INVESTIGATION OF GENETIC DIVERSITY AMONG COWPEA ACCESSIONS USING DIVERSITY ARRAYS TECHNOLOGY SEQUENCING (DARTSEQ) TECHNIQUE

G.V. Nkomo[1, 2], M.M. Sedibe[1], M.A. Mofokeng[2], A.B.A. Assefa[3]

[1]Central University of Technology, Department of Agriculture, Bloemfontein, Republic of South Africa
[2]Agriculture Research Council, Grain Crops Institute, Potchefstroom, Republic of South Africa
[3]Agriculture Research Council, Vegetable and Ornamental Plant Institute, Pretoria, Republic of South Africa

A total of 85 cowpea accessions were investigated for genetic diversity and population structure using 18 284 single nucleotide polymorphisms (SNPs). In this study, only 51 % of the SNPs were polymorphic across the 85 accessions. The genetic distance, estimated based on Nei's index among genotypes, ranged from 0.14 to 0.44, with a mean value of 0.35. The polymorphism information content ranged from 0.024 to 0.50, with a mean value of 0.25. Twenty-six percent of the SNPs had genetic diversity values greater than 0.40, suggesting that the genotypes were highly genetically diversified. A high gene flow of 4.89 was observed between Zimbabwean and South African accessions, indicating a high germplasm exchange among these neighbouring countries. The analysis of molecular variance revealed a highly significant variation ($P < 0.001$) among individual accessions and low variation within individuals. The accessions showed significant ($P < 0.001$) levels of differentiation among geographic regions. Cluster analysis of the 85 accessions generated by the unweighted pair group method with arithmetic mean (UPGMA) clustered the accessions into three groups. Accessions with great potential can be selected and further improved for cultivation.

DArTseq, genetic diversity, single nucleotide polymorphism, *Vigna unguiculata* (L.) Walp.

doi: 10.7160/sab.2022.530102
Received for publication on August 11, 2020
Accepted for publication on January 19, 2022

INTRODUCTION

Cowpea [*Vigna unguiculata* (L.) Walp.] is a primary source of protein in sub-Saharan Africa. It can be consumed as fresh leaves or dry grains and foliage, and also forage (Munoz-Amatriain et al., 2016). The ability of cowpea to produce stable yields under abiotic stress conditions (drought, heat, low soil fertility) and to replenish nitrogen back into the soil (Muchero et al., 2009) makes it an important crop in low-input farming systems. The yield of cowpea is generally low in developing countries because of the lack of improved cultivars, low fertiliser use, and poor management practices (Boahen et al., 2017).

The recent advances in molecular biology have made characterisation of the germplasm, genetic mapping and development of quantitative trait loci (QTL) easier (Boukar et al., 2018) in different cowpea research programmes. In cowpeas, different marker technologies have been used, such as restriction fragment length polymorphism (RFLP) (Boukar et al., 2016), amplified fragment length polymorphism (AFLP) (Kolade et al., 2016), simple sequence repeat (SSR) (Wamalwa et al., 2016), random amplified polymorphic DNA (RAPD) (Damarany et al., 2018), and single nucleotide polymorphisms (SNPs) (Chidebe et al., 2018). Boukar et al. (2018) indicated that molecular marker techniques are now being widely used in genetic diversity, variety identification, phylogenetic analysis, gene mapping, and resource classification. Damarany et al. (2018) used RAPD assays to identify DNA markers in seven cultivars of cowpea and also to evaluate the proportions of genetic similarities among the cultivars. Depending

on the condition, positive and negative markers can confer an advantage to the host, but also inhibit growth (Damarany et al., 2018). In cowpeas, a total of 25 simple sequence repeat (SSR) primers were identified and used to differentiate a new cowpea variety, VBN 3, from other cowpea varieties, such as Vamban 1 and CO (CP) 7 (Ragul et al., 2018).

In molecular biology, next generation sequencing (NGS) technologies have a great impact on crops regarding the analyses of genetic diversity in population, gene, and quantitative trait loci (QTL) mapping. Genetic maps are a basis for QTL and gene mapping, marker-assisted selection, and the assembly of genome sequences (Huang et al., 2018). The first genetic map of cowpeas was constructed using RFLP markers based on individuals derived from a cross between cultivars IT 2246-4 and TVNI 963 (Fatokun et al., 1992). Since then, many genetic maps have been developed using various molecular marker techniques.

Diversity Arrays Technology (DArTseq) is a new hybridisation method that combines DArT complexity reduction methods with a next generation sequencing (NGS) platform (Nemli et al., 2017). Also a DArTseq complexity reduction approach can be used in combination with Illumina short-read sequencing (Hiseq 2000) to enable rapid and accurate sequencing (Kilian et al., 2016). DArTseq has been developed to generate high-density data, scoring thousands of unique genomic-wide DNA fragments in a single low-cost experiment (Kilian et al., 2016). Nemli et al. (2017) identified 43 018 SNPs from 173 common bean accessions of Andean and Mesoamerican origin using SNPs detected by a DArTseq approach.

DArTseq-derived SNPs are now widely used in many genetic diversity studies compared to other markers such as AFLP and SSR (Varshney et al., 2007) because SNPs are abundant in the genomes of plants and other organisms (Deulvot et al., 2010). DArTseq is a high-throughput, highly reproducible, and low-cost microarray hybridisation technology, with no previous sequence information for the detection of loci for a trait of interest (Nadeem et al., 2017). A single reaction assay of as little as 50–100 ng of genomic DNA can genotype several thousand genomic loci (Nadeem et al., 2017). A subset of 298 lines from a mini core collection of 370 landraces was genotyped based on a genotyping-by-sequencing (GBS) assay to assess the genetic diversity of the lines using three different methods of cluster analyses (Fatokun et al., 2018). Xiong et al. (2016) analysed the genetic diversity of cowpea and estimated the population structure of 768 cultivated cowpea genotype collections obtained from the Germplasm Resource Information Network (GRIN), which were originally collected from 56 countries. Three well-differentiated genetic populations were postulated from 768 worldwide cowpea collections based on the model-based ancestry, phylogenetic tree, and principal component analyses. Munoz-Amatriain et al. (2016) developed bacterial artificial chromosome (BAC) libraries and a BAC-based physical map, assembled sequences from 4 355 BACs, as well as a whole-genome shotgun (WGS) assembly using the African cultivar IT97K-499-35. The WGS sequences of further 36 different cowpea accessions developed a genotyping assay for over 50 000 SNPs, which was then applied to five bi-parental recombinant inbred line (RIL) populations to produce a consensus genetic map containing 37 372 SNPs (Munoz-Amatriain et al., 2016). The objectives of this study were to assess the genetic diversity and to examine the population structure of 85 cowpea genotypes collected from different geographic origins using DArTseq genotype by sequencing techniques. These cowpea accessions are commonly grown in Africa hence there were chosen for the study.

Assessment of the genetic diversity within a crop's germplasm is fundamental for crop improvement and selection**.** In South Africa**,** most of studies on cowpea were conducted using very limited number of genotypes, less number of quantitative traits, and single location experiments. Besides, there is very limited information on genetic diversity of South African cowpea accessions by using multivariate analyses. Therefore, the objective of the present study were to determine the extent of and pattern of genetic diversity among 85 cowpea accessions based on phenotypic traits and using multivariate analyses.

MATERIAL AND METHODS

Plant material

A total of 85 cowpea accessions collected from three geographic regions were used in this study, of which 45 accessions were obtained from the International Institute of Tropical Agriculture (IITA) in Nigeria, 25 accessions were from the Agricultural Research Council, Grain Crops in South Africa, and 15 accessions were from smallholder farmers in Buhera District in Zimbabwe (Table 1).

DNA extraction, sequencing, and SNP calling

Seeds of cowpea accessions were planted in 20 cm diameter pots in topsoil mixed with compost (3 : 1) in a greenhouse at the Agriculture Research Council, Potchefstroom, South Africa. At the three-leaf stage, young, fresh, and succulent leaves were harvested from each accession. The leaf samples were excised and freeze dried for three days using a VirTis freeze dryer (SP Scientific, USA). Leaf samples were sealed in a zip-lock bag labelled with the corresponding genotypic code and sent to the Integrated Genotyping Service and Support (IGSS), Biosciences eastern and

Table 1. List of cowpea accessions used in this study obtained from three geographic origins

No.	Cultivar Name	Code	Source	Origin
1	IT99K-573-2-1	GH01	IITA	Nigeria
2	VAM14-143-4-1	GH02	IITA	Nigeria
3	IT07K-292-10	GH03	IITA	Nigeria
4	IT07-318-33	GH16	IITA	Nigeria
5	IT97-499-35	GH17	IITA	Nigeria
6	TVU3000	GH18	IITA	Nigeria
7	IT07K-274-2-9	GH19	IITA	Nigeria
8	TVU 13932	GH08	IITA	Nigeria
9	IT90K-59	GH20	IITA	Nigeria
10	IT89KD-288	GH21	IITA	Nigeria
11	98K-476-8	GH25	IITA	Nigeria
12	TVU9443	GH28	IITA	Nigeria
13	Oleyin	GH29	IITA	Nigeria
14	TVU11986	GH30	IITA	Nigeria
15	IT90K-76	GH34	IITA	Nigeria
16	IT93K-452-1	GH35	IITA	Nigeria
17	IT90K-277-2	GH36	IITA	Nigeria
18	IT08K-150-27	GH37	IITA	Nigeria
19	IT96D-610	GH38	IITA	Nigeria
20	IT90K-207-15	GH39	IITA	Nigeria
21	TVU14190	GH43	IITA	Nigeria
22	98K-503-1	GH45	IITA	Nigeria
23	TVU 9620	GH49	IITA	Nigeria
24	TVU13004	GH50	IITA	Nigeria
25	97K-499-35	GH51	IITA	Nigeria
26	TVU12637	GH52	IITA	Nigeria
27	IT93K-129-4	GH53	IITA	Nigeria
28	IT96D-610	GH54	IITA	Nigeria
29	TVU12746	GH58	IITA	Nigeria
30	TVU9596	GH60	IITA	Nigeria
31	TVU2095	GH61	IITA	Nigeria
32	TVU3416	GH62	IITA	Nigeria
34	98D-1399	GH64	IITA	Nigeria
35	86D-1010	GH66	IITA	Nigeria
36	90K-284-2	GH67	IITA	Nigeria
37	99K-494-6	GH70	IITA	Nigeria
38	TVU13778	GH72	IITA	Nigeria
39	83S-911	GH73	IITA	Zimbabwe
40	TVU9671	GH74	IITA	Nigeria
41	Orelo	GH76	IITA	Nigeria
42	95K-589-2	GH79	Buhera	Nigeria
43	IT98K-506-1	GH83	IITA	Nigeria
44	IT82E-18	GH84	IITA	Nigeria
45	99K-494-6	GH89	IITA	Nigeria
46	RV 344	GH07	ARC	South Africa
47	RV207	GH09	ARC	South Africa
48	RV555	GH10	ARC	South Africa
49	Bechuana White	GH11	ARC	South Africa
50	RV503	GH44	ARC	South Africa
51	RV568	GH12	ARC	South Africa
52	RV194	GH26	ARC	South Africa
53	Dr Saunders	GH27	ARC	South Africa
54	RV500	GH68	ARC	South Africa
55	RV351	GH69	ARC	South Africa
56	Encore	GH71	ARC	South Africa
57	RV558	GH75	ARC	South Africa
58	TVU12746	GH75	ARC	South Africa
59	RV202	GH77	ARC	South Africa
60	RV342	GH81	ARC	South Africa
61	RV204	GH82	ARC	South Africa
62	RV213	GH85	ARC	South Africa
63	PAN311	GH86	ARC	South Africa
64	RV221	GH87	ARC	South Africa
65	RV551	GH46	ARC	South Africa
66	RV553	GH47	ARC	South Africa
67	RV554	GH48	ARC	South Africa
68	CH47	GH55	ARC	South Africa
69	RV343	GH57	ARC	South Africa
70	RV574	GH65	ARC	South Africa
71	Dahwa	GH04	Buhera	Zimbabwe
72	Chibundi chitsvuku	GH05	Buhera	Zimbabwe
73	CBC1	GH06	Buhera	Zimbabwe
74	Barapara purple	GH13	Buhera	Zimbabwe
75	Mupengo dema	GH14	Buhera	Zimbabwe
77	Chibundi chemavara	GH15	Buhera	Zimbabwe
78	Barapara jena	GH22	Buhera	Zimbabwe
79	CBC2	GH23	Buhera	Zimbabwe
80	Mupengo wemavara	GH24	Buhera	Zimbabwe
81	Ziso dema	GH31	Buhera	Zimbabwe
82	Barapara remavara	GH32	Buhera	Zimbabwe
83	IT18	GH40	Buhera	Zimbabwe
84	Mutonono	GH41	Buhera	Zimbabwe
85	Zvenyika	GH42	Buhera	Zimbabwe

IITA = International Institute of Tropical Agriculture, ARC = Agriculture Research Council

central Africa Hub – International Livestock Research Institute (BecA-ILRI), Nairobi, Kenya, for sequencing and SNP analysis.

The DArTseq technique was used to evaluate the genetic diversity of 85 cowpea accessions collected from Nigeria, South Africa, and Zimbabwe. DNA extractions and sequencing were performed using the DArTseq protocol (Diversity Arrays Technology Pty Ltd., Australia). About 1 g of young leaf tissue from each accession was used for genomic DNA extraction. Genomic DNA was isolated from the dried leaves using a modified cetyltrimethyl ammonium bromide (CTAB)/chloroform/isoamyl alcohol method (Doyle, Doyle, 1987). The dried leaf tissue was ground and mixed with 2% pre-warmed (60 °C) CTAB isolation buffer of 1.4 M NaCl, 100 mM Tris (pH 8.0), and 20 mM EDTA (Sigma, USA). The mixture was then transferred to a 2 ml microcentrifuge tube and incubated at 60 °C for 1 h. DNA was extracted once with chloroform/isoamyl alcohol (Chl/IAA; 24 : 1) (Sigma) and precipitated with two volumes of isopropanol. The obtained pellet was washed with 70% EtOH, dried, and dissolved in 100 µl of TE buffer with 50 µg ml^{-1} RNase A (Sigma). The extracted DNA was quantified by 0.8% agarose gel electrophoresis, and was adjusted to 50 ng μl^{-1} for DArT and SNP genotyping.

DArT analysis

DNA was processed in digestion/ligation reactions, as reported by Kilian et al. (2012) by replacing a single *Pst*I-compatible adaptor with two different adaptors corresponding to two different restriction enzyme (RE; *Pst*I and *Sph*I) compatible adaptors. The *Pst*I-compatible adapter was designed to incorporate an Illumina flow cell attachment sequence with staggered sequences of varying length barcode region, similar to the sequence reported by Elshire et al. (2011). The reverse adaptor contained a flow cell attachment region with *Sph*I-compatible overhang sequence. Only 'mixed fragments' (*Pst*I–*Sph*I) were effectively amplified in 30 rounds of PCR using the following reaction conditions: 94 °C for 1 min, then 30 cycles of 94 °C for 20 s, 58 °C for 30 s, 72 °C for 45 s, and 72 °C for 7 min. After PCR, equimolar amounts of amplification products from each sample of the 96-well microtiter plate were bulked and applied to c Bot (Illumina) bridge PCR followed by sequencing on the Illumina Hiseq 2500 system (Illumina, USA). The sequencing (single read) was run for 77 cycles.

Sequences generated from every lane were processed using proprietary DArT analytical pipelines (PLs). In the primary pipeline analysis, fragments of poor-quality sequences with reproducibility below 90 % and read depths lower than 3.5 for SNPs or 5 for presence–absence markers were filtered out. More stringent selection criteria were applied to the barcode region compared to the rest of the sequences. The assignments of the sequences to specific samples carried within the barcode splitting step were very reliable. No samples were dropped because of low coverage across loci; however, individual sequences were removed if they did not meet the above criteria. Approximately 2.5 million sequences per barcode per sample were identified and used in marker calling. The average browsing depth across loci was 9.2 reads per individual per locus for reference alleles and 6.5 for SNP alleles. Finally, identical sequences were collapsed into 'fastqcoll' files. The fastqcoll files were groomed using DArT PL's proprietary algorithm, which corrects low-quality bases from singleton tags into correct bases using collapsed tags with multiple members as a template. The groomed fastqcoll files were utilised in the secondary pipeline for DArT PL's proprietary SNP and SilicoDArT (presence/absence of restriction fragments in representation; PA markers) calling algorithms using DArTsoft14 (Diversity Arrays Technology Pty Ltd.).

SNP calling was performed for all tags from all libraries enclosed within the DArTsoft14 analysis clustered using DArT PL's C^{++} algorithm program at a brink distance of three. Parsing of the clusters into separate SNP loci was performed using a technique called balance of read counts for the allelic pairs. Additional choice criteria were further added to the algorithm program supported by an analysis of roughly 1 000 controlled cross populations. Testing for deviations from the Hardy–Weinberg equilibrium of alleles in these populations was conducted to facilitate the selection of technical parameters to effectively discriminate true allelic variants from paralogous sequences. In addition, multiple samples were processed from DNA to allelic calls as technical replicates, and scoring consistency was used as the main selection criteria for high-quality/low-error rate markers. Calling quality was assured by a high average browse depth per locus (average across all markers was over 30 reads per locus).

Data analysis

Null alleles (those that failed to amplify any fragment), monomorphic SNPs, SNPs with minor allele frequencies of less than 2 %, and SNPs that had missing alleles for more than 20 % of the genotypes were filtered out. Genotypic data were subjected to analyses with various measures of genetic diversity within and among genotypes using GenAlex software version 6.5 (Peakall, Smouse, 2012). Genetic diversity parameters, such as observed heterozygosity (H_o), Shannon Information Index, gene diversity (H_e), and polymorphic information content (PIC) were determined using the protocol of Nei and Li (Nei, Li, 1979). Based on the geographic stratification, genetic diversity analysis within and among populations and

Table 2. Genetic diversity within and among 85 cowpea genotypes based on 7 799 SNPs markers

Chromosome	No. of SNPs used	Polymorphic SNPs	% P	H_o	H_e	F_{IS}	PIC
1	1155	465	40.26	0.050	0.260	0.807	0.259
2	1145	647	56.51	0.034	0.222	0.867	0.217
3	2169	940	43.34	0.048	0.268	0.808	0.267
4	1606	731	45.52	0.049	0.252	0.804	0.251
5	1452	568	39.12	0.052	0.235	0.781	0.233
6	1425	669	46.95	0.043	0.261	0.823	0.259
7	1760	852	48.41	0.045	0.265	0.808	0.264
8	1326	602	45.40	0.056	0.253	0.781	0.251
9	1367	566	41.40	0.049	0.234	0.770	0.233
10	1531	671	43.83	0.061	0.286	0.789	0.284
11	1542	640	41.50	0.055	0.243	0.778	0.242
Unknown	1770	448	25.25	0.069	0.273	0.730	0.272
All chromosomes	1520	650	42.73	0.050	0.255	0.798	0.254
SE	–	–	–	0.001	0.002	0.003	0.002

SNPs = single nucleotide polymorphisms, % P = percentage polymorphic markers per chromosome, H_o = observed gene diversity within genotypes, H_e = average gene diversity within genotypes, F_{IS} = inbreeding coefficient, PIC = polymorphic information content, SE = standard error

the analysis of molecular variance (AMOVA) were performed using GenAlex.

Cluster analysis of the 85 cowpea genotypes was conducted using the Jaccard dissimilarity matrix index. Cluster analysis was performed based on a neighbour-joining algorithm using the unweighted pair group method using arithmetic (UPGMA) in DARwin 5.0 software (Perrier, Jacquemoud-Collet, 2006). A dendrogram was then generated on the dissimilarity matrix. To investigate the genetic relationships among accessions, genetic distances among all pairs of individual accessions were estimated to draw the dendrogram. Bootstrap analysis was performed for node construction using 10 000 bootstrap values.

RESULTS

Genetic diversity and SNP characterisation

The SNPs were filtered by removing the rare alleles (less than 2 %), high-missing ratios (more than 20 %), and monomorphic alleles. Out of a total of 18 284 SNPs tested, only 7 799 (51.15 %) were found to be polymorphic across the 85 accessions and fulfilled the selection criteria. The 7 799 selected SNPs were subjected to genetic analyses, and Table 2 presents the genetic diversity parameters measured from 85 cowpea genotypes. The number of polymorphic SNPs per chromosome varied from 448 on chromosome 5 to 940 on chromosome 3, with an overall mean of 650 SNPs per chromosome. Chromosome 2 had the highest polymorphic loci content (56.51 %), while chromosome 5 had the lowest with 39.12 %. Observed heterozygosity (H_o) ranged from 0.034 to 0.069 per chromosome. Similarly, gene diversity (H_e) ranged from 0.222 to 0.286, with a mean of 0.255. Chromosome 10 and chromosome 2 revealed the highest and lowest H_o and H_e values, respectively. The observed mean fixation rate (F_{IS}) was 79.8 %. Chromosome 2 had the highest F_{IS} value at 0.867, while chromosome 9 had the lowest F_{IS} value at 0.770. Markers on chromosome 10 had the highest PIC value of 0.284, while on chromosome 2 had the lowest value of 0.217, with a mean PIC of 0.254.

The distribution and genetic diversity parameter estimates of the 7 799 SNPs used in this study are presented in Fig. 1A. Observed heterozygosity (H_o) values ranged from 0.00 to 0.914 with a mean of 0.05 (Fig. 1B). The majority of the SNPs (49 %) had H_o values ranging from 1.1 to 5 %, indicating that the alleles of these SNPs were fixed among the cowpea genotypes, while only 67 SNPs had H_o values greater than 5 %. The SNP diversity ranged from 0.024 to 0.505, in which 26.6 % had values greater than 0.40, suggesting that the genotypes were highly genetically diversified (Fig. 1A). Generally, the PIC value ranged from 0.024 to 0.50, with an average value of 0.25. Approximately 40 % of the SNPs used in this study had PIC values exceeding the mean value 0.30 and 27 % of the SNPs had PIC values between 0.40 and 0.50, demonstrating the high discriminatory power of the markers (Fig. 1C). The inbreeding coefficient in contrast, displayed contrasting values ranging from –0.83 to 1.00, with a mean of 0.798. Of the 7 799 SNPs tested, 319 SNPs showed negative F_{IS} values, indicating that these markers were highly heterozygous among genotypes. More than 92 % of

Table 3. Genetic diversity within and among the cowpea populations stratified based on geographic origin

Origin	N	P_A	I	H_o	H_e	F_{IS}	% P
Nigeria	45	475	0.381	0.050	0.247	0.770	92.31
South Africa	23	222	0.358	0.052	0.233	0.717	85.14
Zimbabwe	15	147	0.325	0.047	0.216	0.731	73.59
Overall mean	–	–	0.355	0.050	0.232	0.740	83.68
SE	–	–	0.002	0.001	0.001	0.002	5.45

N = number of genotypes tested per population, P_A = private allele, I = Shannon information index, H_o = average observed gene diversity within genotypes per population, H_e = average gene diversity within genotypes per population, F_{IS} = inbreeding coefficient, % P = percentage of polymorphic loci, SE = standard error

the SNPs had F_{IS} values exceeding 0.50 and 17 % of the SNPs had PIC values of 1.00, demonstrating that the majority of the SNPs were fixed (Fig. 1D).

Genetic diversity of inter- and intra-populations

The genetic diversity was further analysed by geographic origin as a stratification criterion. The average observed gene diversity within genotypes per population (H_o) ranged from 0.047 in the Zimbabwe accessions to 0.052 for accessions from Nigeria, with an overall mean value of 0.050 (Table 3). The mean values of the total Shannon information index ranged from 0.325 to 0.381 for accessions from Nigeria and Zimbabwe, respectively, with an overall mean value of 0.355, while the average gene diversity within genotypes per population (H_e) ranged from 0.216 to 0.247 with an overall mean value of 0.232. The inbreeding coefficient values ranged from 0.717 for South African accessions to 0.770 for accessions from Nigeria, with an overall mean value of 0.740, while the percentage of polymorphic loci (%P) ranged from 73.59 for accessions from Zimbabwe to 92.31 for accessions from Nigeria, with an overall mean value of 83.68 (Table 3). Diversity indices observed among the three geographic origins revealed that the Nigerian accessions had the highest number of private alleles (475), Shannon index (0.381), expected gene diversity (0.247) and percentage of polymorphic loci (92.3 %) (Table 3).

According to the standard guidelines for the interpretation of genetic differentiation (Wright, 1978), the range 0–0.005 indicates little, 0.05–0.15 indicates moderate, 0.15–0.25 indicates great, and above 0.25 indicates very great genetic differentiations. Genetic differentiation (F_{ST}) revealed moderate genetic differentiation among the accessions ranging from 0.04 to 0.07 in all regions (Table 4). Similarly, Slatkin (1989) and Morjan, Rieseberg (2004) indicated that gene flow (N_m) < 1 is considered to be low, while N_m = 1 is considered to be moderate and N_m > 1 is considered to be high. In this study, a very high gene flow (4.89) was observed between Zimbabwean and South African accessions, indicating that there was germplasm exchange between the neighbouring countries. Xiong et al. (2016) observed that the majority of genetic variance exists within instead of among geographic regions and within instead of among

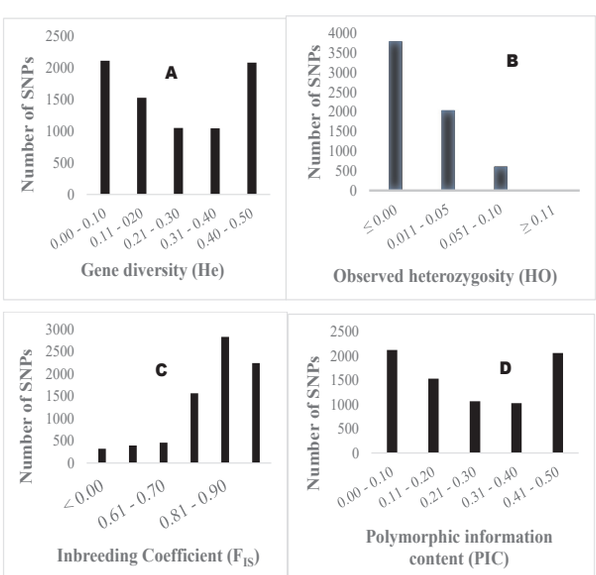

Fig. 1. Distribution of the 7,799 SNPs estimated on 85 cowpea genotypes (A) Gene diversity (He); (B) Observed heterozygosity (Ho); (C) Inbreeding coefficient (FIS); and (D) Polymorphic information content (PIC).

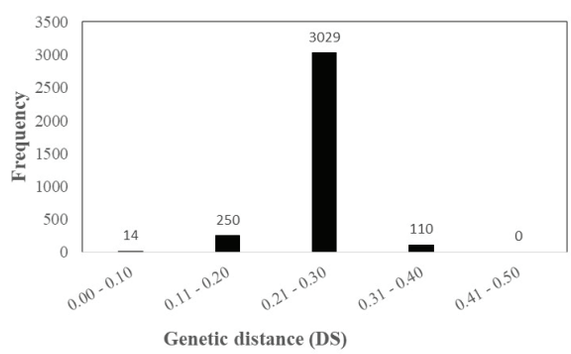

Fig. 2. Genetic distance estimate among 85 cowpea genotypes

Table 4. Pair-wise estimates of genetic differentiation (F_{ST}) (above diagonal off brackets), gene flow (N_m) (above diagonal within brackets); genetic distance GD (lower diagonal off brackets) and genetic identity (GI) (lower diagonal within brackets)

Population	Nigeria	South Africa	Zimbabwe
Nigeria		0.049 (4.89)	0.065 (3.57)
South Africa	0.037 (0.965)		0.048 (4.92)
Zimbabwe	0.050 (0.951)	0.027 (0.974)	

countries. GBS was used to discover SNPs in cowpea and the identified SNP alleles were used to estimate the level of genetic diversity, population structure, and phylogenetic relationships from 768 worldwide cowpea genotypes.

Analysis of molecular variance (AMOVA) was carried out on the three regions, and revealed substantial geographic differentiation in cowpea accessions (Table 5). A highly significant ($P < 0.001$) differentiation was observed among populations, among individuals, and within individuals. The variation between individuals (78 %) was higher than those between populations (8 %) and within individual varieties (15 %). There is a moderate amount of differentiation between the three regions, indicating that the accessions in the three regions are relatively genetically distinct. The significant F_{IS} values observed indicate that the cowpea lines within the regions were inbred lines.

The genetic distance value estimated on the basis of SNP markers ranged from 0.14 to 0.44 with a mean value of 0.35 (Fig. 2). The majority (89 %) of the genetic distance based on differences at marker loci between pairs of accessions ranged from 0.20 to 0.30. Neighbour-joined cluster analysis generated by UPGMA clearly divided the 85 cowpea accessions into three distinct clusters (Fig. 3). Cluster one in black colour was made up of 49 accessions (59 %) admixture from the three geographic origins with most accessions from Zimbabwe and South Africa. Cluster two designated by blue colour had 30 cowpea accessions with majority (27) of the accessions from Nigeria. Cluster three with a red colour designation contained only four accessions, three (GH43, GH45 and GH50) from Nigeria and GH 47 from South Africa.

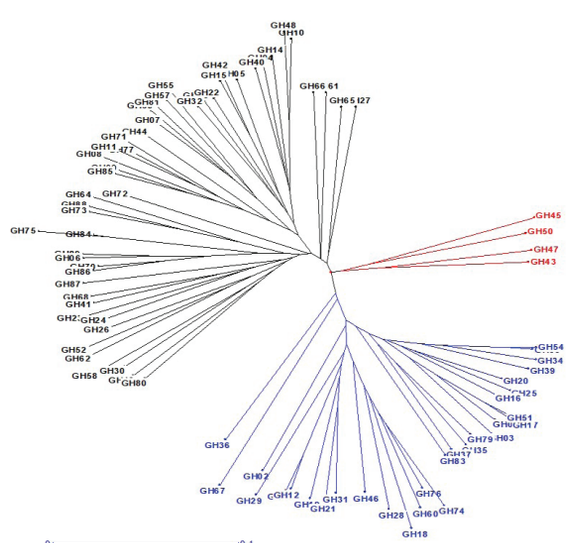

Fig. 3. Clustering patterns of the 85 cowpea accessions constructed based on neighbour-joining algorithm using unweighted pair group method (UPGMA)

DISCUSSION

The assessment of the genetic diversity of cowpea accessions using informative molecular markers is important for its management, genetic improvement, and conservation in plant breeding. From such studies, accessions with great potential can be selected and further improved for cultivation as it is an important crop in the smallholder farming sector of sub-Saharan Africa.

In this study, DArTseq was used to analyse the genomes of 85 cowpea accessions. It is a cheap and efficient platform that allows genome-wide marker discovery through restriction enzyme-mediated genome complexity reduction and sequencing of the restriction fragments (Melville et al., 2017). Although DArTseq yields a lower density of markers (10 000 to 35 000 loci) compared to the GBS approach (over 800 000 loci), DArTseq has substantially higher genome wide coverage and lower missing data (Lambert et al., 2016; Barilli et al., 2018). In addition, DArTseq provides a means to directly score samples as heterozygous/homozygous at each locus with the lower density approach, and provide thousands of short primes with polymorphic loci. Restriction site-associated DNA sequencing (RADseq) typically yields markers of 85 bp or longer, while DArTseq criteria produces shorter sequences of 69 bp (Lambert et al., 2016). It was suggested that GBS markers had low redundancy, and it was the best technique for further diversity analyses and genomic selection.

In this study, over 18 000 SNPs were used to assess the level and pattern of genetic variation among cowpea genotypes collected from three geographic origins. However, only 51 % of the SNPs were polymorphic. The remaining SNPs were either monomorphic or contained rare alleles of less than 2% allele frequency or null alleles and these were eliminated from the analysis.

From the results, the observed genetic distance between pairs of cowpea genotypes based on 7 799 SNP markers ranged from 0.14 to 0.44, with a mean value of 0.35. Similar findings were reported by Fatokun et al. (2018) using 370 accessions sampled from world

Table 5. Analysis of molecular variance (AMOVA) among 85 cowpea accessions classified based geographic origin using 7799 SNPs markers

Source	df	SS	MS	Est.Var.	Per. Var.	F.Statistics
Among populations	2	14178.1	7087.5	98.2	8%	FST ($P \leq 0.001$)
Among individual	81	177615.7	2192.7	1002.7	78%	FIS ($P \leq 0.001$)
Within individual	84	15730.0	187.3	187.3	15%	FIT ($P \leq 0.001$)
Total	167	207517.8	-	1288.2	100%	

df= degree of freedom, SS= sum of squares, MS= mean sum of squares, Est. Var. = estimated variance, Per. Var. = Percentage variation

cowpea collections. However, Huynh et al. (2013) reported a wider range of genetic distances (0.01 to 0.72) based on shared alleles among cowpea landraces collected from 56 countries. Wang et al. (2008), in their study of genetic diversity using gene-derived markers and sequencing on the USDA *Vigna* germplasm collection, reported low genetic diversity and minimal genetic distance among cowpea accessions. Calculating genetic diversity is important because it helps to assess valuable alleles that are of interest in resisting biotic and abiotic stresses. Maintaining diversity gives the population a buffer against change, providing the flexibility to adapt. The low level of polymorphism detected in the present study and in other previous studies may be attributed to the self-pollinated reproduction mechanism of cowpea, and the restriction induced by a single domestication event (Badiane et al., 2004; Diouf, Hilu, 2005; Wamalwa et al., 2016). Xiong et al. (2016) observed that the degree of genetic diversity has a positive correlation with the number of countries from which the accessions were collected. Kambua et al. (2019) observed that when there are more accessions from different places of origin, a higher genetic diversity will be detected. Genetic diversity can also be influenced by population size. Results indicate that Nigeria has a much larger population size than the both South Africa and Zimbabwe, which can largely explain why it is most diverse. In this study, only accessions from three countries (Nigeria, South Africa, and Zimbabwe) were used; hence, the genetic diversity was low. Populations from different areas with similar genetic structures always have a smaller distance and similar genetic diversity.

AMOVA analysis revealed a moderate but significant differentiation in the cowpea accessions collected from three geographic regions. The AMOVA indicated significant differences among populations and individuals, and within individuals ($P < 0.001$). The variance among populations was significantly low (8 % of the total variation), while the variance among individuals was significantly high (78 % of the total variation). Similar results have been reported by Fatokun et al. (2018) and Chen et al. (2017), in which the highest variations were observed among accessions compared to those within accessions and among populations. Thus most of the genetic variation observed in cowpea is attributed to that among individuals rather than geographic alignment. The relatively low variation within accessions can be explained by the low outcrossing rate of cowpea, as its floral structure only promotes inbreeding (Lush, 1979). However, the high individual variation could probably be attributed to the high level of germplasm exchange by smallholder farmers across geographic regions. Additionally, the low level of variation observed among regions could be the result of high gene flow within regions with little time for genetic differentiation along geographical lines (Wamalwa et al., 2016). The low levels of differentiation among geographic regions and the high levels of variation within regions suggest that a large random collection would capture most of the genetic variation within cowpea accessions in each region (Xiong et al., 2016).

The clustering patterns of the 85 cowpea accessions collected from different geographic regions revealed the presence of three distinct groups. The observed clustering pattern was, to some extent, consistent with the geographic origins of the accessions. The study showed a similar genetic makeup among some of the accessions; this was exhibited by the closeness of the accessions in the UPGMA analysis. Cluster 1 had the most diverse accessions (GH43, GH45, and GH50 from Nigeria and GH47 from South Africa). Menssen et al. (2017) reported that the clustering of genotypes originating from different countries on the same cluster could be due to the fact that cowpea played a significant role in human history. Furthermore, these four accessions in Cluster one could be used to infuse new genetic diversity into cowpea breeding programmes, as they are not closely related to either of the two clusters. Mafakheri et al. (2017) assessed 32 cowpea genotypes for 17 morphological traits. The study confirmed the existence of a high morphological variation in cowpea genotypes, which is an important aspect for plant breeding programmes to introduce new and hybrid varieties.

The UPGMA analysis showed three clusters, with Cluster 2 dominated by accessions from Zimbabwe and South Africa, with a few from Nigeria. This implies that there is a high level of import, export, and exchange of accessions through human activity between South Africa and Zimbabwe; they are neighbouring countries that share the same geographic and political boundaries. The fewer Nigerian accessions in this cluster could

have happened because of human migration ages ago. However, Cluster 3 contained mostly Nigerian accessions, which implies that these could have originated from West Africa. Udensi et al. (2016) found six clusters from 20 cowpea accessions. The UPGMA-based cluster analysis revealed that cowpea accessions obtained from the same geographical locations were found on the same cluster. This geographically based clustering of the accessions was affirmed by the genetic distances results. This implies that cowpea accessions found on the same cluster were genetically similar, while those found on different clusters were genetically diverse.

The study revealed great diversity within individuals among the 85 cowpea accessions. Some accessions, according to the UPGMA analysis, were closely related, especially in Cluster 2 and Cluster 3, even though they were from different geographic regions. The differences and similarities of accessions in some clusters as a result of their locations indicate the extent of accession exchange among farmers from different regions (Al-Saady et al., 2018). However, Cluster 2 contains mostly Nigerian accessions, which implies that these accessions evolved in specific environments and shared a similar environmental bottleneck. Cluster 3 had four unique genotypes from Nigeria and South Africa. Menssen et al. (2017) reported that the clustering of genotypes originating from different countries on the same cluster could be due to the fact that cowpea played a significant role in human history. Furthermore, these four accessions in Cluster 3 may be used to infuse new genetic diversity into cowpea breeding programmes, as they are not closely related to either of the two clusters. It is also imperative to widen the genetic base of cowpea accessions, which could be achieved through the use of alleles either from IITA or from closely related wild relatives of cowpea, such as *Vigna triphylla* and *Vigna reticulate* (Ali et al., 2015).

CONCLUSION

In this study, SNP data analysis indicated the existence of high levels of genetic diversity among cowpea accessions that were collected from South Africa, Zimbabwe and Nigeria. The current study also found that the DArTseq-derived SNP markers were efficient in genetic diversity analysis and relatedness in cowpeas. Closely related accessions in Cluster 1 include GH43, GH45, GH47, and GH50. In Cluster 2, GH30, GH58 and GH80 closely resembled each other. GH60, GH76, and GH74 in Cluster 3 were also closely related. GH8, GH45, GH47, GH48, GH75 could also be selected for hybridisation or for use as parents. In order to provide biological meaning to the clusters, molecular data should be backed by morphological data. These can be used in the identification of SNP markers that are associated with desirable agronomic attributes, such as high tolerance to biotic and abiotic stress factors, grain yield and crop quality.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

Ali ZB, Yao KN, Odeny DA, Kyalo M, Skilton R, Eltahir IM (2015): Assessing the genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] accessions from Sudan using simple sequence repeat (SSR) markers. African Journal of Plant Science, 9, 293–304. doi: 10.5897/AJPS2015.1313.

Al-Saady NA, Nadaf SK, Al-Lawati AH, Al-Hinai SA (2018): Germplasm collection and seed diversity of cowpea (*Vigna unguiculata* (L.) Walp.). Asian Journal of Agriculture and Food Sciences, 6, 129–139. doi: 10.24203/ajafs.v6i4.5349.

Badiane FA, Diouf D, Sane D, Diouf O, Goudiaby V, Diallo N (2004): Screening cowpea [*Vigna unguiculata* (L.) Walp.] varieties by inducing water deficit and RAPD analyses. African Journal of Biotechnology, 3, 174–178.

Barilli E, Cobos MJ, Carrillo E, Kilian A, Carling J, Rubiales D (2018): A high-density integrated DArTseq SNP-based genetic map of *Pisum fulvum* and identification of QTLs controlling rust resistance. Frontiers in Plant Science, 9: 167.

Boahen SK, Savala CEN, Chikoye D, Abaidoo R (2017): Growth and yield responses of cowpea to inoculation and phosphorus fertilization in different environments. Frontiers in Plant Science, 8: 646. doi: 10.3389/fpls.2017.00646.

Boukar O, Fatokun CA, Huynh BL, Roberts PA, Close TJ (2016): Genomic tools in cowpea breeding programs: Status and perspectives. Frontiers in Plant Science, 7: 757. doi: 10.3389/fpls.2016.00757.

Boukar O, Belko N, Chamarthi S, Togola A, Batieno J, Owusu E, Haruna M, Diallo S, Umar ML, Olufajo O, Fatokun C (2018): Cowpea (*Vigna unguiculata*): Genetics, genomics and breeding. Plant Breeding, 138: pp 415–424. doi: 10.1111/pbr.12589.

Chen H, Chen H, Hu L, Wang L, Wang S, Wang ML, Chen X (2017): Genetic diversity and a population structure analysis of accessions in the Chinese cowpea [*Vigna unguiculata* (L.) Walp.] germplasm collection. The Crop Journal, 5, 363–372. doi: 10.1016/j.cj.2017.04.002.

Chidebe IN, Jaiswal SK, Dakora FD (2018): Distribution and phylogeny of microsymbionts associated with cowpea (*Vigna unguiculata*) nodulation in three agroecological regions of Mozambique. Applied and Environmental Microbiology, 84: 1–25. doi: 10.1128/AEM.01712-17.

Damarany A, El-Dkeshy MHZ, Attallah SY, Haridy AG (2018): Molecular markers and genetic variation revealed by RAPD-

PCR in seven cowpea (*Vigna unguiculata* (L.) Walp.) cultivars. International Journal of Biosciences, 13, 66–74.

Deulvot C, Charrel H, Marty A, Jacquin F, Donnadieu C, Lejeune-Henaut I, Burstin J, Aubert G (2010): Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. BMC Genomics, 11, 1: 468.

Diouf D, Hilu KW (2005): Microsatellites, RAPD markers to study genetic relationships among cowpea breeding lines and local varieties in Senegal. Genetic Resources and Crop Evolution, 52, 1057–1067.

Doyle JJ, Doyle JL (1987): A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19, 1–15.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchel SE (2011): A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One, 6: e19379.

Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND (1992): Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. Genetics, 1132, 841–846.

Fatokun C, Girma G, Abberton M, Gedil M, Unachukwu N, Oyatomi O, Yusuf M, Rabbi I, Boukar O (2018): Genetic diversity and population structure of a mini-core subset from the world cowpea (*Vigna unguiculata* (L.) Walp.) germplasm collection. Scientific Reports, 8: 16035. doi: 10.1038/41598-018-34555-9.

Huang H, Tan H, Dongmei XD, TangY, Niu Y, Lai Y, Tie M, Huanxiu L (2018): High-density genetic map construction and comparative genome analysis in asparagus bean. Scientific Reports, 8: 4836. doi: 10.1038/s41598-018-23873-7.

Huynh BL, Close TJ, Roberts PA, Hu Z, Wanamaker S, Mitchell R, Lucas MR, Chiulele R, Cisse N, David A, Hearn S, Fatokun C, Diop N, Ehlers JD (2013): Genepools and the genetic architecture of domesticated cowpea. Plant Genome, 6, 1–8. doi: 10.3835/plantgenome2013.03.0005.

Kambua MR, George ME, Mwangi GS, Miinda AE (2019): Genotypic diversity of cowpea from lower eastern Kenya. International Journal of Agronomy and Agricultural Research, 14, 9–19.

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, Aschenbrenner-Kilian M, Evers M, Peng K, Cayla C, Hok P, Uszynski G (2012): Diversity arrays technology: A generic genome profiling technology on open platforms. In: Pompanon F, Bonin A (eds): Data production and analysis in population genomics. Methods in Molecular Biology (Methods and Protocols), vol. 888. Humana Press, Totowa, 67–89. doi: 10.1007/978-1-61779-870-2_5.

Kilian A, Sanewski G, Ko L (2016): The application of DArTseq technology to pineapple. Acta Horticulture, 1111, 181–188.

Kolade OA, Olowolafe MO, Fawole L (2016): Characterisation of mutant cowpea (*Vigna unguiculata*) lines using random amplified polymorphic DNAs (RAPDS) and amplified fragment length polymorphism (AFLP) markers. African Journal of Biotechnology, 15, 2530–2537.

Lambert MR, Skelly DK, Ezaz T (2016): Sex-linked markers in the North American green frog (*Rana clamitans*) developed using DArTseq provide early insight into sex chromosome evolution. BMC Genomics, 17: 844. doi: 10.1186/s12864-016-3209-x.

Lush WM (1979): Floral morphology of wild and cultivated cowpeas. Economic Botany, 33, 442–447.

Mafakheri K, Bihamta MR, Abbasi AR (2017): Assessment of genetic diversity in cowpea (*Vigna unguiculata* L.) using morphological and molecular characterisation. Cogent Food and Agriculture, 3: doi: 10.1080/23311932.1327092.

Melville J, Haines ML, Boysen K, Hodkinson L, Kilian A, Date KLS, Potvin DA, Parris KM (2017): Identifying hybridization and admixture using SNPs: Application of the DAarTseq platform in phylogeographic research on vertebrates. Royal Society Open Science, 4: doi: 10.1098/rsos.161061.

Menssen M, Linde M, Omondia EO, Abukutsa-Onyango M, Dinssad FF, Winkelmann T (2017): Genetic and morphological diversity of cowpea (*Vigna unguiculata* (L.) Walp.) entries from East Africa. Scientia Horticulturae, 226, 268–276. doi: 10.1016/j.scienta.2017.08.003.

Morjan CL, Rieseberg LH (2004): How species evolve collectively: Implications of gene flow and selection for the spread of advantageous alleles. Molecular Ecology, 13, 1341–1356.

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, Close TJ (2009): A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. Proceedings of the National Academy of Sciences of the United States of America, 106, 18159–18164. doi: 10.1073/pnas.0905886106.

Munoz-Amatriain M, Mirebrahim H, Xu P, Wanamaker S, Luo MC, Alhakami H, Alpert M, Atokple I, Batieno BJ, Boukar O, Bozdag S, Cisse N, Drabo I, Ehlers JD, Farmer A, Fatokun C, Gu YQ, Guo YN, Huynh BL, Jackson SA, Kusi F, Lawley CT, Lucas MR, Ma Y, Timko MP, Wu J, You F, Barkley NA, Philip A, Roberts PA, Lonardi S, Timothy J, Close TJ (2016): Genome resources for climate-resilient cowpea, an essential crop for food security. The Plant Journal, 89, 1042–1054. doi: 10.1111/tpj.13404.

Nadeem MA, Nawaz MA, Shahid MQ, Dogan Y, Comertpay G, Yildiz M, Hatipoglu R, Ahmad F, Alsaleh A, Labhane N, Ozkan H, Chung G, Baloch FS (2017): DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. Biotechnology and Biotechnological Equipment, 32, 261–285. doi: 10.1080/13102818.2017.1400401.

Nei M, Li W (1979): Mathematical method for studying genetic variation in terms of restriction endonucleases. Proceedings of National Academy of Sciences of the United States of America, 76, 5269–5273.

Nemli S, Asciogul TK, Ates D, Esiyok D, Tanyolac MB (2017): Diversity and genetic analysis through DArTseq in common bean (*Phaseolus vulgaris* L.) germplasm from Turkey. Turkish Journal of Agriculture and Forestry, 41, 389–404. doi: 10.3906/tar-1707-89.

Peakall R, Smouse PE 2012: GENEALEX 6. Genetic analysis in Excel. Population Genetic Software for Teaching and Research. Molecular Ecology Notes, 6, 288–295.

Perrier X, Jacquemoud-Collet JP (2006): DARwin Software. Dissimilarity analysis and representation for Windows. Available from http://www.darwin.cirad.fr/darwin.html. Accessed on 19 October 2019.

Ragul S, Manivannan N, Mahalingam A, Prasad VBR, Narayanan SL (2018): SSR marker based DNA fingerprinting for cowpea varieties of Tamil Nadu [*Vigna unguiculata* (L.) Walp.]. International Journal of Current Microbiology and Applied Sciences, 7, 641–647. doi: 10.20546/ijcmas.2018.704.072.

Slatkin M (1989): Population structure and evolutionary progress. Genome, 31, 196–202.

Udensi OU, Okon EA, Ikpeme EV, Onung OO, Ogban FU (2016): Assessing the genetic diversity in cowpea (*Vigna unguiculata* L. Walp) accessions obtained from IITA, Nigeria using random amplified polymorphic DNA (RAPD). International Journal of Plant Breeding and Genetics, 10, 12–22.

Varshney RK, Chabane K, Hendre PS, Aggarwal RK, Graner A (2007): A comparative assessment of ESTSSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Science, 173, 638–649.

Wamalwa EN, Muoma J, Wekesa C (2016): Genetic diversity of cowpea (*Vigna unguiculata* (L.) Walp.) accession in Kenya gene bank based on simple sequence repeat markers. International Journal of Genomics, 2016, Article ID 8956412. doi: 10.1155/2016/8956412.

Wang ML, Barkley NA, Gillaspie GA, Pederson GA (2008): Phylogenetic relationships and genetic diversity of the USDA *Vigna* germplasm collection revealed by gene-derived markers and sequencing. Genetic Resources, 90, 467–480.

Wright S (1978): Evolution and the genetics of populations. Volume 4. Variability within and among natural populations. University of Chicago Press, Chicago.

Xiong H, Shi A, Mou B, Qin J, Motes D, Lu W, Ma J, Weng Y, Yang W, Wu D (2016): Genetic diversity and population structure of cowpea (*Vigna unguiculata* L. Walp). PLoS ONE, 11: e0160941. doi: 10.1371/journal.pone.0160941.

Corresponding Author:

Gabriel Vusanimuzi N k o m o (Doctorate of Tech Agriculture, MSc Crop Production, BSc Hon Agriculture and Natural Resources Management, Diploma in Horticulture and Certificate in Horticulture) Agriculture Research Council, 114 Chris Hani Street, Private Bag X1251, Potchefstroom, 2520, Republic of South Africa, phone: +263774135565, e-mail: gvnkomo@gmail.com