Tree Genetics & Genomes (2018) 14: 32
https://doi.org/10.1007/s11295-018-1244-2

ORIGINAL ARTICLE

Genetic identity and diversity of Nigerian cacao genebank collections verified by single nucleotide polymorphisms (SNPs): a guide to field genebank management and utilization

Festus O. Olasupo[1] · Daniel B. Adewale[2] · Peter O. Aikpokpodion[3] · Anna A. Muyiwa[1] · Ranjana Bhattacharjee[4] · Osman A. Gutierrez[5] · Juan Carlos Motamayor[6] · Raymond J. Schnell[6] · Sona Ebai[7] · Dapeng Zhang[8]

Received: 30 September 2017 / Revised: 4 March 2018 / Accepted: 9 March 2018 / Published online: 21 March 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Nigeria is the sixth largest cacao producer in the world. Field performance and quality of cacao hybrid families is largely dependent on the genetic integrity of parental clones obtained in field genebank collections. However, information on the impact of mislabeling on seed garden output in Nigeria is lacking. Using 63 single nucleotide polymorphism (SNP) markers, we analyzed 1457 cacao trees sampled from seven major field genebank plots in Nigeria to assess the genetic integrity in Nigerian cacao germplasm. The procedure of multilocus matching with known reference clones revealed up to 78% mislabeling in recently introduced international germplasm. A high rate of mislabeling was also revealed in the West African local selections and breeding lines, using Bayesian assignment test. The problem of mislabeling has been attributed to errors from the sources of introduction, pre-planting labeling errors, and rootstocks overtaking budded scions due to poor field management. The analysis of genetic diversity revealed a good representation of the available cacao germplasm groups in Nigerian field genebanks, indicating that the genetic base of Nigeria cacao germplasm has been significantly widened through germplasm introductions. However, only a small proportion of the available germplasm in the genebank have been utilized for variety development. This study proved the utility of SNP markers for cleaning up the genebanks and reducing offtypes; thereby providing a strong basis for improving the accuracy and efficiency in cacao genebank management and breeding, as well as for mobilizing improved varieties to cacao farmers in Nigeria.

Keywords Chocolate tree · Genetic integrity · Mislabeling · Off-types · Tropical agriculture · West Africa

Communicated by V. Decroocq

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11295-018-1244-2) contains supplementary material, which is available to authorized users.

✉ Festus O. Olasupo
festusolasupo@gmail.com

[1] Crop Improvement Division, Cocoa Research Institute of Nigeria, Idi-Ayunre, Ibadan, Nigeria

[2] Department of Crop Science and Horticulture, Federal University, Oye, Ekiti State, Nigeria

[3] Department of Genetics and Biotechnology, University of Calabar, Calabar, Cross River State, Nigeria

[4] Bioscience Center, International Institute of Tropical Agriculture, Ibadan, Nigeria

[5] USDA-ARS Subtropical Horticulture Research Station, Miami, FL, USA

[6] Mars Incorporated, Miami, FL, USA

[7] World Cocoa Foundation, Accra, Ghana

[8] USDA-ARS, NEA, BARC, SPCL, Beltsville, MD 20705, USA

www.manaraa.com

Introduction

Cacao (*Theobroma cacao* L.), a fruit tree crop cultivated in the humid tropics of the world, is one of the major agricultural commodities in West Africa. Nigeria is currently the sixth largest producer of cacao in the world after Cote d'Ivoire, Ghana, Indonesia, Ecuador, and Cameroon, with an output of 200,000 MT in 2017 (ICCO 2017). It provides the largest non-oil foreign exchange earnings and contributes largely to Nigeria's gross domestic product (GDP). Cacao production mainly depends on smallholder farmers who cultivate not more than one to five hectares per household. Also, cacao provides livelihood and income for many rural communities whose production has depended on low productive and obsolete cacao cultivars. However, dry bean yield in Nigeria (of about 0.35 t/ha) still remains very low when compared to the crop genetic potential as well as to the yield obtainable in other producing countries. This problem is attributable to inadequate availability and accessibility of improved planting materials among other factors (Aikpokpodion et al. 2009).

Usually, crop breeding programs are initiated by the introduction of germplasm accessions from the crop's center of origin or places of high crop diversification. In 1910, Nigeria's Department of Agriculture started a formal introduction of cacao germplasm into the country, which was established at Moor Plantation Ibadan, following earlier introductions by Chief Squiss Ibaningo, missionaries, and slave traders. These initial introductions were all from the Amelonado population (Aikpokpodion 2007). This was followed by the introduction of Trinitario and Criollo selections from Trinidad and Ceylon, respectively, in 1933 (Jacob et al. 1971). With the establishment of the West African Cocoa Research Institute (WACRI) in Tafo, Ghana, Upper Amazon cacao germplasm from Pound's collection in Trinidad was introduced into Ghana in 1944, and subsequently, from Ghana into Nigeria after the establishment of Cocoa Research Institute of Nigeria (CRIN). The open-pollinated pods from the Trinidad introduction established in Ghana were also introduced into Nigeria to form an open-pollinated F_2 population from which open-pollinated F_3 seeds were generated and distributed to farmers in Nigeria as an intervention against cocoa swollen shoot virus (CSSV) epidemics (Aikpokpodion 2007).

Between 1931 and 1956, the first cacao breeding program of Nigeria conducted progeny trials, double hybrid crosses, and clonal trials, and this led to the selection of genotypes from West African Amelonado and local Trinitario families. These local selections represented the traditional cacao varieties in West Africa before the introduction of Upper Amazon materials, and were known as the "C clones" in the CRIN germplasm collection. Further germplasm introductions into Nigeria, as reported by Aikpokpodion (2007), included a large-scale introduction in the mid-1960s sponsored by the Cocoa Alliance initiated in Nigeria, which consisted of 313 clones and 701 seedling progenies derived from 350 intra-Nanay, intra-Parinari, intra-Iquitos, and inter-P (Pound's selections) crosses. These clonal and hybrid introductions from Trinidad constituted the "T clones" of CRIN germplasm collection. International clonal materials were also acquired from Costa Rica, Indonesia, Fernando Po, Kew Gardens (UK), Wageningen (the Netherlands), and Miami (USA) (Jacob et al. 1971). Between 1998 and 2004, 43 clones were introduced into Nigeria through an international initiative known as "Cocoa Germplasm Utilization and Conservation: A Global Approach," a project sponsored by the United Nations Common Fund for Commodity (CFC), through the supervision of the International Plant Genetic Resources Institute (IPGRI) as the executing agency (Eskes and Efron 2006). These materials were established at the international clone plot at CRIN, Ibadan.

Culturally, West African farmers grow cacao from seeds (beans) obtained from the pods, while the use of clones to establish plantations is a rare practice in the region. Therefore, seed gardens appeared to be the most efficient means of introducing improved cacao varieties to the farmers as planting materials (Adewale et al. 2016). The poor performance of the distributed cacao genetic materials in farmers' plots, leading to the low national productivity of 0.3–0.5 ton/ha, has necessitated the need for verification of the genetic integrity of the parental materials. In the Nigerian cacao germplasm collections, the occurrence of mislabeling and offtypes was first reported by Aikpokpodion et al. (2010). Mislabeling has been identified as one of the key factors contributing to the high rate of unwanted/unproductive progenies produced in seed gardens because this can seriously compromise the quality of seedlings that would be distributed to farmers (Cervantes-Martinez et al. 2006; Aikpokpodion et al. 2010). Moreover, when mislabeling occurs in biclonal and polyclonal seed gardens where hybrid pods are being generated from open-pollinated female parents presumed to be self-incompatible, this irregular situation further complicates breeders' efforts to assure the fidelity and accuracy of hybrid seed/pods produced in the seed gardens (Padi et al. 2015), thereby making the genetic identity of the hybrids unreliable. In addition to these issues, offtypes can hinder varietal performance when wrongly assigned for crossing, and they are unsuitable for scientific study. Offtypes have been observed to affect the accuracy of heritability estimations for black pod resistance in cacao (Adomako 2006). Takrama et al. (2014) attributed the problem of mislabeling in cacao germplasm collections and seed gardens to human errors and stated that this may occur during the introduction of genetic materials from points of clonal seedling collection to points of establishment.

Cacao germplasm collections in CRIN research stations are the main source of genetic diversity for breeders' varietal development, seed garden establishment, and consequently,

www.manaraa.com

hybrid seed (full-sib families) production for farmers. Therefore, verification of genetic integrity of germplasm accessions is critical to any successful breeding program and to germplasm management. The global cacao germplasm collections have been estimated to contain 15 to 44% mislabeled individuals (Motilal 2004; Motilal and Butler 2003; Sounigo et al. 2005; Takrama et al. 2005).

The use of molecular markers for cacao germplasm characterization, which started in the 1980s (Guiltinan et al. 2008), provides the opportunity to verify the identity of plant materials. Dominant molecular markers (RAPD) have been used for the identification of mislabeled accessions (Sounigo et al. 2005; Whitkus et al. 1998; Figueira et al. 1994). Lerceteau et al. (1997) and N'Goran et al. (2000) also verified mislabeled accessions using restriction fragment length polymorphism (RFLP) (codominant DNA markers). Saunders et al. (2004) developed a set of 15 simple sequence repeat (SSR) primers as international molecular standards for DNA fingerprinting of *T. cacao*. These SSR markers were selected for having high reproducibility and consistency and allowing the detection of mislabeled clones in various cacao collections (Zhang et al. 2006, 2009a, b). Aikpokpodion et al. (2010) used 12 SSRs (codominant DNA markers) to analyze the population structure and determine mislabeling and ancestry of 243 cacao accessions used for the breeding program in the germplasm collection of Nigeria. A number of accessions were reported by them as mislabeled or as offtype among the samples studied. Recent progress in the development of cacao genomic resources has led to the use of single nucleotide polymorphisms (SNPs) as markers for cacao DNA fingerprinting (Takrama et al. 2014; Ji et al. 2012; Kuhn et al. 2012; Livingstone et al. 2011; Motamayor et al. 2012). The impact of large-scale SNP fingerprinting on cacao germplasm management has been demonstrated in Ghana (Padi et al. 2015) and Brazil (DuVal et al. 2017).

This study, therefore, is aimed at three main objectives: first, to reveal the level of mislabeling in Nigeria's cacao germplasm collection; second, to verify the true genetic identity of the breeders' active clones used for the breeding program in Nigeria; and last, to correct initial mislabeling of the affected clones using the results from this study. This will help to ascertain uniformity within genotypes and certify genetic material for explicit utility for higher productivity. In addition, it will serve as the basis of expanding the capacity of the seed gardens established with true-to-type clones.

Materials and methods

Plant materials

The seven germplasm collections considered in this study were the most-utilized "C" breeding plots of Nigeria's (CRIN) cacao breeding programs. The collections varied in year of establishment, status, and purpose of establishment (Table 1). Cacao clonal materials collected from these collections include the international clones, T and C clones (Table 2). Five representative individual plants along a row of each clone were tagged in each plot, and fresh young cacao leaf samples were collected. The leaf samples collected from Ondo and Cross River States (which are located farther away from the molecular laboratory) were stored in freshly prepared 10-ml aliquots of NaCl-CTAB-azide solution (Bhattacharjee et al. 2004) in a 25-ml universal bottle. Collected samples were packed in boxes, transported, and delivered to the Bioscience Center of the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria. However, leaf samples from the CRIN headquarters in Ibadan were preserved on ice packs and delivered the same day to the Bioscience Center IITA.

Table 1 Description of seven Nigerian cacao field gene bank plots where leave samples were taken for the present study

Plot name	Location	Year of establishment	Status and purpose of establishment
IBIC	CRIN Ibadan	2003	International clones germplasm plot/budwood garden
IBLC	CRIN Ibadan	2004	Breeders' active clone collection/budwood garden
IBN	CRIN Ibadan	2004	Breeders' active clone collection/ budwood garden
AGS	Ondo State	1974	Polyclonal seed garden
OTU	Ondo State	1974	Biclonal seed garden
ADC	Cross Rivers State	2010	Polyclonal seed garden
AJAS	Cross Rivers State	2010	Biclonal seed garden

IBIC = international clone plot, CRIN, Ibadan, Nigeria; IBLC = local clone plot, CRIN, Ibadan, Nigeria; IBN = clonal plot, CRIN Nursery, Ibadan, Nigeria; AGS = polyclonal plot, Ago Store Ondo State, Nigeria; OTU = biclonal plot, Otu, Ondo State, Nigeria; ADC = polyclonal plot, ADC Camp, Ikom, Cross River State, Nigeria; AJAS = biclonal plot, Ajassor-Ikom, Cross River State, Nigeria

www.manaraa.com

Table 2 Name, origin, sources and plot location of sampled cacao accessions

Clone	Source/pedigree	Plot location	Number of trees sampled
Amaz 15–15	Chalmers Collection, Iquitos, Peru	IBIC	16
APA 4	Amazonico Palmira, Colombia	IBIC	23
BE 10	Belem, Brazil	IBIC	9
CATIE 1000	Pound 12 × Catongo ›››	IBIC	14
EET 59	Nacional × unknown	IBIC	15
GF 24	Amelonado × unknown	IBIC	25
ICS 1	Trintario × unknown	IBIC	46
IFC 5	Local selections from Cote d'Ivoire	IBIC	10
IMC 47	Iquitos, Pound collection	IBIC	17
MAN 15–2	Manaus, Brazil	IBIC	19
MOCORONGO	International clone	IBIC	9
MXC 67	Chiapas, Mexico	IBIC	13
N 38	Trinidad	IBIC, IBN, AGS	17
PA 107	Pound collection, Parinari, Peru	IBIC	23
PA 120	Pound collection, Parinari, Peru	IBIC	22
PA 150	Pound collection, Parinari, Peru	IBIC, IBN, ADC	54
PLAYA ALTA	Orinoco River, Venezuela	IBIC	13
POUND 7	Pound collection, Nanay River, Peru	IBIC, IBN, ADC	65
SCA 6	Pound Collection, Ucayali, River, Peru	IBIC	13
SPEC 54/1	Papuri River, Vaupes, Colombia	IBIC	13
UF 676	Trinitario × unknown, Costa Rica	IBIC	13
VENC 4	Amazonas, Venezuela	IBIC	9
T 9/15	Trinitario	AGS, IBLC	37
T 12/11	SCA 12 OP	IBLC	59
T 16/17	IMC 24 OP	IBLC	11
T 22/28	JA 3/11 OP	IBLC, IBN	7
T 30/13	PA 103 OP	IBLC	11
T 53/5	PA 37 OP	AGS, IBLC, IBN	46
T 53/8	PA 37 OP	IBIC	15
T 57/22	ICS 60 OP	ADC, AJAS, IBLC	42
T 65/7	PA 7 × IMC 47	ADC, AJAS, IBLC, IBN	114
T 65/35	PA 7 × IMC 47	IBLC	18
T 79/4	NA 32 × PA 7	IBN	4
T 82/27	NA 32 × PA 35	IBLC, IBN	25
T 85/799	IMC 60 × NA 34	IBIC	18
T 86/2	PA 35 × PA 7	ADC, AGS, IBLC	252
T 101/15	IMC 76 × NA 32	IBLC, IBN	15
C clones			
C 3	Amelonado × unknown	IBLC	4
C 18	Amelonado × unknown	AGS	36
C 25	Amelonado × unknown	AGS, out	50
C 27	Amelonado × unknown	AGS	43
C 30	Amelonado × unknown	IBLC	8
C 42	T60/887 (PA 7 × NA 32)	IBN	7
C 60	T9/21 (Local Trinitario)	IBLC	6
C 67	T79/501 (NA 32 × PA 7)	ADC, AJAS, IBLC, IBN	149
C 74	T63/971 (PA 35 × NA 32)	AGS	27
C 75	T63/967 (PA 35 × NA 32)	AGS, out	20
C 77	T85/799 (IMC 60 × NA 34)	AGS, AJAS, IBIC, IBLC, IBN	166

IBIC = international clone plot, CRIN, Ibadan, Nigeria; IBLC = local clone plot, CRIN, Ibadan, Nigeria; IBN = clonal plot, CRIN Nursery, Ibadan, Nigeria; AGS = polyclonal plot, Ago Store, Ondo State, Nigeria; OTU = Biclonal plot, Otu, Ondo State, Nigeria; ADC = polyclonal plot, ADC Camp, Ikom, Cross River State, Nigeria; AJAS = biclonal plot, Ajassor-Ikom, Cross River State, Nigeria

DNA extraction

The leaf tissues stored at 4 °C in the NaCl-CTAB-azide buffer were removed with the help of forceps and placed inside 1.2-ml propylene strip tubes with strip caps (Marsh Biomarket, USA, available as 12 × 8-well strips), that contained two pre-chilled 4-mm chrome-plated grinding steel ball bearings. The balls were pre-dispensed using an automatic dispenser.

The DNA extraction and purification of 2000 samples was done in batches of 192 samples per day, with CTAB buffer as described in Bhattacharjee et al. (2004). Extracted DNA was re-suspended in 100-μl low-salt TE and stored at 4 °C for

www.manaraa.com

further use. Aliquots of 3 μl freshly extracted genomic DNA were electrophoresed on 0.8% agarose gel, stained with ethidium bromide, and visualized under a UV transilluminator to assess the quality of DNA. The quantity of DNA was estimated with 1 μl of freshly extracted DNA on a NanoDrop spectrophotometer.

SNP genotyping

A total of 89 SNPs were used to fingerprint the cacao trees. These SNP loci were originally identified from expressed sequence tags (ESTs) of a wide range of cacao tissue and organs that displayed differences in the transcriptome (Argout et al. 2008, 2011; Allegre et al. 2012), and 1536 SNPs were screened using Illumina's GoldenGate Assay (Michel Boccara, unpublished data). The 89 SNPs were selected from these based on call rate, representativeness across the ten chromosomes and heterozygosity, and these have also been applied in previously reported cacao research (Fang et al. 2014; Lukman et al. 2014; Cosme-Reyes et al. 2016). A list of the SNPs and their flanking sequences was presented in Supplementary Table 1. SNP Genotyping was performed using KASP™ assays from LGC Genomics (http://www.lgcgroup.com/kasp). KASP genotyping assays are based on competitive, allele-specific PCR and enable high-throughput genotyping of specific SNPs and InDels. Once the KASP reaction was complete, the resulting fluorescence was measured on a BMG PHERAstar plate reader. The raw data were analyzed using LGC's proprietary Kraken™ software and scored on a Cartesian plot, also known as a cluster plot, in order to assign a genotype to each DNA sample.

Data analysis

Raw data was imported and organized in Microsoft Excel 2007 for each SNP locus and sample call. Descriptive statistics measuring informativeness and quality of the successfully amplified SNP loci were calculated based on the introduced international clones, using the GenAlEx 6.5 program (Peakall and Smouse 2006, 2012). The key descriptive statistics included Shannon's information index (I), observed heterozygosity (H_O), expected heterozygosity (H_E), minor allele frequency (MAF), and inbreeding coefficient (FIS).

Two approaches were used to identify mislabeling (offtypes) in the Nigerian germplasm. The first approach used multilocus matching to directly compare original reference cacao trees with the Nigerian trees. This approach was suitable for introduced international clones, which have original reference trees maintained in the International Cacao Genebanks in Trinidad and Costa Rica or have leaf samples in the collection at USDA Beltsville Agricultural Research Center. The reference SNP profiles were generated in a Fluidigm 48.48 nanofluidic array and the detailed genotyping procedure was reported by Fang et al. 2014. These reference profiles were also deposited in International Cacao Germplasm Database (ICGD; http://www.icgd.rdg.ac.uk/, accessed September 29, 2017). The majority of these samples were also verified previously by SSR fingerprinting (Motamayor et al. 2008; Zhang et al. 2009a, b; Johnson et al. 2009). Pairwise multilocus matching was applied between each individual tree and the reference trees from the international germplasm collections, using the computer program GenAlEx 6.5 (Peakall and Smouse 2006, 2012). Accessions with same names as the reference trees, but with non-matching SNP patterns, were declared offtypes. For the multilocus matching, the option to consider missing data was used. Discriminating power of the SNP loci was computed using the probability of identity (PID) (Waits et al. 2001) option implemented in the same computer program. The second approach dealt with Nigerian cacao accessions that were not introduced as international clones. For these trees, there were no references available in the international genebanks. Therefore, the true-to-typeness of these trees had to be inferred indirectly using Bayesian assignment test (Pritchard et al. 2000), based on their recorded pedigree or other passport data. These accessions included two categories. The first category are the T clones that were hybrid families introduced into West Africa in 1944 from Trinidad (Posnette 1986; Toxopeus 1964, 1985). The second group of trees that do not have reference profiles are the C clones, which are the selected West African Amelonado and local Trinitario selections (C1–C39) or local selections of Upper Amazon parentage (C41 and above) (Lockwood and Gyamfi 1979). For these two groups of trees, the assignment test was applied to infer their hidden membership to a known cacao population or germplasm group, using the STRUCTURE software program (Falush et al. 2003; Pritchard et al. 2000). SNP profiles of 140 reference accessions were included in the analysis. These 140 reference accessions were taken from seven known cacao germplasm populations, which covered all the recorded parentage background for the Nigerian cacao germplasm (Aikpokpodion 2012; Lockwood and Gyamfi 1979; Posnette 1986; Toxopeus 1985). Among the seven germplasm populations, Criollo and Amelonado represent the parentage background of local germplasm in West Africa before the introduction of Upper Amazon Forasteros in 1944, whereas Scavina (SCA), Iquitos Mixed Calabacillo (IMC), Nacional, Nanay (NA), and Parinari (PA) represent the background of the Pound collection (Posnette 1944; Aikpokpodion 2012; Zhang and Motilal 2016). Molecular classification of these germplasm groups have been reported by Motamayor et al. (2008) and Zhang et al. (2006, 2009a). The full list of the 140 reference accessions was presented in Supplemental Table 2. Detailed information of SNP profiles of these accessions, as well as the genotyping protocol, has been reported by Cosme-Reyes et al. (2016). Out of the 96 SNP markers in the reference populations, there were 37 overlapped with the 63 SNPs

www.manaraa.com

generated by the present study for the Nigerian trees. These 37 common SNPs were used to run the Bayesian assignment test.

The assignment test was conducted in two steps. Firstly, allele frequency of the 37 SNPs was computed for each of the seven reference populations. Based on the resultant allele frequency, the size of each reference population was brought to 200 using the SIMULATION procedure implemented in the computer program ONCOR (Kalinowski et al. 2007). The simulated reference populations were then analyzed together with the Nigerian tree samples using the computer program STRUCTURE (Pritchard et al. 2000). Ten independent runs were assessed for $K = 7$. Of the 10 independent runs, the one with the highest Ln Pr ($X|K$) value (log probability or log likelihood) was chosen and represented as a bar plot. The result of the assignment test was compared with the multilocus matching analysis to ensure consistency.

To assess genetic diversity among the core breeders' collections as revealed by STRUCTURE, a tridimensional graph was generated by the first three principal component axes using the procedure PRINCOMP and g3d, respectively, in SAS-version 9.3 (SAS 2011).

Results

SNP fingerprinting

From the 89 SNP panels chosen to fingerprint the Nigerian cacao germplasm, 63 SNPs were polymorphic and generated high call rates (> 75%) across all tested cacao samples. The failures of 26 SNPs were likely due to sequence complexities and those of genotypes were due to DNA quality and problems with PCR amplification. For the 63 SNP markers that generated consistent results, a total of 1457 samples (72.9%) had an SNP call rate higher than 75% and were used for further analysis.

Summary statistics for these 63 SNPs was computed based on the 462 trees of international clones and the result was presented in Supplemental Table 3. The observed heterozygosity (H_o) ranged from 0.057 (for the TcSNP1392 locus) to 0.835 (for TcSNP277 locus) with an overall H_{ob} average of 0.248. H_E ranged from 0.093 for the TcSNP1392 locus to 0.501 for several loci and averaged 0.413. Polymorphic information content ranged from 0.089 for TcSNP1392 to 0.375 for several loci and averaged 0.321. Minor allele frequency ranged from 0.049 for TcSNP1392 to 0.500 for the TcSNP645 and TcSNP723 loci and averaged 0.337.

Identified mislabeling in introduced international clones by multilocus matching with reference clones

Out of the 63 SNP markers, there were 28 that overlapped with USDA's reference SNP profiles generated by the Fluidigm genotyping panel of 48 markers. Therefore, these 28 SNPs were used to compare the 462 trees, representing 22 introduced international clones, with the reference trees. An example of the multilocus matching was presented in Table 3 and the full result of the identified offtypes in all tested international clones was presented in Supplementary Table 4. Mislabeling occurred both within and among stations. Among the screened international clones, the mislabeling rate ranged from 0.55 (Playa Alta) to 1.00 (APA 4, MOCORONGO, PA 107, SPEC 54/1, and VENC 4). These trees were defined as offtype or homonymous mislabeling because they shared the same name with the reference tree but differed in multilocus SNP profiles. With all 28 loci considered, the combined probability of identity was in the order of 10^{-6}, demonstrating that the 28 SNPs were sufficient to verify the true-to-type trees (Fig. 1).

Validation of the simulated reference populations

At $K = 7$, the Bayesian clustering analysis could not fully separate the simulated population into seven respective genetic groups: IMC, NA, SCA, PA, Nacional, Amelonado, and Criollo. The populations of IMC and NA could not be distinguished from each other, due to the relatively small number of SNP markers (37) used in the analysis. Therefore, the clustering analysis was re-run at $K = 6$, by treating IMC and NA as one single reference group. The correctly identified international clones from the Nigerian cacao core breeders' collection were included in the analysis.

The assignment result of the international clones at $K = 6$ was fully consistent with the previously reported classification of these clones by SSR analysis (Table 4) (Zhang et al. 2009a; Motamayor et al. 2008). Accessions from reference germplasm group Nacional, Refractario (Nacional hybrid), Parinari, Scavina, Amelonado, and Trinitario were all properly assigned to their recorded groups, respectively. However, Nanay and Criollo were not observed among the materials from introduced international clone plot. The assignment of the hybrid clones such as EET 59, ICS 1, MXC 67, and UF 676 was consistent with the recorded pedigree and large proportion of admixture was found between clusters (Table 4). The result thus proved the validity of using the simulated reference populations to examine the recorded pedigree of other Nigerian germplasm using assignment test.

Identified mislabeling in the T clones using assignment test

Five hundred and sixty trees, representing 15 T clones and GF 24 were examined in four plots (Fig. 2). There were no reference SNP profiles for these accessions; however, their identities were examined based on their recorded pedigree by assignment test. The level of mislabeling varied among the

www.manaraa.com

Table 3 Examples of DNA fingerprints based on multilocus matching of 28 SNPs between original references and Nigerian cacao collection (showing truncated profiles)

Genotype	Sample ID	Assessment	Tc32	Tc144	Tc193	Tc230	Tc242	Tc372	Tc529	Tc560	Tc591
AMAZ 15–15	ICGT, Trinidad	Reference	TT	AC	AC	AA	CT	AA	AC	GG	AA
AMAZ 15–15	IBAM1515-04_NIG132	True to type	TT	AC	AC	AA	CT	AA	AC	GG	AA
AMAZ 15–15	IBAM1515-02_NIG1918	Offtype	TT	AC	AA	AA	CC	AA	AC	GG	AC
AMAZ 15–15	IBAM1515-01_NIG1901	Offtype	TT	CC	AA	AG	CC	AA	CC	GT	AA
AMAZ 15–15	IBAM1515-03_NIG1876	Offtype	AT	CC	AA	AG	TT	AA	AC	GT	00
PA 150	ICGT, Trinidad	Reference	AA	AC	AA	GG	CC	AA	AC	TT	AC
PA 150	IBPA15005B_NIG1947	True to type	AA	AC	AA	GG	CC	AA	AC	TT	AC
PA 150	IBPA15005_NIG387	True to type	AA	AC	AA	GG	CC	AA	AC	TT	AC
PA 150	IBPA15009_NIG391	True to type	AA	AC	AA	GG	CC	AA	AC	TT	AC
PA 150	IBPA150-12B_NIG208	Offtype	AA	AC	AA	GG	TT	AT	CC	GT	AC
PA 150	IBPA15022B_NIG1955	Offtype	TT	AC	AA	AG	CC	AA	CC	TT	CC
PA 150	ADPA15013_NIG723	Offtype	AA	AC	00	AG	CT	AT	AC	GG	AA
Playa Alta	CATIE, Costa Rica	Reference	TT	CC	AA	AG	CC	AT	CC	TT	AC
Playa Alta	IBPLA04_NIG392	True to type	TT	CC	AA	AG	CC	AT	CC	TT	AC
Playa Alta	IBPLA03_NIG393	True to type	TT	CC	AA	AG	CC	AT	CC	TT	AC
Playa Alta	IBPLAY-12_NIG450	True to type	TT	CC	AA	AG	CC	AT	CC	TT	AC
Playa Alta	IBPLA11_NIG1873	Offtype	AT	AA	AA	AA	CC	AA	AA	TT	AC
Playa Alta	IBPLA010_NIG1897	Offtype	AT	CC	AA	AG	TT	AA	AC	GT	AC
Playa Alta	IBPLA09_NIG1898	Offtype	AT	AA	AC	AG	CT	AT	AC	GT	CC
SCA 6	ICGT, Trinidad	Reference	AA	AC	CC	AA	CT	AA	AA	GG	CC
SCA 6	IBSCA605_NIG383	True to type	AA	AC	CC	AA	CT	AA	AA	GG	CC
SCA 6	IBSCA604_NIG388	True to type	AA	AC	CC	AA	CT	AA	AA	GG	CC
SCA 6	IBSCA602_NIG458	True to type	AA	AC	CC	AA	CT	AA	AA	GG	CC
SCA 6	IBSCA6-08_NIG196	Offtype	AA	AC	CC	AA	CT	TT	AC	TT	AC
SCA 6	IBSCA6-11_NIG1882	Offtype	TT	CC	AA	AG	CC	AA	CC	GT	AA
SCA 6	IBSCA607_NIG1891	Offtype	TT	CC	AA	AG	TT	AA	AA	GG	CC
SCA 6	IBSCA6-06_NIG1944	Offtype	TT	CC	AA	AG	CT	AA	AA	GT	AC

Genotype	Tc619	Tc645	Tc723	Tc872	Tc878	Tc917	Tc929	Tc994	Tc998	Tc1060	Tc1062
AMAZ 15–15	CT	GG	TT	CG	CC	CC	GG	CC	AA	CT	GG
AMAZ 15–15	CT	GG	00	CG	CC	CC	GG	CC	AA	CT	GG
AMAZ 15–15	TT	AG	GG	CG	CC	00	CG	CC	AG	CT	GG
AMAZ 15–15	CC	AA	GT	CG	CG	CT	GG	CC	AG	CC	AG
AMAZ 15–15	CC	AA	GG	GG	CC	CT	GG	CT	AA	CT	AG
PA 150	TT	AA	GG	GG	CC	CT	CC	CC	AG	CC	GG
PA 150	TT	AA	GG	GG	CC	CT	CC	CC	AG	CC	GG
PA 150	TT	AA	GG	GG	CC	CT	CC	CC	AG	CC	00
PA 150	TT	AA	GG	GG	CC	CT	CC	CC	AG	CC	GG
PA 150	CT	AA	GG	CC	CC	CT	GG	TT	AG	CC	GG
PA 150	TT	AG	GG	CG	CC	00	GG	CT	AG	CC	AG
PA 150	TT	AG	GG	CG	CC	CC	GG	CT	AG	CT	AG
Playa Alta	TT	GG	GG	CG	CC	TT	GG	CC	AA	CT	AA
Playa Alta	TT	GG	GG	CG	CC	TT	GG	CC	AA	CT	AA
Playa Alta	TT	GG	GG	CG	CC	TT	GG	CC	AA	CT	AA
Playa Alta	TT	GG	GG	CG	CC	TT	GG	CC	AA	CT	AA
Playa Alta	TT	GG	GG	CG	00	CT	CG	CT	AA	CC	GG

www.manaraa.com

Table 3 (continued)

Genotype	Tc619	Tc645	Tc723	Tc872	Tc878	Tc917	Tc929	Tc994	Tc998	Tc1060	Tc1062
Playa Alta	CC	AA	GG	GG	CC	CT	GG	CT	AA	CT	AG
Playa Alta	TT	GG	GT	CG	CC	CT	GG	00	AA	CC	AG
SCA 6	TT	AA	GG	CC	CG	CC	CC	CC	AA	CT	GG
SCA 6	TT	AA	GG	CC	CG	CC	CC	CC	AA	CT	GG
SCA 6	TT	AA	GG	CC	CG	CC	CC	CC	AA	CT	00
SCA 6	TT	AA	GG	CC	CG	CC	CC	CC	AA	CT	GG
SCA 6	TT	GG	GT	GG	CG	TT	GG	TT	AA	CT	AG
SCA 6	CC	AA	GT	CG	CG	CT	GG	CC	AG	CC	AG
SCA 6	TT	AA	GG	CC	CC	CT	GG	CT	AA	CC	AG
SCA 6	CT	AA	GG	CC	CC	CC	GG	TT	AA	CC	AG

clones and across plots in both the budwood and seed gardens (Table 5). Rate of mislabeling/offtype ranged from 15% (T65/35) to 100% (T16/17, T65/7, T79/4, and T82/27) in the budwood gardens, whereas a range of 10% (T86/2) to 73% (T65/7) was observed among the clones in the seed gardens. Two possible genotypes were identified based on the parentage of the following clones: T9/15 (Trinitario and Trinitario × Nacional hybrid), T22/28 (Amelonado hybrid and Trinitario), and T30/13 (PA × IMC and PA × SCA). Correct trees of each clone were, respectively, identified as T9/15-Type1 and T9/15-Type2, T22/28-Type1 and T22/28-Type2, and T30/13-Type1 and T30/13-Type2. Similarly, 67% mislabeling was observed in GF24, while the pedigree of correctly identified trees was assigned five new genotypes as: GF24-Type1, GF24-Type2, GF24-Type3, GF24-Type4, and GF24-Type5 based on population structure (Table 6).

Identified mislabeling in local selections by assignment test

An assignment test was performed on 529 trees representing 11 clones belonging to the C clones. The analysis presented a varied level of mislabeling and offtypes among the clones and across plots in both the budwood and seed gardens (Table 7). Labeling errors ranged from 0% (C67) to 100% (C25, C27, C60, and C74) among the cacao C clones sampled from the Nigerian germplasm collections. The clone labeled C18 at AGS was observed to be basically of IMC background. A correctly identified tree of C18 was found to have Trinitario × Nacional hybrid pedigree. High rates of mislabeling were also found associated with clones C25 (100%), C27 (100%), C74 (100%), and C77 (80%) at AGS. A similar trend was observed in sampled clones from other seed gardens (OTU, ADC, and AJAS) except for C75 at AGS (39%) and OTU (28%). Generally, lower rates (0–50%) labeling errors were observed in all the breeders' active clone collection plots. However at IBLC labeling errors were 68% for C77, 86% for C30, and 100% for C60. Interestingly, the predominant contaminant of mislabeled C18 at AGS seed garden plot was of IMC ancestry, which was C77 (Table 7) as revealed by an assignment test. Similar observations were made at AGS and AJAS seed garden plots in C27 and C67, respectively, where mislabeled samples were confirmed to be of IMC ancestry (i.e., C77). The clones C74 and C75 were wrongly labeled as C77 at AGS seed garden plot. Also at OTU seed garden, C75 sampled trees were confirmed to be C77 whereas, at AJAS and IBLC, C75 were mislabeled as C77. The main contaminant of mislabeled C67 at ADC seed garden plot had SCA × PA parentage, which happened to be T12/11, while IMC × PA (i.e., C75) were confirmed as the predominant contaminant of C25 at AGS and C77 at AJAS and IBLC plot (Table 7).

www.manaraa.com

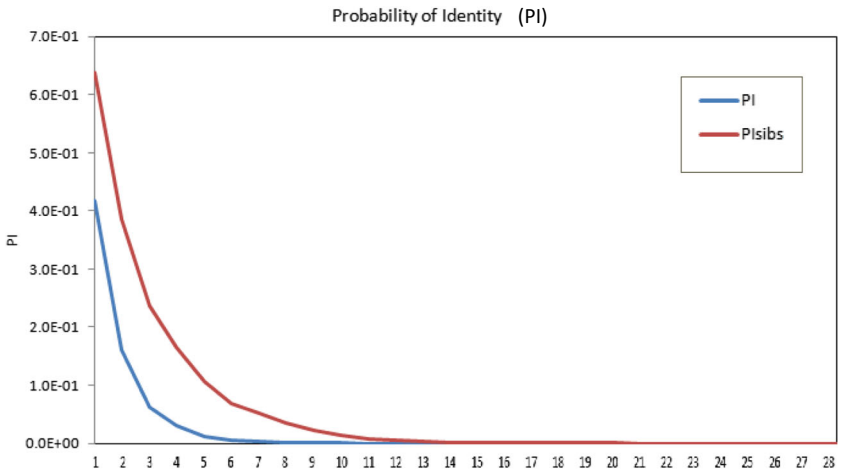

Fig. 1 Probabilities of identity (PID) based on 28 SNP markers used for offtype identification in Nigerian cacao germplasm

Genetic diversity of the Nigerian cacao germplasm collections

The first three principal component (PC) axes explained 78% as the total similarity among the correctly labeled clones in the active breeders' collection (Fig. 3). Seven basic groups were visible in the figure with the exception of C30. Cluster I had 14 clones (BE10, ICS1, IFC5, N38, Playa Alta, UF676, C3, C18, C25, T9/15-Type1, T9/15-Type2, T22/28-Type2, T57/22, and GF24-Type1), which were mainly Trinitarios. Five genotypes (EET59, MXC67, C27, T22/28-Type1, and GF24-Type5) were observed among the clones in cluster II. Two clones each clustered in groups III (T16/17 and GF24-Type2) and IV (GF24-Type3 and GF24-Type4). In cluster V, five Upper Amazon materials (PA150, SCA6, T12/11, T30/13-Type2, and T86/2) were visible while nine clones (C42, C67, C75, T30/13-Type1, T53/5, T53/8, T65/7, T65/35, and T82/27) were observed in cluster VI. Six clones (Amaz15-15, IMC47, P7, C77, T85/799, and T101/15), which are Upper Amazon hybrids, clustered in group VII. Based on the Gower distance similarity coefficient (table not provided), very high mean similarities (0.98, 0.97, and 0.96) were obtained among clones within clusters VI, VII, and III. The similarity coefficient for clones in cluster I was 0.9 while 0.79 was observed in cluster IV.

Discussion

Our result revealed fingerprinting with SNP markers to be a highly efficient tool to clean up labeling errors in seed gardens and germplasm plots. Consistent with Livingstone et al. (2012), SNPs also have high-throughput efficiency like SSRs in determining offtypes among clonal populations.

Table 4 Assigned memberships (Q value) of correctly identified international clones in Nigeria field genebank using seven simulated reference populations at $K = 6$

Clone	Scavina	IMC	Parinari	Nacional	Amelonado	Criollo	Confirmed genetic group
Amaz 15–15	0.013	0.956	0.021	0.004	0.004	0.002	IMC
BE 10	0.002	0.007	0.005	0.002	0.982	0.002	Amelonado
EET 59	0.028	0.016	0.017	0.428	0.500	0.011	Nacional hybrid (or Refractario)
ICS 1	0.002	0.004	0.003	0.002	0.506	0.482	Trinitario
IFC 5	0.009	0.008	0.007	0.003	0.799	0.173	Amelonado
IMC 47	0.074	0.784	0.007	0.124	0.005	0.006	IMC
MXC 67	0.013	0.010	0.009	0.435	0.515	0.018	Nacional hybrid
N 38	0.005	0.008	0.013	0.002	0.964	0.008	Amelonado
P 7	0.004	0.966	0.005	0.002	0.020	0.002	IMC
PA 150	0.019	0.013	0.847	0.004	0.115	0.002	Parinari
Playa Alta	0.014	0.041	0.053	0.007	0.618	0.266	Trinitario
SCA 6	0.928	0.005	0.010	0.051	0.004	0.002	Scavina
UF 676	0.006	0.009	0.008	0.004	0.509	0.465	Trinitario

www.manaraa.com

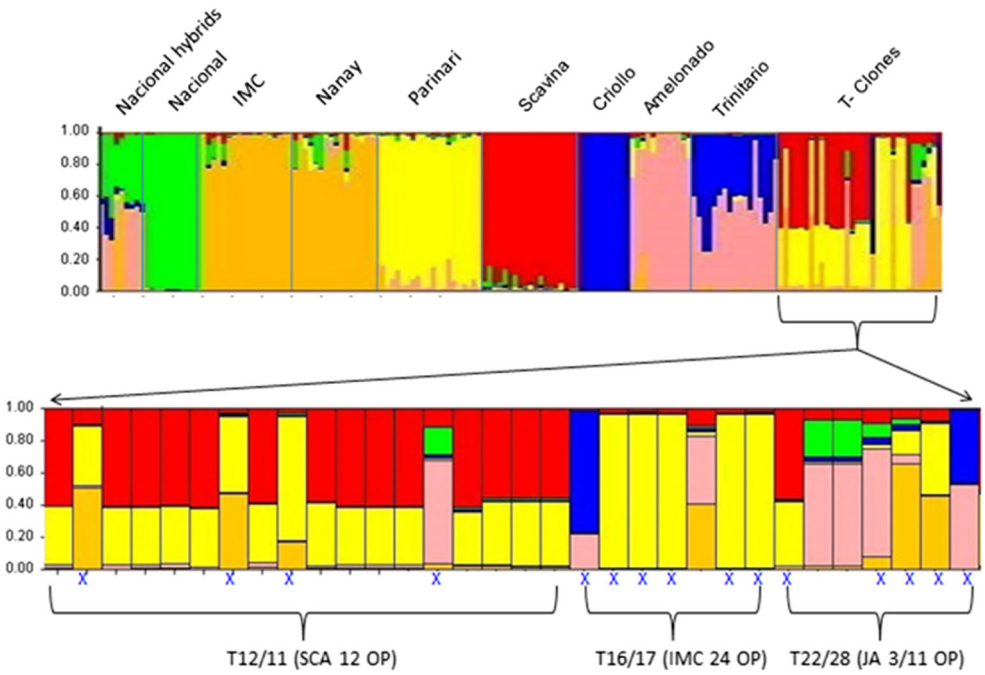

Fig. 2 An example of detecting offtype trees in three T clones of cacao using assignment test. From STRUCTURE, *K* is the number of genetic clusters that exist in the overall sample of individuals. Each vertical line represents one individual multilocus genotype. Individuals with multiple colors have admixed genotypes from multiple clusters. Each color represents the most likely ancestry of the cluster from which the genotype or partial genotype was derived. Clusters of individuals are represented by colors. The identified offtype trees were marked by "X"

The clonal materials studied here represent an important part of the cacao germplasm core breeders' collections as well as seed garden materials that have been used to produce planting materials for distribution to farmers in Nigeria. Detection of mislabeling helps us to resolve one of the fundamental issues of cacao breeding output, in which the use of wrongly identified materials as parents to produce hybrid varieties (full-sib families) leads to poor performance of the trees in farmers' field. The present study reaffirms the work of Aikpokpodion et al. (2010), which employed the use of microsatellite markers (SSRs) for the verification and identification of mislabeling in the Nigerian cacao collection. Although application of microsatellite markers has greatly increased the efficiency and capacity for cacao fingerprinting and resulted in a wide application of cacao genotype identification in the recent past (Aikpokpodion et al. 2005; Efombagan et al. 2008; Zhang et al. 2009b; Motilal et al. 2010), combining SSR data from different laboratories has been difficult, due to the different genotyping platforms. The same allele may be binned differently, causing confusion and leading to false conclusions (Takrama et al. 2014). Compared with SSR genotyping, the assays of SNPs were done without requiring separation of DNA by gel electrophoresis. The diallelic nature of SNPs resulted in a lower error rate in allele calling and it is cost-effective. The use of SNP as markers for cacao DNA fingerprinting and offtype identification in this study has demonstrated improved accuracy and efficiency. SNPs profiles of international reference trees have been used for clones' genetic identity and offtype detection of cacao field genebanks without ambiguity.

High levels of labeling error observed among the international clone materials is worrisome and implies that an urgent corrective measure is necessary. Although the plot does not serve as an active clone collection for Nigerian cacao breeding, the degree and spread of mislabeling among the clones are consistent and could have emanated mainly from human actions, mishandling, and misidentification. Furthermore, human mislabeling error at the nursery level may not be ruled out as reported by Padi et al. (2015). The implication of labeling error of the introduced international clones is huge since it will lead to the wrong prediction of performance of a genotype in Nigeria's growing conditions. In addition, these errors are multiplied when budwoods from mislabeled clones are used for the establishment of new seed gardens. A good example is the multiplied effect of 58% mislabeling of PA150 at IBIC. This error may have translated into higher (100%) error observed at ADC seed garden. However, the low level of mislabeling and uniformity observed by this study in a few breeders' active clone collections in IBN is noteworthy. Moreover, this suggests the need to confirm the identity of remaining clones in the same plot.

Observed contaminants of the C clones in most of the seed gardens could be used as indicators of the origin of labeling errors encountered in different locations. Almost all the C18 sampled from AGS plot in this study were observed to be basically of IMC ancestry instead of local Amelonado or local Trinitario. A similar observation had been reported by Aikpokpodion et al. (2010); some local clones shared allelic profiles with Upper Amazon's Nanay, Parinari, and Iquitos Mixed Calabacillo primary populations. Therefore, they

www.manaraa.com

Table 5 Level of mislabeling in the T clones of cacao breeders' plots in Nigeria

Clone label in plot	Plot		Number of trees sampled	Number of mislabeled trees	Identity of correct clone	Identity of contaminant[a]	Unidentified contaminant (%)
	Budwood garden	Seed garden					
T9/15*		AGS	10	6	Type 1 (Trinitario), Type 2 (Trinitario × Nacional hybrid)	IMC × PA (2), IMC (1), PA × Amelonado (1)	20
	IBLC		16	10	Type 1 (Trinitario), Type 2 (Trinitario × Nacional hybrid)	IMC × PA (2), SCA × PA (4), SCA × Amelonado (2)	13
T12/11	IBLC		46	23	SCA × PA	IMC × PA (8), IMC (3), IMC × Amelonado (2), PA (1), PA × Amelonado (1), PA hybrid (1), Trinitario (1), SCA × Amelonado (1)	2
T16/17	IBLC		11	11	–	PA (5), IMC × Amelonado (1), Criollo hybrid (1)	36
T22/28*	IBLC		3	2	Type 2 (Trinitario)	IMC (1), IMC × PA (1)	0
	IBN		3	1	Type 1 (Amelonado hybrid)		33
T30/13*	IBLC		10	5	Type 1 (PA × IMC), Type 2 (PA × SCA)	Trinitario (3), SCA × Amelonado (1), SCA × IMC (1)	0
T53/5		AGS	16	8	PA × IMC	IMC (7), PA (1)	0
	IBLC		20	8	PA × IMC	IMC (5), PA (1), PA × SCA (1), IMC × Amelonado (1)	0
	IBN		3	1	PA × IMC	IMC (1)	0
T53/8	IBIC		11	9	PA × IMC	IMC (2), PA (1), PA × SCA (1), PA × Amelonado (1), IMC × Amelonado (1), Trinitario (1), Trinitario × Nacional hybrid (1)	9
T57/22		ADC	14	6	Trinitario × Nacional hybrid, Trinitario	PA (4), Amelonado (1), PA × IMC (1)	0
		AJAS	7	4	Trinitario	PA (2), Amelonado (1), PA × Amelonado (1)	0
	IBLC		4	1	Trinitario	IMC (1)	0
T65/7		ADC	11	8	PA × IMC	PA (3), PA × SCA (3), IMC (1), Trinitario (1)	0
		AJAS	68	35	PA × IMC	IMC (32), PA (3)	0
	IBLC		17	11	PA × IMC	IMC (6), PA (3), PA × SCA (1), IMC × Nacional (1)	0
	IBN		4	4	–	PA (2), IMC (2)	0
T65/35	IBLC		13	2	PA × IMC	IMC (1), PA (1)	0
T79/4	IBN		4	4	–	IMC × Amelonado (4)	0
T82/27	IBLC		18	15	IMC × PA	IMC (7), SCA × PA (3), IMC × Amelonado (1)	22
	IBN		6	6	–	PA (4), SCA × PA (1), Trinitario × Nacional hybrid (1)	0
T85/799	IBIC		13	6	IMC	IMC × PA (2), SCA (1), IMC × SCA (1)	23
T86/2		ADC	89	9	PA	IMC (3), Trinitario × Nacional hybrid (2), PA × SCA (1), Trinitario × Nacional (1)	2
		AGS	108	37	PA	IMC (16), PA × IMC (8), Trinitario × NA hybrid (6), Amelonado (3), Trinitario (1)	2
	IBLC		13	5	PA	PA × IMC (3), PA × SCA (1)	8
T101/15	IBLC		9	4	IMC	Trinitario (1), PA (1)	22

www.manaraa.com

Table 5 (continued)

Clone label in plot	Plot		Number of trees sampled	Number of mislabeled trees	Identity of correct clone	Identity of contaminant[a]	Unidentified contaminant (%)
	Budwood garden	Seed garden					
	IBN		4	2	IMC	PA (2)	0
GF 24*	IBIC		14	9	Type 1 (Trinitario), Type 2 (Amelonado × IMC), Type 3 (Amelonado × SCA), Type 4 (Amelonado × PA), Type 5 (Nacional hybrid)	IMC (4), SCA (1)	43

IBIC = international clone plot, CRIN, Ibadan, Nigeria; IBN = clonal plot, CRIN Nursery, Ibadan, Nigeria; AGS = polyclonal plot, Ago Store, Ondo State, Nigeria; ADC = polyclonal plot, ADC Camp, Ikom, Cross River State, Nigeria

*Clones with more than one correct pedigree based on assigned *Q* values were classified into Types

[a] The value in parenthesis indicates the number of contaminant attributable to the particular clone

proposed that materials of Upper Amazon origin were already in existence in West Africa before the Posnette's introduction into the region in 1944. However, the present study showed that wrong materials were probably sampled from the seed garden due to mislabeling. The local materials, C1 to C39, were often used as male parents in Nigerian seed garden establishment (Aikpokpodion and Adewale, personal communication). We therefore identified the following as some of the major factors for the high mislabeling: poor handling of budding technique, e.g., budding above the cotyledonary scar; poor germplasm; and poor plot management, e.g., root stocks overgrowing some of the scions, loss of the correct male trees in the seed garden because they were less vigorous than Upper Amazon materials (used as female parents), replacement of some of the missing stands with wrong materials from the available clones, and open-pollinated trees from fallen beans of other trees. The same possible explanations hold for the labeling error observed at AGS in C25 and C27 where IMC × PA (i.e., C75) and IMC (i.e., C77) were confirmed as contaminants. Except for a few, most of the materials labeled as C67 at ADC and AJAS were not C67. The sampled C67 trees at ADC were confirmed to be T12/11, whereas at AJAS C77, they were wrongly labeled as C67. The biclonal seed garden at AGS was originally established with C74 and C18 in block 1, C75 and C25 in block 2, C75 and C14 in block 3, and C77 and C27 in block 4. The plot at OTU was also established with C75 and C25 in block 1 and C74 and C23 in block 2. This indicates that the labeling error of C77 observed in these plots originated from the source of budwood collection or from the propagation shed in the nursery. Mislabeling of parental stocks may be responsible for the low progress recorded in cacao breeding over the years since the clones used for crosses in seed gardens to generate new full-sib families were wrongly identified.

The results of assignment test not only identified mislabeling among the clones but also revealed the proportion of parentages contributed by each germplasm population, thus providing guidance to breeders for the choice of crosses that will produce desired hybrids in the future. Genetic profiles of candidate clones of correctly identified trees in the present study can be used as reference set of cacao collections that do not have a reference tree in the international germplasm collections. The present study shows that assignment test is a useful technique for determining the population origin of a single individual through Bayesian methodology (Falush et al. 2003; Pritchard et al. 2000). Recently introduced international clones into Nigerian germplasm collections could have been responsible for the observation in this study of genetic groups that cut across the major primary populations. However, only a small proportion of the available genetic diversity in our germplasm collection has been utilized for the development of improved varieties (Aikpokpodion et al. 2010), thereby resulting in a very narrow genetic base of

www.manaraa.com

Table 6 Assigned memberships (Q value) of correctly identified genotypes of GF24 in Nigeria germplasm

Sample ID in plot	Recorded pedigree	Scavina	IMC	Parinari	Nacional	Amelonado	Criollo	Confirmed pedigree	Identity
GF24_NIG1843	Amelonado × unknown	0.005	0.011	0.005	0.004	0.489	0.486	Trinitario	GF24-Type1
GF24_NIG421	Amelonado × unknown	0.007	0.014	0.006	0.003	0.732	0.238	Trinitario	GF24-Type1
GF24_NIG533	Amelonado × unknown	0.006	0.533	0.013	0.004	0.428	0.016	Amelonado × IMC	GF24-Type2
GF24_NIG206	Amelonado × unknown	0.429	0.175	0.030	0.017	0.339	0.009	Amelonado × SCA	GF24-Type3
GF24_NIG561	Amelonado × unknown	0.014	0.014	0.578	0.005	0.388	0.002	Amelonado × PA	GF24-Type4
GF24_NIG551	Amelonado × unknown	0.012	0.010	0.586	0.005	0.386	0.002	Amelonado × PA	GF24-Type4
GF24_NIG1915	Amelonado × unknown	0.030	0.008	0.015	0.476	0.436	0.035	Nacional hybrid	GF24-Type5

commercial cultivars (Schnell et al. 2005; Warren and Kennedy 1991). Aikpokpodion et al. (2010) further studied the level of utilization of the genetic diversity present in the cacao collections and populations that have had influence on

Table 7 Level of mislabeling in the C clones of Nigeria cacao breeders' collections

Clone label in plot	Plot		No. of trees sampled	Mislabeled trees	Recorded pedigree	Identity of contaminant[a]	Unidentified contaminant (%)
	Budwood garden	Seed garden					
C18		AGS	29	28	Trinitario × Nacional hybrid	IMC (26), PA (1), IMC × PA (1)	0
C25		AGS	37	37	–	IMC × PA (18), IMC (12), Amelonado (4), PA hybrid (1), Amelonado × PA (1), IMC × Amelonado (1)	0
		OTU	7	6	Trinitario	IMC (3), Amelonado (2), IMC × PA (1)	0
C27		AGS	36	36	–	PA (17), IMC (12), IMC × PA (2), Amelonado × PA (1)	8
C3	IBIC		2	1	Trinitario	Amelonado × PA (1)	0
C30	IBLC		7	6	Amelonado × unknown	IMC × PA (2), Amelonado × IMC (1), IMC (1), PA (1)	14
C42	IBN		7	2	PA × IMC	IMC (1), Amelonado (1)	0
C60	IBLC		5	5	–	IMC (1), SCA × PA (1), SCA × Amelonado (1), IMC × Amelonado (1), SCA hybrid (1)	0
C67		ADC	80	78	IMC × PA	SCA × PA (72), IMC (2), SCA hybrid (2), IMC hybrid (1), SCA × Trinitario (1)	0
		AJAS	22	21	IMC × PA	IMC (10), Amelonado (2), Trinitario (2), SCA × PA (2), PA (1), IMC × Amelonado (1), Nacional hybrid (1), IMC hybrid (1), Trinitario × UAF (1)	0
	IBLC		24	11	IMC × PA	PA (5), IMC (3), Trinitario (1), SCA × PA (1), IMC hybrid (1	0
	IBN		5	0	IMC × PA		0
C74		AGS	22	22	–	IMC (18), Amelonado (3), PA (1)	0
C75		AGS	57	22	PA × IMC	IMC (11), Amelonado (6), Trinitario (3), Trinitario × Nacional hybrid (1), IMC × Amelonado (1)	0
		OTU	53	15	PA × IMC	IMC (13), PA (2)	0
C77		AGS	15	12	IMC	Amelonado (6), IMC × PA (4), PA (1), IMC hybrid (1)	0
		AJAS	62	54	IMC	IMC × PA (29), PA (22), Trinitario (2), Amelonado (1)	0
	IBIC		27	12	IMC	Trinitario (3), IMC × PA (5), PA (2), IMC hybrid (2)	0
	IBLC		28	19	IMC	IMC × PA (15), PA (1), PA hybrid (1), Nacional hybrid (1) Trinitario (1)	0
	IBN		4	1	IMC	IMC × PA (1)	0

IBIC = international clone plot, CRIN, Ibadan, Nigeria; IBN = clonal plot, CRIN Nursery, Ibadan, Nigeria; IBLC = local clone plot, CRIN, Ibadan, Nigeria; AGS = polyclonal plot, Ago Store, Ondo State, Nigeria; OTU = biclonal plot, Otu, Ondo State, Nigeria; ADC = polyclonal plot, ADC Camp, Ikom, Cross River State, Nigeria; AJAS = biclonal plot, Ajassor-Ikom, Cross River state, Nigeria

[a] The number in parenthesis indicates the amount of contaminant attributable to the particular clone

www.manaraa.com

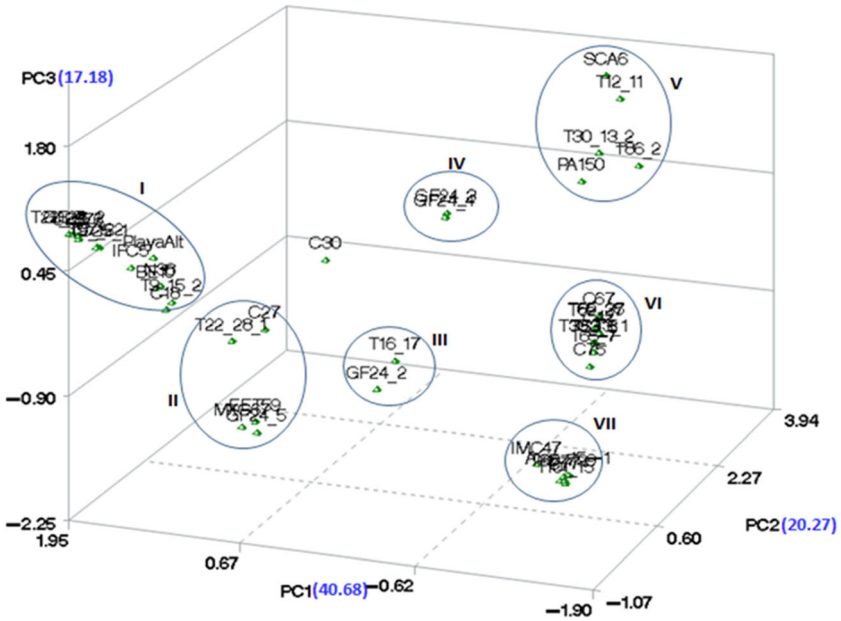

Fig. 3 Three-dimensional spatial plot of breeders' cacao collection showing their genetic similarity as revealed by STRUCTURE

varieties that were developed and distributed to Nigerian farmers by CRIN. Their results showed that Upper Amazons populations such as Scavina and Iquitos Mixed Calabacillo are yet to be significantly exploited in development efforts with varieties. This may be responsible for the prevalence of low yields experienced in commercial plantations as influenced by pests and diseases. Therefore, there is the need for targeted exploitation of useful underutilized genetic resources available in our germplasm collections for variety development in future breeding program.

In summary, this study detected a high level of mislabeling in recently introduced international germplasm in Nigeria. The problem of mislabeling has been attributed to errors from the sources of introduction, pre-planting labeling errors, and rootstocks overtaking budded scions due to poor field management. We therefore suggest the following as a remedy to the labeling errors and offtypes observed in field genebanks and seed gardens as well as improved management strategies of the diversity present in Nigerian cacao collections. First, great cautions are required in nursery practices in order to ensure accuracy in labeling and correct identities of clones generated for parental gardens and clonal plots establishment. Training on new propagation techniques with respect to the rooting of cuttings is necessary as the best alternative to budding and grafting methods which predispose juvenile clones to the challenge of rootstocks overtaking the scions under poor management on the field. There is the need for immediate correction of mislabeled trees in breeders' active clone collection and seed garden plots as well as replacement of dead stands in the plots using clones that have been confirmed as true-to-type for the purpose of obtaining better outputs from cacao breeding investments. Permanent labels using barcoding technique needs to be adopted to eliminate the problem of losses and fading of tree labeling tags on the field. Padi et al. (2015) also suggested the necessity of mainstreaming SNP fingerprinting in the breeding program through SNP marker auditing of sampled progenies from manual pollinators as a way to checkmate pollen contamination during hybridization. Conservation of cacao genetic materials in Nigeria needs to be reorganized by establishing all correctly identified clones in a new breeders' core collection germplasm plot. This proposed germplasm plot should be adequately equipped with drip irrigation facilities as a mitigation strategy against losses associated with severe drought in order to ensure efficient conservation and utility of these valuable genetic resources.

Acknowledgements The authors express appreciation to the Executive Director of the Cocoa Research Institute of Nigeria for the permission granted to publish this paper. The invaluable support of field and laboratory staff during the execution of this work is gratefully acknowledged. The authors also thank Drs. Sue Mischke and Lyndel Meinhardt, USDA-ARS, for reviewing and editing this manuscript.

Data archiving statement All row data for the Nigerian germplasm and the reference trees are being submitted to the International Cacao germplasm Database (http://www.icgd.rdg.ac.uk/). The full list of SNP markers and reference cacao accessions are included as Supplementary Materials of this manuscript.

Funding The authors wish to acknowledge the World Cocoa Foundation's African Cocoa Initiative for funding this study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

www.manaraa.com

References

Adewale B, Adeigbe O, Muyiwa A (2016) Cocoa seed garden: a means to disseminating improved planting materials for enhanced national productivity: a review. Agric Rev 37:205–212

Adomako B (2006) Combining ability analysis of blackpod disease incidence in cocoa genotypes in Ghana. Trop Sci 46:201–204

Aikpokpodion P (2007) Genetic diversity in Nigerian cacao, (Theobroma cacao L.) collections as revealed by phenotypic and simple sequence repeats marker. The University of Ibadan, Nigeria, Ibadan

Aikpokpodion PO (2012) Defining genetic diversity in the chocolate tree, Theobroma cacao L. grown in West and Central Africa. Genetic diversity in plants. ISBN: 978-953-51-0185-7, InTech https://doi.org/10.5772/33101. http://www.intechopen.com/books/genetic-diversity-in-plants/defining-genetic-diversity-in-the-chocolate-tree-theobroma-cacao-l-grown-in-west-and-central-africa. Accessed 28 July 2014

Aikpokpodion PO, Adetimirin VO, Ingelbrecht I, Schnell RJ, Kolesnikova-Allen M (2005) Assessment of genetic diversity of cacao, Theobroma cacao L. collections in Nigeria using simple sequence repeat markers. In: Denamany G, Lamin K, Ling A, Maisin N, Ahmad AC, Saripah B, Nuraziawati MY (eds) Sustainable cocoa economy through increase in productivity, efficiency and quality: proceedings of 4th Malaysian International Cocoa Conference, Kuala Lumpur, Malaysia 18th–19th July 2005, Malaysian Cocoa Board, Kota Kinubalu, pp 83–86

Aikpokpodion PO, Motamayor JC, Adetimirin VO, Adu-Ampomah Y, Ingelbrecht I, Eskes AB, Schnell RJ, Kolesnikova-Allen M (2009) Genetic diversity assessment of sub-samples of cacao, Theobroma cacao L. collections in West Africa using simple sequence repeats marker. Tree Genet Genom 5:699–711

Aikpokpodion PO, Kolesnikova-Allen M, Adetimirin VO, Guiltinan MJ, Eskes A, Motamayor JC, Schnell RJ (2010) Population structure and molecular characterization of Nigerian field genebank collections of cacao, Theobroma cacao L. Silvae Genet 59:273–285

Allegre M, Argout X, Boccara M, Fouet O, Roguet Y, Berard A, Thevenin JM, Chauveau A, Rivallan R, Clement D, Courtois B, Gramacho K, Boland-Auge A, Tahi M, Umaharan P, Brunel D, Lanaud C (2012) Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of Theobroma cacao L. DNA Res 19:23–35

Argout X, Fouet O, Wincker P, Gramacho K, Legavre T, Sabau X, Risterucci A, da Silva C, Cascardo J, Allegre M, Kuhn D, Verica J, Courtois B, Loor G, Babin R, Sounigo O, Ducamp M, Guiltinan MJ, Ruiz M, Alemanno L, Machado R, Phillips W, Schnell R, Gilmour M, Rosenquist E, Butler D, Maximova S, Lanaud C (2008) Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of Theobroma cacao L. generated from various tissues and under various conditions. BMC Genomics 9:512

Bhattacharjee R, Kolesnikova-Allen M, Aikpokpodion P, Taiwo S, Ingelbrecht I (2004) An improved semiautomated rapid method of extracting genomic DNA for molecular marker analysis in cocoa, Theobroma cacao L. Plant Mol Biol Report 22:435–436

Cervantes-Martinez C, Brown JS, Schnell RJ, Phillips-Mora W, Takrama JF, Motamayor JC (2006) Combining ability for disease resistance, yield, and horticultural traits of cacao (Theobroma cacao L.) clones. J Am Soc Hortic Sci 131:231–241

Cosme-Reyes SM, Cuevas HE, Zhang D, Oleksyk TK, Irish BM (2016) Genetic diversity of naturalized cacao (Theobroma cacao L.) in Puerto Rico. Tree Genet Genomes 12:88. https://doi.org/10.1007/s11295-016-1045-4

DuVal A, Gezan SA, Mustiga G, Stack C, Marelli JP, Chaparro J, Livingstone D, Royaert S, Motamayor JC (2017) Genetic parameters and the impact of off-types for *Theobroma cacao* L. in a breeding program in Brazil. Front Plant Sci 8:2059. https://doi.org/10.3389/fpls.2017.02059 eCollection 2017

Efombagan IB, Motamayor JC, Sounigo O, Eskes AB, Nyasse S, Cilas C, Schnell RJ, Manzanares-Dauleux M, Kolesnikova-Allen M (2008) Genetic diversity and structure of farm and genebank accessions of cacao (Theobroma cacao L.) in Cameroon revealed by microsatellite markers. Tree Genet Genomes 4:821–831

Eskes AB, Efron Y (2006) Global approaches to cocoa germplasm utilization and conservation. Final report of the CFC/ICCO/IPGRI project on "Cocoa germplasm utilization and conservation: a global approach" (1998–2004). CFC, ICCO, IBPGR, Amsterdam

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Fang W, Meinhardt LW, Mischke BS, Bellato C, Motilal L, Zhang D (2014) Accurate determination of genetic identity for a single cacao bean, using molecular markers with a nanofluidic system, ensures cocoa authenticity and traceability. J Agric Food Chem 62:481–487

Figueira A, Janick J, Levy M, Goldsbrough P (1994) Reexamining the classification of Theobroma cacao L. using molecular markers. J Am Soc Hortic Sci 119:1073–1082

Guiltinan MJ, Verica J, Zhang D, Figueira A (2008) Genomics of Theobroma cacao, "the food of the gods". In: Moore PH, Ming R (eds) Genomics of tropical crop plants. Springer, New York, pp 145–170

ICCO (2017) Production of cocoa beans. ICCO quarterly bulletin of cocoa statistics, Vol. XLIII, No. 1, Cocoa year 2016/17

Jacob V, Atanda O, Opeke L (1971) Cacao breeding in Nigeria. In: Progress in tree crops research in Nigeria. CRIN commemorative publication pp 23–33

Ji K, Zhang D, Motilal L, Boccara M, Lachenaud P, Meinhardt LW (2012) Genetic diversity and parentage in farmer varieties of cacao (Theobroma cacao L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. Genet Resour Crop Evol 60:441–453. https://doi.org/10.1007/s10722-012-9847-1

Johnson ES, Bekele FL, Brown SJ, Song Q, Zhang D, Meinhardt LW, Schnell RJ (2009) Population structure and genetic diversity of the Trinitario cacao (L.) from Trinidad and Tobago. Crop Sci 49:564–564

Kalinowski ST, Manlove KR, Taper ML (2007) ONCOR a computer program for genetic stock identification. Department of Ecology, Montana State University, Bozeman MT 59717. Available: http://www.montana.edu/kalinowski

Kuhn DN, Livingstone D, Main D, Zheng P, Saski C, Feltus FA, Mockaitis K, Farmer AD, May GD, Schnell RJ, Motamayor JC (2012) Identification and mapping of conserved ortholog set (COS) II sequences of cacao and their conversion to SNP markers for marker-assisted selection in Theobroma cacao and comparative genomics studies. Tree Genet Genomes 8:97–111

Lerceteau E, Robert T, Pétiard V, Crouzillat D (1997) Evaluation of the extent of genetic variability among Theobroma cacao accessions using RAPD and RFLP markers. Theor Appl Genet 95:10–19

Livingstone DS, Motamayor JC, Schnell RJ, Cariaga K, Freeman B, Meerow AW, Brown JS, Kuhn DN (2011) Development of single nucleotide polymorphism markers in Theobroma cacao and comparison to simple sequence repeat markers for genotyping of Cameroon clones. Mol Breed 27:93–106

Livingstone DS, Freeman B, Motamayor JC, Schnell RJ, Royaert S, Takrama J, Meerow AW, Kuhn DN (2012) Optimization of a SNP assay for genotyping Theobroma cacao under field conditions. Mol Breed 30:33–52

Lockwood G, Gyamfi MMO (1979) A note on codes at Tafo, Cocoa Research Institute, Ghana, technical Bulletin No 10. pp 1–61

Lukman ZD, Susilo A, Dinarti D, Bailey BA, Mischke BS, Meinhardt LW (2014) Genetic identity, ancestry and parentage in farmer

www.manaraa.com

selections of cacao from Aceh, Indonesia revealed by single nucleotide polymorphism (SNP) markers. Trop Plant Biol 7:133–143

Motamayor JC, Lachenaud P, da Silva e Mota JW, Loor R, Kuhn DN, Brown JS, Schnell RJ (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (Theobroma cacao L). PLoS ONE 3:e3311. https://doi.org/10.1371/journal.pone.0003311

Motamayor J, Schnell R, Kuhn D (2012) Applying SNP marker technology in the cacao breeding programme in Ghana. Afr Crop Sci J 20: 67–75

Motilal LA (2004) The potential of cacao microsatellites amplification across diverse plant taxa. In: Genetic Resources and Biotechnology. Edited by D T, T P, PA B, vol. 2. New Delhi: Regency Publications, pp 24–49

Motilal LA, Butler D (2003) Verification of identities in global cacao germplasm collections. Genet Resour Crop Evol 50:799–807

Motilal LA, Zhang D, Umaharan P, Mischke S, Mooleedhar V, Meinhardt LW (2010) The relic Criollo cacao in Belize—genetic diversity and relationship with Trinitario and other cacao clones held in the International Cocoa Genebank, Trinidad. Plant Genet Resourc 8: 106–115

N'Goran JAK, Laurent V, Risterucci AM, Lanaud C (2000) The genetic structure of cocoa populations (Theobroma cacao. L.) revealed by RFLP analysis. Euphytica 115(2):83–90

Padi FK, Ofori A, Takrama J, Djan E, Opoku SY, Dadzie AM, Bhattacharjee R, Motamayor JC, Zhang D (2015) The impact of SNP fingerprinting and parentage analysis on the effectiveness of variety recommendations in cacao. Tree Genet Genomes 11:1–14

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 (2012) genetic analysis in excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539

Posnette AF (1944) Pollination of cacao in Trinidad. Trop Agric (Trinidad) 21(6):115–118

Posnette AF (1986) Fifty years of cocoa research in Trinidad and Tobago. Cocoa Research Unit, University of the West Indies, St. Augustine

Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

SAS (2011) SAS/STAT software, version 9.3. SAS Institute Inc, Cary

Saunders JA, Mischke S, Leamy EA, Hemeida AA (2004) Selection of international molecular standard for DNA fingerprinting. Theor Appl Genet 110:41–47

Schnell R, Brown J, Kuhn D, Cervantes-Martinez C, Borrone J, Phillips W, Johnson E, Monteverde-Penso E, Motamayor J, Amores F (2005) Current challenges of tropical tree crop improvement: integrating genomics into an applied cacao breeding program. In: International symposium on biotechnology of temperate fruit crops and tropical species 738:129–144

Sounigo O, Umaharan R, Christopher Y, Sankar A, Ramdahin S (2005) Assessing the genetic diversity in the International Cocoa Genebank, Trinidad (ICG, T) using isozyme electrophoresis and RAPD. Genet Resour Crop Evol 52:1111–1120

Takrama J, Cervantes-Martinez C, Phillips-Mora W, Brown J, Motamayor J, Schnell R (2005) Determination of off-types in a cocoa breeding programme using microsatellites. Ingenic. Newsletter 10:2–8

Takrama J, Kun J, Meinhardt L, Mischke S, Opuku S, Padi FK, Zhang D (2014) Verification of genetic identity of introduced cacao germplasm in Ghana using single nucleotide polymorphism (SNP) markers. Afr J Biotechnol 13:2127–2136

Toxopeus H (1964). F3 Amazon in Nigeria. In Annual report of the cocoa research Institute of Nigeria, Ibadan, 1963/64, 13–23. Reprinted 1982: archives of cocoa research 1:179–191

Toxopeus H (1985) Botany, types and populations. In: Wood GAR, Lass RA (eds) Cocoa, 4th edn. Blackwell Science, Oxford, pp 11–37

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol 10:249–256

Warren J, Kennedy A (1991) Cocoa breeding revisited. Cocoa Growers' Bull (United Kingdom) 44:18–24

Whitkus R, de la Cruz M, Mota-Bravo L, Gómez-Pompa A (1998) Genetic diversity and relationships of cacao (Theobroma cacao L.) in southern Mexico. Theor Appl Genet 96:621–627

Zhang D, Arevalo-Gardini E, Mischke S, Zuñiga-Cernades L, Barreto-Chavez A, Del Aguila JA (2006) Genetic diversity and structure of managed and semi-natural populations of cocoa (Theobroma cacao) in the Huallaga and Ucayali valleys of Peru. Ann Bot 98:647–655

Zhang D, Boccara M, Motilal L, Mischke S, Johnson ES, Butler DR, Bailey B, Meinhardt L (2009a) Molecular characterization of an earliest cacao (Theobroma cacao L.) collection from upper Amazon using microsatellite DNA markers. Tree Genet Genomes 5:595–607

Zhang D, Mischke S, Johnson ES, Phillips-Mora W, Meinhardt L (2009b) Molecular characterization of an international cacao collection using microsatellite markers. Tree Genet Genomes 5:1–10

Zhang D, Motilal L (2016) Origin, dispersal and current global distribution of cacao genetic diversity. In: Baley BA, Meinhardt LW (eds) Cacao Diseases. Springer International Publishing, Switzerland, p 3–31

المنارة للاستشارات
www.manaraa.com

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.

www.manaraa.com