Genet Resour Crop Evol (2007) 54:1301–1313
DOI 10.1007/s10722-006-9111-7

RESEARCH ARTICLE

Field Guide efficacy in the identification of reallocated clonally propagated accessions of cacao (*Theobroma cacao* L.)

Elizabeth S. Johnson · Antonio Mora · Raymond J. Schnell

Received: 6 March 2006 / Accepted: 28 August 2006 / Published online: 27 February 2007

Abstract Misidentification is a major constraint to the conservation and utilization of cacao genetic resources. One solution for rapid and accurate identification is to produce a Field Guide for each recognized cacao early generation original population. Each Field Guide in this series of compendia would be comprised of pod digital images with complementary morphological descriptors, molecular fingerprint data with kinship analyses and agronomic data useful to cacao cultivar development. This Field Guide Concept culminated in the production of the first compendium entitled, Field Guide to the ICS Clones of Trinidad. In the current study, this Field Guide was used to verify the identity of 69 trees representing a possible 15 ICS clone genotypes introduced nearly 50 years ago to Costa Rica. Phase one of this study involved identity verification of the trees in Costa Rica using the pod digital image and morphological descriptor data in the Field Guide to the ICS Clones of Trinidad. The error rate was 3.5% for misidentification in the field in Costa Rica. In the second and final phase, SSR fingerprint data were generated for each of the 69 trees and analyzed for verification to the original ICS genotype in the Field Guide. Misidentification was reduced to 0% with the addition of SSR fingerprint analyses. The 69 trees in this study clustered into two groups or sub-populations clearly differentiated by Discriminant analyses with six SSR primers. A 46% cost reduction in SSR fingerprinting of the ICS clones was realized by combining use of the pod image and morphological descriptor data in the Field Guide with genetic diversity estimates derived from these six primers. A Field Guide approach to the identification of reference genotypes for cacao germplasm is discussed.

Keywords Cost reduction of cacao SSR Fingerprinting · Genotype verification of genetic resources · Germplasm misidentification · Reference cacao genotype · SSR fingerprinting · *Theobroma cacao* L.

E. S. Johnson (✉)
CABI Bioscience/USDA, c/o Laboratorio de Biotecnología, CATIE 7170, Turrialba, Costa Rica C.A.
e-mail: johnsoel@ba.ars.usda.gov

A. Mora
Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), CATIE 7170, Turrialba, Costa Rica C.A.

R. J. Schnell
USDA-ARS Subtropical Horticultural Research Station, 13601 Old Cutler Rd., Miami, FL 33158, USA

Introduction

The beans of the neotropical understory tree, *Theobroma cacao* L., provides the basic ingredients

cocoa solids and cocoa butter for the multi-billion dollar chocolate and cosmetic industries. Currently, the crop is produced mainly by farmers with small acreages in developing countries around the world with tropical climates. However, the products made from the cacao bean have fascinated man from as early as the 12th century (Lerceteau et al. 1997). Despite such a long history of cultivation and selection, current cacao cultivars are quite similar to the originally cultivated cacao germplasm.

One of the reasons for slow genetic improvement and under utilization of genetic resources of this crop is the use of misidentified parents in cacao breeding programs. This has impeded resolution of the genetic basis underlying agronomic traits as the resulting mixed populations of progeny produce inconsistent results and conflicting genetic estimates over time and location. Many factors have contributed to the problem of misidentification of cacao genetic resources ranging from difficulties in applying current methodologies of identification, germplasm exchange and maintenance in collections to the economic sustainability of cacao programs all interwoven with the peculiarities of the species.

Cacao germplasm must be maintained as living tree gene banks. Two cacao germplasm collections, one maintained in the International Cocoa Genebank, Trinidad (ICG, T) with over 2,300 accessions (Bekele et al. 2006) and the second maintained at the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) in Costa Rica with over 750 accessions, have been recognized as genetically complementary and designated ''Universal Collection Repositories'' for *Theobroma cacao* L. (IBPGR 1981).

Cacao production is severely constrained by major fungal diseases such as *Moniliophthora* or frosty pod rot (MPR), Vascular streak dieback (VSD) and Witches' broom (WB), an Oomycete which cause Black pod or *Phytophthora* pod rot (PPR), and the viral disease cacao swollen shoot (CSSV) (Purdy et al. 1998). These diseases are individually capable of near to total devastation of the cacao crop and are present in different producing regions (Bowers et al. 2001; Frison et al. 2000). Although black pod is pandemic, it is caused by different species of *Phytophthora* in different producing regions. A quarantine system for the exchange of cacao germplasm has been developed, is currently in use, and has been very successful in limiting the spread of the aforementioned and other cacao diseases (Frison et al. 2000).

The proper exchange of cacao germplasm through quarantine requires a minimum of 3 years and is a multi-step process. It is initiated with the collection of budwood from trees in the donor country, which are treated (Frison et al. 2000) and shipped to the current quarantine facilities in Reading University, UK. At the quarantine facility the germplasm is regenerated by grafting onto rootstock and the scion or rootstock, for witches' broom or swollen shoot virus disease symptoms, respectively, closely observed for a minimum of 2 years (Frison et al. 2000). Budwood is harvested from disease free plants after the quarantine period and shipped to the recipient country. Upon arrival, the materials are again regenerated by grafting onto rootstock and maintained in the greenhouse until the scion is well established. It could be several months or up to a year later before the introduced materials are transplanted into the field in the recipient country. It will be at least another 2 years before the exchanged germplasm begins to produce sufficient pods, flowers and flush leaves for effective verification by morphological descriptors.

The use of morphological descriptors to identify cacao germplasm is well established with the most comprehensive compilation of this data available in the public domain at the International Cocoa Germplasm Database (ICGD 2003). Initiated in 1990, the database now holds information on over 14,000 cacao accessions with morphological descriptors for 5,000 genotypes.

Cacao is a predominantly cross-pollinating species and the hybrid nature of the early generation materials with a combination of phenotypic characteristics limits the use of anecdotal and morphological descriptors for identification purposes. Moreover, some morphological descriptors can also be influenced by environment (Engels 1992; Bekele 1992; Bekele and Butler 1998). The growth cycle of cacao makes simultaneous application of all three forms (flower, pod and flush leaf) of morphological descriptors near impossible.

The energy demands of pod development and maturity inhibits the production of flush leaves or new growth in the cacao tree. As such the complete identification of a cacao genotype with the use of morphological descriptors can take up to a year.

Among the most frequently exchanged and thus widely distributed cacao accessions are early generation lines or populations (Ronning and Schnell 1994) such as the ICS clones of Trinidad as they possess agronomic traits of interest for the development of new cacao cultivars. However, the latest estimate of the misidentification problem of cacao accessions in germplasm collections was in the range of 15–44%, the problem being more severe with widely distributed accessions (Motilal and Butler 2003).

The Field Guide Concept (Johnson et al. 2004) was to produce a tool designed to facilitate rapid and accurate identification, in both field and laboratory applications, of each recognized cacao early generation original population. Utility of the population, and as such cacao genetic resources, could be improved by incorporating information to aid in the decision-making process in the selection of appropriate parents for the development of new cacao cultivars. These populations are composed of researchers' and/or farmers' selections from outcross or early generation progeny from cacao breeding programs.

Each Field Guide would be a compendium comprised of two parts, the first containing summarized background information of the subject matter or population and kinship analyses depicting relationships amongst the individuals. The second part is a Pod Image Library consisting of mature ripe and unripe pod digital images for each individual juxtaposed by complementary morphological descriptors, agronomic data useful to cacao cultivar development and molecular fingerprint data in tabular format. The Field Guide Concept culminated in the production of the first in the series of compendia entitled, Field Guide to the ICS Clones of Trinidad based on the Imperial College Selection (ICS) Clones.

The objectives of this study were to literally put the Field Guide to the ICS Clones of Trinidad in practice and determine its utility in the identification of the ICS clones introduced from Trinidad in 1959, planted and maintained in the cacao germplasm collection at CATIE in Costa Rica.

Materials and methods

Plant materials

In 1959, CATIE introduced 27 of the ICS clones from the Republic of Trinidad and Tobago, which were planted on-station in the cacao germplasm fields called Turriabla Cabiria-2 and -5. Only the accession names from the passport data of the introduced clones are available at CATIE. However, 14 of the 27 (ICS 1, 6, 8, 16, 40, 43, 44, 46, 60, 84, 89, 95, 98, 100) are represented in the compendium, Field Guide to the ICS Clones of Trinidad. The ICS clones in this Field Guide are those located at the San Juan Estate in Trinidad, the original field planted to these clones, and are used as the Reference genotypes in this study. A 15th genotype was also included as a Reference for ICS 100. This tree is from the germplasm cacao collection in Trinidad, located in Campus Field 4 at position X12,Y12, on the campus of the University of the West Indies, St. Augustine. Currently, a total of 70 trees (Table 1) are labeled as members of the 15 Reference ICS clones in Cabiria 2 and Cabiria 5 in Costa Rica.

Application of the compendium, Field Guide to the ICS clones of Trinidad, to single tree identity verification in the field in Cabiria, Costa Rica

Pod image and morphological descriptors for the 15 ICS reference clones, in the Pod Image Library of the Field Guide to the ICS Clones of Trinidad, were used to evaluate the pod morphology of each tree possessing mature fruits and representing that ICS clone in Cabiria (Table 1). The pod image and morphological descriptors in the Field Guide complement each other to visually represent the recommended procedure for characterization of cacao germplasm by pod descriptors (IBPGR 1981, p. 25; Bekele and Butler 1998).

Table 1 The results of application of the compendium ''Field Guide of the ICS Clones of Trinidad'' in the genotype verification of trees labeled as ICS clones in Cabiria 2 and 5 (only TC-859) at CATIE in Turrialba, Costa Rica. The ''?'' designate trees of as yet unknown identity

Sample ID	Row label in Cabiria	Lot and tree number	Pod descriptor	Genotype by SSR markers
#TC-824	ICS-1	Lot E, Tree 5	Off-type	
TC-825[a]	ICS-1	Lot E, Tree 6	NF	ICS-1
#TC-826	ICS-1	Lot E, Tree 7	Off-type	
#TC-827	ICS-1	Lot E, Tree 8	Off-type	
#TC-828	ICS-1	Lot E, Tree 11	Off-type	
#TC-829	ICS-1	Lot E, Tree 13	NF	
#TC-830	ICS-1	Lot E, Tree 14	Off-type	
#TC-831	ICS-16	Lot E, Tree 5	NF	
TC-832	ICS-40	Lot E, Tree 5	OK	ICS-40
#TC-833	ICS-40	Lot E, Tree 8	Off-type	
TC-834	ICS-40	Lot E, Tree 9	YF	ICS-40
TC-835	ICS-40	Lot E, Tree 11	NF	ICS-40
#TC-836[a]	ICS-46	Lot E, Tree 2	Off-type	? Not Catongo Blanco
#TC-837	ICS-46	Lot E, Tree 5	Off-type	? Not Catongo Blanco
#TC-838	ICS-46	Lot E, Tree 11	Off-type	? Not Catongo Blanco
#TC-839	ICS-46	Lot E, Tree 13	NF	? Not Catongo Blanco
TC-840	ICS-60	Lot E, Tree 5	NF	Progeny of ICS-40
#TC-841[a]	ICS-60	Lot E, Tree 10	NF	
TC-842	ICS-60	Lot E, Tree 11	YF	Progeny of ICS-40
TC-843	ICS-95	Lot E, Tree 1	NF	ICS-95
TC-844[a]	ICS-95	Lot E, Tree 4	NF	Possibly hybrid (ICS-95 X ICS-60)
TC-845	ICS-95	Lot E, Tree 7	YF	Possibly progeny of ICS-40
TC-846	ICS-95	Lot E, Tree 8	OK	Possibly progeny of ICS-40
TC-847	ICS-95	Lot E, Tree 9	OK	ICS-95
#TC-848	ICS-95	Lot E, Tree 10	NF	
TC-849	ICS-84	Lot E, Tree 2	NF	ICS-60
TC-850[a]	ICS-84	Lot E, Tree 4	NF	ICS-60
#TC-851	ICS-84	Lot E, Tree 12	YF	
TC-852[a]	ICS-6	Lot E, Tree 2	NF	ICS-6
TC-853	ICS-6	Lot E, Tree 3	OK	ICS-6
TC-854	ICS-6	Lot E, Tree 4	OK	ICS-6
TC-855	ICS-6	Lot E, Tree 6	OK	ICS-6
#TC-856	ICS-6	Lot E, Tree 7	YF	
TC-857	ICS-6	Lot E, Tree 8	OK	ICS-6
TC-858	ICS-6	Lot E, Tree 11	OK	ICS-6
TC-859[a]	ICS-8	Cabiria 5,Tree 1	OK	ICS-8
#TC-860[a]	ICS-43	Lot E, Tree 2	NF	
#TC-861	ICS-43	Lot E, Tree 3	Off-type	
TC-862	ICS-43	Lot E, Tree 4	NF	ICS-95
TC-863	ICS-43	Lot E, Tree 5	Off-type	ICS-95
TC-864	ICS-43	Lot E, Tree 6	NF	ICS-95
TC-865	ICS-43	Lot E, Tree 7	NF	ICS-95
TC-866	ICS-43	Lot E, Tree 10	Off-type	ICS-95
#TC-867	ICS-44	Lot E, Tree 1	NF	
#TC-868	ICS-44	Lot E, Tree 2	NF	
#TC-869[a]	ICS-44	Lot E, Tree 4	NF	
#TC-870	ICS-44	Lot E, Tree 5	Off-type	
#TC-871	ICS-44	Lot E, Tree 6	NF	
#TC-872	ICS-44	Lot E, Tree 8	NF	
#TC-873	ICS-44	Lot E, Tree 9	NF	
#TC-874	ICS-44	Lot E, Tree 10	NF	
#TC-875	ICS-44	Lot E, Tree 11	NF	
#TC-876	ICS-44	Lot E, Tree 12	NF	
TC-877[a]	ICS-89	Lot E, Tree 1	OK	ICS-89

Table 1 continued

Sample ID	Row label in Cabiria	Lot and tree number	Pod descriptor	Genotype by SSR markers
TC-878	ICS-89	Lot E, Tree 2	OK	ICS-89
TC-879	ICS-89	Lot E, Tree 4	OK	ICS-89
TC-880	ICS-89	Lot E, Tree 5	OK	ICS-89
TC-881	ICS-89	Lot E, Tree 8	OK	ICS-89
TC-883	ICS-89	Lot E, Tree 14	OK	ICS-89
#TC-884	ICS-98	Lot D, Tree 3	YF	
#TC-885	ICS-98	Lot D, Tree 6	Off-type	
#TC-886	ICS-98	Lot D, Tree 9	Off-type	
TC-887	ICS-100	Lot E, Tree 2	NF	ICS-100 UWI
TC-888	ICS-100	Lot E, Tree 8	YF	ICS-100 UWI
TC-889[a]	ICS-100	Lot E, Tree 9	NF	ICS-100 UWI
TC-890	ICS-100	Lot E, Tree 10	NF	ICS-100 UWI
TC-891	ICS-100	Lot E, Tree 11	NF	ICS-100 UWI
#TC-892	ICS-100	Lot E, Tree 12	NF	
TC-893	ICS-100	Lot E, Tree 13	NF	ICS-100 UWI

[a] 11 Trees selected by Jim Saunders in 2000

NF, No Fruits Present; #TC, Trees confirmed as Non-ICS Genotypes (Off-type) by Set 1 primers; YF, Young Fruits present

Application of the compendium, Field Guide to the ICS clones of Trinidad, to single tree identity verification in the lab

DNA extraction, PCR amplification and Capillary electrophoresis

In the lab, the SSR Fingerprint data in the Field Guide of the ICS Clones of Trinidad was used as the reference for comparing that of the 70 trees in Costa Rica representing each of the 15 corresponding ICS clones. DNA extraction, PCR amplification of microsatellite loci and capillary electrophoresis of products, allele calling and binning were as previously described (Johnson et al. 2004; Schnell et al. 2004) and only briefly summarized here.

Total genomic DNA was extracted from 150 mg dried leaves of Inter-flush 2 developmental stage (leaves erect, recently hardened off and have lost all traces of flush color or are light green in color) (Greathouse et al. 1971) for each of the 70 trees in Cabiria representing the 15 ICS clones and the accession Catongo Blanco, as a process control using the FastPrep FP120 and Fast DNA Kits and protocol (Qbiogene, Inc., Carlsbad, CA) and standardized to 2.5 ng/µl.

PCR amplifications were performed with 14 fluorescent labeled microsatellite primers divided into two sets (Table 2). Five of the primers in Set1 and two of Set2 were recommended as standards for DNA fingerprinting of cacao (Saunders et al. 2004). The primers, developed by CIRAD (Lanaud et al. 1999; Pugh et al. 2004), were used in 10 µl or 25 µl PCR reactions, for single and multiplexed reactions, respectively, containing at final concentration 1× PCR buffer with 15 mM $MgCl_2$ (Applied Biosystems, Inc., Foster City, CA), 200 µM dNTP, 1 mg BSA, 200 nM of each Forward and Reverse primer, 0.5 U AmpliTaq DNA polymerase (Applied Biosystems, Inc., Foster City, CA), 2.5 ng of DNA template.

PCR reaction amplifications were performed using the cycling protocol 1 cycle of 94°C for 4 min, 32 cycles of 94°C for 30 s, primer Tm °C for 1 min and 72°C for 1 min. For multiplexed primers, amplification was performed for an additional cycle of 65°C for 3 min before the 4°C holding cycle. For electrophoretic separation of each sample 1 µl of amplification product, 12 µl Hi-Di Formamide and 0.2 µl Rox 500 size standard (Applied Biosystems, Inc., Foster City, CA) was denatured at 95°C for 5 min and placed immediately on ice. Capillary electrophoresis was performed on the ABI 310 automated sequencer in 36 cm capillaries using POP4 polymer at 60°C, injection parameters of 15 kV for 5 s and run parameters 15 kV.

Table 2 Microsatellite primers used to verify the of identity 69 trees in Cabiria fields 2 and 5 at CATIE in Costa Rica labeled as 15 Reference ICS Clones introduced from Trinidad nearly 50 years ago

Microsatellite primers in multiplex groups	Linkage group	T_m (°C)	Alleles identified by discriminant analyses
(mTcCIR6*, mTcCIR25*)[a]	6, 6	46	CIR6_228, CIR6_246, CIR25_150
(mTcCIR12*, mTcCIR15*, mTcCIR21)[a]	4, 1, 3	46	CIR12_187, CIR12_209, CIR15_240, CIR15_248
(mTcCIR24*)[a]	9	46	CIR24_192, CIR24_196, CIR24_200
(mTcCIR26*)[a]	8	46	CIR26_296
(mTcCIR3, mTcCIR19, mTcCIR264)[b]	2, 2, 1	46	
(mTcCIR1, mTcCIR9)[b]	8, 6	51	
(mTcCIR17, mTcCIR18)[b]	4, 4	51	

[a] Primers of Set1

[b] Primers of Set2

[*] Indicate primers shown by discriminant analyses to distinguish ICS from Non-ICS type trees in Cabiria

Allele calling and sizing were accomplished using GeneScan and Genotyper 3.7 software (Applied Biosystems, Inc., Foster City, CA). Allele sizes were standardized between runs by binning or grouping, using allele bins developed at the USDA-SHRS in Miami.

Set1 primers—microsatellite marker and statistical analyses

All statistical analyses were performed using InfoGen/P ver 1.0 (InfoGen/P 2003). Bootstrap analyses were performed with the Dice coefficient for 800 permutations using WinBoot (Yap and Nelson 1996). Cluster analyses, using the UPGMA algorithm with Dice coefficient, were performed on alleles generated by Set1 primers for 69 of the 70 cacao trees, the 15 Reference ICS clones from Trinidad and the process control, Catongo Blanco. PCR amplification failed for most primers with the DNA from tree TC-882, which subsequently died in Cabiria and was eliminated from further analyses.

Cluster analyses indicated a putative sub-population structure in the 69 trees in Cabiria, which was further investigated using Genetic diversity measures (Nei 1973, 1978). Discriminant analyses (Lindeman et al. 1980) were used to identify the primers or alleles, which best distinguish the sub-populations. The absolute value of the canonical discriminant function for each allele was used as the selection criterion for subsequent rounds of analyses. Alleles with low canonical scores were eliminated until a subset of alleles was identified that distinguished the sub-populations with 0% misclassification error.

Set2 primers—microsatellite marker and statistical analyses

A second round of SSR analyses were performed using Set2 primers (Table 2) on the sub-population of 37 cacao trees from Cabiria shown by Set1 primers to be most closely related to the 15 Reference ICS clones from Trinidad and the four trees most similar to the control Catongo Blanco (Table 1). The alleles generated by both Set1 and Set2 primers for these 37 trees were used as the SSR fingerprint for comparison to that of the 15 reference ICS clone genotypes in the Field Guide. These comparisons were performed by cluster analyses using the same algorithm and coefficient as stated previously to verify the identity of each of these 37 cacao trees in Cabiria. Principal Coordinate Analyses (PCO) were used to further clarify relationships of trees in Cabiria that were not clones of, but are closely related to the Reference ICS clones from Trinidad.

Results

Single tree identity verification using the compendium, Field Guide to the ICS clones of Trinidad

Application in Cabiria, Costa Rica

The Pod Image Library of the Field Guide to the ICS clones of Trinidad facilitated the *in situ* field identification of the 29 (42%) trees in Cabiria

with mature pods at the time of characterization. All 15 (52%) of these trees identified as ''Off-type'' (not the Reference ICS clone) in the Field Guide, were later corroborated by SSR analyses (Table 1). Three of these Off-type trees (TC-836 to TC-838) possessed pods which were very similar in appearance to that of the control, Catongo Blanco. Tree TC-846 was assessed as true-to-type for ICS 95 in the Field Guide. However, SSR analyses proved this to be a misidentification giving a possible 3.5% error rate if using the Pod Image Library in the Field Guide exclusively for tree genotype verification.

Set1 primers indicated a sub-population structure amongst the 69 trees in Costa Rica

The 7 primers of Set1 (Table 2) generated a total of 48 alleles, two of which were eliminated as non-informative or having duplicate information content. Cluster and Bootstrap analyses using the 46 informative alleles for the 15 Reference ICS clones from Trinidad, 69 trees (Table 1) and the process control Catongo Blanco from Cabiria showed that 32 (46%) of the trees were not closely related to the ICS clones. There was no significant (1–17%) bootstrap support for similarity to the ICS clones from Trinidad for these trees. Henceforth, this sub-population of 32 trees will be referred to as the ''Non-ICS Type'' in this study. Cluster and bootstrap analyses also confirmed the field observation, that four of the Non-ICS Type trees (TC-836 to TC-839) were most likely (97.3% bootstrap support) the control Catongo Blanco. Bootstrap support ranging from 47% to 98% for relatedness to the ICS clones from Trinidad was accepted for the second sub-population of 37 trees in Cabiria, which were called the ''ICS-Type'' in this study.

Genetic diversity analyses (Table 3) further supported the differentiation of the 69 trees into the two groups or sub-populations, called the Non-ICS Type and ICS-Type. The 37 trees belonging to the ICS-Type sub-population possessed a narrower genetic base compared to the 32 trees from the Non-ICS Type (Table 3). Percentage of polymorphic loci was 28% compared to 85%, the Nei's unbiased observed heterozygosity measure (*Ho*) was 0.09 compared to 0.22 with an average number of alleles per locus of 1.46 compared to 2.00 for the ICS-Type and Non-ICS Type sub-populations, respectively.

Discriminant analyses were performed with all 46 alleles from the 7 Set1 primers to determine the minimum number of SSR primers required to classify a tree as an ''Off-type'' or Non-ICS Type. Eleven alleles (CIR6_228, CIR6_246, CIR12_187, CIR24_196, CIR25_150, CIR26_296 for presence in ICS-Type and CIR 12_209, CIR15_240, CIR15_248, CIR24_192, CIR24_200 for absence in ICS-Type) from 6 of the Set1 primers (Table 2) clearly differentiated (0% misclassification error rate) the ICS-Type from the Non-ICS Type trees in Costa Rica. Based on the results from the preceding analyses the 32 trees from the Non-ICS Type sub-population, labeled with ''#'' in front of the Sample ID designation in Table 1, were eliminated from further analyses.

Set2 primers—SSR fingerprint comparison of the 37 ICS-type trees in Costa Rica to the reference in the Field Guide to the ICS clones of Trinidad

The 7 primers of Set2 (Table 2) generated 27 alleles, which were combined with the 22 alleles generated by the Set1 primers for the 37 trees in Cabiria classified as the ICS-Type sub-population.

Table 3 Genetic diversity measures for the 69 trees, Catongo Blanco (control) in Cabiria and the 15 ICS clones from Trinidad grouped according to Cluster and Bootstrap analyses with Set1 primers for 46 alleles

Statistic	Groups	
	$n = 33$ Non-ICS type trees	$n = 52$ ICS-type trees
Polymorphic loci (95)	0.85 ± 0.00	0.28 ± 0.07
Nei unbiased heterozygosity	0.22 ± 0.00	0.09 ± 0.05
Average number of alleles	2.00 ± 0.00	1.46 ± 0.07
Effective numbers of alleles	1.33 ± 0.00	1.13 ± 0.07

Standard errors estimated by Bootstrap for 850 simulations

A cluster analysis using Spearman's correlation as a similarity index was used to identify alleles having the same information content for all 37 trees. Alleles with the same information content clustered into groups each having a within cluster correlation coefficient ($r = 1$) indicating redundancy. As such only one allele from each cluster is required for analyses, resulting in the elimination of 19 of the 49 alleles generated by the combined Set1 and Set2 primers for the 37 ICS-Type trees in Cabiria. The 30 informative alleles generated by all 14 SSR primers were used in subsequent analyses.

SSR fingerprint verification of identity to the 15 ICS reference clones from Trinidad were performed using cluster analyses for the 37 ICS-Type trees, the process control Catango Blanco and its four closely related trees (TC-836 to TC-839). Only 32 of the 37 ICS-Type trees in Cabiria possessed identical SSR fingerprints to or are clones of eight of the original ICS genotypes (ICS 1, 6, 8, 60, 89, 95 and 100) introduced from Trinidad (Fig. 1). The genotype of the ICS 100 clone in Costa Rica is that of the tree located in the cacao germplasm collection on the UWI campus in St. Augustine, Trinidad.

Principal Coordinate Analyses (PCO) were conducted for the five ICS-Type trees and the four Non-ICS Type trees closely related to the control Catongo Blanco in Cabiria using the ICS Clones from Trinidad as the reference genotypes. One of the ICS-Type trees in Cabiria, TC-844, is possibly a hybrid or progeny from a cross (ICS-60TT X ICS-95TT). The other four ICS-Type trees, TC-840 and TC-842, and TC-845 and TC-846 share identical SSR fingerprints, respectively, for the 14 primers used in this study and are possibly open-pollinated progeny of ICS-40. The four Non-ICS Type trees, TC-836 to TC-839 all share the same SSR fingerprint but are not clones of the process control Catongo Blanco. Results of these analyses further demonstrate (with 93% Bootstrap support) the findings derived from the data of the Set1 primers, that the four trees TC-836 to TC-839 and Catongo Blanco belong to a different or Non-ICS Type genetic pool (Fig. 2).

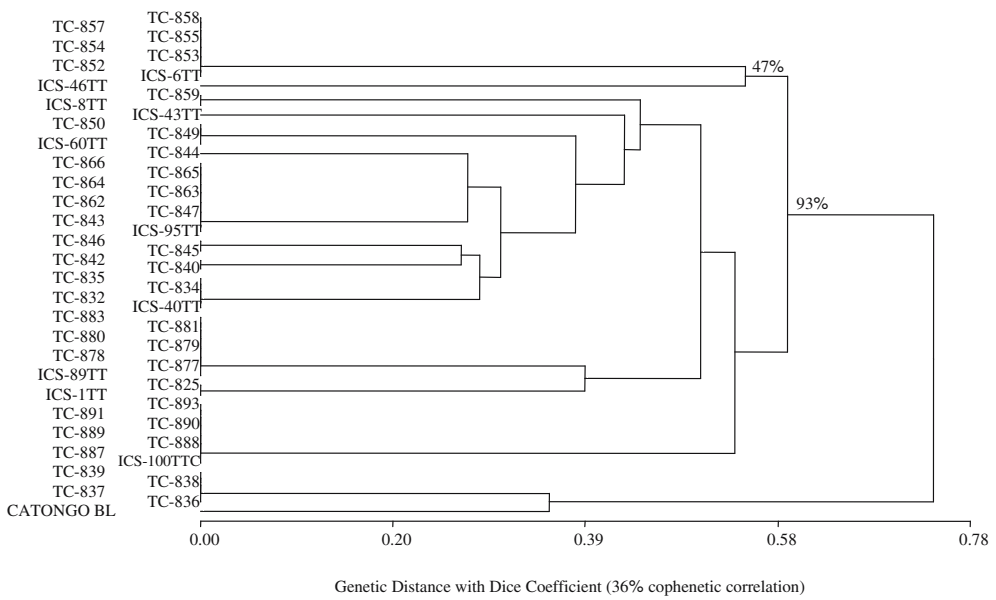

Fig. 1 Cluster analyses using Dice coefficient on 30 alleles produced by 14 primers (Set1 and Set2) for 10 Reference ICS Clones from Trinidad (clone name ending TT), 37 ICS-Type Trees, the process control Catongo Blanco and its four closely related trees all from the Cabiria field at CATIE in Costa Rica. ICS 100TTC indicates the reference genotype located in the cacao germplasm collection located on the UWI campus, St. Augustine. Trees in Cabiria that are clones of the Reference ICS genotype from Trinidad possessing the same SSR fingerprints are represented by a group or cluster joined by a bar. Catongo Blanco and four closely related trees clustered separately from the ICS-Type trees 93% of the times in 800 bootstrap permutations performed using Winboot

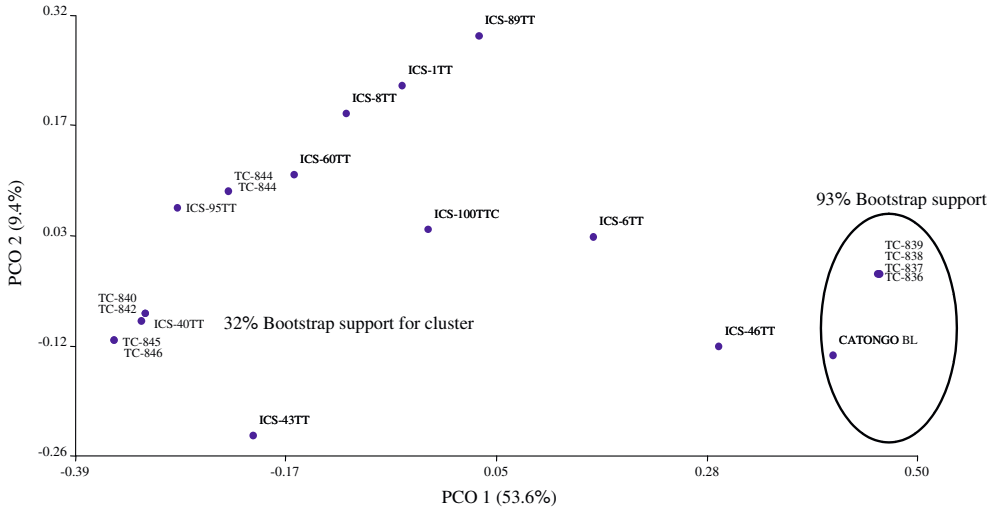

Fig. 2 Principal Coordinate Analysis with Simple matching coefficient for five trees in Cabiria closely related to the Reference ICS clones (clone name ending TT) from Trinidad and TC-836 to TC-839 closely related to Catongo Blanco. PCO (1) and (2) explains 63% of the total genetic variation. There is a 93% Bootstrap support for Catongo Blanco and the four related trees clustering separately or belonging to the Non-ICS Type sub-population. ICS-40TT clustered with its progeny (TC-840, TC-842, TC-845 and TC-846) 32% of the times in 800 bootstrap permutations

Discussion

The Field Guide in single tree identity verification

The Field Guide Concept was conceived to provide a convenient method for the rapid and accurate identification of and to facilitate efficient utilization of early generation cacao pre-breeding populations. The resulting compendium, the Field Guide of the ICS Clones of Trinidad, was successfully applied in Costa Rica and paramount to the verification of identity of the ICS Clones introduced from Trinidad over 46 years ago. The pictorial format in the Pod Image Library of the Field Guide, for the reference ICS clones, facilitated *in situ* field identification of trees in Costa Rica with mature pods. The SSR fingerprint data in the Field Guide was used for both the verification of field observations and to identify trees possessing immature or no pods at the time of field evaluation. The seasonal availability of mature pods is one of the major limitations to the use of morphological descriptors for the identification of cacao genetic resources. Although the availability of mature pods is also a limitation to the in-situ field application of the Field Guide, the SSR fingerprint data provided overcomes this problem.

A total of 69 trees in the cacao germplasm collection at CATIE in Costa Rica representing the 14 introduced ICS clones from Trinidad were analyzed in this study. Application of the Field Guide in identity verification of each tree resulted in the discovery that only 32 of these trees, from the ICS-Type sub-population in this study, were clones of eight of the original 14 introduced ICS genotypes. Mixtures of genotypes within rows were clearly identified with the two further discoveries, ICS 60 and ICS 95 are in rows designated as ICS 84 and ICS 43, respectively, in Cabiria. This could result from either mislabeling of budwood materials sent from a quarantine facility, mislabeling of the materials during grafting or remapping of the fields in Cabiria without proper documentation on field maps.

The other five trees classified in the ICS-Type sub-population were in fact progeny of ICS clones. Trees TC-845 and TC-846 differed from ICS 95 for one of the 14 primers, mTcCIR19. ICS 95 is homozygous with allele (375) for mTcCIR19, while, TC-845 and TC-846 are heterozygous (369, 375) at this locus. TC-845 and TC-846 also share the same fingerprint for all 14 primers used in this

study as does TC-840 and TC-842. Principal coordinate analyses revealed that these four trees are possibly seedlings or progeny of ICS 40. The fifth tree TC-844 is most likely a hybrid of ICS-60 and ICS-95.

Tree TC-846 was misidentified in the field in Costa Rica as ICS 95 according to the digital image and morphological descriptors for the pod of this clone in the Field Guide. As such, the use of only the pod images and morphological descriptors in the Profile of the Field Guide, can reduce the error rate in the misidentification of the ICS Clones to 3.5% from the 40% to 44% previously reported for cacao (Motilal and Butler 2003; Sounigo et al. 2000).

Both field application and use of SSR fingerprint data in the Field Guide clearly identified the 32 ''Off-Type'' trees designated as the Non-ICS Type sub-population. Genetic diversity estimates derived from six discriminant SSR primers showed that the Non-ICS Type sub-population was more genetically diverse than the Trinitario group to which the ICS clones belong. Gene frequencies in combination with the information in the Pod Image Library of the Field Guide to the ICS Clones of Trinidad, differentiated the Non-ICS Type trees in Costa Rica with 0% error. This justified the elimination of the 32 trees classified in the Non-ICS Type sub-population from complete SSR fingerprinting analyses. As a result a 46% reduction in the cost of SSR fingerprinting analyses was realized in this study using the Field Guide.

Gene frequencies between botanical groups have proven useful in the classification of recognized cacao populations. Cheesman (1944) reclassified cacao genetic resources into the botanical or horticultural groups Criollo, Upper and Lower Amazon Forastero and the hybrid Trinitario, between the Criollo and Lower Amazon groups (Laurent et al.1994). The Criollo group has been shown to have a narrow genetic base, in part, attributable to millennia of cultivation and selection for specific traits, including flavor, by pre-Colombian civilizations in Central and South America (Whitkus et al. 1998; Motamayor et al. 2002). Low genetic diversity has also been reported for the Lower Amazon Group (Laurent et al. 1993).

Not surprisingly, the resultant hybrid Trinitario group, also subjected to over half a millennium of farmer selection (Cope and Bartley 1954), has been shown to possess a narrow genetic base compared to the Forastero group (N'Goran et al. 2000; Motamayor et al. 2003). A low level of genetic diversity also proved useful in the differentiation of the original Ecuadorian Nacional cacao, also cultivated and selected over millennia by the Incas, from the more genetically diverse hybrids formed between this group and Trinitario materials introduced in 1890 (Lerceteau et al. 1997). However, the hybrid nature of the Trinitario group makes them very variable in appearance making identification based solely on morphological descriptors difficult (Figueira et al. 1994).

Molecular or SSR fingerprinting of cacao genetic resources is highly desirable but can become cost prohibitive when considering SSR fingerprinting of each tree representing an accession in cacao germplasm collections. The application of the Field Guide in this study cut the cost of SSR fingerprinting by approximately half for the verification of single tree identity. The USDA in Beltsville, MD has undertaken the rationalization of the 3,050 cacao accessions currently maintained in the two Universal Collection Repositories recognized by the IBPGR (Saunders et al. 2004). This is of tremendous value and an admirable service to the cacao research community. However, the full benefits of this effort will not be realized if the necessary steps are not taken to correct the problems identified in the collections.

In this aspect CATIE has taken the lead to reduce the probability of future errors being committed when using the ICS clones in their collection. Each of the 32 trees verified as a clone of one of the eight reference ICS clone genotypes from Trinidad has been individually labeled. A permanent aluminum tag painted with blue automotive paint is being used as a long-lasting field label, stating the name of the clone, tree number in the row and field location. This is to designate that the genotype of these 32 trees were verified by SSR fingerprinting and morphological descriptors to a Reference genotype from the country of origin, Trinidad in this

case. The Off-Type trees, identified in this study, will be removed from within rows for re-planting and possible re-labeling (Turnbull et al. 2004) in another field.

The lengthy process of cacao germplasm introduction and costs incurred over years of maintenance in a living tree gene bank makes chopping down healthy cacao trees in a collection a very difficult decision. The alternative is to relocate trees identified as ''Off-types'' to another site in the collection. The Field Guide format aids in the detection of ''Off-type'' trees that are not the original or reference genotype and as such provides the justification for tree excision within rows. The Field Guide is also useful for verification of clones during replication within a country.

The second interesting outcome of this study concerns the report of resistance to Monilia disease in both ICS 95 and ICS 43 in Cabiria (Phillips 1996). The results of this study demonstrated that the accession believed to be ICS 43, is in fact ICS 95, which explains its resistance to Monilia or frosty pod rot (*Moniliophthora roreri*). The discovery of frosty pod rot resistance in the ICS population is significant and further supports the potential and utility of these early generation selections to current cacao breeding programs. The report of resistance in two ICS clones, when in fact they are one and the same, further emphasizes the importance of the ongoing USDA program to rationalize cacao germplasm collections if progress is to be made in the genetic improvement of cacao varieties for agronomic traits of interest.

Encountering frosty pod rot resistance in the ICS population may not be as surprising when considering the Forastero and Criollo heritage in Trinitario materials and high tolerance to witches' broom disease. Vegetative susceptibility with good pod resistance was reported (Cheesman 1947) and observed to still hold in ICS 60 after 60 years of disease pressure, at the same location, in the San Juan Estate (E. Johnson field observation at San Juan Estate in 2001). It was recently shown that the basidiomycetes causing both witches' broom and frosty pod rot diseases are very closely related with the suggestion that *Crinipellis perniciosa* should be renamed *Moniliophthora perniciosa* (Aime and Phillips-Mora 2005).

Field Guides in assigning reference genotypes for widely distributed cacao accessions

The problems encountered in the ICS materials in Costa Rica presents one scenario for widely distributed cacao accessions planted around 50 years ago in cacao germplasm collections worldwide. The problems that may be encountered in germplasm collections include lost of or incomplete passport data, mislabeling or inadequate labeling of field plots/rows or of trees within a plot/row. Mislabeling or insufficient labeling of harvested and in-process propagation materials also results in poor record keeping of the movement of germplasm. The situation can be further complicated by out dated or under utilization of field maps and lack of use of available descriptor information. Mixtures of genotypes within a row representing a cacao accession is known to occur which could be a consequence of re-planting to replace dead trees to maintain the canopy.

In addition to these technical issues, there are logistical problems including the longevity of a cacao tree which will usually outlast the lifetime career of researchers, changes in focus and personnel in cacao conservation and breeding programs and discontinuous and/or limited funding for the maintenance of this type of genetic resource. These factors have resulted in alternating periods of good and poor maintenance of cacao germplasm collections, which all contributes to an accumulation of misidentification errors in cacao accessions.

The misidentification problem is perpetuated and compounded when budwood is harvested from multiple trees within a row for renovation, duplication and exchange of an accession, or flowers collected, again from multiple trees within a row, to be used as the pollen parent when making crosses. Therefore a multitude of and/or different combinations of factors could lead to the misidentification of cacao germplasm in a given collection. The consequence is a unique history behind each widely distributed cacao accession and even for different assemblages of accessions or field blocks within a collection. This makes broad generalizations as to the level of misidentification in the collection as a whole invalid.

The problems cited are not unique to this crop. As such, older plantings of cacao germplasm cannot be destroyed because of misidentification problems, until the original or ''Reference'' genotype of each cacao accession has been assigned. These older plantings of cacao germplasm maybe the source of, or greatly assist in the assignment of the Reference genotype for an accession; case in point, ICS 100 from Trinidad.

There are two genotypes called ICS 100 in Trinidad, one located in Block 5 at the San Juan Estate and is in the Field Guide, the other genotype is located in Campus Field 4 at the University campus of the University of the West Indies (UWI), St. Augustine. The results of this study show that the ICS 100 in Costa Rica is the same as the genotype at the UWI campus. The ICS Clones at CATIE were introduced into Cabiria from Trinidad in 1959, after characterization in Trinidad during the period 1945–1951. Combining the results of morphological descriptors, SSR and cluster analyses with that of the history of the ICS Clones, it stands to reason that the ICS 100 genotype on the UWI campus is most likely the original and should be designated as the ''Reference'' genotype for ICS 100.

The search for or designation of a ''Reference'' genotype for each accession of cacao maintained in the two Universal Collection Repositories is currently being undertaken as a joint initiative between the USDA and its collaborators worldwide. The previous ICS 100 example may prove useful in locating and/or designating the original or ''Reference'' genotype for some of the most widely distributed cacao accessions. The relevance of this initiative cannot be overstated as the use of incorrect genotypes has resulted in limited progress in cacao breeding programs. A ''Reference'' genotype will also facilitate the use of molecular fingerprinting for genotype verification during any or all stages of multiplication and movement of cacao germplasm.

Acknowledgements The authors wish to thank staff members of CATIE for their kind cooperation in making this work possible. In particular, Allan Meneses financially supported by this project for technical assistance with field and molecular work and Dr. Wilbert Phillips for reviewing this article. The assistance of Dr. Fernando Casanoves in the precise application of InfoGen/P for performing analyses is greatly appreciated. This work was funded by CABI Bioscience and the USDA.

References

Aime MC, Phillips-Mora W (2005) The causal agents of witches' broom and frosty pod rot of cocoa (chocolate, *Theobroma cacao*) form a new lineage of Marasmiaceae. Mycologia 97:1012–1022

Bekele F (1992) Use of botanical descriptors for cocoa characterization: CRU experiences. In: Proceedings of the International Workshop on Conservation, Characterization and Utilization of Cocoa Genetic Resources in the 21st Century. CRU, Port-of-Spain, Trinidad, September 13–17, pp 77–91

Bekele F, Butler DR (1998) Proposed short list of cocoa descriptors for characterization. Working procedures for cocoa germplasm evaluation and selection. In: Eskes AB, Engels JMM, Lass RA (eds) Proceedings of the CFC/ICCO/IPGRI Project Workshop, CIRAD, Montpellier, France, February 1–6, 1998, pp 41–48

Bekele FL, Bekele I, Butler DR, Bidaisee G (2006) Patterns of morphological variation in a sample of cacao (*Theobroma cacao* L.) germplasm from the International Cocoa Genebank, Trinidad. Genet Resour Crop Evol 53:933–948

Bowers JH, Bailey BA, Hebbar PK, Sanogo S, Lumsden RD (2001) The impact of plant diseases on world cocoa production. Online. Plant Health Progress doi:10. 1094/PHP-2001-0709-01-RV. http://www.apsnet.org/online/feature/cacao

Cheesman EE (1944) Notes on the nomenclature, classification and possible relationships of cacao populations. Trop Agric 21:144–159

Cheesman EE (1947) Results of cacao experiments in 1944–45. Trop Agric 23:63–65

Cope FW, Bartley BGD (1954) Some aspects of the plant improvement programme. Imperial College of Tropical Agriculture. Ann Rep Cocoa Res 1954:27–31

Engels JMM (1992) The use of botanical descriptors for cacao characterization: CATIE experiences. In: Proceedings of the International Workshop on Conservation, Characterization and Utilization of Cocoa Genetic Resources in the 21st Century. CRU, Port-of-Spain, Trinidad, September 13–17, pp 69–76

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

Frison EA, Diekman M, Nowell D (2000) FAO/IPGRI Technical guidelines for the safe movement of germplasm No. 20. Cacao. FAO, Rome/International Plant Genetic Resources Institute, Rome, Italy, p 32. ISBN 92-9043-399-6

Greathouse DC, Laetsch WM, Phinney BO (1971) The shoot-growth rhythm of a tropical tree (*Theobroma cacao* L.). Amer J Bot 58:281–286

IBPGR (1981) Genetic resources of cocoa (AGP:IBPGR/80/56). IBPGR Working Group on Genetic Resources of Cocoa, IBPGR Secretariat, Rome, Italy, 25 pp

ICGD (2003) http://www.icgd.rdg.ac.uk/

InfoGen/P (2003) InfoGen/P statistical software version 1.0. Universidad Nacional de Córdoba, Argentina

Johnson ES, Bekele FL, Schnell RJ (2004) Field guide to the ICS clones of Trinidad. CATIE, Serie Técnica Manual Técnico, No. 54. ISBN 9977-57-398-0, 32 pp

Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, Lagoda PJL (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. Mol Ecol 8:2141–2143

Laurent V, Risterucci AM, Lanaud C (1993) Variability for nuclear ribosomal genes within *Theobroma cacao*. Heredity 71:96–103

Laurent V, Risterucci AM, Lanaud C (1994) Genetic diversity in cocoa revealed by cDNA probes. Theor Appl Genet 88:193–198

Lerceteau E, Quiroz J, Soria J, Filpo S, Pétiard V, Crouzilat D (1997) Genetic differentiation among Ecuadorian *Theobroma cacao* L. accessions using DNA and morphological analyzes. Euphytica 95:77–87

Lindeman HR, Merenda PF, Gold RZ (1980) Discriminant analysis. In: Introduction to bivariate and multivariate analysis. Scott, Foresman and Company, Palo Alto, California, USA, pp 170–219

Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. Heredity 89:380–386

Motamayor JC, Risterucci AM, Heath M, Lanaud C (2003) Cocoa domestication II. Progenitor germplasm of the Trinitario cacao cultivar. Heredity 91:322–330

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

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci 70:3321–3323

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

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

Phillips W (1996) Studies at CATIE on Moniliasis resistance (*Moniliophthora roreri* (Cif. & Par.) Evans et al.). In: Proceedings of the International Workshop on the Contribution of Disease Resistance to Cocoa Variety Improvement. Salvador, Brazil, November 24–26, pp 111–117

Pugh T, Fouet O, Risterucci AM, Brottier P, Abouladze M, Deletrez C, Courtois B, Clement D, Larmande P, N'Goran JAK, Lanaud C (2004) A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. Theor Appl Genet 108:1151–1161

Purdy LH, Schmidt RA, Gramacho KP (1998) Common names of plant diseases: diseases of Cacao (*Theobroma cacao* L.). http://www.apsnet.org/online/common/names/cacao.asp

Ronning CM, Schnell RJ (1994) Allozyme diversity in a germplasm collection of *Theobroma cacao*. L. J Heredity 85:291–295

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

Schnell RJ, Heath MA, Johnson ES, Brown JS, Olano CT, Motamayor JC (2004) Frequency of off-type progeny among the original ICS 1 X SCA 6 reciprocal families and parental clones used for disease resistance selection in Trinidad. INGENIC Newslett 9:34–39

Sounigo O, Christopher Y, Bekele F, Mooledhar V, Hosein F (2000) The detection of mislabeled trees in the International Cocoa Genebank, Trinidad (ICG,T) and options for a global strategy for identification of accessions. In: INGENIC Workshop from Proceedings of the International Workshop on New Technologies and Cocoa Breeding. Kota Kinabalu, Sabah, Malaysia, October 16–17, pp 34–39

Turnbull CJ, Butler DR, Cryer NC, Zhang D, Lanaud C, Daymond AJ, Ford CS, Wilkinson MJ, Hadley P (2004) Tackling mislabeling in cocoa germplasm collections. INGENIC Newslett 9:8–11

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

Yap IV, Nelson RJ (1996) WinBoot: UPGMA bootstrapping for binary data. IRRI Discussion Paper, Series No. 14, pp 20