

Full Length Research Paper

Genetic diversity and relatedness of oil palm progenies determined by microsatellite and agronomic markers

Maxwell N. Okoye^{1,3*}, Claude Bakoumé², Michael I. Uguru³ and Rajinder Singh⁴

¹Plant Breeding Division, Nigerian Institute for Oil Palm Research (NIFOR), P.M.B. 1030 300001, Benin City, Edo State, Nigeria.

²Maxi Productivity Sarl, P. O. Box 2137 (or 24240), Douala, Cameroon.

³Department of Crop Science, Faculty of Agriculture, University of Nigeria, Nsukka, Enugu State, Nigeria.

⁴Advanced Biotechnology and Breeding Centre, Malaysian Palm Oil Board (MPOB), P.O. Box 10620, 50720 Kuala Lumpur, Malaysia.

Received 24 July, 2017; Accepted 28 September, 2017

Genetic diversity in oil palm has been usually determined using agro-morphological markers (traits) before the advent of DNA-based markers including microsatellite markers, which have the advantage of being environmentally independent. The current study aimed to compare genetic diversity and relationships among 10 Nigerian Institute for Oil Palm Research (NIFOR) *dura* × *tenera* oil palm progenies using simple sequence repeats (SSR) and agronomic markers. 114 individual palms representing the NIFOR progenies were screened for polymorphism at five agronomic traits and 16 SSR loci. Coefficients of variation and genetic diversity parameters were calculated to compare the trait-based variation with genetic (SSR) diversity. Hierarchical clustering and principal coordinate analyses (PCA) were performed with agronomic and molecular datasets generated from the progenies. Agronomic traits showed wide range of variation from 5.6% for oil-to-mesocarp ratio to 40.5% for bunch number. The SSR markers deployed showed 100% polymorphism and high genetic diversity ($H_e = 0.661$, $H_o = 0.580$) among the progenies. While SSR data discriminated the progenies with respect to pedigree or shared ancestry of the parents, the clustering pattern based on agronomic data predominantly reflected the differences in agronomic traits. Results of this study suggest that agronomic trait data are insufficient in selecting parents for crossing and that genotypic data are more informative.

Key words: Agronomic traits, *Elaeis guineensis*, genetic diversity, Nigerian Institute for Oil Palm Research (NIFOR), simple sequence repeats (SSR) markers.

INTRODUCTION

The African oil palm (*Elaeis guineensis* Jacq.) is the most productive oil bearing crop yielding more than five times oil per hectare of any annual oil crop. In West Africa in general, and particularly in Nigeria, palm oil is the most

valuable natural oil in the local diets both as crude red palm oil and as refined palm oil; olein (Corley and Tinker, 2003). The oil palm industry is a major source of employment and income to a substantial proportion of

*Corresponding author. E-mail: maxwellokoye@gmail.com.

Nigerian populace (Omoti, 2009). The world palm oil production in 2016 was 58.72 million tons, with a projected yield increase of up to 6.93%. Nigeria's stagnated 970,000 metric tons of annual palm oil production for the past 6 years from an estimated 2.6 million hectares of oil palm production area, ranks first in the African oil palm belt and fifth in the world after Indonesia, Malaysia, Thailand and Colombia (United States Department of Agriculture (USDA) Statistics, 2017). Notwithstanding its fifth position, the country's average yield is low compared to those of South East Asian palm oil producing countries. In addition, the demand for vegetable oil in Nigeria outstrips domestic supply leading to a deficit of about 500,000 metric tons. Continuous yield improvement to cope with the ever increasing demand for palm oil and its products is currently on-going at the Nigerian Institute for Oil Palm Research (NIFOR), and substantial progress has already been made in this direction. Presently, the commercial variety (*tenera*) yields 20 to 25 mt of fresh fruit bunch ha^{-1} year $^{-1}$ and 3 to 3.5 tons oil ha^{-1} year $^{-1}$ in mature plantations (Okwuagwu et al., 2005). Therefore, replanting old areas with improved varieties in conjunction with proper management and cultural practices would be more than enough in boosting oil palm productivity in Nigeria.

Germplasm collections of oil palm and related richness in new genes are valuable sources of traits of agronomic importance for subsequent development of new improved varieties. The on-going oil palm Main Breeding Programme carried out by NIFOR is presently testing several oil palm genotypes for high yield and adaptation to different agro-ecological areas.

Genetic diversity in crop species can be determined indirectly by phenotypic markers (agro-morphological and biochemical markers) and genotypic (DNA-based) markers (Mohammadi and Prasanna, 2003). Several of these agro-morphological trait based studies have assessed genetic diversity in oil palm populations (Kushairi et al., 1999; Ataga et al., 2005; Okwuagwu et al., 2008, 2011). However, this approach is limited by the long juvenile phase, confounding effects of developmental stage of the plant, long term field evaluation, and vulnerability to environmental effects. Currently, a range of molecular marker techniques are available for measuring genetic diversity. They assess genetic variation at the DNA level and are renowned to be less environment-dependent. The most routinely used are restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and microsatellites or simple sequence repeats (SSRs) (Purba et al., 2000; Rajanaidu et al., 2000; Maizura et al., 2006; Bakoumé et al., 2015).

Microsatellites or SSRs markers provide an important source of polymorphism because they are neutral and widely distributed throughout species genome, thus suitable in assessing genetic diversity within populations.

Though SSRs are considered selectively neutral, they have selective value or are linked to genes of selective value. Therefore, it has been assumed here that the fraction of genetic diversity detected by the SSRs can be reflected by the variation revealed by related agronomic traits if any. Several works on genetic diversity, progeny legitimacy test, ortet-ramets relationship, and genetic mapping using microsatellite markers, have been reported in oil palm (Billotte et al., 2001, 2005; Bakoumé et al., 2007, 2011; 2015; Singh et al., 2008; Cochard et al., 2009; Arias et al., 2012; 2013; Okoye et al., 2016). This work aimed at establishing if the amount of genetic variation and structure of genetic variation detected in oil palm parental materials and progenies of the NIFOR Main Breeding Programme by SSR markers agree with those revealed by highly heritable agronomic traits of interest for oil palm breeding and for oil palm industry.

MATERIALS AND METHODS

Plant material

Ten *dura* × *tenera* (D × T) oil palm progenies of the NIFOR second cycle Main Breeding Programme were evaluated in this study. The progenies were derived from 11 bi-parental crosses comprising four NIFOR Deli *dura* (DD), two *dura* (AD), and five *tenera* (T) parents collected from Nigerian oil palm groves and maintained at NIFOR. A total of 114 oil palms that is, 8 to 17 palms per progeny whose agronomic traits were recorded were used for SSR analysis (Table 1).

Agronomic data collection

The highly heritable agronomic traits recorded per individual palm basis included: (i) number of harvested bunches (BN), (ii) average bunch weight (ABW), (iii) mesocarp to fruit ratio (M/F), (iv) oil to mesocarp ratio (O/M), and (v) palm height. Average bunch weight (ABW) was obtained as the ratio of fresh fruit bunch (FFB) yield to BN. Mesocarp to fruit ratio and oil to mesocarp ratio was evaluated through fresh fruit bunch analyses according to standard method outlined by Blaak et al. (1963). Seven years (1999 to 2005) mature yield data from the NIFOR oil palm Main Breeding Programme were used for this study. Palm height was measured from the ground level to the base of the leaf or leaf subtending the last harvested bunch for mature palms.

DNA extraction and quantification

DNA was isolated from the 114 oil palm samples at the Bioscience Centre, International Institute of Tropical Agriculture (IITA) Ibadan, using CTAB DNA isolation protocol of Doyle and Doyle (1990), with minor modifications. DNA quantification was determined by electrophoresis and NANODROP® (ND-1000) Spectrophotometer (Thermo Fisher Scientific Inc., Denver). All DNA samples were stored at -20°C until microsatellite analysis at the Genomics Unit of Advanced Biotechnology and Breeding Centre (ABBC), Malaysian Palm Oil Board (MPOB) Selangor, Malaysia. The DNA samples were diluted to an optimum concentration of 25 ng/μl by addition of sterile distilled water or appropriate amount of TE (Tris-EDTA) buffer and stored at 4°C until polymerase chain reaction (PCR) amplification.

Table 1. Progenies derived from 11 NIFOR oil palm breeding parent trees.

S/N	Cross	Progeny code	Number of sample	Dura parent/pedigree	Tenera parent/pedigree
1	DD1 × T2	DT1	12	NIFOR (ex Serdang Avenue Malaysia)	Calabar
2	DD1 × T6	DT2	17	NIFOR (ex Serdang Avenue Malaysia)	Umuabi OP
3	T1 × DD2	DT3	14	NIFOR (ex Serdang Ave. X IRHO – Pobe)	Ufuma (ex Aba)
4	T3 × DD3	DT4	10	Ulu Remis Deli x ex Sabah	Aba (ex Calabar)
5	DD3 × T5	DT5	9	Ulu Remis Deli x ex Sabah	Ufuma
6	DD4 × T2	DT6	12	Ecuador Deli	Calabar
7	DD4 × T6	DT7	13	Ecuador Deli	Umuabi OP
8	AD1 × T6	DT8	9	Calabar	Umuabi OP
9	AD3 × T6	DT9	8	Ufuma	Umuabi OP
10	AD3 × T5	DT10	10	Ufuma	Ufuma
Total			114		

S/N, Serial number.

Microsatellite analysis

A total of 16 microsatellite markers comprising nine markers developed at the Genomics Unit of ABBC-MPOB and seven from the French Centre de Coopération Recherche Agronomique pour le Développement (CIRAD) were used for the PCR amplification. MPOB SSRs were developed from the oil palm expressed sequence tags (ESTs) and genomic sequences reported by Singh et al. (2008) and Ting et al. (2010). SSRs developed from *E. guineensis* ESTs, *E. guineensis* genomic sequences, and *E. oleifera* genomic sequences were labeled sEg, sMg, and sMo, respectively. SSR primer sequences from CIRAD were downloaded from the Trop GENE database (<http://tropgenedb.cirad.fr/html/oilpalm/Marker.html>) and originally labeled as mEgCIR (Billotte et al., 2005). Details of the 16 SSR markers are presented in Table 2. The amplifications of microsatellite loci with fluorescently labeled polymerase chain reaction (PCR) primers were performed as described by Ting et al. (2010) in a Perkin Elmer 9700 thermocycler (Life Technologies, Thermo-Fisher Scientific, USA). PCR products were analyzed on an ABI 3730 Genetic Analyzer and sized using GeneMapper® 4.1 (Applied Biosystems Inc., Foster City, CA, USA) and GS LIZ 500 size standard. Electropherogram profiles (sample plots) were generated and genotype data for all the SSR markers was exported as data table for scoring. The scoring of the genotype data was performed manually with reference to allele and peak size.

Data analyses

For the agronomic data, progeny means as well as population mean, range, and standard error of mean were determined using GenStat software. Coefficients of variation were calculated within progeny for each of the 5 agronomic traits. Clustering of progenies was performed using an average linkage algorithm (UPGMA, Unweighted Pair Group Method with Arithmetic mean) based on Manhattan dissimilarity coefficients (MD) (Sokal and Michener, 1958). Principal coordinate analysis (PCoA) (Gower, 1966) was performed to better depict the relatedness among progenies or individual genotypes. The pair-wise comparisons of progenies and PCoA were facilitated by DARwin 6.0.4 programme (Perrier and Jacquemoud-Collet, 2006). For the microsatellite data, the genotype data of the 10 D × T progenies at all SSR loci were used to assess the number of alleles (A_o), percentage of polymorphic alleles (%P), observed (H_o) and expected (H_e) heterozygosities using Genetic Analysis in Excel (GenAlEx) version 6.5 (Peakall and

Smouse, 2006, 2012). Rogers' dissimilarity coefficients (Rogers, 1972) were calculated and the dissimilarity coefficient matrices were again subjected to cluster analysis and PCoA to explore and establish similarity or dissimilarity among D × T progenies. All genetic distance calculations and construction of dendrograms were performed using PowerMarker v3.25 (Liu and Muse, 2005) and MEGA software v4.0 (Tamura et al., 2007), respectively.

RESULTS

Variability in progenies of NIFOR parents of the main breeding programme revealed by SSR markers and agronomic traits

The genetic diversity parameters in progenies of NIFOR parents of the Main Breeding Programme using SSR markers and agronomic markers are given in Table 3. Among the five traits evaluated, wider range of variation was observed across the ten different oil palm progenies for most of the traits. The number of harvested bunch (BN) was in the range of two to 14. The minimum (3 kg/p/yr) and maximum (26.3 kg/p/yr) average bunch weight (ABW) was observed in DT5 and DT10 progenies, respectively. The annual growth rate (height) was maximum (90 cm/yr) in DT10 and minimum (31.9 cm/yr) in DT7. Mesocarp to fruit ratio varied from 40 in DT2 to 92.5 in DT9 while oil to mesocarp ratio (O/M) was as small as 45.1 in DT10 and as large as 65.8 in DT2, with a mean of 57.23. The coefficients of variation (CV) for the progenies with respect to the agronomic traits revealed wide range of values. BN was found to be the most variable character for all the progenies evaluated. The CV% ranged from 27.76% in DT1 to 59.29% in DT5. This is followed by height (8.23% in DT8 to 31.94% in DT10) and ABW (22.64% in DT9 to 39.47% in DT7), respectively. The CV within the progenies was moderate or low for M/F (16.63% in DT8 to 26.17% in DT2) and O/M (3.339% in DT1 to 7.231% in DT4).

Based on SSR markers derived from oil palm, the level polymorphism was 100% in all the D × T progenies

Table 2. Microsatellite primer pairs used for population genetic analysis of oil palm

S/N	SSR LOCI	Linkage group	Ta (°C)	Primer sequence	SSR repeat motif	Expected fragment size (BP)	Accession number
1	mEgCIR3813	1	52	F-CATACCCTGCTTATCTTTC R- GTAGATACCCGTTAGTTGAC	(GA)19	167	AJ578734
2	mEgCIR0793	2	56	F-GTACTTCGCAACTATTCCTTTTCTT R- AGTTGATCGTGGTGCCTGAC	(GA)15	149	AJ578545
3	mEgCIR0425	3	58	F-AGCAAAGAGCAAGAGCAGAACT R- CTTGGGGGCTTCGCTATC	(CCG)9	232	AJ578521
4	sMg00156	4	50	F-GGTGTCATAACTTCGTTGTTGCT R- ATGCTCAAAAGTGGGTTTCTCTC	(CT)15	237	Pr010615888*
5	mEgCIR3828	5	50	F-AGCCAGATGGAATACAC R- GTGCGATAAAGAGGAGAGT	(GA)23	282	AJ578738
6	sEg00154	6	57	F-TCCCCAATACTCATCATGC R- TGATCGACGGTTGTACATT	(CAG)5	238	EY410356**
7	sMo00102	7	53	F-ATGAGATGGGACAAATCAAAC R- ACCATACCACTAGAGAACTAAACA	(AG)11	235	Pr010615939*
8	sMg00228	8	54	F-CACGTATATGAGCAGGATTTGA R- CTCCAAACCACTAGAGCTGA	(AT)25	205	Pr010615913*
9	sMg00016	9	52	F-GCGATTCCGGTTATCTTTAG R- GAGTTTTTGTGTGATGATTAG	(GA)13	274	Pr010615861*
10	mEgCIR3519	10	52	F-CCACTGCTTCAAATTTACTAG R- GCGTCCAAACATAAATCAC	(GA)15(GT)8	236	AJ578672
11	sMg00120	11	54	F-GATCAATGCGAGAAATCAGG R- GATCATGCTTATCCTTTCCAAGT	(AT)11	152	Pr010615881*
12	mEgCIR0790	12	52	F-TTGGTGGTCCTTTTGAATATC R- ACAAACCCAGCACTTAAAATAAC	(GA)19	215	AJ578544
13	sEg00151	13	57	F-ATCACAACAGCAGCAGCATC R- CGCATCAAGAAACATGGAGA	(CAG)8	219	EY411661**
14	sMg00179	14	54	F-AACCCCTTTTTTCATGCTCTAA R- CTGATTTTGAATCAGAGGTG	(AAAAG)6	214	Pr010615893*
15	sMg00087	15	58	F-CACCTAAAAACGGCAAGGAAC R- GGAGGAGAGAAATGGAAGACG	(AG)19AA(AG)	212	Pr010615880*
16	mEgCIR3745	16	52	F-GGAAGTCTTGATGTTGAAAG R- ATCAAGCAGTCGCATAATAC	(GA)18	260	AJ578718

*Probe Unique Identifiers (PUIDs) of NCBI Probe Database. **Accession numbers of NCBI GenBank

evaluated. The number of alleles per progeny (A_o) varied from 3.2 (DT9) to 4.1 (DT10), with a population mean value of 3.770. Observed (H_o) and expected (H_e) heterozygosities were greater than 0.5 in all but one

progeny, DT3 ($H_o=0.448$) and DT4 ($H_e=0.488$), respectively with an average H_o of 0.580 and H_e of 0.661. Progeny DT1 recorded the highest H_o (0.662) and H_e (0.614).

Table 3. Diversity in progenies of NIFOR parents of the main breeding programme using SSR markers and agronomic markers.

Progeny	SSR markers				Agronomic markers									
	Å	%P	H _o	H _e	BN		ABW		M/F		O/M		Height	
					Mean	CV (%)	Mean (kg)	CV (%)	Mean (%)	CV (%)	Mean (%)	CV (%)	Mean (cm/yr)	CV (%)
DT1	4.000	100	0.662	0.614	5.909	27.76	11.582	35.70	63.12	23.14	57.38	3.339	61.46	13.59
DT2	3.900	100	0.622	0.537	6.059	35.24	9.495	31.04	63.39	26.17	57.98	5.764	51.62	19.61
DT3	3.500	100	0.448	0.517	5.929	44.61	12.82	29.50	68.36	22.97	56.54	5.280	66.62	11.16
DT4	3.500	100	0.532	0.488	4.1	42.17	14.98	33.20	61.37	22.87	57.00	7.231	77.19	12.75
DT5	3.900	100	0.510	0.561	5.333	59.29	16.36	31.70	67.88	19.78	58.03	7.143	70.72	15.96
DT6	3.800	100	0.580	0.568	5.417	29.93	11.28	33.09	60.27	24.04	57.20	6.515	47.05	16.55
DT7	4.400	100	0.642	0.576	6.154	34.37	8.3	39.47	68.9	22.73	56.90	5.429	44.47	17.03
DT8	3.400	100	0.520	0.526	7.444	46.59	10.27	27.44	68.51	16.63	57.03	4.654	50.26	8.23
DT9	3.200	100	0.538	0.512	4	32.73	16.62	22.64	70.81	23.71	60.84	4.448	61.03	13.26
DT10	4.100	100	0.607	0.573	5	52.49	11.27	38.56	70.0	24.70	53.99	5.764	59.79	31.94
Total Popul.														
Mean	7.6	100	0.580	0.661	5.611	40.52	11.917	32.23	65.953	22.67	57.23	5.56	58.22	16.01
Minimum						2		3	40		45.1		31.92	
Maximum						14		26.3	92.5		65.8		90.12	
SE	0.476		0.041	0.034	0.228		0.427		1.409		0.341		0.156	

Å=Number of allele; H_o=Observed heterozygosity; H_e=Expected heterozygosity; %P=Percentage of polymorphic loci; BN=Number of harvested bunches; ABW=Average bunch weight; M/F=Mesocarp to fruit ratio; O/M=Oil to mesocarp ratio; CV=Coefficient of variation; SE=Standard error of mean. Popul: Population.

Relatedness of progenies of NIFOR parents of the main breeding programme revealed by agronomic markers and SSR markers

Using the mean values of the 5 agronomic markers (number of bunches (BN), average bunch weight (ABW) (kg/p/year), mesocarp to fruit ratio (M/F), oil to mesocarp ratio (O/M), and height (cm/yr)), Manhattan dissimilarity coefficients (MD) were calculated by pair-wise comparisons of progenies using DARwin software. Manhattan dissimilarity coefficients ranged from 0.145 to 0.617 with an average of 0.4. The UPGMA based dendrogram grouped the progenies in two major clusters (Figure 1). The progeny DT9 was clearly differentiated from all other progenies in Group I. Pedigree analysis of DT9 revealed that T6 and AD3, the Umuabi *tenera* and Ufuma *dura* parents are involved in the development of this progeny. DT9 is characterized by high ABW, M/F, and O/M that separated it from the rest of the progenies. Group II was constituted by progenies classified either by different agronomic traits or pedigree. Five additional sub-clusters varying from 1 to 3 progenies per sub-cluster were observed within this group. Sub-clusters IIA and IIB comprised only one progeny DT4 and DT10, respectively. These progenies are unique for palms with high increment in height (tall palms) and high M/F, respectively. Progeny DT3 and DT5 fell in sub-cluster IIC due to the common origin of the male parents (T1 and T5); both parents are from Ufuma. Similarly in sub-cluster IID, DT7 and DT8 share the same male parent (T6) from

Umuabi. Finally, sub-cluster IIE included progenies sharing same female parent (DD1) and male parent (T2).

Clustering pattern of 10 D x T progenies based on the SSR markers (Figure 2) was not entirely different from the patterns obtained from agronomic markers. The UPGMA dendrogram based on the Rogers' dissimilarity among the progenies showed two main clusters. Out of 10 progenies, 9 were grouped into a single cluster with four sub-clusters. The first cluster contained progeny DT9 derived from Ufuma *dura* (AD3) and Umuabi *tenera* (T6) parents. The second main cluster was of four sub-clusters IIA, IIB, IIC and IID. Sub-clusters IIC and IID contained progenies which either share one of the parents directly or through the ancestry. For instance, DT5, DT10, DT8, and DT7 share the same male parent (T6) from Umuabi while DT5 share the same origin with DT10. Both the male parent of DT5 (T5) and the female parent of DT10 (AD3) are from the same origin/geographical location, Ufuma. Also, DT1 and DT2 in sub-cluster IID share the same female parent (DD1); a NIFOR Deli from Serdang Avenue, Malaysia. The inclusion of DT4 in this sub-cluster was probably because of the common Calabar origin of the male parent (T3) and T2 in DT1. DT6 and DT3 were distinct sub-clusters IIA and IIB respectively.

Principal coordinate analysis

To visualize the similarity or dissimilarity among

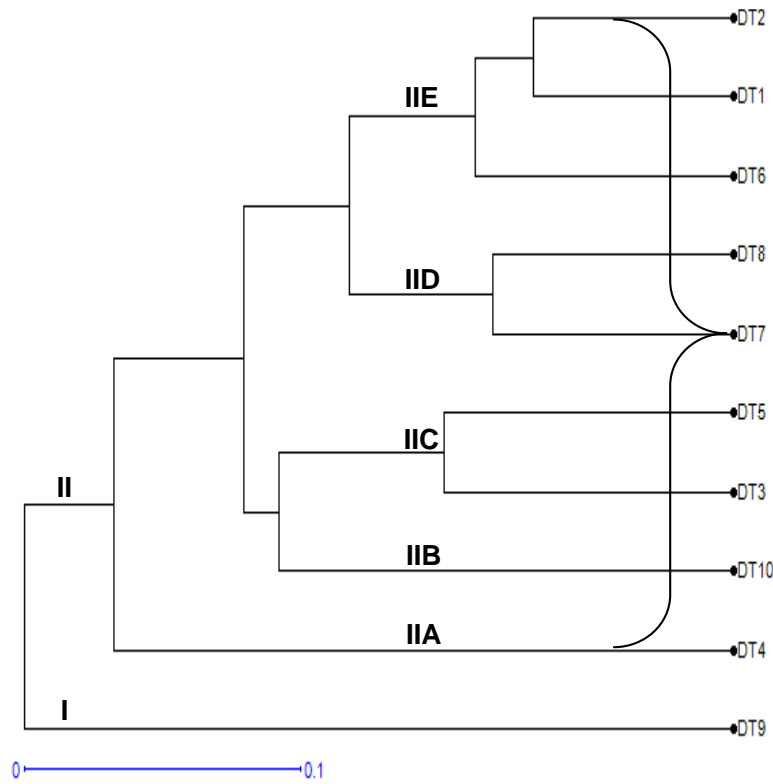


Figure 1. Dendrogram based on Manhattan dissimilarity coefficients demonstrating association among 10 D x T NIFOR oil palm progenies.

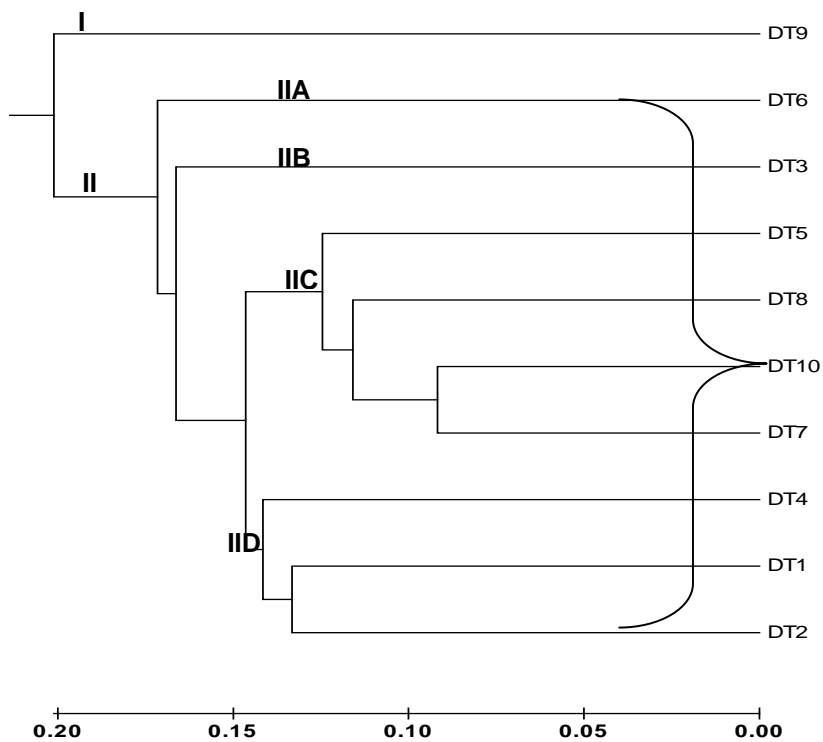


Figure 2. Dendrogram based on UPGMA clustering of 10 D x T NIFOR oil palm progenies based on Rogers' dissimilarity.

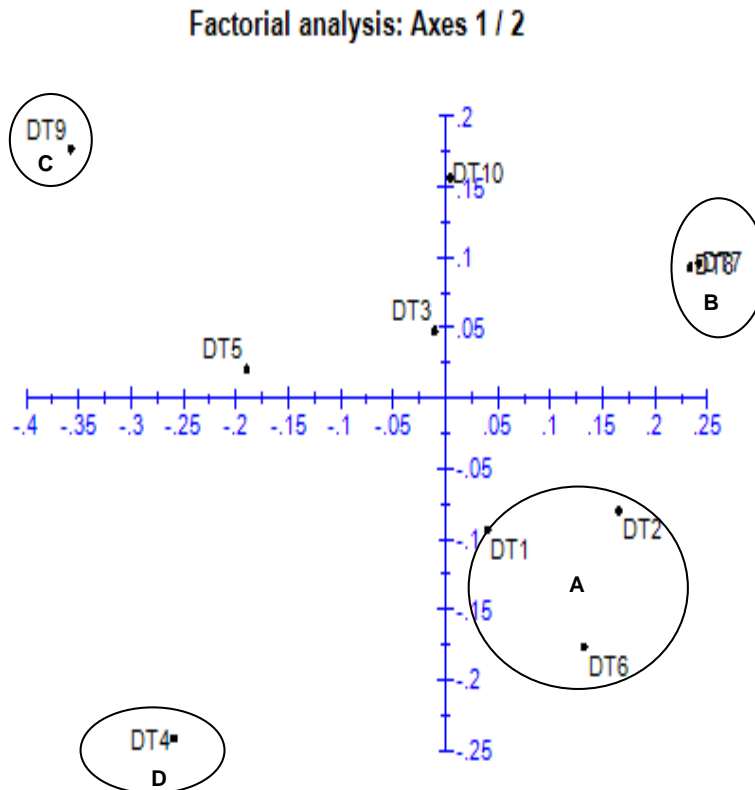


Figure 3. Principal Coordinates Analysis using genetic distance matrix based on 5 agronomic traits.

progenies or individual genotypes, principal coordinate analysis (PCoA) was performed using DARwin software version 6.0.4 programme (Figures 3 and 4). The PCoA analysis further validated the positions and grouping of progenies. PCoA based on genetic distance matrix of agronomic data (Figure 3) explained 74.17% of the diversity on the first two principal components (PC1 = 50.66% and PC2 = 23.51%). There was, however, a tendency that the progenies clustered according to their agronomic performance or shared ancestry. The derivatives of the Deli *dura* (DT1, DT2 and DT6) were distributed in one cluster (A) based on high BN and average O/M while DT7 and DT8 were grouped together in cluster B with respect to high BN and M/F, average O/M and plant height. Noteworthy is the shared pedigree of either a common male or female parent among the progenies in clusters A and B. Progenies DT9 and DT4 were separated from the rest of the progenies in clusters C and D. DT9 exhibited the highest ABW, M/F and O/M compared to the very tall palms in progeny DT4. The rest of the progenies were scattered as individuals or smaller groups on the basis of intermediate agronomic traits very much similar to the dendrogram.

In comparison to the grouping based on agronomic data, the grouping of progenies derived from SSR marker data is observed to be more incisive and compelling

(Figure 4). The only progeny (DT9) derived from Ufuma *dura* (AD3) and Umuabi *tenera* (T6) was solely placed in the first cluster (A) while the progenies (DT1, DT3, and DT6) were grouped in cluster B. DT1 and DT3 share the same female Serdang Avenue Deli grandparent (19 x 65) and DT6 share the same male parent (T2) with DT1. Cluster C contain progenies with common male parent (T6) as well as some of the progenies (DT8 and DT10) in cluster D. DT4 and DT5 sharing same female parent (DD3) were also grouped together with DT8 and DT10 in cluster D. These two sets of progenies were classified together due to the common Ufuma origin of the male parent (T5) in DT5 and female parent (AD3) in DT10. The PCoA based on molecular data is better in discriminating related progenies of common origin and parentage.

Simple correlation between phenotypic variation, estimated by Manhattan distances using all agronomic characters and SSR marker based distance matrices was low ($r = 0.2989$) and non-significant.

DISCUSSION

It is important to advance NIFOR oil palm breeding to meet the growing domestic demand for palm oil and its products. In view of the successes of the conventional

Factorial analysis: Axes 1 / 2

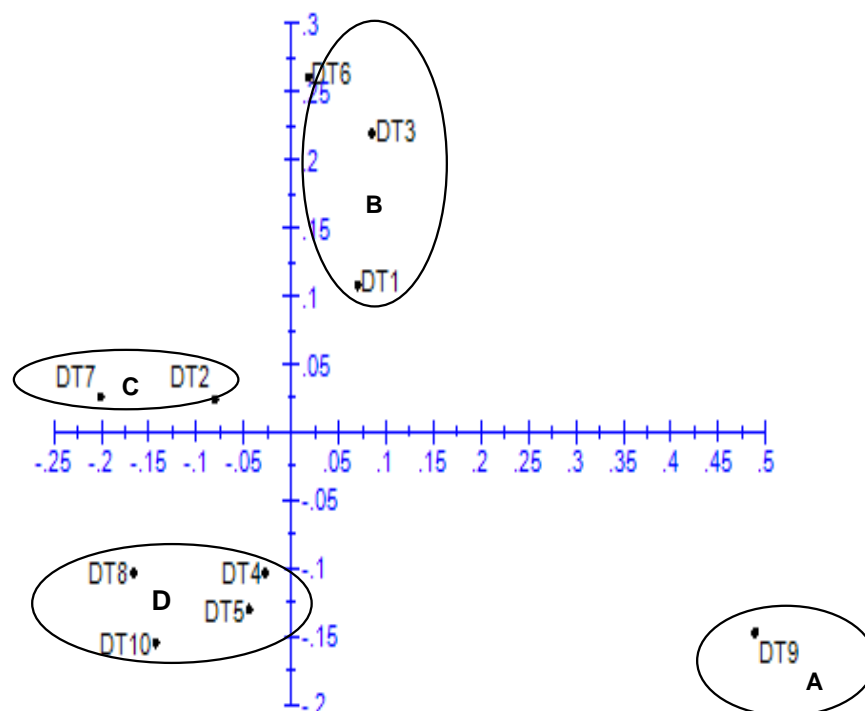


Figure 4. Principal Coordinates Analysis using SSR marker based similarity coefficient matrix of 10 NIFOR oil palm D x T progenies.

breeding methodology over the years, oil palm hybrid breeding remains a choice method. Better understanding on the genetic diversity provides possibilities for breeders to select desired individuals for the plant improvement, to produce more performing progenies that assemble the good parental characteristics. In the present study, oil palm progenies screened with SSR markers have been subjected to analysis of variations for agronomic traits.

Agronomic and microsatellite variability in the 10 D x T progenies

The results of this investigation showed that the 10 NIFOR D x T oil palm progenies were variable for most of the evaluated agronomic traits and microsatellite markers. Among the progenies, substantial variation was observed for number of harvested bunch (2 to 14), average bunch weight (3 to 26.3 kg⁻¹/year⁻¹), and annual growth rate (height = 31.92 to 90.12 cm/yr). Similar results on wide range of phenotypic variability of agronomic traits have been reported in the evaluation of introgressed progenies of Nigerian origin in MPOB by Noh et al. (2014). The coefficient of variation (CV) values of these traits varied from 27.76 to 59.29% between DT1

and DT5, 22.64 to 39.47% for DT9 and DT7, and 8.23 to 31.94% in favour of DT8 and DT10, respectively. The high coefficient of variation observed among the progenies for the respective traits evaluated indicates high genetic diversity existing in the NIFOR oil palm main breeding population. Differences in CV among the progenies could be explained by genotype, environment, or genotype-environment interaction. The presence and importance of genotype-environment interaction in the NIFOR oil palm progenies has been mentioned by several authors (Ataga 1993; Okoye et al., 2008). The broad range in the population mean of the various traits implies great potential for grouping the oil palm progenies into various groups of poor performers and good performers. Although agronomic traits have been used for grouping the oil palm genotypes, agro-morphological traits with high heritability were considered in the present study. The Ufuma x Umuabi progeny (DT9) ranked highest for average bunch weight production with the least CV suggesting the stability of this genotype for this agronomic trait. Such significant genetic variation has also been reported (Okwuagwu and Okoye, 2006; Okwuagwu et al., 2008; Okoye et al., 2008) on agronomic traits in oil palm. Better understanding on the influence of environment on these quantitative traits would help to

group the genotypes with better accuracy. Comparatively, microsatellite analysis revealed higher genetic diversity among the 10 D × T oil palm progenies than the agronomic analysis. The percentage of polymorphic loci was very high (100%) across all the progenies, indicating the superiority of SSR markers over agronomic markers. Agronomic traits are often limited by high cost and long term field evaluation, low polymorphism and the influences of environmental factors. Thus, these traits may not adequately represent the genetic diversity among genotypes.

However, the three progenies (DT5, DT7, and DT10) with the highest CV over the five agronomic traits also exhibited high values of genetic diversity parameters demonstrating the high genetic diversity existing in the progenies. Relatively higher genetic diversity ($A_o = 4.4$, $H_o = 0.642$ and $H_e = 0.576$) was observed in Ecuador × Umuabi progeny (DT7) compared to the other progenies. Interestingly, this progeny also had higher values for agronomic traits such as number of harvested bunch (6.154), mesocarp to fruit ratio (68.9), and most importantly, comprising palms with the least annual growth rate (height = 44.47 cm/year). This finding could suggest a possible correlation between the extent of genetic variation within progenies and agronomic traits in oil palm. Therefore, palms from this progeny should get priority for breeding high bunch yielding and dwarf planting materials. Variable efficiencies of different marker systems for detecting genetic diversity in oil palm have been reported in oil palm using isozyme (mean $A_o = 1.6$; Hayati et al., 2004) and RFLP technique (mean $A_o = 1.8$; Maizura et al., 2006). High genetic diversity has been reported in both oil palm breeding materials and natural collections irrespective of both the country of origin and the genetic marker technique used (Bakoumé, 2016).

Relatedness of 10 D × T progenies based on agronomic markers and SSR markers

Results of Manhattan and Rogers' dissimilarity coefficients for the respective agronomic traits and SSR marker data were comparable. The highest genetic distance (highest genetic diversity) for agronomic data (0.617) and molecular data (0.4575) corresponded well to progenies DT6 (Ecuador Deli × Calabar) and DT9 (Ufuma × Umuabi). Parents of DT9 and the female parent of DT6 are new introductions in the breeding programme hence the very high genetic diversity in the progenies. Both methods of analyses classified the ten progenies into two main clusters with some disagreements in the grouping of progenies. Although not always satisfactory, Manhattan dissimilarity coefficients differentiated the genotypes predominantly on the basis of agronomic traits. This is an indication of the extent of variation across the oil palm genotypes for the quantitative traits. Similar results were reported by Arias et al. (2013) in the evaluation of 43 oil

palm progenies from Angola with nine phenotypic traits and 30 SSR markers. Similarly, Rogers' dissimilarity coefficients based on SSR marker data differentiated the oil palm progenies into two different groups with respect to their pedigree or shared ancestry. This demonstrates the effectiveness of SSR markers in identifying close pedigree relationship in breeding material.

On comparing agronomic and molecular clustering patterns, molecular markers have proved to be efficient in pedigree characterization of the NIFOR oil palm progenies. The progenies clustered according to the shared pedigree of either a common male or female parent. For instance, DT8, and DT7 share the same male parent (T6) from Umuabi while DT5 and DT10 share the same male parent (T5) from Ufuma. The grouping according to pedigree or shared ancestry is supported by previous findings in oil palm using molecular markers. Norziha et al. (2008) classified 16 D × P oil palm progenies from MPOB into 4 major clusters based on pedigree information with the aid of nine microsatellite markers. However, a tendency of clustering of progenies based on either agronomic performance or pedigree was observed using agronomic markers. For example, DT7 and DT8 were grouped in sub-cluster IID based on high bunch number in addition to sharing the same male parent palm (T6) from Umuabi. Similar result regarding effectiveness of SSR markers in monitoring genetic diversity for yield component traits as well as quality traits have also been reported in oil palm (Abdullah et al., 2011; Solin et al., 2014). Both agronomic and microsatellite clusters distinguished DT9 from the other progenies, thereby establishing high genetic diversity of the genotype. Progeny DT9 was derived from a high bunch weight Ufuma *dura* parent AD3, crossed to a high BN and O/M Umuabi open pollinated *tenera* parent (T6). The more recent assessment of the NIFOR main breeding parent genotypes by Okoye et al. (2016) with 10 microsatellite markers revealed the prevalence of private alleles in the Ufuma *dura* (AD3) and Umuabi *tenera* (T6) parents. It is possible that the pedo-climatic conditions of the parents' provenances may have an adaptive value on the genotypes to justify the presence of private alleles as proposed by Zeng et al. (2004). The highest number of private alleles was reported in the *tenera* parents from Umuabi. This provenance is derived from savannah ecology and generally regarded as marginal for oil palm production with rainfall of about <2000 mm per annum and a well-drained sandy clay loam soil. The private alleles found in such a marginal environment may have an adaptive value. The palms are characterized by slow stem increment, high bunch yield, and palm wine (alcoholic beverage from oil palm sap) production. Then again, Ufuma is a rainforest with rich clayey loam soil and high rainfall (>2000 mm). This area is characterized by high yielding palms with good fruit and bunch composition traits in addition to the unusually high proportion of *tenera* (thin-shelled) palms. Additional

support on the adaptive genetic variant is likely in sub-cluster IID comprising DT1, DT2, and DT4 (Figure 2). This is similar to the results of the agronomic analysis for the Deli *dura* derived progenies (sub-cluster IIE = DT1, DT2, and DT6; Figure 1), except that DT6 is included with the other Serdang avenue Deli progenies. The similarity between phenotypic and molecular marker analysis could suggest that SSR markers may be highlighting expressed traits with adaptive significance.

The grouping of the progenies obtained through PCoA on the basis of SSR marker data did not confirm that obtained by UPGMA cluster analysis in contrast to the agronomic dataset. Besides, the clustering provided by the dendrograms failed to reflect the genetic relatedness of the progenies observed in the PCoA plots. The PCoA plot based on SSR data is more revealing in the grouping of progenies with respect to their parentage or shared ancestry; for instance, the grouping of progenies DT1, DT3, and DT6 in cluster B. DT1 is associated with DT3 by a common female Serdang Avenue Deli grandparent (19 × 65) while DT6 share similar Calabar *tenera* male parent (T2) with DT1 (Figure 4). Additional support is the classification of DT7 and DT2 in cluster D, both of which contain the same male *tenera* parent (T6) from Umuabi. The genetic information based on molecular data enables the accurate grouping of genotypes sharing common lineage or genotypes developed for specific objectives. Cochard et al. (2009) also used PCoA to delineate and visualize 318 individuals from 26 origins and eight countries into three groups. With few exceptions, the agronomic based PCoA showed very little grouping according to known pedigree of the progenies (Figure 3). The grouping of DT1, DT2, and DT6 in cluster A and DT7 and DT8 in cluster B revealed some relatedness with respect to parentage of the progenies. DT7 and DT8 share the same male parent (T6) from Umuabi irrespective of their agronomic similarity in terms of mesocarp to fruit and oil to mesocarp ratios. Overall, both PCoA methods classified the 10 oil palm progenies into four groups with some deviations in the grouping pattern.

The low correlation between genetic distances calculated from the two approaches could be due to the fact that DNA markers reports genetic variation also in non-coding regions which hardly have an effect on phenotype. On the other hand, quantitative traits are influenced by environmental factors and their phenotype is a product of genotype × environment interaction. Plants may be morphologically similar, but this does not necessarily imply genetic similarity, since different genetic bases can result in similar phenotypic expression (Khan et al., 2009). A large portion of variation detected by molecular markers is non-adaptive and is, therefore, not subject to either natural or artificial selection as compared with phenotypic characters, which in addition to pressure selection are influenced by the environment (Vieira et al., 2007). Nonetheless, the low or no correspondence between variation estimated by

molecular markers and agronomic characters should not be considered a limitation given the fact that genetic and morphological diversity work in different ways to determine the relationships among populations. Besides, the disparity as in the case of the present study suggests that progeny classification and selection of parents for crosses in NIFOR oil palm breeding programme should not be relied on only one method of evaluation (agronomic traits).

Conclusion

Although both agronomic and molecular analysis revealed differences in genotype clustering, they shared several common aspects, such as high diversity between DT9 and DT6 genotypes. The genotypes analyzed in the present work had been previously selected and evaluated in NIFOR breeding programme, and had exhibited good performance. Therefore, prior screening of the most divergent genotype pairs identified through both methods is suggested for evaluation of the relative agronomic performance of their hybrids to prevent inbreeding depression. In a conventional breeding programme, numerous crosses are normally performed and evaluated in field trials. According to the data described in this study, SSR-based genetic distances could be useful in selecting superior crosses between oil palm trees derived from a population with a broad genetic base. Hence, the application of SSR markers in NIFOR oil palm breeding could be instrumental in reducing the number of single-cross hybrids to be evaluated.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the facilities provided by the Director-General of Malaysian Palm Oil Board (MPOB) for microsatellite genetic analyses at the Advanced Biotechnology and Breeding Centre (ABBC) Selangor, Malaysia. The data on which this paper is based formed part of the Ph.D. thesis submitted by the first author to the University of Nigeria, Nsukka. The assistance of Mrs. Maryann Okoye, Ms. Colette Chizitalum, Messrs. Innocent Ani, Sylvester Odifiri, and Jeremy Okoye in field work is also acknowledged.

REFERENCES

- Abdullah N, Yusop MR, Maizura I, Ghizan S, Latif MA (2011). Genetic variability of oil palm parental genotypes and performance of its' progenies as revealed by molecular markers and quantitative traits. *C. R. Biol.* 334:290-299.

- Arias D, González M, Prada F, Restrepo E, Romero H (2013). Morpho-agronomic and molecular characterisation of oil palm *Elaeis guineensis* Jacq. material from Angola. *Tree Genet. Genomes* 9:1283-1294.
- Arias D, Montoya C, Rey L, Romero H (2012). Genetic similarity among commercial oil palm materials based on microsatellite markers. *Agron. Colomb.* 30(2):188-195.
- Ataga CD (1993). Genotype-environment interaction and stability analyses for bunch yield in the oilpalm (*E. guineensis* Jacq.). *Oleagineux* 2:59-63.
- Ataga CD, Okolo EC, Okwuagwu CO (2005). Evaluation of an oil palm *Elaeis guineensis* Jacq genetic material collected from marginal zone of Nigeria. In. *Proceeding of PIPOC International Palm Oil Congress*, Malaysian Palm Oil Board, Kuala Lumpur.
- Bakoumé C (2016). Genetic Diversity, Erosion, and Conservation in Oil Palm (*Elaeis guineensis* Jacq.). In. Ahuja MR, Mohan Jain S (eds) *Genetic Diversity and Erosion in Plants: Case Histories* 2:1-33.
- Bakoumé C, Aziah MY, Praveena T, Teh CK, Suzaini Y, Hamidah M, Jangi MS, Basiran MN, Khairudin H, Harikrishna K (2011). DNA Sequence-Based Markers for Verification of Ramet-to-Ortet Relationship in Oil Palm (*Elaeis guineensis* Jacq.). *Am. J. Plant Sci.* 2:539-548.
- Bakoumé C, Wickneswari R, Rajanaidu N, Kushairi A, Amblard P, Billotte N (2007). Allelic diversity of natural oil palm (*Elaeis guineensis* Jacq.) populations detected by microsatellite markers: implication for conservation. *Plant Genet. Res.* 5(2):104-107.
- Bakoumé C, Wickneswari R, Siju S, Rajanaidu N, Kushairi A, Billotte N (2015). Genetic diversity of the world's largest oil palm (*Elaeis guineensis* Jacq.) field genebank accessions using microsatellite markers. *Genet. Resour. Crop Evol.* 62:349-360.
- Billotte N, Marseillac N, Risterucci AM, Adon B, Brottier P, Baurens FC, Singh R, Herra'n A, Asmady H, Billot C, Amblard P, Durand-Gasselin T, Courtois, B, Asmono D, Cheah SC, Rohde W, Ritter E, Charrier A (2005). Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* 110:754-765.
- Billotte N, Rusterucci A, Barcelos E, Noyer J, Amblard P, Baurens F (2001). Development, characterization, and across-taxa utility of oil palm (*Elaeis guineensis* Jacq.) microsatellite markers. *Genome* 44(3):413-425.
- Blaak G, Sparnaaij LD, Menendez T (1963). Breeding and inheritance in the oil palm (*Elaeis guineensis* Jacq.). Part II. Methods of bunch quality analysis. *J. West Afr. Inst. Oil Palm Res.* 4:146-155.
- Cochard B, Adon B, Rekima S, Billotte N, Desmier De Chenon R, Koutou A, Nouy B, Omoré A, Razak P, Glaszmann JC, Noyer JL (2009). Geographic and genetic structure of African oil palm diversity suggests new approaches to breeding. *Tree Genet. Genomes* 5(3):493-504.
- Corley RHV, Tinker PB (2003). Selection and Breeding. In: *The oil palm*, 4th edition. Blackwell Science.
- Doyle JJ, Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* 12:13-15.
- Gower JC (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 63:315-328.
- Hayati A, Wickneswari R, Maizura I, Rajanaidu N (2004). Genetic diversity of oil palm (*Elaeis guineensis* Jacq.) germplasm collections from Africa: implications for improvement and conservation of genetic resources. *Theor. Appl. Genet.* 108:1274-1284.
- Khan M A, von Witzke-Ehbrecht S, Maass BL, Becker HC (2009). Relationships among different geographical groups, agromorphology, fatty acid composition and RAPD marker diversity in safflower (*Carthamus tinctorius*). *Genet. Resour. Crop Evol.* 56:19-30.
- Kushairi A, Rajanaidu N, Jalani BS, Zakri AH (1999). Agronomic performance and genetic variability of *dura* x *pisifera* progenies. *J. Oil Palm Res.* 11(2):1-24.
- Liu K, Muse SV (2005). PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128-2129.
- Maizura I, Rajanaidu N, Zakri A, Cheah S (2006). Assessment of Genetic Diversity in Oil Palm (*Elaeis guineensis* Jacq) using restriction fragment length polymorphism (RFLP). *Genet. Res. Crop Evol.* 53(1):187-195.
- Mohammadi SA, Prasanna BM (2003). Analysis of genetic diversity in crop plants salient statistical tools and considerations. *Review and Interpretation. Crop Sci.* 43:1235-1248.
- Noh A, Rafii MY, Mohd Din A, Kushairi A, Norziha A, Rajanaidu N, Latif MA, Malek MA (2014). Variability and performance evaluation of introgressed Nigerian *dura* x Deli *dura* oil palm progenies. *Genet. Mol. Res.* 13(2):2426-2437.
- Norziha A, Rafii M, Maizura I, Ghizann S (2008). Genetic variation among oil palm parent genotypes and their progenies based on microsatellite markers. *J. Oil Palm Res.* 20:533-541.
- Okoye MN, Okwuagwu CO, Uguru MI (2008). Genotype and genotype by environment (GGE) biplot analysis of fresh fruit bunch yield and yield components of oil palm (*Elaeis guineensis* Jacq.). *J. Appl. Biosci.* 8(1):288-303.
- Okoye MN, Uguru MI, Bakoumé C, Singh R, Okwuagwu CO (2016). Assessment of Genetic Diversity of NIFOR Oil Palm Main Breeding Parent Genotypes Using Microsatellite Markers. *Am. J. Plant Sci.* 7:218-237.
- Okwuagwu CO, Ataga CD, Okoye MN and Okolo EC (2011). Germplasm Collection of Highland Palms of Afikpo in Eastern Nigeria. *Bayero J. Pure Appl. Sci.* 4(1):112-114.
- Okwuagwu CO, Ataga, CD, Okolo EC, Ikenobe CE, Ugbah MM (2005). The production of NIFOR Elite *Tenera* hybrid planting material – The NIFOR EWS. Technical Report, Nigerian Institute for Oil Palm Research (NIFOR), Benin City, Nigeria.
- Okwuagwu CO, Okoye MN (2006). Developments from the NIFOR Second Cycle Oil Palm Main Breeding and Selection Programme. Seminar paper presented in June 2006 at ERLSD, Hall, NIFOR BENIN.
- Okwuagwu CO, Okoye MN, Okolo EC, Ataga CD, Uguru MI (2008). Genetic variability of fresh fruit bunch yield in Deli/*dura* x *tenera* breeding populations of oil palm (*Elaeis guineensis* Jacq.) in Nigeria. *J. Trop. Agric.* 46(1-2):40-45.
- Omoti U (2009). Oil palm sector analysis in Nigeria. Volume 1-Main Report. Submitted to the United Nations Industrial Development Organization (UNIDO), Abuja, 64 + xviii + 275pp.
- Peakall R, Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6:288-295.
- Peakall R, Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28:2537-2539.
- Perrier X, Jacquemoud-Collet JP (2006). DARwin software <http://darwin.cirad.fr/>.
- Purba AR, Noyer JL, Baudouin L, Perrier X, Hamon S, Lagoda PJJ (2000). A new aspect of genetic diversity in Indonesian oil palm (*Elaeis guineensis* Jacq.) revealed by isoenzyme and AFLP markers and its consequences for breeding. *Theor. Appl. Genet.* 101:956-961.
- Rajanaidu N, Kushairi A, Rafii M, Din M, Maizura I, Jalani BS (2000). Oil palm breeding and genetic resources. In: Basiron Y, Jalani BS, Chan KW(eds) *Advances in Oil Palm Research*. Malaysian Palm Oil Board, Kuala Lumpur. pp. 171-227.
- Rogers JS (1972). Measures of genetic similarity and genetic distance. In. *Studies in Genetics VII*. University of Texas Publication 7213, Austin, Texas. pp. 145-153.
- Singh R, Noorhariza MZ, Ting NC, Rozana R, Tan SG, Low LET, Ithnin M, Cheah SC (2008). Exploiting an oil palm EST the development of gene-derived and their exploitation for assessment of genetic diversity. *Biologia* 63:1-9.
- Sokal RR, Michener CD (1958). A statistical method for evaluating systematic relationships. *The University of Kansas Science Bulletin* 38:1409-1438.
- Solin M, Sobir, NuritaToruan-Mathius N (2014). Genetic Diversity of D x P Population Yield Component in Oil Palm's Paternal Half-Sib Family Based on Microsatellite Markers. *Energy Procedia* 47:196-203.
- Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24:1596-1599.
- Ting NC, Noorhariza MZ, Rozana R, Low ET, Ithnin M, Cheah SC, Tan SG, Singh R (2010). SSR mining in oil palm EST database: application in oil palm germplasm diversity studies. *J. Genet.* 89:135-145.
- USDA (2017). United States Department of Agriculture. Foreign

Agricultural Service/USDA Office of Global Analysis. April 2017.
Vieira E, Carvalho F, Bertan I, Kopp M, Zimmer P, Benin G, da Silva J, Hartwig I, Malone G, de Oliveira A (2007). Association between genetic distances in wheat (*Triticum aestivum* L.) as estimated by AFLP and morphological markers. *Genet. Mol. Biol.* 30:392-399.