

RESEARCH ARTICLE

High-density DArT-based SilicoDArT and SNP markers for genetic diversity and population structure studies in cassava (*Manihot esculenta* Crantz)

Bright Gyamfi Adu^{1*}, Richard Akromah², Stephen Amoah², Daniel Nyadanu³, Alex Yeboah⁴, Lawrence Missah Aboagye¹, Richard Adu Amoah¹, Eva Gyamfuaa Owusu⁵

1 Council for Scientific and Industrial Research-Plant Genetics Resources Research Institute, Bunso, Ghana, **2** Department of Crop and Soil Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, **3** Cocoa Research Institute of Ghana, New Tafo-Akim, Ghana, **4** Council for Scientific and Industrial Research -Savanna Agricultural Research Institute, Tamale, Ghana, **5** Department of Statistics and Actuarial Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

☯ These authors contributed equally to this work.

‡ DN, AY, LMA, RAA and EGO also contributed equally to this work.

* adubright1992@gmail.com



OPEN ACCESS

Citation: Adu BG, Akromah R, Amoah S, Nyadanu D, Yeboah A, Aboagye LM, et al. (2021) High-density DArT-based SilicoDArT and SNP markers for genetic diversity and population structure studies in cassava (*Manihot esculenta* Crantz). PLoS ONE 16(7): e0255290. <https://doi.org/10.1371/journal.pone.0255290>

Editor: Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

Received: March 1, 2021

Accepted: July 14, 2021

Published: July 27, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0255290>

Copyright: © 2021 Adu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Abstract

Cassava (*Manihot esculenta* Crantz) is an important industrial and staple crop due to its high starch content, low input requirement, and resilience which makes it an ideal crop for sustainable agricultural systems and marginal lands in the tropics. However, the lack of genomic information on local genetic resources has impeded efficient conservation and improvement of the crop and the exploration of its full agronomic and breeding potential. This work was carried out to obtain information on population structure and extent of genetic variability among some local landraces conserved at the Plant Genetic Resources Research Institute, Ghana and exotic cassava accessions with Diversity Array Technology based SilicoDArT and SNP markers to infer how the relatedness in the genetic materials can be used to enhance germplasm curation and future breeding efforts. A total of 10521 SilicoDArT and 10808 SNP markers were used with varying polymorphic information content (PIC) values. The average PIC was 0.36 and 0.28 for the SilicoDArT and SNPs respectively. Population structure and average linkage hierarchical clustering based on SNPs revealed two distinct subpopulations and a large number of admixtures. Both DArT platforms identified 22 landraces as potential duplicates based on Gower's genetic dissimilarity. The expected heterozygosity which defines the genetic variation within each subpopulation was 0.008 for subpop1 which were mainly landraces and 0.391 for subpop2 indicating the homogeneous and admixture nature of the two subpopulations. Further analysis upon removal of the duplicates increased the expected heterozygosity of subpop1 from 0.008 to 0.357. A mantel test indicated strong interdependence ($r = 0.970$; $P < 0.001$) between SilicoDArT and DArTSeq SNP genotypic data suggesting both marker platforms as a robust system for genomic studies in cassava. These findings provide important information for efficient *ex-situ*

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

conservation of cassava, future heterosis breeding, and marker-assisted selection (MAS) to enhance cassava improvement.

Introduction

Cassava (*Manihot esculenta* Crantz) is the third most important source of calories in the tropics after rice and maize with millions of people depending on it in the world [1, 2]. Mostly grown in marginal ecologies, the crop is usually cultivated by smallholder farmers due to its ability to grow and yield in unfavourable conditions with poor soil fertility and low rainfall [1, 3]. The starchy roots contain mainly carbohydrates and the leaves are also used as a vegetable in some African countries which are cheap but a rich source of proteins, vitamins A, B and C, and other minerals [3, 4]. The cultivation of the crop continues to spread in Ghana and around the globe due to its storage roots which serve as the main raw material for industrial starch and alcohol production [5, 6]. However, marginal yields continue to be realised at the farm level which is partly due to the lack of improved varieties and the use of low yielding and environmentally sensitive lines by farmers [7]. There are also new emerging and diversified markets demand for cassava in Ghana which further suggests breeding for improved cultivars to meet specific domestic and industrial needs [8].

Germplasm banks and farmer's fields serve as reservoirs of genetic variability of crops. These collections harbour genes or germplasm with the potential to improve productivity and adaptation or tolerance to abiotic and biotic stresses [9, 10] which is particularly relevant in the current frame of climatic change and global warming. It is therefore imperative to identify true biodiversity in biological resources for efficient management, including conservation and selection of genetically divergent accessions to optimize breeding programs [11]. Genetic diversity and population structure analysis are important for characterizing the natural selection history and genetic relationships among accessions [12]. A comprehensive understanding of the genetic variability and relationships in available germplasm is a determining factor towards efficient conservation and designing breeding programmes and/or achieving breeding objectives. Several reports have highlighted the significant genetic variability within the cassava genepool for several traits associated with yield, disease resistance and drought that can be exploited for crop improvement [3, 8, 13–16].

Diversity analysis is an important component of plant breeding and genetics, conservation and evolution [17]. Most cassava diversity analyses have been based on phenotypic characters [18–22] which are mostly not reliable and environmentally plastic [23–25]. Therefore, genomic techniques may be useful in diversity assessment and selection. The use of molecular tools in plant genetic analyses and crop improvement cannot be overemphasized [26–29]. Molecular tools have proven to be more reliable in identifying duplicates among accessions during characterization which is important towards revealing genuine variability for breeding and reducing space and maintenance cost at gene banks [30, 31]. Molecular marker technologies such as Random Amplified Polymorphic DNAs (RAPDs), allozymes, single nucleotide polymorphisms (SNPs), Simple Sequence Repeats (SSRs) are available for genetic study in cassava from Ghana and other countries [32–35]. However, several degrees of limitations are also associated with these gel-based molecular marker systems which include lack of reliability and resolution (RAPD markers), poor genome coverage or labour intensive and not amenable for throughput genotyping (AFLP, RFLP) or less cost-effectiveness or require sequence information (SSRs) [36, 37]. These factors limit their applicability for many crops, especially for 'orphan' crops and polyploid species [37]. To this end, DArT markers developed through Next

Generation Sequencing platforms (NGS) are the prime alternative for molecular studies since they cover wide section of the genome with high-throughput and cost effective [38].

DArT markers (Diversity Array Technology Pty Ltd) was developed as one of the ultra-high-throughput, sequence independent, cost effective, whole-genome genotyping technique with large number of markers that cover the entire genome [36]. DArT markers have been applied successfully in genomic studies in many species including those with large and complex genomes such as barley, sugarcane, wheat, oat and strawberry [37, 39–43]. DArT markers are developed through the use of combinations of restriction enzyme digestions to reduce genome complexity, followed by next-generation sequencing of complexity reduced representations or fragments to identify DNA polymorphisms and SNPs leading to the production of thousands of polymorphic loci in a single assay [38, 44, 45]. Currently, DArT platform generates two variants of markers (SilicoDArT and DArTSeq SNP markers). SilicoDArT markers are dominant and are mostly scored for the absence (0) or presence (1) of a single allele while as DArTSeq SNPs are co-dominant markers [38, 44].

Xia et al. [46] used DArT for high-throughput genotyping of cassava and its wild relatives and suggested *PstI/TaqI* and *PstI/BstNI* as the best complexity reduction method. DArTSeq was recently used to generate a garlic core collection from the accessions kept in garlic germplasm bank, Cordoba, Spain by revealing that 31.5% of the accessions were genetically redundant [30]. These cassava germplasm have been phenotypically characterized, nonetheless, redundancy may be expected [26, 27]. Herein, the DArT markers (SilicoDArT and DArTseq SNPs) were used to analyse the genetic dissimilarities among a collection of cassava germplasm Ghana and examine the structure of the population. This will lay a foundation for efficient curation of cassava germplasm and parental selection in future breeding programs.

Materials and methods

Plant materials

A total of 87 cassava accessions were used in the study (Table 1). Sixty-two (62) of these were local cultivars comprising of local landraces (41) and locally improved/released (21) cultivars. The landraces were obtained from CSIR-Plant Genetic Resource and Research Institute of Ghana (PGRRI-Bunso). The improved cultivars were also obtained from CSIR-Crops Research Institute (Kumasi, Ghana). The remaining 25 exotic genotypes including drought tolerant populations from International Institute of Tropical Agriculture, Nigeria (IITA) were obtained through CSIR-Savanna Agricultural Research Institute, Nyankpala (CSIR-SARI, Ghana). The 87 cassava genotypes were established on the field for morphological characterization using a standard cassava descriptor in the year 2018 [25].

Molecular characterization

Extraction and quantification of DNA were carried out at the Agricultural Biotechnology Laboratory, Faculty of Agriculture, Kwame Nkrumah University of Science and Technology, Ghana while the genotyping by sequencing was done at the Diversity Array Technology Laboratory, University of Canberra, Australia.

Genomic DNA extraction. DNA samples were extracted from the youngest fully expanded leaves of each of the 87 cassava genotypes two weeks after planting using the DArT DNA extraction protocol [38]. The concentration of extracted DNA was checked using the Nanodrop 2000c spectrophotometer (NanoDrop Lite, LT2878, Thermo Scientific, USA). DNA samples were diluted between 50–100 ng/μl based on the Nanodrop readings. The samples were then packaged and shipped to Diversity Array Technology incorporation, Australia for DArTSeq genotyping.

Table 1. List of cassava accessions used in the study and some morphological attributes.

Accession Number	Name	Source/region of collection	Colour of apical leaves	Average % roots dry matter
1	Sika bankye	Locally released variety	Purplish green	34.67
2	UCC-01-464	Local, Landrace, Central Ghana	Purplish green	35.78
3	DMA-00-024	Local, Landrace, Brong Ahafo, Ghana	Dark green	41.00
4	UCC-01-144	Local, Landrace, Central Ghana	Purplish green	40.68
5	1090090	IITA	Purplish green	37.45
6	DMA-00-070	Local, Landrace, Brong Ahafo, Ghana	Dark green	41.03
7	IBA010040	IITA	Light green	38.55
8	IBA070134	IITA	Dark green	40.99
9	UCC-01-507	Local, Landrace, Central Ghana	Purple	40.57
10	UCC-01-015	Local, Landrace, Central Ghana	Dark green	33.21
11	BD-96-065	Local, Landrace, Eastern, Ghana	Dark green	39.92
12	KSI-00-191	Local, Landrace, Ashanti, Ghana	Dark green	36.42
13	AGBELIFIA	Locally released variety	Purplish green	29.75
14	BEDIAKO	Locally released variety	Light green	39.24
15	AGRA	Locally released variety	Purplish green	32.19
16	UCC-01-157	Local, Landrace, Central Ghana	Dark green	38.65
17	Doku duade	Locally released variety	Dark green	39.74
18	Abrabopa	Locally released variety	Purple	39.36
19	KSI-00-179	Local, Landrace, Ashanti, Ghana	Dark green	44.24
20	Essambankye	Locally released variety	Light green	37.21
21	IBA950289	IITA	Purplish green	34.66
22	1070557	IITA	Purplish green	30.29
23	UCC-01-195	Local, Landrace, Central Ghana	Purple	33.15
24	Broni	Locally released variety	Purplish green	36.85
25	TEKBANKYE	Locally released variety	Purplish green	38.88
26	IBA96/0581	IITA	Purplish green	32.98
27	IBA98/0505	IITA	Dark green	30.88
28	1011797	IITA	Dark green	37.35
29	TA-97-047	Local, Landrace, Brong Ahafo, Ghana	Dark green	40.13
30	Amasen	Locally released variety	Purplish green	39.22
31	KSI-00-001	Local, Landrace, Ashanti, Ghana	Purplish green	39.38
32	1082264	IITA	Dark green	29.1
33	IBA9102324	IITA	Purplish green	34.00
34	UCC BANKYE	Local, Landrace, Central Ghana	Dark green	35.32
35	IBA011368	IITA	Purplish green	33.25
36	IBA020431	IITA	Light green	39.36
37	IBA993073	IITA	Purplish green	34.06
38	Duade kpakpa	Locally released variety	Purple	39.95
39	SW-00-006	Local, Landrace, Ashanti, Ghana	Dark green	41.77
40	UCC-00-002	Local, Landrace, Central Ghana	Purplish green	37.15
41	Ampong	Locally released variety	Dark green	37.8
42	TME 149	IITA	Dark green	23.06
43	Fil-Indiakonia	Locally released variety	Purplish green	36.63
44	UCC-01-004	Local, Landrace, Central Ghana	Dark green	43.54
45	ADE-00-097	Local, Landrace, Ashanti, Ghana	Purple	27.74
46	Bankyehemaa	Locally released variety	Purplish green	40.83
47	OTUHIA	Locally released variety	Purplish green	21.58

(Continued)

Table 1. (Continued)

Accession Number	Name	Source/region of collection	Colour of apical leaves	Average % roots dry matter
48	UCC-01-296	Local, Landrace, Central Ghana	Dark green	41.74
49	DUDZI	Locally released variety	Purplish green	34.14
50	O.D	Local, Landrace, Ashanti, Ghana	Dark green	38.05
51	1083724	IITA	Purplish green	35.58
52	BD-96-157	Local, Landrace, Eastern, Ghana	Purplish green	38.56
53	IBA 011371	IITA	Purplish green	35.68
54	UCC-01-243	Local, Landrace, Central Ghana	Dark green	34.04
55	AFS-00-050	Local, Landrace, Central Ghana	Dark green	40.77
56	UCC-01-184	Local, Landrace, Central Ghana	Purplish green	30.91
57	1090151	IITA	Light green	29.36
58	Debor	Local, Landrace, Ashanti, Ghana	Dark green	35.15
59	BD-96-115	Local, Landrace, Eastern, Ghana	Purplish green	40.55
60	BIABESE	Local, Landrace, Tamale, Northern Ghana	Purplish green	32.46
61	OFF-00-023	Local, Landrace, Ashanti, Ghana	Dark green	39.8
62	Nyerikogba	Locally released variety	Purplish green	36.71
63	96/1613	IITA	Purplish green	40.27
64	IBA010034	IITA	Dark green	37.03
65	01/1206	IITA	Light green	31.97
66	IBA30572	IITA	Purplish green	35.73
67	SW-00-220	Local, Landrace, Ashanti, Ghana	Purplish green	42.22
68	ADE-00-046	Local, Landrace, Ashanti, Ghana	Light green	40.25
69	WCH-00-037	Local, Landrace, Brong Ahafo, Ghana	Purplish green	38.29
70	Afisafi	Locally released variety	Purplish green	38.81
71	UCC-01-249	Local, Landrace, Central Ghana	Dark green	42.32
72	ADE-00-038	Local, Landrace, Ashanti, Ghana	Dark green	42.94
73	1083703	IITA	Purplish green	36.1
74	BD-96-141	Local, Landrace, Eastern, Ghana	Dark green	43.75
75	Eskamaye	Locally released variety	Purplish green	40.33
76	1083774	IITA	Purplish green	38.15
77	ABASAFITA	Locally released variety	Dark green	41.11
78	OFF-00-087	Local, Landrace, Ashanti, Ghana	Dark green	36.85
79	NKABOM	Locally released variety	Dark green	36.8
80	KSI-00-036	Local, Landrace, Ashanti, Ghana	Dark green	38.88
81	IBA061635	IITA	Purplish green	37.1
82	KW-00-148	Local, Landrace, Ashanti, Ghana	Dark green	40.8
83	UCC-01-461	Local, Landrace, Central Ghana	Purple	34.99
84	NKZ-00-034	Local, Landrace, Brong Ahafo, Ghana	Purple	42.29
85	UCC-01-218	Local, Landrace, Central Ghana	Dark green	41.21
86	1085392	IITA	Light green	34.87
87	UCC-01-011	Local, Landrace, Central Ghana	Dark green	42.49

<https://doi.org/10.1371/journal.pone.0255290.t001>

DArT procedure. The DArT arrays were produced from libraries prepared from *PstI*-based genomic representations [38]. The genomic representations were generated by digesting 100 ng mixtures of DNA samples with 2 U *PstI* and a frequent cutter (*BstNI* or *TaqI*) (NEB) in a buffer containing 10 mM Tris-OAc, 50 mM KOAc, 10 mM Mg(OAc)₂ and 5mM DTT as suggested by Xia et al. [38, 46, 47] for genome complexity reduction. Fragments were sequenced on HiSeq 2500 (Illumina). Libraries were sequenced from one end by performing

single read sequencing runs [38]. The SNP markers were searched and filtered using algorithms. The sequenced data were analyzed using DarTsoft14, an automated genotypic data analysis program and DarTdb, a laboratory management system. Markers were scored '1' for presence, and '0' for absence and '-' for calls with non-zero count but too low counts to score confidently as "1" for the SilicoDArT while the DArTSeq SNPs were scored '0' for reference allele homozygote, '1' for SNP allele homozygote and '2' for heterozygote.

Marker quality parameters. Both marker systems were tested for their PIC, reproducibility (%) and call rate (%). PIC indicates the diversity of the marker in the population and showed the usefulness of the marker for linkage analysis while reproducibility involved the proportion of technical replicate assay pairs for which the marker score exhibited consistency [48]. Call rate (%) was also used to eliminate markers with $\geq 5\%$ missing data.

Genetic relationship analysis. Genetic relationships among the accessions were estimated based on Gower's dissimilarity index for the set of DArT markers [49]. Accessions were declared potential duplicates if the dissimilarity between them fell within the threshold of the replicated DNAs. The "pvclust" package was used to generate the Gower's dissimilarity matrix and "cluster" package in R was used to construct average linkage hierarchical dendrogram for SilicoDArT and DArTSeq SNPs data [50, 51]. Correlation between both marker systems was determined using the Mantel test as implemented in the "vegan" package of statistical software program 'R' by employing 10,000 random iterations in the non-parametric test calculator while the "ggplot2" package was used to generate the Mantel test scatterplot [52, 53].

Genetic diversity and population structure analysis. STRUCTURE v.2.3.4 [54] was used to analyse the genetic structure of the initial 87 cassava collection and 67 individuals upon removing off the potential duplicates. The number of hypothetical subpopulations (K) was estimated with the STRUCTURE software through the application of a Bayesian clustering approach for the organisation of genetically similar cultivars into the same subgroups. Admixture and shared allele frequencies model was used to determine the number of clusters (K) in the range from 1 to 10. For each run, the initial burn-in period was set to 20,000 followed by 30,000 MCMC (Markov chain Monte Carlo) iterations, with no prior information on the origin of individuals. Longer burn-in or MCMC has been reported not to change significantly the results [55]. The ΔK method was used to determine the most suitable value of K as implemented in Structure Harvester [55]. The structure results for the assumed population (1–10) were subsequently analysed online using the STRUCTURE HARVESTER [56] to identify a distinct peak of the curve in the change of likelihood (ΔK) at the true value of K. Principal Coordinate Analysis (PCoA) of the DArTseq markers was performed using PAST software v.3.14 [57].

Results

Quality parameters of markers

A total of 31865 SNP and 32377 SilicoDArT markers were identified upon application of the complexity reduction method with a call rate in the range of 81–100% and 70–100% respectively (S1 Fig). Around 22516 SNPs and 27182 SilicoDArT markers were assigned to 18 haploid chromosomes of cassava after aligning to cassava_v61 and v8 model reference. These ranged from 910 markers on chromosome18 (chr18) to 2158 on chromosome1 (Fig 1) for the dominant SilicoDArT markers. A total of 10521 SilicoDArT markers (S1 Table) passed the quality parameters with 100% reproducibility, and 97.4% mean call rate. The selected 10521 markers were very informative with a PIC range of 0.2–0.5 and an average of 0.36 (Fig 2).

Also, 10808 SNPs (S2 Table) with 100% reproducibility and call rate passed the quality test and were selected for further analysis. The mean PIC of these selected markers was 0.28 which

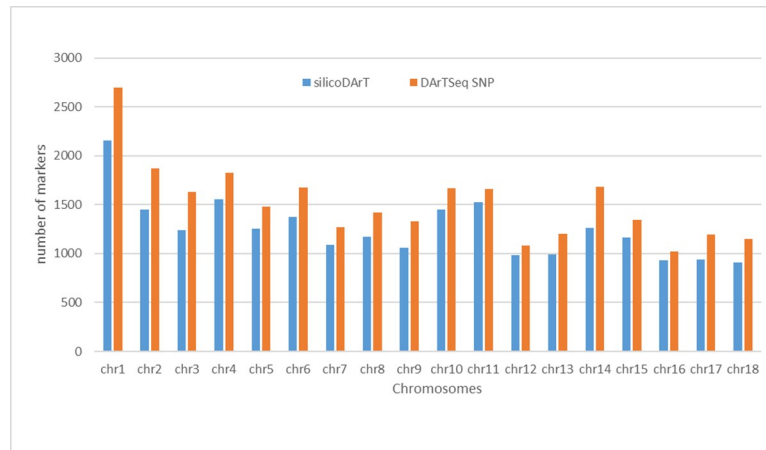


Fig 1. Genomic distribution of markers across the chromosome of cassava.

<https://doi.org/10.1371/journal.pone.0255290.g001>

was relatively lower than that of the SilicoDArT markers. Around 20.65% of the SNPs were in the lowest PIC value range of $0 < 0.10$ while 32.42% were in the highest range of $0.4-0.5$ (Fig 2). A genome-wide SNP density plot revealed that 1019 to 2696 SNPs were physically mapped to chromosome16 (chr16) and chromosome (chr1) respectively (Fig 1).

Analysis of the type of SNP (Table 2) in the selected SNPs revealed that transition SNPs (50.60%) were closely similar to transversions (49.40%). Among the six SNP types (S2 Table), A/G transitions (0.256) had the highest frequency though it was similar to C/T transitions (0.250) while G/C transversions were the least (0.101).

Genetic relationships among the cassava accessions

Genetic relationships were estimated among the accessions through their genetic distances using the selected SilicoDArT and SNP data based on Gower's genetic dissimilarity.

The overall genetic distance ranged from 0.00 to 0.41 with an average of 0.30 as revealed by the DArTSeq SNPs (S3 Table). The range of genetic dissimilarity among the IITA lines was between 0.01–0.40 with a mean of 0.33 which was similar to that of the improved ones (0.34). The lowest average genetic distance was found within the landraces (0.20). The critical distance threshold to declare whether two genotypes are duplicates/clones was empirically determined

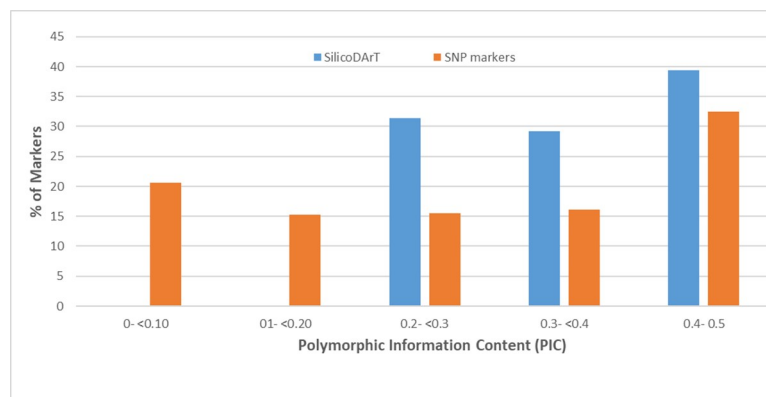


Fig 2. Percentage distribution of PIC values of selected SilicoDArT and DArTSeq markers.

<https://doi.org/10.1371/journal.pone.0255290.g002>

Table 2. Distribution of transition and transversion SNPs across the cassava genome.

Type of SNP	Transitions		Transversions			
	A/G	C/T	AC	AT	GC	GT
Number of allelic sites	2764	2706	1379	1576	1090	1293
Individual frequencies	0.256	0.250	0.128	0.145	0.101	0.120
Total	5470 (0.506)		5338 (0.494)			

<https://doi.org/10.1371/journal.pone.0255290.t002>

from the distribution of pairwise distances between replicated DNAs from three samples. The genetic dissimilarity index between the duplicated DNAs fell within the range of 0.00–0.01 and 0.00–0.02 for the SNPs and SilicoDArT respectively hence accessions with pairwise distances within these ranges were referred to as being potentially redundant. Therefore 22 landraces (UCC-BANKYE, O.D: DMA-2000-024, UCC-2001-015, BD-96-065, KSI-2000-191, UCC-2001-157, KSI-2000-179, TA-97-047, UCC-2001-004, UCC-2001-296, UCC-2001-243, AFS-2000-050, Debor, OFF-2000-023, UCC-2001-249, ADE-2000-038, OFF-2000-087, KSI-2000-036, KW-2000-148, UCC-2001-218, UCC-2001-011) constituting 53% of the total landraces obtained from the genebank were found to be potential duplicates. Debor, a known landrace was found to be similar to 19 other landraces collected from different regions. The IITA lines were generally closely related to the improved varieties than the landraces. Grouping of the accessions based on average linkage hierarchical clustering gave two main clusters (C1 and C2) containing related cassava accessions with common origin or shared parental lines (Fig 3). Several of the IITA lines (64%) predominantly grouped in C2 (green) while the landraces (90%) grouped in C1 (red). The improved varieties were evenly distributed among the two clusters.

After employing the SilicoDArT data, the genetic distance among the 87 accessions ranged between 0.0 to 0.61 (S4 Table). The average genetic dissimilarity of 0.38 among the accessions revealed by the SilicoDArT was higher than that of the SNP markers (0.30). Also, a low mean genetic dissimilarity of 0.24 was found among the landraces. Again, the 22 landraces with pairwise distance within the 0.00–0.02 range or threshold were found to be redundant confirming what was revealed by the SNPs. Ten out of these 22 came from one region (UCC, central region, Table 1) manifesting the sharing of plant material among farmers within the vicinity. Generally, the genetic distance between the groups of accessions was higher with the SilicoDArT than that observed through the SNP markers. The dendrogram obtained with the

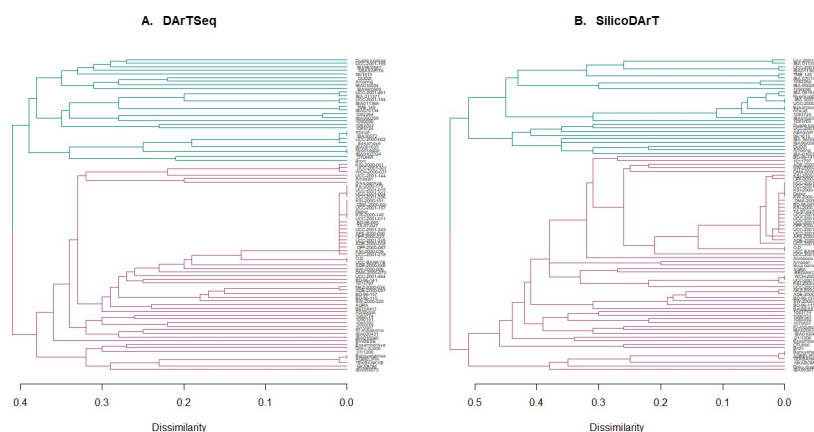


Fig 3. Average linkage hierarchical dendrogram showing genetic relationship among 87 cassava genotypes (C1 = red, C2 = green). A) Based on DArTSeq SNPs. B) Based on SilicoDArT markers.

<https://doi.org/10.1371/journal.pone.0255290.g003>

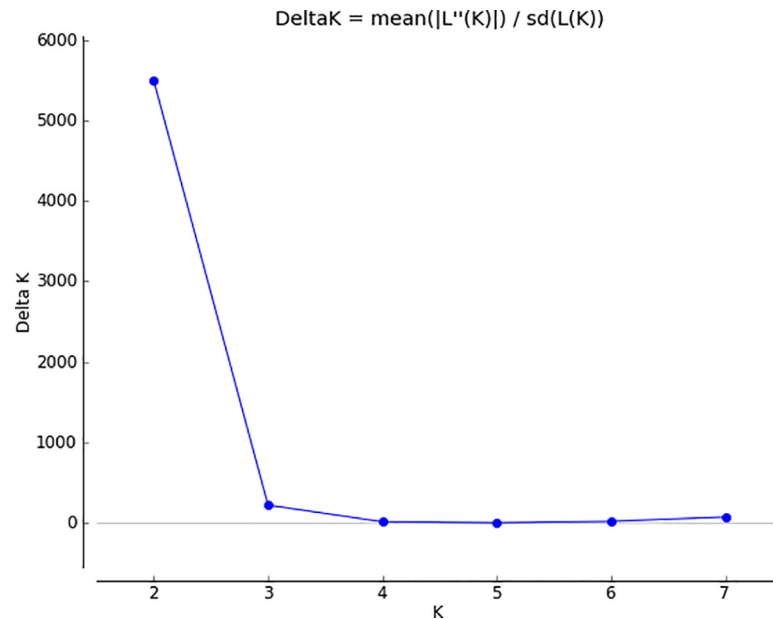


Fig 4. Delta K for different numbers of subpopulations (K).

<https://doi.org/10.1371/journal.pone.0255290.g004>

SilicoDArT markers produced two main clusters (C1 and C2) and its adjoining subclusters (Fig 3). The IITA lines mainly grouped in C2 (green) while the landraces (90%) grouped in C1 (red) just as was observed with the SNP markers.

Diversity and population structure based on DArTSeq SNP markers

The SNP markers were used for estimating the genetic structure of the cassava population using the Bayesian clustering model implemented in the computer software STRUCTURE. The simulations (logarithm probability relative to standard deviation, ΔK) estimated from the SNP markers showed a sharp peak at $K = 2$ (Fig 4) which is the real K similar to what was observed by Hampton et al. [58]. This means that the optimum number of subpopulations is two.

The population assignment test with the total 87 samples in the structure analysis shows the overall proportion of membership of the sample in each of the two clusters as illustrated in the bar plot for $K = 2$ (Fig 5). The two subpopulations (subpop1 and subpop2) consisted of 35.4% and 64.6% members respectively (Table 3). The hypothetical founder population seen in subpop1 (red) is represented by landraces while that of subpop2 (green) consisted mainly of IITA

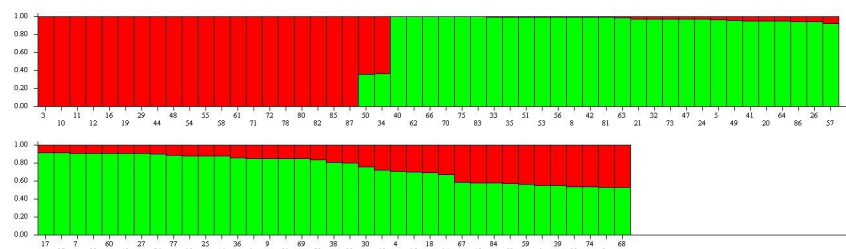


Fig 5. Estimated population structure of 87 cassava accessions on $K = 2$. Accessions in red were clustered into subpop1 and the ones in green were clustered into subpop2.

<https://doi.org/10.1371/journal.pone.0255290.g005>

and improved lines similar to what was revealed by the average linkage hierarchical clustering. Twenty accessions which were mainly landraces from Ghana were fully assigned in subpop1. These accessions were evident as C1 in the dendrogram and were the least divergent accessions based on Gower's dissimilarity. In addition, these 20 landraces were fully conserved in one group even when higher K (K = 9) was assumed (S2 Fig) indicating these accessions could be clones. On other hand, 15 lines comprising of nine IITA lines (IBA070134, IBA9102324, IBA011368, TME 149, 1083724, IBA 011371, IBA30572, IBA061635, 96/1613), three improved (Nyerikogba, Afisiafi, Eskamaye) and three putative landraces (UCC-2000-002, UCC-2001-184, UCC-2001-461) were also entirely assigned to subpop2. About 52 accessions showed admixtures (with $\geq 1\%$ of ancestry from subpop1 or subpop2) of subpop1 and subpop2 genetic composition. Majority of the IITA lines (68%) were found to be in admixture. These results conformed with the average linkage hierarchical clustering developed through the set of marker systems (Fig 3).

The expected heterozygosity from the STRUCTURE analysis was used to define the average genetic diversity between individuals within each subpopulation (subpop). The expected heterozygosity (Table 3) of subpop1 was 0.008 while that of subpop2 was 0.391 indicating the homogenous and the diverse nature of subpop1 and subpop2 respectively. Divergence among the two subpopulations shown by Net nucleotide distance (0.22) revealed that subpop1 was moderately distantly related to subpop2 (Table 3).

Further population structure analysis was performed among 67 accessions upon purging the potential duplicates. The optimum number of subpopulations was again found to be two mainly made of landraces in SUBPOP1 and IITA and improved in SUBPOP2 while 45 accessions were in admixture (S3 Fig). The expected heterozygosity increased from 0.008 to 0.357 for subpop1 while that of subpop2 remained almost the same (0.368) (Table 3)

A principal coordinate analysis (PCoA) based on the pairwise Gower's genetic distance matrix among the accessions was performed to depict the genetic divergence in the cassava lines using the two variants of DArT markers. Using the SNP markers, 38.2% of the total genetic variation was explained by the first two axes of the PCoA (Fig 6A). The first two axes of the PCoA based on the SilicoDArT (Fig 6B) explained 47.2% of the total genetic divergence. The distribution of the accessions based on the two marker systems was similar which was also consistent with the average linkage hierarchical clustering (Fig 3) and the structure analysis (Fig 5). The landraces (black) clustered together while the IITA lines (green) also clustered together. Though the improved varieties (blue) showed wide diversity, they were predominantly clustered with the IITA lines.

Association among the two DArT markers systems

Comparisons of the relationship among the accessions based on Gower's distance matrices derived from the SilicoDArT and DArTseq SNP markers depicted high association ($r = 0.970$; $P < 0.001$) between both markers systems through the Mantel correlation test (S4 Fig). The result showed a good fit between SilicoDArT and DArTseq SNP markers data sets in assessing genetic diversity in the cassava germplasm.

Table 3. Net nucleotide distance (diversity between populations) and expected heterozygosity (average diversity within populations) and proportion of membership in each subpopulation.

Subpopulations	Net Nucleotide Distance	Expected heterozygosity	Proportion of Membership	Expected Heterozygosity of 67 samples
	subpop2	of 87 samples	(87 samples)	
subpop1	0.22	0.008	0.354	0.357
subpop2	-	0.391	0.646	0.368

<https://doi.org/10.1371/journal.pone.0255290.t003>

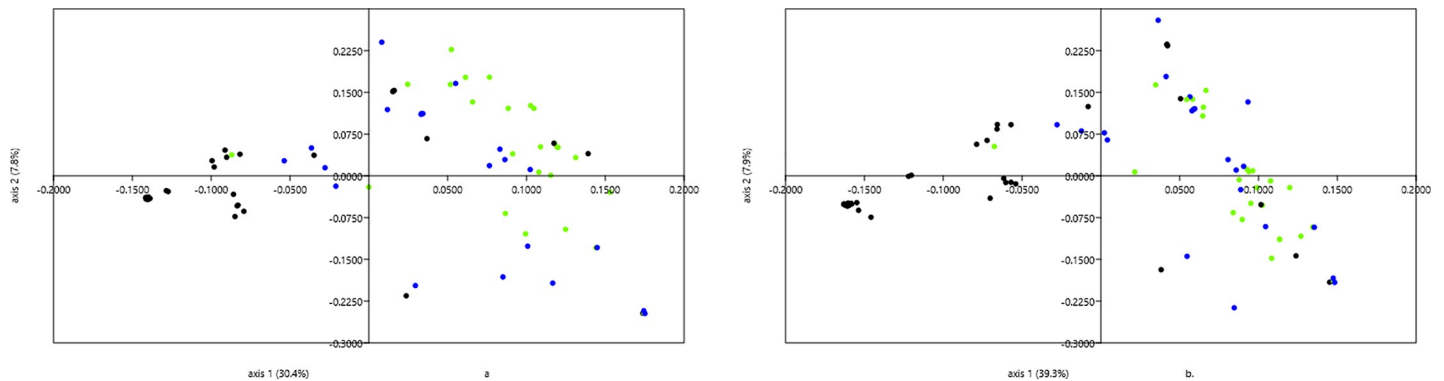


Fig 6. Principal coordinate analysis of cassava accessions (landraces = black, improved = blue, green = IITA) (a) based DArTSeq SNPs (b) based SilicoDArT markers.

<https://doi.org/10.1371/journal.pone.0255290.g006>

Discussion

Genome-wide marker discovery and quality analysis

Genome level profiling of crop germplasm collections is a critical initial step in the identification of duplicates and divergent parents for effective conservation and utilization in breeding programs. Cassava is known worldwide for domestic and industrial use however, its classification and conservation in germplasm banks is challenging, due to phenotypic plasticity of cassava being further complicated by its asexual life-cycle [25, 59]. This study highlights the potential of highly informative and selective SilicoDArT and DArTSeq SNP markers for genomic studies in cassava which might underpin conservation and future breeding efforts. A total of 10521 SilicoDArT markers and 10808 informative DArTSeq SNPs were used for genetic diversity studies and population structure analysis. Future cassava breeding programs depend on the usage of a large number of these high-throughput markers for effective selection and genome association studies. The quality parameters of the selected markers were comparable with that of others plant species. The mean PIC of 0.36 for the dominant SilicoDArT markers was lower than the 0.42 identified in cassava and its wild relatives by Xia *et al.* [46] and 0.41 found in sorghum [60] but higher than the 0.28 in *Beta vulgaris* [61] and 0.29 in macadamia [48]. The average PIC of the SNPs (0.28) on the other hand was higher than the 0.228 identified in cassava [35], 0.21 in macadamia and 0.265 in Durum wheat [62]. The average PIC value of SilicoDArT was greater than that of SNP markers similar to that reported by Alam *et al.* [48] in Macadamia indicating the SilicoDArT were more informative than the SNP markers. The use of these high-density SilicoDArT and SNP markers may achieve better genome coverage through the sampling of a greater number of points in the whole genome, as marker density is reported to have a high correlation with gene density [38, 63]. Earlier reports on diversity studies on cassava utilized relatively smaller number of molecular markers; 35 SSR [33], 26 SNPs [35] and 4 RAPD [64], hence these high-density SilicoDArT and SNP markers may better suit for robust genomic and conservation activities in cassava. Additionally, the co-dominant inheritance pattern of SNP markers may increase the utility of DArT platforms for genetic population analysis. Relative to other marker technologies, DArT markers are suitable for high-throughput work as well as being cost effective [46, 65].

Similar to previous studies with relatively fewer SNPs within genes involved in cyanogenesis (CYP79D2), starch metabolism (GBSSII) and defense pathways within cassava, transition SNPs were similar to transversions with the most frequent transition and transversion being A/G and A/T, respectively [35]. Higher DNA polymorphisms are expected from out-crossing and inbreeding-sensitive crops like cassava partly due to the inherently high number of loci

maintained in a heterozygous state. This contradicts what was reported in crops like *Camelina sativa* [12], rubber tree [66], *Brassica napus* [67] where transitional SNPs substantially exceeded transversion SNPs.

Genetic relationship among accessions

The assessment of genetic diversity is a vital pre-breeding activity towards crop improvement and efficient conservation of crop biodiversity. A genetic distance approach based on Silico-DArT and DArTSeq SNPs were successfully used to ascertain the relationship among the accessions as well as revealing potential duplicates. Similar to an earlier report by Alam et al. [48], the average genetic distance among the cassava accessions was higher with the dominant SilicoDArT than the SNPs (S3 and S4 Tables). Two major clusters were formed using both marker systems. The dendrogram (Fig 3) created using the average linkage hierarchical clustering method for both sets of markers grouped the landraces into C1 (red) and the IITA lines in C2 (green) showing their relationship with their pedigrees.

Base on the markers, the highest average dissimilarity index was found among the improved lines which were uniformly distributed among the two clusters probably due to their diverse pedigree or origin. Ferguson et al. [68] found elite lines from ESC Africa and IITA breeding lines to be more closely related and indicated that, this could be as a result of the movement of germplasm from IITA to ESC Africa through collaborations. Similar to the situation in Ghana, the known improved cassava varieties documented in the *Catalogue of crop varieties released and registered in Ghana* had their pedigree/line from IITA and/or the local landraces so it was not surprising the improved lines were evenly distributed among the two clusters [69]. Cassava, unlike crops such as maize, landraces are often indistinguishable from improved clones and are often considered for release hence, breeding has not significantly separated improved clones from landraces [68].

The landraces on the other hand had the lowest mean dissimilarity index due to the high level of duplicates that were detected within the group. Both marker systems identified 22 landraces to be within the distance threshold of the replicated DNAs. These were identified as potential duplicates. The lack of a formal seed system for cassava production promotes the sharing of planting materials among farmers as most farmers use their own planting materials (usually stem cuttings from the preceding crop) or they source stem cuttings from neighboring farmers especially for varieties with good culinary attributes for cultivation [70, 71]. These result in the renaming of the accessions leading to increased synonymy and homonymy which may confound the true diversity within the accessions when relying on the use of local/vernacular names alone [70, 72, 73]. Most of these landraces classified as potential duplicates were independently collected from different regions of Ghana and therefore came with different accession names. Though morphologically characterized, the low resolution of morphological markers could not reveal the redundant accessions [25, 59, 74]. Debor has excellent culinary traits, two of which are mealiness after boiling and relatively sweet taste hence it was not surprising it was found to be redundant with 19 landraces. This aligns with the report by Rabbi et al. [71] between Debor and other cassava genotypes. The complementary results by the SilicoDArT and SNP markers provide enough evidence to cull/rogue off redundant lines for efficient conservation and subsequent breeding.

Structure of the population

Population structure analysis provides helpful information in maintaining and monitoring the genetic diversity required for a robust breeding program [75]. Results of population structure analysis among the original 87 samples revealed only two major subpopulations (Fig 5)

analogous to the report by Adjabeng-Danquah et al. [33]. They studied some of these accessions with 35 SSR markers but indicated a clear genetic divergence based on the origin of the cassava genotypes. The grouping of the landraces based on the Bayesian model showed similar results as reflected in average linkage hierarchical clustering (Fig 3) and principal coordinate analysis. The genetic structure present in the population meets our expectations based on the sources/pedigree of the genotypes. All the genotypes were originally collected from two different locations (1. Landraces + improved = local, 2. IITA). Subpop1 consisted of only landraces while the IITA and other genotypes were clustered in subpop2.

In addition to these distinct two subpopulations, a large number of admixtures was obtained among the cassava population studied [12, 33]. The expected heterozygosity was used to express the genetic diversity between individuals within each subpopulation (Table 3). The expected heterozygosity of subpop1 was 0.008 which was very low indicating the large homogeneity of the individuals in the subgroup. This was also seen in the low level of admixture in subpop1 as revealed by the STRUCTURE results (Fig 5). As evidenced in the cluster analysis, the 22 landraces found to be redundant based on Gower's dissimilarity constituted subpop1 hence the low genetic diversity between the genotypes in the subgroup was not surprising. This shows a true reflection of the exchange of plant materials with good economical and culinary traits among farmers within the country resulting in the generation of such duplicates and the high resolution of molecular markers over morphological ones [59, 76, 77]. In contrast to subpop1, the average diversity between individuals within subpop2 was 0.391 which shows the significant divergence within the subgroup (Table 3). This level of differentiation of individuals inside the group indicates that even among individuals clustered by genetic proximity, there is still high variability [78]. The high level of admixture found in subpop2 again reflects the genetic variation within the group [62]. The high level of admixture found in the IITA lines in subpop2 corroborates what was reported by Adjabeng-Danquah et al. [33]. Also, the diverse background (source) of the accessions in subpop2 relative to subpop1 could influence the amount of variation between individuals in the subpopulation. The collections obtained from IITA could likely be made up of accessions originating from various geographical areas in West Africa [68]. As witnessed in most of the released improved cassava varieties in Ghana, they had their parental sources from IITA, indicating the formal exchange of breeding materials across borders [69] which could lead to 'secondary' mixing of the gene pool of the collections through hybridization and enhanced variation.

The variation between the two subpopulations defined by the Net nucleotide distance (Allele-frequency divergence among populations, 0.22) showed a moderate genetic differentiation between subpop1 and subpop2 (Table 3). Therefore, the genetic variation in the cassava germplasm is captured within and between subpopulations while a high level of similarity also exists within the subpopulation (subpop1). Adjabeng-Danquah et al. [33] reported a high level of variation within cassava groups similar to what was found here.

Further population structure analysis without the potential duplicates gave a higher average diversity within subpop1 (expected heterozygosity, 0.357 for SUBPOP1 and 0.368 for SUBPOP2) which was similar to the 0.333 average reported by Ferguson et al. [68] within groups. This supports the earlier revelation of a high level of duplicates among the landraces resulting in low expected heterozygosity in subpop1 and the need to purge them. This seems to be a common phenomenon among gene banks since the earlier characterizations of collections were based on agro-morphology which are phenotypically plastic or limited DNA-based markers [24, 25]. Similarly, a recent study found several duplicates within the gene bank at IITA, and efforts are being made to purge these duplicates [68]. This information is necessary particularly for the selection of individuals to serve as parents in breeding programs.

The set of DArT platforms were very informative in revealing the genetic variations in the germplasm as earlier reported [38, 48]. This was confirmed by the high correlation index revealed by the Mantel test that was conducted to check the association between both marker systems (S3 Fig). The consistency in genetic relationships among the cassava by the Silico-DArT and SNPs suggest that the two DArT marker systems are highly reliable for genetic diversity study in cassava. The available population structure is a step towards efficient germplasm conservation and parental selection to take advantage of heterosis breeding.

Conclusion

In this study, high-density dominant SilicoDArT and codominant DArTSeq SNP markers were used to explore genetic diversity and population structure among a collection of cassava. Both DArT markers successfully revealed the parental relationships and the extent of diversity in the population by showing some degree of genetic diversity and duplicates in the collection. There was a high level of redundancy in the local landraces (53%) compared to those obtained from IITA. This level of genetic diversity could be the basis for developing new cassava cultivars with desirable characteristics while improving the *ex situ* cassava germplasm conservation activities at the gene bank. In addition, our study identified two subpopulations that could be explained by the pedigrees of the genotypes. The Structure analysis indicated subpop2 to be more diverse than subpop1 (0.008) based on expected heterozygosity. Further analysis following the removal of the potential duplicates increased the expected heterozygosity from 0.008 to 0.357. Both DArT platforms seem an inexpensive and robust option for genomics studies in cassava. The large number of highly polymorphic markers developed and the knowledge of population structure and genetic diversity of cassava accessions will be important for cassava germplasm curation, heterosis breeding and future marker-traits association studies and genomic selection.

Supporting information

S1 Fig. Call rate distribution of the two marker system.

(TIF)

S2 Fig. Population structure bar plot of 87 cassava accessions on $K = 9$.

(TIF)

S3 Fig. Population structure bar plot of 67 cassava accessions on $K = 2$.

(TIF)

S4 Fig. Mantel correlation test between SilicoDArT and SNP.

(TIF)

S1 Table. List of silicoDArT markers.

(XLSX)

S2 Table. List of DArT-based SNP markers.

(XLSX)

S3 Table. Gower's dissimilarity index among the cassava accessions based on SNP markers using the "cluster" package in R software.

(CSV)

S4 Table. Gower's dissimilarity index among the cassava accessions based on SilicoDArT markers using the "cluster" package in R software.

(CSV)

Acknowledgments

The authors thank Fuseini Mohammed and Edwin Appiah Moses in the extraction of DNA, data management and the anonymous reviewers for their insightful suggestions. The support from Alliance for Green Revolution Africa (AGRA) through the IMCDA programme at KNUST (2018) is also highly acknowledged.

Author Contributions

Conceptualization: Bright Gyamfi Adu.

Data curation: Bright Gyamfi Adu, Stephen Amoah, Alex Yeboah, Richard Adu Amoah, Eva Gyamfuaa Owusu.

Formal analysis: Bright Gyamfi Adu, Daniel Nyadanu, Richard Adu Amoah, Eva Gyamfuaa Owusu.

Funding acquisition: Richard Akromah, Stephen Amoah.

Investigation: Bright Gyamfi Adu.

Methodology: Bright Gyamfi Adu, Richard Akromah, Stephen Amoah, Eva Gyamfuaa Owusu.

Project administration: Richard Akromah.

Resources: Lawrence Missah Aboagye.

Software: Bright Gyamfi Adu, Alex Yeboah, Richard Adu Amoah, Eva Gyamfuaa Owusu.

Supervision: Richard Akromah, Stephen Amoah.

Validation: Richard Akromah, Stephen Amoah, Lawrence Missah Aboagye.

Visualization: Stephen Amoah, Daniel Nyadanu.

Writing – original draft: Bright Gyamfi Adu, Lawrence Missah Aboagye.

Writing – review & editing: Bright Gyamfi Adu, Stephen Amoah, Daniel Nyadanu, Alex Yeboah.

References

1. Ceballos H, Morante N, Sanchez T, Ortiz D, Aragon I, Chavez AL, et al. Rapid cycling recurrent selection for increased carotenoids content in cassava roots. *Crop Sci.* 2013; 53: 2342–2351.
2. FAO, “FAO Statistics (FAOSTAT)”, 2014. ><http://faostat.fao.org/CountryProfiles> (Accessed 20 May, 2018)
3. El-Sharkawy MA. Physiological characteristics of cassava tolerance to prolonged drought in the tropics: implications for breeding cultivars adapted to seasonally dry and semiarid environments, *Braz. J. Plant Physiol.* 2007; 19: 257–286.
4. Fregene M, Bernal A, Duque M, Dixon A, Tohme J. AFLP analysis of African cassava (*Manihot esculenta*, Crantz) germplasm resistant to the cassava mosaic disease (CMD). *Theor Appl Genet.* 2000; 100: 678–685.
5. Balagopalan C. Cassava Utilization in Food and Feed Industry. In: Hillocks RJ, Thresh JM, Beelloti AC, Eds. *Biology, Production and Utilization.* CAB International, Cassava. 2001; 301–318. <https://doi.org/10.1006/dbio.2001.0162> PMID: 11237466
6. Sam J, Dapaah H. West Africa Agricultural Productivity Programme (WAAPP) Ghana, Baseline Survey Report. 2009.
7. Akinwale MG, Akinyele BO, Odiyi AC, Dixon AGO. Genotype x environment interaction and yield performance of 43 improved cassava (*Manihot esculenta* Crantz) genotypes at three agro-climatic zones in Nigeria. *Br. Biotechnol. J.* 2011; 1: 68–84.

8. Ampong MG, Afuakwah JS, Lamptey J, Adu-Meansah J, Ohemeng-Dapaah S, Asare-Bediako A, et al. Released of cassava varieties, a project sponsored by CSIR-CRI, IFAD/MOFA(RTIP) and IITA. A document presented to the Ghana national varietal release committee. 2004; 1–15. <https://doi.org/10.1148/radiol.2332031487> PMID: 15459322
9. Swaminathan MS. Gene banks for a warming planet. *Science*. 2009; 325: 517. <https://doi.org/10.1126/science.1177070> PMID: 19644077
10. Tanksley SD, McCouch SR. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*. 1997; 277: 1063–1066. <https://doi.org/10.1126/science.277.5329.1063> PMID: 9262467
11. Olukolu B. A., Mayes S., Stadler F., Ng N. Q., Fawole I., Dominique D., et al. Genetic diversity in Bambara groundnut (*Vigna subterranea* (L.) Verdc.) as revealed by phenotypic descriptors and DArT marker analysis. *Genet. Res. Crop Evol.* 2012; 59: 347–358. <https://doi.org/10.1007/s10722-011-9686-5>
12. Luo Z, Brock J, Dyer JM, Kutchan T, Schachtman D, Augustin M, et al. Genetic Diversity and Population Structure of a Camelina sativa Spring Panel. *Front. Plant Sci.* 2019; 10:184. <https://doi.org/10.3389/fpls.2019.00184> PMID: 30842785
13. Rabbi IY, Udoh LI, Wolfe M, Parkes EY, Gedil MA, Dixon A, et al. Genome-Wide association mapping of correlated traits in cassava: dry matter and total carotenoid content. *Plant Genome*. 2017; 10:3. <https://doi.org/10.3835/plantgenome2016.09.0094> PMID: 29293815
14. Boonchanawiat A, Sraphet S, Whankaew S, Boonseng O, Smith R. Mapping of quantitative trait loci underlying resistance to cassava anthracnose disease. *J Agric Sci.* 2016; 154:1209–1217. <https://doi.org/10.1017/S0021859615001057>
15. Wang B, Guo X, Zhao P, Ruan M, Yu X, Zou L, et al. Molecular diversity analysis, drought related marker-traits association mapping and discovery of excellent alleles for 100-day old plants by EST-SSRs in cassava germplasm (*Manihot esculenta* Crantz). *PLoS One*. 2017; 12:e0177456. <https://doi.org/10.1371/journal.pone.0177456> PMID: 28493955
16. Okogbenin E, Ekanayake IJ, Porto MCM. Genotypic variability in adaptation responses of selected clones of cassava to drought stress in the Sudan Savannah Zone of Nigeria. *J Agron Crop Sci.* 2003; 189: 376–389.
17. Peterson GW, Dong Y, Horbach C, Fu YB. Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. *Diversity*. 2014; 6: 665–680.
18. Adjebeng-Danquah J, Gracen VE, Offei SK, Asante IK, Manu-Aduening J. Agronomic performance and genotypic diversity for morphological traits among cassava genotypes in the Guinea Savannah Ecology of Ghana. *J. Crop Sci. Biotechnol.* 2016; 19: 99–108.
19. Agre AP, Dansi A, Rabbi IY, Battachargee R, Dansi M, Melaku G, et al. Agromorphological Characterization of Elite Cassava (*Manihot esculenta* Crantz) Cultivars Collected in Benin. *Int. J. Curr. Res. Bioscience and Plant Biol.* 2015; 2: 1–14.
20. Nadjiam D, Sarr PS, Naïtormaïdé M, Mbaïguinam JMM, Guisse A. Agro-Morphological Characterization of Cassava (*Manihot esculenta* Crantz) Cultivars from Chad. *J Agric Sci.* 2016; 7:479–492.
21. Pinton F, Emperaire L. Le manioc en Amazonie brésilienne: diversitévariétale et marché, *GENET SEL EVOL.* 2001; 33:491–512.
22. Benesi RM, Labuschagne MT, Herselman L, Mahungu NM, Saka JK. The effect of genotype, location and season on cassava starch extraction. *Euphytica*. 2008; 160: 59–74.
23. Noerwijati K, Nasrullah T, Djoko P. Breeding value estimation of fifteen related cassava genotypes using blup. *J Agric Biol Sci.* 2013; 4: 317–321.
24. Volk GM, Henk AD, Richards CM. Genetic diversity among U.S. Garlic clones as detected using AFLP methods. *J. Am. Soc. Hortic. Sci.* 2004; 129: 559–569.
25. Adu BG, Yeboah A, Akromah R, Bobobee E, Amoah S, Wireko-Kena A. Whole Genome SNPs and Phenotypic Characterization of Cassava (*Manihot esculenta* Crantz) Germplasm in the Semi-deciduous Forest Ecology of Ghana. *Ecological Genetics and Genomics*. 2020; 17: 100068. <https://doi.org/10.1016/j.egg.2020.100068>
26. Mondini L, Noorani A, Pagnotta MA. Assessing plant genetic diversity by molecular tools. *Diversity*. 2009; 1: 19–35.
27. Hayward AC, Tollenaere R, Dalton-Morgan J, Batley J. Molecular Marker Applications in Plants. In: Jacqueline Batley (ed.), *Plant Genotyping: Methods and Protocols*. *Methods in Molecular Biology*. 2015; 1245:13–27. https://doi.org/10.1007/978-1-4939-1966-6_2 PMID: 25373746
28. Govindaraj M, Vetriventhan M, Srinivasan M. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet. Res. Int.* 2015; 2015: 431487. <https://doi.org/10.1155/2015/431487> PMID: 25874132

29. Agarwal M, Shrivastava N, Padh H. Advances in molecular marker techniques and their applications in plant sciences. *Plant cell reports*. 2008; 27(4):617–631. <https://doi.org/10.1007/s00299-008-0507-z> PMID: 18246355
30. Egea LA, Mérida-García R, Kilian A, Hernandez P, Dorado G. Assessment of Genetic Diversity and Structure of Large Garlic (*Allium sativum*) Germplasm Bank by Diversity Arrays Technology “Genotyping-by-Sequencing”MPlatform (DArTseq). *Front Genet*. 2017; 8:98. <https://doi.org/10.3389/fgene.2017.00098> PMID: 28775737
31. Zhao WG, Chung JW, Lee GA, Ma KH, Kim HH, Kim KT, et al. Molecular genetic diversity and population structure of a selected core set in garlic and its relatives using novel SSR markers. *Plant Breed*. 2010; 130: 46–54. <https://doi.org/10.1111/j.1439-0523.2010.01805.x>
32. Hurtado P, Olsen KM, Buitrago C, Ospina C, Marin J, Duque M, et al. Comparison of simple sequence repeat (SSR) and diversity array technology (DArT) markers for assessing genetic diversity in cassava (*Manihot esculenta* Crantz). *PLANT GENET RESOUR-C*. 2008; 6:208–214.
33. Adjebeng-Danquah J, Manu-Aduening J, Asante IK, Agyare RY, Gracen V, Offei SK. Genetic diversity and population structure analysis of Ghanaian and exotic cassava accessions using simple sequence repeat (SSR) markers. *Heliyon*. 2020; 6: e03154. <https://doi.org/10.1016/j.heliyon.2019.e03154> PMID: 32042951
34. Ocampo C, Hershey C, Iglesias C, Inawaga M. Esterase enzyme fingerprinting of cassava germplasm held at CIAT. In: W. Roca, A.M. Thro (Eds.), *Proc. First Int. Sci. Meet in CIAT, Cali, Colombia*. Cassava Biotechnol. Netw. (CBN), 1992; 81–89.
35. Kawuki RS, Ferguson M, Labuschagne M, Herselman L, Kim D-J. Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). *Mol Breeding*. 2009; 23:669–684. <https://doi.org/10.1007/s11032-009-9264-0>
36. Jaccoud D, Peng KM, Feinstein D, Kilian A. Diversity Arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research*. 2001; 29(4):e25. <https://doi.org/10.1093/nar/29.4.e25> PMID: 11160945
37. Sanchez-Sevilla JF, Horvath A, Botella MA, Gaston A, Folta K, Kilian A, et al. Diversity Arrays Technology (DArT) Marker Platforms for Diversity Analysis and Linkage Mapping in a Complex Crop, the Octoploid Cultivated Strawberry (*Fragaria x ananassa*). *Plos One*. 2015; 10(12):22.
38. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H. Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods in Molecular Biology*. 2012; 888: 67–89. https://doi.org/10.1007/978-1-61779-870-2_5 PMID: 22665276
39. Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proceedings of the National Academy of Science USA*. 2004; 101: 9915–9920. <https://doi.org/10.1073/pnas.0401076101> PMID: 15192146
40. Heller-Uszynska K, Uszynski G, Huttner E, Evers M, Carlig J, Caig V. Diversity Arrays Technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane. *Molecular Breeding*. 2010; 28: 37–55.
41. Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, et al. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theoretical and Applied Genetics*. 2006; 113:1409–1420. <https://doi.org/10.1007/s00122-006-0365-4> PMID: 17033786
42. Tinker NA, Kilian A, Wight CP, Heller-Uszynska K, Wenzl P, Rines HW. New DArT markers for oat provide enhanced map coverage and global germplasm characterization. *BMC Genomics*. 2009; 10: 39. <https://doi.org/10.1186/1471-2164-10-39> PMID: 19159465
43. Schouten HJ, Weg WE, Carling J, Khan SA, McKay SJ, Kaauwen MPW. Diversity arrays technology (DArT) markers in apple for genetic linkage maps. *Molecular Breeding*. 2012; 29: 645–660. <https://doi.org/10.1007/s11032-011-9579-5> PMID: 22408382
44. Cruz VMV, Kilian A, Dierig DA. Development of DArT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop lesquerella and related species. *PLoS ONE*. 2013; 8: e64062. <https://doi.org/10.1371/journal.pone.0064062> PMID: 23724020
45. Raman H, Raman R, Kilian A, Detering F, Carling J, Coombes N. Genome-wide delineation of natural variation for pod shatter resistance in *Brassica napus*. *PLoS ONE*. 2014; 9:e101673. <https://doi.org/10.1371/journal.pone.0101673> PMID: 25006804
46. Xia L, Peng K, Yang S, Wenzl P, de Vicente MC, Fregene M, et al. DArT for high throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives. *Theor. Appl. Genet*. 2005; 110: 1092–1098. <https://doi.org/10.1007/s00122-005-1937-4> PMID: 15742202
47. Dorado G, Besnard G, Unver T, Hernández P. “Polymerase chain reaction (PCR),” in *Reference Module in Biomedical Sciences*, ed. Caplan M. (Amsterdam: Elsevier). 2015.

48. Alam M, Neal J, O'Connor K, Kilian A, Topp B. Ultra-high-throughput DArTseqbased silicoDArT and SNP markers for genomic studies in macadamia. *PLoS One*. 2018; 13(8): e0203465. <https://doi.org/10.1371/journal.pone.0203465> PMID: 30169500
49. Gower JC. A general coefficient of similarity and some of its properties. *Biometrics*. 1971; 27: 857–871. <https://doi.org/10.2307/2528823>
50. Suzuki R, Shimodaira H. pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. 2015. <http://stat.sys.i.kyoto-u.ac.jp/prog/pvclust/>
51. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. (2013). *cluster*: Cluster Analysis Basics and Extensions. R package version 1.14.4
52. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, Ohara RB, et al. *Vegan*: Community Ecology Package. R package Version 2.4–3. 2013. <https://CRAN.R-project.org/package=vegan>
53. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. 2016; ISBN 978-3-319-24277-4. <https://ggplot2.tidyverse.org>.
54. Pritchard J, Wen X, Falush D. Documentation for structure software: Version 2.3. University of Chicago, Chicago, IL. 2010.
55. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*. 2005; 14: 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x> PMID: 15969739
56. Earl DA, Vonholdt BM. Structure harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conserv Genet Resour*. 2012; 4: 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
57. Hammer Ø, Harper ATD, Ryan PD. “Past: Paleontological Statistics Software Package for Education and Data Analysis”. *Palaeontologia Electronica*. 2001; 4 (1): 4–9.
58. Hampton JO, Spencer PBS, Alpers DL, Twigg LE, Woolnough AP, Doust J, et al. Molecular techniques, wildlife management and the importance of genetic population structure and dispersal: a case study with feral pigs. *Journal of Applied Ecology*. 2004; 41: 735–743.
59. Asare PA, Galyuon IKA, Sarfo JK, Tetteh JP. Morphological and molecular based diversity studies of some cassava (*Manihot esculenta* crantz) germplasm in Ghana. *Afr J. Biotechnol*. 2011; 10: 13900–13908.
60. Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, et al. DArT markers: diversity analyses and mapping in Sorghum bicolor. *BMC Genomics*. 2008; 9(1):1. <https://doi.org/10.1186/1471-2164-9-26> PMID: 18208620
61. Simko I, Eujayl I, van Hintum TJ. Empirical evaluation of DArT, SNP, and SSR marker-systems for genotyping, clustering, and assigning sugar beet hybrid varieties into populations. *Plant science*. 2012; 184:54–62. <https://doi.org/10.1016/j.plantsci.2011.12.009> PMID: 22284710
62. Baloch FS, Alsaleh A, Shahid MQ, Ciftci V, Sa'enz de Miera EL, Aasim M, et al. A Whole Genome DArTseq and SNP Analysis for Genetic Diversity Assessment in Durum Wheat from Central Fertile Crescent. *PLoS ONE*. 2017; 12(1):e0167821. <https://doi.org/10.1371/journal.pone.0167821> PMID: 28099442
63. Dierig D, Ray DT. New crops breeding: Lesquerella. In: Vollmann J, Rajcan I, editors. *Oil Crops*. Springer. 2009: 507–516. <https://doi.org/10.1354/vp.46-1-34> PMID: 19112112
64. Asante IK, Offei SK. RAPD-based genetic diversity study of fifty cassava (*Manihot esculenta*, Crantz) genotypes. *Euphytica*. 2003; 131:113–119.
65. Kilian A, Huttner E, Wenzl P, Jaccoud D, Carling J, Caig V, et al. The Fast and the Cheap: SNP and DArT-Based Whole Genome Profiling for Crop Improvement. *Proceedings of the international congress in the wake of the double helix: from the green revolution to the gene revolution*; 2003:443–461.
66. Mantello C. C, Cardoso-Silva CB, da Silva CC, de Souza LM, Scaloppi Junior EJ, de Souza Goncalves P., et al. De novo assembly and transcriptome analysis of the rubber tree (*Hevea brasiliensis*) and SNP markers development for rubber biosynthesis pathways. *PLoS One*. 2014; 9:e102665. <https://doi.org/10.1371/journal.pone.0102665> PMID: 25048025
67. Huang SM, Deng LB, Guan M, Li J, Lu K, Wang H Z, et al. Identification of genome-wide single nucleotide polymorphisms in allopolyploid crop Brassica napus. *BMC Genomics*. 2013; 14:717. <https://doi.org/10.1186/1471-2164-14-717> PMID: 24138473
68. Ferguson ME, Shah T, Kulakow P, Ceballos H. A global overview of cassava genetic diversity. *PLoS ONE*. 2019; 14(11): e0224763. <https://doi.org/10.1371/journal.pone.0224763> PMID: 31693686
69. Ministry of Food and Agriculture (MOFA). Catalogue of Crop Varieties Released and Registered in Ghana. 2019; pp. 58–63.

70. Kyamanywa S, Kashaija I, Getu E, Amata R, Senkeshu N, Kullaya A. Enhancing food security through improved seed systems of appropriate varieties of cassava, potato and sweetpotato resilient to climate change in Eastern Africa. Nairobi, Kenya: ILRI. 2011: 1–28.
71. Rabbi IY, Kulakow AP, Manu-Aduening JA, Dankyi AA, Asibuo JY, Parkes EY, et al. Tracking crop varieties using genotyping by-sequencing markers: a case study using cassava (*Manihot esculenta* Crantz). *BMC Genetics*. 2015; 16:115. <https://doi.org/10.1186/s12863-015-0273-1> PMID: 26395668
72. Sadiki M, Jarvis D, Rijal DK, Bajracharya J, Hue NN, CamachoVilla C, et al. Variety names an entry point to crop genetic diversity and distribution in agroecosystems in Agricultural Ecosystems. Columbia University Press, New York. 2006. <https://doi.org/10.1021/la061160c> PMID: 17014119
73. Agre AP, Bhattacharjee R, Rabbi IY, Alaba OY, Unachukwu NN, Ayanan MAT, et al. Classification of elite cassava varieties (*Manihot esculenta* Crantz) cultivated in Benin Republic using farmers' knowledge, morphological traits and simple sequence repeat (SSR) markers. *Genetic Resources Crop Evolution*. 2017; 64:307–320. <https://doi.org/10.1007/s10722-017-0550-0>
74. Duminil J, Di Michele M. Plant species delimitation: a comparison of morphological and molecular markers. *Plant Biosystems*. 2009; 143(3): 528–542.
75. Eltaher S, Sallam A, Belamkar V, Emara HA, Nower AA, Salem KFM, et al. Genetic diversity and population structure of F3:6 nebraska winter wheat genotypes using genotyping-by-sequencing. *Front Genet*. 2018; 9:76. <https://doi.org/10.3389/fgene.2018.00076> PMID: 29593779
76. Kizito EB, Chiwona-Kartun L, Egwang T, Fregene M, Westerbergh A. Genetic diversity and variety composition of cassava on small-scale farms in Uganda: an interdisciplinary study using genetic markers and farmer interviews. *Genetics*. 2007; 130: 301–318.
77. Ortiz A, Rocha V, Moiana L, Gonçalves-Vidigal M, Galvan M, Vidigal-Filho P. Population structure and genetic diversity in sweet cassava cultivars from Parana Brazil. *Plant Molecular Biology Report*. 2016; 34: 1153–1166.
78. Ortiz AHT, Filho PSV, Rocha VPC, Ferreira RCU, Gonçalves TM, Gonçalves-Vidigal MC. Population Structure and Genetic Diversity of Sweet Cassava Accessions from the Midwestern, Southeastern and Southern Regions of Brazil. *Brazilian Archives of Biology and Technology*. 2019; 62: e19180556. <http://dx.doi.org/10.1590/1678-4324-2019180556>