals in same cluster. (b) Genetic variation within the subpopulations (Fst).

<https://doi.org/10.1371/journal.pone.0271565.g006>

of markers involved in the different studies or the collection size [53–55]. All these findings reinforced the existence of a new structuration in pigeonpea genetic diversity in Benin.

Population structure analysis clarifies genetic diversity studies [8]. This study showed that Beninese's pigeonpea were divided into four clear subpopulations with a high genetic differentiation between these subpopulations as revealed by a low rate of admixed varieties (18.18%). The Weighted Neighbor Joining tree agreed structure results by giving similar result. The presence of structure in pigeonpea landraces has been reported early in South American by Sousa et al. [56], in Malawi by Njung'e et al. [14] and in Benin by Zavinon et al. [37]. This result supports the findings of Sousa et al. [56] among 77 tested pigeonpea. However, our results contradicted Zavinon et al. [37] who reported three subpopulations with high rate (20%) of admixed genotypes. The presence of a structure within pigeonpea as identified in the present study is of paramount importance. Thus, our results could lay the basis for Genome Wide Association Studies and consequently Marker-Assisted Selection to enhance genetic gain in pigeonpea breeding programs [8]. They could therefore facilitate genes of interest discovery, molecular breeding and allow rapid identification of heterogeneous groups for the development of hybrids with high agronomic performance [57]. Gene flow like selection and genetic drift plays a major role in shaping the genetic structure [58]. In this study, AMOVA results indicated a high level of genetic diversity (93%) within the four subpopulations whereas the rest of the variation (7%) was among subpopulations suggesting a high level of differentiation within subpopulations. The selection for specific agronomic traits by farmers could be the main reason for this high variation within subpopulations. The Nm value (7.13) observed was very high. Knowing that an Nm value less than 1 indicates limited gene flow among subpopulations [59], our result suggested that a high genetic exchange or high gene flow [8] may occur and confirmed that the genetic variation among subpopulations was low when compared to the second level of variation. This result coincided with the PCoA results for which most of the genotypes from subpopulations 1, 2 and 3 showed intermixing among themselves. These findings were in line with those of Kinhoégbè et al. [26] on pigeonpea seed system management in Benin. According to the authors, pigeonpea seed system is informal which favored gene flow through seeds exchanges [26]. They corroborated Radosavljevic et al., [58] who reported that selection and gene flow constituted the main factors of the population structure's dynamic over time. Hence, these two factors were the main factors that influenced Beninese pigeonpea structure. However, by relying on pigeonpea's strong tendency to autogamy [60] which is known to contribute to the low genetic variation [61], the low rate (18.18%) of admixed varieties as reported in this study justified the fact that the relatively great diversity obtained in Beninese pigeonpea was due to the selection based on specific traits rather than gene flow.

The diversity pattern revealed by the mean values of total number of alleles, number of loci with private alleles and percentage of polymorphic loci within subpopulations provided insight to genetic diversity within populations [62] indicated a signature in populations [57]. We observed that subpopulation 2 had the highest mean value of the total number of alleles. The loci with private alleles and the percentage of polymorphic loci followed the same trend while the genotypes from subpopulation 4 were the most highly differentiated ($I = 0.28$; $H = 0.20$) from those of the remaining populations. This is probably because subpopulation 2 had the largest number of genotypes indicating the high genetic diversity that existed among the genotypes [8]. As a result, the subpopulation 2 represented the main gene pool and could be serve as a source for diverse parents' selection useful to improve the existing landraces.

Conclusion

In this study, GBS approach was used to determine the level of genetic diversity, population structure of a collection of pigeonpea landraces grown in Benin. Results showed significant genetic diversity structured and maintained by the selection of cultivars based on specific traits. Four clear subpopulations were identified. The subpopulation 2 exhibited the highest mean values of total number of alleles, number of loci with private alleles and percentage of polymorphic loci. It represented the main gene pool in the analyzed collection and could contain desirable traits, such as biotic or abiotic stress tolerance. These findings laid the basis for Genome Wide Association Studies and consequently Marker-Assisted Selection to enhance genetic gain in pigeonpea breeding programs in Benin.

Supporting information

S1 Table. List of pigeonpea genotypes with corresponding codes and collection sites (District and Village).

(XLSX)

S2 Table. Genomic distribution of 688 SNPs mapped on the 11 pigeonpea chromosomes.

(XLSX)

S3 Table. Likelihood value raw summary.

(XLSX)

S4 Table. Estimated ancestry membership coefficients (Q) for genotypes.

(XLSX)

S5 Table. Pair-wise genetic distances matrix across 44 pigeonpea genotypes.

(XLSX)

S6 Table. List of genotypes with one or more private alleles.

(XLSX)

Acknowledgments

The authors are grateful to the Centre of Excellence in Genomics and Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) of India for their helpful assistance during the study. We thank the anonymous reviewers and editors for their comments on previous versions of the manuscript. We express our sincere gratitude to Doctors Ruth TAMADAHO and Aliu IBUKUN for English Language Editing. Our sincere grateful to Dr Vanika Garg from Murdoch University, Murdoch, Western Australia, Australia for their help when it has been to prepare and to upload Raw sequencing data in NCBI Sequence Read Archive.

Author Contributions

Conceptualization: Géofroy Kinhoégbè, Rajeev Kumar Varshney.

Data curation: Géofroy Kinhoégbè, Gustave Djèdatin, Rachit Kumar Saxena, Anu Chitikineni, Prasad Bajaj, Johiruddin Molla.

Formal analysis: Géofroy Kinhoégbè, Prasad Bajaj, Johiruddin Molla.

Methodology: Géofroy Kinhoégbè, Gustave Djèdatin, Rachit Kumar Saxena, Anu Chitikineni, Rajeev Kumar Varshney.

Resources: Géofroy Kinhoégbè, Rajeev Kumar Varshney.

Supervision: Gustave Djèdatin, Rachit Kumar Saxena, Clément Agbangla, Alexandre Dansi, Rajeev Kumar Varshney.

Validation: Clément Agbangla, Alexandre Dansi.

Visualization: Géofroy Kinhoégbè, Rajeev Kumar Varshney.

Writing – original draft: Géofroy Kinhoégbè.

Writing – review & editing: Géofroy Kinhoégbè.

References

1. 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(3):393–408. <https://doi.org/10.1007/s11032-009-9327-2> PMID: 20976284
2. 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*. 2012; 30(1): 83–89. <https://doi.org/10.1038/nbt.2022> PMID: 22057054
3. Upadhyaya KC and Maneesha. Analysis of genetic diversity in pigeon pea germplasm using retrotransposon-based molecular markers. *J Genet*. 2017; 96(4): 551–561. <https://doi.org/10.1007/s12041-017-0802-y> PMID: 28947703
4. Saxena KB, Kumar RV and Rao PV. Pigeonpea nutrition and its improvement. In: Basra AS, Randhawa IS, editors. *Quality Improvement in Field Crops*. New-York: Food Products Press; 2002. pp. 227–260. https://doi.org/10.1300/J144v05n01_10
5. Saxena KB and Sawargaonkar SL. Genetic enhancement of seed proteins in pigeonpea—methodologies, accomplishments, and opportunities. *Int J Sci Res*. 2015; 4(5): 3–7.
6. Dansi A, Vodouhe R, Azokpota P, Yedomonhan H, Assogba P, Adjatin A, et al. Diversity of the neglected and underutilized crop species of importance in Benin. *Sci World J*. 2012. <https://doi.org/10.1100/2012/932947> PMID: 22593712
7. Kinhoégbè G, Djèdatin G, Loko LEY, Favi AG, Adomou A, Agbangla C, et al. On-farm management and participatory evaluation of pigeonpea (*Cajanus cajan* [L.] Millspaugh) diversity across the agro-ecological zones of the Republic of Benin. *J Ethnobiology and Ethnomedicine*. 2020a; 16:24. <https://doi.org/10.1186/s13002-020-00378-0> PMID: 32404139
8. Eltaher S, Sallam A, Belamkar V, Emara HA, Nower AA, Salem KF M, et al. Genetic diversity and population structure of F3:6 nebraska winter wheat genotypes using genotyping-by-sequencing. *Front Genet*. 2018; 9:76. <https://doi.org/10.3389/fgene.2018.00076> PMID: 29593779
9. Govindaraj M, Rai KN, Kanatti A, Upadhyaya HD, Shivade H, Rao AS. Exploring the genetic variability and diversity of pearl millet core collection germplasm for grain nutritional traits improvement. *Sci Rep*. 2020; 10:21177. <https://doi.org/10.1038/s41598-020-77818-0> PMID: 33273504
10. Muniswamy S, Lokesh R, Dharmaraj PS, Yamanura JR. Morphological characterization and assessment of genetic diversity in minicore collection of pigeonpea (*Cajanus cajan* (L.) Millsp). *Eur J Pharm Biopharm*. 2014; 5(2):179–186.
11. Mohammadi SA and Prasanna BM. Analysis of Genetic Diversity in Crop Plants—Salient Statistical Tools and Considerations. *Crop Sci*. 2003; 43:1235–48. <https://doi.org/10.2135/cropsci2003.1235>
12. Sneller CH, Nelson RL, Carter TE, Cui Z. Genetic diversity in crop improvement. *J Crop Improv*. 2005; 14:103–144. https://doi.org/10.1300/J411v14n01_06

13. Varshney RK, Graner A and Sorrells ME. Genomics-assisted breeding for crop improvement. *Trends Plant Sci.* 2005; 10(12):621–630. <https://doi.org/10.1016/j.tplants.2005.10.004> PMID: 16290213
14. Njung'e V, Deshpande S, Siambi M, Jones R, Silim S and Villiers S. SSR genetic diversity assessment of popular pigeonpea varieties in Malawi reveals unique fingerprints. *Electron J Biotechnol.* 2016; 21:65–71. <https://doi.org/10.1016/j.ejbt.2016.02.004>
15. Bajracharya J, Steele KA, Jarvis DI, Sthapit BR, Witcombe JR. Rice landrace diversity in Nepal: Variability of agro-morphological traits and SSR markers in landraces from a high-altitude site. *Field Crops Res.* 2006; 95:327–335. <https://doi.org/10.1016/j.fcr.2005.04.014>
16. Mitu De. Use of Descriptor Codes in Agro-Morphological Characterization: Qualitative assessment of 20 Land Races of Rice (*Oryza sativa* L.) from West Bengal. *Int J Adv Life Sci.* 2019; 2(2): 21–26.
17. Upadhyaya HD, Pundir RPS, Gowda CLL, Reddy KN and Singh S. Geographical patterns of diversity for qualitative and quantitative traits in the pigeonpea germplasm collection. *Plant Genet Resour.* 2005; 3(3):331–352. <https://doi.org/https://doi.org/10.1079/PGR200587>
18. Silim NS, Bramel JP, Akonaay HB, Mligo JK, Christiansen JL. Cropping systems, Uses and primary in situ characterization of Tanzania pigeonpea [*Cajanus cajan* (L.) Millsp.] landraces. *Genet Resour Crop Evol.* 2005; 52:645–654. <https://doi.org/10.1007/s10722-003-6015-7>
19. Manyasa EO, Silim SN, Githiri SM and Christiansen JL. Diversity in Tanzanian pigeonpea (*Cajanus cajan* (L.) Millsp.) landraces and their response to environments. *Genet Resour Crop Evol.* 2008; 55(3): 379–387. <https://doi.org/10.1007/s10722-007-9245-2>
20. Manyasa EO, Silim SN and Christiansen JL. Variability patterns in Ugandan pigeonpea landraces. *J SAT Agric Res.* 2009; 7:1–9.
21. Gwata and Silim SN. Utilization of landraces for the genetic enhancement of pigeon pea in Eastern and Southern Africa. *J Food Agric Environ.* 2009; 7(2):803–806.
22. Vange T, Egbe MO. Studies on Genetic Characteristics of Pigeon Pea Germplasm at Otobi, Benue State of Nigeria. *World J Agric Sci.* 2009; 5:714–719.
23. Kundy AC, Mponda M, Mkandawile C, Mkamillo G. Yield evaluation of eighteen pigeonpea (*Cajanus cajan* (L.) Millsp.) genotypes in South Eastern Tanzania. *European Journal of Physical and Agricultural Sciences.* 2015; 3(2):9–15.
24. Quenum FJB, Djaboutou MC, Houedjissin SS, Sinha MG, Doko R, Cacaï GH, et al. Diagnosis of production in support of the evaluation of the pigeon pea seeds quality (*Cajanus cajan* (L.) Millsp.) in Benin. *Bulletin de la Recherche Agronomique du Bénin.* 2016; 80:34–46.
25. Zavinon F, Adoukonou-Sagbadja H, Bossikponnon A, Dossa H and Ahanhanzo C. Phenotypic diversity for agro-morphological traits in pigeon pea landraces [(*Cajanus cajan* L.) Millsp.] cultivated in southern Benin. *Open Agric.* 2019; 1(4):487–499. <https://doi.org/10.1515/opag-2019-0046>
26. Kinhoégbè G, Djèdatin G, Loko LEY, Agbo RI, Saxena RK, Varshney RK, et al. Agro-morphological characterization of pigeonpea (*Cajanus cajan* L. Millspaugh) landraces grown in Benin: Implications for breeding and conservation. *J Plant Breed Crop Sci.* 2020b; 12(1):34–49. <https://doi.org/10.5897/JPCS2019.0836>
27. Sharma S, Dastagiri MB, Reddy MN. Morphological Variation and Evaluation of Gladiolus (*Gladiolus Hybridus hort.*) Cultivars. *J Hortic.* 2017; 4:4. <https://doi.org/10.4172/2376-0354.1000212>
28. Mishra KK, Fougat RS, Ballani A, Vinita T, Yachana J, Madhumati B. Potential and application of molecular markers techniques for plant genome analysis. *Indian J Pure Appl Biosci.* 2014; 2:169–188.
29. Sarkar B, Chakravarthy VSK, Varalaxmi Y, Yadav SK, Vanaja M and Maheswari M. Genetic Diversity among Pigeonpea (*Cajanus cajan* L. Millsp.) Genotypes Using Genic SSRs with Putative Function for Drought Tolerance. *Int J Curr Microbiol Appl Sci.* 2017; 6(4):1804–1814. <https://doi.org/10.20546/ijcmas.2017.604.216>
30. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *Plos One.* 2008; 3(10):e3376. <https://doi.org/10.1371/journal.pone.0003376> PMID: 18852878
31. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *Plos One.* 2011; 6(5):e19379. <https://doi.org/10.1371/journal.pone.0019379> PMID: 21573248
32. Poland JA and Rife TW. Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome.* 2012; 5(3):92–102. <https://doi.org/10.3835/plantgenome2012.05.0005>
33. Jaganathan D, Thudi M, Kale S, Azam S, Roorkiwal M, Gaur PM, et al. Genotyping-by-sequencing based intra-specific genetic map refines a "QTL-hotspot" region for drought tolerance in chickpea. *Mol Genet Genom.* 2015; 290(2):559–571. <https://doi.org/10.1007/s00438-014-0932-3> PMID: 25344290

34. Hart JP and Griffiths PD. Genotyping-by-sequencing enabled mapping and marker development for the by-2 potyvirus resistance allele in *Common bean*. *Plant Genome*. 2015; 8(1). <https://doi.org/10.3835/plantgenome2014.09.0058> PMID: 33228290
35. Gao L, Kielsmeier-Cook J, Bajgain P, Zhang X, Chao S, Rouse MN, et al. Development of genotyping by sequencing (GBS) and array-derived SNP markers for stem rust resistance gene Sr42. *Mol Breed*. 2015; 35:1–13. <https://doi.org/10.1007/s11032-015-0404-4>
36. Saxena RK, Kale SM, Kumar V, Parupalli S, Joshi S, Singh VK, et al. Genotyping by Sequencing of three mapping populations for identification of candidate genomic regions for resistance to sterility mosaic disease in pigeonpea. *Sci Rep*. 2017; 7(1):1813. <https://doi.org/10.1038/s41598-017-01535-4> PMID: 28500330
37. Zavinon F, Adoukonou-Sagbadja H, Keilwagen J, Lehnert H, Ordon F, Perovic. Genetic diversity and population structure in Beninese pigeon pea [*Cajanus cajan* (L.) Huth] landraces collection revealed by SSR and genome wide SNP markers. *Genet Resour Crop Evol*. 2020; 67: 191–208. <https://doi.org/10.1007/s10722-019-00864-9>
38. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y and Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007; 23(19): 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308> PMID: 17586829
39. Langmead B and Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9(4): 357–359. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
40. Zhang X, Pérez-Rodríguez P, Semagn K, Beyene Y, Babu R, López-Cruz MA, et al. Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. *Heredity* (Edinb). 2015; 114(3): 291–299; <https://doi.org/10.1038/hdy.2014.99> PMID: 25407079
41. Hussain W, Baenziger PS, Belamkar V, Guttieri MJ, Venegas JP, Easterly A, et al. Genotyping-by-sequencing derived high-density linkage map and its application to QTL mapping of flag leaf traits in bread wheat. *Sci Rep*. 2017; 7(1):16394. <https://doi.org/10.1038/s41598-017-16006-z> PMID: 29180623
42. Liu K and Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 2005; 21(9):2128–2129. <https://doi.org/10.1093/bioinformatics/bti282> PMID: 15705655
43. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155(2):945–959. <https://doi.org/10.1093/genetics/155.2.945> PMID: 10835412
44. Rosenberg NA, Burke T, Eloe K, Feldman MW, Freidlin PJ, Groenen MA, et al. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics*. 2001; 159(2): 699–713. <https://doi.org/10.1093/genetics/159.2.699> PMID: 11606545
45. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*. 2012; 28: 2537–9. <https://doi.org/10.1093/bioinformatics/bts460> PMID: 22820204
46. Govindaraj M, Vetriventhan M, Srinivasan M. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet Res Int*. 2015; 2015:431487. <https://doi.org/10.1155/2015/431487> PMID: 25874132
47. Carvalho M, Muñoz-Amatriaín M, Castro I, Lino-Neto T, Matos M, Egea-Cortines M, et al. Genetic diversity and structure of Iberian Peninsula cowpeas compared to world-wide cowpea genotypes using high density SNP markers. *BMC Genomic*. 2017; 18:891. <https://doi.org/10.1186/s12864-017-4295-0> PMID: 29162034
48. Rufo R, Alvaro F, Royo C and Soriano JM. From landraces to improved cultivars: assessment of genetic diversity and population structure of Mediterranean wheat using SNP markers. *Plos One*. 2019; 14(7): e0219867. <https://doi.org/10.1371/journal.pone.0219867> PMID: 31306459
49. Saxena RK, von Wettberg E, Upadhyaya HD, Sanchez V, Songok S, Saxena K, et al. Genetic Diversity and Demographic History of *Cajanus* spp. Illustrated from Genome-Wide SNPs. *PLoS One*. 2014; 9(2): e88568; <https://doi.org/10.1371/journal.pone.0088568> PMID: 24533111
50. Louwaars NP. Plant breeding and diversity: A troubled relationship? *Euphytica*. 2018; 214: 114. <https://doi.org/10.1007/s10681-018-2192-5> PMID: 30996394
51. Niu S, Song Q, Koiwa H, Qiao D, Zhao D, Chen Z, et al. Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (*Camellia sinensis*) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. *BMC Plant Biol*. 2019; 19(1):328. <https://doi.org/10.1186/s12870-019-1917-5> PMID: 31337341
52. Ren X, Jiang H, Yan Z, Chen Y, Zhou X, Huang L, et al. Genetic diversity and population structure of the major peanut (*Arachis hypogaea* L.) cultivars grown in China by SSR markers. *Plos One*. 2014; 9(2): e88091. <https://doi.org/10.1371/journal.pone.0088091> PMID: 24520347

53. Xia XC, Reif JC, Hoisington DA, Melchinger AE, Frisch M, Warburton ML. Genetic diversity among CIM-MYT maize inbred lines investigated with SSR markers: I. Lowland tropical maize. *Crop Sci.* 2004; 44(6): 2230–2237. <https://doi.org/10.2135/cropsci2004.2230>
54. Yan J, Yang X, Shah T, Sánchez-Villeda H, Li J, Warburton M, et al. High-throughput SNP genotyping with the Goldengate assay in Maize. *Mol Breed.* 2010; 25(3): 441–451.
55. Adu BG, Badu-Apraku B, Akromah R, Garcia-Oliveira LA, Awuku FJ, Gedil M. Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. *Plos One.* 2019; 14(4):e0214810. <https://doi.org/10.1371/journal.pone.0214810> PMID: 30964890
56. Sousa ACB, Godoy R, Sforça DA, de Campos T, Zucchi MI, Jank L et al. Genetic diversity analysis among pigeon pea genotypes adapted to South American regions based on microsatellites markers. *Sci Agric.* 2011; 68(4): 431–439. <https://doi.org/10.1590/S0103-90162011000400007>
57. Serba DD, Muleta KT, St Amand P, Bernardo A, Bai G, Perumal R, et al. Genetic Diversity, Population Structure, and Linkage Disequilibrium of Pearl Millet. *Plant Genome.* 2019; 12(3): 1–12. <https://doi.org/10.3835/plantgenome2018.11.0091> PMID: 33016587
58. Radosavljevic I, Satovic Z and Liber Z. Causes and consequences of contrasting genetic structure in sympatrically growing and closely related species. *AoB Plants.* 2015; 7: plv106. <https://doi.org/10.1093/aobpla/plv106> PMID: 26333826
59. Arora A, Kundu S, Dilbaghi N, Sharma I, Tiwari R. Population structure and genetic diversity among Indian wheat varieties using microsatellite (SSR) markers. *Aust J Crop Sci.* 2014; 8:1281–1289.
60. Saxena K, Bohra A, Choudhary KA, Sultana R, Sharma M, Pazhamala TL, et al. The alternative breeding approaches for improving yield gains and stress response in pigeonpea (*Cajanus cajan*). *Plant Breed.* 2020; 140(1):74–86. <https://doi.org/10.1111/pbr.12863>
61. Gong W, Gu L, Zhang D. Low genetic diversity and high genetic divergence caused by inbreeding and geographical isolation in the populations of endangered species *Loropetalum subcordatum* (Hamamelidaceae) endemic to China. *Conserv Genet.* 2010; 11:2281–2288. <https://doi.org/10.1007/s10592-010-0113-9>
62. Luo Z, Brock J, Dyer JM, Kutchan T, Schachtman D, Augustin M, et al. Genetic Diversity and Population Structure of a *Camelina sativa* Spring Panel. *Front Plant Sci.* 2019; 10:184. <https://doi.org/10.3389/fpls.2019.00184> PMID: 30842785