

DArT-seq based SNP analysis of diversity, population structure and linkage disequilibrium among 274 cowpea (*Vigna unguiculata* L. Walp.) accessions

Frejus Ariel Kpedetin Sodedji

Non-Timber Forest Products and Orphan Crops species Unit, Laboratory of Applied Ecology (LEA), University of Abomey-Calavi (UAC), 01 BP:526, Cotonou, Benin

Symphorien Agbahoungba (✉ agbasympho@gmail.com)

Laboratory of Applied Ecology, University of Abomey-Calavi <https://orcid.org/0000-0001-8454-7297>

Eric Echikintho Agoyi

Non-Timber Forest Products and Orphan Crops species Unit, Laboratory of Applied Ecology (LEA), University of Abomey-Calavi (UAC), 01 BP: 526, Cotonou, Benin

Konoutan Médard Kafoutchoni

Non-Timber Forest Products and Orphan Crops Species Unit, Laboratory of Applied Ecology (LEA), University of Abomey-Calavi (UAC), 01 BP: 526, Cotonou, Benin

Ho-Youn Kim

Smart Farm Research Center, Korea Institute of Science and Technology(KIST), Gangneung, Gangwon, 25451, Republic of Korea

Simon-Pierre Assanvo Nguetta

West Africa Center of Excellence in Climate Change, Biodiversity and Sustainable Agriculture (CEA-CCBAD), Bioscience Research Unit, University Felix Houphouet-Boigny, 22 BP:461, Abidjan, Côte d'Ivoire

Achille Ephrem Assogbadjo

Non-Timber Forest Products and Orphan Crop species Unit, Laboratory of Applied Ecology (LEA), University of Abomey-Calavi (UAC), 01 BP: 526, Cotonou, Benin

Research article

Keywords: Grain legume, Gene diversity, Next generation sequencing, Diversity array technology, Genetic differentiation

Posted Date: November 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-50796/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Genetic diversity in a germplasm is crucial for continuous improvement of crop varieties. A panel of 274 cowpea (*Vigna unguiculata* L.) accessions of unknown genetic diversity was assembled from diverse sources. This study used 3127 SNP markers, generated with the diversity array technology (DArT), to assess genetic diversity, population structure and linkage disequilibrium (LD) in the assembled germplasm.

Results: The population structure analysis inferred three subpopulations within the germplasm, which was confirmed by Neighbour-Joining (NJ) clustering and principal component analysis (PCA). Low genetic distances (0.005 to 0.44) were observed between accessions. Accessions from Africa; West and Central Africa (113 accessions), East and Central Africa (93 accessions), and Asia (53 accessions) were predominant in the germplasm; and distributed across all subpopulations. High fixation indexes ($0.48 \leq F_{ST} \leq 0.56$) were obtained for the inferred subpopulations. AMOVA revealed a very large contribution of within subpopulations variation to the observed genetic variation in the germplasm. However, the expected heterozygosity (H_e) was higher than the observed heterozygosity (H_o), indicating high proportion of inbred lines in the germplasm. Linkage Disequilibrium (LD) was observed in the germplasm, which showed a low decay at longer physical distance between markers in the genome.

Conclusions: Significant genetic structuration exists in the assembled cowpea germplasm which shows that there is a potential for improvement of the crop. The subgroups consisted mainly of inbred lines which, although from different geographical regions shared alleles in common reflecting high movement of seeds and exchange of germplasm between regions. The presence of linkage disequilibrium in the germplasm paves a way for prospective whole genome-wide association studies in cowpea for quality attributes and important agronomic traits.

Background

Diversity in plant genetic resources creates an avenue for plant breeders to develop improved varieties with desirable attributes to cope with the ever-changing environments [1, 2]. Cowpea [*Vigna unguiculata* (L.) Walp; $2n = 2x = 22$] is an important grain legume crop in the world [3]. Grain, leaves and pods are the most used parts of cowpea plant, and their characteristics vary among cultivars [4, 5]. Cowpea contributes to food and nutrition security and income generation of millions of households in semi-arid tropics, including Asia, Africa, and Latin America [3, 6, 7]. Although cowpea thrives in drought prone environments and on poor soils [8], it is highly susceptible to pest and diseases, which leads to the low yields (25-600Kg/ha) reported in the production areas [3, 9–11]. This is a threat to world food security, and calls for constant efforts by breeding programmes to explore, create and use diversity within the species in order to overcome the various biotic and abiotic constraints and meet consumers' preferences.

Genetic diversity of a given plant population reflects its evolutionary potential and determines its capacity to be improved [2]. It informs on the patterns and magnitude of population structure which is driven by

the combined effects of evolutionary processes such as recombination, mutation, genetic drift, demographic history, and natural selection [12]. Knowledge of the genetic structure can provide valuable guidelines for the formulation of breeding strategies [13, 14]. Investigating the genetic structure of a population entails a thorough analysis of the allelic patterns between individuals within the population to make use of the genetic variation for cultivar development [15].

To date, a range of molecular and quantitative methods have been developed for easy and effective assessment of genetic diversity [2]. The rapid advances in sequencing technologies provides many possibilities to decipher the organization of natural populations [16]. High throughput genotyping such as Diversity Arrays Technology (DArT) has emerged as technology of choice for genetic diversity analysis and genomic studies because of its efficiency and low cost [14]. DArT offers high-throughput markers system for genome analysis and has successfully been deployed to assess genetic diversity in a number of legume crops including cowpea [17, 18], common bean [19], and pigeonpea [20].

Previous studies indicated the presence of population structure within the cultivated cowpea (*V. unguiculata*) [17, 21, 22]. Two to four subgroups with varying levels of genetic diversity were identified depending on the germplasm used [17, 18, 21–25]. Meanwhile, the reproductive nature of cowpea, which is primarily a self-pollinated plant [6], increases the degree of inbreeding with individuals becoming more homozygous for many alleles, consequently narrow genetic base and genetic distance were reported in the species [23, 26]. Furthermore, the high degree of inbreeding in this species also increases chances of linkage disequilibrium (LD) between loci [27], which is a determining factor in marker-trait association analysis [28], hence should be assessed in a plant germplasm collection designated for long-term breeding.

In 2018-2019, a set of 274 cowpea accessions was acquired from different origins and will serve as genetic material for developing cowpea breeding programme in Benin. The present study used DArT-seq generated SNP markers for diversity, population structure and linkage disequilibrium analyses in the germplasm for effective breeding decision-making.

Results

Profile and diversity of SNP markers

A total of 12,689 SNPs markers were generated from the DArT-seq genotyping of 274 cowpea accessions (Table 1) from 33 countries across the globe (Fig. 1). High number of markers (9562 SNPs) were removed during filtering based on low minor allele frequency ($MAF < 0.05$), high percentage ($> 20\%$) of missing data, and lack of information on their position in the genome. The rest of the markers, 3217 SNPs (24.65%) spanning the 11 linkage groups of cowpea matched the quality criteria and were used in the diversity analyses.

The profiles of the 3127 SNP markers are presented in Table 2. The markers showed high reproducibility (0.99) with a mean call-rate value of 0.87. Markers diversity analysis revealed that these markers had an

average minor allele frequency (MAF) of 0.22, and the majority (93.51%) had a PIC value above 0.1 with a mean value of 0.24 (Fig. 2, Table 2). The mean expected heterozygosity (0.23) of the markers was higher than the average observed heterozygosity (0.07)

Population structure

Two different approaches (STRUCTURE and DAPC) were used to identify the fitting number of clusters within the cowpea germplasm. Results of the structure analysis showed that the likelihood of DeltaK (ΔK) peaked at $K=3$ (Fig. 3a), indicating that three clusters contribute to the total variation in the diversity panel under study. Consequently, the 274 cowpea accessions can be grouped into 3 subpopulations/clusters. The distribution of the cowpea accessions among different clusters revealed that cluster 1 had the highest percentage of membership (53.2%), followed by cluster 2 (31.8 %) and cluster 3 which recorded the lowest percentage (14.9 %). However, as shown in the inferred ancestry bar plot (Fig. 3b), some accessions were in admixture i.e. represent a sum of variation from more than one cluster. Based on the probability value for assignment of individual accession to a specific cluster, 31 accessions fell into the group of admixed individuals (Additional file 2).

In the DAPC approach, the curve of BIC versus number of clusters shows a rapid decline from $K=1$ to $K=3$ followed by a very slow increase (Fig. 4 a), suggesting that $K=3$ is the optimum number of clusters inferred through this approach. Furthermore, 2 discriminant functions (DA) were detected, which explained 59.09% and 24.92 % of the variation in the dataset respectively (Fig. 4b). The graph of probability values of assignment of each accession to a specific cluster showed a perfect inference of the accessions with no admixed individuals (Fig. 4c). The plot of the densities of individuals on the first discriminant function showed that cluster 1 was distant from the two other clusters (Fig. 4d). Cluster 2 had the highest number of membership (42.70%; 117 accessions), followed by cluster 3 (28.83%; 79 accessions), and cluster 3 (28.46%; 78 accessions) (Supplementary file2).

Diversity and genetic relationship between accessions

The grouping of the accessions according to their regions of origin showed that West and Central Africa (113 accessions), East and central Africa (93 accessions) and Asia (53 accessions) were the three dominant origins of the accessions while the rest (15 accessions) of the accessions were from North Africa, America, and Oceania (Table 1). The genetic distance values between pairs of accessions based on differences at marker loci varied from 0.005 to 0.44 (Additional file 1). The Neighbour Joining phylogenetic tree depicted three main sub-roots (Fig. 3 c), which confirms the presence of three clusters in the germplasm. Cluster 1 consisted of 53.28% (146 accessions) of the accessions, while cluster 2 and clusters3 had 36.86% (101 accessions), and 9.85% (27 accessions), respectively.

The phylogenetic relationship showed that accessions from all regions, except the accessions from Oceania, were distributed in two or three subpopulations (Additional file 2). Accessions from West and Central Africa, and accessions from East and Southern Africa were predominant in cluster 2 and Cluster 1, respectively. In terms of within country diversity, Nigerian accessions were highly represented in cluster

1 (20 accessions) and cluster 2 (17 accessions). The majority of accessions from Benin were found in cluster 2 (27 accessions). Ugandan accessions were predominant (76 accessions) among accessions from East and Southern Africa region, with the majority (53 accessions) in cluster 1. Indian accessions (49 accessions), the most predominant in the group of accessions from Asia, were distributed in cluster 1 (28 accessions), cluster 2 (15 accessions) and cluster 3 (6 accessions).

Principal Component Analysis (PCA) was performed and the scatter plot of the 274 accessions based on first two principal component axes, which together explained 23.57% of the variation among accessions, showed that the accessions can be grouped into 3 main subpopulations with some of them in admixture (Fig. 3 d). These patterns are consistent with the structure analysis.

Genetic diversity and population differentiation of observed groups

Genetic diversity and population differentiation in the germplasm were examined by estimating diversity parameters and analysing molecular variance among and between inferred clusters/subpopulations (Table 3, Table 4 and Table 5). From the structure analysis, the genetic variability among the three inferred subpopulations represented herein by Nei's net nucleotide distance (Table 3), varied from 0.17 (between cluster 1 and cluster 2) to 0.26 (between cluster 2 and cluster 3). This indicates a degree of relatedness between subpopulations, with cluster 1 more related to cluster 2 than cluster 3 while cluster 2 and cluster 3 were more distant. The highest within population variation (expected heterozygosity) was observed in cluster 3 (0.32), followed by cluster 1 (0.26). Cluster 2 contained the highest proportion of genetic variance ($F_{ST}=0.56$) whilst cluster 1 and cluster 3 had similar mean values of population variance ($F_{ST} = 0.49$ and 0.48). Results from the NJ clustering showed that the number of effective alleles (N_e) and polymorphic information content (PIC) values varied across the three subpopulations (Table 4). The highest number of effective allele ($N_e=1.48$) was recorded in cluster 3 while the highest PIC mean value was observed in cluster 2 (PIC=0.23). The observed heterozygosity (H_o) values were generally lower than the expected heterozygosity (H_e) across subpopulations (Table 4). Cluster 3 had the highest observed heterozygosity ($H_o=0.11$) while cluster 1 and cluster 2 recorded the lowest values for this parameter, 0.07 and 0.06, respectively.

The genetic diversity estimates based on the DAPC clustering method (Table 4) showed that the number of effective alleles ranged from 1.48 in cluster 3 to 1.44 in cluster 2. Cluster 2 had the highest PIC mean value (0.23) while the lowest value for this parameter (0.19) was observed in cluster 3. Overall, expected heterozygosity ($0.23 \leq H_e \leq 0.29$) was higher than observed heterozygosity ($0.05 \leq H_o \leq 0.10$) in all subpopulations. The highest observed heterozygosity was obtained in cluster 2 ($H_o=0.10$).

Analysis of molecular variance (AMOVA) revealed similar patterns in the repartition of the genetic variance irrespective of the clustering approaches used (Table 5). The total genetic variation in the germplasm was partitioned mainly into among accessions variation and among subpopulations variation. Low contribution of between subpopulations variance was observed; 1% and 19% following the

NJ clustering and DAPC, respectively. The inbreeding coefficient (F_{IS}) was very high implying that the inferred subpopulations are mainly composed of inbred lines.

Linkage disequilibrium

Linkage disequilibrium (LD) was analysed across the cowpea genome. Out of 3127 SNPs, 1754 SNPs spanning the whole genome were in true LD based on comparison between pairs of markers in window size of 500kb ($r=0.2$ threshold). These markers were almost evenly distributed across the genomes with the highest numbers of SNPs in LD with each other registered on chromosomes 3 and 7 (Fig5.a). LD estimates (R_{VS}^2), corrected by structure and genetic relationship observed in the cowpea germplasm, were plotted along physical genetic distance between markers across the genome (Fig5.b). The LD decay plot showed that on average LD estimates were in general low ($R_{VS}^2=0.02$). However, few high LD values were observed over a short physical genetic distance which decayed rapidly, with very low decline at longer distances between markers across the genome. LD decayed below $R_{VS}^2=0.2$, more precisely to $R_{VS}^2=0.1$, with an increasing physical distance of 8.75 to 25.16 kb (Fig. 5b).

Discussion

Markers diversity

The present study investigated genetic diversity and population structure of a panel of 274 cowpea accessions from diverse origins using a set of 3127 SNP markers. The SNP markers were polymorphic and informative suggesting they are amenable to reliable fingerprinting and good inference of genetic variation within the germplasm. The results showed that the markers were highly reproducible (0.99), scored high call rate (0.87) with average values of minor allele frequency (MAF) and polymorphism information content (PIC) of 0.22 and 0.24, respectively. These values are within the range of values reported in previous SNP-based genetic diversity studies conducted in some important food crops including cowpea [18, 21], common bean [19], and maize [29]. However, low heterozygosity ($H_o=0.07$) was observed in these markers. Low mean of observed heterozygosity values ($H_o=0.06$; $H_o=0.075$) was also reported in worldwide germplasm collections of cowpea maintained at USDA GRIN [21] and IITA [18]. The mean heterozygosity, calculated across a number of loci, is a true indicator of the degree of genetic variation within a population [30], implying there is low genetic variation in cowpea. This is expected since cowpea is primarily a self-pollinated crop exhibiting high degree of inbreeding [6, 23, 31], which reduces genetic diversity.

Population structure and relationships between accessions

Population structure analysis is important in understanding genetic diversity and association mapping in a germplasm [15]. The cowpea collection was divided into three subgroups irrespective of the approaches used, DAPC and STRUCTURE, and this was further confirmed by both PCA and Neighbour-Joining clustering analyses. Previous studies in worldwide germplasm (size= 298-768 accessions) of cowpea have

also reported the presence of three clusters [18, 25, 32], suggesting that our collection, to some extent, has captured the diversity in the crop. This can be very useful considering the use of the germplasm in breeding activities. Conducting evaluation on the whole panel for specific trait of interest may be very informative, and could also guide selection of genotypes to use as a training population in genomic prediction studies. However, this may require more high-density genome wide markers [33]. Differences were observed between the DAPC and Structure analysis in the size of the subgroups which can be attributed to the presence of admixed individuals in the population. Thirty-one (31) accessions were in admixture when the cut-off value of coefficient ancestry was set at 0.52, and this number could increase with higher thresholds, a factor that leads to discrepancies in the results of these analyses [34, 35].

Genetic differentiation and allelic patterns of the subpopulations

Low genetic distances were observed between accessions. The genetic distance varied from 0.005 to 0.44 between pairs of accessions in the set of the germplasm, confirming that some accessions within the germplasm shared many alleles. The result corroborates the findings of Fatokun et al. [18] who reported a low genetic distance, ranging from 0.0096 to 0.462, between pairs of 298 cowpea lines. The low genetic distance may limit the progress in developing superior crop varieties through simple hybridization between accessions. As highlighted earlier, the movements of seeds across geographical areas which promotes genes flow between breeding germplasms can affect existing genetic boundaries, which reduces genetic distance among individuals and populations differentiation.

High F_{ST} mean values ($0.48 \leq F_{ST} \leq 0.56$) were observed for the subpopulations, suggesting there is a remarkable level of genetic differentiation of the subpopulations. This was further supported by the results of analysis of molecular variance, with high contribution of within subpopulations variance to the total genetic variation in the germplasm. Previous studies [18, 39] also indicated that genetic variation in cowpea is mainly due to within subpopulations variation. Based on the observed heterozygosity (H_o) values, subpopulation 3 ($H_o=0.11$) and Subpopulation 2 ($H_o=0.10$) were the most diverse among all clusters following the Neighbour-Joining clustering and DAPC. Nevertheless, the expected heterozygosity (H_e) was moderately low and, in general higher than the observed heterozygosity (H_o), for all subpopulations. Fatokun et al. [18] also observed a similar trend in a mini-core subset of the world collection. Low H_o implies a high proportion of inbred lines within sub-populations [1], which is confirmed by the high inbreeding coefficient value ($F_{IS}=1$).

Patterns of Linkage disequilibrium

Linkage disequilibrium (LD) is very important to investigate in population genetic and genomic studies. On average, a low linkage disequilibrium characterized the cowpea germplasm collection with slow LD decay over long distances of physical distance between markers, and this can be explained by the selfing nature of the crop, which limits the effectiveness of recombination events and delays LD decay [40]. Deploying techniques such as mutagenesis and crossing between genetically distant individuals can help improve the recombination rate and subsequently increase genetic diversity among the germplasm.

Nonetheless, the LD analysis in line with previous studies [41] indicates a faster LD decay in a population of high number of inbred lines. The faster is the LD decay the better is the genetic mapping resolution [42]. In the present study, LD declined rapidly below 0.2, reaching 0.1 at a distance of 26.16 kb, suggesting there is potential for genome wide association studies and candidate gene selection [43]. A number of markers distributed on different chromosomes in the genome were in true LD and could serve the purpose. There have been reports of quantitative traits loci (QTLs) and or markers associated with different traits on these chromosomes in cowpea, including pod fiber contents [46], perenniality and floral scent, seed size [47] and resistance to biotic stresses (*Fusarium oxysporum* f.sp. *tracheiphilum* Race 3, *Striga gesnerioides* race 1) [44, 45]. Besides, some accessions in our collection were reported to have good attributes for important traits such as yield and resistance to flower thrips [48] and bruchid [49]. Hence, in-depth investigation of LD patterns in the germplasm can help to map genomic regions associated with these and other preferred traits in cowpea.

Conclusions

This study used 3127 high-quality DArT-seq SNP markers to genotype and analyse genetic diversity within a large collection of 274 cowpea accessions. Important genetic structuration was observed within the germplasm. Each of the subgroups identified exhibits a level of genetic diversity that can be leveraged in developing cowpea varieties with desirable attributes. The subgroups consisted mainly of inbred lines which, although from different geographical regions shared alleles in common that implies significant exchange of germplasm between regions. The presence of structure and linkage disequilibrium within the collection provides valuable insights into the future use of the germplasm in genome-wide association studies and its exploitation in cowpea breeding programmes.

Methods

Plant materials

The cowpea germplasm comprises 274 accessions from 33 countries (Fig. 1, Additional file 1). Seeds of the accessions were obtained from different sources including International Institute of Tropical Agriculture (IITA, Nigeria), Institut de l' Environnement et de Recherches Agricoles (INERA, Burkina-Faso) , Laboratory of Applied Ecology (LEA), University Naguia Abrogoua (Côte d'Ivoire), Makerere Regional Center for Crop Improvement (MaRCCI, Uganda) and United States Department of Agriculture-Agricultural Research Service (USDA-ARS). Seeds were sowed in plastic bags for leaf tissues sampling in greenhouse at the University of Abomey-Calavi (Benin).

DNA extraction and genotyping

Fresh leaf samples were collected from 14 days old plants of the 274 cowpea accessions into three 96 wells sample collection plates and shipped to the Integrated Genotyping Service and Support (IGSS) of Biosciences eastern and central Africa (BeCA)-ILRI Hub in Kenya for genotyping. Briefly, DNA was

extracted from the leaves' tissues using the Nucleomag Plant Genomic DNA extraction kit and, quality and quantity control checked on 0.8% agarose. Genotyping was done using the Diversity Arrays Technology Sequencing (DArT-seq). Genomic DNA Library construction was done using genomic complexity reduction technology [50]. The library was purified and quantified for cluster generation in an automated clonal amplification system (cBOT Illumina) followed by Next Generation Sequencing (NGS) in the sequencer HiSeq 2500 (Illumina). The reads were aligned to the cowpea reference genome *Vunguiculata_469_v1.0*, publicly accessible on Phytozome [51].

Data quality control, filtering imputation, and markers diversity analysis

Data quality control and filtering were performed using R package dartR [52]. SNP markers with more than 20% missing data, having a minor allele frequency (MAF) <0.05 and of unknown position were removed. Data imputation was done using the expectation maximization (EM) algorithm, which recorded the highest simple matching coefficient (SMC= 0.76) among other imputations algorithms as implemented in KDCompute pipeline [53]. Summary statistics of the SNP markers were generated in PowerMarker V3.25 [54]. The computed statistics included allele frequencies, expected heterozygosity (He), observed heterozygosity (Ho) and polymorphic information content (PIC).

Population structure analysis

Filtered SNPs were used to infer the population structure within the germplasm. The structure analysis was performed using the Bayesian clustering approach Bayesian clustering approach in STRUCTURE V2.3.4. [55]. The structure analysis was run considering a burn-in period of 10,000 Markov Chain Monte Carlo (MCMC) iterations and 100,000-run length, with an admixture model following Hardy-Weinberg equilibrium and correlated allele frequencies. Ten independent runs were performed for each value of K (number of clusters), ranging from 1 to 11. The outputs from structure were analysed in Structure Harvester [56], which enabled the identification of the best K-value as the distinct peak in the change of likelihood (ΔK). The fixation index (FST) of each of cluster was retrieved and interpreted according to Wright [57], considering an FST value above 0.25 as a very large genetic differentiation. The accessions were assigned to their respective cluster using the coefficient of ancestry values generated from the Structure software (Additional file 2), with the assumption that an individual is a true member of a given cluster if its coefficient of ancestry in this cluster is above 0.52 [37].

Discriminant analysis of principal components (DAPC) was performed to confirm the best fitting clusters (K) among the cowpea germplasm. DAPC is a multivariate method which uses sequential K-means and model selection to infer and describe clusters in populations of genetic related individuals [16]. In this approach, the optimum K was identified as the minimum number of clusters after which the Bayesian Information Criterion (BIC) increases or decreases by a negligible amount [16]. DAPC was implemented using adegenet package [58] in R V3.5.0 [59].

Genetic relationships and diversity analysis

To examine the phylogenetic relationships between accessions and confirm the number of clusters, an identity by state (IBS) distance matrix (Additional file 1) was generated in Tassel V5.2.60 [60]. A phylogenetic tree was constructed using the Neighbour-Joining (NJ) algorithm in Darwin V6.0.2 [61], which was exported in FigTree V1.4.3 [62] for annotation. Prior the analysis, the 274 accessions were grouped according to their geographical regions of origin to describe the composition of the identified clusters (Table 1). A principal component analysis (PCA) was done in Tassel V5.2.60 which enabled to construct a scatter plot of the cowpea accessions using the first two principal component axes (PC1 and PC2).

Estimates of genetic diversity parameters (H_e , H_o , PIC, N_e) of the identified subpopulations through the NJ and DAPC clustering approaches were computed using PowerMarker V3.25 and GenAlEx 6.41 [63]. Analysis of molecular variance (AMOVA) was also implemented in GenAlEx 6.41 using the SNP markers and the repartition of the 274 cowpea accessions into different subpopulations as revealed by the clustering analysis. Prior to the analysis, the markers dataset were numerically coded (A = 1, C = 2, T = 3, G = 4, see Additional file 1) as suggested in GenAlEx manual [64].

Linkage disequilibrium analysis

Linkage disequilibrium (LD) analysis was performed among markers across the genome. The LD estimates were generated in R LDcorsSv package [59, 65] which provides LD measure ($R_{v_s}^2$) corrected by both structure and relatedness between accessions in a population [66]. SNPs markers were pruned to identify markers which are in true linkage disequilibrium in window size of 500kb ($r=0.2$ threshold) using the R package SNPRelate [67]. The rate of LD decay was depicted as a graph of variation of $R_{v_s}^2$ along physical genetic distance (Mbp) between pairs of SNP markers; and the average distances at which LD decayed at $R_{v_s}^2=0.1$ and $R_{v_s}^2=0.2$ were estimated in R [59, 68].

Abbreviations

AMOVA: Analysis of molecular variance

BIC: Bayesian Information Criterion

DNA: Deoxyribonucleic acid

DArt: Diversity array technology

DArT-seq: Diversity array technology sequencing

DAPC: Discriminant analysis of principal component

EM: Expectation maximization

Fst: Fixation index

F_{IS}: Inbreeding coefficient

He: expected heterozygosity

Ho: Observed Heterozygosity

MAF: Minor allele frequency

Ne: Number of effective allele

MCMC: Markov Chain Monte Carlo

NGS: Next Generation Sequencing

NJ: Neighbour-Joining clustering

IBS: Identity by state

LD: Linkage disequilibrium

LEA: Laboratory of Applied Ecology

PCA: Principal Component Axis

PIC: Polymorphism Information Content

SNP: Single nucleotide polymorphism

SSR: Simple sequence repeats

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests

Funding

The sequencing of the 274 cowpea accessions used in this study was funded by the World Academy of Sciences (TWAS) grant 18-238 RG/BIO/AF/AC_G-FR3240303667.

Authors' contributions

FAKS assembled the cowpea accessions from the diverse sources, carried out the study, performed the statistical and bioinformatics analyses and drafted the original manuscript. SA assisted in germplasm acquisition, design of the study, compilation and proofreading of the manuscript. KMK helped in data analysis and reviewed the final draft. AEA, HK, SAN, AS, and EEA supervised the research, read and improved the original draft. All authors read and approved the final manuscript.

Acknowledgements

The supports provided by PASET-RSIF (Partnership for skills in Applied Sciences, Engineering, and Technology- Regional Scholarship and Innovation Fund) and the Carnegie Cooperation of New York through RUFORUM (Regional Universities Forum for Capacity Building in Agriculture) grant RU/2018/TQA/38 for the germplasm acquisition and the statistical analyses are acknowledged. The authors are grateful to the different institutions that provided the accessions used in this study: IITA-Nigeria, INERA-Burkina-Faso, Laboratory of Applied Ecology/UAC-Benin, University Naguia Abrogoua, MaRCCI-Uganda and USDA-ARS. We thank [Boris Mahule Elyse Alladassi](#), Mathieu Anatole Ayenon and Dr Jaeyoung Choi for the valuable discussion and advice.

References

1. 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;2015:1–14. doi:10.1155/2015/431487.
2. Frankham R, Ballou JD, Briscoe DA, McInnes KH. *Introduction to Conservation Genetics*. Cambridge University Press; 2002. doi:10.1017/CBO9780511808999.
3. Boukar O, Belko N, Chamarthi S, Togola A, Batiemo J, Owusu E, et al. Cowpea (*Vigna unguiculata*): Genetics, genomics and breeding. *Plant Breed*. 2019;138:415–24. doi:10.1111/pbr.12589.
4. Herniter IA, Lo R, Muñoz-Amatriaín M, Lo S, Guo Y-N, Huynh B-L, et al. Seed Coat Pattern QTL and Development in Cowpea (*Vigna unguiculata* [L.] Walp.). *Front Plant Sci*. 2019;10:1–12. doi:10.3389/fpls.2019.01346.
5. Gonçalves A, Goufo P, Barros A, Domínguez-Perles R, Trindade H, Rosa EAS, et al. Cowpea (*Vigna unguiculata* L. Walp), a renewed multipurpose crop for a more sustainable agri-food system: nutritional advantages and constraints. *J Sci Food Agric*. 2016;96:2941–51. doi:10.1002/jsfa.7644.

6. Timko MP, Singh BB. Cowpea, a Multifunctional Legume. In: Genomics of Tropical Crop Plants. New York, NY: Springer New York; 2008. p. 227–58. doi:10.1007/978-0-387-71219-2_10.
7. Fatokun CA, Tarawali SA, Singh BB, Kormawa PM, Tamò M. Challenges and Opportunities for Enhancing Sustainable Cowpea Production. In: Proceedings of the World Cowpea Conference III held at the International Institute of Tropical Agriculture (IITA). Ibadan, Nigeria, Nigeria: IITA; 2002. p. 433.
8. Carvalho M, Lino-Neto T, Rosa E, Carnide V. Cowpea: a legume crop for a challenging environment. J Sci Food Agric. 2017;97:4273–84. doi:10.1002/jsfa.8250.
9. Kamara AY, Omoigui LO, Kamai N, Ewansiha SU, Ajeigbe HA. Improving cultivation of cowpea in West Africa. In: Achieving sustainable cultivation of grain legumes. Cambridge, UK: Burleigh Dodds Science Publishing; 2018. p. 235–52. doi:10.19103/AS.2017.0023.30.
10. Togola A, Boukar O, Belko N, Chamarthi SK, Fatokun C, Tamo M, et al. Host plant resistance to insect pests of cowpea (*Vigna unguiculata* L. Walp.): achievements and future prospects. Euphytica. 2017;213:239. doi:10.1007/s10681-017-2030-1.
11. Sodedji FAK, Agbahoungba S, Nguetta SPA, Agoyi EE, Ayanan MAT, Sossou SH, et al. Resistance to legume pod borer (*Maruca vitrata* Fabricius) in cowpea: genetic advances, challenges, and future prospects. J Crop Improv. 2020;34:238–67. doi:10.1080/15427528.2019.1680471.
12. Andam CP, Challagundla L, Azarian T, Hanage WP, Robinson DA. Population Structure of Pathogenic Bacteria. Elsevier Inc.; 2017. doi:10.1016/B978-0-12-799942-5.00003-2.
13. Hayward MD, Breese EL. Population structure and variability. In: Plant Breeding. Dordrecht: Springer Netherlands; 1993. p. 16–29. doi:10.1007/978-94-011-1524-7_3.
14. Mogga M, Sibiyi J, Shimelis H, Lamo J, Yao N. Diversity analysis and genome-wide association studies of grain shape and eating quality traits in rice (*Oryza sativa* L.) using DArT markers. PLoS One. 2018;13:e0198012. doi:10.1371/journal.pone.0198012.
15. Eltahir S, Sallam A, Belamkar V, Emara HA, Nower AA, Salem KFM, et al. Genetic Diversity and Population Structure of F3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing. Front Genet. 2018;9:1–9. doi:10.3389/fgene.2018.00076.
16. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010;11:94. doi:10.1186/1471-2156-11-94.
17. Qin J, Shi A, Xiong H, Mou B, Motes DR, Lu W, et al. Population structure analysis and association mapping of seed antioxidant content in USDA cowpea (*Vigna unguiculata* L. Walp.) core collection using SNPs. Can J Plant Sci. 2016;96:CJPS-2016-0090. doi:10.1139/CJPS-2016-0090.
18. Fatokun C, Girma G, Abberton M, Gedil M, Unachukwu N, Oyatomi O, et al. Genetic diversity and population structure of a mini-core subset from the world cowpea (*Vigna unguiculata* (L.) Walp.) germplasm collection. Sci Rep. 2018;8:1–10. doi:10.1038/s41598-018-34555-9.
19. Nemli S, Kaygisiz Ascioğul T, Ates D, Esiyok D, Tanyolac MB. Diversity and genetic analysis through DArTseq in common bean (*Phaseolus vulgaris* L.) germplasm from Turkey. TURKISH J Agric For. 2017;41:389–404. doi:10.3906/tar-1707-89.

20. Yang S, Pang W, Ash G, Harper J, Carling J, Wenzl P, et al. Low level of genetic diversity in cultivated Pigeonpea compared to its wild relatives is revealed by diversity arrays technology. *Theor Appl Genet.* 2006;113:585–95. doi:10.1007/s00122-006-0317-z.
21. Xiong H, Shi A, Mou B, Qin J, Motes D, Lu W, et al. Genetic Diversity and Population Structure of Cowpea (*Vigna unguiculata* L. Walp). *PLoS One.* 2016;11:e0160941. doi:10.1371/journal.pone.0160941.
22. Ravelombola WS, Shi A, Weng Y, Clark J, Motes D, Chen P, et al. Evaluation of Salt Tolerance at Germination Stage in Cowpea [*Vigna unguiculata* (L.) Walp]. *HortScience.* 2017;52:1168–76. doi:10.21273/HORTSCI12195-17.
23. Chen H, Chen H, Hu L, Wang L, Wang S, Wang ML, et al. Genetic diversity and a population structure analysis of accessions in the Chinese cowpea [*Vigna unguiculata* (L.) Walp.] germplasm collection. *Crop J.* 2017;5:363–72. doi:10.1016/j.cj.2017.04.002.
24. Shi A, Buckley B, Mou B, Motes D, Morris JB, Ma J, et al. Association analysis of cowpea bacterial blight resistance in USDA cowpea germplasm. *Euphytica.* 2016;208:143–55. doi:10.1007/s10681-015-1610-1.
25. Qin J, Shi A, Mou B, Bhattarai G, Yang W, Weng Y, et al. Association mapping of aphid resistance in USDA cowpea (*Vigna unguiculata* L. Walp.) core collection using SNPs. *Euphytica.* 2017;213:36. doi:10.1007/s10681-016-1830-z.
26. Wamalwa EN, Muoma J, Wekesa C. Genetic Diversity of Cowpea (*Vigna unguiculata* (L.) Walp.) Accession in Kenya Gene Bank Based on Simple Sequence Repeat Markers. *Int J Genomics.* 2016;2016:1–5. doi:10.1155/2016/8956412.
27. Kovi MR, Fjellheim S, Sandve SR, Larsen A, Rudi H, Asp T, et al. Population Structure, Genetic Variation, and Linkage Disequilibrium in Perennial Ryegrass Populations Divergently Selected for Freezing Tolerance. *Front Plant Sci.* 2015;6:1–13. doi:10.3389/fpls.2015.00929.
28. Laidò G, Marone D, Russo MA, Colecchia SA, Mastrangelo AM, De Vita P, et al. Linkage Disequilibrium and Genome-Wide Association Mapping in Tetraploid Wheat (*Triticum turgidum* L.). *PLoS One.* 2014;9:e95211. doi:10.1371/journal.pone.0095211.
29. Adu BG, Badu-Apraku B, Akromah R, Garcia-Oliveira AL, Awuku FJ, Gedil M. Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. *PLoS One.* 2019;14:e0214810. doi:10.1371/journal.pone.0214810.
30. Sbordoni V, Allegrucci G, Cesaroni D. Population structure. In: *Encyclopedia of Caves.* Second Edi. Elsevier Inc.; 2012. p. 608–18. doi:10.1016/B978-0-12-383832-2.00090-6.
31. Farahani, Maleki, Mehrabi, Kanouni, Scheben, Batley, et al. Whole Genome Diversity, Population Structure, and Linkage Disequilibrium Analysis of Chickpea (*Cicer arietinum* L.) Genotypes Using Genome-Wide DArTseq-Based SNP Markers. *Genes (Basel).* 2019;10:676. doi:10.3390/genes10090676.
32. Shi A, Buckley B, Mou B, Motes D, Morris JB, Ma J, et al. Association analysis of cowpea bacterial blight resistance in USDA cowpea germplasm. *Euphytica.* 2016;208:143–55. doi:10.1007/s10681-

015-1610-1.

33. He T, Li C. Harness the power of genomic selection and the potential of germplasm in crop breeding for global food security in the era with rapid climate change. *Crop J.* 2020;8:688–700. doi:10.1016/j.cj.2020.04.005.
34. Campoy JA, Lerigoleur-Balsemin E, Christmann H, Beauvieux R, Girollet N, Quero-García J, et al. Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. *BMC Plant Biol.* 2016;16:49. doi:10.1186/s12870-016-0712-9.
35. Ketema S, Tesfaye B, Keneni G, Fenta BA, Assefa E, Greliche N, et al. DArTSeq SNP-based markers revealed high genetic diversity and structured population in Ethiopian cowpea [*Vigna unguiculata* (L.) Walp.] germplasms. *PLoS Genet.* 2020;1–20. doi:10.1371/journal.pone.0239122.
36. Lush WM, Evans LT. The domestication and improvement of cowpeas (*Vigna unguiculata* (L.) Walp.). *Euphytica.* 1981;30:579–87. doi:10.1007/BF00038783.
37. Basak M, Uzun B, Yol E. Genetic diversity and population structure of the Mediterranean sesame core collection with use of genome-wide SNPs developed by double digest RAD-Seq. *PLoS One.* 2019;14:1–15. doi:10.1371/journal.pone.0223757.
38. Gomez Carlos P. COWPEA Post-harvest Operations. Book. 2004;1–70.
39. Kouam EB, Pasquet RS, Campagne P, Tignegre J, Thoen K, Gaudin R, et al. Genetic structure and mating system of wild cowpea populations in West Africa Genetic structure and mating system of wild cowpea populations in West Africa. *BMC Plant Biol.* 2012;12. <http://www.biomedcentral.com/1471-2229/12/113%0APage>.
40. Morrell PL, Toleno DM, Lundy KE, Clegg MT. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proc Natl Acad Sci U S A.* 2005;102:2442–7.
41. Xu P, Wu X, Muñoz-Amatriaín M, Wang B, Wu X, Hu Y, et al. Genomic regions, cellular components and gene regulatory basis underlying pod length variations in cowpea (*V. unguiculata* L. Walp). *Plant Biotechnol J.* 2017;15:547–57.
42. Hindu V, Palacios-Rojas N, Babu R, Suwarno WB, Rashid Z, Usha R, et al. Identification and validation of genomic regions influencing kernel zinc and iron in maize. *Theor Appl Genet.* 2018;131:1443–57. doi:10.1007/s00122-018-3089-3.
43. Badji A, Otim M, Machida L, Odong T, Kwemoui DB, Okii D, et al. Maize Combined Insect Resistance Genomic Regions and Their Co-localization With Cell Wall Constituents Revealed by Tissue-Specific QTL Meta-Analyses. *Front Plant Sci.* 2018;9 June. doi:10.3389/fpls.2018.00895.
44. Pottorff M, Wanamaker S, Ma YQ, Ehlers JD, Roberts PA, Close TJ. Genetic and physical mapping of candidate genes for resistance to fusarium oxysporum f.sp. *tracheiphilum* race 3 in cowpea [*vigna unguiculata* (L.) walp]. *PLoS One.* 2012;7:e41600.
45. Ouédraogo JT, Tignegre J-B, Timko MP, Belzile FJ. AFLP markers linked to resistance against *Striga gesnerioides* race 1 in cowpea (*Vigna unguiculata*). *Genome.* 2002;45:787–93. doi:10.1139/g02-

043.

46. Watcharatpong P, Kaga A, Chen X, Somta P. Narrowing down a major QTL region conferring pod fiber contents in yardlong bean (*Vigna unguiculata*), a vegetable cowpea. *Genes (Basel)*. 2020;11.
47. Huynh B, Ehlers JD, Huang BE, Muñoz-Amatriaín M, Lonardi S, Santos JRP, et al. A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *Plant J*. 2018;93:1129–42. doi:10.1111/tpj.13827.
48. Agbahoungba S, Karungi J, Odong TL, Badji A, Sadik K, Rubaihayo PR. Stability and extent of resistance of cowpea lines to flower bud thrips in Uganda. *African Crop Sci J*. 2017;25:1.
49. Kpoviessi AD, Datinon B, Agbahoungba S, Agoyi EE, Chougourou DC, Sodedji FKA, et al. Source of Resistance among Cowpea Accessions to Bruchid, *Callosobruchus maculatus* F. Coleoptera: Chrysomelidae, in Benin. *African Crop Sci J*. 2020;28:49–65.
50. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Caig V, et al. Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. Springer S. New York: Springer Science+Business Media New York; 2012. doi:10.1007/978-1-61779-870-2.
51. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40:D1178–86. doi:10.1093/nar/gkr944.
52. Gruber B, Unmack PJ, Berry OF, Georges A. dartR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol Ecol Resour*. 2018;18:691–9. doi:10.1111/1755-0998.12745.
53. Diversity Arrays Technology. KDcompute. 2017.
54. Liu K, Muse S V. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 2005;21:2128–9. doi:10.1093/bioinformatics/bti282.
55. Bennett RA, Thiagarajah MR, King JR, Rahman MH. Interspecific cross of *Brassica oleracea* var. alboglabra and *B. napus*: effects of growth condition and silique age on the efficiency of hybrid production, and inheritance of erucic acid in the self-pollinated backcross generation. *Euphytica*. 2008;164:593–601. doi:10.1007/s10681-008-9788-0.
56. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*. 2012;4:359–61. doi:10.1007/s12686-011-9548-7.
57. Wright S. *Evolution and the genetics of populations: variability within and among natural populations*. Chicago: University of Chicago press; 1984.
58. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24:1403–5. doi:10.1093/bioinformatics/btn129.
59. Team RC. R v. 3.5. 0: A language and environment for statistical computing. 2018.
60. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007;23:2633–5. doi:10.1093/bioinformatics/btm308.

61. Perrier X, Jacquemoud-Collet JP. Darwin software. 2016. <http://darwin.cirad.fr/>.
62. Rambaut A. FigTree. 2016. <http://tree.bio.ed.ac.uk/software/figtree/>.
63. PEAKALL R, SMOUSE PE. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 2006;6:288–95. doi:10.1111/j.1471-8286.2005.01155.x.
64. Blyton MDJ, Flanagan NS. A comprehensive guide to GenAEx 6.5. Australian national University; 2006. <http://biology.anu.edu.au/GenAEx/>.
65. Desrousseaux D, Sandron F, Siberchicot A, Cierco-Ayrolles C, Brigitte Mangin. Linkage Disequilibrium Corrected by the Structure and the Relatedness. 2020;108:285–91. doi:10.1038/hdy.2011.73.
66. Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. Heredity (Edinb). 2012;108:285–91. doi:10.1038/hdy.2011.73.
67. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28:3326–8.
68. Badji A, Kwemoui DB, Machida L, Okii D, Mwila N, Agbahoungba S, et al. Genetic basis of maize resistance to multiple insect pests: Integrated genome-wide comparative mapping and candidate gene prioritization. Genes (Basel). 2020;11:1–27.

Tables

Table 1 Geographical distribution of the 274 cowpea accessions

| Regions | Accessions | Countries of origin |
|--------------------------|------------|--|
| West and Central Africa | 113 | Benin, Burkina-Faso, Côte d’Ivoire, Ghana; Liberia; Mali; Niger; Nigeria; Senegal, Central Africa Republic |
| East and Southern Africa | 93 | Kenya, Malawi, Mozambique, Uganda, Botswana, Lesotho, South Africa, Swaziland |
| North Africa | 3 | Egypt and Mauritania |
| Asia | 53 | India, Sri Lanka, Iran, Pakistan, Yemen |
| America | 9 | Brazil, Colombia, Guatemala, Honduras, Nicaragua , Puerto Rico, US |
| Oceania | 3 | Australia |

Table 2 Quality and diversity of SNP markers used to investigate genetic diversity and population structure of the cowpea germplasm

| 1-Markers quality parameters | Mean | Min | Max |
|-------------------------------------|-------------|------------|------------|
| Call rate | 0.87 | 0.80 | 0.98 |
| One ratio—reference allele | 0.66 | 0.05 | 1.00 |
| One ratio—SNP allele | 0.37 | 0.05 | 0.99 |
| Reproducibility | 0.99 | 0.91 | 1.00 |

| 2-Markers diversity | Mean | Min | Max |
|----------------------------------|-------------|------------|------------|
| Major Allele Frequency | 0.78 | 0.50 | 0.96 |
| Minor Allele Frequency | 0.22 | 0.05 | 0.5 |
| Expected Heterozygosity (He) | 0.30 | 0.08 | 0.5 |
| Observed Heterozygosity (Ho) | 0.07 | 0 | 0.40 |
| Polymorphism information content | 0.24 | 0.08 | 0.37 |

Table 3 Genetic variability among (net nucleotide distance) and within (expected heterozygosity) populations, proportion of membership, and mean value of the fixation index (Fst) observed from the population structure of 274 cowpea cultivars

| Population | Net Nucleotide Distance | | Expected Heterozygosity | Mean Fixation Index (Fst) | % of Individuals |
|-------------------|--------------------------------|------------------|--------------------------------|----------------------------------|-------------------------|
| | Cluster 2 | Cluster 3 | | | |
| Cluster 1 | 0.17 | 0.21 | 0.26 | 0.49 | 53.2 |
| Cluster 2 | | 0.26 | 0.22 | 0.56 | 31.8 |
| Cluster 3 | | | 0.32 | 0.48 | 14.9 |

Table 4 Diversity parameters of the identified subpopulations in the Neighbour-Joining (NJ) clustering and discriminant analysis of principal components (DAPC)

| Subpopulations | Subpopulation Size | Ne | He | Ho | PIC |
|------------------------|---------------------------|-----------|-----------|-----------|------------|
| NJ clustering | | | | | |
| Cluster 1 | 146 | 1.46 | 0.27 | 0.07 | 0.22 |
| Cluster 2 | 101 | 1.46 | 0.28 | 0.06 | 0.23 |
| Cluster 3 | 27 | 1.48 | 0.27 | 0.11 | 0.22 |
| DAPC clustering | | | | | |
| Cluster 1 | 78 | 1.47 | 0.26 | 0.05 | 0.21 |
| Cluster 2 | 117 | 1.44 | 0.29 | 0.10 | 0.23 |
| Cluster 3 | 79 | 1.48 | 0.23 | 0.06 | 0.19 |

Ne = Number of Effective Alleles, He = Expected Heterozygosity, Ho = Observed Heterozygosity

Table 5 Analysis of molecular variance (AMOVA) for variation among and within sub populations of 274 cowpea accessions

| Source | df | SS | MS | % Est.Var. | F-Statistics (F_{IS}) | Probability |
|----------------------------|-----|-----------|----------|------------|---------------------------|-------------|
| NJ based grouping | | | | | | |
| Among subpopulations | 2 | 2552.44 | 1276.22 | 1 | | 0.001 |
| Among accessions | 271 | 243531.44 | 898.64 | 99 | 1 | |
| Total | 273 | 246083.88 | | 100 | | |
| DAPC based grouping | | | | | | |
| Among subpopulations | 2 | 33573.48 | 16786.74 | 19 | | |
| Among accessions | 271 | 212513.58 | 784.18 | 81 | 1 | 0.001 |
| Total | 273 | 246085.11 | | 100 | | |

F_{IS} = Inbreeding coefficient, % Est.Var. = percentage of estimated variance

Figures

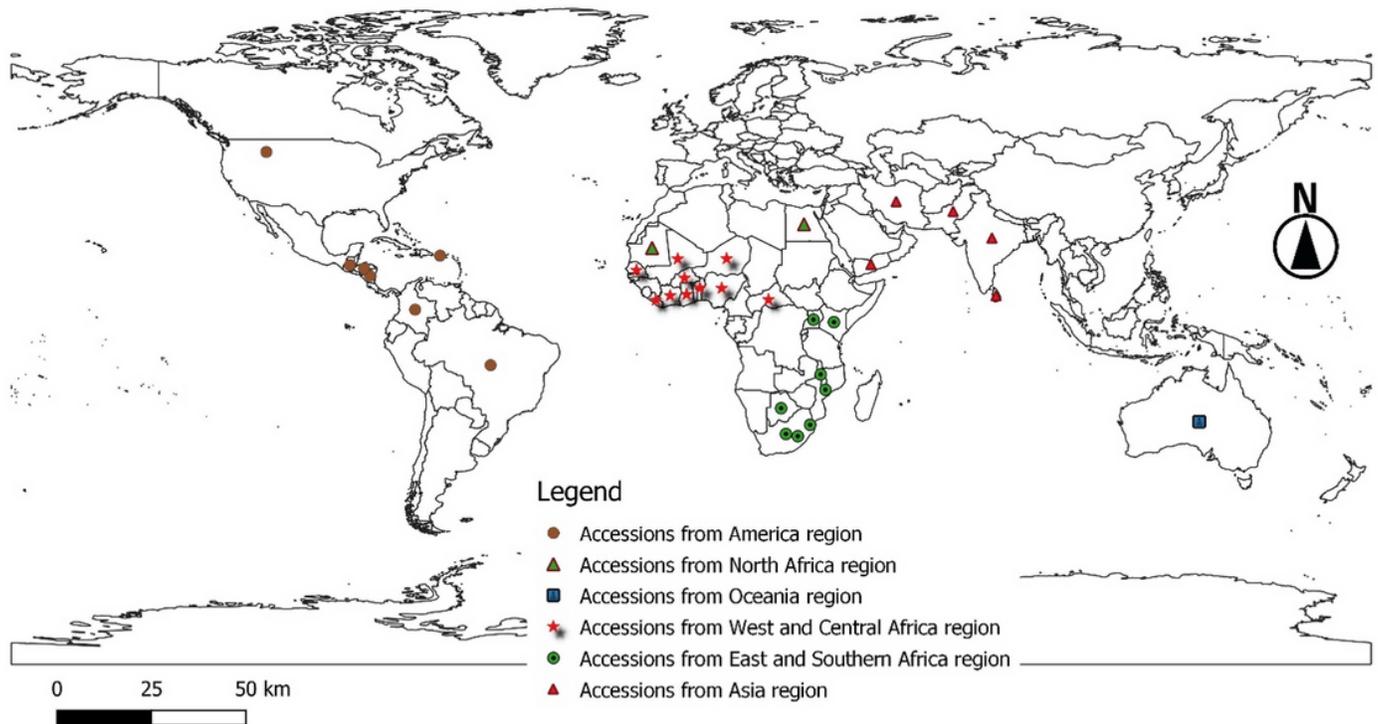


Figure 1

Geographical origins of the accessions. The map was drawn using QGIS version 3.4.4 (<https://www.qgis.org/en/site/about/index.html>). Each marker on the map represents a single country and accessions from the same regions are depicted with identical markers. A total of 274 accessions from 33 countries and 6 regions were included in the study. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

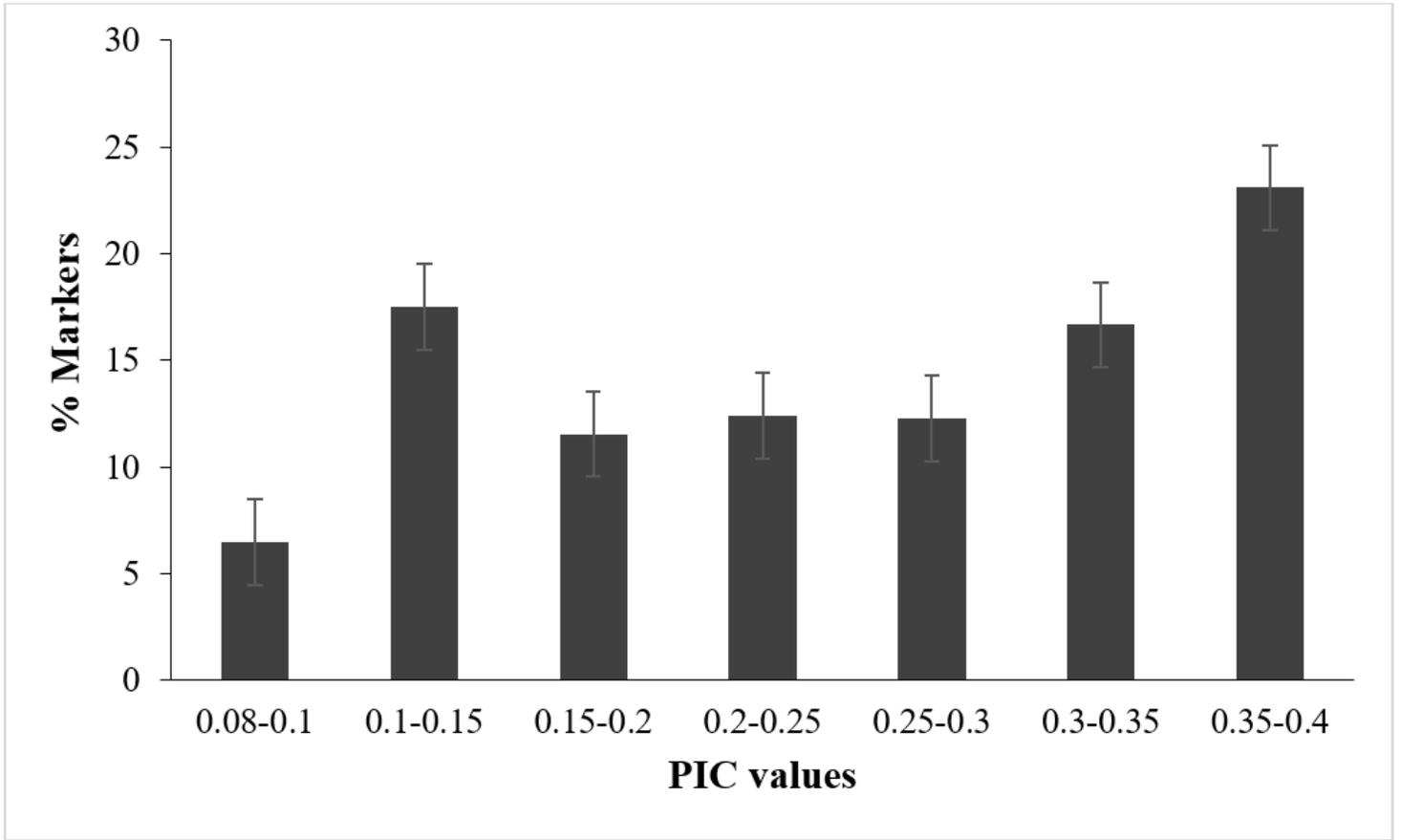


Figure 2

Distribution of PIC values of SNP markers

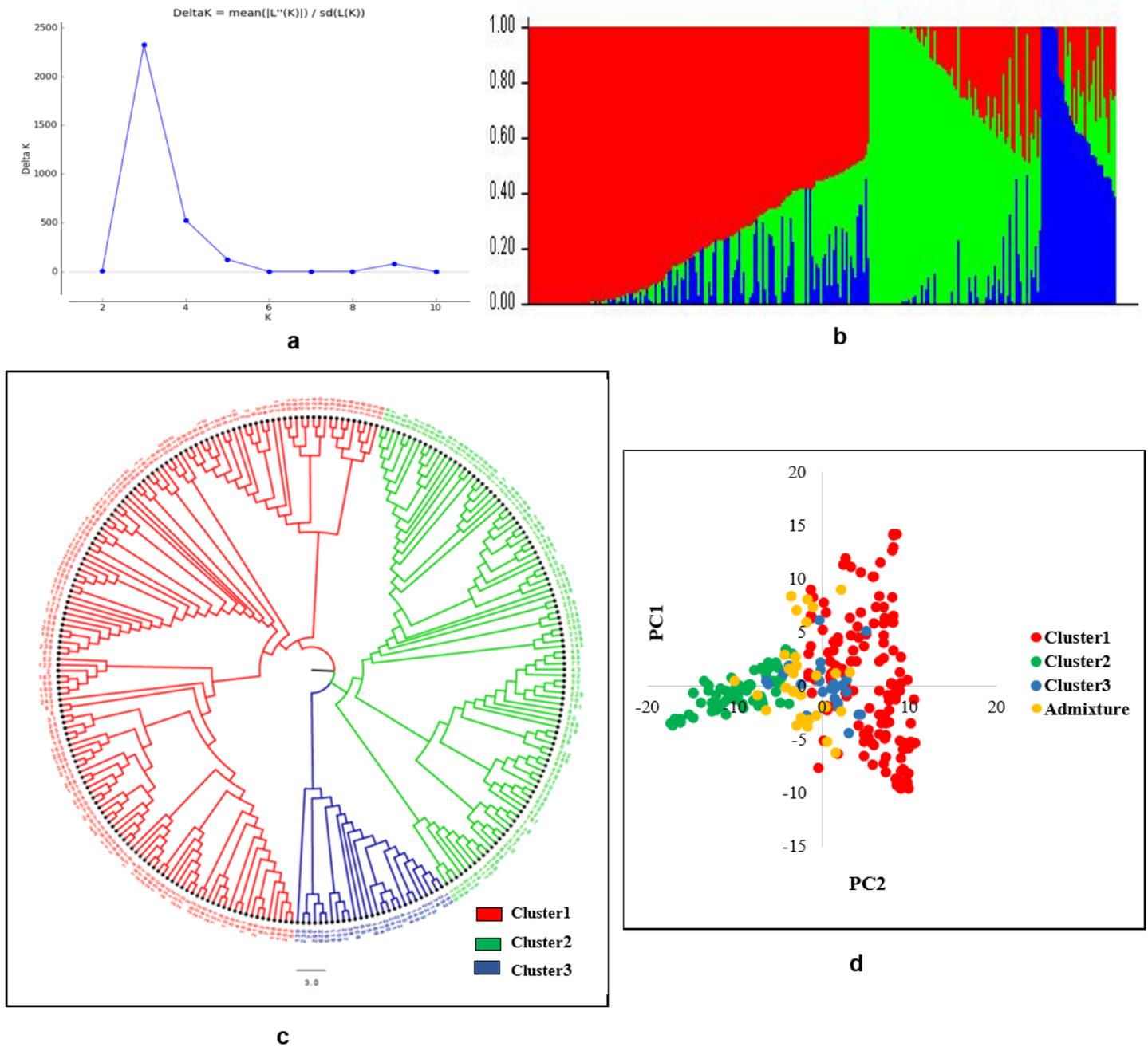


Figure 3

Population structure and phylogenetic relationships of 274 cowpea accessions based on 3127 SNP markers. (a) Likelihood of ΔK showing the best K value (K=3) from the Bayesian clustering approach in STRUCTURE V2.3.4 , (b) Bar plot of accessions assigned into K=3 clusters (cluster1, cluster2, cluster3 are represented by red, green and blue colours, respectively) as identified by STRUCTURE and STURCTURE Harvester, (c) Neighbour-Joining phylogenetic tree which groups accessions into 3 main branches depicting a similitude to the structure analysis and (d) scatter plot of the 274 cowpea accessions along with PC1 and PC2 which inferred the membership of the accessions to 3 clusters with some admixed individuals (Yellow colour) showing similar pattern to the structure analysis.

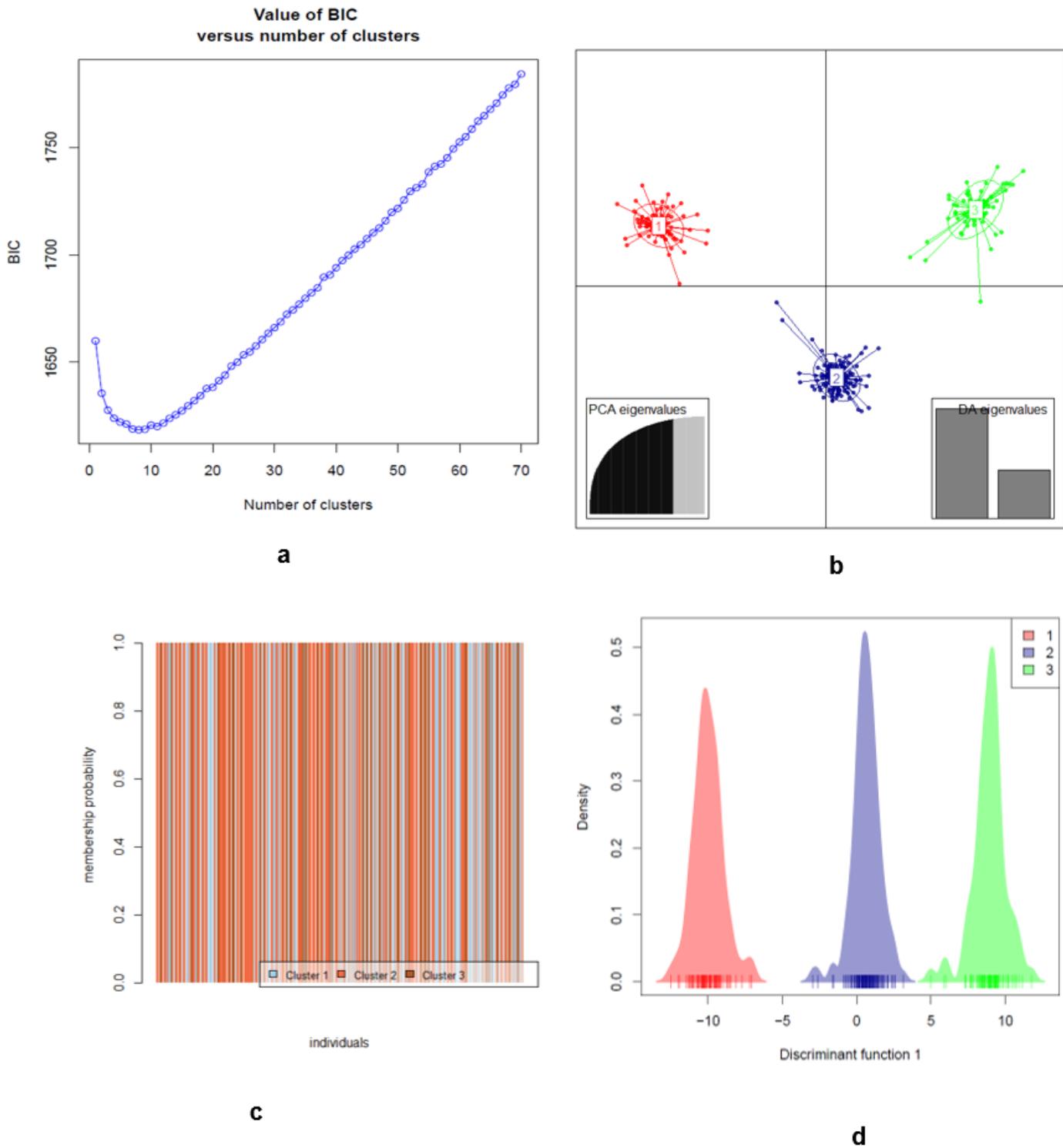
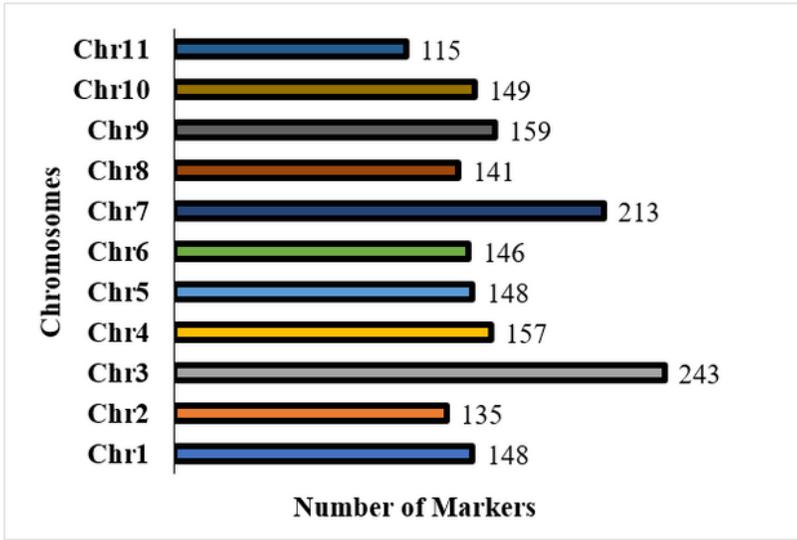
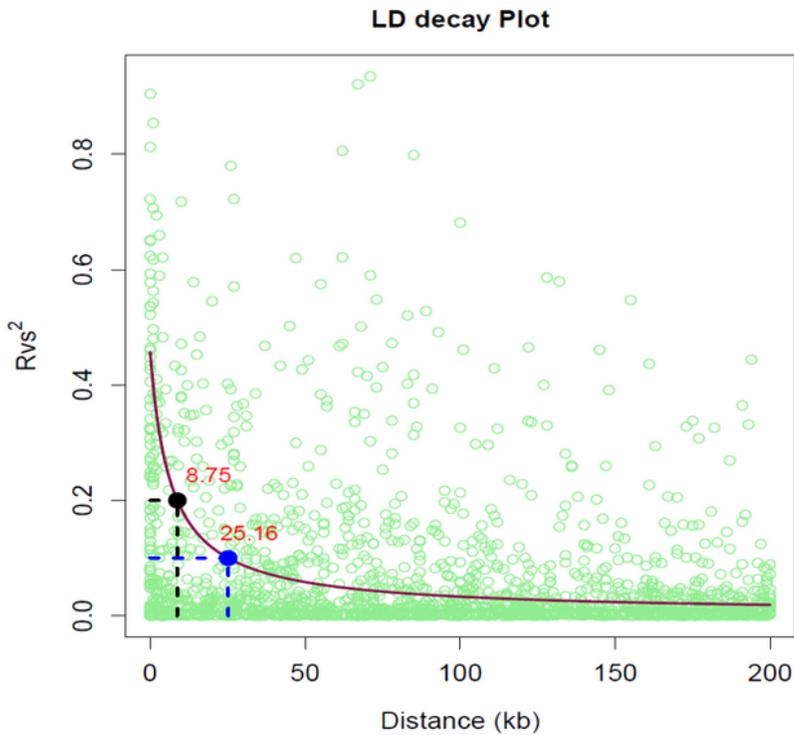


Figure 4

Discriminant analysis of principal components (DAPC). (a) The graph of Bayesian Information Criterion versus number of clusters indicates the optimum number of clusters ($K=3$) inferred in the cowpea diversity panel, (b) Scatter plot of the 274 cowpea accessions grouped into 3 clusters using 2 discriminant functions, (c) Probabilities values of inferred clusters and (d) Plot of the densities of individuals on the first discriminant function which displays clusters differences.



a



b

Figure 5

Linkage disequilibrium (LD) analysis showing: (a) the distribution of 1754 markers in true LD with each other across the cowpea genome and (b) the average LD decay in the cowpea diversity panel. The values on the Y-axis represent the squared correlation coefficient (R_{vs}^2) corrected by both population structure and relatedness among individuals and the x axis represents the physical distance in Kb.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.xlsx](#)
- [Additionalfile2revised.xlsx](#)
- [Additionalfile3.xlsx](#)