Check for updates

Genetic diversity and population structure of some Nigerian accessions of Bambara groundnut (*Vigna subterranea* (L.) Verdc.,) using DArT SNP markers

Opeyemi T. Osundare · Benjamin O. Akinyele ·
Alexander C. Odiyi · Rajneesh Paliwal ·
Olaniyi A. Oyatomi · Michael T. Abberton

Abstract Bambara groundnut is one of the crops with inadequate molecular research to show its full potentials. Previous studies showed morphological diversity with inadequate information to confirm genetic variations. In the quest to reveal the genetic potentials, deoxyribonucleic acid (DNA) of the selected accessions was extracted through leaf samples at 3 weeks old, using Dellaporta Miniprep for Plant DNA Isolation procedure. The high quality DNA was sequenced using Diversity Arrays Technology (DArT) markers to unlock diversity among Bambara groundnut of Nigerian origin. Cluster analysis (neighbor-joining clustering) of the single nucleotide polymorphisms (SNP's) were used to generate sub-population to show relatedness and differences. Seven sub-populations were generated with 5927 (50.13%) high quality DArT markers out of the 11,821 SNPs generated. This revealed high genetic diversity existed among the selected Bambara groundnut accessions in Nigeria. This also revealed that DArT markers were highly efficient in classifying the accessions based on molecular expressions. This study also identified markers responsible for genetic variation that could facilitate the characterization of larger collections for further utilization of genetic resources and most importantly Bambara groundnut for the purpose of crop improvement.

Keywords Diversity · DArT markers · Bambara groundnut · Nigerian accessions

O. T. Osundare · B. O. Akinyele · A. C. Odiyi
Department of Crop, Soil and Pest Management, Federal University of Technology, Akure, Nigeria

O. T. Osundare · R. Paliwal · O. A. Oyatomi ·
M. T. Abberton (✉)
Genetic Resources Centre, International Institute of Tropical Agriculture, Ibadan, Nigeria
e-mail: m.abberton@cgiar.org

O. T. Osundare
Department of Crop Science and Horticulture, Federal University Oye-Ekiti, Oye-Ekiti, Ekiti State, Nigeria

Introduction

Bambara groundnut (*Vigna subterranea* [L.] Verdc., Syn: *Voandzeia subterranea* [L.] Thouars) is an under-utilized grain legume grown in Africa, for food security (Ntundu et al. 2006). The crop is an important legume in Africa after cowpea (*Vigna unguiculata* [L.] Walp.) (Atoyebi et al. 2017). Molecular research provides new insight on the population structure of African Bambara groundnut germplasm which will help in conservation strategy and management of the crop (Uba et al. 2021). Published research works relating to African orphan crops emphasized on the need for more molecular research to harness the hidden potentials (Mayes et al. 2013; Ntundu et al. 2006). Bambara groundnut seed is regarded as a balanced food because when compared to most food legumes, it is rich in iron and its protein contains

high level of lysine and methionine (Adu-Dapaah and Sangwan, 2004; Massawe et al. 2005). This promising crop is a weapon against food and nutritional insecurity ravaging in Africa. The current world population is over 7 billion, Africa accounts for over 1.2 billion and Nigeria being the most populated African country, accounts for over 200 million, with an average population growth above 1% yearly (World population prospect, 2019). The logical connective is an increasing food demands and decreasing arable land in the world, hence, increasing number of people and decreasing arable land. This, if not checked, everyone becomes vulnerable to food and nutritional insecurity. Bambara groundnut can be considered as one of the crop with inadequate molecular research to show its full potential for improvement (Ntundu et al. 2006; Massawe et al. 2005), when compared with cereal crops. Most available researches are based on agronomic characterization (Mayes et al. 2013; Goli 1997), while some show results of tolerance to abiotic stresses (Kouassi and Zoro 2010). Bambara groundnut is cleistogamous, highly inbreeding and has 11 pairs of chromosomes ($2n = 2x = 22$). There are numerous Bambara groundnut accessions in the genebanks that needs to be evaluated for possible improvement in identifying traits with better performance and associated genes to the existing or identified ones. Amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR) and random amplified polymorphic DNA markers (RAPD) have been reported to reveal genetic diversity but high throughput markers are needed to study genetic diversity, especially, when studying larger diverse germplasm collections (Atoyebi et al. 2017; Mohammed et al. 2014; Olukolu et al. 2012; Amadou et al. 2001). Diversity Arrays Technology (DArT) was originally developed to detect polymorphism at the recognition sites of methylation sensitive restriction enzymes (Wenzl et al. 2004). A combination of DArT approaches with next generation sequencing technologies have resulted in low cost and high throughput genotyping by sequencing, facilitating detection of variations in single nucleotide polymorphisms (SNP) in a genome (Kilian et al. 2012). SNP as a co-dominant marker in DArT will help to unlock genetic diversity and understand the population structure of some Nigeria accessions of Bambara groundnut. This work focuses on the use of DArTseq to identify genetic diversity and associated significant

genes to agro-morphological variation in Bambara groundnut to facilitate the utilization of existing genetic resources for enhanced adaptability to climate change.

Materials and methods

The study was carried out at the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria and International Institute of Tropical Agriculture, research station, located at the Institute of Agricultural Research and Training (IAR&T) Ikenne, Nigeria for three years. The site (Ibadan) is located on coordinate's 7.38°N latitude and 3.94°E longitude and it is situated at elevation 181 m above sea level. The average annual temperature is 26.5 °C and about 1311 mm of precipitation falls annually with 81% mean relative humidity. Ikenne is on coordinate's 6.87°N latitude and 3.71°E longitude and 235.2 m above sea level, has an annual rainfall of 1200 mm, 65% mean relative humidity and 21.4 °C mean temperature respectively. Seeds of one hundred accessions of Bambara groundnut were collected from the Genetic Resources Centre (GRC), of IITA in 2017 and used for the field experiments (Table 1). The accessions were collected from the regions of Nigeria for years and conserved in the genebank of IITA for germplasm exchange.

Experimental design

The experiments were carried out in two locations (Ibadan and Ikenne) for three years (2017/2018, 2018/2019 and 2019/2020). The experiments were laid out in a Randomized Complete Block Design (RCBD) with three replications and total block size was 21 m × 50 m and each plot was 1 m × 2.5 m. Inter and intra row spacing was 1.00 m and 0.25 m respectively. Data were collected on 28 traits (Table 2).

DNA extraction and genotyping

Three weeks after planting, leaf samples of each accession were collected and plant stands from which samples were collected was tagged. DNA extraction was done at the Bioscience Centre, International Institute of Tropical Agriculture, Ibadan following the procedure of Dellaporta Miniprep for Plant DNA Isolation (Weigel and Glazebrook 2009).

Table 1 List of the 100 Bambara groundnut accessions used for the experiment

S/N	Accessions	S/N	Accessions	S/N	Accessions	S/N	Accessions
1	TVSu–336	26	TVSu–570	51	TVSu–2101	76	TVSu–650
2	TVSu–644	27	TVSu–334	52	TVSu–2106	77	TVSu–579
3	TVSu–355	28	TVSu–670	53	TVSu–277	78	TVSu–590
4	TVSu–2093	29	TVSu–651	54	TVSu–330	79	TVSu–356
5	TVSu–589	30	TVSu–2094	55	TVSu–2090	80	TVSu–2102
6	TVSu–267	31	TVSu–357	56	TVSu–2105	81	TVSu–275
7	TVSu–127	32	TVSu–656	57	TVSu–576	82	TVSu–2108
8	TVSu–2104	33	TVSu–647	58	TVSu–2095	83	TVSu–361
9	TVSu–1241	34	TVSu–1239	59	TVSu–640	84	TVSu–261
10	TVSu–119	35	TVSu–256	60	TVSu–675	85	TVSu–34
11	TVSu–349	36	TVSu–359	61	TVSu–662	86	TVSu–262
12	TVSu–844	37	TVSu–331	62	TVSu–2099	87	TVSu–12
13	TVSu–278	38	TVSu–577	63	TVSu–2109	88	TVSu–586
14	TVSu–2110	39	TVSu–838	64	TVSu–179	89	TVSu–1252
15	TVSu–348	40	TVSu–366	65	TVSu–367	90	TVSu–351
16	TVSu–1222	41	TVSu–633	66	TVSu–365	91	TVSu–268
17	TVSu–1242	42	TVSu–269	67	TVSu–83	92	TVSu–347
18	TVSu–602	43	TVSu–572	68	TVSu–627	93	TVSu–2100
19	TVSu–82	44	TVSu–181	69	TVSu–346	94	TVSu–368
20	TVSu–2112	45	TVSu–287	70	TVSu–81	95	TVSu–1245
21	TVSu–173	46	TVSu–178	71	TVSu–350	96	TVSu–363
22	TVSu–263	47	TVSu–271	72	TVSu–639	97	TVSu–1246
23	TVSu–659	48	TVSu–273	73	TVSu–280	98	TVSu–2103
24	TVSu–667	49	TVSu–285	74	TVSu–129	99	TVSu–585
25	TVSu–14	50	TVSu–340	75	TVSu–333	100	TVSu–594

DNA was quantified at 260 nm with a Nanodrop spectrophotometer (NANODROP 8000, thermo scientific) and its quality was checked at 260/230 and 260/280 nm absorbance ratios and agarose gel electrophoresis respectively. The samples were sent to Diversity Arrays Technologies commercial service Ltd., Australia (www.diversityarrays.com) for a high throughput single nucleotide polymorphism (SNP) genotyping. The obtained DArT marker set was filtered on the basis of individual marker-related statistics by removing markers with inappropriate quality control parameters with call rate $\leq 80\%$ and missing data $\geq 20\%$ in TASSEL 5.0 software (Bradbury et al. 2007). The informativeness of the marker was determined using the polymorphic information content (PIC). The remaining high quality markers of 5927 (50.13%) out of 11,821 SNPs were retained for data quality and used in the analysis. DArTs were coded as 0/1 (presence and absence) and were used as different entries in TASSEL 5.0 software to construct neighbor

join phylograms. The genome wide association study (GWAS) was conducted using the phenotypic data means for the three years (2017–2020), using the GLM and MLM-PCA batch commands in TASSEL 5.0 software (Bradbury et al. 2007). Marker-trait associations (MTAs) were analyzed for the quantitative traits using a generalized linear model and a population stratification (structure) correction based on principal component analysis (PCA) (Price et al. 2006). The significance of associations between SNPs and traits was based on the threshold $P < 1.68 \times 10^{-4}$, calculated by dividing 1 by the total number of SNPs (5927) in the analysis (Li et al. 2016).

Analysis of population structure

Clustering approach was used to estimate the real number of subpopulations with correlated allele frequencies based on marker data by neighbor-joining method using TASSEL 5.0 software (Bradbury et al.

Table 2 Quantitative traits observed on Bambara groundnut accessions

SN	Traits	Measurement	Unit
1	Peduncle length	Length of peduncle using meter rule	Cm
2	Petiole length	Length of petiole from ground level to leaf stalk using meter rule	Mm
3	Number of flowers per peduncle	Number of flower present on each peduncle	Visual assessment
4	Number of trifoliate leaves	Counting of trifoliate leaves present on each selected plant	Visual assessment
5	Banner length	Length of flag or banner of flower using meter rule	Mm
6	Plant height	Distance from the ground level to longest terminal leaf using meter rule	cm
7	Internode length	Length of nodes from point to point using meter rule	mm
8	Plant spread	Taken as the widest ends of the plant using meter rule	cm
9	Terminal leaflet length	Measured as the distance from the leaf tip to the leaf stalk using meter rule	Mm
10	Terminal leaflet width	Widest ends across the leaf blade using meter rule	mm
11	Number of days to 50% flowering	Counting of days to flowering of 5 plants	Visual assessment
12	Number of days to first flowering	Counting of days to first flower emergence	Visual assessment
13	Days to maturity	Number of days maturity senescence shown	Visual assessment
14	Yield /ha	Weight of pods harvested /ha using weighing balance	kg
15	Yield/plant	Weight of pods harvested / plant using weighing balance	g
16	Number of seeds per pod	Number of seed(s) present in a pod	Visual assessment
17	Number of pods/area(Plot)	Number of pods harvested per plot	Visual assessment
18	Pod length	Length of pod in perpendicular position using Digital Vernier Caliper	mm
19	Pod width	Length of pod across ends traverse position using Digital Vernier Caliper	mm
20	Seed length	Length of pod in perpendicular position using Digital Vernier Caliper	mm
21	Seed width	Length of pod across ends traverse position using Digital Vernier Caliper	mm
22	Shell thickness	Measurement of shell thickness using Digital Vernier Caliper	mm
23	100-seed weight	Weight of 100 seeds using weighing balance	g
24	Chaff weight	Weight of shell removed using weighing balance	g
25	Seed weight per plant	Weight of seeds harvested per plant using weighing balance	g
26	Seed weight per plot	Weight of seeds harvested per plot using weighing balance	g
27	Shelling percentage	Weight of seeds/weight of harvested pods × 100	%
28	Shelled harvest	Shelling % / Total pod weight × 100	%

2007). Each genotype was assigned to subpopulation based on its membership probability.

2-Dimensional principal component analysis (2DPCA matrix quadrant)

2DPCA matrix quadrants (PC 1 and PC 2) were also carried out to reveal the distribution of the accessions based on their morphological traits expression. Classification based on 2-Dimensional principal component analysis is based on 2D image matrix, so the image matrix does not need to be transformed into a vector prior to feature extraction. Therefore, an image is constructed with the covariance matrix and its eigenvectors derived for image feature extraction which uses the first two highest contributing principal components (PC 1 and PC 2) to ensure that traits of the highest importance in contribution to variation are captured based on their influence and an image of subgroup membership of the genetic resources were constructed to reveal variations (Yanai and Ishii 2010).

Results

Population stratification

DArTseq showed that 100 accessions of Bambara groundnut were grouped into seven clusters (Table 3 and Fig. 1). DArTseq based on SNP revealed 7 clusters to enrich the selection of accessions with different colour denoting subgroup membership, indicating genetic variation. Cluster I had twenty-six accessions that had similar SNPs, cluster II had nineteen accessions, cluster III had six accessions, cluster IV had nine accessions, cluster V had fourteen accessions, cluster VI had twelve accessions and cluster VII had twelve accessions, while the remaining 7 accessions had mixed allelic patterns that could not be assigned to any of the clusters or probably could not produce enough information to meet the quality check. The accessions were clustered in accordance with relatedness and differences in their genetic constitution. Accessions in the same cluster reflect genetic relationships and most diverse accessions were out grouped.

Comparison between population structure as revealed by DArT markers and 2DPCA in three years

Figures 1, 2 Table 3 and revealed the accessions that were genetically and morphologically grouped in the same or different cluster as revealed by DArTseq or

Table 3 Clustering of Bambara Groundnut accessions based on Single Nucleotide Polymorphism (SNP) genotyping

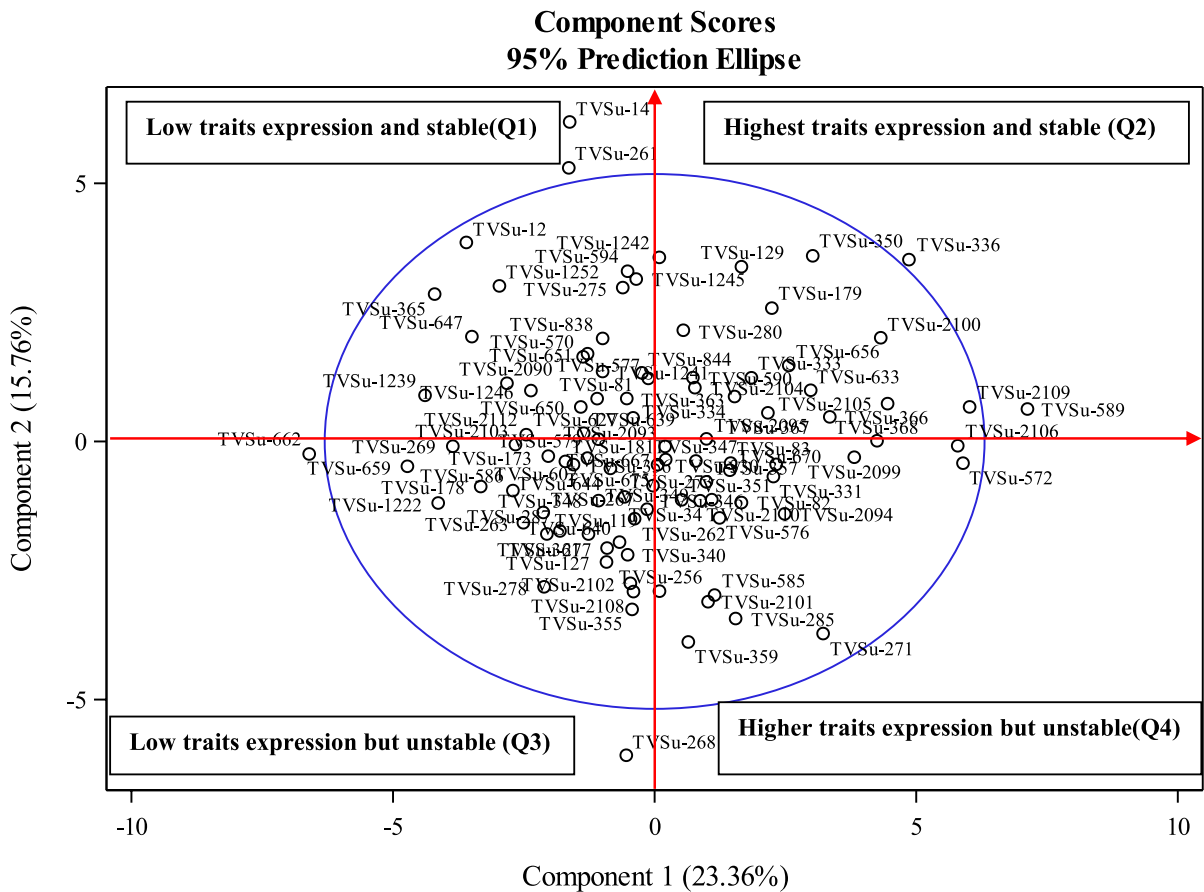| S/N | Group of clusters and the colours of differentiation | | | | | | |
	Cluster 1 (Blue)	Cluster 2 (Green)	Cluster 3 (Pink)	Cluster 4 (Purple)	Cluster 5 (Yellow)	Cluster 6 (Black)	Cluster 7 (Red)
1	TVSu-269	TVSu-262	TVSu-350	TVSu-14	TVSu-2090	TVSu-2102	TVSu-2094
2	TVSu-586	TVSu-285	TVSu-334	TVSu-261	TVSu-2112	TVSu-2104	TVSu-271
3	TVSu-2101	TVSu-277	TVSu-347	TVSu-365	TVSu-659	TVSu-349	TVSu-572
4	TVSu-119	TVSu-275	TVSu-346	TVSu-12	TVSu-662	TVSu-2093	TVSu-83
5	TVSu-585	TVSu-267	TVSu-340	TVSu-1246	TVSu-178	TVSu-2108	TVSu-34
6	TVSu-576	TVSu-278	TVSu-330	TVSu-1242	TVSu-647	TVSu-2110	TVSu-2095
7	TVSu-590	TVSu-181		TVSu-1252	TVSu-2103	TVSu-366	TVSu-639
8	TVSu-675	TVSu-273		TVSu-1241	TVSu-602	TVSu-357	TVSu-640
9	TVSu-570	TVSu-594		TVSu-1245	TVSu-627	TVSu-359	TVSu-656
10	TVSu-333	TVSu-173			TVSu-651	TVSu-355	TVSu-644
11	TVSu-331	TVSu-179			TVSu-650	TVSu-368	TVSu-82
12	TVSu-336	TVSu-268			TVSu-81	TVSu-367	TVSu-633
13	TVSu-348	TVSu-256			TVSu-1222		
14	TVSu-361	TVSu-287			TVSu-667		
15	TVSu-356	TVSu-280					
16	TVSu-351	TVSu-577					
17	TVSu-2100	TVSu-579					
18	TVSu-2105	TVSu-838					
19	TVSu-2109	TVSu-1239					
20	TVSu-589						
21	TVSu-2106						
22	TVSu-129						
23	TVSu-670						
24	TVSu-263						
25	TVSu-363						
26	TVSu-127						

Minimum genetic distance = 0.01 cM: 5927

Fig. 1 Cluster analysis of Bambara groundnut collection based on 5927 DArTseq

2DPCA quadrant. This investigation is necessary to ensure if the observed morphological variation in the selected accessions were totally the same as the classification based on molecular variation or occurred as a result of chance. In cluster or subpopulation 1 based on DArTseq, accessions TVSu-333, TVSu-336, TVSu-2100, TVSu-2105, TVSu-2109, TVSu-589, TVSu-2106, TVSu-670 were classified in the same

2DPCA matrix quadrant (Q2), based on traits contribution to morphological diversity, meaning that the aforementioned accessions were not only morphologically related but also genetically related. In cluster IV based on DArTseq, accessions TVSu-14, TVSu-261, TVSu-12, TVSu-1242, TVSu-1252 and TVSu-365 were classified in the same 2DPCA matrix quadrant (Q1) based on morphological grouping, cluster II

Component Scores
95% Prediction Ellipse

Fig. 2 2D images of the first two principal components showing overall variation among hundred accessions of Bambara groundnut in the locations and years (2017/2018, 2018/2019 and 2019/2020)

based on DArTseq showed that accession TVSu-262 and TVSu-256 were also morphologically clustered in quadrant 4 (Q4). Cluster I, II, III, V and VI also showed that accessions TVSu-269, TVSu-127, TVSu-263, TVSu-173, TVSu-181, TVSu-178, TVSu-1222, TVSu-659, TVSu-2108 and TVSu-355 were grouped in quadrant 3 (Q3) which indicated interrelationship in the population. This revealed that expression of morphological traits might be genetic and also indicated genetic similarities in accessions that were grouped in the same cluster and quadrant.

Allele frequencies and proportion based on Single nucleotide polymorphism (SNPs) genotyping

The selected accessions used in this study showed allelic variation across multiple loci and such differences in the allele frequencies produced the genetic

variation in the population. However, the values of proportion of alleles and diploids observed were small and allele frequencies very similar (Table 4). DNA quality ranges between 1.21 and 2.22.

Marker-trait association

The filtered DArTseq genotyping produced 5927SNPs from 98 lines. In the 5927 SNP filtered, 4095 SNP was unaligned on 'Mung bean genome'. In the analysis 1832 SNPs (30.90%) aligned on 'Mung bean genome'. Chromosome 1 had 158 aligned markers (8.62%), chromosome 2 had 119 aligned markers (6.49%), chromosome 3 had 120 aligned markers (6.55%), chromosome 4 had 86 aligned markers (4.69%), chromosome 5 had 203 aligned markers (11.08%), chromosome 6 had 148 aligned markers (8.07%), chromosome 7 had 203 aligned markers (11.08%), chromosome 8 had

Table 4 Gene summary based on Single Nucleotide Polymorphism genotyping

Alleles	Number	Proportion	Frequency
C	142,970	0.24614	0.25092
G	139,951	0.24094	0.24562
A	138,784	0.23893	0.24357
T	137,216	0.23623	0.24082
N	11,066	0.01905	0.01942
Y	3758	0.00647	0.0066
R	3245	0.00559	0.0057
W	1123	0.00193	0.00197
M	1118	0.00192	0.00196
S	924	0.00159	0.00162
K	691	0.00119	0.00121
C:T	856	0.14442	
G:A	805	0.13582	
T:C	783	0.13211	
A:G	776	0.13093	
T:A	293	0.04943	
G:C	266	0.04488	
A:T	265	0.04471	
C:G	261	0.04404	
A:C	251	0.04235	
C:A	245	0.04134	
G:T	244	0.04117	
T:G	217	0.03661	
G:G	172	0.02902	
C:C	168	0.02834	
A:A	163	0.0275	
T:T	162	0.02733	

Percentage of missing data and unidentified SNPs were coded N,Y,R,W,M,S,K less than 20%

Available SNPs were called at 0.80 to make 5927 SNPs

199 aligned markers (10.86%), chromosome 9 had 98 aligned markers (5.34%), chromosome 10 had 132 aligned markers (7.20%) and chromosome 11 had 134 aligned markers (7.31%) while the remaining markers were scaffolds. General linear model (GLM) revealed 1504 markers were significant. Mixed linear model (MLM) revealed 611 markers were significant and MDS revealed 3589 markers were significant in the sequence at $-\log_{10}4$ for 21 traits. Total unfiltered SNP was 11,821 and 3387 SNP (28.65%) aligned on Mung bean genome. Genome-wide association studies indicated significant SNP markers that revealed quantitative traits loci (QTLs) associated with phenotypic traits. However, many of the quantitative traits used in this study associated with more than one marker. Table 5 revealed that marker 24,383,534|F|0–6:G > A-6:G > A associated with Chaff weight, plant height, shelled harvest per Plot and shelling percentage. Marker 37,320, 448|F|0–15:G > A-15:G > A also associated with Chaff weight, plant height, shelled harvest per Plot and shelling percentage. Marker 24,383,752|F|0–61:G > T-61:G > T associated with 100-seed weight, Number of pods per plot, shelled harvest per plot and shelling percentage. Marker 24,346,601|F|0–67:T > C-67:T > C only associated with number of days to flowering. Marker 4,183,841|F|0–29:A > G-29:A > G only associated with number of flower per peduncle. Marker 24,383,752|F|0–61:G > T-61:G > T associated with number of pods per area, shelled harvest per plot, shelling percentage and 100-seed weight. Marker 24,385,974|F|0–26:C > G-26:C > G associated with number of seed per pod, terminal leaflet length and terminal leaflet width. Marker 24,346,965|F|0–27:C > T-27:C > T associated with pod length and terminal leaflet width. Marker 24,384,331|F|0–20:G > A-20:G > A only associated with pod width. Marker 27,640,259|F|0–9:A > G-9:A > G associated with petiole length, terminal leaflet length and terminal leaflet width. Marker 27,641,016|F|0–24:G > T-24:G > T associated with plant height, shelled harvest per plot, shelling percentage, terminal leaflet length, terminal leaflet width and petiole length. Marker 27,641,679|F|0–9:C > T-9:C > T only associated with seed length. Marker 24,383,723|F|0–24:C > T-24:C > T associated with only seed width. Marker 4,181,685|F|0–19:T > A-19:T > A associated with seed weight per plot and yield per hectare. Markers 24,383,752|F|0–61:G > T-61:G > T associated with Shelled harvest per plot, Shelling percentage, 100-seed weight and number of pods per area. Marker 24,383,752|F|0–61:G > T-61:G > T associated with Shelled harvest per plot, Shelling percentage, 100-seed weight and number of pods per area. Marker 24,346,321|F|0–14:G > A-14:G > A associated with terminal leaflet length and terminal leaflet width. Marker 24,346,321|F|0–14:G > A-14:G > A associated with terminal leaflet width and terminal leaflet length. Marker 4,181,685|F|0–19:T > A-19:T > A associated with seed weight per plot and yield per hectare. Marker 24,346,252|F|0–35:A > G-35:A > G only associated with yield per plant. MLM revealed 611 significant markers with associated phenotypic trait. Nineteen of the 21 phenotypic traits had significant markers associated with the traits in the analysis. Chaff weight per plot had two significant markers associated,

Table 5 DArTseq SNP markers having significant association with more than one morphological trait

Evaluated in six environments (Based on MLM)

SN	Traits	Markers	Chr. Positions	*P*-value	Marker R^2
1	Chfwt, PHT, shdhvP, shperc	24,383,534|F|0–6:G > A-6:G > A	0	6.90E − 04	0.05686
2	Chfwt, PHT, shdhvP, shperc	37,320,448|F|0–15:G > A-15:G > A	0	6.90E − 04	0.05686
3	Hsdwt, Npdspar, shdhvP, shperc	24,383,752|F|0–61:G > T-61:G > T	5	2.20E − 05	0.0512
4	NFpP	4,183,841|F|0–29:A > G-29:A > G	5	2.51E − 04	0.13848
5	NDtoF	24,346,601|F|0–67:T > C-67:T > C	4	8.98E − 04	0.13103
6	Npdspar, shdhvP, shperc, Hsdwt	24,383,752|F|0–61:G > T-61:G > T	5	4.40E − 05	3.13E-04
7	Nsdpd, TLL, TLW	24,385,974|F|0–26:C > G-26:C > G	3	4.73E − 04	0.14784
8	PdL, TLW	24,346,965|F|0–27:C > T-27:C > T	8	1.56E − 04	0.1135
9	PdW	24,384,331|F|0–20:G > A-20:G > A	0	9.23E − 04	0.06279
10	PetL, TLL, TLW	27,640,259|F|0–9:A > G-9:A > G	5	9.68E − 04	0.01115
11	PHT, shdhvP, shperc, TLL, TLW, PetL	27,641,016|F|0–24:G > T-24:G > T	11	2.02E − 04	0.01528
12	SdL	27,641,679|F|0–9:C > T-9:C > T	0	8.46E − 04	0.08146
13	SdW	24,383,723|F|0–24:C > T-24:C > T	8	2.56E − 06	0.18536
14	Sdwt, Ykgpha,	4,181,685|F|0–19:T > A-19:T > A	11	4.58E − 04	0.0256
15	ShdhvP, Shperc, Hsdwt, Npdspar	24,383,752|F|0–61:G > T-61:G > T	5	7.77E − 04	0.02701
16	Shperc, Hsdwt,Npdspar, ShdhvP	24,383,752|F|0–61:G > T-61:G > T	5	6.34E − 04	0.02654
17	TLL, TLW	24,346,321|F|0–14:G > A-14:G > A	3	2.24E − 04	0.11874
18	TLW, TLL	24,346,321|F|0–14:G > A-14:G > A	3	9.12E − 05	0.13762
19	Ykgpha, sdwt	4,181,685|F|0–19:T > A-19:T > A	11	4.02E − 05	1.90E − 05
20	Yplant	24,346,252|F|0–35:A > G-35:A > G	7	4.72E − 04	0.04166

Legend: *Chfwt* chaff weight, *PHT* Plant height, *shdhvP* shelled harvest, *Hsdwt* 100-seed weight, *Npdspar* Number of pods per area, *Shperc* shelling percentage, *NFpP* Number of flower per peduncle, *NDtoF* Number of days to flowering, *Nsdpd* Number of seed per pod, *TLL* Terminal leaflet length, *TLW* Terminal leaflet width, *PdL* Pod length, *PdW* Pod width, *PetL* Petiole length, *SdL* Seed length, *SdW* Seed width, *Ykgpha* Yield per hectare, *Yplant* Yield per plant

100-seed weight had 11 significant markers associated, number of flower per peduncle had only 1 significant marker associated, number of days to flowering had 5 significant markers associated, number of pods per plot had 9 significant markers associated, number of seed per pod had 32 significant markers associated, pod length had 45 significant markers associated, pod width had 3 significant markers associated, petiole length had 4 significant markers associated, plant height had 4 significant markers associated, seed length had 7 significant markers associated, seed width had 25 significant markers associated, seed weight per plot had 5 significant markers associated, shelled harvest per plot had 13 significant markers associated, shelling percentage per plot had 13 significant markers associated, terminal leaflet length had 119 significant markers associated, terminal leaflet width had 179 significant markers associated, yield per hectare had 3 significant markers associated and yield per plant had 131 significant markers associated. In the associated traits, terminal leaflet width had the highest associated markers, followed by yield per plant and terminal leaflet length respectively. Figures 3, 4, 5, 6, 7 and 8 showed Manhattan plots of $-\log_{10}{}^4$ versus chromosomal position of SNP markers associated with some of the phenotypic traits. A red line represents the significant threshold ($-\log_{10}{}^4$) and higher chromosomal positions above the threshold showed loci significance to variations.

Discussion

DArTseq based on single nucleotide polymorphisms (SNPs) shown seven subpopulations, indicating wide genetic variation. This reiterated the suitability of DArT high throughput sequencing for diversity studies. Olukolu et al. (2012) also reported genetic diversity for Bambara groundnut accessions using DArT
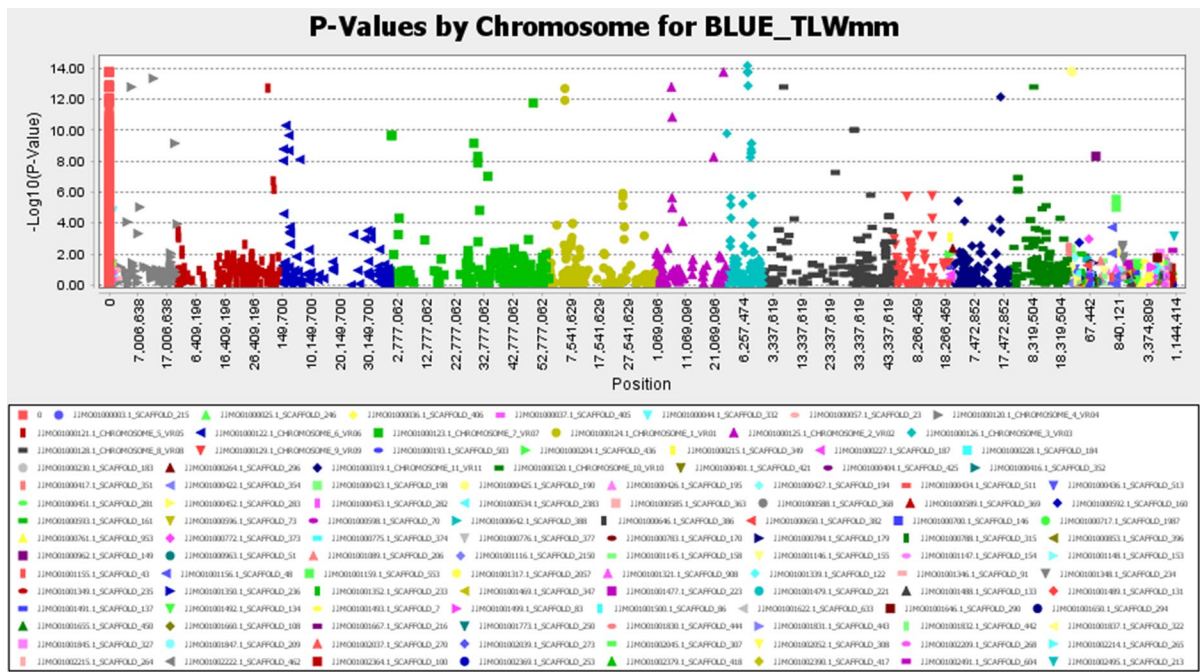
Fig. 3 Manhattan plots of $-\log_{10}^{4}$ versus chromosomal position of SNP markers associated with plant height (PHT). Significant threshold ($-\log_{10}^{4}$)

Fig. 4 Manhattan plots of $-\log_{10}^{4}$ versus chromosomal position of SNP markers associated with Terminal leaflet length (TLL). Significant threshold ($-\log_{10}^{4}$)
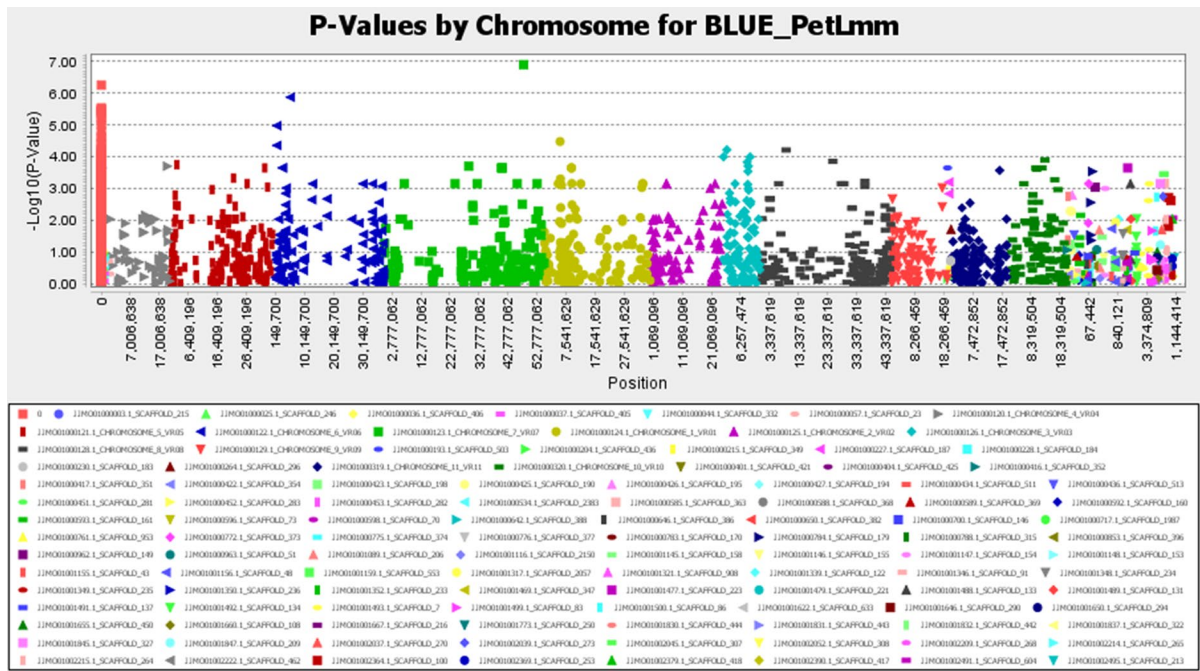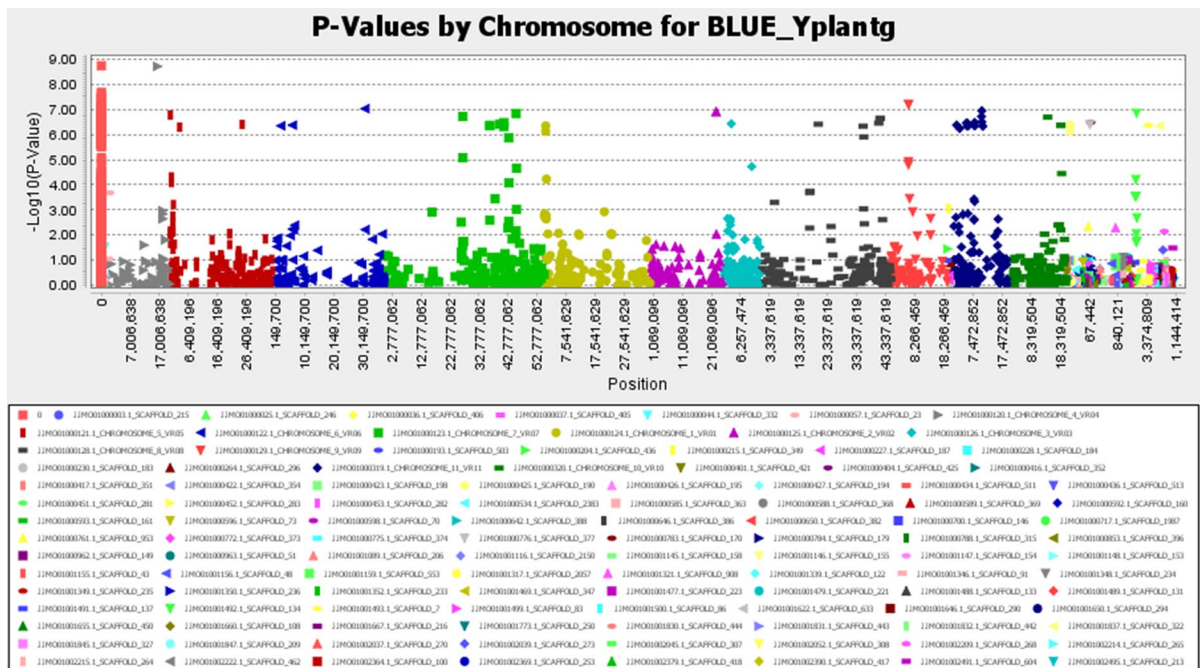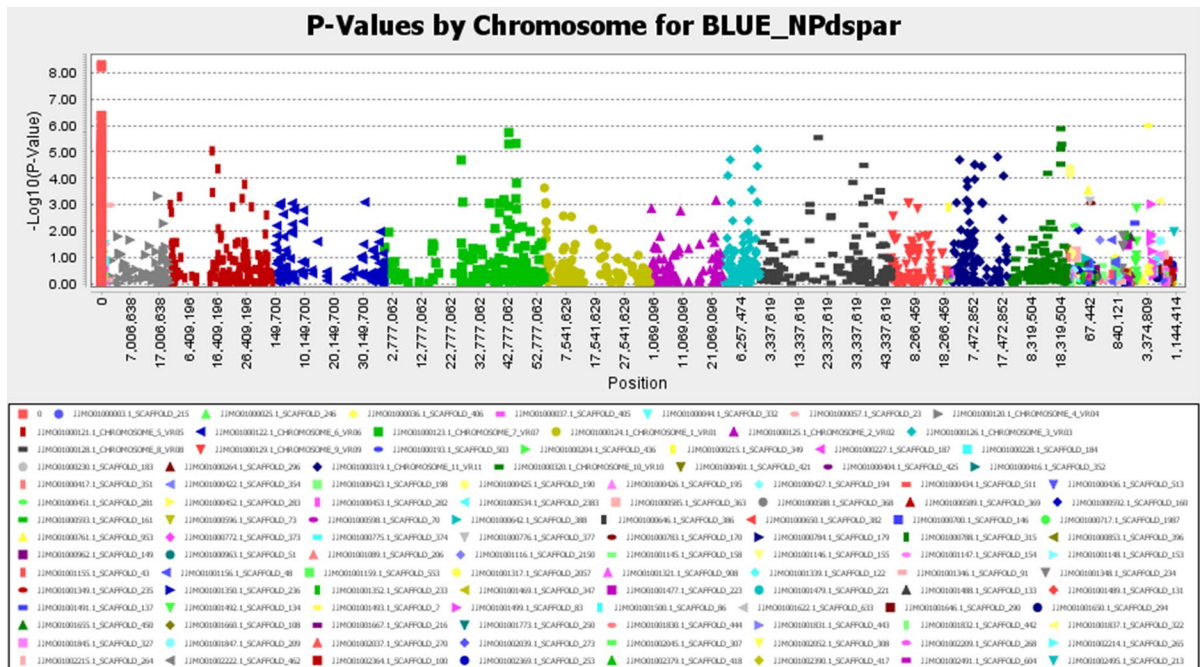
Fig. 5 Manhattan plots of $-\log_{10}4$ versus chromosomal position of SNP markers associated with Terminal leaflet width (TLW). Significant threshold ($-\log_{10}4$)

Fig. 6 Manhattan plots of $-\log_{10}4$ versus chromosomal position of SNP markers associated with petiole length (PetL mm). Significant threshold ($-\log_{10}4$)

Fig. 7 Manhattan plots of $-\log_{10}{}^4$ versus chromosomal position of SNP markers associated with yield per plant (Yplant g). Significant threshold ($-\log_{10}{}^4$)

Fig. 8 Manhattan plots of $-\log_{10}{}^4$ versus chromosomal position of SNP markers associated with number of pods per plot (NPdspar). Significant threshold ($-\log_{10}{}^4$)

markers. DArTseq revealed that accessions formed different subpopulation, indicating high genetic variation in the selected population. DArTseq also revealed some accessions in the same subpopulation; this might be due to related accession ancestry or resulting from generational out-crossing, hence, related genetic constitution. Highest number of accessions (26) in the subpopulation structure (cluster I), indicated that the highest genetic relationship existed in that subpopulation structure. Cluster III had the lowest number of accessions (6) in the subpopulation structure, indicating that the lowest genetic relationship existed in that subpopulation structure. This was also reported by Rex Bernardo, 2020 that using molecular data is a veritable tool in unlocking diversity. This indicated that accessions clustered in the same group have genetic relationship in the population. Diversity among the selected accessions shown in different subpopulation (cluster I-VII), in relation to variation expressed also indicated dissimilarities in the genetic makeup of the population. The consistency in the grouping of the accessions as expressed in the clusters indicated that the DNA bases expressed in the genetic makeup form the basis of heredity, hence, responsible for traits transfer to sustaining diversity in Bambara groundnut. Alleles percentage variation showed ≥1% genetic variation in the population data, where Cytosine (C), Guanine (G), Adenine (A) and Thymine (T) ranged from (23–24%), C and G had (24%) while A and T (23%) in the SNPs, which indicated large source of sequence variation in Bambara groundnut. Hence, SNPs are confirmative means to reveal genetic diversity. Cytosine and Guanine are the most common bases in the regions of the genome as well as in the genes with over 24% of the genetic makeup. This indicated that the differing allele frequencies are responsible for genetic variations. This was also supported by Matthew et al. (2004), who reported that SNPs with MAF greater than 5% in populations are responsible for genetic variations. Genome-wide Association Studies (GWAS) identified stable loci for 19 phenotypic traits; this indicated that these loci may be veritable tools in developing new cultivars with high yield and yield stability. The associated significant markers indicated high level of genetic variance in Bambara groundnut. This supports the findings of Choudhary et al. (2013), Huynh et al. (2013), Aliyu et al. (2016) that molecular genetic diversity analyses aid breeding

decisions in crop species. The differing significant marker sites indicated variations in the haplotypes and identified exons. Stadler (2009), Somta et al. (2011) also reported high level of allelic diversity in Bambara groundnut. However, some of the significant markers associated to more than one trait confirmed pleiotropy or linkage of the genes. The alignment of the sequence polymorphic markers on 'Mung bean genome' by some of the significant markers associated with phenotypic traits indicated that chromosome positions 5, 7 and 8 had the strongest associations with the phenotypic traits and is crucial to be considered in genetic variation of Bambara groundnut which accounted for 33.02% of variations identified through the chromosomes. Similar result was also reported by Gahlaut et al. (2019) in wheat. The alignment of the Bambara groundnut sequence on the 'Mung bean genome also indicated that 1832 SNPs (30.90%) aligned, nevertheless, some of the markers were less informative and could not be assigned to chromosome locations in the sequence, hence tagged scaffolds, this reiterated the result of Ho et al. (2017) that 48% of Bambara groundnut sequence marker tags were mapped on the 'Mung bean genome'. GWAS also indicated that many of the phenotypic traits are associated with more than one marker in different or the same chromosomal positions, meaning that the phenotypic traits are polygenic; they are determined by the interaction between several alleles. This further reiterated the fact that each marker compared the trend in the quantitative traits to the trend in the genotypes and identify loci causing variation in the genome. Further observation showed that some of the markers coded for more than one trait, i.e. single marker coding for two or more phenotypic traits, this indicated the pleiotropic nature of genes. This was also supported by Wang et al. (2008), Yongsheng Chen (2011) who corroborated that genetic trait correlation can be due to pleiotropy or linkage disequilibrium. Although, discussion of genetic traits correlation is controversial but Falconer (1960) resolved that pleiotropy is possible to be the reason for traits association than linkage. In contrast Mather and Jinks (1971) advocated that associations between traits are due to linkage. Further observation in this study as indicated at the sequence polymorphic level, showed that genetic traits association is due to pleiotropy. This was also supported by Sibov et al. (2003), Zygier et al. (2005), Chen et al. (2008) and Wang et al.

(2008). This study further revealed that some traits are associated with the same sequence polymorphic markers and chromosome position, meaning that the traits are correlated.

Conclusion

This study confirmed DArT genotyping based on SNP is effective to detect the genetic diversity and population structure of Bambara groundnut accessions. DArT markers used also revealed wide range of genetic diversity and making the selected population a potential source for Bambara groundnut improvement. Selection of high contributing morphological traits in Bambara groundnut showed differences between populations. The differing identified subpopulation of Bambara groundnut could be used in breeding improved cultivars in enhancing responsiveness to climate change. This could further be reaffirmed by using other high throughput markers in future research.

Data availability The data used to support the findings of this research are available from the corresponding author upon request.

Declarations

Conflict of interest The authors declare that we have no conflict of interest.

References

Adu-Dapaah HK, Sangwan RS (2004) Improving Bambara groundnut productivity using gamma irradiation and in vitro techniques. Afr J Biotechnol 3:260–265

Aliyu S, Massawe F, Mayes S (2016) Genetic diversity and population structure of Bambara groundnut (*Vigna subterranea* (L.) Verdc.): synopsis of the past two decades of analysis and implications for crop improvement programmes. Genet Resour Crop Evol 63:925–943. https://doi.org/10.1007/s10722-016-0406-z

Amadou HI, Bebeli PJ, Kaltsikes PJ (2001) Genetic diversity in Bambara groundnut (*Vigna subterranea* L. germplasm revealed by RAPD markers. Genome 44(6):995–999

Atoyebi JO, Oyatomi OA, Odutola O, Adebawo O, Abberton MT (2017) Morphological characterisation of selected African Accessions of Bambara Groundnut (*Vigna subterranea* (L.) Verdc.). Int J Plant Res 7(2):29–35. https://doi.org/10.5923/j.plant.20170702.01

Bernado R (2020) Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. The Genetics Society, Heredity. https://doi.org/10.1038/s41437-020-0312-1

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics (Oxford, England) 23:2633–2635

Chen J, Xu L, Cai Y, Xu J (2008) QTL mapping of phosphorus efficiency and relative biologic characteristics in maize (Zea mays L.) at two sites. Plant Soil 313:251–266

Choudhary G, Ranjitkumar N, Surapaneni M, Deborah AD, Anuradha G, Siddiq EA, Vemireddy LR (2013) Molecular genetic diversity of major Indian rice cultivars over decadal periods. PLoS ONE. https://doi.org/10.1371/journal.pone.0066197

Falconer DS (1960) An introduction to quantitative genetics. Ronald Press, New York

Gahlaut V, Jaiswal V, Singh S, Balyan HS, Gupta PK (2019) Multi-locus genome wide association mapping for yield and its contributing traits in hexaploid wheat under different water regimes. Sci Rep 9:19486. https://doi.org/10.1038/s41598-019-55520-0

Goli AE (1997) Bibliograghical Review. In: Proceedings of workshop on conservation and improvement of Bambara groundnut (*Vigna subterranea* (L.) Verdc.) IPGRI and IPK, Harare, Zimbabwe. P.4–10

Hailegiorgis D, Mesfin M, Genet T (2011) Genetic divergence analysis on some bread wheat genotypes grown in Ethiopia. J Central European Agric 12(2):344–352

Ho Wai Kuan, Chai Hui Hui, Kendabie Presidor, Ahmad Nariman Salih, Jani Jaeyres, Massawe Festo, Kilian Andrzej, Mayes Sean (2017) Integrating genetic maps in Bambara groundnut [Vigna subterranea (L) Verdc] and their syntenic relationships among closely related legumes. BMC Genomics 18:192. https://doi.org/10.1186/s12864-016-3393-8

Huynh B-L, Close TJ, Roberts PA (2013) Gene pools and the genetic architecture of domesticated cowpea. Plant Genome. https://doi.org/10.3835/plantgenome2013.03.0005

IPGRI, IITA, BAMNET (2000). Descriptors for Bambara groundnut (*Vigna subterranea*) International Institute of Tropical Agriculture, Ibadan, Nigeria/International Plant Genetic Resources Institute and International Bambara Network, Germany, ISBN:92-9043-461-9

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H (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. Humana Press, Totowa, NJ, pp 67–89. https://doi.org/10.1007/978-1-61779-870-2_5

Kouassi NJ, Zoro IA (2010) Effect of sowing density and seed bed type on yield and yield components in Bambara groundnut (*Vigna subterranea* (L.) Verdc.) in Woodland Savannas of Cote D'Ivoire. Exp Agricult 46(1):99–110

Kumar A, Gupta T, Berzsenyi S, Giangrande A (2015) N-cadherin negatively regulates collective Drosophila migration through actin cytoskeleton remodeling. J Cell Sci 128(5):900–912

Li F, Chen B, Xu K, Gao G, Yan G, Qiao J, Li J, Li H, Li L, Xiao X, Zhang T (2016) A genome-wide association study of plant height and primary branch number in rapeseed (Brassi canapus). Plant Sci 242:169–177

Massawe F, Mwale S, Azam-Ali S and Roberts J (2005) Breeding in bambara groundnut (Vigna subterranea (L.) Verdc.): Strategic considerations. African Journal of Biotechnology, Volume 4(6)463-471. ISSN 1684–5315, Academic Journals

Mather K, Jinks JL (1971) Biometrical Genetics, 2nd edn. Chapman and Hall, London

Mayes S, Basu S, Molosiwa O, Redjeki E (2013) Molecular analysis of Bambara roundnut, an underutilized African Legume Crop as part of the BAMLINK Project-What Lessons can we Learn? Acta Horticulture 979

Mohammed J, Bortolamiol-Becet D, Flynt AS, Gronau I, Siepel A, Lai EC (2014) Adaptive evolution of testis-specific, recently evolved, clustered miRNA in Drosophila. RNA 20(8):1195–1209

Mutava R, Prasad V, Tuinstra M, Kofoid K (2011) Characterisation of Sorghum genotypes for traits related to drought tolerance. Field Crop Res 123(1):10–18

Nelson MR, Marnellos G, Kammerer S, Hoyal CR, Shi MM, Cantor CR, Braun A (2004) Large-scale validation of single nucleotide polymorphisms in gene regions. Genome Res 14(8):1664–1668. https://doi.org/10.1101/gr.2421604

Ntundu WH, Shillah SA, Marandu WY, Christiansen JL (2006) Morphological Diversity of Bambara Groundnut [*Vigna subterranea* (L.) Verdc.] Landraces in Tanzania. Genetic Resour Crop Evol 53(2):367–378. https://doi.org/10.1007/s10722-004-0580-2

Olukolu B, Mayes S, Stadler F, Nyat Q, Fawole I, Dumet D, Azam-Ali S, Abbott A, Kole C (2012) Genetic diversity in Bambara groundnut (*Vigna subterranea* (L.) Verdc.) as revealed by phenotypic descriptors and DArT marker analysis. Genetic Resour Crop Evol 59(3):347–358. https://doi.org/10.1007/S10722-011-9686-5

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

Ravishanker D, Rajora AK, Greco F, Osborn HM (2013) Flavonoids as prospective compounds for anti-cancer therapy. Int J Biochem Cell Biology 45(12):2821–2831

SAS. Statistical Analysis System, version 9.4 edition (2017) SAS User's guide. SAS Institute Inc, Cary

Shegro AG, Rensburg WS, Adebola P (2013) Assessment of genetic variability in Bambara groundnut (*Vigna subterranea* (L.) Verdc.) using morphological quantitative traits. Academia J Agric Res 1(3):45–51. ISSN:2315-7739

Sibov ST, de Souza CL, Garcia AA, Silva AR, Garcia AF, Mangolin CA, Benchimol LL, de Souza AP (2003) Molecular mapping in tropical maize (*Zea mays* L.) using microsatellite markers. 2. Quantitative trait loci (QTL) for grain yield, plant height, ear height and grain moisture. Hereditas 139:107–115

Somta P, Chankaew S, Rungnoi O, Srinives P (2011) Genetic diversity of Bambara groundnut (*Vigna subterranea* (L.) Verdc.) as assessed by SSR markers. Genome 54(11):898–910

Stadler F (2009) Analysis of differential gene expression under water-deficit stress and genetic diversity in Bambara groundnut [*Vigna subterranea* (L.) Verdc.] using novel high-throughput technologies. *PhD Thesis*, Technische Universität München, Germany

Subramanian A, Subbaraman N (2010) Hierarchical cluster analysis of genetic diversity in Maize germplasm. Electron J Plant Breed 1(4):431–436

Uba CU, Oselebe HO, Tesfaye AA, Abtew 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

Wang L, Zhang L, Zhao C, Shan F, Tian Z, Li G (2008) QTL mapping of Flag leaf stage and heading date of wheat. Molecular Plant Breeding 6:689–694

Weigel D, Glazebrook J (2009) Quick miniprep for plant DNA isolation. Cold Spring Harb Protoc. https://doi.org/10.1101/pdb.prot5179

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A et al (2004) Diversity arrays technology (DArT) for whole genome profiling of barley. Proc Natl Acad Sci 101:9915–9920. https://doi.org/10.1073/pnas.0401076101

Yanai H, Ishii H (2010) International encyclopedia of education on principal component analysis. 3rd edition. ISBN: 978-0-08-044894-7. Elsevier Science

Yongsheng C (2011) Pleiotropic effects of genes involved in cell wall lignification on agronomic characters. A dissertation submitted to the graduate faculty in partial fulfillment of the requirements for the degree of Doctor of philosophy, Iowa State University, Ames, Iowa

Zygier S, Chaim AB, Efrati A, Kaluzky G, Borovsky Y, Paran I (2005) QTLs mapping for fruit size and shape in chromosomes 2 and 4 in pepper and a comparison of the pepper QTL map with that of tomato. Theor Appl Genet 111:437–445