Core ideas

- Study informativeness of a cowpea mid-density genotyping panel
- The panel correctly ascertained genetic relatedness, diversity, and population structure
- The panel correctly depicted slow LD decay in a bi-parental population relative to other populations
- The panel enabled re-mapping of known genomic regions for seed and flower colors
- The panel's usefulness for quality control in a breeding program was confirmed

A medium density DArTag single nucleotide polymorphism panel for genetic dissections and deployment in cowpea improvement

Patrick Obia Ongom, https://orcid.org/0000-0002-5303-3602

Christian Fatokun, https://orcid.org/0000-0002-8428-7939

Abou Togola, https://orcid.org/0000-0001-6155-8292

Ana Luisa Garcia-Oliveira, https://orcid.org/0000-0001-8561-4172

Eng Hwa NG, https://orcid.org/0000-0003-0493-6212

Andrzej Kilian, https://orcid.org/0000-0002-2730-7462

Stefano Lonardi, https://orcid.org/0000-0002-2696-7274

Timothy J Close, https://orcid.org/0000-0002-9759-3775

Ousmane Boukar, https://orcid.org/0000-0003-0234-4264

Affiliations: P. O. Ongom, A. Togola and O. Boukar, International Institute of Tropical Agriculture (IITA), PMB 3112 Kano, Nigeria; C. Fatokun, International Institute of Tropical Agriculture (IITA), PMB 5320, Ibadan, Nigeria; A. L. Garcia-Oliveira, International Maize and Wheat Improvement Center (CIMMYT), ICRAF House, UN Avenue, PO Box, Nairobi 1041-00621, Kenya & Department of Molecular Biology, Biotechnology and Bioinformatics, College of Basic Sciences and Humanities, CCS Haryana Agricultural University, Hisar, Haryana, India; E. H. NG, Excellence in Breeding Platform, International Maize and Wheat Improvement Center (CIMMYT), 56237 El Batan, Mexico; A. Kilian, Diversity Arrays Tchenology Pty Ltd, University of Canberra, Monana st., Bruce, ACT 2617, Australia; S. Lonardi, Department of Computer Science and Engineering, University of California, 900 University Avenue, Riverside, CA, 92521 USA; T. J. Close, Department of Botany and Plant Sciences, University of California, 900 University Avenue, Riverside, CA, 92521 USA.

Received _____. *Corresponding author (P.Ongom@cgiar.org).

Abbreviations: DArT, Diversity array technology; DArTag, Diversity array technology targeted genotyping; RILs, Recombinant inbred lines; MAGIC, Multi-parent advanced generation intercross; PCA, Principal component analysis; DAPC, Discriminant analysis of principal components; AMOVA, Analysis of molecular variance; SNP, Single nucleotide polymorphism; MnAF, Minor allele frequency; MAF, Major allele frequency; LD, Linkage disequilibrium; GWAS, Genome wide association study.

Abstract

Genomic tools are increasingly being deployed to unlock factors affecting genetic gain. Here, we report the utility of a mid-density marker panel for genetic studies and other applications in cowpea breeding. The 2,602-marker panel was used to genotype 376 cowpea materials pooled from four different genetic backgrounds. The panel was informative with over 78% SNPs exceeding minor allele frequency of 0.20. The panel correctly deciphered co-ancestry among lines, identifying 0.04 % of all pairwise relationships as identical lines, 0.01% as parent-offspring, 0.12% as half-sibs, 39.19% as unrelated and 60.64% with distant relationships. STRUCTURE, principal component analysis (PCA) and discriminant analysis of principal components (DAPC) inferred two major groups, with all the bi-parental RILs placed in a single gene pool. Excluding bi-parental RILs exposed sub-structure within the remaining diverse lines. Variance within populations was higher (16.64%) than that between populations (8.38%). Linkage disequilibrium (LD) decay was correctly depicted as being slower in bi-parental RILs than in other populations. Overall, LD dissipated to $r^2 = 0.1$ at 1.25Mb. In addition, we mapped a region on chromosome VU07 known to be associated with both seed and flower colors in cowpea. This region harbors several genes including *Vigun07g110700*, a basic helix-loop-helix (bHLH) DNA-binding superfamily protein that regulates plant pigmentation. The panel revealed unexpected heterozygosity within some lines and authenticated the hybridity of F_1 progenies. This study demonstrated the panel's effectiveness for molecular applications in cowpea, and that the accessions that were used are genetically diverse and suitable for trait discovery and breeding.

Introduction

Cowpea [*Vigna unguiculata* (L.) Walp.] is a diploid species (2n = 2 x = 22) with a genome size estimated at 640.6 Mbp based on cytometry (Lonardi et al., 2019). Cowpea is globally recognized as a key food and nutritional security legume in sub-Saharan Africa (SSA). It is grown in areas regarded as marginal for many other crops owing to its inherent drought and heat tolerance and ability to fix soil nitrogen (Boukar et al., 2019). Cowpea feeds more than 200 million people in SSA where the major producers are smallholder farmers, and the crop is grown often as an intercrop with cereals (Boukar et al., 2019). Farmers grow this protein-rich crop for its grains, tender leaves and pods which are consumed as food, while the crop residues are used for fodder or added back to the soil to improve fertility (Singh, 2014). Despite its significance, cowpea suffers yield penalties from several biotic and abiotic stresses, including diseases (bacterial,

fungal, and viral), insect pests, parasitic weeds, and extreme drought and heat (Boukar et al., 2018; Sodedji et al., 2019; Nkomo et al., 2021). These challenges can be mitigated through development and deployment of improved varieties that are stress resilient.

Given the strategic placement as a food security crop, cowpea is gaining more research attention globally. Such efforts have in the past few years led to the development of excellent genetic resources that are being tapped to improve the crop's productivity. For instance, the International Institute of Tropical Agriculture (IITA) maintains over 15,000 accessions of cowpea from which a core and mini-core subsets representing the global diversity have been constituted (Fatokun et al., 2018). In addition, different genetic resources have been developed at IITA and collaborative institutions including Multi-parent Advanced Generation Inter-cross (MAGIC) population developed by University of California Riverside (UCR), bi-parental recombinant inbred lines, and many elite breeding lines developed by IITA. These resources have been tapped for trait improvement by several cowpea breeding programs across the world.

Efforts to accelerate genetic gain in cowpea have also led to development of genomic resources. The genome of cowpea has been dissected beginning with a single reference genome based on IITA line 1T97K-499-35 (Lonardi et al., 2019), currently expanded to seven reference genomes termed as the pan-genome of domesticated cowpea (https://phytozome-next.jgi.doe.gov/cowpeapan). To best exploit these genomes and broad genetic diversity within cowpea germplasm, genotyping platforms for cowpea have also been developed. The first was the 1536-SNP GoldenGate assay (Muchero et al., 2009) which has been exploited for linkage mapping and QTL analyses (Lucas et al., 2011; Muchero et al., 2013; Pottorff et al., 2014) and assessment of genetic diversity (Huynh et al., 2013). Another platform with high density markers is the Illumina Cowpea iSelect Consortium Array which represents a publicly accessible resource for screening 51,128 single nucleotide polymorphisms (SNPs) (Muñoz-Amatriaín et al., 2017). Due to cost limitations associated with these platforms, focus has recently shifted to Diversity array Technology (DArT), described as low-cost high-throughput, robust system with minimal DNA sample requirement capable of providing comprehensive genome coverage even in organisms without any prior DNA sequence information (Jaccoud et al., 2001). Since invention, the DArT platform has been extensively utilized in various crops including cowpea for different purposes: QTL mapping for grain yield traits using DArTseq platform (Garcia-Oliveira et al., 2020), genetic diversity and population structure analysis using DArTseq SNPs (Ketema et al., 2020; Nkhoma et al., 2020a).

DArT has been evolving leading to the development of multiple options that can be tailored to specific breeding needs. Among the suites of DArT options that have recently been developed is the targeted genotyping (DArTag) which allows genotyping using selected marker sets (https://www.diversityarrays.com/technology-and-resources/targeted-genotyping/). DArTag is a variant of many of the targeted genotyping suites developed by the DArT company. With

DArTag, any SNP (or a small indel) can be targeted if there is some genomic sequence available around the variant base/indel. DArTag offers cost efficiency and reduced bioinformatics load, well suited for high-throughput scenario.

In the present study, we described and demonstrated the application of a medium-density marker panel, here referred to as the '*Cowpea mid-density genotyping panel V1.0.*', for genetical studies and utilization in molecular breeding. This marker panel has 2,602 SNPs, custom-designed from 51,128-SNP cowpea iSelect Consortium Array (Muñoz-Amatriaín et al., 2017). The specific objective was to assess the performance of this custom-made SNP panel in diversity studies, potential applications in gene discovery and quality control (QC) using a set of 376 diverse cowpea genotypes. The study demonstrated the usefulness of this genomic resource to the cowpea community, facilitating adoption and deployment of molecular markers in cowpea improvement.

Materials and Methods

Plant genetic materials

The cowpea genotypes used in this study were constituted from groups of cowpea genotypes with different genetic backgrounds. The genetic groups included elite breeding lines, germplasm accessions from the IITA Genetic Resources Center, multi- and bi-parental recombinant inbreds, making a total of 376 genotypes **(Table 1)**.

Table 1 Descriptions of the cowpea genetic materials used in the study

Type of material	Size	Description
Breeding lines	123	Favourite breeding materials including released varieties and land races often use
Accessions	22	Favourite materials selected from the IITA mini-core which are part of a world c
Multi-parental lines	100	Randomly sampled from the UCR cowpea MAGIC recombinant inbred lines. Ha
Bi-parental lines	101	Randomly sampled from IITA recombinant inbred lines segregation for aphid res
F1 progenies	30	IITA crosses combining multiple traits including high yield, resistances to striga
Total	376	

The first group of cowpea genotypes used consisted of 123 elite breeding lines from IITA that are generally used as parents in several cowpea breeding programs. These lines are high yielding, drought tolerant, heat tolerant, striga resistant and have several seed quality traits demanded by farmers in SSA. The second category included 22 accessions which have been selected from the IITA cowpea mini-core population. The cowpea mini-core is a sub-set of a world cowpea germplasm collection maintained at IITA crop genetic resource gene bank and they are good sources for traits of economic importance in cowpea (Fatokun et al., 2018; Muñoz-Amatriaín et al., 2021). The third group consisted of 100 cowpea multi-parent advanced generation inter-cross (MAGIC) inbred lines previously described by (Huynh et al., 2018), here on referred to as multi-parental RILs. These recombinant inbred lines derived from eight diverse parents com-

bine many abiotic and biotic stress resistances, seed quality and agronomic traits relevant to cowpea in sub-Saharan Africa. A fourth group was a random sample of 101 bi-parental RILs derived from a cross between aphid resistant wild relative TVNu1158 and elite IITA line IT99K-573-1-1. The fifth category included 30 F_1 progenies derived from different crosses in the breeding program, mainly included to help in verifying the sensitivity of the marker panel in differentiating between heterozygous and homozygous genotypes.

Sample preparation

The 376 cowpea genotypes were planted in the screenhouse (Latitude11°58'51.5"N; Longitude 8°33'28.3"E) in pots of size 24 cm (height) x 25.4 cm (diameter), three-quarter filled with sterilized topsoil, placed on the crossing benches. Three seeds were sown per pot and thinned to one seedling a week after emergence. Two weeks after, a young trifoliate leaf from each plant was sampled for DNA analysis. The sampling was done according to the procedure described by Intertek-Agritech laboratory (Intertek-Agritech, 2016). First, unique sample identifications (UIDs) were generated from the EiB crop galaxy website (http://cropgalaxy.excellenceinbreeding.org/?tool_id=UIDs_generator&version=1.0.0&___identifer=a5g1e7h
and these were used to track the samples throughout from the time of sampling up to genotyping and data analysis. The plants to be sampled were labeled and a single hole-puncher (6.0 mm diameter) was used to punch and collect 4 leaf discs per sample from young and healthy newly developed trifoliate leaf. The punched leaf discs from each sample were transferred with forceps directly into a single well of 96-well plates (1.2 ml AbGene Storage Plate, number AB0564, Thermo Scientific). The forceps and the hole-puncher were wiped with 75% ethanol before and after placing each sample in a well to avoid cross contamination. The sample plates were later placed in a box containing silica gel which allowed the leaf discs to dry slowly for two weeks. The 96-well sample plates were then sealed with sealing mats (AB 0674, Thermo Scientific), wrapped in plastic bags, secured firmly, and shipped to Intertek laboratory South Australia.

DNA isolation and genotyping

Total genomic DNA was isolated at the Intertek laboratory Australia and the samples were forwarded to Diversity Arrays Technology (DArT) facility for genotyping. Genotyping was done by employing DArTag technology, one of the targeted genotyping approaches which offers the capacity to genotype materials using specific or selected sets of SNP markers (https://www.diversityarrays.com/technology-and-resources/targeted-genotyping/). For the 376 leaf samples, a panel of 2,602 SNP markers regarded as the *Cowpea mid-density genotyping panel V1.0.* was used. This marker platform has an average density of about 3 SNPs per cM (or 4 per Mbp) throughout the 11 cowpea chromosomes. Marker density is higher by physical distance in high recombination regions and lower in low recombination regions. These markers are a subset from the 51,128-SNP Cowpea iSelect Consortium Array (Muñoz-Amatriaín et al, 2017), and were selected based on iSelect data from 2,714 diverse cultivated cowpea accessions,

with extra weight given to 184 accessions most used in African breeding programs. The criteria used for marker selection (with some exceptions to fill gaps) were: (1) iSelect missing data rate less than 5%, (2) iSelect data MAF > 0.2 (all but 21 SNPs), and (3) even spacing along the genetic linkage map (by cM) weighted slightly by Mbp/cM (as calculated in Data S1 of Lonardi et al. 2019) to improve coverage of low recombination regions. A total of 2,753 SNPs matching these criteria were included in a DArTag test set against 376 cowpea DNA samples **(Supplemental Table 1)**. Data were obtained from 362 of these samples, including 307 that were previously genotyped using the iSelect array. These 362 samples included 191 RILs, 31 F1s and 140 diverse accessions. When considering the DArTag data only from 107 diverse accessions that were previously genotyped using the iSelect array, a total of 2,073 SNPs had no missing data, 2,435 SNPs had less than 10% missing data, and 2,602 SNPs had less than 50% missing data. The mid-density genotyping service information is also available at the EiB website (https://excellenceinbreeding.org/toolbox/services/cowpea-mid-density-genotyping-services). Further information on SNP markers and other aspects of the cowpea genome can be retrieved from CowpeaPan (https://phytozome-next.jgi.doe.gov/cowpeapan/), the Legume Information System (https://www.legumeinfo.org/), the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/genome/?term=cowpea), EnsemblPlants (http://plants.ensembl.org/index.html) and Lonardi et al. (2019).

DArTag genotyping was accomplished using special molecular probes that select the small target regions containing sequence variants. The targeted regions were then amplified and, in parallel, the sample-specific barcode was attached. The libraries generated were sequenced on the Next Generation Sequencing (NGS) equipment, Illumina Hiseq2500/Novaseq with 1,200,000 reads per sample. The resulting sequences processed using DArT PL's proprietary pipeline, that includes sequence alignment to sequences matching fragments of the IITA cowpea IT97K-499-35 reference genome (Lonardi et al., 2019) *Vigna unguiculata* v1.1, publicly accessible on Phytozome (https://phytozome-next.jgi.doe.gov/info/Vunguiculata_v1_1) delineated by the DArTag oligos from the panel and allele calling based on counts of alternative alleles for each sample and marker. DArTag genotyping details are accessible from DArT website (https://www.diversityarrays.com/technology-and-resources/targeted-genotyping/).

Data filtering

The data received from the DArT facility contained 2,753 SNPs genotyped across 362 out of 376 cowpea genotypes which included both the F_1 progenies and the lines. DArT report was not generated for 14 genotypes due to extreme missing data. Upon receipt, data were filtered using TASSEL v.5.2.79 (Bradbury et al., 2007) for missingness and low minor allele frequency (MnAF) with the following criteria: SNPs with $> 20\%$ missing data and MnAF < 0.05 were removed, leaving 2,435 SNPs. This data set was then used to test the marker panel for application in breeding as quality control (QC) markers. In the sec-

ond filtering step, the 30 F_1s were excluded from the data set and the remaining data were filtered against high heterozygosity, where genotypes with >0.3 heterozygosity were removed. The resulting data, consisting of 2,230 SNPs and 330 cowpea genotypes were used for identity by descent (IBD) analysis, PCA, structure and linkage disequilibrium analyses.

Nucleotide diversity analysis

The cowpea mid-density SNP panel was used to evaluate genetic diversity among individuals of the population. Major allele frequency (MAF), minor allele frequency (MnAF), proportion of heterozygous loci and proportion of missing data were calculated using the software TASSEL v.5.2.79 based on the default settings. The average pairwise divergence among genotypes, which represents the nucleotide diversity per base pair, (PiPerBP) and the expected number of polymorphic sites per nucleotide, (ThetaPerBP), were also estimated using TASSEL v.5.2.79. In addition, the normalized measure of difference between the observed ( ) and expected ( ) nucleotide diversity, Tajima's D (TajimaD), was computed.

Analysis of genetic relatedness

We dissected the extent of relatedness between every pair of cowpea line excluding the F_1 progeny to ascertain if the mid-density SNP panel can identify unexpected relatedness and/or duplicates among individuals in a population. Identity-by-descent (IBD) coefficients (k_0 and k_1), the probabilities of sharing 0, 1 or 2 alleles that are identical by descent, were estimated using Maximum Likelihood Estimation (MLE) procedure implemented in *SNPRelate* R package (Zheng et al., 2012). For this analysis, we utilized the function *snpgdsVCF2GDS()* in *gdsfmt* package to reformat the data to GDS (Graphic Data System) file required by *SNPRelate* package and performed LD pruning prior to IBD analysis. LD pruning was performed using the function *snpgdsLDpruning()* in SNPRelate package, setting the LD threshold at 0.2. Selected SNPs were extracted for use in IBD and subsequent analyses.

Population structure analysis

Filtered SNP data in HapMap format was uploaded in TASSEL v.5.2.79 and saved as a VCF (Variant Call Format) file. The software PGDSpider v.2.1.1.5 (Lischer & Excoffier, 2012) was used to convert the VCF file to STRUCTURE format. STRUCTURE 2.3.4 (Pritchard et al., 2000a) was used to conduct structure analysis. During the structure analysis, the parameters were configured and set to 5000 Burnin period, while the number of Markov Chain Monte Carlo (MCMC) repetitions after Burnin was 50000 and choosing the Admixture Model. A simulation was then run with the number of assumed populations (K) set from 1 to 10 and with 20 iterations for each K. The results file was zipped and uploaded to STRUCTURE Harvester, Web v0.6.94 (Earl & vonHoldt, 2012) where delta K was calculated and plots for the median value of lnPr(K) for each K were generated. This allowed the identification of the possible number of sub-groups (K) based on the second order rate of change of

the likelihood function with respect to K as described by Evanno et al. (2005). In addition, replicated results from structure program were summarized using CLUMPP (Cluster Matching and Permutation Program) version 1.1.2 (Jakobsson & Rosenberg, 2007). CLUMPP program reduces the stochastic effect of replicating STRUCTURE runs by calculating medians. The resulting output from CLUMPP was then used in software DISTRUCT version 1.1 (Rosenberg, 2004) to generate a graphical visualization of the population structure. Additional STUCTURE analysis was performed after excluding bi-parental RILs from the data to check any potential obscurity in detecting sub-structure within the remaining diverse lines. Further, PCA was conducted in R using the LD pruned SNPs and after removing duplicated lines based on IBD coefficient. Before performing PCA analysis, missing data was imputed using *missMDA R package,* based on the kfold method (Josse & Husson, 2016). The function *estimncpPCA()* from the *missMDA package* was used to estimate the number of components from incomplete data. The function *imputePCA()* was then used to invoke the iterative PCA algorithm for imputation. The imputed data were used for PCA computation, utilizing the *prcomp()* function in R. Visualizations of PCA were accomplished using factoextra package (Kassambara & Mundt, 2020) which generated a scree plot and a 2D PCA plot. Hierarchical cluster analysis was conducted using the *pheatmap* package. In this analysis, the function *pheatmap()* was used on the imputed and scaled SNP data, with the distance matrix based on Euclidean distance. Data imputation was done as described for PCA using the *missMDA* package and scaled using the function *scale()* from the *pheatmap* package.

Population differentiation analysis

To further demonstrate the applicability of the mid-density SNP panel in distinguishing between different populations, differentiation statistics and analysis of molecular variance (AMOVA) were used to compare diversity within and between four genetic populations, after excluding the F1 progenies. We considered the four different categories of genetic materials included in the study and the STRUCTURE-inferred groups or gene pools as unique populations for these analyses. SNP data were coded as: 1= major allele, 0 = alternative allele and 0.5 = heterozygotes. A separate file for population information was prepared, with columns listing all the cowpea lines along with the genetic population categorization of each into: Breeding lines, Bi-parental RILs, Multi-parental RILs and Accessions, and an additional column specifying the groups inferred from STRUCTURE analysis. Utilizing *adegenet* R package, the SNP data were converted into a '*genid object'* using the function *df2genind ()* and population information merged to the object. Measures of population differentiation (F_{ST} and G_{ST}) (Meirmans & Hedrick, 2011) were generated using the function *diff_stats()* of the *mmod* package. Gene flow (N_m) was estimated from F_{ST} according to island model (Wright, 1931; Wang, 2012) as follow: $N_m \quad 0.25(1 - F_{ST})/F_{ST}$. AMOVA was conducted in *poppr* package after converting the data from the *genid object* to *genclone object* using the function *as.genclone()*. The function *poppr.amova ()* was then used to generate AMOVA output. AMOVA was con-

ducted separately for the genetic populations and for the gene pools inferred from STRUCTURE. Finally, we conducted Discriminant Analysis of Principal Components (DAPC) to check which population has significant structure (Jombart et al., 2010). DAPC was done by invoking the function *dapc()* from *adegenet* package. The number of PCs retained was set to 100 after inspecting the curve of variance explained by PCA while the number of discriminant clusters was determined from the function *find.clusters()* which uses the K-means algorithm where the number of clusters corresponds to the lowest Bayesian Information Criterion (BIC) value (Jombart et al., 2010).

Linkage disequilibrium analysis

We estimated the rate of LD decay in the four cowpea genetic populations. A measure of LD (r^2) and pairwise distance between SNPs were generated in TASSEL v.5.2.79 and the rate of decay on each of the 11 cowpea chromosomes were visualized with graphics generated with *ggplot2* package in R. Mean LD per chromosome was calculated after every 0.5Mb interval, and the average genome-wide decay rate estimated by averaging LD in each interval across all chromosomes. LD was also computed for each of the four genetic populations separately to be able to decode the difference in the rate of decay within each population and the entire population. A line graph was used to clearly display the overlay of chromosome and population specific LD as well as the mean genome-wide LD decay rates.

GWAS analysis

To test the capability of the *Cowpea mid-density genotyping panel V1.0.* in trait mapping, we conducted genome-wide association analysis (GWAS analysis) using flower and seed color phenotypes. The 330 cowpea lines were planted at the IITA Minjibir research farm, in Kano state, Nigeria (12.1924° N, 8.6284° E). The nursery was established with 1-meter, single row plots, arranged in augmented design. At reproductive stage, flower colors were scored visually and later encoded into numeric values. Given that only two major flower colors were exhibited by these cowpea lines, white and purple colors were scored as 1 and 0 respectively. Similarly, after harvest, major seed colors were identified, recorded visually, and encoded into numeric values as follows: White =1, brown=2, black=3, purple=4, speckled=5 and mosaic colors=6.

GWAS was conducted in TASSEL v.5.2.79 where three different models were used to find which one best fits the data set with minimal spurious association based on the general formula:

$$y = \mu + \mathrm{X} + \mathrm{M} + \mathrm{Zk} + e$$

where y is a response vector for phenotypic values, μ is the total mean, is a vector of fixed effects regarding population structure estimated using principal components (PCs), α is the vector of fixed effect for markers, k is the vector of random effects for kinship and e is the vector of residuals, while X, M and Z are the incidence matrices relating individuals to , α and k, respectively. In-

dividual terms in the general formula above were excluded accordingly when fitting the three models: (i) Naive model: General Linear Model (GLM) without accounting for both structure and kinship; (ii) Q-model: GLM with PCs as correction for population structure; (iii) Q+K-model: Mixed linear model (MLM) with PCs and kinship (K)-matrix as correction for population structure. Model statistics from TASSEL v.5.2.79 were exported to R statistical package (R Core Team, 2014) where scripts were written to generate Manhattan and quantile-quantile (QQ) plots. The position of the SNPs spanning regions with significant association signals were used to locate model genes via phytozome (https://phytozome-next.jgi.doe.gov/info/Vunguiculata_v1_2).

Impurity and hybridity analysis

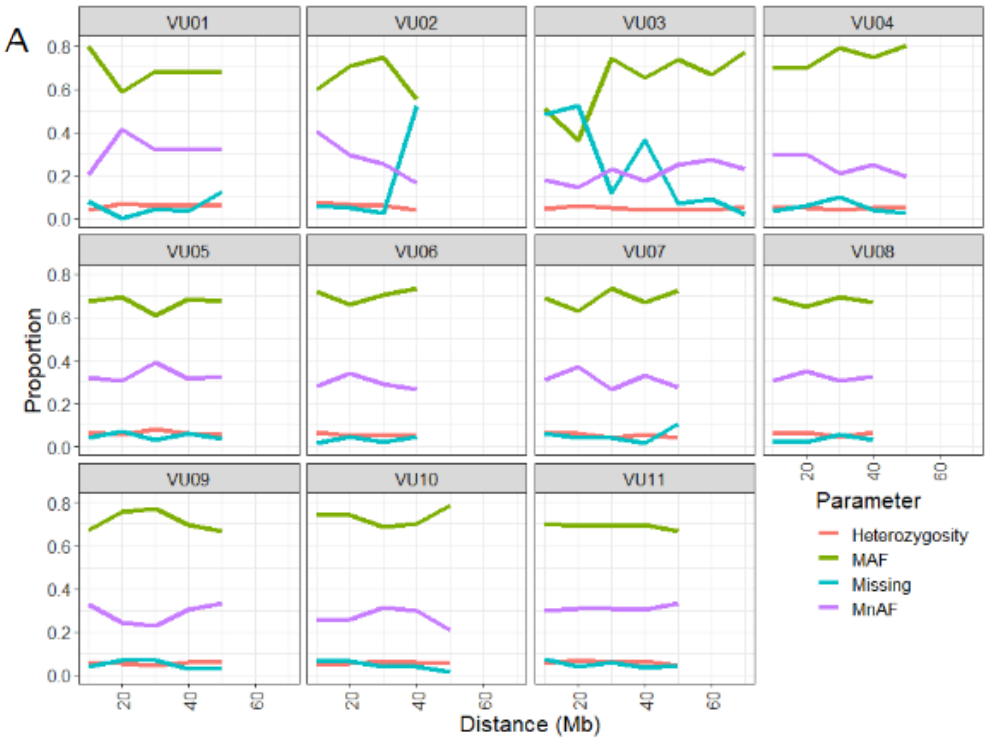
To evaluate the application of the *Cowpea mid-density genotyping panel V1.0.* for quality control/quality assurance (QC/QA) in breeding programs, we compared heterozygosity level of individual genotypes from the different germplasm categories included in the study: breeding lines, F_1 progeny, accessions, multi- and bi-parental RILs. TASSEL v.5.2.79 was used to compute proportion of heterozygous loci for each cowpea genotype. Box plots and a faceted dot graph were generated in R to depict the distribution of heterozygosity for the different categories of cowpea genotypes included in the sample. In addition, the 30 F_1s and their parents were considered separately for hybridity analysis. First, cluster analysis to determine diversity among the parents of F_1s was conducted in TASSEL v.5.2.79 using 2,163 filtered SNPs. For this analysis, a neighbor-joining method (Nei & Saitou, 1987) was used to generate the genetic distances and phylogenetic tree using archaeopteryx in TASSEL v.5.2.79 to visualize the clustering among parents. This was followed by analysis of marker polymorphism between every pair of parents used in making the 30 F_1s as previously described (Ongom et al., 2021). Markers that were found to be polymorphic between the parental pairs were then used to assess the level of hybridity among the F1s. Hybridity was expressed as a ratio of the number of polymorphic markers that detected a particular F_1 as being heterozygous to the total number of polymorphic markers between the parents of that cross (Ongom et al., 2021). Further, SNP marker efficiency was assessed by determining how frequent a marker was polymorphic across the 30 pairs of parents (Ongom et al., 2021), and this allowed us to identify a set of SNPs that are suitable for hybridity testing and parental fingerprinting.

Results

Polymorphism and nucleotide diversity

We examined the informativeness of the *Cowpea mid-density genotyping panel V1.0.* based on heterozygosity of loci, allele frequencies, nucleotide density and diversity. Chromosome-wide distribution of allele frequencies, and heterozygosity are presented in **Figure 1A.** These genetic parameters varied along chromosomes but generally exhibited high major allele frequency (MAF) followed by the minor allele frequency (MnAF) while heterozygosity proportions remained

low across all chromosomes. The proportion of missing marker data were also generally low, except on chromosomes VU02 and VU03 where the regions at 40 Mb and 20 Mb respectively had high missing data. Mean proportion of heterozygous loci ranged from 0.047 on chromosome VU03 to 0.065 on chromosome VU05, with a chromosome-wide average of 0.056 (**Supplementary Table 2**). The mean MAF ranged from 0.64 on chromosome VU03 to 0.75 on chromosome VU04 with a genome-wide average of 0.69. Meanwhile, MnAF ranged from 0.21 on chromosome VU03 to 0.33 on chromosome VU05 and a chromosome-wide average of 0.29 (**Supplementary Table 2**). Overall, 78% of SNPs had MnAF above 0.2 while the remaining 22% had MnAF that were 0.2 but still above 0.05 (**Figure 1B**).


A
VU01
VU02
VU03
VU04
VU05
VU06
VU07
VU08
VU09
VU10
VU11
Proportion
0.0
0.2
0.4
0.6
0.8
20
40
60
Distance (Mb)
Parameter
Heterozygosity
MAF
Missing
MnAF

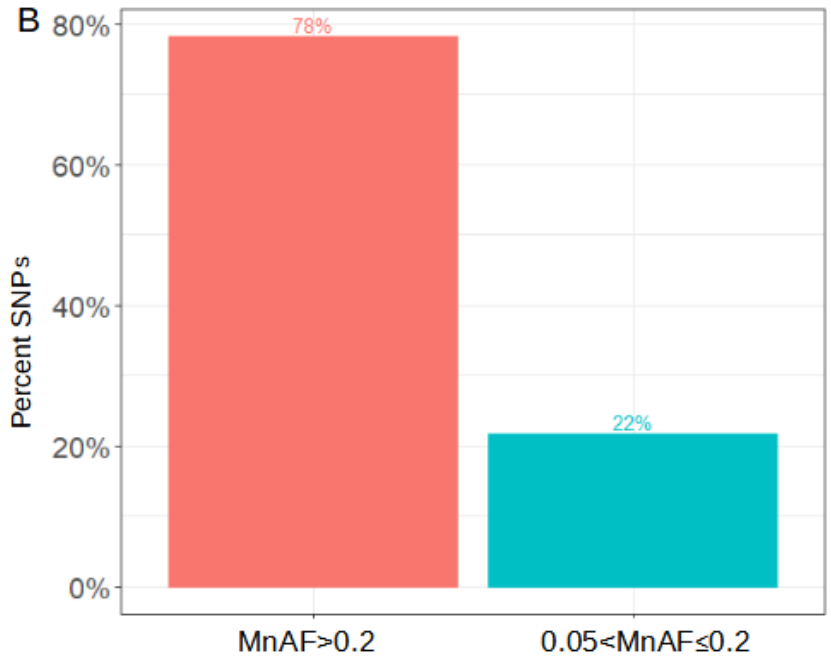

B
Percent SNPs
80%
60%
40%
20%
0%
78%
22%
MnAF>0.2
0.05<MnAF≤0.2

Figure 1 Distribution of allele frequency. (A) Chromosome-wide distribution of major and minor allele frequency proportions, including proportion of heterozygous loci and missing data. (B) Percentage distribution of informative SNP markers as defined by proportion of minor allele frequency (MnAF) above and below 0.2.

The distribution of SNPs per chromosome based on the number of SNPs within 1Mb window size is presented in **Figure 2.** Chromosome lengths varied with the shortest and longest chromosomes being VU02 (33.75Mb long) and VU03 (64.99Mb long) respectively (**Figure 2, Supplementary Table 3**). Chromosome VU03 had the highest number of SNPs (295) while chromosome VU10 had the lowest number of SNPs (165), with a genome-wide average number of SNPs per chromosome being 202.73 (**Supplementary Table 3**). Considering the varying chromosome lengths, chromosome VU07 which harbors 256 SNPs, registered the highest SNP density of 6 SNPs per Mb, while VU05 had the lowest density of approximately 4 SNPs per Mb and a chromosome-wide average SNP density of 4.79 SNPs per Mb. Consequently, the chromosome-wide average distance between SNPs was estimated at 0.2Mb (i.e., one SNP every 200Kb), with a range of 0.16Mb on chromosome VU07 to 0.26Mb on chromosome VU05 (**Supplementary Table 3**).

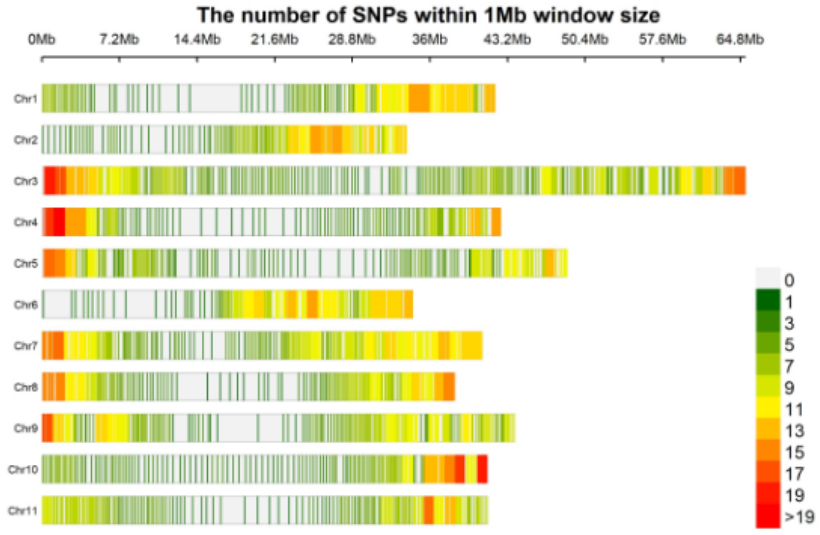


Figure 2 Chromosome-wide distribution the *Cowpea mid-density genotyping panel V1.0.* along the 11 chromosomes of cowpea (*Vigna unguiculata*). Color code key indicates regions of high-density SNPs (red) and low-density SNPs (gray). Data on Number of SNPs per chromosome is provided in supplementary Table 2.

Further, based on nucleotide diversity analysis, the *Cowpea DArTag panel* was able to clearly scrutinize the degree of polymorphism within the four cowpea genetic populations used in this study. The distributions of genetic diversity within the genetic populations and in the entire population based on Nei's nucleotide diversity (PiPerBP) and Tajima's diversity (TajimaD) is presented in **Figure 3.** Chromosome-wide distribution of nucleotide diversity revealed differences in diversity along chromosomes but with a clear separation among the four genetic populations **(Figure 3)**. Mean nucleotide diversity was high in the bi-parental RILs (PiPerBP= 0.47, TajimaD = 4.58), followed by the breeding lines (PiPerBP= 0.42, TajimaD = 3.96), multi-parental RILs (PiPerBP= 0.38, TajimaD = 3.10) and the accessions (PiPerBP= 0.38, TajimaD = 1.34), while the mean diversity remained high (PiPerBP= 0.41, TajimaD = 4.74) when all populations were considered together **(Figure 3, Supplementary Table 4).**

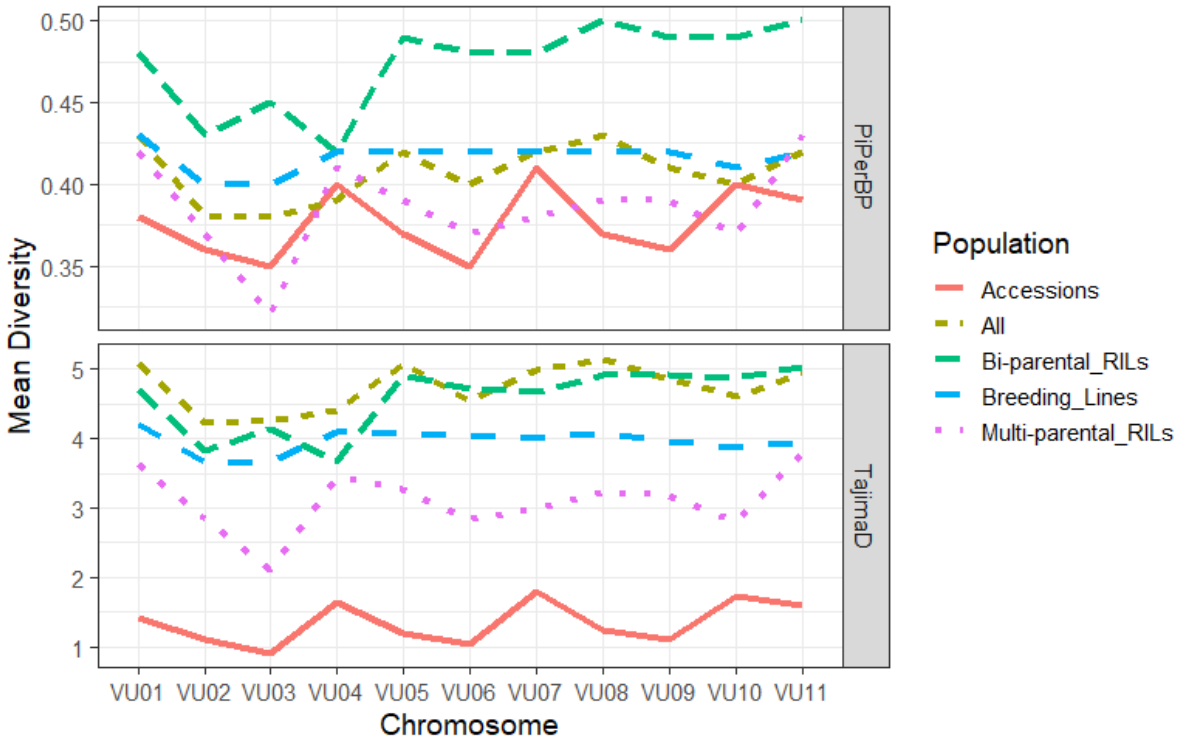


Figure 3 Genetic diversity in the population. The figure displays a chromosome-wide distribution of nucleotide diversity within four genetic populations and in the entire population. Diversity was measured using Nei's nucleotide diversity (PiPerBP) and Tajima's diversity (TajimaD).

Genetic relationships

Pairwise Identity-by- descent (IBD) analysis was able to uncover the relation-

ships among individuals in the populations studied. **Figure 4A** shows the (k_0, k_1)-plot for the pairs of cowpea lines, depicting the IBD relationship based on the Maximum Likelihood Estimation (MLE) method. Of the 54,285 pairs investigated, 4 pairs exhibited parent-offspring (PO) relationship (k_0=0, k_1=1), 65 pairs had half-sib (HS) relationships (k_0=0.5, k_1=0.5) while 24 pairs had identical relationships (k_0=0, k_1=0). The next category consisted of pairs that were unrelated (k_0=1, k_1=0) having 21,276 pairs of lines while the last category involved either complex or distant relationships (**Figure 4A**). The list of actual cowpea genotypes involved in this relationship and summary statistics are provided in **Supplementary Table 5**.

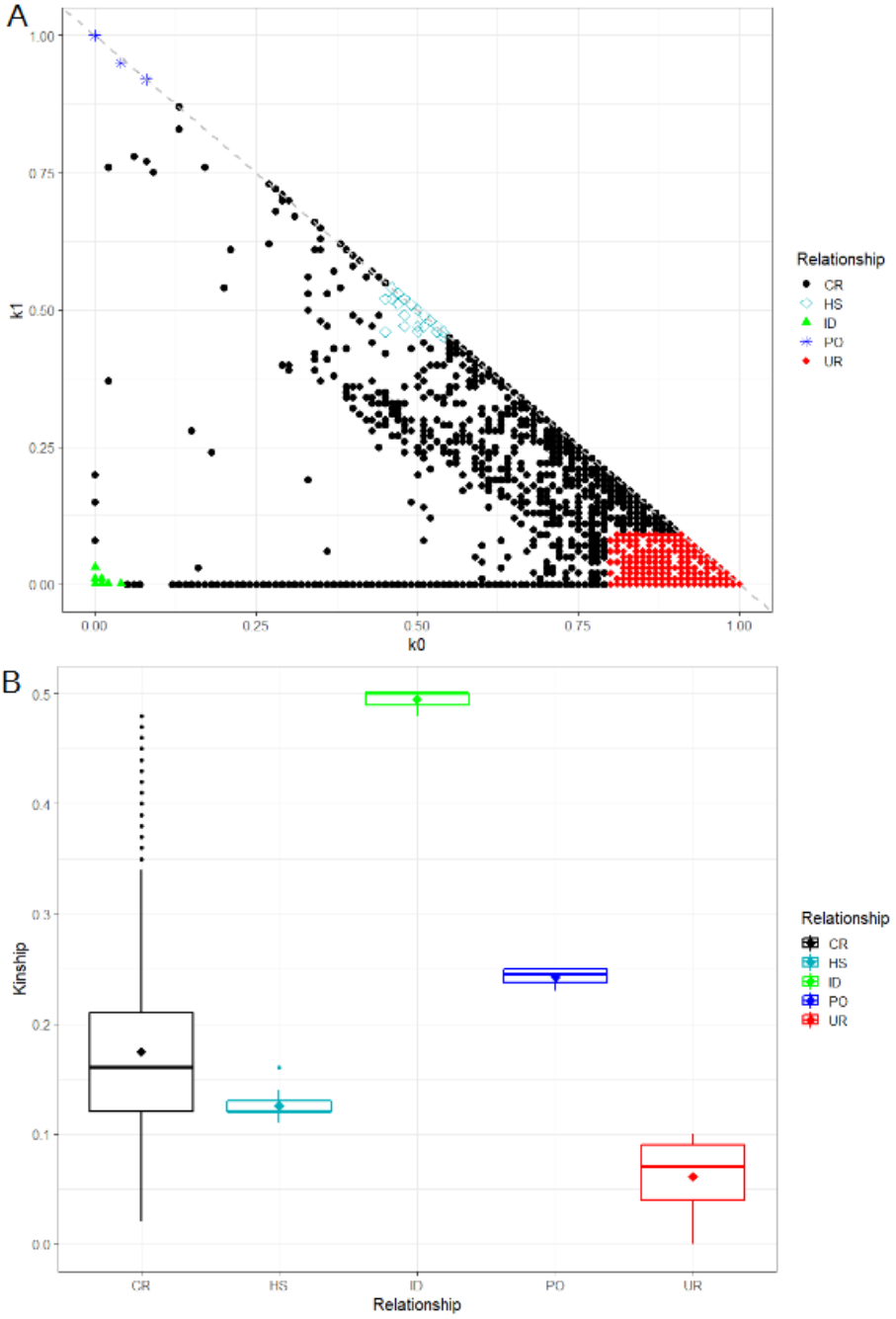


Figure 4 Genetic relatedness among cowpea genotypes in the entire population. (A) Plot of K1 versus K0 for Identity by descent (IBD) based on the Maximum Likelihood Estimation (MLE) method. (B) Distribution of relationships with

respect to kinship coefficient. The relationships in both A and B are explained by IBD coefficients K1 and K0, where, PO (parent-offspring) relationship is defined by IBD coefficients k0=0 and k1=1; Hs (Half-sib) relationship is defined by IBD coefficients k0=0.5, k1=0.5; ID (Identical or duplicate) is defined by IBD coefficients k0=0, k1=0; UR (Unrelated) is defined by IBD coefficients k0=1, k1=0 and the last group had complex (CR) or distant relationships.

The four pairs of cowpea genotypes that showed PO relationship were: (i) IT00K-1263 vs IT97K-556-43, (ii) IT00K-1263 vs IT97K-556-61, (iii) IT00K-1263 vs MAGIC350, and (iv) MAGIC110 vs MAGIC183, while the example of lines detected as being identical included: IT99K-573-1-1 vs IT99K-573-1-2, IT97K-568-11 vs IT97K-568-191, MAGIC345 vs MAGIC346 which are known sister lines. Boxplots of kinship estimator shown in **Figure 4B** clearly depict the differences in the median, mean and variance of each relationship category, with the identical pairs having the highest median and mean kinship, sharply contrasting the unrelated pairs. The complex relationships category had the highest dispersion but with the distribution shifted towards low kinship. The half-sib (HS) and parent-offspring (PO) relationships were also clearly distinguishable and falling within the expected respective kinship ranges. Based on this IBD output, one of each identical pair was randomly discarded before performing other downstream analysis as these would be considered duplicates.

Population structure and differentiation

Population structure was assessed using STRUCTURE software. The most probable number of sub-groupings when all the 330 genotypes were considered together was K=2 as depicted by the *Delta K* vs *K* plot **(Figure 5A)**, complemented by a clear group assignment depicted by the STRUCTURE bar plots in **Figure 5B.** It is evident from these bar plots that the bi-parental RILs were assigned to one group while the rest of the genotypes form a second large and diverse group. Further investigation based on probability of group assignment revealed that sub-group one was made up of 30% of the total population and out of this, 97% were purely the bi-parental RILs, the remaining 3% of group one consisted of 2 breeding lines and 1 multi-parental line **(Supplementary Table 6)**.

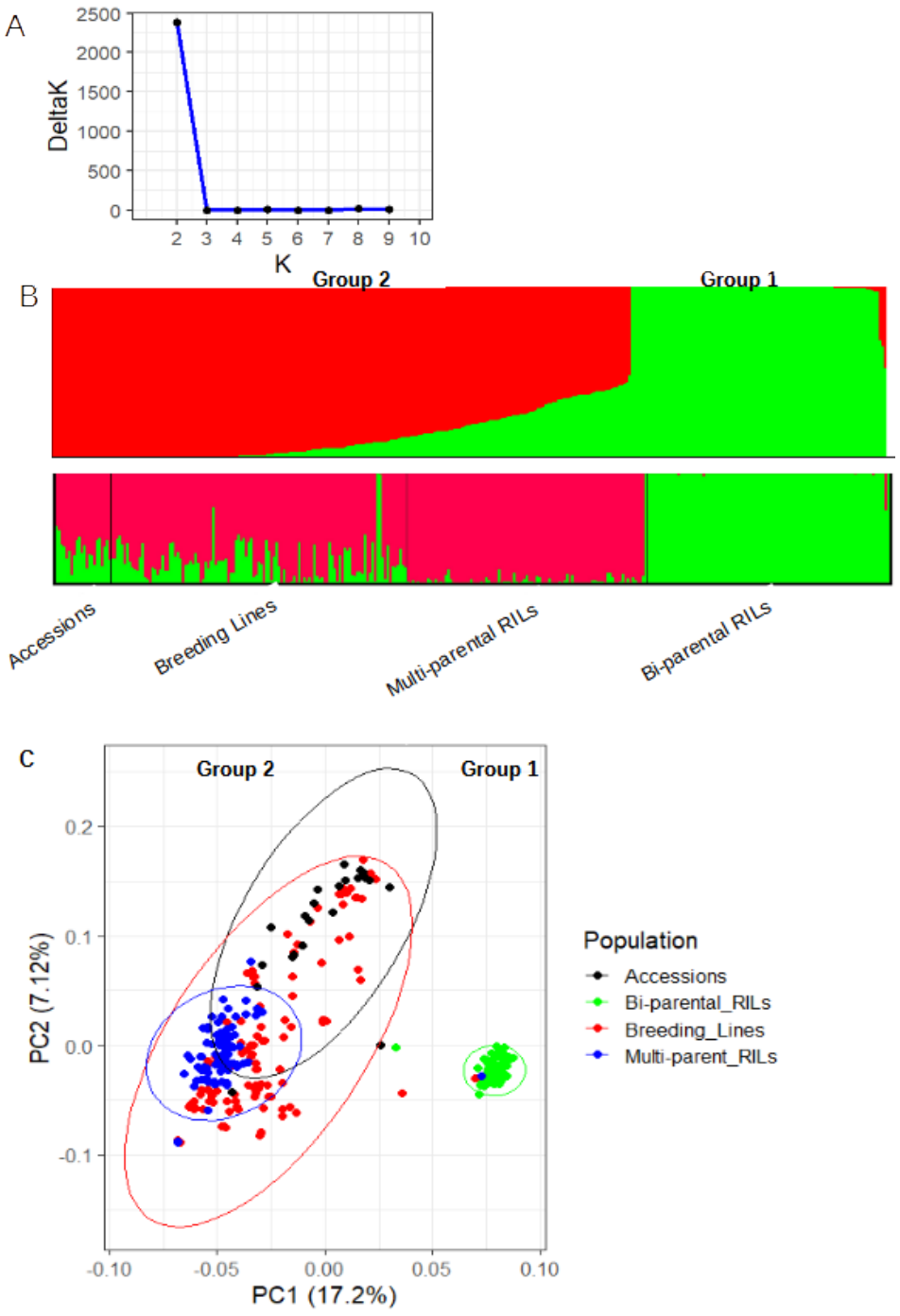


Figure 5 Population structure of the 330 cowpea genotypes constituted from a

sample of lines coming from four different genetic backgrounds. (A) Plot of *K versus DeltaK* showing the most probable number of sub-groupings (K=2). (B) STRUCTURE bar plots depicting two major gene pools in the constituted population, with bi-parental recombinant inbred lines (RILs) forming one group and the other three genetic populations forming one bigger diverse group. (C) Principal component analysis (PCA) displaying the scattering of genotypes along X (PC1) and Y(PC2) axes.

The second sub-group constituted 59% of the population and it contained about 90 breeding lines, 94 multi-parental RILs and 11 accessions. The remaining 11% were those that were categorized as admixed and were made up of 11 accessions and 26 breeding lines **(Supplementary Table 6)**. A heatmap showing the relationship among the four genetic populations and how they fit in the STRUCTURE inferred groups is presented in **Supplementary Figure 1.** The heatmap overlaid the four genetic populations on the group assignments inferred by STRUCTURE software, revealing that all bi-parent RILs belong to group 1 of the STRUCTURE inference while group 2 harbors the remaining three populations, supporting the observation of two major groups in this constituted population. Similar group assignments were depicted by PCA, substantiating the STRUCTURE results (**Figure 5C**). PCA further revealed that within group 2, the breeding lines and accessions were the most scattered while multi-parental lines were closer together (**Figure 5C**). After excluding the bi-parental RILs STRUCTURE analysis revealed two sub-groups within the remaining genetic populations (**Supplementary Figure 2**). Group 1 had a total of 40 lines, 58% of which were breeding lines, 43% were accessions while multi-parental RILs had zero membership in this group. Group 2 was the largest with 158 lines, 52% being multi-parental RILs, 47% were the breeding lines while only 1 accession was a member of this group. A total of 36 lines were categorized as admixed since they had almost equal probability of belonging to both groups.

To gauge how well the SNP panel can discern the differentiation between and within populations, we computed the pairwise genetic distances between populations, followed by an analysis of molecular variance (AMOVA). There was clear differentiation among both the genetic populations and groupings inferred based on STRUCTURE analysis. Genetic distance and differentiation measures ranged from Dist.=8.38, $F_{ST} = 0.06$, and $G_{ST} = 0.04$ to Dist.=22.26, F_{ST} =0.41, and $G_{ST} = 0.27$, with the low values recorded among breeding lines, multi-parental RILs and accessions while high values were registered when bi-parental RILs were compared with the rest of the genetic populations **(Table 2)**. The above observations also reflected the outcome of genetic distances among the groups inferred from STRUCTURE analysis. This implies that genetic distance between Group 1 (which is composed mostly of bi-parental RILs) and Group 2 was higher (Euclidean Dist =21) compared to the distance between Group 2 versus the Admixed group (Euclidean Dist = 16.05) **(Table 2)**. Pairwise gene flow (N_m) among the four genetic populations ranged from N_m =0.36 (bi-parental RILs vs accessions) to N_m =3.92 (breeding lines vs multi-parent RILs) and mean of N_m =1.3 **(Table 2)**. The gene flow estimate between groups in-

ferred by STRUCTURE varied from $N_m = 0.46$ (Group 1 vs Admixed) to N_m =1.07 (Group 2 vs Admixed), with a mean of $N_m = 0.67$. Overall, low gene flow estimates were registered between bi-parental RILs and all other genetic populations, a pattern that interestingly corresponded with high differentiation measures **(Table 2)**.

Table 2 Pairwise genetic distance and differentiation between four genetic populations of cowpea and between groups inferred from STRUCTURE analysis

Genetic population						
Comparison	Ec.Dist[a]	Gst[b]	FstLB[c]	FstUB[d]	Fst[e]	N_m
Bi-parental_RILs vs Accessions	20.95	0.24	0.39	0.42	0.41	0.36
Bi-parental_RILs vs Breeding Lines	19.36	0.20	0.29	0.31	0.30	0.58
Bi-parental_RILs vs Multi-parent_RILs	22.26	0.27	0.39	0.41	0.40	0.38
Breeding_Lines vs Accessions	13.56	0.11	0.12	0.13	0.13	1.67
Breeding_Lines vs Multi-parent_RILs	8.38	0.04	0.05	0.06	0.06	3.92
Multi-parent_RILs vs Accessions	16.89	0.17	0.21	0.22	0.22	0.89
Number of populations	4.00					
Average no. of genotypes per population	82.50					
Number of loci	2753					
Minimum	8.38	0.04	0.05	0.06	0.06	0.36
Maximum	22.26	0.27	0.39	0.42	0.41	3.92
Average	16.90	0.17	0.24	0.26	0.25	1.30
STRUCTURE Inferred grouping						
Comparison	Ec.Dist[a]	Gst[b]	FstLB[c]	FstUB[d]	Fst	Nm
Group 1 vs Group 2	21.00	0.24	0.32	0.35	0.34	0.49
Group 1 vs Admixed	19.10	0.19	0.34	0.37	0.35	0.46
Group 2 vs Admixed	16.05	0.15	0.19	0.2	0.19	1.07
Number of populations	3.00					
Average no. of genotypes per population	110.00					
Number of loci	2753					
Minimum	16.05	0.15	0.19	0.20	0.19	0.46
Maximum	21.00	0.24	0.34	0.37	0.35	1.07
Average	18.71	0.19	0.28	0.31	0.29	0.67

[a]Euclidean genetic distance; [b]Nei's differentiation measure; [c]Lower bound confidence interval

[d]upper bound confidence interval; [e]Wright's differentiation measure; N_m is the gene flow between populations, calculated as $N_m = 0.25(1 - F_{\mathrm{ST}})/F_{\mathrm{ST}}$

AMOVA **(Table 3)** revealed significant genetic variations within the genetic populations (P=0.01) while variation between the populations was not significant (P =0.18). The overall variation among genotypes across all four populations was highly significant (P=0.01). Variation within populations accounted for 16.64% of total variation while that between populations accounted for only

8.38%, while variability among all genotypes across populations accounted for 74.98%. Population differentiation statistic (phi) was similarly higher within population (phi = 0.18) and among genotypes across all populations (phi= 0.25) compared to between population variations (phi= 0.08). Considering inferred groups, variation within genotypes across groups explained 70% of total variation, followed by variance between groups (24%) and lastly variance within the groups (5.4%).

Table 3 Analysis of molecular variance (AMOVA) showing variation within and between cowpea populations

Biological population							
Source of variation	Df	SS	MS	Sigma	%Var	Phi	P-value
Between-Populations	3	56010.49	18670.16	87.72	8.38	0.08	0.18
Within-Populations	5	13912.60	2782.52	174.17	16.64	0.18	0.01
Within-genotypes	321	251915.73	784.78	784.78	74.98	0.25	0.01
Total	329	321838.82	978.23	1046.67	100.00		
Inferred group							
Between-groups	2	52677.83	26338.91	246.07	24.00	0.24	0.01
Within-groups	6	11555.42	1925.90	55.39	5.40	0.07	0.01
Within-genotypes	321	232406.04	724.01	724.01	70.60	0.29	0.01
Total	329	296639.29	901.64	1025.46	100.00		

Discriminant Analysis of Principal Components (DAPC) **(Figure 6A)** further depicted clear differentiation between the populations. Bi-parental RILs were clearly distant from the rest of the populations and DAPC decoded a clear separation between the multi-parent RILs and breeding lines, while the accessions remained closer to the breeding lines. A further investigation of discriminant clusters revealed that breeding lines were the most stratified and diverse, followed by the accessions, multi-parent RILs and the bi-parental RILs were the least structured **(Figure 6B)**. Bayesian information criteria (BIC) plot from DAPC analysis detected six clusters in the population which were used in determining the extent of the structure within each genetic population **(Figure 6C)**.

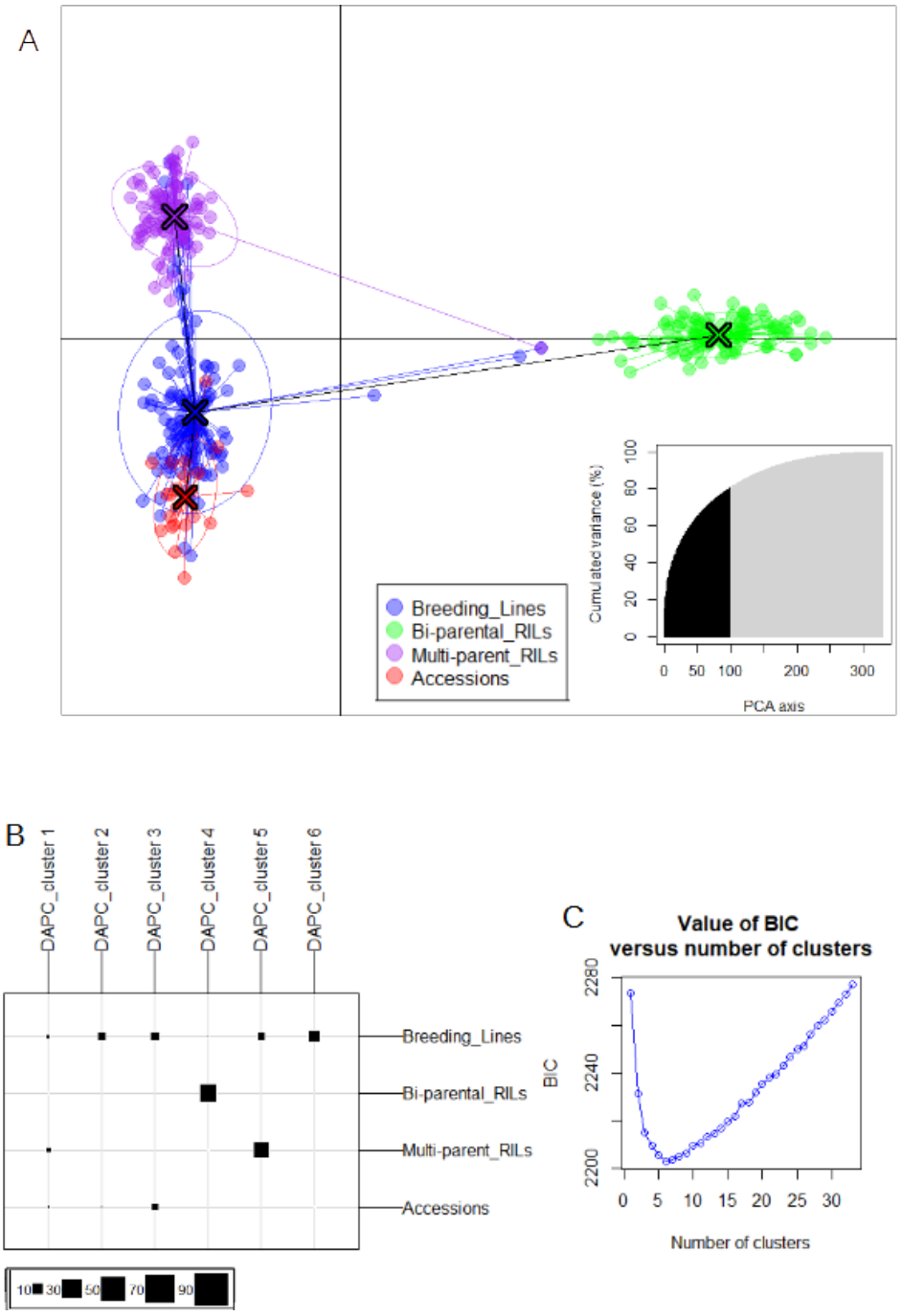


Figure 6 Discriminant Analysis of Principal Components (DAPC). (A) DAPC depicts a clear differentiation between bi-parental RILs and the other three pop-

ulations but precisely separates multi-parental RILs from breeding lines while accessions remained together with the breeding lines (B) DAPC clustering showing the extent of structure within each genetic population: breeding lines being the most structured having multiple black blocks, and bi-parental RILs being the least structured, having just one black block (C) Bayesian information criteria (BIC) versus number of clusters, depicting appropriate number of clusters used in determining extent of structure within each genetic population. The optimal number of clusters correspond to the lowest point of the curve where BIC is at its lowest value.

Linkage disequilibrium decay and GWAS

To elucidate potential use of this constituted set of 330 cowpea lines and the DArTag markers for QTL and gene discovery, we examined LD decay within each genetic population and in the entire population **(Figure 7)**. First, the marker panel was able to correctly detect the rate of LD decay within each of the four genetic populations with bi-parental RILs registering the slowest decay rate followed by multi-parental RILs, while breeding lines and accessions displayed the fastest LD decay rates **(Figure 7A)**. Chromosome-wide LD decay for the entire population showed variable LD decay rates on each chromosome, and when averaged across the genome, LD in the entire 330 cowpea genotypes decayed down to 0.1 at 1.25 Mb **(Figure 7B)**.

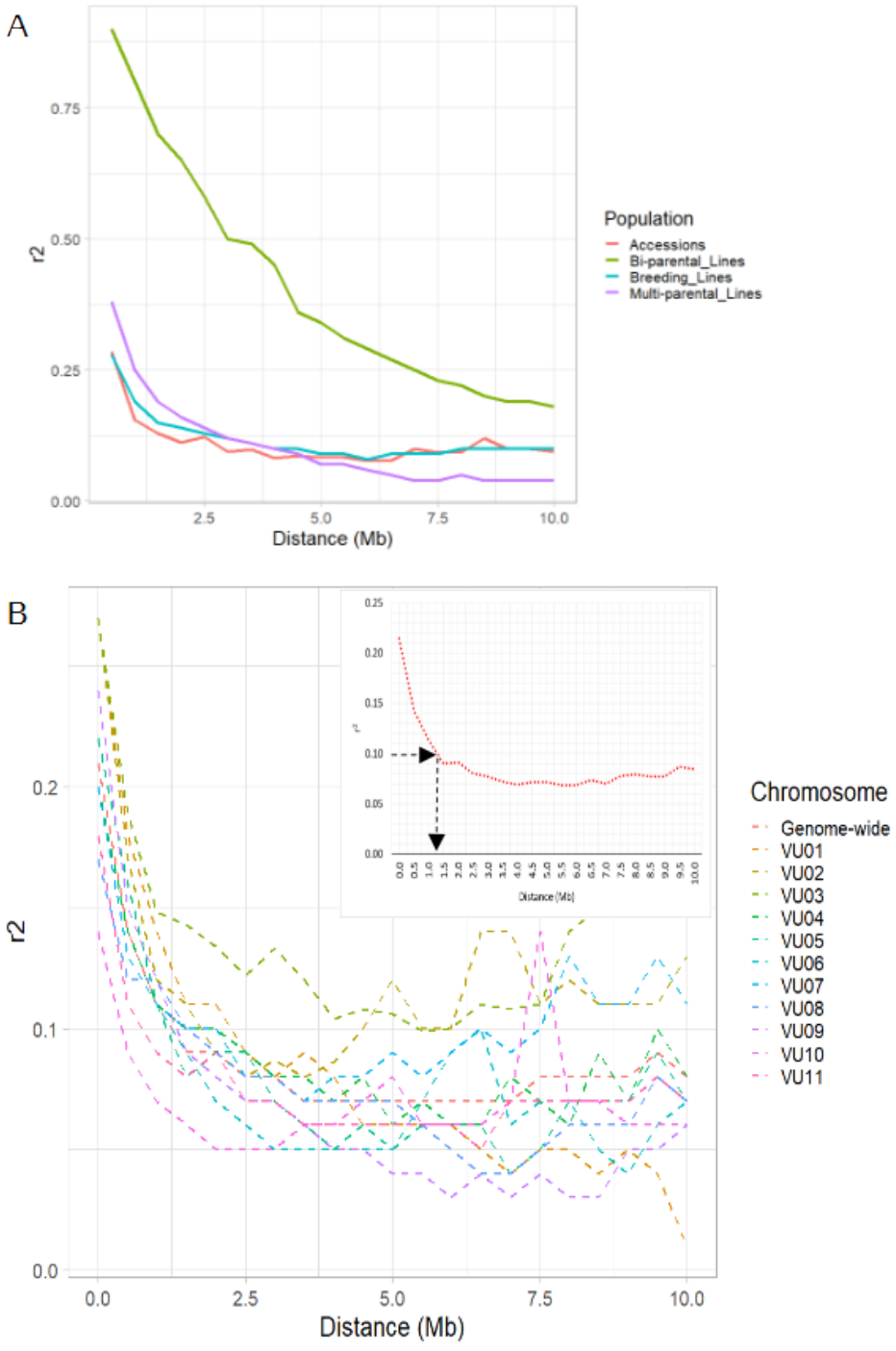


Figure 7 Linkage disequilibrium decay (LD) in a constituted population of cow-

pea. (A) LD decay within four genetic populations of cowpea: bi-parental RILs (slowest decay rate), multi-parental RILs, breeding lines and accessions (fastest decay rate) (B) Chromosome-wide LD decay showing dissipation of LD along each of the 11 cowpea chromosomes; insert on the top right corners displays genome-wide LD decay at $r^2 = 0.1$ within 1.25Mb between pairs of markers. For both figures, the X-axis is the LD measure based on correlation coefficient r^2 and Y-axis is physical distance (Mb).

To further test the gene mapping capability of the *Cowpea mid-density genotyping panel V1.0.,* GWAS was conducted, and it identified significant association signals for seed color and flower color, spanning the same genomic region on chromosome VU07 **(Figure 8)**.

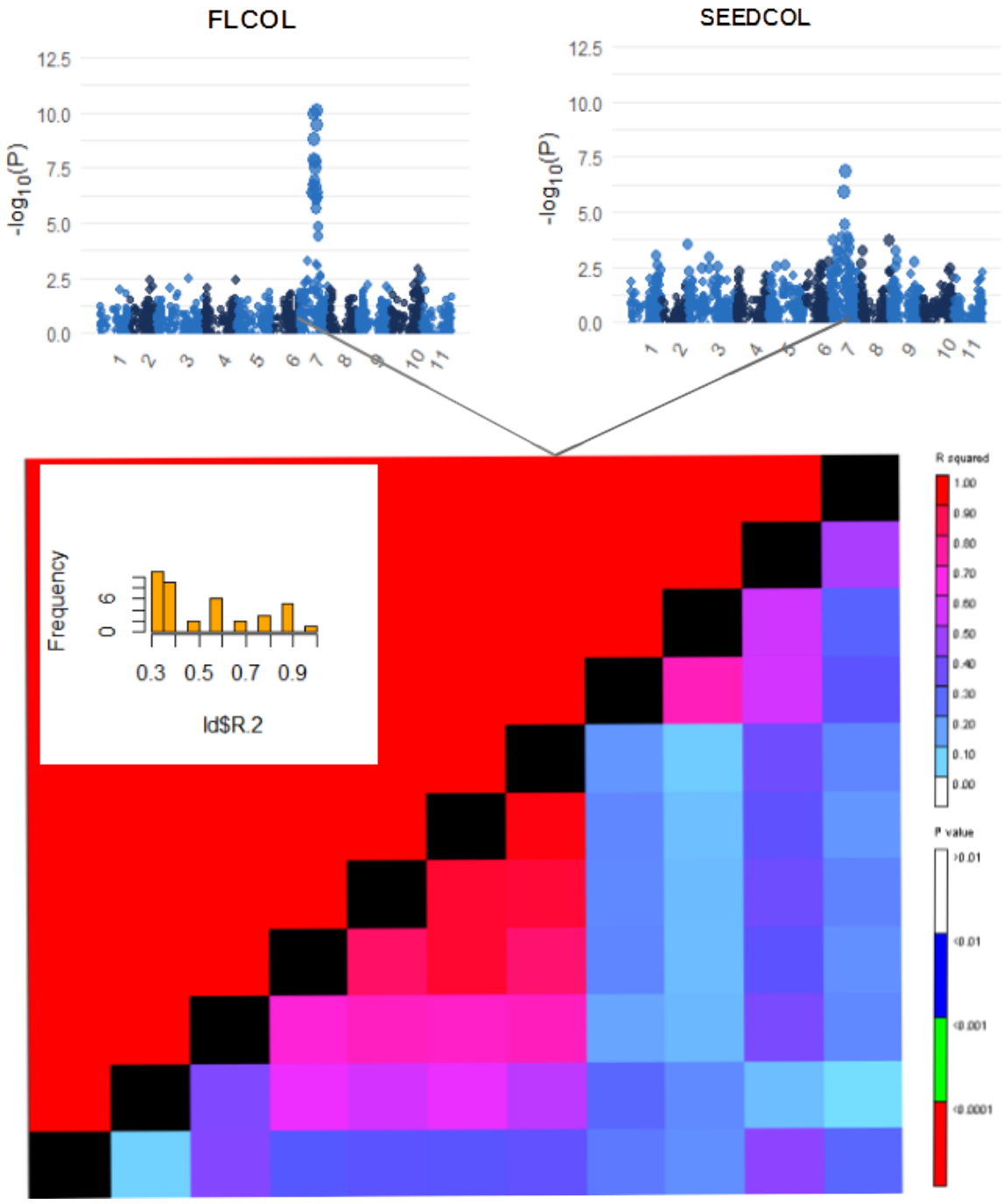


Figure 8 Genome-wide association signals for flower color and seed color traits in cowpea. Linkage disequilibrium matrix is presented below the Manhattan

plots depicting the pattern of pairwise correlation between significant SNPs spanning the signal region on chromosome VU07. Insert on the top left corner of LD plot is a histogram depicting the distribution of LD (r^2) values for SNPs within this region.

The mapped region harbored a total of 26 significant SNPs, three of which overlapped for both flower and seed color (**Table 4**). These SNPs displayed moderate to high linkage disequilibrium and pairwise LD (r^2) ranged from 0.3 to 1.0 with a mean of 0.5 **(Figure 8).** The peak SNPs 2_34565 and 2_06783 on chromosome VU07 explained 18% and 12% of variation in flower color and seed color respectively **(Table 4).** These flower and seed color association signals spanned a region harboring several model genes. One of the genes in this region is *Vigun07g110700* which is a basic helix-loop-helix (bHLH) DNA-binding superfamily protein known to be involved in pigment regulation. The list of all the genes within the association signal region is provided in **Supplementary Table 7**.

Table 4 Significant SNP markers associated with flower and seed colors in cowpea on chromosome VU07

Trait	Marker ID	Chromosome	Pos(bp)	-Log10(p)	PVAR
FLCOL	2_34565	VU07	23,705,735	10.14	18%
	2_47670	VU07	20,629,436	9.94	18%
	2_47424	VU07	24,060,891	9.45	28%
	2_17108	VU07	20,465,839	8.82	16%
	2_01670	VU07	20,808,628	7.87	14%
	2_19077	VU07	22,712,648	7.77	14%
	2_18459	VU07	21174521	7.55	13%
	2_43619	VU07	22202114	6.95	12%
	2_55172	VU07	20295282	6.74	12%
	*2_06783**	VU07	19694195	6.70	12%
	2_12758	VU07	23341686	6.65	12%
	2_13172	VU07	25261206	6.38	11%
	*2_47143**	VU07	17922038	6.35	11%
	*2_51319**	VU07	19490375	6.32	11%
	2_14370	VU07	25158752	6.18	11%
	2_03953	VU07	24658559	6.13	11%
	2_12882	VU07	24521329	6.10	11%
	2_03283	VU07	23239607	5.66	10%
	2_04843	VU07	25438449	4.89	8%
	2_09527	VU07	26085154	4.44	8%
SEEDCOL	*2_06783**	VU07	19,694,195	6.86	12%
	*2_47143**	VU07	17,922,038	5.96	10%
	*2_51319**	VU07	19,490,375	4.44	8%
	2_54172	VU07	15,774,834	3.84	7%
	2_12490	VU07	24,420,693	3.83	7%

Trait	Marker ID	Chromosome	Pos(bp)	-Log10(p)	PVAR
	2_19423	VU07	25,863,610	3.74	6%

[a]Flower color; [b]Seed color; *SNPs significantly associated with both flower and seed color, PVAR refers to percent variance explained by the SNP

Quality control application in breeding

We assessed the ability of the SNP panel for QC/QA application in cowpea breeding by testing how well the marker panel can discern contamination and/or the level of genetic purity among cowpea genotypes. The *Cowpea mid-density genotyping panel V1.0.* was able to clearly identify highly heterozygous individuals from the essentially homozygous others in each population (**Figure 9**). As expected, heterozygosity distribution showed F_1s to be skewed towards the highest proportion relative to the other categories **(Figure 9, Supplementary Figure 3).**

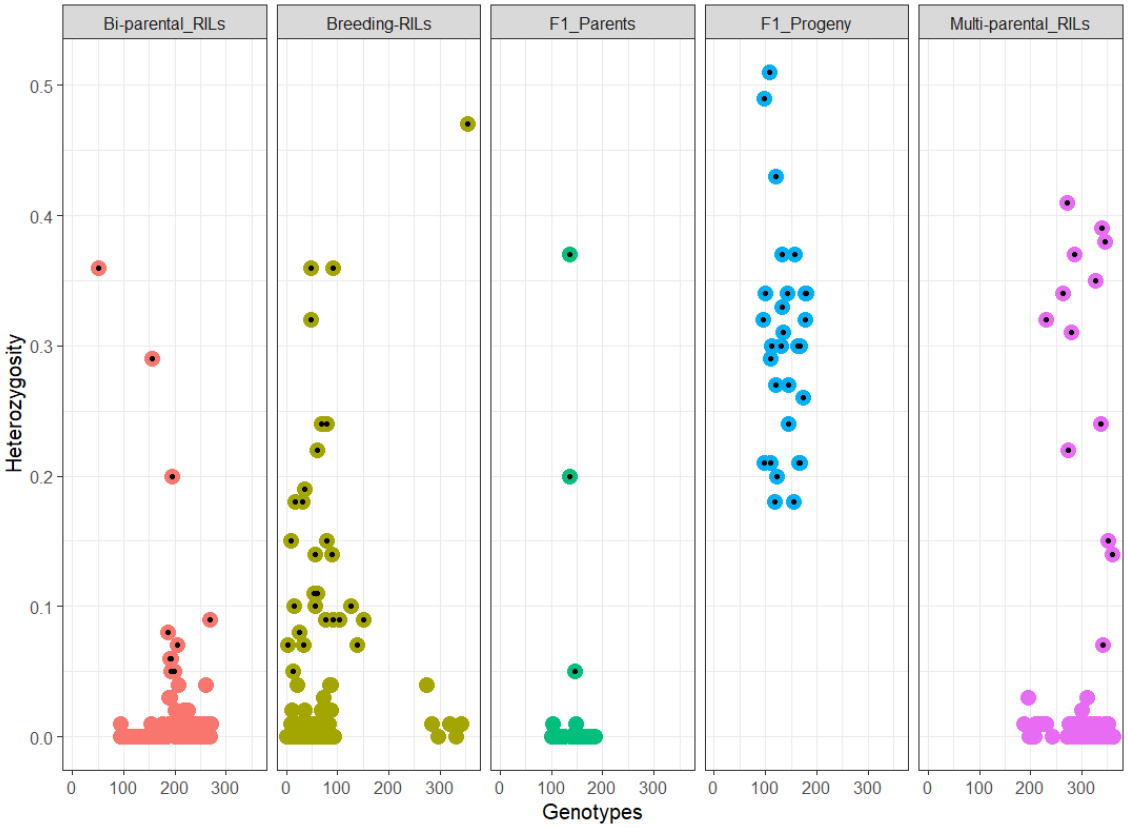


Figure 9 Distribution of the proportion of heterozygosity within five groups of cowpea genotypes. The dot plots depicting the percentage of genotypes in each

group with heterozygosity level above 0.05. The overlaid black dots represent individuals whose heterozygosity level exceed 0.05.

The other four categories (Parents, breeding lines, Bi-parental and Multi-parental RILs) exhibited low proportion of heterozygosity, however, there were outliers representing individuals that were heterozygous in these categories. Individuals detected as being heterozygous are presented in **Figure 9B**, where 100% of the F_1 progenies were above the heterozygosity threshold of 0.05. Among the categories that are expected to be highly homozygous and homogenous, the bi-parental RILs had the lowest percent (10%) of individuals with heterozygosity above 0.05, followed by the parents of F_1 progenies (11%), multi-parental RILs (14%) and breeding lines (26%).

Further, a cluster analysis based on 2,163 filtered SNPs revealed high genetic diversity among the parents of the 30 F_1 progenies, with the parents being placed into three clusters (**Supplementary Figure 4**). Strikingly, two known IITA sister lines IT99K-573-1-1 and IT99K-573-2-1 were grouped closely together in cluster III. One hundred ninety-one (191) SNPs were 70% polymorphic between the 30 parental pairs. A further, 742 SNPs had intermediate polymorphism (50-60%) between the parental pairs while the rest of the SNPs were less than 50% polymorphic **(Supplementary Table 8)**. The lowest proportion (16%) of polymorphic markers was registered between parents IT15K-2241-2 and IT99K-573-2-1 while the highest (61%) was recorded between IT04K-267-8 and SANZI (**Supplementary Figure 5)**. Using polymorphic SNPs only, the levels of hybridity of the 30 F_1 progenies were assessed, and the distribution is presented in **Supplementary Figure 6**. Hybridity was found to range from 23% in a cross of IT15K-2241-2 x IT99K-573-2-1 to 97% in a cross of IT97K-568-11 x IT90K-76. Overall, 40% of the F_1s had hybridity 70%, while 57% had intermediate hybridity (30-60%) and 3% had hybridity below 30% (**Supplementary Figure 6; Supplementary Table 9).**

Discussion

Crop improvement through breeding has been the major tool to lift people out of poverty and to increase global food supply. With the projected population pressure and climate change threats, breeding must be done in a more innovative and precise way to meet the global demand for food security. This has triggered attention towards ground-breaking crop manipulation approaches in the struggle towards achieving sustained increase in genetic gain. Developing and mining crop genetic and genomic resources play crucial roles in enhancing genetic gain through maximization of diversity and discovery of molecular tools that will accelerate breeding for traits of economic importance. Such efforts, in cowpea, have led to the development of excellent genetic resources including over 15,000 gene bank accessions (Fatokun et al., 2018), cowpea MAGIC population (Huynh et al., 2018), mini-core populations (Fatokun et al., 2018; Muñoz-Amatriaín et al., 2021) in addition to elite breeding lines from breeding programs. To best utilize these germplasm resources, molecular platforms have also been developed including but not limited to GoldenGate assay (Muchero et al., 2009) and Cow-

pea iSelect Consortium Array (Muñoz-Amatriaín et al., 2017) that have been used as genotyping platforms by the cowpea research community. Despite these resources, routine application of genomics in cowpea breeding is still limited and this is partly attributable to the relatively high cost of existing high density genotyping platforms. This calls for the development of cost-effective platforms that can be utilized by breeders in the developing world. The *Cowpea mid-density genotyping panel V1.0.* described here has 2,602 SNPs, custom-designed from 51,128-SNP cowpea iSelect Consortium Array. It is a medium-density marker panel which is cost effective coupled with reduced bioinformatics load. The SNPs were selected by considering iSelect missing data rate less than 5%, iSelect data MnAF > 0.2, and an even spacing along the genetic linkage map (https://excellenceinbreeding.org/toolbox/services/cowpea-mid-density-genotyping-services). In the present study, we dissected the properties of the marker panel and its potential utility in cowpea genetic improvement.

A total of 376 cowpea genetic materials were genotyped and after filtering, 362 genotypes were retained. The cowpea population was constituted from groups of materials coming from different genetic backgrounds viz, 123 elite breeding lines, 22 germplasm accessions, 100 multi-parental RILs, 101 bi-parental RILs and 30 F_1 progenies. These were purposely included to test the sensitivity of the panel to heterozygosity and contamination detection in the genetic materials. The choice of genetic populations included in this study was deliberately made to enable accurate assessment of the SNP panel performance, given the expected genetic features of these populations. Our intention was to evaluate how well the panel would decipher the unique genetic parameters within the genetic groups and across the entire population.

Informativeness of the marker panel

We started by examining how informative the marker panel is by looking at the distribution of minor allele frequency in the entire population. About 78% of the SNP markers in the panel (1,882 SNPs) had MnAF above 0.2, with the remaining 22% SNPs still having MnAF above 0.05. Minor allele frequency is widely used in population genetics studies because it provides information to differentiate between common and rare variants in the population (Uemoto et al., 2015; Dussault & Boulding, 2018). It also determines allele diversity and heritability in the population, and it has been shown that markers with high MnAF have high resolution power and are good in detecting quantitative trait loci (QTL) (Uemoto et al., 2015; Dussault & Boulding, 2018). The moderate to high MnAF observed in the present study suggested that the cowpea mid-density SNP panel is informative, making it a useful genetic resource for the cowpea scientific community. The raw data of cowpea mid-density DArTag SNP genotypes have been provided in **Supplementary Table 9**.

When we scrutinized SNP distributions on each chromosome, even coverage of markers was depicted with an average density of 203 SNPs per chromosome, and approximately one SNP every 200kb. This marker density and distribution is modest for dissection of molecular diversity, genetic relatedness, population

structure, linkage disequilibrium, genomic selection and even a medium resolution QTL discovery. Marker density of less than 5,000 SNPs that are well distributed across the genome have been deployed successfully to decipher genetic diversity and other molecular and genetic applications in crops (Akohoue et al., 2020; Chander et al., 2021; Gbedevi et al., 2021).

Using two nucleotide diversity measures; Nei's diversity (PiPerBP) (Nei & Li, 1979), and Tajima's diversity (TajimaD) (Tajima, 1989), the SNP panel was further able to discern the genetic diversity within four genetic populations of cowpea and across the entire set after excluding the F_1 progenies. Nucleotide diversity is a molecular genetics concept which is used to measure the degree of polymorphism within a population and the mathematical model was first developed by (Nei & Li, 1979). It measures the average number of nucleotide differences per site between two DNA sequences in all possible pairs in the sample population, and hence, it is a measure of population genetic variation (Kilian et al., 2007). Interestingly, our study showed nucleotide diversity to be high within the bi-parental RILs and was a bit reduced within the accessions. The 101 bi-parental RILs included as part of the bigger population used in this study were derived from a cross between a wild cowpea TVNu1158 and an elite breeding line IT99K-573-1-1. The wide genetic distance between the parents of this bi-parental RILs explains the observed high nucleotide diversity. It is known that wide hybridization or introgression from previously isolated populations increase genetic diversity (Bhandari et al., 2017). On the other hand, we observed unexpectedly low genetic diversity within the tested accessions. It should be noted that there were 22 accessions included in this study and these were sampled from the IITA cowpea mini-core population which is part of a world collection of cowpea germplasm (Fatokun et al., 2018). The observed reduced nucleotide diversity within the accessions could possibly be due to the small sample size used. It has been shown that using human and animal populations, allelic diversity measures such as Nei's and Tajima's diversity, depend on the number of alleles per locus and the population size (Pruett et al., 2008; Subramanian, 2016). The authors highlighted that at samples of less than 40, the number of alleles per locus is a poor measure of diversity. In four outcrossing conifer tree species and simulated data for self-mating system, allelic richness is reported to gradually increase with increase in sample size (Bashalkhanov et al., 2009). It is interesting that the mid-density SNP panel detected similar patterns in cowpea. When the entire cowpea populations were considered, nucleotide diversity remained reasonably high, suggesting that this constituted population is rich in alleles and could be exploited for traits discovery.

Detection of genetic relatedness

Based on theoretical values of IBD coefficients (Wagner et al., 2006; Galván-Femenía et al., 2021), the co-ancestry among the genetic populations were dissected, where 0.01% had PO relationships, 0.12% were HS, 0.04% were identical and 39% were unrelated while 60.64% had either distant or complex relationships. Estimation of genetic relatedness is fundamental in the study of quanti-

tative traits where the proportion of trait variability explained by shared alleles indicates the strength of the genetic component of the trait (Sethuraman, 2018). In several applied fields, accurate estimation of genetic relatedness is critical. For instance, in a breeding program, identification of genetic relatedness among parents is key in prioritizing crosses that will maximize diversity for eventual increased genetic gain (Ongom et al., 2021). Also, association studies and linkage analyses without accounting for the increased relatedness due to population genetic structure could lead to spurious associations (Pritchard et al., 2000b).

Discerning population structure and differentiation

The genetic structure of a population is defined as a group of individuals sharing a common gene pool, and it determines its capacity to be improved or changed by selection (Hayward & Breese, 1993). Assessing population structure, therefore, is fundamental both in guiding breeding options as well as in association studies leading to traits discoveries. The present study identified two major gene pools in the entire population. Interestingly, all the bi-parental RILs formed one group while the second group harbored a mixture of the breeding lines, accessions and multi-parental RILs. This outcome is expected given that the parentage of breeding lines includes most of the accessions and, the multi-parental lines were also derived from some of the elite breeding lines included in these populations (Huynh et al., 2018). The bi-parental RILs share a wild relative's alleles from TVNu1158, as such, they constituted a unique gene pool. However, when the bi-parental RILs were excluded from the population, two sub-groups were detected within the remaining diverse lines suggesting that this stratification was confounded by the bi-parental RILs.

The DAPC was also able to clearly differentiate among the breeding lines, multi-parental RILs and the accessions even though PCA had depicted these as forming a single diverse group. This was further corroborated by pairwise differentiation measures (F_{ST} and G_{ST}) and Euclidian genetic distance between the four genetic populations, which depicted higher differentiation between the bi-parental RILs and the rest of the groups. This suggested that breeding lines, multi-parental RILs and accessions are genetically similar among themselves but distant from the biparental RILs. These results were also supported by pairwise Euclidean genetic distances and gene flow estimates which revealed the same pattern of genetic relationships among these four populations. Wright's (1951) F_{ST} and Nei's (1973) G_{ST} are statistics that measure the proportion of genetic diversity in a population (Culley et al., 2002). These two statistics are equivalent when there are only two alleles at a locus and, in the case of multiple alleles, G_{ST} is equivalent to the weighted average of F_{ST} for all alleles (Nei, 1973). Consequently, in the present study, the two statistics depicted the same pattern of differentiation among populations. Some past genetic studies in cowpea have used F_{ST} to assess the extent of differentiation between sub-populations. For instance, Gbedevi et al. (2021) reported low to moderate pairwise F_{ST} values in the range of 0.014 to 0.117 and a mean 0.072 among six sub-populations of cowpea accessions grouped by geographic regions in Togo. Using 15 SSR markers,

Sarr et al. (2021) reported genetic differentiation (F_{ST}) to vary from 0.018 to 0.100 among cowpea accessions collected from different regions of Senegal. Average $F_{ST} = 0.075$ was reported among cowpea accessions collected from Ethiopia (Desalegne et al., 2016). Low F_{ST} values (low differentiation) indicate that little variation is proportioned between populations while high values denote that a large amount of variation is found among populations (Culley et al., 2002). The afore listed studies attributed the cause of observed low F_{ST} values to short distances between geographical regions of collection that facilitated easy exchange of genetic materials between regions. Generally, self-pollinated crops tend to have low genetic diversity and it has been observed that differentiation among populations of self-pollinated crops like cowpea are generally low (Desalegne et al., 2016; Wamalwa et al., 2016; Carvalho et al., 2017). In the present study, the observed high differentiation between the bi-parental RILs and the rest of the populations was expected given that one of the parents of the bi-parental RILs is a wild relative, hence, this population has a unique gene pool, which explains why it is highly differentiated from the rest of the genetic populations. Cowpea is reported to have evolved from a few progenitors and it exhibits very limited gene flow between wild and cultivated types (Ba et al., 2004; Pasquet, 2000; Boukar et al., 2020; Pasquet et al., 2021). In fact, gene flow estimates (N_m) in the present study were high among breeding lines, multi-parental RILs and accessions (N_m =0.89 to 3.9) compared to that between bi-parental RILs and the rest of the populations (N_m =0.36 to 0.58). Upon checking pedigree records from our breeding program, it was evident that the parents of most breeding lines came from the accessions, while that of the multi-parental RILs came from the elite breeding lines (Huynh et al., 2018). Indeed, multi-parental RILs and the breeding lines had the highest gene flow (N_m =3.9) and strikingly, the lowest genetic distance (Dist.=8.38) between them, confirming that these two populations share common alleles. These results implied that the *Cowpea mid-density genotyping panel V1.0.* correctly identified the structure in the population and also detected an overall significant genetic variation and differentiation among the genotypes included in the population.

Results of AMOVA further revealed higher variation within the genetic populations than between the populations. A recent study using 255 cowpea accessions collected from six regions in Togo reported significant genetic variations among and within populations, with variations among individuals that were within each of the six geographic origins explaining the highest percentage (78%) of total variability (Gbedevi et al., 2021). Similar studies have been reported by several authors. For instance, using 671 cowpea accessions obtained from 8 regions of Senegal, variance among individuals within the regions accounted for 75% of total variation, followed by variance within accessions (14%) and between populations (11%) (Sarr et al., 2021). Higher variations within population *vis a vis* between populations were also reported in cowpea (Chen et al., 2017; Mafakheri et al., 2017; Xiong et al., 2018; Nkhoma et al., 2020b). The higher genetic variance within populations than between populations have been explained in terms of possible gene flow between populations, through germplasm sharing across

geographic regions (Chen et al., 2017; Mafakheri et al., 2017; Nkhoma et al., 2020b; Gbedevi et al., 2021; Sarr et al., 2021). In the present study, the use of materials from genetic populations in routine breeding must have facilitated gene flow between the four groups, leading to the observed higher genetic variance within than between the groups.

Application in gene discovery

Linkage disequilibrium decay was examined in each of the four genetic populations and in the entire population. LD decay rates varied across chromosomes, with Vu03 and Vu09 showing the lowest and fasted LD decay rate respectively. Recombination frequency, a factor that determines LD decay rate, was found to vary along the 11 chromosomes of cowpea (Lonardi et al., 2019). The pattern of this recombination rate corresponded with the chromosome-wise LD decay rates observed in the present study. Given the known genetic background of members of these populations, it was possible to validate the efficiency of the *Cowpea mid-density genotyping panel V1.0.* in estimating LD decay rate. For instance, LD displayed a slower dissipation in the bi-parental RILs than in the other three genetic populations. This is an expected outcome, given that bi-parental populations are limited in the number of genetic recombination and alleles. On the other hand, the accessions, breeding lines and multi-parental RILs have much higher recombination rates than bi-parental RILs (Ongom & Ejeta, 2018; Stadlmeier et al., 2018), and were correctly depicted to show faster LD decay. Overall, the genome-wide LD decay in the entire population extended to 1.25Mb, that is, the LD between any two markers dissipated when the markers were approximately 1Mb apart. This LD decay is moderate and is typical of self-pollinated crops that have limited chances of recombination from natural out-crossing (Flint-Garcia et al., 2003). In a population of 274 cowpea accessions, using 3,127 SNPs, LD decay rate of 100 kb, smaller than what is observed in the present study, was reported (Sodedji et al., 2021). It should be noted that, in the present study, populations consisted of genetic groups combining both high and low recombination frequency backgrounds, a possible reason for the average genome wide LD decay to extend up to ~1Mb. Nevertheless, the LD decay rate in the present study falls in the ranges reported in the cowpea mini-core population, where chromosome-wide LD varied from 809 kb (~0.8Mb) to 4705kb (~4.6Mb) (Muñoz-Amatriaín et al., 2021). In asparagus bean (*Vigna. unguiculata ssp. sesquipedialis),* a relatively high LD of 1.88 Mb was reported (Xu et al., 2012). Overall, these observations indicate that the LD decay distances are fairly long in autogamous species. In contrast, LD declines very rapidly in allogamous species where physical recombination is more common. For instance, LD decays within only a few kilobases in maize (Yan et al., 2009) and only 200 bp in a wild sunflower population (Liu & Burke, 2006). Despite the observed slow decay rate in cowpea populations, such populations have been used to successfully map quantitative trait loci for several traits (Herniter et al., 2018, 2019; Lo et al., 2018, 2019; Paudel et al., 2021)

A Genome-wide scan in the present study demonstrated the capability of the

panel in QTL mapping. This was evident by mapping a region on chromosome Vu07 which was previously identified to harbor *Vigun07g110700* (bHLH), proposed to regulate pigmentation constriction (Herniter et al., 2019). This region was reported to contain genes regulating both flower and seed color (Herniter et al., 2019; Lo et al., 2019). Prior to these studies, it was reported that a lack of pigment in the flower was often associated with a lack of pigment in the seed coat, suggesting a pleiotropic effect of the gene (Harland, 1920). Our study provided a proof of concept, asserting that the marker panel and the population can be mined by the cowpea breeding community for any trait that shows variation in the population.

Application in quality control

The potential of the marker panel for QC/QA in cowpea was also assessed. Molecular determination of parental purity prior to hybridization and detection of true hybridity are becoming routine in modern breeding programs. Our results showed the *Cowpea mid-density genotyping panel V1.0.* to excellently deduce heterogeneity within different categories of cowpea genotypes. As expected, the F_1 progenies displayed the highest level of heterozygosity implying they were true hybrids. Interestingly, elite breeding lines and RILs showed low proportion of heterozygosity and yet potentially impure individuals were detectable within each category, with the elite breeding lines exhibiting the highest percentage (26%) of individuals that were heterozygous at some loci. The observed high level of heterogeneity among the inbred lines suggested the need to purify these lines prior to using them as parents in the breeding program and further demonstrated that the marker panel is effective in detecting parental purity.

Knowing that the display of heterozygosity among F_1s may not necessarily determine whether they are true hybrids, we assessed the polymorphism of markers between each parental combination of the F_1s and used the polymorphic SNPs to authenticate hybridity. Moderate to high marker polymorphisms were recorded among the parental pairs implying that these sets of polymorphic markers would also delineate the hybridity of F_1 progenies with high accuracy. In fact, we identified over 191 SNPs that had high polymorphism across these parental pairs. Moreover, the parents were also genetically diverse, meaning that these 191 SNPs could be considered as additional QC/QA SNPs for cowpea to the 17 KASP-based previously described for cowpea (Ongom et al., 2021). High degree of hybridity (above 70%) was recorded in more than 40% of the F_1s and about 57% had intermediate hybridity, consequently 97% of the F_1s had moderate to high hybridity and only one F_1 progeny registered hybridity of less than 30%. A previous study on hybridity using 17 KASP-based SNPs detected 79% true F_1s and 14% self-fertilization in a sample of 1,436 F_1 plants (Ongom et al., 2021). Genetic purity of parental lines and hybridity authentication are important quality control criteria in breeding, that directly affect the quality of lines and varieties being developed (Ertiro et al., 2017; Ongom et al., 2021). Our results further demonstrate the effectiveness of this cost-efficient marker panel

for genetic purity assessment and other QC needs in the breeding pipeline.

CONCLUSIONS

This study deployed a total of 2,602 DArTag SNPs in a population of 376 cowpea genetic materials with the objective of deciphering the usefulness of this low cost, medium density marker panel in the analysis of genetic relatedness, diversity assessment, gene discovery and quality control in cowpea breeding. It was discovered that the *Cowpea mid-density genotyping panel V1.0.* contains informative SNPs with high polymorphisms among diverse cowpea lines and modest density of about one SNP after every 200kb. Indeed, this cost-effective mid-density marker panel displayed good potential for genetic diversity assessment, linkage disequilibrium and genome-wide mapping of QTL and potential application as QC markers in the breeding program. The study further unearthed the resourcefulness of the constituted set of cowpea population in terms of high genetic and trait variation which will be exploited to improve this crop. It is hoped that the findings presented here will advance the practice and knowledge of molecular marker deployment to improve economic traits in crop plants, and particularly, the application of genomic-aided breeding in cowpea.

Acknowledgments

This work was supported in part, by the Bill & Melinda Gates Foundation [OPP1198373/INV-009649]. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. Design of the SNP panel by Timothy J. Close fulfilled an objective of NSF IOS-1543963 (Advancing the Cowpea Genome for Food Security). Authors appreciate DArT for supporting targeted genotyping and related bioinformatics. We further, acknowledge Excellence in Breeding (EiB) platform through EiB-INTERTEK-DA agreement services for supporting the genotyping work. We thank all our field technicians, especially Mr. Abdullahi D. Haruna and Ms Safiya Abubakar who supported sample collection and preliminary processing.

Conflict of Interest

The authors declare no conflict of interest

Supplemental Material

Supplementary Table 1. Design sequences *Cowpea mid-density genotyping panel V1.0.*

Supplementary Table 2. Heterozygosity and allele frequencies

Supplementary Table 3. SNP density

Supplementary Table 4. Chromosome-wide Nucleotide diversity

Supplementary Table 5. IBD relationships among cowpea genotypes

Supplemental Table 6. STRUCTURE group assignment

Supplemental Table 7. Candidate genes for seed and flower colors in cowpea

Supplemental Table 8. SNP marker polymorphism scores between 30 pairs of parents

Supplemental Table 9. Scores of Hybridity among 30 F_1 progenies

Supplementary Table 10. SNP data of *Cowpea mid-density genotyping panel V1.0.*

Supplementary Figure 1. Heatmap and dendrogram classifying the 330 cowpea lines

Supplementary Figure 2. Sub-structure depiction after excluding the bi-parental RILs

Supplementary Figure 3. Box plot depicting heterozygosity dispersion within populations.

Supplementary Figure 4. Phylogenetic relationship among parents of the 30 F_1s

Supplementary Figure 5. Distribution of marker polymorphism between parents

Supplementary Figure 6. Distribution of the level of hybridity of the 30 F_1s

data availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files)

Authors' contributions

POO: corresponding author, wrote the paper, organized samples for genotyping and conducted the statistical analysis. CF, OB, AT, ALGO, EHN & TJC: revised the paper. OB, CF and TJC: Assembled the populations and designed the study. TJC, SL and AK designed the SNP panel. ALGO and EHN: provided technical support for genotyping. All authors reviewed and approved the manuscript.

References

Akohoue, F., Achigan-Dako, E.G., Sneller, C., Deynze, A. Van, & Sibiya, J. (2020). Genetic diversity, SNP-Trait associations and genomic selection accuracy in a west African collection of Kersting's groundnut [*Macrotyloma geocarpum*(Harms) Maréchal & Baudet]. *PLoS ONE*, *15*, e0234769. https://doi.org/10.1371/journal.pone.0234769

Bashalkhanov, S., Pandey, M., & Rajora, O.P. (2009). A simple method for estimating genetic diversity in large populations from finite sample sizes. *BMC Genetics*, *10*, 1–10. https://doi.org/10.1186/1471-2156-10-84/FIGURES/5

Bhandari, H., Nishant Bhanu, A., Srivastava, K., Singh, M., & Hemantaranjan, A. (2017). Assessment of genetic diversity in crop plants - an overview. *Advances in Plants & Agriculture Research*, *Volume 7*. https://doi.org/10.15406/APAR.2017.07.00255

Boukar, O., Abberton, M., Oyatomi, O., Togola, A., Tripathi, L., & Fatokun, C. (2020). Introgression breeding in cowpea [*Vigna unguiculata* (L.) Walp.]. *Frontiers in Plant Science*, *11*, 1439. https://doi.org/10.3389/FPLS.2020.567425/BIBTEX

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

Boukar, O., Togola, A., Chamarthi, S., Belko, N., Ishikawa, H., Suzuki, K., & Fatokun, C. (2019). Cowpea [*Vigna unguiculata* (L.) Walp.] breeding. *Advances in Plant Breeding Strategies: Legumes* (pp. 201–243). Springer International Publishing.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., & Buckler, E.S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics (Oxford, England)*, *23*, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Carvalho, M., Muñoz-Amatriaín, M., Castro, I., Lino-Neto, T., Matos, M., Egea-Cortines, M., Rosa, E., Close, T., & Carnide, V. (2017). Genetic diversity and structure of Iberian Peninsula cowpeas compared to world-wide cowpea accessions using high density SNP markers. *BMC Genomics*, *18*, 1–9. https://doi.org/10.1186/S12864-017-4295-0/TABLES/2

Chander, S., Garcia-Oliveira, A.L., Gedil, M., Shah, T., Otusanya, G.O., Asiedu, R., & Chigeza, G. (2021). Genetic diversity and population structure of soybean lines adapted to sub-saharan africa using single nucleotide polymorphism (Snp) markers. *Agronomy*, *11*, 604. https://doi.org/10.3390/agronomy11030604

Chen, H., Chen, H., Hu, L., Wang, L., Wang, S., Wang, M.L., & Cheng, 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. https://doi.org/10.1016/J.CJ.2017.04.002

Culley, T.M., Wallace, L.E., Gengler-Nowak, K.M., & Crawford, D.J. (2002). A comparison of two methods of calculating GST, a genetic measure of population differentiation. *American Journal of Botany*, *89*, 460–465

Desalegne, B.A., Mohammed, S., Dagne, K., & Timko, M.P. (2016). Assessment of genetic diversity in Ethiopian cowpea [*Vigna unguiculata* (L.) Walp.] germplasm using simple sequence repeat markers. *Plant Molecular Biology Reporter 2016 34:5*, *34*, 978–992. https://doi.org/10.1007/S11105-016-0979-X

Dussault, F.M., & Boulding, E.G. (2018). Effect of minor allele frequency on the number of single nucleotide polymorphisms needed for accurate parentage

assignment: A methodology illustrated using Atlantic salmon. *Aquaculture Research*, *49*, 1368–1372. https://doi.org/10.1111/are.13566

Earl, D.A., & vonHoldt, B.M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*. https://doi.org/10.1007/s12686-011-9548-7

Ertiro, B.T., Semagn, K., Das, B., Olsen, M., Labuschagne, M., Worku, M., Wegary, D., Azmach, G., Ogugo, V., Keno, T., Abebe, B., Chibsa, T., & Menkir, A. (2017). Genetic variation and population structure of maize inbred lines adapted to the mid-altitude sub-humid maize agro-ecology of Ethiopia using single nucleotide polymorphic (SNP) markers. *BMC Genomics*, *18*. https://doi.org/10.1186/s12864-017-4173-9

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*. https://doi.org/10.1111/j.1365-294X.2005.02553.x

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. https://doi.org/10.1038/s41598-018-34555-9

Flint-Garcia, S.A., Thornsberry, J.M., & Edward IV, S.B. (2003). Structure of Linkage Disequilibrium in Plants. *http://dx.doi.org/10.1146/annurev.arplant.54.031902.134907*, *54*, 357–374. https://doi.org/10.1146/ANNUREV.ARPLANT.54.031902.134907

Galván-Femenía, I., Barceló-Vidal, C., Sumoy, L., Moreno, V., de Cid, R., & Graffelman, J. (2021). A likelihood ratio approach for identifying three-quarter siblings in genetic databases. *Heredity*, *126*. https://doi.org/10.1038/s41437-020-00392-8

Garcia-Oliveira, A.L., Zate, Z.Z., Olasanmi, B., Boukar, O., Gedil, M., & Fatokun, C. (2020). Genetic dissection of yield associated traits in a cross between cowpea and yard-long bean (*Vigna unguiculata* (L.) Walp.) based on DArT markers. *Journal of Genetics 2020 99:1*, *99*, 1–13. https://doi.org/10.1007/S12041-020-01216-8

Gbedevi, K.M., Boukar, O., Ishikawa, H., Abe, A., Ongom, P.O., Unachukwu, N., Rabbi, I., & Fatokun, C. (2021). Genetic diversity and population structure of cowpea [*Vigna unguiculata* (L.) walp.] germplasm collected from Togo based on DArT markers. *Genes*, *12*, 1451. https://doi.org/10.3390/genes12091451

Harland, S.C. (1920). Inheritance of certain characters in the cowpea (*Vigna sinensis*). II. *Journal of Genetics 1920 10:3*, *10*, 193–205. https://doi.org/10.1007/BF03007981

Hayward, M.D., & Breese, E.L. (1993). Population structure and variability. In M.D. Hayward, N.O. Bosemark, I. Romagosa & M. Cerezo (Eds.), *Plant*

Breeding (pp. 16–29). Springer, Dordrecht.

Herniter, I.A., Lo, R., Muñoz-Amatriaín, M., Lo, S., Guo, Y.-N., Huynh, B.-L., Lucas, M., Jia, Z., Roberts, P.A., Lonardi, S., & Close, T.J. (2019). Seed coat pattern QTL and development in cowpea (*Vigna unguiculata* [L.] Walp.). *Frontiers in Plant Science*, *0*, 1346. https://doi.org/10.3389/FPLS.2019.01346

Herniter, I.A., Muñoz-Amatriaín, M., Lo, S., Guo, Y.N., & Close, T.J. (2018). Identification of candidate genes controlling black seed coat and pod tip color in cowpea (*Vigna unguiculata* [L.] Walp). *G3 GenesGenomesGenetics*, *8*, 3347–3355. https://doi.org/10.1534/G3.118.200521

Huynh, B., Ehlers, J.D., Huang, B.E., Muñoz-Amatriaín, M., Lonardi, S., Santos, J.R.P., Ndeve, A., Batieno, B.J., Boukar, O., Cisse, N., Drabo, I., Fatokun, C., Kusi, F., Agyare, R.Y., Guo, Y., Herniter, I., Lo, S., Wanamaker, S.I., Xu, S., Close, T.J., & Roberts, P.A. (2018)(a). A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *The Plant Journal*, *93*, 1129–1142. https://doi.org/10.1111/tpj.13827

Huynh, B.-L., Close, T.J., Roberts, P.A., Hu, Z., Wanamaker, S., Lucas, M.R., Chiulele, R., Cissé, N., David, A., Hearne, S., Fatokun, C., Diop, N.N., & Ehlers, J.D. (2013). Gene pools and the genetic architecture of domesticated cowpea. *The Plant Genome*, *6*, 1–8. https://doi.org/10.3835/plantgenome2013.03.0005

Huynh, B.-L., Ehlers, J.D., Huang, B.E., Muñoz-Amatriaín, M., Lonardi, S., Santos, J.R.P., Ndeve, A., Batieno, B.J., Boukar, O., Cisse, N., Drabo, I., Fatokun, C., Kusi, F., Agyare, R.Y., Guo, Y.-N., Herniter, I., Lo, S., Wanamaker, S.I., Xu, S., Close, T.J., & Roberts, P.A. (2018)(b). A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *The Plant Journal*, *93*, 1129–1142. https://doi.org/10.1111/TPJ.13827

Intertek-Agritech. (2016). Agri-Services CGIAR HTPG PROJECT Sampling instructions for SNP verification and routine SNP analysis.

Jaccoud, D., Peng, K., Feinstein, D., & Kilian, A. (2001). Diversity Arrays: a solid-state technology for sequence information independent genotyping. *Nucleic Acids Research*, *29*, 25

Jakobsson, M., & Rosenberg, N.A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*. https://doi.org/10.1093/bioinformatics/btm233

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*. https://doi.org/10.1186/1471-2156-11-94

Josse, J., & Husson, F. (2016). missMDA: A package for handling missing values in multivariate data analysis. *Journal of Statistical Software*, *70*.

https://doi.org/10.18637/jss.v070.i01

Kassambara, A., & Mundt, F. (2020). Factoextra: Extract and visualize the results of multivariate data analyses [R Package Version 1.0.7.]

Ketema, S., Tesfaye, B., Keneni, G., Fenta, B.A., Assefa, E., Greliche, N., Machuka, E., & Yao, N. (2020). DArTSeq SNP-based markers revealed high genetic diversity and structured population in Ethiopian cowpea [*Vigna unguiculata* (L.) Walp] germplasms. *PLOS ONE*, *15*, e0239122. https://doi.org/10.1371/JOURNAL.PONE.0239122

Kilian, B., Özkan, H., Walther, A., Kohl, J., Dagan, T., Salamini, F., & Martin, W. (2007). Molecular diversity at 18 Loci in 321 wild and 92 domesticate lines Reveal no reduction of nucleotide diversity during *Triticum monococcum* (Einkorn) domestication: Implications for the origin of agriculture. *Molecular Biology and Evolution*, *24*, 2657–2668. https://doi.org/10.1093/MOLBEV/MSM192

Lischer, H.E.L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*. https://doi.org/10.1093/bioinformatics/btr642

Liu, A., & Burke, J.M. (2006). Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics*, *173*, 321–330. https://doi.org/10.1534/GENETICS.105.051110

Lo, S., Muñoz-Amatriaín, M., Boukar, O., Herniter, I., Cisse, N., Guo, Y.N., Roberts, P.A., Xu, S., Fatokun, C., & Close, T.J. (2018). Identification of QTL controlling domestication-related traits in cowpea (*Vigna unguiculata* L. Walp). *Scientific Reports 2018 8:1*, *8*, 1–9. https://doi.org/10.1038/s41598-018-24349-4

Lo, S., Muñoz-Amatriaín, M., Hokin, S.A., Cisse, N., Roberts, P.A., Farmer, A.D., Xu, S., & Close, T.J. (2019). A genome-wide association and meta-analysis reveal regions associated with seed size in cowpea [*Vigna unguiculata* (L.) Walp]. *Theoretical and Applied Genetics*, *132*, 3079–3087. https://doi.org/10.1007/s00122-019-03407-z

Lonardi, S., Muñoz-Amatriaín, M., Liang, Q., Shu, S., Wanamaker, S.I., Lo, S., Tanskanen, J., Schulman, A.H., Zhu, T., Luo, M.C., Alhakami, H., Ounit, R., Hasan, A.M., Verdier, J., Roberts, P.A., Santos, J.R.P., Ndeve, A., Doležel, J., Vrána, J., Hokin, S.A., Farmer, A.D., Cannon, S.B., & Close, T.J. (2019). The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant Journal*, *98*, 767–782. https://doi.org/10.1111/tpj.14349

Lucas, M.R., Diop, N.-N., Wanamaker, S., Ehlers, J.D., Roberts, P.A., & Close, T.J. (2011). Cowpea–soybean synteny clarified through an improved genetic Map. *The Plant Genome*, *4*, 218–225. https://doi.org/10.3835/PLANTGENOME2011.06.0019

Mafakheri, K., Bihamta, M.R., & Abbasi, A.R. (2017). Assessment of genetic diversity in cowpea (*Vigna unguiculata* L.) germplasm using morphological and molecular characterisation. *http://www.editorialmanager.com/cogentagri*, *3*, 1327092. https://doi.org/10.1080/23311932.2017.1327092

Meirmans, P.G., & Hedrick, P.W. (2011). Assessing population structure: FST and related measures. *Molecular Ecology Resources*, *11*. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Muchero, W., Diop, N.N., Bhat, P.R., Fenton, R.D., Wanamaker, S., Pottorff, M., Hearne, S., Cisse, N., Fatokun, C., Ehlers, J.D., Roberts, P.A., & Close, T.J. (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. https://doi.org/10.1073/pnas.0905886106

Muchero, W., Roberts, P.A., Diop, N.N., Drabo, I., Cisse, N., Close, T.J., Muranaka, S., Boukar, O., & Ehlers, J.D. (2013). Genetic architecture of delayed senescence, biomass, and grain yield under drought stress in cowpea. *PLOS ONE*, *8*, e70041. https://doi.org/10.1371/JOURNAL.PONE.0070041

Muñoz-Amatriaín, M., Lo, S., Herniter, I.A., Boukar, O., Fatokun, C., Carvalho, M., Castro, I., Guo, Y.N., Huynh, B.L., Roberts, P.A., Carnide, V., & Close, T.J. (2021). The UCR Minicore: a resource for cowpea research and breeding. *Legume Science*, *3*, e95. https://doi.org/10.1002/LEG3.95

Muñoz-Amatriaín, M., Mirebrahim, H., Xu, P., Wanamaker, S.I., Luo, M.C., Alhakami, H., Alpert, M., Atokple, I., Batieno, B.J., Boukar, O., Bozdag, S., Cisse, N., Drabo, I., Ehlers, J.D., Farmer, A., Fatokun, C., Gu, Y.Q., Guo, Y.N., Huynh, B.L., Jackson, S.A., Kusi, F., Lawley, C.T., Lucas, M.R., Ma, Y., Timko, M.P., Wu, J., You, F., Barkley, N.A., Roberts, P.A., Lonardi, S., & Close, T.J. (2017). Genome resources for climate-resilient cowpea, an essential crop for food security. *Plant Journal*, *89*, 1042–1054. https://doi.org/10.1111/tpj.13404

Nei, M., & Li, W.H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*, 5269. https://doi.org/10.1073/PNAS.76.10.5269

Nei, M., & Saitou, N. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, *4*, 406–25

Nkhoma, N., Shimelis, H., Laing, M.D., Shayanowako, A., & Mathew, I. (2020)(a). Assessing the genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] germplasm collections using phenotypic traits and SNP markers. *BMC Genetics 2020 21:1*, *21*, 1–16. https://doi.org/10.1186/S12863-020-00914-7

Nkhoma, N., Shimelis, H., Laing, M.D., Shayanowako, A., & Mathew, I. (2020)(b). Assessing the genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] germplasm collections using phenotypic traits and SNP markers. *BMC Genetics*, *21*, 1–16. https://doi.org/10.1186/S12863-020-00914-7/TABLES/9

Nkomo, G. v., Sedibe, M.M., & Mofokeng, M.A. (2021). Production constraints and improvement strategies of cowpea (*Vigna unguiculata* L. Walp.)

genotypes for drought tolerance. *International Journal of Agronomy*, *2021*. https://doi.org/10.1155/2021/5536417

Ongom, P.O., & Ejeta, G. (2018). Mating design and genetic structure of a multi-parent advanced generation intercross (MAGIC) population of sorghum (*Sorghum bicolor* (L.) moench). *G3: Genes, Genomes, Genetics*, *8*, 331–341. https://doi.org/10.1534/g3.117.300248

Ongom, P.O., Fatokun, C., Togola, A., Salvo, S., Oyebode, O.G., Ahmad, M.S., Jockson, I.D., Bala, G., & Boukar, O. (2021). Molecular fingerprinting and hybridity authentication in cowpea using single nucleotide polymorphism based kompetitive allele-specific PCR assay. *Frontiers in Plant Science*, *12*, 2205. https://doi.org/10.3389/FPLS.2021.734117/BIBTEX

Pasquet, R.S. (2000). Allozyme diversity of cultivated cowpea *Vigna unguiculata* (L.) Walp.. *Theoretical and Applied Genetics 2000 101:1*, *101*, 211–219. https://doi.org/10.1007/S001220051471

Pasquet, R.S., Feleke, Y., & Gepts, P. (2021). Cowpea [*Vigna unguiculata* (L.) Walp.] maternal lineages, chloroplast captures, and wild cowpea evolution. *Genetic Resources and Crop Evolution*. https://doi.org/10.1007/s10722-021-01155-y

Paudel, D., Dareus, R., Rosenwald, J., Muñoz-Amatriaín, M., & Rios, E.F. (2021). Genome-wide association study reveals candidate genes for flowering time in cowpea (*Vigna unguiculata* [L.] Walp.). *Frontiers in Genetics*, *12*. https://doi.org/10.3389/FGENE.2021.667038/FULL

Pottorff, M.O., Li, G., Ehlers, J.D., Close, T.J., & Roberts, P.A. (2014). Genetic mapping, synteny, and physical location of two loci for *Fusarium oxysporum* f. sp. tracheiphilum race 4 resistance in cowpea [*Vigna unguiculata* (L.) Walp]. *Molecular breeding: new strategies in plant improvement*, *33*, 779–791. https://doi.org/10.1007/S11032-013-9991-0

Pritchard, J.K., Stephens, M., & Donnelly, P. (2000)(a). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Pritchard, J.K., Stephens, M., Rosenberg, N.A., & Donnelly, P. (2000)(b). Association mapping in structured populations. *American Journal of Human Genetics*, *67*, 170. https://doi.org/10.1086/302959

Pruett, C.L., Winker, K., Pruett, C.L., & Winker, K. (2008). The effects of sample size on population genetic diversity estimates in song sparrows (*Melospiza melodia)*. *J. Avian Biol*, *39*, 252–256. https://doi.org/10.1111/j.2008.0908-8857.04094.x

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rosenberg, N.A. (2004). DISTRUCT: A program for the graphical display of

population structure. *Molecular Ecology Notes*, *4*. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Sarr, A., Bodian, A., Gbedevi, K.M., Ndir, K.N., Ajewole, O.O., Gueye, B., Foncéka, D., Diop, E.A.M.C., Diop, B.M., Cissé, N., & Diouf, D. (2021). Genetic diversity and population structure analyses of wild relatives and cultivated cowpea (*Vigna unguiculata* (L.) Walp.) from Senegal using simple sequence repeat markers. *Plant Molecular Biology Reporter*, *39*, 112–124. https://doi.org/10.1007/S11105-020-01232-Z/FIGURES/7

Sethuraman, A. (2018). Estimating Genetic Relatedness in Admixed Populations. *G3: GenesGenomesGenetics*, *8*, 3203. https://doi.org/10.1534/G3.118.200485

Singh, B.B. (2014). *Cowpea: The Food Legume of the 21st Century. Cowpea: The Food Legume of the 21st Century* Crop Science Society of America, Madison, WI, USA.

Sodedji, F.A.K., Agbahoungba, S., Agoyi, E.E., Kafoutchoni, M.K., Choi, J., Nguetta, S.P.A., Assogbadjo, A.E., & Kim, H.Y. (2021). Diversity, population structure, and linkage disequilibrium among cowpea accessions. *The Plant Genome*, *14*, e20113. https://doi.org/10.1002/TPG2.20113

Sodedji, F.A.K., Agbahoungba, S., Nguetta, S.-P.A., Agoyi, E.E., Ayenan, M.A.T., Sossou, S.H., Mamadou, C., Assogbadjo, A.E., & Kone, D. (2019). Resistance to legume pod borer (Maruca vitrata Fabricius) in cowpea: genetic advances, challenges, and future prospects. *https://doi.org/10.1080/15427528.2019.1680471*, *34*, 238–267. https://doi.org/10.1080/15427528.2019.1680471

Stadlmeier, M., Hartl, L., & Mohler, V. (2018). Usefulness of a multiparent advanced generation intercross population with a greatly reduced mating design for genetic studies in winter wheat. *Frontiers in Plant Science*, *871*, 1825. https://doi.org/10.3389/FPLS.2018.01825/BIBTEX

Subramanian, S. (2016). The effects of sample size on population genomic analyses – implications for the tests of neutrality. *BMC Genomics*, *17*. https://doi.org/10.1186/S12864-016-2441-8

Ba, F.S., Pasquet, R.S., & Gepts, P. (2004). Genetic diversity in cowpea [*Vigna unguiculata* (L.) Walp.] as revealed by RAPD markers. Genetic Resources and Crop Evolution, 51, 539–550. https://doi.org/10.1023/B:GRES.0000024158.83190.4e

Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*, *123*, 585. https://doi.org/10.1093/genetics/123.3.585

Uemoto, Y., Sasaki, S., Kojima, T., Sugimoto, Y., & Watanabe, T. (2015). Impact of QTL minor allele frequency on genomic evaluation using real genotype data and simulated phenotypes in Japanese Black cattle. *BMC Genetics*, *16*. https://doi.org/10.1186/S12863-015-0287-8

Wagner, A.P., Creel, S., & Kalinowski, S.T. (2006). Estimating relatedness and relationships using microsatellite loci with null alleles. *Heredity*, *97*.

https://doi.org/10.1038/sj.hdy.6800865

Wamalwa, E.N., 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*. https://doi.org/10.1155/2016/8956412

Wang, J. (2012). On the measurements of genetic differentiation among populations. *Genetics Research Cambridge*, *94*, 275–289. https://doi.org/10.1017/S0016672312000481

Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, *16*, 97. https://doi.org/10.1093/GENETICS/16.2.97

Xiong, H., Qin, J., Shi, A., Mou, B., Wu, D., Sun, J., Shu, X., Wang, Z., Lu, W., Ma, J., Weng, Y., & Yang, W. (2018). Genetic differentiation and diversity upon genotype and phenotype in cowpea (*Vigna unguiculata* L. Walp.). *Euphytica*, *214*, 1–15. https://doi.org/10.1007/S10681-017-2088-9/FIGURES/4

Xu, P., Wu, X., Wang, B., Luo, J., Liu, Y., Ehlers, J.D., Close, T.J., Roberts, P.A., Lu, Z., Wang, S., & Li, G. (2012). Genome wide linkage disequilibrium in Chinese asparagus bean (*Vigna. unguiculata* ssp. *sesquipedialis*) germplasm: implications for domestication history and genome wide association studies. *Heredity 2012 109:1*, *109*, 34–40. https://doi.org/10.1038/hdy.2012.8

Yan, J., Shah, T., Warburton, M.L., Buckler, E.S., McMullen, M.D., & Crouch, J. (2009). genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLOS ONE*, *4*, e8451. https://doi.org/10.1371/JOURNAL.PONE.0008451

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., & Weir, B.S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*, 3326. https://doi.org/10.1093/BIOINFORMATICS/BTS606