MUIR, G., A. J. LOWE, C. C. FLEMING and C. VOGL (2004): High nuclear genetic diversity, high levels of outcrossing and low differentiation among remnant populations of *Quercus petraea* at the margin of its range in Ireland. Ann. Bot. **93**, 691–697.

NARUM, S. R., M. BANKS, T. D. BEACHAM, M. R. BELLINGER, M. R. CAMPBELL, J. DEKONING, A. ELZ, C. M. GUTHRIE, C. KOZFKAY, K. M. MILLER, P. MORAN, R. PHILLIPS, L. W. SEEB, C. T. SMITH, K. WARHEIT, S. F. YOUNG and J. C. GARZA (2008): Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. Mol. Ecol. **17**, 3464–3477.

OGDEN, R. (2008): Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. Fish and Fisheries **9**, 462–472.

ORRU, L., G. CATILLO, F. NAPOLITANO, G. DE MATTEIS, M. C. SCATA, F. SIGNORELLI and B. MOIOLI (2009): Characterization of a SNPs panel for meat traceability in six cattle breeds. Food Control **20**, 856–860.

PIRY, S., A. ALAPETITE, J. M. CORNUET, D. PAETKAU, L. BAUDOUIN and A. ESTOUP (2004): GENECLASS2: A software for genetic assignment and first-generation migrant detection. J. Hered. **95**, 536–539.

RANNALA, B. and J. L. MOUNTAIN (1997): Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci. U.S.A. **94**, 9197–9201.

STEINKELLNER, H., S. FLUCH, E. TURETSCHEK, C. LEXER, R. STREIFF, A. KREMER, K. BURG and J. GLÖSSL (1997): Identification and characterization of $(GA/CT)_n$ microsatellite loci from *Quercus petraea*. Plant Mol. Biol. **33**, 1093–1096.

WASSER, S. K., C. MAILAND, R. BOOTH, B. MUTAYOBA, E. KISAMO, B. CLARK and M. STEPHENS (2007): Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban. Proc. Natl. Acad. Sci. U.S.A **104**, 4228–4233.

WASSER, S. K., A. M. SHEDLOCK, K. COMSTOCK, E. A. OSTRANDER, B. MUTAYOBA and M. STEPHENS (2004): Assigning African elephant DNA to geographic region of origin: Applications to the ivory trade. Proc. Natl. Acad. Sci. U.S.A. **101**, 14847–14852.

WEIR, B. S. (1990): Genetic Data Analysis. Sinauer Associates, Sunderland, MA.

XING, J. C., W. S. WATKINS, D. J. WITHERSPOON, Y. H. ZHANG, S. L. GUTHERY, R. THARA, B. J. MOWRY, K. BULAYEVA, R. B. WEISS and L. B. JORDE (2009): Fine-scaled human genetic structure revealed by SNP microarrays. Genome Research **19**, 815–825.

Population Structure and Molecular Characterization of Nigerian Field Genebank Collections of Cacao, *Theobroma cacao* L.

By P. O. AIKPOKPODION[1),2),4),*)], M. KOLESNIKOVA-ALLEN[2)], V. O. ADETIMIRIN[3)], M. J. GUILTINAN[4)], A. B. ESKES[5)], J.-C. MOTAMAYOR[6)] and R. J. SCHNELL[7)]

(Received 2nd October 2008)

Abstract

Inadequate knowledge of the population structure and diversity present often hamper the efficient use of germplasm collections. Using a high through-put system, twelve microsatellite loci were used to analyze genetic diversity and population structure in a national field genebank repository of 243 cacao accessions grouped into 11 populations based on their known sources. Based on multi-locus profiles, the Bayesian method was used for individual assignment to verify membership in each population, determine mislabeling and ancestry of some important accessions used in breeding program. A total of 218 alleles was revealed with a mean number of 18.2 alleles per locus. Gene diversity (H_e = 0.70) and allelic richness (4.34 alleles per locus) were highest in the F_1 hybrid population. Differential mating system was suggested as responsible for the observed deficit and excess of heterozygotes observed among the populations. Analysis of molecular variance showed that within-population variance accounted for 63.0% of the total variance while the rest 37% was accounted for by the among-population variance. Cluster dendrogram based on UPGMA revealed two main subsets. The first group was made up of the Amelonado/Trinitario ancestry and the other of Nanay/Parinari ancestry. We found that Nanay and Parinari populations were the major source of Upper Amazon genes utilized while a large proportion of genetic diversity in the field genebank remained under-utilized in development of improved cultivars released to

1) Plant Breeding Division, Cocoa Research Institute of Nigeria PMB 5244 Ibadan Nigeria.

2) Central Biotechnology Laboratory, International Institute of Tropical Agriculture PMB 5320 Ibadan Nigeria.

3) Department of Agronomy, University of Ibadan, Nigeria.

4) Department of Horticulture, 422 Life Sciences Building, Pennsylvania State University, University Park, PA 16802, USA.

5) Bioversity International, Parc Scientifique Agropolis II, 34397 Montpellier Cedex 5, France.

6) Mars Incorporated, Subtropical Horticulture Research Station, USDA-ARS, 13601 Old Cutler Road, FL33158.

7) Subtropical Horticulture Research Station, USDA/ARS, 13601 Old Cutler Road, FL33158.

*) Corresponding author's current address: Department of Genetics and Biotechnology, University of Calabar, PMB 1115, Calabar, Cross River State, Nigeria. E-Mail: paikpokpodion@yahoo.com

farmers in Nigeria. This study showed that the presence of alleles of the Upper Amazon Forasteros (Nanay, Parinari and Iquitos Mixed Calabacillo) genetic materials in the locally available accessions predated the formal large scale introduction of Upper Amazon materials in 1944. This is the first report of population structure of field genebank collections of cacao in Nigeria since more than seven decades of formal cacao breeding research.

Key words: cacao, fixation index, germplasm, cultivars, gene diversity, Hardy-Weinberg equilibrium.

Introduction

Cacao, *Theobroma cacao*, a diploid fruit tree species (2n = 20) originally designated a member of the Sterculiaceae family (PURSEGLOVE, 1974), but recently re-classified into the Malvaceae (ALVERSON et al., 1999) is native to the humid tropical regions of the northern parts of South America, and the northern parts of Central America (CHEESMAN, 1944; CUATRECASAS, 1964). However, since introduction in the late 19th century, the West and Central African region has become the largest producer accounting for more than 70% of the world's cocoa output of 3.592 million metric tons (ICCO, 2008). Dried cocoa beans obtained from the cacao tree are the main products used in the chocolate and confectioneries industries worth an estimated US$70 billion annually in the United States alone and exports from producing countries worth US$5 – 6 billion annually (GUILTINAN, 2007). In Africa, cocoa production is predominantly a smallholders' enterprise with several hundred-thousand families depending on this cash crop for their livelihood and significant foreign exchange earnings for producing countries (RICE and GREENBERG, 2000, MOTAMAYOR et al., 2008). Nigeria currently ranks fourth globally among cocoa producers, and in comparison with other agricultural commodities, cocoa makes the largest non-oil contribution to the nation's economic development and accounted for 65% of total agricultural export in 2004 (AIKPOKPODION, 2007).

Since the first introduction of cacao into Nigeria in 1874, when Squiss Ibaningo transported pods of '*Amelonado*' cocoa from Fernando Po, there has been a series of additional germplasm introductions as reviewed by (BARTLEY, 2005; AIKPOKPODION, 2007). However, formal selection programs and germplasm conservation started around 1931 by O.J. VOELCKLER at the Nigerian Department of Agriculture in Moor Plantation, Ibadan by the Colonial Administration. Further germplasm introductions of some Trinitario and Criollo selections from Trinidad and Ceylon, respectively, were made in 1933 (JACOBS et al., 1971). Several materials belonging to Upper Amazon Forastero and Trinitario populations were also introduced from Trinidad by the West African Cocoa Research Institute (WACRI) into Tafo, Ghana in 1944 (TOXOPEUS, 1964). In the mid-1960's, large scale introductions sponsored by the Cocoa Alliance of London was initiated in Nigeria which consisted of 313 clones and 701 seedling progenies derived from some 350 intra-Nanay, intra-Parinari, intra-Iquitos and inter-P (Pound's selections) crosses. Cocoa germplasm materials were also introduced from Costa Rica, Indonesia, Fernando Po, Kew Gardens (United Kingdom), Wageningen (The Netherlands) and Miami (USA) (JACOBS et al., 1971; OLATOYE and ESAN, 1992). Although several hundred clones and accessions are now existing in the Nigerian field collection, there has been no study on the genetic diversity of this collection as is the case with many cacao germplasm collections worldwide (ZHANG et al., 2006a, ZHANG et al., 2009; MOTILAL et al., 2009, JOHNSON et al., 2009, MOTAMAYOR et al., 2008).

A rational and efficient use of available germplasm collection depends largely on the knowledge of the nature and amount of genetic diversity present in the collections as well as the relationships among various accessions. This knowledge can also guide in the formulation of appropriate breeding strategies for the development of improved planting materials and integration of useful diversity into the breeding program. Considering the current climate variability, disease pressure, land pressure due to increasing population and urbanization that makes less land available for new plantings, increasing cost of production due to increased labour wages, old age of farms and farmers, the use of improved germplasm for new plantings and rehabilitation of old plantations becomes imperative. This requires a judicious and an optimum use of available genetic resources to develop well-adapted improved cultivars.

The use of molecular markers provides the most efficient means of assessing the extent of genetic diversity in germplasm collections. An international consortium of scientists in academic centers and government sponsored laboratories involved in the cocoa industry have agreed to the use of a set of standardized simple sequence repeat (SSR) primers for characterizing cacao germplasm collections (SAUNDERS, 2004, ZHANG et al., 2006a). The SSR marker system developed in cacao (LANAUD et al., 1999) has played significant role in increasing our understanding of the genetic structure and diversity in cacao germplasm (LANAUD et al., 2001; SERENO et al., 2006; ZHANG et al., 2006b; AIKPOKPODION et al., 2009), clone identification (CRYER et al., 2006, SAUNDERS et al., 2004) identification of mislabelled clones (ZHANG et al., 2006a), parentage analysis (SCHNELL et al., 2005) and of the origin and dispersal of cacao (MOTAMAYOR et al., 2002, 2003, 2008).

In view of the growing interdependence of all cocoa-growing countries for testing and utilization of cacao genetic resources, which makes the characterization of these resources a matter of global importance, we used microsatellite markers to investigate the genetic diversity of field genebank collections in Nigeria. This information will be useful to determine how much of the global cacao genetic resources are represented in this collection, to what extent these have been used to develop varieties released to farmers and how to exploit useful germplasm for development of improved varieties in future breeding programs.

Materials and Methods

Plant materials and sample collection

Leaf samples used for generating DNA fingerprinting profiles were taken from trees of 243 cacao clones and accessions in field genebank locations of the Cocoa

Table 1. – Field genebank accessions assigned to their 'Self' and 'Other' population by Bayesian clustering method.

Population	No. genotyped accessions	No. correctly assigned	Other population	% Error rate	No. verified identity
IMC	7	7	0	0	13
NA	10	8	2	20.0	20
PA	10	8	2	20.0	17
SCA	6	6	0	0	10
Amelonado	7	7	0	0	22
Trinitario	10	9	1	10.0	29
Local Selection	14	5	9	64.3	15
F_1 Hybrid	104	28	76	73.1	28
F_2 Hybrid	20	19	1	5.0	41
GB Selections	48	22	26	54.2	22
GBTRDI	14	10	4	28.6	17
Total	250	129	121	46.4	234

Research Institute of Nigeria (CRIN) headquarters, Ibadan (07.02 °N, 03.09 °E; 122 m above sea level <asl>) and her four sub-stations in Owena, Ondo State (07.20 °N, 05.03 °E; 263 m asl), Uhonmora, Edo State (06.81 °N, 06.02 °E; 160 m asl), Ibeku, Abia State (05.54 °N, 07.58 °E; 82 m asl) and Ajassor, Cross River State (05.88 °N, 08.82 °E; 140 m asl). These were classified into 11 main groups.

1. Local Selections ('LOCAL') collected from the local population of West Africa cacao landraces available in the country at the inception of cocoa research in the early 1930s by earlier workers (LOCKWOOD and GWAMFI, 1979). These were coded in the genebank collection as C-clones (C1–C38). This population is a mixture of both local Amelonado (green pods) and local Trinitario (red pods).

2. Trinitario materials ('TRINITARIO') are Imperial College Selections ('ICS') clones obtained from Trinidad and have been used as parents in some crosses made available to farmers as planting materials.

3. Amelonado ('AMEL') population representing the Bahian Amelonado introduced by the Portuguese to Sao Tome and Principe which formed the base population first cultivated in West Africa.

4. Parinari ('PA') population,

5. Nanay ('NA') population,

6. Iquitos Mixed Calabacillo ('IMC') population

7. Scavina ('SCA') population.

The PA, NA, IMC and SCA populations are the main Upper Amazon Forastero progenitors of the introduced germplasm from Trinidad in the early 1940s, that became the basis of Amazon cocoa released in all West African countries breeding programmes.

8. F_1-Hybrid clones ('F_1-HYB') are the first generation progenies derived from crosses among Upper Amazon clones as well as their crosses with other populations such as Trinitario, introduced from Trinidad to West African Cocoa Research Institute (WACRI) Headquarters in Tafo (now Cocoa Research Institute of Ghana) and its Ibadan substation (now CRIN) in 1944 (POSNETTE and TODD, 1951). These were the clones principally used as parents in various national breeding programmes to develop hybrid selections distributed to farmers such as the WACRI Series I & II selections in Ghana and Nigeria, These hybrids were coded 'T' and 'C' clones.

9. F_2-Hybrid progenies ('F_2-HYB') were developed from either open-pollinated or controlled crosses of various F_1-clones. These have been used to develop the F_3-Amazon 'Synthetic' cultivar distributed to farmers in Nigeria.

10. Genebank Selections ('GB-S') are progenies in the genebank derived from several crosses including double cross and adaptability hybrids derived from the 1967 Intra-Nanay and Intra-Parinari crosses, open-pollinated derived crosses from Catongo Blanc and 'Adaptability Hybrids' involving crosses of Introduced Upper Amazon crosses with local Amelonado.

11. 1967 Trinidad Introduction ('GB-TRD') materials are bi-parental hybrid progenies of intra-Nanay and intra-Parinari clones introduced into Nigeria in 1967 (ATANDA, 1975).

DNA extraction, Polymerase chain reaction (PCR) and Electrophoresis

Extraction of total genomic DNA from recently fully expanded leaves about two months old was performed following the procedure outlined by BHATTACHARJEE et al. (2004), DNA concentration was adjusted to 2.5 ng.µL^{-1}. Twelve microsatellite markers previously reported (LANAUD et al., 1999) were used in this study. PCR amplification reactions were performed in a MJ research PTC 200 thermal cycler (MJ Research, Watertown, Mass) with 5 µL total volume, containing 0.5 µL of cacao DNA (~2.5 ng.µL^{-1}). All PCR reactions contained 0.5 µL of 10X PCR buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl), 0.5 µL of 25 mM $MgCl_2$, 0.125 µL each of forward and reverse primer (10 mM), 0.1 µL of 10 mM dNTPs), 0.05 U.µL^{-1} of Taq polymerase (Bioline). The thermal cycling profile consisted of the following: 4 min denaturation at 94 °C; followed by 32 repeats of the fol-

lowing cycle: 94°C for 30 s, 1 min at 46°C or 51°C annealing temperature depending on primer, 1 min extension at 72°C, with a final 7 min 72°C extension. Capillary electrophoresis (CE) was performed on ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems) in a 36 cm capillary array using POP 4 (Applied Biosystems) as previously described by AIKPOKPODION et al. (2009). Resulting data were analyzed with GeneMapper™ software version 3.5 (Applied Biosystems) for internal standard and fragment size determination.

Table 2. – Bayesian clustering analysis (PRITCHARD et al., 2000) of Local Selection population indicating probability of membership in group (– = less than 0.05 level of probability).

Clone	Probability of membership				
	NA	PA	IMC	Amelonado	TRINITARIO
C13	-	-	-	0.240	0.731
C14	-	-	-	-	0.922
C16	-	-	-	0.299	0.673
C17	-	-	-	0.964	-
C19	0.604	0.308	-	-	-
C20	-	-	0.482	0.139	0.336
C21	-	-	-	0.964	-
C22	0.554	0.372	-	-	-
C23	-	-	-	0.671	0.301
C24	-	-	-	0.283	0.692
C25	-	-	-	0.707	0.265
C26	-	-	-	0.926	-
C27	0.602	0.361	-	-	-
N38	-	-	-	0.962	-

Table 3. – Summary statistics of 12 microsatellite markers used in field genebank accessions of cacao in Nigeria.

Marker	LG	H_e	H_o	*PIC*	*K*	F_{is}	F_{it}	F_{st}	*HWE*
mTcCIR12	4	0.863	0.722	0.847	19	-0.077	0.162	0.222	0.001
mTcCIR15	1	0.894	0.807	0.882	17	-0.033	0.115	0.144	0.000
mTcCIR17	4	0.556	0.560	0.537	14	-0.070	0.167	0.222	0.125^{NS}
mTcCIR18	4	0.820	0.748	0.800	24	-0.149	0.123	0.237	0.099^{NS}
mTcCIR19	2	0.805	0.652	0.785	17	-0.031	0.236	0.260	0.000
mTcCIR21	3	0.788	0.760	0.773	23	-0.128	0.032	0.143	0.792^{NS}
mTcCIR24	9	0.512	0.536	0.497	18	-0.066	0.134	0.191	0.042
mTcCIR25	6	0.841	0.767	0.823	17	-0.194	0.099	0.245	0.006
mTcCIR26	8	0.782	0.772	0.751	14	-0.124	0.076	0.178	0.326^{NS}
mTcCIR3	2	0.903	0.789	0.892	23	-0.098	0.127	0.203	0.000
mTcCIR6	6	0.777	0.726	0.747	15	-0.156	0.127	0.245	0.011
mTcCIR9	6	0.739	0.642	0.715	17	-0.066	0.178	0.230	0.000
Mean over loci									
Jackknifing over all loci (mean±std. deviation)		0.773	0.707	0.754	18.2	-0.10±0.016	0.135±0.015	0.213±0.012	
Bootstrapping over all loci (1000BS)									
95% CI						-0.13 – -0.07	0.108 – 0.163	0.019 –0.234	
99% CI						-0.139 – -0.061	0.099 – 0.173	0.183 –0.239	

LG = Linkage group, H_{nb} = unbiased gene diversity, H_o = observed heterozygosity, PIC = polymorphism information content, K = total no of alleles, A_r = Allelic richness, F_{is} = inbreeding coefficient of individuals within subpopulations, F_{it} = inbreeding coefficient of individuals within total population, F_{st} = amount of variation due to differentiation between subpopulations, CI = Confidence Interval, NS = not significant at $P < 0.005$.

Table 4. – Diversity parameters of 11 cacao populations in Nigerian field genebank based on 12 microsatellite markers.

	n	H_e	H_o	$P_{(0.95)}$	K	A	A_e	A_p	F_{is}
F_1 Hybrid	28	0.70	0.79	1.00	67	7.25	4.34	0.42	-0.108
F_2 Hybrid	41	0.60	0.69	1.00	47	6.08	3.27	0.17	-0.138
GB-S	22	0.57	0.56	1.00	46	5.75	2.70	0.25	0.050
GB-TRD	17	0.61	0.68	0.92	44	5.67	3.03	0.58	-0.087
LOCAL	15	0.50	0.59	1.00	36	3.92	2.18	0.17	-0.145
AMEL	22	0.20	0.18	0.50	20	3.25	1.34	0.17	0.142
IMC	13	0.57	0.73	1.00	40	4.00	2.55	0.0	-0.236
NA	20	0.51	0.53	0.92	40	5.08	2.52	0.0	-0.001
PA	17	0.50	0.51	0.92	35	3.58	2.26	0.08	0.013
SCA	10	0.55	0.66	1.0	50	4.17	2.46	0.75	-0.152
TRIN	29	0.58	0.73	1.0	33	4.67	2.47	0.25	-0.230

n = number of samples, H_e = expected heterozygosity, H_o = observed heterozygosity, $P_{(0.95)}$ = proportion of polymorphic loci when most frequent allele does not exceed 95.0%, K = number of alleles with frequency > 5.0% in each population, A = Mean number of alleles per locus, A_e = effective number of alleles, F_{is} = inbreeding coefficient of individuals within subpopulations obtained from 1000 Bootstrap at probability level at 95%.

Data analysis

In the first step of analysis, an assignment test based on Bayesian method (*Table 1*) was carried out to verify genetic identity of accessions in the 11 populations including parental clones representing primary reference populations such as Nanay (NA), Parinari (PA), Scavina (SCA), Trinitario (TRIN), Amelonado (AMEL) and Local Selection (LOCAL) populations. Five Amelonado reference samples (SIAL 20, SIAL 70, SIAL 84, SIC 19 and SIC 23) were included in the assignment test. The assignment probabilities were computed for each individual to determine the degree to which its genome was classified into each cluster. Assignment of individuals into the *a priori* primary populations (PA, NA, IMC, SCA, TRINITARIO and LOCAL) was set at not <90% probability. Individuals in the *a priori* grouping that did not meet the criteria were considered as mislabeled and excluded from subsequent analysis. The program GENALEX ver. 6 (PEAKALL and SMOUSE, 2005) was used for the assignment test.

Genetic diversity within and among loci and populations was estimated using the software package GENALEX ver. 6, GENETIX 4.0.2 (BELKHIR, 2001) and FSTAT ver. 2.9.3 (GOUDET, 2001) for the following statistics: allelic frequencies, effective number of alleles (A_e) per locus for each population (the measure of the num-

Table 5. – Analysis of molecular variance (AMOVA) for microsatellite variation in 11 populations of cacao in Nigerian field genebank.

Source	df	SSD	MSD	Variance component	% Total	P-value
Among pops	10	793.64	79.36	3.50	37.0	0.0001
Within pops	223	1345.04	6.03	6.03	63.0	0.0001
F_1 Hybrid	28	206.02	7.36	-	-	-
F_2 Hybrid	41	249.43	6.08	-	-	-
GB-S	22	155.90	7.09	-	-	-
GB-TRD	17	110.17	6.48	-	-	-
LOCAL	15	73.95	4.93	-	-	-
AMEL	22	60.12	2.73	-	-	-
IMC	13	64.76	4.98	-	-	-
NA	20	119.85	5.99	-	-	-
PA	17	99.46	5.85	-	-	-
SCA	10	52.30	5.23	-	-	-
TRINITARIO	29	153.09	5.28	-	-	-
Total	234	2138.69	9.14	-	-	-

PhiPT = 0.367; N0 = 20.94 obtained from 9999 pairwise poplation permutations.

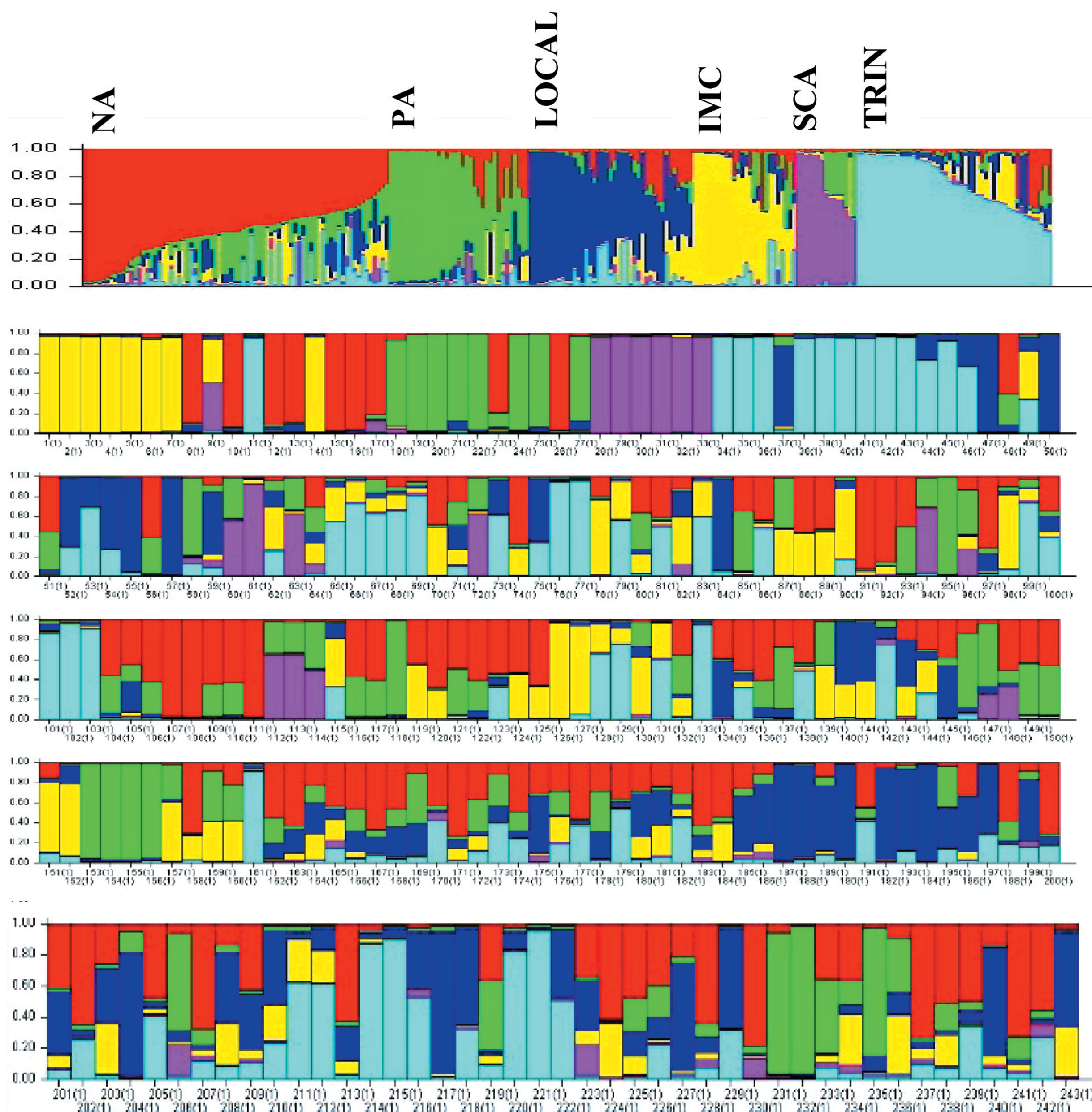

Figure 1. – Optimum alignment generated with CLUMPP (JAKOBSSON and ROSENBERG, 2007) from analysis of 243 accessions set at with k = 6 at 200,000 iterations after a burn-in period of 100,000 using STRUCTURE (PRITCHARD et al., 2000). Label 1-7 = IMC, 8-17 = NA, 18-27 = PA, 28-33 = SCA, 34-43 = TRIN, 44-57 = LOCAL, 58-161 = F_1 Hybrid, 162-181 = F_2 Hybrid, 182-229 = GB-S, 230-243 = GB-TRDI. NA = Nanay, PA = Parinari, LOCAL = Local Selections, SCA = Scavina, IMC = Iquitos Mixed Calabacillo, and, TRIN = Trinitario.

ber of alleles per locus independent of sample size), expected (H_e) and observed heterozygosity (H_o) (NEI, 1978). Standard errors for the above parameters and confidence intervals at 95% and 99% levels were estimated over all loci by bootstrapping (1000 bootstraps) and jackknifing (QUENOILLE, 1956; EFRON, 1982) using 1000 replications. Genetic differentiation for polymorphism between pairs of populations was analyzed by F-statistics estimator, $F_{st}(\theta)$ (WRIGHT, 1965), as described by WEIR and COCKERHAM (1984) based on 1000 permutations. $F_{st}(\theta)$ values were then subjected to standard Bonferroni corrections (HOLM, 1979; RICE, 1989) to guide against type I error. Exact test of deviation of fixation indices for each locus from Hardy-Weinberg equilibrium was done with TFPGA (MILLER, 1997) software. For each SSR marker, the polymorphism information content (PIC) value was calculated according to POWELL et al. (1996) using the computer program CERVUS (MARSHALL, 1998). Genetic distance between pairs of population was estimated based on unbiased minimum genetic distance (D_j) according to NEI (1978) and clustered based on UPGMA using TFPGA (MILLER, 1997).

Genetic structure in the genebank was determined by a hierarchical analysis of molecular variance (AMOVA, EXCOFFIER et al., 1992), implemented in GENALEX ver.6. The total molecular variance was partitioned as among population and individuals within population. The significance of Φ statistics was tested by permuta-

tion with 1,000 randomizations for probability of non-differentiation.

Population structure inference and ancestry of 243 samples were assessed with a Bayesian model-based clustering method, implemented in STRUCTURE (PRITCHARD et al., 2000). Based on the known number of parental groups, the *k* value was set from 5 to 7 and the analysis was carried out in 200,000 iterations after a burn-in period of 100,000. Twenty independent runs were assessed for each *k* value. Results of the five iterations with highest score in each replicated run were combined using CLUMPP to generate the optimum alignment (JAKOBSSON and ROSENBERG, 2007). Pair-wise genetic distances among individuals were computed in GENETIX 4.0.2 (BELKHIR, 2001) and presented on a three-dimensional scale plot using factorial analysis of correspondences (FAC) implemented in the same program.

Results

Assignment test and identification of mislabeling

With the prior classification of the genebank accessions into 11 population groups, the assignment test (*Table 1*) showed an error rate between 0.0% (IMC, SCA, AMEL) and 20% (NA, PA) in the parental accessions but higher error rate between 5.0% (F_2 Hybrid) and 73.1% (F_1 hybrid) in the rest populations. For instance, in the NA population, two accessions, NA 48 and Ghn-NA31 were detected as off-types belonging to Trinitario and IMC populations, respectively. Accessions labeled PA 35 and PA7 in the field genebank were also detected as off-types. An accession labeled Nig-ICS1 in the Trinitario population was assigned to the Local Selection group. In spite of this high error rate in the *a priori* classification, the assignment test was able to correctly assign 93.6% of individuals in their appropriate population groups. Sixteen incorrectly assigned and mislabeled individuals were excluded from subsequent analysis.

The assignment test also gave a clear differentiation among accessions in the Local Selections group (*Table 2*) originally classified based on pod colour as 'local Amelonado' (green pods) or 'local Trinitario' (red pods). Accessions C17, C21, C26 and N38 were classified as truly Amelonado sharing profiles with the reference SIC and SIAL Amelonado reference samples. Clone C14 was classified into Trintario group, sharing profiles with

Table 6. – Pairwise estimate of F_{st} (θ) for differentiation between pairs of populations over 12 microsatellite loci in Nigeria's *Theobroma cacao* genebank collection.

	F_2 Hyb	GB-S	GB-TRD	LOCAL	AMEL	IMC	NA	PA	SCA	TRIN
F_1 Hybrid	0.072	0.085	0.048	0.165	0.322	0.115	0.111	0.110	0.197	0.183
F_2 Hybrid		0.059	0.037	0.129	0.257	0.216	0.091	0.161	0.285	0.191
GB-S			0.121	0.074	0.166	0.221	0.204	0.189	0.288	0.156
GB-TRD				0.199	0.391	0.200	0.086	0.124	0.262	0.232
LOCAL					0.116	0.282	0.307	0.246	0.355	0.120
AMEL						0.502	0.491	0.403	0.565	0.339
IMC							0.218	0.322	0.302	0.304
NA								0.295	0.325	0.328
PA									0.369	0.281
SCA										0.338

Values ($p < 0.0001$) obtained from 10,000 permutations.

Table 7. – Estimate of NEI's (1978) unbiased minimum distance between primary cacao populations and groups in Nigeria's germplasm collection. LOCAL = Local selections, TRIN = Trinitario, IMC = Iquitos Mixed Calabacillo, NA = Nanay, PA = Parinari, SCA = Scavina, GB = Genebank, TRD = Trinidad Introduction.

	Nei's minimum distance						
	LOCAL	AMEL	IMC	NA	PA	SCA	TRIN
F_1 Hybrid	0.127	0.23	0.086	0.079	0.079	0.165	0.144
F_2 Hybrid	0.085	0.159	0.163	0.057	0.110	0.235	0.139
GB-S	0.045	0.079	0.166	0.144	0.130	0.236	0.108
GB-TRD	0.139	0.239	0.149	0.053	0.079	0.213	0.180
Local	-	0.041	0.212	0.230	0.166	0.293	0.074

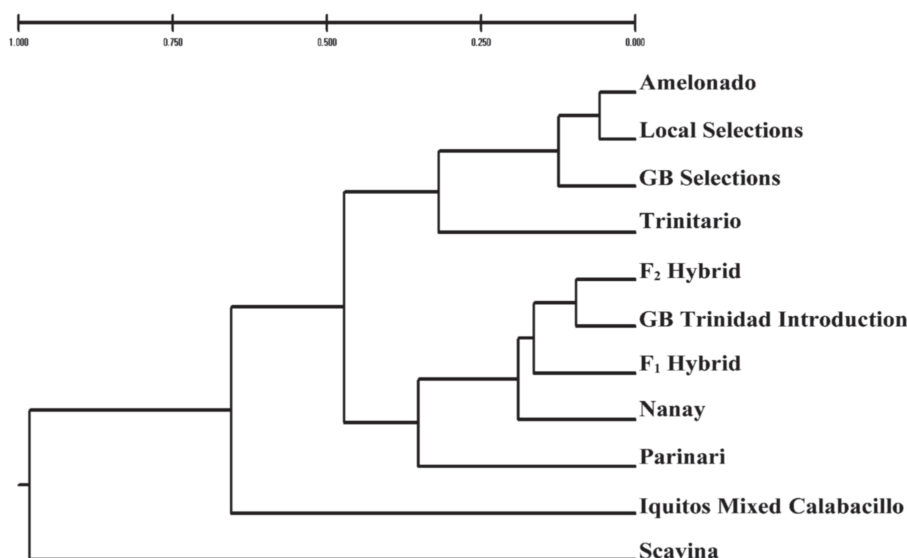

Figure 2. – Dendrogram of 11 cacao populations (including 234 accessions) in Nigerian cacao field genebank collection.

Trinidad's ICS reference clones. However, clones C13, C16, C23, C24 and C25, originally classified as 'Local Trinitario', were found to share alleles with local Amelonado. On the other hand, Clones C19, C22 and C27 were assigned as Upper Amazon's NA x PA hybrids and C20 as IMC x Trinitario hybrid from the Bayesian clustering implemented in STRUCTURE program (*Table 2*).

Genetic diversity assessment

A total of 218 alleles were obtained from the 234 accessions using 12 microsatellite loci (*Table 3*). The mean number of alleles per SSR locus (18.2) ranged from 14 for *mTcCIR17* to 24 for *mTcCIR18*. However, the mean expected heterozygosity ranged from 0.512 for *mTcCIR24* to 0.903 for *mTcCIR3*. The mean polymorphism information content (*PIC*) of the markers was 0.754 and ranged from 0.497 for *mTcCIR24* to 0.892 for *mTcCIR3*. Fixation index (F_{is}) indicated excess of heterozyzotes for all 12 loci, eight of which showed significant deviation from Hardy-Weinberg equilibrium (*Table 4*). However, the average fixation index indicated a general deficit of heterozygotes among sub-populations (F_{st} = 0.213) and total population (F_{it} = 0.135).

The mean number of alleles per locus for all 11 populations ranged from 3.25 for Amelonado population to 7.25 for F_1 Hybrid population (*Table 4*). The effective number of alleles per locus also ranged from 1.34 in Amelonado population to 4.34 in the F_1 Hybrid population. However, the number of private alleles was highest in the Scavina population (0.75), while the IMC and NA population had none. The mean gene diversity (H_e) ranged from 0.20 for the Amelonado to 0.70 for F_1 Hybrid population. The pattern of observed heterozygosity (H_o) was not different from the expected (H_e) but generally lower than the expected heterozygosity, except for Amelonado and Genebank Selection populations (*Table 4*). The number of alleles present was highest in the F_1 hybrid' population (k = 67) and followed by Scavina (k = 50). It was lowest in Amelonado (k = 20) and followed by Trinitario (k = 33). The Fixation index (F_{is}) was positive and indicated deficiency of heterozygotes in the Amelonado (0.142), Genebank Selection (0.050) and PA (0.013) populations (*Table 4*). On the other hand, it was negative for the rest populations such as IMC (–0.236), Trinitario (–0.230) and Local Selection (–0.145). This indicated that there was a significant excess of heterozygotes in these populations.

Population structure and genetic relationships

AMOVA showed that majority of the molecular variance (63.0%) was contributed by the within-population variance (*Table 5*). The among-population variance accounted for 37.0% of total variance. The population structure inference and ancestry assessed with a Bayesian model clustering method (*Figure 1*), based on known parental groups implemented in STRUCTURE (PRITCHARD et al., 2000) distinctly separated the IMC, NA, PA, SCA, Trinitario and Local Selection populations. However, the ancestry inference of accessions in the rest genebank populations indicated that they were mainly hybrids. Mislabeled and wrongly classified accessions were also identified.

Genetic differentiation between pairs of population (*Table 6*) estimated with F_{st} (theta) showed that the F_1 Hybrid population showed less differentiation from F_2 Hybrid, Genebank Trinidad Introduction and Genebank Selection than the other populations. Local and Genebank Selection also showed much less differentiation than others. The Nanay population showed the least differentiation from F_2 Hybrid and Genebank Trinidad Introduction than the other populations. The estimated unbiased minimum distance between pair of populations (*Table 7*) showed that F_1 Hybrid population had greater relationship with the PA, NA and IMC populations, while F_2 hybrid population was more related to NA and Local Selection populations. The NA and PA parental populations were more related to the Genebank Trinidad Introduction, while Local Selection and Amelonado populations were more related to the Genebank Selection population than others. The Local

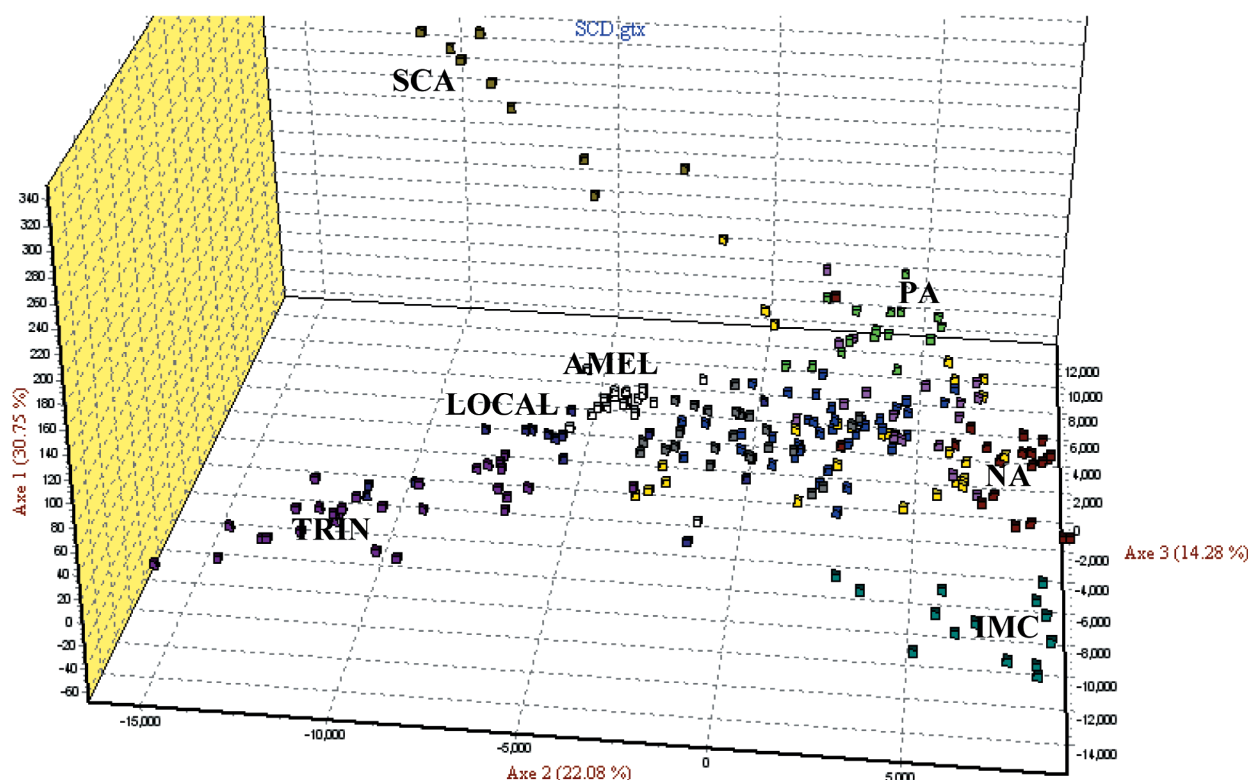

Figure 3. – Factorial analysis of correspondence (FAC) accounting for 67.1% of total variation showing relationships among accessions in the 11 population groups of cacao in the Nigerian field genebank collections. Legends are ■ = F_1 hybrid, ■ = F_2 hybrids, ■ = Genebank selections, and ■ = 1967 Genebank Trinidad Introduction. Other populations, Trinitario (TRIN), Scavina (SCA), Parinari (PA), Nanay (NA), Amelonado (AMEL), Local Selection (LOCAL) and Iquitos Mixed Calabacillo (IMC) are shown on the graph.

Selection population was most related to the Amelonado and Trintario populations.

The genetic relationship among the 11 populations is illustrated by the dendrogram in *Figure 2*. The 11 populations were divided into two main subsets. The first subset consisted of the Local Selection, Genebank Selection, Amelonado and Trinitario populations. Amelonado and Trinitario populations appeared to have provided parentage to accessions of the Local and Genebank Selection population groups in this subset. The second subset consisted of F_1 Hybrid, F_2 Hybrid, Genebank Trinidad Introduction, which appeared to have drawn parentage largely from the Nanay and Parinari populations. The Scavina and IMC populations appeared distantly related to the two main subsets. Factorial analysis of correspondence (*Figure 3*) accounting for 67.1% of total variance illustrates the relationships among field genebank accessions in the various population groups. It showed that accessions in the field genebank were more closely related to Nanay, Parinari, and Local Selections than to the Trinitario, Scavina and IMC populations which were clearly separated.

Discussion

The knowledge of the nature and amount of genetic diversity present in a germplasm collection is a major pre-requisite for exploitation of useful diversity in development of improved cultivars in breeding programs. The development of cacao microsatellite markers (Lanaud et al., 1999; Pugh et al., 2004), and the optimization of fingerprinting procedures (Zhang et al., 2006a; Cryer et al., 2006) have opened new opportunities to use microsatellite markers to both fingerprint and determine genetic diversity of germplasm collections. Until this study, there was no information available on the genetic structure of cacao genetic resources in Nigeria that has been used in more than seven decades of formal selection and breeding programme.

Verification of genetic identity and identification of mislabeled accessions

Although mislabeling and misidentification has been known in cacao collections, it was only recently that these have been estimated with the use of molecular markers (Motilal and Butler, 2003; Boccara and Zhang, 2006; Zhang et al., 2008). The assignment test determines the population of origin of a single individual through Bayesian method (Pritchard et al., 2000). This method needs only relatively small number of loci to detect a very strong signal of population structure and assign individuals appropriately. This has been used successfully to determine the structure of the cacao primary genepool (Motamayor et al., 2008), 'Refractarios' (Zhang et al., 2008). Until this study, there has been no such information on cacao germplasm collections maintained in Nigeria.

In this study, we found that a number of accessions that were wrongly identified. For instance, accessions labeled Ghn-NA31, PA 35 and PA7 was found to be off-

types. The genetic structure of the Local Selection population, originally classified based on pod colour as 'local Amelonado' (green pods) or 'local Trinitario' (red pods) was also elucidated with the assignment test used in this study. Accessions C17, C21, C26 and N38 were classified as truly Amelonado sharing profiles with SIC and SIAL Amelonado reference samples, while C14 was classified as Trintario, sharing profiles with Trinidad's ICS Trinitario reference clones. However, clones C13, C16, C23, C24 and C25 which were also originally referred to as Local Trinitario were found having Amelonado ancestry, indicating the hybrid nature of these clones. These clones could therefore be classified as 'Trintario-Amelonado' hybrids, and, distinguished from C14 with same SSR profiles as Trinitario ICS clones. A more surprising finding was the ancestry of clones C19, C22 and C27 that were assigned as Upper Amazon's NA x PA hybrids and C20 as IMC x Trinitario hybrid. These clones were originally believed to be Local Amelonado or Local Trinitario. This provides a very important clue to a limitation in past breeding programmes where these wrongly identified materials had been used in variety development. ROB LOCKWOOD (*personal communication*) had observed that the tree morphology and growth pattern of C19, C22 and C27, was atypical of local Amelonado, and had cast some doubt on their identity as Amelonado. However, till this study, there had been no scientific proof to support this observation. This finding that these local clones shared allelic profiles with Upper Amazon's Nanay, Parinari and Iquitos Mixed Calabacillo primary populations, therefore, indicated that cultivation of cacao having some Upper Amazon genetic background by local farmers in West Africa, especially Ghana and Nigeria predated the Posnette's 1944 Introduction, widely believed to be the first source of materials from the Upper Amazon population in West Africa. It also showed that apart from the Bahian 'Comum' Amelonado cocoa and red-podded Trinitario cocoa that were regarded as the base population in farmers' fields, some materials of Upper Amazon population had also reached farmers' plots before the inception of organized breeding programme in the early 1930s by the colonial administration in Nigeria. BARTLEY (2005) mentioned that although there had not been records of the original plantings, some varieties from Ecuador, Trinidad and Venezuela were introduced into Sao Tome around 1880, besides earlier introduction of Amelonado from Bahia in 1822. Seeds from these introduced materials were distributed to several farms during a period of active cultivation expansion. We can therefore suggest that some materials of Upper Amazon Forastero origin were among earlier materials introduced from Fernando Po into the West African mainland. These non-Amelonado materials could also have been introduced into the country through undocumented means or from other sources as the Kew Botanical Gardens by the colonial administration.

Since the structure analysis showed that a large percentage of the accessions were hybrids, it would be necessary to determine the ancestry of these accessions in order to facilitate their effective utilization in breeding program. The correctly verified accessions will also be useful as reference materials in the collection while mislabeled ones are removed from the genebank. This study provides a basis for the rationalization of the Nigerian field genebank collection of cacao in view of the cost and mislabeling problems involved in maintaining living tree collections.

Population structure and genetic variation

In this study, a total of 218 alleles were revealed from 12 microsatellite markers used to determine the genetic diversity in 11 populations consisting of 234 clones and accessions sampled from the germplasm repository of the Cocoa Research Institute of Nigeria. High levels of polymorphism were observed in the loci used with the number of alleles generally higher than reported by LANAUD et al., 1999 (*Table 2*). While equal number of alleles as observed by LANAUD et al. (1999) for *mTcCIR6*, *mTcCIR18* and *mTcCIR25* were obtained in the samples studied, certain markers exhibited an up to three-fold increase in numbers of alleles. For instance, 23 alleles per locus were obtained for *mTcCIR3* against 6 alleles per locus reported by LANAUD et al. (1999). The higher number of alleles obtained in this study could be explained by the higher number of samples, different genotypes and use of capillary electrophoresis (CE) system which allowed the detection of more alleles as could not be done with gel-based staining system, as was also observed by SERENO et al. (2006). Capillary electrophoresis is a high throughput system that allowed detection of a single base pair difference and is highly reproducible between laboratories (ZHANG et al., 2006a; CRYER et al., 2006).

Although gene diversity was low for the Amelonado population, it was generally high in the other field genebank collections ranging from 0.500 in the Local Selection to 0.70 in the F_1 Hybrid population. The F_1 hybrid population introduced into West Africa from Trinidad in 1944 (POSNETTE and TODD, 1951) presents a unique genetic makeup. The high genetic diversity (H_{nb} = 0.70) indicated high amount of genetic diversity in this population and the highest number of alleles per locus (7.25). This is an important source of genetic diversity that could be exploited to deal with production problems such as resistance to *Phytophthora* pod rot and drought. This wide genetic background which cuts across major primary populations could have been responsible for the large amount of gene diversity observed in this population. Although largely of Upper Amazon origin, a good proportion of the F_1 hybrid population which constituted the T-Clones used in the West African cocoa breeding programmes were of Trinitario and Criollo origin. Out of 59 hybrid crosses of the first batch of this population where only the mother tree is known, 27 are of Criollo and Trinitario origin, 15 of Upper Amazon, and eight were obtained from Ecuador. In the second batch of bi-parental crosses of known parents, 16 were of full sib Upper Amazon origin; five of Upper Amazon x Criollo or Trinitario, and eight of Trinitario or non-Upper Amazon origin. This wide genetic background could have been responsible for the large amount of gene diversity observed in this population and the large number of rare and distinct alleles not present in their Upper Amazon Forastero or Trinitario parents. This could be explained by a large proportion of

this population that was obtained from open pollinated mother trees (T1–T59) (Toxopeus, 1964). It would be important to further study this unique population not present anywhere else in the world for any unique attributes these clones might possess. Sereno et al. (2006) also stressed the importance of germplasm collections when they found several distinct alleles in CAB (Cacao Amazon Brazil) accessions collected from Lower Amazon region such as Alenquer, Para State in Brazil (Bartley, 2005).

Examination of the fixation index showed a significant deficiency of heterozygotes in Amelonado population which is consistent with reports by other workers (Motamayor et al., 2008; Sereno et al., 2006). The inbreeding coefficient (F_{is}) is one of the most important parameters used to describe the mating systems in flowering plants and inference is drawn from the distribution of allelic variation into genotypes and by comparison of observed genotypic proportions with those expected at Hardy-Weinberg equilibrium (Hayati et al., 2004; Ohtsuka et al., 2005). Iquitos Mixed Calabacillo, Scavina, F_1 Hybrid, F_2 Hybrid, Genebank Trinidad Introduction and Local Selection populations exhibited significant excess of heterozygotes. This indicated an outcrossing mating system and preponderance of self-incompatibility in these populations. On the other hand, Parinari, Genebank Selection and Amelonado populations showed deficit of heterozygotes. This suggests a tolerance to inbreeding in these populations. Studies have shown that degree of selfing varies often within a genus or within a species (Sweigart et al., 1999; Ohtsuka et al., 2005). This implies a need to study the compatibility status of the genebank entries in order to determine the self-compatibility status, or otherwise, of the accessions to facilitate their effective utilization in breeding program.

Estimate of pair wise genetic differentiation among the populations indicated that significant differentiation occur between the pairs of population. However, it appeared that F_2 hybrid population and Genebank Trinidad introduction are less differentiated as the case with the Local Selection and Genebank Selections' populations. The latter situation could have been due to the use of clones in the Local Selections group in the development of population in the Genebank Selections population, such as the WACRI Series II variety which was a cross between Local Selections and some F_1 hybrid clones (Atanda, 1975). The genetic differentiation of the Genebank Selections population from the F_1 hybrid, F_2 Hybrid and Upper Amazon populations indicated that only a small proportion of the genetic diversity in these groups have been utilized in variety development in more than four decades of cocoa breeding research at the Cocoa Research institute of Nigeria.

In an attempt to determine the level of utilization of the genetic diversity present in the genebank and populations that have had influence on varieties developed and distributed to farmers, results obtained showed clearly that the Upper Amazons Nanay, Parinari and Iquitos Mixed Calabacillo are the main ancestral progenitors of F_1 Hybrid population. However, the Nanay population contributed largely to the F_2 Hybrid population. The Genebank Selections derived mainly from Local and Amelonado. The Local Selection population derived mainly from the Amelonado and Trinitario population. The cluster diagram also showed that Upper Amazons Scavina, Iquitos Mixed Calabacillo and Trinitario populations are yet to be significantly exploited in variety development efforts. Moreover, the higher proportion of private alleles in Scavina, Genebank Trinidad Introduction and F_1 Hybrid populations showed that they are yet to be fully exploited for variety development. Important alleles locked up in these populations may be useful to address key production constraints. For instance, useful resistance to *Phytophthora* pod rot and witches broom associated with the Scavina population and the large bean size of Iquitos Mixed Calabacillo population may have been little utilized in the breeding program. In this study, we showed that only a small proportion of genetic diversity available in the Nigerian germplasm collection has been utilized for the development of improved varieties distributed to farmers. Most of clones that have been used in cacao breeding at CRIN have originated mainly from the Parinari and Nanay populations and the bulk of the 1967 Trinidad Introduction materials used were mainly intra-Nanay and intra-Parinari crosses (Atanda, 1975; Ojo et al., 1991). This study revealed the need for guided exploitation of useful diversity in Scavina and Iquitos Mixed Calabacillo populations for development of cultivars to address production problems such as *Phytophthora* pod rot, swollen shoot virus disease and bean quality in future breeding program.

Acknowledgement

This study was part of the first author's Ph.D research and financially supported by United States Department of Agriculture (USDA), Mars Incorporated and United States Agency for International Development (USAID) through the Norman Borlaug LEAP (Leadership Enhancement in Agricultural Program) Fellowship Award to the first author. The field and laboratory support of Messrs. Raji LO, Efunla MS (CRIN), Sunday Taiwo (IITA), Wilber Quinternella, Nannette Langevin and Cecile Olano (USDA-ARS SHRS Miami) are highly appreciated. The authors are also grateful to Dr. Dapeng Zhang (USDA-ARS Beltsville) for STRUCTURE analysis of the data. The useful suggestions of the two anonymous referees are gratefully acknowledged.

References

Aikpokpodion, P. O. (2007): Genetic diversity in Nigerian cacao, *Theobroma cacao* L. collections as revealed by phenotypic and simple sequence repeats marker. PhD Thesis submitted to the University of Ibadan, Nigeria. 147 pp.

Aikpokpodion, P. O., J. C. Motamayor, V. O. Adetimirin, Y. Adu-Ampomah, I. Ingelbrecht, A. B. Eskes, R. J. Schnell and M. Kolesnikova-Allen (2009): Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. Tree Genetics & Genomes **5**: 699–711.

ALVERSON, W. S., B. A. WHITLOCK, R. NYFELLER, C. BAYER and D. A. BAUM (1999): Phylogeny of the core Malvales: Evidence from *ndhF* sequence data. American Journal of Botany **86**: 1474–1486.

ATANDA, O. A. (1975): Assessment of WACRI Series I and II varieties of cocoa in Nigeria. 2. Pod value, butter fat content and flavour assessment. Ghana Journal of Agricultural Science **8**: 3–10.

BARTLEY, B. G. D. (2005): The genetic diversity of cacao and its utilization CABI Publishing, United Kingdom 341pp.

BELKHIR, K., P. BORSA, L. CHIKHI, N. RAUFASTE and F. BONHOMME (2001): GENETIX 4.02, logiciel sous Windows TM pour la génétique des populations, Laboratoire Génome, Populations, Interactions: CNRS UMR. 5,000, Université de Montpellier II, Montpellier, France

BHATTACHARJEE, R., M. KOLSNIKOVA-ALLEN, P. O. AIKPOKPODION, S. TAIWO and I. INGELBRECHT (2004): An improved semi-automated rapid method of extracting genomic DNA for molecular marker analysis in cocoa, *Theobroma cacao* L. Plant Molecular Biology Reporter **22**: 435a–435h.

BOCCARA, M. and D. ZHANG (2006): Progress in resolving identity issues among the Parinari accessions held in Trinidad: the contribution of the collaborative USDA/CRU project. *In:* Proceedings of the CRU annual report 2005, Cacao Research Unit, The University of the West Indies, St. Augustine, Trinidad and Tobago.

CHEESMAN, E. E. (1944): Notes on nomenclature, classification and possible relationships of cacao populations. Tropical Agriculture **21**: 144–159.

CUATRECASAS, J. (1964): Cacao and its allies: A taxonomic revision of the genus *Theobroma*. Contributions to the U.S. Natural Herbarium **35**(6): 379–614.

CRYER, N. C., M. G. E. FENN, C. J. TURNBULL and M. J. WILKINSON (2006): Allelic Size Standards and Reference Genotypes to Unify International Cocoa (*Theobroma cacao* L.) Microsatellite Data. Genetic Resources and Crop Evolution **53**: 1643–1652.

EFRON, B. (1982): The jackknife, the bootstrap and other resampling plans. NSF-CBMS Regional Conference Series in Applied Mathematics, Monograph 38. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania.

EXCOFFIER, L., P. E. SMOUSE and J. M. QUATTRO (1992): Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131**: 479–491.

GOUDET, J. (2001): FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Updated from Goudet (1995) Available from http://www.unilch/izea/softwares/fstat.html.

GUILTINAN, M. J. (2007): Cacao. Biotechnology in Agriculture and Forestry, Vol. 60. *In:* Transgenic Crops V (ed. PUA, E. C., DAVEY, M. R.). Springer-Verlag Berlin Heidelberg

HAYATI, A., R. WICKNESWARI, I. MAIZURA and N. RAJANAIDU (2004): Genetic diversity in oil palm (*Elaeis guineensis* Jacq.) germplasm collections from Africa: implications for improvement and conservation of genetic resources. Theoretical and Applied Genetics **108**: 1274–1284.

HOLM, S. (1979): A simple sequentially rejective multiple test procedure. Scand J. Stat **6**: 65–70.

ICCO: INTERNATIONAL COCOA ORGANIZATION (2008): Quarterly Bulletin Cocoa Statistics. Vol. XXXIV. 2007/2008.

JACOBS, V. J., O. A. ATANDA and L. K. OPEKE (1971): Cacao breeding in Nigeria. *In:* CRIN Commemorative Publication – Progress in Tree Crops Research in Nigeria pp. 9–12.

JAKOBSSON, M. and N. A. ROSENBERG (2007): *CLUMPP*: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure, Bioinformatics **23**: 1801–1806.

JOHNSON, E. S., F. L. BEKELE, S. J. BROWN, Q. SONGE, D. ZHANG, L. MEINHARDT and R. J. SCHNELL (2009): Population structure and genetic diversity of the Trinitario cacao (*Theobroma cacao* L.) from Trinidad and Tobago. Crop Sci **49**: 564–572.

LANAUD, C., A. M. RISTERUCCI, I. PIERETTI, M. FALQUE, A. BOUET and P. J. L. LAGODA (1999): Isolation and characterization of microsatellites in *Theobroma cacao* L. Molecular Ecology **8**: 2141–2143.

LOCKWOOD, G. and M. M. O. GYAMFI (1979): The CRIG cocoa germplasm collection with notes on codes used in the breeding programme at Tafo and elsewhere. Cocoa Research Institute, Ghana Technical Bulletin No. 10 Tafo.

MARSHALL, T. (1998): CERVUS v.2.0. University of Edinburgh 1998–2001.

MILLER, M. P. (1997): Tools for Population Genetics analysis (TFPGA) version 1.3: a window program for the analysis of allozyme and molecular population genetic data©. Flagstaff Northern Arizona University, Flagstaff, AZ.

MOTAMAYOR, J.-C., A. M. RISTERUCCI, P. A. LOPEZ, C. F. ORTIZ, A. MORENO and C. LANAUD (2002): Cacao domestication I: the origin of the cacao cultivated by the Mayas. Heredity **89**: 380–386.

MOTAMAYOR, J.-C., A. M. RISTERUCCI, M. HEATH and C. LANAUD (2003): Cacao domestication II: progenitor germplasm of the Trinitario cacao cultivar. Heredity **91**: 322–330.

MOTAMAYOR, J.-C., P. LACHENAUD, J. WALLACE, G. LOOR, D. N. KUHN, J. S. BROWN and R. J. SCHNELL II. (2008): Geographic and genetic population differentiation of the Amazonian chocolate tree. PLoS One 3: e3311. doi:10.1007/s12042-008-9011-4

MOTILAL, L. and D. BUTLER (2003): Verification of identities in global cacao germplasm collections. Genet Resour Crop Evol **50**: 799–807.

MOTILAL, L., D. ZHANG, P. UMAHARAN, S. MISCHKE, V. MOOLEEDHAR and L. MEINHARDT (2009): The relic Criollo cacao in Belize – genetic diversity and relationship with Trinitario and other cacao clones held in the International Cocoa Genebank, Trinidad Plant Genetic Resources.

NEI, M. (1978): Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89**: 583–590.

OHTSUKA, A., M. WATANABE and T. YAHARA (2005): Inbreeding coefficients in six species of Ainsliaea and two species of Pertya (Asteraceae). Plant Syst. Evol. **251**: 143–151.

OJO, A. A., J. O. SANWO and E. B. ESAN (1991): Early results of evaluation of Nanay and Parinari *Theobroma cacao* double cross progenies in Nigeria. Field Crops Research **27**: 257–266.

OLATOYE, S. T. and E. B. ESAN (1992): Recent innovation in cacao genetic resources conservation in Nigeria. *In:* Proceedings of the International Workshop on Conservation, Characterization and Utilization of Cacao Genetic Resources in the 21st Century. Cocoa Research Unit, Port of Spain, Trinidad 13-17 September, 1992 pp. 281–291.

PEAKALL, R. and P. E. SMOUSE (2006): Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes **6**: 288–295.

POSNETTE, A. F. and J. M. A. TODD (1951): Virus diseases of cacao in West Africa. VIII. The search for virus resistant cacao. Annals of Applied Biology **38**: 785–800.

POWELL, W., M. MORGANTE, C. ANDRE, M. HANATEY, J. VOGEL, S. TINGEY and A. RAFALSKI (1996): The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding **2**: 225–238.

PRITCHARD, J. K., M. STEPHENS and P. DONNELLY (2000): Inference of population structure from multilocus genotype data. Genetics **155**: 945–959.

PUGH, T., O. FOUET, A. M. RISTERUCCI, P. BROTTIER, M. ABOULADZE, C. DELETREZ, B. COURTOIS, D. CLEMENT, P. LARMANDE, J. A. K. NGORAN and C. LANAUD (2004): A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers Theor. Appl Genet **108**: 1151–1161.

PURSEGLOVE, J. W. (1974): Tropical crops, dicotyledons. John Wiley & Sons, New York.

QUENOILLE, M. (1956): Notes on bias in estimation. Biometrika **43**: 253–260.

RICE, W. R. (1989): Analyzing tables of statistical tests. Evolution **43**: 223–225.

RICE, R. A. and R. GREENBERG (2000): Cacao cultivation and the conservation of biological diversity. Ambio **29**(3): 167–173.

SAUNDERS, J. A., S. MISCHKE, E. A. LEAMY and A. A. HEMEIDA (2004): Selection of international molecular standard for DNA fingerprinting. Theor. Appl. Genet. **110**: 41–47.

SCHNELL, R. J., C. T. OLANO, J. S. BROWN, A. W. MEEROW, C. CERVANTES-MARTINEZ, C. NAGAI and J.-C. MOTAMAYOR (2005): Retrospective determination of the parental population of superior cacao (*Theobroma cacao* L.) seedlings and association of microsatellite alleles with productivity. J. Amer. Soc. Hort. Sci. **130**: 181–190.

SERENO, M., P. P. ALBUQUERQUE, R. VENCOVSKY and A. FIGUERA (2006): Genetic diversity and natural population structure of cacao, *Theobroma cacao* L. from the Brazilian Amazon evaluated by microsatellite markers. Conservation Genetics **7**: 13–24.

SWEIGART, A., K. KAROLY, A. JONES and J. H. WILLIS (1999): The distribution of individual inbreeding coefficients and an pairwise relatedness in a population of *Mimulus guttatus*. Heredity **83**: 625–632.

TOXOPEUS, H. (1964): F_3 Amazon cacao in Nigeria. Rep. Cacao Res. Inst. Nigeria, 1963/64, pp.13–22.

WEIR, B. S. and C. C. COCKERHAM (1984): Estimating F-statistics for the analysis of population structure. Evolution **38**: 1358–1370.

WRIGHT, S. (1965): The interpretation of genetic structure by F-statistics with special regard to systems of mating. Evolution **19**: 355.

ZHANG, D., E. AREVALO, B. MISCHKE, L. ZUNIGA and A. BARRETO (2006a): Genetic diversity and population structure of cocoa, *Theobroma cacao* L. in the Hullaga and Ucayali valleys of Peru. Annals of Botany **98**: 647–655.

ZHANG, D., E. AREVALO, S. MISCHKE, R. GOENAGA, A. A. HEMEIDA and J. A. SAUNDERS (2006b): Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. Crop Science **46**: 2084–2092.

ZHANG, D., M. BOCCARA, L. MOTILAL, D. R. BUTLER, P. UMAHARAN, S. MISCHKE and L. MEINHARDT (2008): Microsatellite variation and population structure in the "refractario" cacao of Equador. Conserv Genet **9**: 327–337.

ZHANG, D., S. MISCHKE, E. S. JOHNSON, W. PHILLIPS-MORA and L. MEINHARDT (2009): Molecular characterization of an International cocoa collection using microsatellite markers. Tree Genetics & Genomes **5**: 1–10.

Cross-amplification and Characterization of Polymorphic Microsatellite Markers From *Acacia* (Senegalia) *mellifera* and *Acacia brevispica* to *Acacia senegal* (L.) Willd.

By S. F. OMONDI[1),3),*)], O. G. DANGASUK[2)], D. W. ODEE[3),4)], S. CAVERS[4)] and D. P. KHASA[5)]

(Received 6th May 2009)

[1)] Department of Forestry and Wood Science, Moi University, P.O. Box 1125 Eldoret, Kenya.

[2)] Department of Biological sciences, Moi University, P.O. Box 1125 Eldoret, Kenya.

[3)] Kenya Forestry Research Institute, P.O. Box 20412-00200 Nairobi, Kenya.

[4)] Centre for Ecology and Hydrology, Bush Estate, Penicuik, Midlothian, EH26 0QB, UK.

[5)] Centre for Forest Research and Canada Research Chair in Forest and Environmental Genomics, Université Laval, Sainte-Foy, Québec, Canada G1V 0A6.

[*)] Author of correspondence: STEPHEN F. OMONDI.
E-Mail: stephenf.omondi@gmail.com

Abstract

Seven polymorphic microsatellite markers isolated from *Acacia brevispica* and *Acacia mellifera* were successfully cross-amplified in *Acacia senegal*. The loci were surveyed for polymorphism using 30 samples. Allelic diversity ranged from 4 (*Ame02*, *Ab06* and *Ab18*) to 13 (*Ab26*) per locus. The expected heterozygosity (H_E) ranged from 0.543 (*Ame02*) to 0.868 (*Ab26*) while observed heterozygosity (H_O) ranged from 0.516 (*Ame05*) to 0.800 (*Ame03*). Cross-amplification of these loci represents a potential source of co-dominant markers and will be useful in the study of genetic diversity, structure,