Genetic Diversity and Population Structure of some Nigerian Accessions of Bambara Groundnut (Vigna subterranea (L.) Verdc) using DArT SNP Markers

OPEYEMI TUNDE OSUNDARE (opeyemi.osundare@fuoye.edu.ng)
Federal University Oye-Ekiti

Benjamin O. Akinyele
Federal University of Technology Akure

Alexander C. Odiyi
Federal University of Technology Akure

Rajneesh Paliwal
IITA: International Institute of Tropical Agriculture

Olaniyi A. Oyatomi
IITA: International Institute of Tropical Agriculture

Michael T. Abberton
IITA: International Institute of Tropical Agriculture

Research Article

Keywords: Diversity, DArT markers, Bambara groundnut and Nigerian accessions

Posted Date: February 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1292086/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Genetic Diversity and Population Structure of some Nigerian Accessions of Bambara Groundnut (*Vigna subterranea* (L.) Verdc) using DArT SNP Markers

Opeyemi T. Osundare¹,², Benjamin O. Akinyele¹, Alexander C. Odiyi¹, Rajneesh Paliwal², Olaniyi A. Oyatomi²

Michael T. Abberton²*

¹Department of Crop, Soil and Pest management, Federal University of Technology, Akure, Nigeria
²Genetic Resources Centre, International Institute of Tropical Agriculture, Ibadan, Nigeria

*Corresponding author: m.abberton@cgiar.org

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 were 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 5,927 (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 and Nigerian accessions

Introduction

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. Markers based on amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR) and rapid amplified polymorphic DNA markers (RAPD) have been reported to reveal genetic diversity (Atoyebi et al., 2017; Mohammed et al., 2014; Olukolu et al., 2012; Amadou et al., 2001) but concluded that it is essential to investigate larger diverse collections with high throughput markers to revealing comprehensive genetic diversity with associated traits. Diversity Array Technology (DArT) was originally developed to detect polymorphism at the recognition sites of methylation sensitive restriction enzymes (Wenzl et al., 2004) and a combination of DArT approaches with next generation sequencing technologies have resulted in low cost and high throughput genotyping by sequencing, showing the 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 enhance 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 meters 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 meters 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 I). 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 x 50 m and each plot was 1 m x 2.5 m. Inter and intra row spacing was 1.00 m and 0.25 m respectively. Twenty eight quantitative data were collected.

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). 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 ≤ 80 % and missing data ≥ 20 %. The informativeness of the marker was determined using the polymorphic information content (PIC). The remaining high quality markers of 5,
927 (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 (Bukler Lab for Maize genetics and diversity, 2018). 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 II and Figure 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

Figure 1, Table II and Figure 2 revealed the accessions that were genetically and morphologically grouped in the same or different cluster as revealed by DArTseq or 2DPCA quadrant. This investigation is necessary to ensure if the observed morphological variation in the selected accessions were genetic 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 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 genetic differentiation in the population. However, the values of proportion of alleles and diploids observed were small and allele frequencies very similar. Table IV showed that the addition of the interactions and frequency distribution of the diploid used was the same with the ‘called’ SNPs (5, 927) in this
study, 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, 0.04888), A:T (265, 0.04471), C:G (261, 0.04404), A:C (251, 0.04235), C:A (245, 0.04135), 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) and T:T (162, 0.02733).

Alleles percentage in the SNPs (C, G, A and T) ranged from 0.23 – 0.24 (23 – 24 percent), C and G had 0.24 (24 percent) while A and T had 0.23 (23 percent) in the SNPs.

Marker Data

The filtered DArTseq genotyping produced 5,927 SNPs from 98 lines. In the 5,927 SNP filtered, 4,095 SNP was unaligned on ‘Mung bean genome’. In the analysis 1,832 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 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 1,504 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 3,387 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 V revealed that marker 24383534|F|0-6:G>A-6:G>A controlled Chaff weight, plant height, shelled harvest per Plot and shelling percentage. Marker 37320448|F|0-15:G>A-15:G>A also controlled Chaff weight, plant height, shelled harvest per Plot and shelling percentage. Marker 24383752|F|0-61:G>T-61:G>T controlled 100-seed weight, Number of pods per plot, shelled harvest per plot and shelling percentage. Marker 24346601|F|0-67:T>C-67:T>C only controlled number of days to flowering. Marker 4183841|F|0-29:A>G-29:A>G only controlled number of flower per peduncle. Marker 24383752|F|0-61:G>T-61:G>T controlled number of pods per area, shelled harvest per plot, shelling percentage and 100-seed weight. Marker 24385974|F|0-26:C>G-26:C>G controlled number of seed per pod, terminal leaflet length and terminal leaflet width. Marker 24346965|F|0-27:C>T-27:C>T controlled pod length and terminal leaflet width. Marker 24384331|F|0-20:G>A-20:G>A only controlled
pod width. Marker 27640259[F]0-9:A>G-9:A>G controlled petiole length, terminal leaflet length and terminal leaflet width. Marker 27641016[F]0-24:G>T-24:G>T controlled plant height, shelled harvest per plot, shelling percentage, terminal leaflet length, terminal leaflet width and petiole length. Marker 27641679[F]0-9:C>T-9:C>T only controlled seed length. Marker 24383723[F]0-24:C>T-24:C>T controlled only seed width. Marker 4181685[F]0-19:T>A-19:T>A controlled seed weight per plot and yield per hectare. Markers 24383752[F]0-61:G>T-61:G>T controlled Shelled harvest per plot, Shelling percentage, 100-seed weight and number of pods per area. Marker 24383752[F]0-61:G>T-61:G>T controlled Shelled harvest per plot, Shelling percentage, 100-seed weight and number of pods per area. Marker 24346321[F]0-14:G>A-14:G>A controlled terminal leaflet length and terminal leaflet width. Marker 24346321[F]0-14:G>A-14:G>A controlled terminal leaflet width and terminal leaflet length. Marker 4181685[F]0-19:T>A-19:T>A controlled seed weight per plot and yield per hectare. Marker 24346252[F]0-35:A>G-35:A>G only controlled 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, 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. Figure 3-8 showed manhattan plots of $-\log_{10}^4$ vs. 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) showed 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 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 Bernado (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. The subsequent show of 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 mean 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 1,832 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 controlled by 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 genetically associated traits controlled by the same sequence polymorphic markers and chromosome position are tightly 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 genetic diversity and making the selected population a potential source for Bambara groundnut improvement. Selection of high contributing morphological traits in Bambara groundnut is genetic. 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.

**Funding**

This paper was financed by The Global Crop Diversity Trust (Crop Trust), through the Consortium of International Agriculture Research (CGIAR) and the Genetic Resources Centre, International Institute of Tropical Agriculture, Ibadan, Nigeria.

**Acknowledgement**

Appreciation to the Genetic Resources Centre, International Institute of Tropical Agriculture, Ibadan, Nigeria for the research fellowship offered during this research work and their technical assistance throughout the period.

**Data Availability**

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

**Conflict of interest**

The authors declare that we have no conflict of interest.

**References**


Bukler Laboratory for Maize genetics and diversity. 2018. TASSEL software for diversity analysis. NY


Rex Bernado. 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


groundnut \(Vigna\ subterranea\) (L.) Verdc.] using novel high-throughput technologies. PhD Thesis, Technische Universita’t Mu’nchen, Germany


Yongsheng Chen. 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

chromosomes 2 and 4 in pepper and a comparison of the pepper QTL map with that of tomato. *Theor Appl Genet* 111: 437-445
Figures

Figure 1

Please See image above for figure legend.

Figure 2

Please See image above for figure legend.

Figure 3

Please See image above for figure legend.

Figure 4

Please See image above for figure legend.
Fig. 5 Manhattan plots of $-\log_{10}^4$ vs. chromosomal position of SNP markers associated with Terminal leaflet width (TLW). A red line represents the significant threshold ($-\log_{10}^4$).

Figure 5

Please See image above for figure legend.
Figure 6 Manhattan plots of $-\log_{10}^4$ vs. chromosomal position of SNP markers associated with petiole length (PetL mm). A red line represents the significant threshold ($-\log_{10}^4$).

Figure 6

Please see image above for figure legend.

Figure 7

Please see image above for figure legend.
Fig. 8 Manhattan plots of $-\log_{10}^4$ vs. chromosomal position of SNP markers associated with number of pods per plot (NPdspar). A red line represents the significant threshold ($-\log_{10}^4$).

Figure 8

Please See image above for figure legend.