

Diversity analysis and genome-wide association studies of seed weight trait in Bambara groundnut (*Vigna subterranea* (L.) Verdc.) using DArTseq-derived SNP markers

Phyllis Muturi

`muturi.phyllis@embuni.ac.ke`

University of Embu

Martina Kyallo

Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI) Hub

Edmore Gasura

University of Zimbabwe

Nasser Yao

Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI) Hub

Research Article

Keywords: DArT markers, genetic diversity, population structure, genome wide association, *Vigna subterranea*, SNP markers

Posted Date: August 28th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4829431/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Euphytica on March 10th, 2025. See the published version at <https://doi.org/10.1007/s10681-025-03478-8>.

Abstract

Bambara groundnut is an indigenous legume cultivated in Africa and referred to as 'poor man's crop'. Seed weight is an important trait that end-users consider in Bambara groundnut because it is associated with better yield. The crop is drought tolerant compared to other legumes. In Kenya, the crop is grown by women and the diversity of the accessions grown is unknown. Lack of information on the crop's diversity and population structure make selection and genetic improvement of the crop difficult. The objective of this study was to: i. determine the genetic diversity and population structure of 86 Bambara groundnut accessions from western and coastal Kenya using 4,399 SNP markers and ii) identify SNPs associated with the seed weight of the Bambara groundnut collection. The DArT complexity reduction approach in combination with Illumina short-read sequencing (HiSeq 2000) was applied. Population structure analysis revealed three distinct genetic clusters implying that the accessions were from three major gene pools. Analysis of molecular variance showed a low level of differentiation among the populations implying the need to broaden the genetic base of the collection tested in this study. The study revealed 8 significant DArTseq derived SNP markers ($P < 0.05$) distributed across Bambara groundnut contigs 1, 2, 3, and 5 associated with seed weight. The study demonstrated low genetic variation among the populations but a high genetic variability within population. There will be a need to validate the generated markers and conduct whole genome association studies with other economically important traits in Bambara groundnut.

Introduction

Bambara groundnut (*Vigna subterranea* (L) Verdc.) is an indigenous legume well adapted in semi-arid lands of sub-Saharan Africa (SSA). The crop is mainly grown by smallholder resource-constrained farmers, mainly women (Massawe et al., 2016). Climate change and global warming are continuous threats to food and nutritional security, especially in developing countries. Bambara groundnut landraces have been found to have higher stability and adaptation to climate variability in low-input agriculture thus cultivation of the crop may contribute to household resilience in periods of food production deficits (Govindaraj et al., 2013). The crop is diploid, with 11 pairs of chromosomes ($2n = 2x = 22$) and is highly inbred as reported by Massawe et al., (2016). The origin of the crop is believed to be in West Africa, with Burkina Faso being the highest producer, followed by Cameroon, Democratic Republic of Congo, Mali, Niger, and Togo (FAOSTAT, 2022). High seed weight is an important trait in Bambara groundnut that is considered by both consumers and farmers thus a major target trait in Bambara groundnut breeding. Seed weight positively correlates with high grain yield, improved seed germination, and enhanced seedling vigour (Somta et al., 2015). Seed weight is a quantitative trait with average heritability of over 90% in *Vigna* spp. (Thouseem et al., 2018).

Estimates of food insecurity and malnutrition indicate that SSA has the highest proportion of malnourished people in the world (Xie et., 2021). There has been a growing awareness in recent years on the potential of Bambara groundnuts to contribute to increased food, nutrition, and income securities in SSA. Hot and dry climate, erratic rainfall, and nutrient-deficient soils have been identified as the major

challenges to improved food production in this region (Govindaraj et al., 2013). About 4.3 million (30%) people in Kenya live in arid and semi-arid lands (ASALs) which occupy 80% of the landmass (World Food Programme, 2012). Climate change variability and overdependence on a few exotic staples such as common beans that often fail due to inadequate rainfall may have contributed to food and nutritional insecurity in these areas. Climate change brings in new challenges to ASALs as it further deteriorates the prevailing dry and hot conditions translating to food and nutritional insecurities in these areas (Massawe et al. 2016). The types of crops cultivated in ASALs have drastically reduced and few drought-tolerant indigenous crops such as Bambara groundnuts are in the hands of farmers. Bambara groundnut is an indigenous, nutritious climate-smart legume that is well adapted to arid and semi-arid lands (ASALs) (Bamshaiye et al., 2011). The crop is rich in lysine which is deficient in many kinds of cereal such as sorghum which are grown in ASALs (NRC, 2006). Bambara groundnut has, therefore, enormous potential in reducing chronic malnutrition, starvation, and hunger which are commonly experienced in low agricultural potential areas. Severe and persistent droughts associated with climate change in Kenya require climate-smart and nutritious crops such as Bambara groundnut.

Bambara groundnut is mainly grown in western Kenya in Bungoma, Kakamega, Busia, Vihiga counties and at the Coast (Ndiang et al., 2014). Bambara groundnut seeds contain approximately 63% carbohydrate, 19% protein, and 6.5% oil thus referred to as complete food (NRC, 2006). The crop holds considerable commercial potential since the crop fetches a premium in the market compared to other legumes such as common beans and other types of beans. Bambara groundnut has high levels of methionine than most legumes thus can be a candidate to manage malnutrition (Bamshaiye et al., 2011) and serves as an important source of protein among poor people who cannot afford animal protein (Ndiang et al., 2014). Bambara groundnut is a sustainable crop because; it requires no spraying to control pests and diseases, boosts soil fertility, conditions impoverished soils, tolerates acidic soils that are highly toxic for other types of legumes, and has antagonistic effects on *Striga* weed (Bamshaiye et al., 2011; Massawe et al. 2016). The crop is considered as one of the orphaned and underutilized species in Kenya (Njoroge et al., 2010). Bambara groundnut has not received attention in research compared to other legumes such as common beans thus the reason why farmers grow unimproved genotypes and most communities do not know that this superior bean exists (NRC, 2006). The crop is a dual purpose whose grains are consumed either as roasted or boiled and the haulms are used as animal fodder. The crop is important for the small-scale farmers in rural households due to its considerable commercial potential and can raise communities from poverty, food, and nutritional insecurities (Ndiang et al., 2014).

Knowledge of the degree of genetic relationships between Bambara groundnut is important for crop improvement and may help to establish a core collection as part of the germplasm collection management (Bamshaiye et al., 2011). Moreover, the utilization of landraces is valued as they can contain favourable alleles for many agronomic traits. A core collection has the maximum genetic variation in a minimum number of accessions thus economical in conservation. Diversity in plant genetic resources provides breeders with opportunities to develop new cultivars with end-user desirable characteristics (Govindaraj et al, 2015). These traits include farmer-preferred traits, breeders preferred traits, and market preferred traits. Scanty information on genetic diversity of Bambara groundnut has

made the selection, genetic improvement, and conservation of the crop difficult for breeders. The development of improved varieties with desirable traits in crops depends on the available germplasm (Bernardo, 2008). Such collections are sources of novel genes desirable for enhancing productivity and resilience of newly developed varieties (Govindaraj et al., 2013). Generally, there are often large numbers of uncharacterized accessions in gene banks in Sub-Saharan Africa. To overcome this bottleneck, it is advisable to work with core collections since they represent the crop's genetic diversity and is more cost-effective to work with (Upadhyaya et al., 2008). Traditional breeding of Bambara groundnut has been difficult since the crop is an extreme inbreeder and flowers are formed at the soil ground level. The use of molecular markers in recent years has revolutionized the precision of plant genetic analysis and molecular breeding of crops (Bernardo, 2008; Mayes et al. 2011). Molecular markers such as DArT based markers have been used to reveal the genetic diversity in different populations such as sorghum. Molecular markers such as single nucleotide polymorphisms (SNPs) are at the epicentre of molecular genetics. SNPs are the best suited for diversity studies due to; their abundance in the genomes, amenability to high-throughput detection platforms, their stability, efficiency, cost-effectiveness, and are very near to or even within the gene of interest (Rafalski, 2002; Morin et al., 2004; Syvanen, 2005; Bernardo, 2008; Govindaraj et al., 2013).

Genome wide association study (GWAS) is a powerful tool for the identification of candidate regions associated with quantitative traits (Kamfwa et al., 2015). The approach searches the whole genome for causal genetic variation and does not require previous information on candidate genes (Kamfwa et al., 2015). GWAS is one of several approaches developed to identify marker-trait associations taking advantage of linkage disequilibrium (Bernardo, 2008). The approach was initially used in human and animal research and has been adapted in plant research (Morin et al., 2004). GWAS can reveal associations between genomic loci and advantageous traits such as seed weight (Kilian and Graner, 2012). These loci are key targets for crop improvement by the breeder. The chief benefit of GWAS is the potential to capture greater diversity by including a large number of unrelated individuals with distinct genetic backgrounds (Kamfwa et al., 2015). GWAS usually narrows down the candidate regions using natural populations (Mayes et al. 2011). Genotyping by sequencing (GBS) is one of the genotyping methods used for GWAS, and GBS-GWAS approaches have been successfully applied to the identification of candidate genes controlling quantitative traits in legumes such as soybean (*Glycine max*) (Sonah et al., 2015).

Recently, there has been effort to study Bambara groundnut diversity using SNPs markers and the study concluded that DArT SNP marker can be widely used for molecular analysis of Bambara groundnut (Uba et al., 2021). The fast advancement of next-generation sequencing technology has assisted the large-scale discovery of SNPs in various plant species (Valdisser et al., 2017). Genotyping-by-sequencing (GBS) is a next-generation sequencing (NGS) based method that employs reduced representation to enable high-throughput genotyping of a large number of SNP markers (Elshire et al., 2011; Poland and Rife, 2012). Microarrays have facilitated the development of SNP platforms enabling genotyping of thousands of markers in one experiment (Gupta et al., 2008). Array-based assays such as Diversity Arrays Technology (DArT) have been utilized to provide the desired ultra-high throughput and cost-

effectiveness (Valdisser et al., 2017). These microarray-based markers have been the markers of choice and have been used for the construction of high-density maps, quantitative trait loci (QTL) mapping, and genetic diversity analysis with limited time and funds (Gupta et al., 2008). DArTseq technology offers a gain in the reduction of genome complexity through intelligent selection of genome fraction corresponding predominantly to active genes in genome characterization (Valdisser et al., 2017). The technology is among the most appropriate system to discover hundreds of polymorphic genomic loci, scoring thousands of unique genomic-wide DNA fragments in a single experiment, without requiring existing DNA sequence information (Valdisser et al., 2017). The DArT complexity reduction approach in combination with Illumina short-read sequencing using HiSeq2000 was applied in this study. The objective of this study was to i) determine the genetic diversity and population structure of 92 Bambara groundnut accessions collected from western and Coastal Kenya using 4,399 SNP markers and ii) identify SNPs associated with seed weight of the Bambara groundnut collection.

Materials and methods

Bambara groundnut germplasm

Eighty-six Bambara groundnut accessions were collected in March 2018 from farmers from a total of five counties i.e., four counties in western highlands of Kenya namely Busia, Kakamega, Bungoma, Vihiga, and from Kilifi in the coastal region where the crop is grown by smallholder farmers (Table 1). The geographical variation of the 5 counties is shown in Table 2. The seeds were sorted according to the seed coat colour/patterns and the place of collection of the germplasm. The health of the seeds was determined visually, and healthy seeds were placed in separate dispensing envelopes. The Bambara groundnut accessions from each county were assigned accession codes for identification purposes.

Table 1
List of Bambara groundnut accessions used in this study

#	Accession code	Seed coat colour /pattern	Seed source	#	Accession code	Seed coat colour /pattern	Seed source
1	KAK101	Brick red with black specks	Kakamega	44	VHG303	Cream	Vihiga
2	KAK102	Cream with black specks	Kakamega	45	VHG304	Dark reddish	Vihiga
3	KAK103	Brown	Kakamega	46	VHG305	Blackish red	Vihiga
4	KAK104	Light brown	Kakamega	47	VHG306	Blackish red	Vihiga
5	KAK105	Light brown	Kakamega	48	VHG307	Creamish	Vihiga
6	KAK107	Dark brown	Kakamega	49	VHG308	Light brown	Vihiga
7	KAK108	Brown with black specks	Kakamega	50	VHG309	Kakhish	Vihiga
8	KAK109	Cream with brown specks	Kakamega	51	VHG310	Brick red with black specks	Vihiga
9	KAK110	Brick red	Kakamega	52	VHG311	Brown	Vihiga
10	KAK111	Light brown	Kakamega	53	VHG312	Reddish	Vihiga
11	KAK112	Dull red	Kakamega	54	VHG313	Brown	Vihiga
12	KAK113	Dark brown	Kakamega	55	VHG314	Light red	Vihiga
13	KAK114	Dark brown	Kakamega	56	KLF401	Reddish brown	Kilifi
14	KAK115	Cream	Kakamega	57	KLF402	Brown	Kilifi
15	KAK116	Light brown	Kakamega	58	KLF403	Light brown	Kilifi
16	KAK117	Purplish	Kakamega	59	KLF404	Dark brown	Kilifi
17	KAK118	Black	Kakamega	60	KLF405	Greyish red with black specks	Kilifi
18	KAK119	Cream	Kakamega	61	KLF406	Cream	Kilifi
19	KAK120	Black	Kakamega	62	KLF407	Cream	Kilifi
20	KAK121	Cream	Kakamega	63	KLF408	Cream with black specks	Kilifi
21	KAK122	Dark red	Kakamega	64	KLF409	Black	Kilifi

#	Accession code	Seed coat colour /pattern	Seed source	#	Accession code	Seed coat colour /pattern	Seed source
22	KAK124	Cream	Kakamega	65	KLF411	Cream	Kilifi
23	KAK125	Black	Kakamega	66	KLF412	Cream	Kilifi
24	KAK126	Dark brown	Kakamega	67	KLF413	Red	Kilifi
25	KAK127	Light brown	Kakamega	68	KLF414	Light brown	Kilifi
26	BGM201	Beige	Bungoma	69	KLF415	Cream	Kilifi
27	BGM202	Dark beige	Bungoma	70	BSA501	Dark purplish	Busia
28	BGM203	Whitish	Bungoma	71	BSA502	Black	Busia
29	BGM204	White with brown specks	Bungoma	72	BSA504	Dark cream	Busia
30	BGM205	Beige	Bungoma	73	BSA505	Light brown	Busia
31	BGM206	Greyish with black specks	Bungoma	74	BSA506	Light brown with black specks	Busia
32	BGM207	Dark beige	Bungoma	75	BSA508	Light brown	Busia
33	BGM208	Dark brown	Bungoma	76	BSA509	Light brown	Busia
34	BGM209	Black	Bungoma	77	BSA510	brown	Busia
35	BGM210	White with brown specks	Bungoma	78	BSA511	Reddish	Busia
36	BGM211	Whitish with black specks	Bungoma	79	BSA513	Reddish	Busia
37	BGM212	Dark brown	Bungoma	80	BSA514	Black	Busia
38	BGM213	Cream	Bungoma	81	BSA515	Light brown	Busia
39	BGM214	Black	Bungoma	82	BSA517	Light brown	Busia
40	BGM215	Reddish	Bungoma	83	BSA518	Dark brown	Busia
41	BGM216	Cream with black specks	Bungoma	84	BSA519	Light brown	Busia
42	VHG301	Cream	Vihiga	85	BSA521	Light brown	Busia
43	VHG302	Brick red with black specks	Vihiga	86	BSA522	Black	Busia

Table 2
Geographical information of counties sampled in this study

County	Elevation (Meters above sea level (MASL))	Ecology	Latitude	Longitude	No. samples collected
Kilifi	30	Coastal lowland	3.5107°S	39.9093°E	14
Vihiga	1,200	Western highland	0.0816°N	34.7229°E	14
Busia	1,227	Western highland	0.4608°N	34.1115°E	17
Bungoma	1,385	Western highland	0.5695°N	34.5584°E	16
Kakamega	1,535	Western highland	0.2842°N	34.7523°E	25

Source: Jaetzold and Schmidt, 1982.

DNA extraction

Three seed samples of Bambara groundnut from each of the 86 accessions were grown in pots in the screen house at the BecA-ILRI Hub in Nairobi, Kenya for DNA extraction. Total DNA was extracted from approximately 30mg of leaves from three weeks old Bambara groundnut seedlings using the Quick-DNA Plant/Seed kit (Zymo Research, USA) following the manufacturer's protocol. The integrity of DNA was checked using agarose gel electrophoresis while the concentration of DNA was estimated using the NanoDrop 1000 spectrophotometer (NanoDrop Technologies, Thermo Scientific, Wilmington, DE, USA). The extracted DNA was normalized to 50 ng/μl with triple distilled water and stored at -20°C before use. DArTseq was done by the Integrated Genotyping Support and Service (IGSS) unit hosted at BecA-ILRI Hub.

DArTseq SNP data scoring and filtering

Since Bambara groundnut and cowpea (*Vigna unguiculata*) are close relatives, the genome of the latter was used as the reference genome for SNP mining. A total of 8,798 SNPs was generated before filtering. DArTseq detected silico DArT markers are dominant and were scored in a binary fashion where 1 represented presence and 0 represented absence of a restriction fragment in the genomic representation of each Bambara groundnut sample (Kilian et al., 2012). The markers were screened according to call rate, polymorphism information content (PIC), marker reducibility, missing SNP data, and minor allele frequency according to Valdisser et al., (2017). A total of 4,399 DArTseq informative SNPs were considered for the current study after quality control.

Genetic diversity and population structure analysis

Bayesian population structure analysis was done using Structure software version 2.3.3 (Falush et al., 2003). A Bayesian clustering approach was used to estimate the number of subpopulations (K) using 100,000 Markov Chain Monte Carlo (MCMC) simulations with ten replicates for each K-value from 1 to 10 with a burn-in length of 10,000. The admixture model and correlated allele frequencies were applied (Pritchard et al., 2003). The online STRUCTURE HARVESTER application (<http://taylor0.biology.ucla.edu/structureHarvester/>) was used to analyse the structure results. The two model choice criteria used to detect the most probable value of K were the $\text{LnP}(D)$ value for each given K and ΔK , an *ad hoc* quantity related to the second-order change of the log probability of data with respect to the number of clusters inferred by Structure (Evanno et al., 2005). Population differentiation due to genetic structure was assessed using a Neighbour Joining (NJ) tree method and Principal Component Analysis (PCA) generated by R statistical software version 3.5.1. (R Core Team, 2013).

Genetic diversity and relationships analysis

Genetic distances among the evaluated Bambara groundnut genotypes were calculated from the proportion of shared alleles obtained from DArTseq SNPs by using Euclidean genetic distance coefficients according to Nadeem et al. (2018). The genotypic data was subjected to various within and among groups genetic diversity measures using GenAlEx version 6.5 software (Peakall and Smouse, 2012). Genetic diversity parameters namely, number of different alleles (N_a), number of effective alleles (N_e), Shannon's information index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), and unbiased expected heterozygosity (uH_e) were calculated using the same software (Fatokun et al., 2018). Total genetic variation was partitioned into within and among populations through molecular analysis of variance (AMOVA) using GenAlEx genetic software (Fatokun et al., 2018). Genetic distances between each pair of accessions and between pre-grouped populations were measured based on Nei's genetic distance using the same software (Nei et al., 1983). Genetic distance matrices for each locus were summed across loci assuming statistical independence (Nei et al., 1983). Pair-wise genetic frequency-based dissimilarity or distance matrix between individuals was calculated according to Nei et al., 1983. The resulting dissimilarity matrix was subjected to tree construction using the unweighted pair group method analysis (UPGMA) using R software (R Core Team, 2013).

Genome-wide association for Bambara groundnut seed weight

Seed weight was considered for genome-wide association studies (GWAS) using 4,399 DArTseq SNP-derived markers. Association analysis was performed with R software using Mixed Linear Model (MLM) method (R Core Team, 2013). Manhattan plots for seed weight were generated in GWAS and showed the most significant associations ($-\log(p\text{-value}) > 2$). Marker alleles with P-values ≤ 0.001 in MLM model were declared significantly associated with seed weight in Bambara groundnut.

Results

Population structure

Results of population structure analysis of 86 Bambara groundnut genotypes using STRUCTURE, for K ranging from 1 to 10, and by inferring on Delta K of Evanno et al., (2005) identified K = 3 as the most suitable K value for determining the genetic cluster as shown in Fig. 1.

A dendrogram was constructed from dissimilarity using 4,399 DArTseq markers and results are shown in Fig. 2.

Principal component analysis (PCoA) was done based on 4,399 DArTseq SNPs, using GenAlEx analysis (Fig. 4). PCA 1 accounted for 25.27% of the variation observed while PCA2 accounted for 16.41% and the accessions clustered in three groups. Interestingly, cluster 1 had accessions from all the counties except Kilifi, cluster 2 had accessions from all the five regions sampled while cluster 3 had accessions from Bungoma and Kakamega.

Gene diversity, polymorphism information content (PIC), and heterozygosity were determined and shown in Table 2. The mean PIC values for each SNP locus in Bambara groundnut accessions from Bungoma, Vihiga, Kilifi, Busia, and Kakamega were 77.13%, 57.58%, 59.90%, 69.99%, and 38.92%, respectively. The mean number of different alleles for each population was 1.77, 1.57, 1.60, 1.70, and 1.38, respectively. Bambara groundnut population from Busia had the highest number of effective alleles (1.33), heterozygosity (0.21), and Shannon Index (0.32). Nevertheless, populations from Bungoma showed the highest level of different alleles (1.77) and the highest percentage of polymorphic loci (77.13%). Populations from Kakamega showed the lowest level of all the measures of genetic diversity measured in this study. The genetic diversity measure of populations from Bungoma and Vihiga were similar in this study (0.28).

Table 2
Estimation of population diversity indices in the 86 Bambara groundnut accessions studied

Population	(N)	(Na)	(Ne)	(P %)	(Ho)	(He)	(uHe)	(I)
Bungoma	16	1.77	1.27	77.13%	0.19	0.17	0.17	0.28
Vihiga	14	1.57	1.3	57.58%	0.21	0.18	0.19	0.28
Kilifi	14	1.6	1.2	59.90%	0.15	0.14	0.14	0.23
Busia	17	1.7	1.33	69.99%	0.24	0.2	0.21	0.32
Kakamega	25	1.38	1.18	38.92%	0.15	0.12	0.13	0.18

(N); No. of samples analysed, (Na); Number of different alleles, (Ne); Number of effective alleles, (P %); Percentage of polymorphic loci, (Ho); Observed Heterozygosity, (He); Expected heterozygosity, (uHe); Unbiased Expected Heterozygosity, (I); Shannon's Information index

The genetic distance among the different Bambara groundnut populations was estimated with 4,399 DArTseq derived SNP markers and the results are shown in Table 3. The greatest genetic distance was observed between populations from Vihiga and Kilifi (0.045), Vihiga and Kakamega (0.045), and Vihiga and Busia (0.042) in that order. The least genetic distance was observed between populations from Bungoma and Vihiga (0.021), Kilifi and Busia (0.021), Kilifi and Kakamega (0.021 and Bungoma and Kakamega (0.023) in that order.

Table 3
Nei genetic distance between populations

Populations	Bungoma	Vihiga	Kilifi	Busia	Kakamega
Bungoma	0				
Vihiga	0.021	0			
Kilifi	0.027	0.045	0		
Busia	0.035	0.042	0.021	0	
Kakamega	0.023	0.045	0.021	0.036	0

The mean fixation index (Fst) was 0.063.

Table 4 shows the results of the analysis of molecular variance (AMOVA) among the 86 Bambara groundnut accessions. The AMOVA was calculated using R software. It was observed that 0.16% of the variance was due to genetic differentiation among the populations and 19% of the variance was as a result of genetic differentiation among individuals within populations, and 81% variation was within individuals in a population.

Table 4
Analysis of molecular variance based on 4,399 DArTseq SNP derived markers.

Source of variation	Degrees of freedom	Sum of squares	Mean squares	Estimated variance	Variation
Among populations	4	8976.097	2244.024	100.370	0.16%
Among individuals	4	16759.03	4189.7579	201.8491	19%
Within individuals	79	67216.13	850.8371	850.8371	81%
Total	87	92951.257	7284.619	1153.056	100%
Fst = 0.1917467					

Fst = 0.1917467

Genome-wide association study on seed weight in Bambara groundnut

Seed weight was considered for genome-wide association study using 4,399 DArTseq SNP-derived markers. Manhattan plots for seed weight was generated in GWAS indicating the most

significant associations ($-\log(p\text{-value}) > 2$) (Figure 6). A quantile-quantile (Q-Q) plot confirmed a normal distribution of the trait and the model effectively controlled false positives (Figure 6).

Association mapping was performed using R statistical software using Genome Association and Prediction Integrated Tool (GAPIT) as described by Lipka et al., (2012) (Table 5). The significant SNPs, their allelic effects, and their respective minor allele frequencies (MAF) in the population are shown in Table 5. P values were used to determine the significant association of the trait with markers (Table 5). Nine significant ($P < 0.001$) associations were detected for grain seed weight on contigs 1, 2, 3, and 5 as shown in Table 5 contributing to 10% – 15% of phenotypic variance. Nine significant SNPs were observed on contig 1, two SNPs were on contig 2 and 5, and one SNP was observed on contig 3. These markers ranged in minor allele frequency from 0.01 to 0.299. Two SNP markers bolded in Table 5 overlapped on the same position on contig 1.

Table 5
Genome wide associations of single nucleotide polymorphisms (SNPs) with seed weight

Sn.	SNP	Contig.	Position	*Allele	P value	MAF	R ² with SNP	Allelic effect estimate
1	100003989	1	36158329	G/C	0.004	0.212	0.12	7.05
2	100006122	1	36148794	A/T	0.003	0.136	0.128	-3.63
3	100014020	1	4.1E + 07	G/T	0.003	0.011	0.127	3.79
4	100014021	1	4.1E + 07	G/A	0.011	0.071	0.1	-4.64
5	4182577	2	40978951	T/C	0.001	0.293	0.152	2.16
6	4184000	2	40967085	T/C	0.004	0.31	0.12	-0.36
7	4184163	3	7342052	A/G	0.003	0.223	0.13	-6.85
8	4182351	5	39550886	C/G	0.013	0.299	0.097	6.85
9	1E + 08	5	1657936	C/G	0.003	0.011	0.127	-8.72

SNP; Single nucleotide polymorphism, *Allele corresponding to seed weight based on 92 Bambara groundnut genotypes studied in this study, MAF; Major allele frequency

Discussion

Genetic diversity information is important for crop improvement since the data enhances the efficient use of genetic variations in breeding programs by supporting the selection of the ideal parents (Bernardo, 2008). The use of molecular markers with high throughput with whole genome coverage to assess genetic diversity implies a substantial revolution (Valdisser et al., 2017). The present study demonstrated the potential of highly informative and selective DArTseq derived SNP markers for genetic diversity analysis and genome wide association study in an 'orphaned' indigenous species whose genetic sequence information is not yet available. The current study is the second attempt to study Bambara groundnut genetic diversity and population structure analysis using DArTseq SNP-derived markers.

Results of the diversity analysis indicated that Bambara groundnut accessions from Bungoma had the highest polymorphic information content (77%) (Table 2). The values were high in Busia and intermediary for Vihiga, Kilifi, and Kakamega. These results suggested that most of the Bambara groundnut accessions cultivated by farmers in western and Coastal Kenya possibly originated from one source (Bungoma). The mixture of accessions in the three clusters implied that the accessions formed a heterogenic, uneven group with variable genetic backgrounds. This suggested that farmers might have exchanged germplasm across the regions and interbreeding of the various accessions might have occurred over time. This observation is supported by the fact that Bambara groundnut originates from West Africa and was brought to East Africa by Bantus as they were moving south. In Kenya, the crop is cultivated by Bantu communities mainly the Luhya's, Kamba, and Giriama (Bamshaiye et al., 2011). The highest genetic distance was observed to be similar between populations from Vihiga and Kilifi and Vihiga and Kakamega (0.045) respectively (Table 3). This implies that though Vihiga is geographically distant from Kilifi and Kakamega, Bambara germplasm sampled from these areas are genetically similar. This could be explained by the fact that Bambara groundnut is grown by farmers as a minor crop and since it is indigenous, the same germplasm must have been exchanged by farmers in the early years and even today during barter trade, cultural functions such as dowry payment, trade among other events. The least genetic distance of 0.021 was observed between populations from Bungoma and Vihiga; Kilifi and Busia; and Kilifi and Kakamega, respectively, implying a close genetic relatedness between the populations.

Results of population structure analysis revealed three major clusters of the accessions, confirming the results of population structure analysis. The clustering of the genotypes did not follow any distinct characteristic such as geographical origin or the seed testa pattern. This implied that the germplasm sampled from western and Coastal Kenya showed a high level of homogeneity and evenness. A low mean fixation index (F_{st}) estimate value of 0.063 and a small percentage variation (6%) among populations as revealed by analysis of molecular variance were recorded (Table 4). The low mean fixation index suggested a low degree of differentiation among populations and increased levels of admixtures. It is also highly likely that alleles in the studied Bambara groundnut population were related. The low degree of differentiation among sampled populations suggested the need for widening the genetic base. Broadening of the genetic base could be achieved by the introduction of landraces from other countries such as West Africa which is the primary centre of origin of the crop. The low mean

fixation index could be attributed to the fact that Bambara groundnut is cleistogamous and highly inbreeding resulting in high percentages of selfing as reported by several authors (Mayes et al., 2019; Onwubiko et al., 2011; Kuan et al., 2017).

The low level of variation observed in this study is supported by other studies. For example, Odongo et al. (2015) who studied genetic diversity in Bambara groundnut using SSR markers reported that the highest proportion of the total variation (98%) was among individuals within accessions, and variation among accessions accounted for less than 2%. A study conducted by Ntundu et al. (2004) on analysis of genetic diversity in Bambara groundnut landraces in Tanzania using amplified fragment length polymorphism (AFLP) markers reported that genetic distances between all accessions based on Jaccard's variability index ranged from 0.1 to 0.68, with a total average of 0.3. Pasquet et al. (1999) used isozyme markers and reported that both wild and domesticated Bambara groundnut accessions were characterized by low genetic diversity, indicating that wild Bambara groundnut is the progenitor of the domesticated type. However, it is worth noting that isozymes are generally limited by the low levels of polymorphism detectable and may fail to discriminate cultivars differing only slightly in the genetic make-up (Pasquet et al., 1999. Amadou et al. (2001) also reported considerable genetic diversity among 25 African Bambara groundnut accessions from International Institute for Tropical Agriculture (IITA), Ibadan, Nigeria, using Random Amplified Polymorphic DNA (RAPD) markers, and demonstrated two main groups of accessions mainly along the lines of their geographic origin.

GWAS was conducted to detect regions of the scaffolds associated with seed weight in Bambara groundnut. The MLM model in GAPIT was selected for the analysis because the model takes into account both population structure and relative kinship—there is a significant reduction of false positives compared with the general linear model (GLM) that considers population structure only and doesn't take into account population structure or relative kinship (Kamfwa et al, 2015). The Manhattan plots for seed weight were generated in genome wide association study (GWAS) indicating the most significant associations (Table 5 and Fig. 6). The amount of genetic variance explained by these signals ranged from 10–15%. This indicated that these associations had significant effects and the genomic regions associated with seed weight in Bambara groundnut need to be investigated further. The R^2 values reported in this study that ranged from 10–15% are consistent with the genetic complexity of traits such as legume seed yield traits controlled by several genes with small but cumulative effect as described by Kamfwa et al., (2015). Nine significant ($P < 0.001$) associations were detected for grain seed weight on contigs 1, 2, 3, and 5. This observation is supported by Somta et al. (2015) who reported that QTLs on contigs 2 and 5 were associated with seed weight in mung bean *Vigna radiate* (L.) Wilczek, Azuki bean (*V. angularis* (Ohwi) Ohwi and Ohashi) and rice bean (*V. umbellata* (Thunb.) Ohwi and Ohashi) suggesting that the region of the chromosome conditioning seed weight in *Vigna* species is conserved. These regions could be novel in marker-assisted breeding of Bambara groundnut. Previous GWAS reports on seed weight in common bean had reported that seed weight is controlled by regions in contigs 2, 3, 6, 7, 8, and 11 (Wen et al., 2019).

Conclusion

Results of the present study suggested that DArTseq derived SNP markers can be used for genotyping 'orphaned' indigenous species whose sequence information is not yet available. The study revealed that Bambara groundnut accessions used in this study constituted three major gene pools. The study revealed low genetic variation among the populations but a high genetic variability among individuals and within population that can be exploited by breeders to improve the crops. Single nucleotide polymorphisms (SNPs) associated with seed weight in the studied Bambara groundnut population were found on contigs 1, 2, 3, and 5. The identified QTLs in this study could potentially be used as candidates for marker-assisted selection, to accelerate gains in breeding for seed yield in Bambara groundnut. Future studies using segregating populations at the significant SNP loci may be necessary to validate the QTLs identified in this study for grain yield and determine their value in breeding.

Declarations

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

Author Contribution

P.M conceived, conceptualized the idea of the research, and wrote the manuscript, M.K and E.G did data analyses, M.K assisted in running the laboratory experiments and N.Y supervised the work. All authors reviewed the manuscript.

Acknowledgment

This work was supported by the BecA-ILRI Hub through the Africa Biosciences Challenge Fund (ABCF) program. The ABCF Program is funded by the Australian Department for Foreign Affairs and Trade (DFAT) through the BecA-CSIRO partnership; the Syngenta Foundation for Sustainable Agriculture (SFSA); the Bill & Melinda Gates Foundation (BMGF); the UK Department for International Development (DFID) and the Swedish International Development Cooperation Agency (Sida).

Data Availability

Most of the data used to support the findings of this study are included in the article. Additional data are available from the corresponding author upon request.

References

1. Bamishaiye OM, Adegbola JA, Bamishaiye EI (2011) Bambara groundnut: an under-utilized nut in Africa. *Biotechnol Adv.* pp.60-72
2. Bernardo R. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Sci.* 2008. 48(5),1649-1664. [oi:org/10.2135/cropsci2008.03.0131](https://doi.org/10.2135/cropsci2008.03.0131)
3. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species (2011) *PLoS One.* 6,1–9. doi: 10.1371/journal.pone.0019379
4. Evanno G, Regnaut S, Goudet J.(2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14:2611–2620
5. FAO. Agricultural Statistics Database (2013) Rome: Wold Agricultural. Information Center. Available online: <https://www.fao.org/faostat/en/#data/QCL>. (Accessed on 26 March 2024)
6. Falush D, Stephens M, Pritchard J K. (2003) Inference of population structure: Extensions to linked loci and correlated allele frequencies. *Genetics.* 164:1567
7. Fatokun C, Gezahegn G, Michael A, Melaku G, Nnanna U, Olaniyi O, Muyideen Y, Ismail R, Ousmane B. (2018) Genetic diversity and population structure of a mini-core subset from the world cowpea (*Vigna unguiculata* (L.) Walp.) germplasm collection. *Sci. Rep.* 8:16035. doi:10.1038/s41598-018-34555-9
8. Govindaraj M, Vetriventhan M, Srinivasan M.(2015) Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet. Res.*, Article ID 431487, doi: [org/10.1155/2015/431487](https://doi.org/10.1155/2015/431487)
9. Gupta PK, Rustgi S, Mir RR. (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity* (Edinb). 101(1):5-18. doi:10.1038/hdy.2008.35
10. Hillocks RJ, Bennett C, Mponda OM (2012) Bambara groundnut: A review of utilization, market potential and crop improvement. *Afr. Crop Sci. J.* 20 (1): 1 –16
11. <https://www.stat.berkeley.edu/~spector/s133/Clus.html>. Performing and interpreting cluster analysis. Retrieved on 31st May 2019
12. Jaetzold, R. and Schmidt, H. (1982) Farm management handbook of Kenya. Natural conditions and farm management information. Part A, B and C. Ministry of Agriculture, Kenya
13. Kamfwa K, Cichy KA, Kelly JD (2015) Genome-wide association study of agronomic traits in common bean. *Plant Genome.* doi:10.3835/plantgenome2014.09.0059
14. Kilian A et al. (2012) Diversity Arrays Technology: a generic genome profiling technology on open platforms. In: Pompanon F., Bonin A. (eds) *Data production and analysis in population genomics. methods in molecular biology (Methods and Protocols)*, vol 888. Humana Press, Totowa, NJ
15. Kilian B, Graner A (2012) NGS technologies for analyzing germplasm diversity in gene banks. *Brief. Funct. Genomics.* 11(1):38–50. doi:10.1093/bfgp/elr046
16. Kuan HW, Chai HH, Kendabie P, Ahmad NS, Jani J, Massawe F, Kilian A, Mayes S. (2017) Integrating genetic maps in Bambara groundnut [*Vigna subterranea* (L) Verdc.] and their syntenic relationships

- among closely related legumes. BMC Genet. 18:192. doi: 10.1186/s12864-016-3393-8
17. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang (2012) GAPIT: Genome association and prediction integrated tool. J. Bioinform. 28:2397–2399. doi:10.1093/bioinformatics/bts444
 18. Massawe F, Mayes S, Cheng (2016) A. Crop diversity: an unexploited treasure for food security. Trends Plant Sci. 21(5):365-368. doi:10.1016/j.tplants
 19. Mayes S, Kuan WH, Hui H, Gao X, Kundy AC, Mateva KI, Zahrulakmal M, Hahiree M. KIM, Kendabie P, Licea LCS, Massawe F, Mabhaudh T, Modi AT, Berchie N, Amoah S, Faloye B, Abberton M, Olaniyi O, Azam-Ali SN (2019) Bambara groundnut: an exemplar underutilised legume for resilience under climate change. Planta. doi: org/10.1007/s00425-019-03191-6
 20. Mayes S, Massawe FJ, Alderson PG, Roberts JA, Azam-Ali SN, Hermann M (2011) The potential for underutilized crops to improve security of food production. J Exp Bot. 63:1075–1079. doi:10.1093/jxb/err396
 21. Morin PA. Luikart G, Wayne RK (2004) SNP in ecology, evolution and conservation. Trend Ecol Evol. 19:208–216
 22. Nadeem MA, Habyarimana EC., Çiftçi V, Nawaz MA, Karakoç T, Comertpay G, et al. (2018) Characterization of genetic diversity in Turkish common bean gene pool using phenotypic and whole-genome DArTseq-generated silicoDArT marker information. PLoS One. 13(10): e0205363. <https://doi.org/10.1371/journal.pone.0205363>
 23. National Research Council 2006 Lost Crops of Africa, Volume II: Vegetables. Policy and global affairs, Washington. New York, Academic Press. 354 pp
 24. Ndiang Z, Bell JM, Fokam PE, Ouattara B, Simo C, Dibong DS (2014) Agro-morphological variability in twelve Bambara groundnut (*Vigna subterranea* (L.) Verdc.) accessions in Cameroon. Sciences, Technologies et Development. 16; 38-45
 25. Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data II. Gene frequency data. J Mol Evol. 19:153–170. doi: 10.1007/bf02300753
 26. Ntundu WH, Inga CB, Jørgen L, Christiansen Sven B, Andersen (2004) Analysis of genetic diversity in Bambara groundnut [*Vigna subterranea* (L.) Verdc] landraces using amplified fragment length polymorphism (AFLP) markers. Afr. J. Biotechnol. Vol.3(4), pp. 220-225
 27. Odongo FO, Oyoo ME, Wasike V, Owuoch JO, Karanja L, Korir P (2015) Genetic diversity of Bambara groundnut (*Vigna subterranea* (L.) verdc.) landraces in Kenya using microsatellite markers. Afr. J. Biotechnol. vol.14 (4). pp. 283-291
 28. Onwubiko NC, Uguru MI, Ngwuta AA, Inyang ET, Nnajieme O J (2011) Floral biology of Bambara groundnut [*Vigna subterranea* (L.) Verdc]. J. Crop Sci. Vol. 3(11). pp. 293-295, 15 October
 29. Pasquet RS, Schwedes S, Gepts P (1999) Isozyme diversity in Bambara groundnut. Crop Sci. 39:1228-1236.
 30. Peakall R, Smouse PE (2012) GenAlEx 6.5. Genetic analysis in Excel. Population genetic software for teaching and research – an update. J. Bioinform. 28,2537-2539

31. Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome*. 5:92±102
32. Pritchard JK, Wen W, Falush D (2003) Documentation for structure software: 2003. Version 2
33. R Core Team (2013). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. URL <http://www.R-project.org/>
34. Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol*. 5:94–100
35. Somta P, Sompong C, Alisa K, Peerasak S (2015) QTLs controlling seed weight and days to flowering in mungbean [*Vigna radiata* (L.) Wilczek], their conservation in Azuki bean [*V. angularis* (Ohwi) Ohwi & Ohashi] and rice bean [*V. umbellata* (Thunb.) Ohwi & Ohashi]. *Agrivita*. Volume 37 No. 2
36. Sonah H, O'Donoghue L, Cober E, Rajcan I, Belzile F (2015) Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol. J*. 13, 211–221
37. Syvanen AC (2005) Toward genome-wide SNP genotyping. *Nat. Genet*. 2005;37: s5–s10
38. Thouseem N, Beena T, Namitha E (2018) Genetic parameters in seed yield components of Cowpea (*Vigna unguiculata*(L.) Walp.) N. *Int J Curr Microbiol Appl Sci*. 7(2): 2268-2274
39. Uba CU, Oselebe HO, Tesfaye AA, Abteu WG (2021) Genetic diversity and population structure analysis of Bambara groundnut (*Vigna subterrenea* L) landraces using DArT SNP markers. *PLoS ONE* 16(7): e0253600. <https://doi.org/10.1371/journal.pone.0253600>
40. Upadhyaya HD, Gowda CLL, Sastry DVS (2008) Management of germplasm collections and enhancing their use by mini core and molecular approaches. APEC-ATCWG workshop. Capacity building for risk management systems on genetic resources. p. 35–70
41. Valdisser PAMR, Wendell PJ, Almeida JE, Filho MBSF, Coelho GRC, Menezes IPP, de, VJPG, Zucchi MI, Lanna AC, Coelho ASG, Oliveir JP, Moraes AC, Brondani C, Vianello RP (2017) In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. *BMC Genomics*. 18:423. doi: 10.1186/s12864-017-3805-4
42. Wen L, Hao-Xun C, Patrick JB, Leslie L, Domier GL, Hartman (2019) Genome-wide association and genomic prediction identifies soybean cyst nematode resistance in common bean including a syntenic region to soybean Rhg1 locus. *Hortic. Res*.6:9. doi: 10.1038/s41438-018-0085-3
43. World Food Programme (2019). Protecting and rebuilding livelihoods in arid and semi-arid areas. <http://executiveboard.wfp.org>. 2012. Accessed on 15th August 2019
44. Xie H, Wen Y, Choi Y, Zhang X (2021) Global trends on food security research: A bibliometric analysis. *Land*. 10(2):119. <https://doi.org/10.3390/land10020119>

Figures

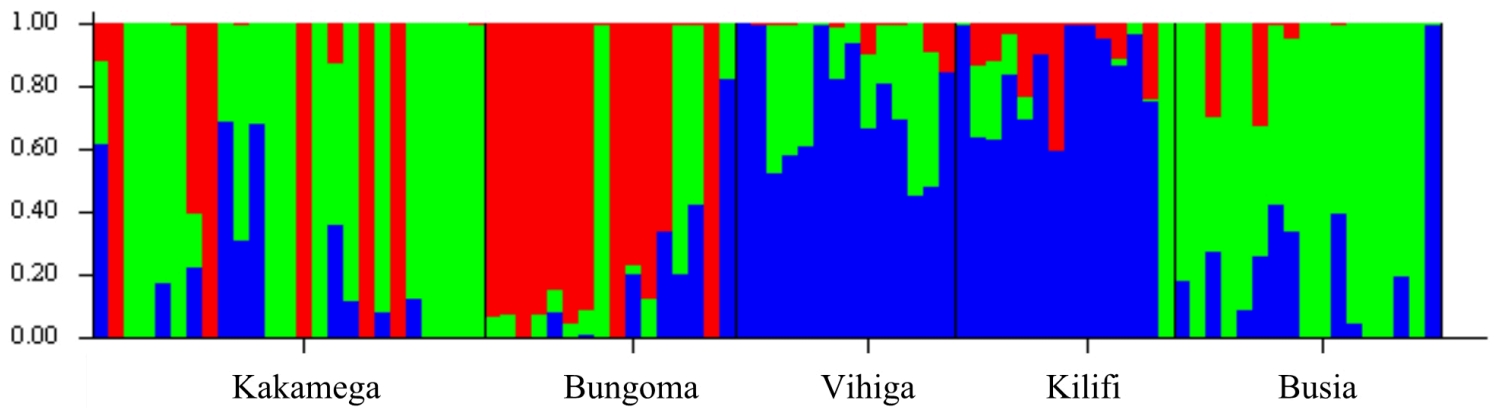


Figure 1

Distribution pattern of 86 Bambara groundnut accessions based on 4,399 DArTseq-derived SNP marker. Each colour represents a different gene pool. The barplot represents each accession as a single vertical bar broken into K colour segments, with lengths proportional to the estimated probability of membership in each inferred cluster.

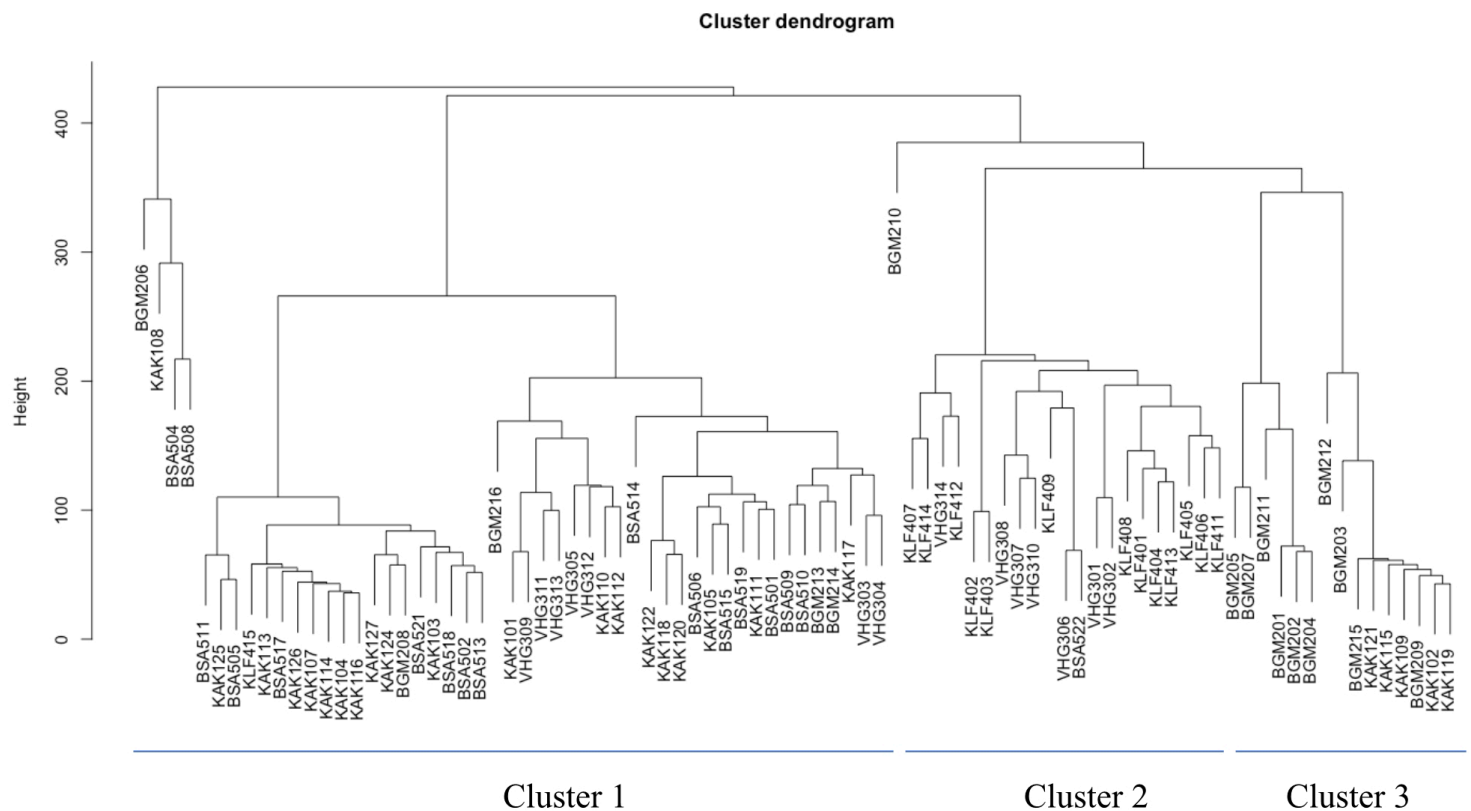


Figure 2

Hierarchical clustering of 86 Bambara groundnut accessions based on 4,399 DArTseq-derived SNP marker.

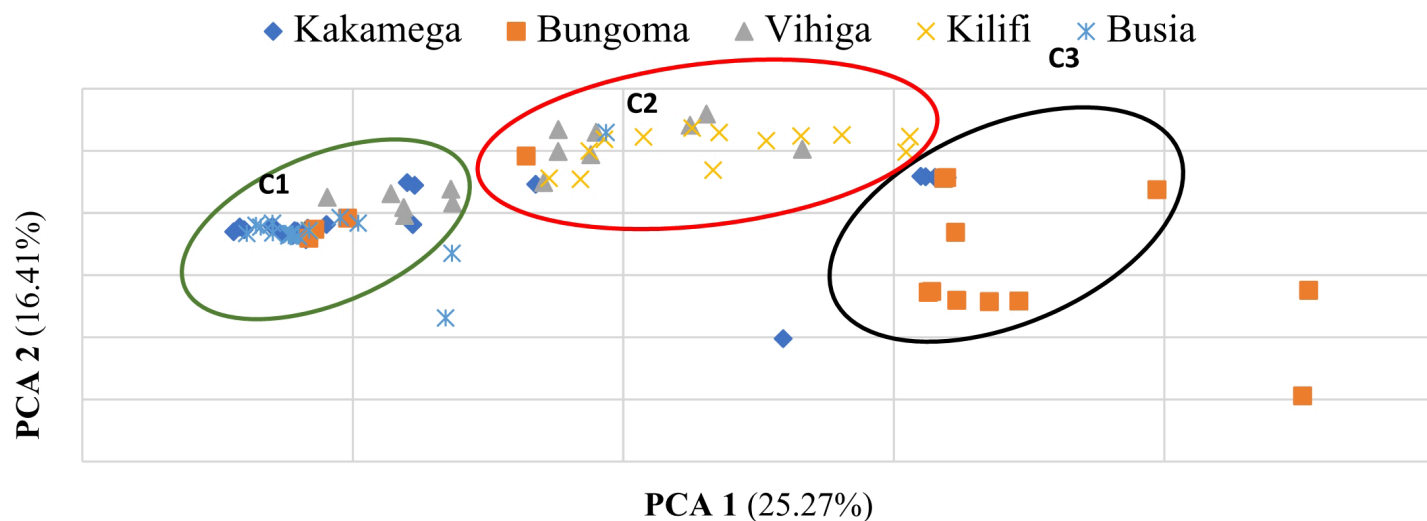


Figure 3

PCoA based on 4,399 DArTseq-derived SNP markers.

Seed weight was considered for genome-wide association study using 4,399 DArTseq SNP-derived markers. Manhattan plots for seed weight was generated in GWAS indicating the most

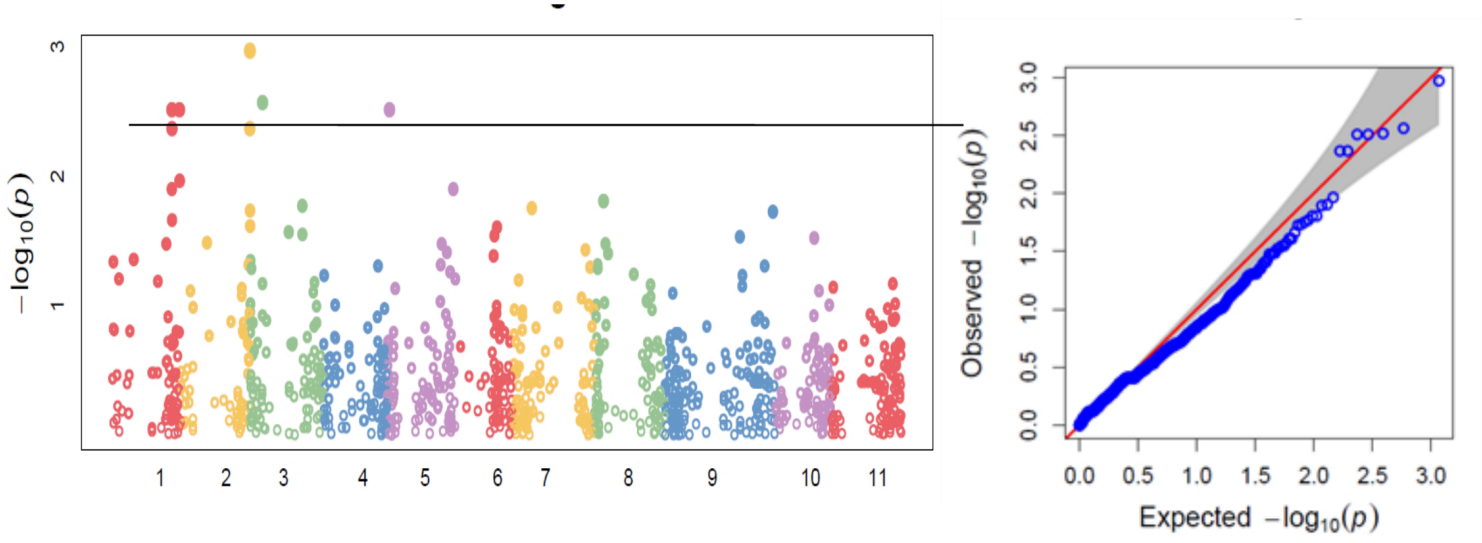


Figure 4

Figure 6. Manhattan plot and respective quantile–quantile (Q-Q) plot of hundred seed weight