RESEARCH

Population Structure and Genetic Diversity of the Trinitario Cacao (*Theobroma cacao* L.) from Trinidad and Tobago

Elizabeth S. Johnson,* Frances L. Bekele, Steven J. Brown, Qijian Song,
Dapeng Zhang, Lyndel W. Meinhardt, and Raymond J. Schnell

ABSTRACT

Population structure of the old Trinitario cacao (*Theobroma cacao* L.) in Trinidad was evaluated from the 35-microsatellite multilocus profile of 32 relict clones from abandoned cocoa estates in Trinidad (TRD) and 88 Imperial College selections (ICS) clones conserved in the International Cocoa Genebank, Trinidad. Ancestry was derived by comparison to 34 representative genotypes of Criollo, Central American, and Lower and Upper Amazon origin. The 154 individuals separated into four populations in this study, labeled Population 1—Upper Amazon; Population 2—Ecuadorian; Population 3—Lower Amazon; and Population 4—Trinitario. Over 90% of the individuals demonstrated 70 to 99% membership to their respective populations. The ICS and TRD clones formed a genetically homogenous group, 84% of which clustered in Population 4 with clones of pure Criollo descent. Another 10% clustered in Population 1 containing clones of Upper Amazon descent. The cacao population of Trinidad is genetically different from clones of Ecuadorian and Lower Amazon descent, Populations 2 and 3, respectively. The 10 most discriminating loci for each population were different as shown by polymorphism information content values for each of the 35 microsatellite loci. This is the first report to provide a strong genetic basis for the industry flavor distinction of Trinitario and Ecuadorian cacaos. These findings significantly impact germplasm fingerprinting and curatorship and maintenance of industry flavor classes while breeding for disease resistance in cacao.

E.S. Johnson, D. Zhang, and L.W. Meinhardt, USDA-ARS PSI SPCL, Room 223, Bldg. 001 BARC-WEST, 10300 Baltimore Ave., Beltsville, MD 20705; E.S. Johnson, current address: CABI Caribbean and Latin America, Gordon Street, Curepe, Trinidad and Tobago, W.I.; F.L. Bekele, Cocoa Research Unit, The Univ. of the West Indies, St. Augustine, Republic of Trinidad and Tobago, W.I.; S.J. Brown and R.J. Schnell, USDA-ARS Subtropical Horticultural Research Station, 13601 Old Cutler Rd., Miami, FL 33158; Q. Song, Plant Science & Landscape Architecture, Univ. of Maryland, College Park, MD 20742, c/o Soybean Genomics and Improvement Lab., Beltsville Agricultural Research Center, USDA-ARS, Beltsville, MD 20705. Received 5 Mar. 2008. *Corresponding author (l.johnson@cabi.org).

Abbreviations: IC3, International Cacao Collection, CATIE, Costa Rica; ICGT, International Cocoa Genebank, Trinidad; ICS, Imperial College selections; PCR, polymerase chain reaction; PIC, polymorphism information content; SJE, San Juan Estate; SSR, simple sequence repeat; TRD, relict clones from abandoned cocoa estates in Trinidad; TSH, Trinidad Selected Hybrids; TT_UCRS, Trinidad and Tobago_ University Cocoa Research Station; UWI, University of the West Indies, St. Augustine campus.

CACAO BEANS ARE THE SOURCE of the raw materials, cocoa powder and cocoa butter, used in the multibillion dollar chocolate, confectionery, and cosmetic industries. The fine chocolate niche market requires cacao beans with distinctive flavor characteristics that are a culmination of genotype, growing region, bean fermentation, and drying processes, to name a few.

The Imperial College selections (ICS) of Trinidad and Tobago were made by F.J. Pound from 1933 to 1935 (Posnette, 1986) and recognized for their distinctive Trinitario flavor. The ICS clones are among the most widely distributed and utilized cacao

---

Published in Crop Sci. 49:564–572 (2009).
doi: 10.2135/cropsci2008.03.0128

677 S. Segoe Rd., Madison, WI 53711 USA

germplasm accessions, and form the base of the Trinitario cacao population in existence at present. A total of 100 selections (labeled ICS-1 to ICS-100) were made after a survey of >50,000 preselected trees estimated to be 15 to 30 yr old and in production on cocoa estates in Trinidad at that time. Thirty-six percent were selected for high yield and good quality measured by pod index, annual yield, and bean dry weight >1.0 g. The other 64 selections were made based on yield and quality, with emphasis placed on disease tolerance and representation of the observed genetic diversity in the Trinitario population (Pound, 1934, 1935, 1936).

At the time F.J. Pound made his selections, the Trinidad cacao population or Trinitario complex showed considerable variation for yield and susceptibility to witches' broom disease (Pound, 1931; Montserin et al., 1957). It was described as a composite of descendants of hybrids between the numerous introductions to the island, which had undergone >500 yr of cultivation, and farmer selection (Cope and Bartley, 1954). Ciferri and Ciferri (1957) surmised that the distinction between what was called Forastero cacao of Trinidad and Criollo cacao of Venezuela was really based on the commercial bean grading system in Venezuela and that the two populations actually shared a common genealogy.

Diseases have also been a major biotic factor influencing the genetic makeup of the Trinitario cacao population of Trinidad and Tobago. By the last quarter of 1928, witches' broom disease, caused by the fungus *Moniliophthora perniciosa* (Stahel) (Aime and Phillips-Mora, 2005), was impacting yields, as >28% of the cacao cultivations were infected in Trinidad (Bekele, 2004). The ICS selections were made to obtain the best individuals to use as parents in a witches' broom disease resistance breeding program, with resistant cacao that would be introduced by Pound from Peru and Ecuador (Upper Amazon) in 1937–1938 and 1942–1943. The best performers were identified by vegetative cloning of the 100 ICS trees and screening in replicated regional trials in Trinidad from 1939 to 1945 (Cheesman, 1947). Two clones in particular, ICS-1 and ICS-95, along with other ICS clones, outperformed the rest in these trials and were hybridized with selections made from the Upper Amazon cacaos from Peru. These selections were based on demonstrated resistance to witches' broom disease from screening under constant high levels of inoculum pressure at the Marper Farm (Montserin et al., 1957).

This effort resulted in the launch of the Trinidad Selected Hybrids (TSH) cacao breeding program in 1949 by the then Ministry of Agriculture (currently Ministry of Agriculture Land and Marine Resources) of Trinidad. By 1965, >1.5 million clones of TSH cultivars with resistance to witches' broom were produced and distributed to the cacao farmers of Trinidad and Tobago. Currently, the TSH clones along with the old Trinitario cacao produce the distinctive Trinitario cocoa flavor classified as a fine or flavor cocoa by the chocolate industry. It is described as fruity with floral attributes and pleasant ancillary flavors such as molasses, licorice, caramel, and raisin, for which farmers receive a premium price (Bekele, 2004; Sukha and Butler, 2005). The TSH cacao breeding program was (Kennedy et al., 1987) and remains one of the most successful and long-lived cacao breeding programs in the history of the crop. Previous to the TSH clones, the fine flavor classification of Trinidad cacao was produced by the base population of the ICS clones.

In 1991–1992, John Warren and Thakurani Cassie attempted to recapture some of this base population by collecting on abandoned cocoa estates in Trinidad. The aim was to collect 10 trees from cocoa estates located in the producing areas at the four cardinal points of the island, including Lopinot, Sangre Grande (area in which witches' broom was first reported in 1928), Cumaca, Blanchisseuse, Maracas (St. Joseph), Mount St. Benedict, Tabaquite, and Rio Claro. A total of 119 trees were collected, of which 68 are currently conserved in the International Cocoa Genebank, Trinidad (ICGT) (Bekele et al., 2007).

Frosty pod rot disease of cacao, caused by the basidiomycete *Moniliophthora roreri* (Cif.) H.C. Evans, Stalpers, Samson & Benny, is now in its invasive phase, rapidly spreading through South and Central America and working its way up through Mesoamerica (Aime and Phillips-Mora, 2005; Phillips-Mora et al., 2006a, 2006b). The most cost-effective and environmentally friendly strategies to control the spread of frosty pod rot involve the development of new cacao cultivars with resistance to frosty pod, as was done for witches' broom. Two clones, ICS-95 and UF-273 Type 1, have been shown to be resistant to the seven known isolates of the fungus (Phillips-Mora et al., 2005) and are currently being used to incorporate resistance to frosty pod into susceptible cultivars.

Industry and consumer cocoa flavor preferences determine the economic returns to farmers. Information on the genetic structure of specific cacao populations can assist breeders in selecting parents to incorporate desired traits without losing established flavor characteristics of existing cultivars. This information will also assist germplasm curators to maximize genetic diversity in a minimum number of conserved genotypes.

The first objective of this study was to establish the population structure and determine the ancestry of the old Trinitario cacao population of Trinidad. The second objective was to assess the possibility of obtaining a minimum number of microsatellite loci to discriminate among the closely related clones. The third objective was to assess the level of genetic diversity captured in the conserved ICS clones and relict clones from abandoned cocoa estates in Trinidad (TRD) to reduce genetic redundancy in the ICGT.

MATERIALS AND METHODS

Plant Materials

A total of 154 genotypes, comprising 88 ICS clones, 32 TRD clones from the ICGT, and 34 accessions as representatives of Criollo, and Lower and Upper Amazon Forastero and Trinitario types of cacaos spanning a diverse range of origin (Supplemental Table 1) were analyzed. DNA samples of FSC-15, SIAL-169_Al, SIAL-169_CEP, SIAL-70, SIC-19, SIC-23, SIC-328, SIC-801, SIC-806, numbered 132 to 140, respectively (Supplemental Table 2), were kindly provided by J.C. Motamayor of MARS Inc./USDA, Miami, FL. The rest of the accessions were sampled from the International Cacao Collection, CATIE, Costa Rica (IC3).

Table 1. Properties and polymorphism information content (PIC) of the 35 microsatellite primers used to determine the genetic makeup of the Trinitario cacao population in Trinidad. PIC-1 through PIC-4 list the values for Populations 1 through 4, respectively, with the ancestry of the members represented as UAF (Upper Amazon), ECU (Ecuadorian), LAF (Lower Amazon), and TRI (Trinitario). T_a = annealing temperature.

GenBank code	Locus	Linkage group	T_a	Size range	No. of alleles per locus	UAF origin PIC-1	ECU origin PIC-2	LAF origin PIC-3	TRI origin PIC-4
			°C	bp					
Y16883	mTcCIR1†	8	51	122–156	4	0.19	**0.25§**	**0.12§**	**0.37§**
Y16977	mTcCIR3	2	46	196–279	12	0.14	0.16	**0.22§**	0.23
Y16980	mTcCIR6†	6	46	218–251	10	0.19	**0.23§**	0.12	0.19
Y16982	mTcCIR8†	9	46	276–322	7	0.2	**0.24§**	0.07	**0.25§**
Y16983	mTcCIR9	6	51	250–297	11	0.22	0.19	0.07	0.23
Y16984	mTcCIR10	5	46	200–218	5	**0.28§**	0.22	**0.12§**	**0.25§**
Y16986	mTcCIR12†	4	46	187–273	9	0.23	0.21	**0.16§**	0.24
Y16988	mTcCIR15†	1	46	220–256	9	0.16	0.18	**0.29§**	**0.30§**
Y16990	mTcCIR17	4	51	256–295	5	**0.28§**	0.2	0.07	**0.37§**
Y16991	mTcCIR18†	4	51	328–354	8	0.15	0.19	0.07	0.19
Y16992	mTcCIR19	2	46	343–387	10	0.15	0.18	**0.20§**	0.16
Y16994	mTcCIR21	3	46	138–182	9	0.19	0.14	0.07	**0.25§**
AJ271995	mTcCIR22†	1	46	270–292	5	0.22	0.21	0.07	**0.37§**
Y16996	mTcCIR24†	9	46	184–206	6	**0.26§**	0.2	M¶	0.23
Y16997	mTcCIR25	6	46	126–172	11	0.17	0.16	0.07	0.24
Y16998	mTcCIR26†	8	46	292–346	7	**0.25§**	0.22	0.12	**0.36§**
AJ271826	mTcCIR33†	4	51	294–348	11	0.18	0.17	0.07	0.1
AJ271827	mTcCIR35	9	46	214–242	5	**0.26§**	0.23	M	0.23
AJ271942	mTcCIR37†	10	46	132–184	11	0.16	0.2	0.07	0.14
AJ271946	mTcCIR44	2	51	165–177	3	0.22	**0.28§**	0.07	0.25
AJ271948	mTcCIR46	7	46	184–213	7	0.19	**0.28§**	0.12	0.17
AJ271952	mTcCIR53	6	51	106–153	13	0.14	0.15	**0.23§**	0.15
AJ271953	mTcCIR54	1	46	139–171	5	0.24	0.23	0.12	0.24
AJ271955	mTcCIR56	7	46	310–368	6	0.2	0.2	0.12	0.18
AJ271958	mTcCIR60†	2	51	187–210	8	0.2	**0.24§**	**0.17§**	0.25
AJ566451	mTcCIR107	4	46	109–123	4	**0.28§**	**0.26§**	0.07	0.25
AJ566488	mTcCIR157	9	48	142–178	5	**0.26§**	0.2	M	0.21
AJ566490	mTcCIR160	9	48	289–293	2	**0.37§**	**0.36§**	M	M
AJ566544	mTcCIR223	10	46	175–199	6	0.22	0.18	**0.27§**	0.19
AJ566553	mTcCIR232	5	51	181–218	6	0.23	0.2	0.07	0.02
AJ566560	mTcCIR239	5	46	199–205	5	0.2	**0.29§**	M	0.31§
AJ566570	mTcCIR249	1	48	171–243	19	0.15	0.12	**0.17§**	0.11
AJ566584	mTcCIR264	1	46	189–234	8	0.16	0.22	0.12	0.14
AJ566594	mTcCIR275	1	53	140–150	3	**0.28§**	**0.28§**	0.07	0.25
unknown	SHRSTc11‡	5	51	298–312	5	**0.26§**	0.23	0.07	**0.33§**
Total					260				

†Primers developed by CIRAD and used in routine fingerprinting of cacao germplasm.

‡Primer developed by USDA-SHRS in Miami. Forward: 5′-CCACCACCCTTACCTTTGAAAC-3′; Reverse: 5′-ACGAGACTCCAAAACGGACG-3′.

§Ten most informative simple sequence repeat loci in each population.

¶M signifies all alleles for this primer were monomorphic in this population.

DNA Extraction, PCR Amplification, and Capillary Electrophoresis

DNA extraction, polymerase chain reaction (PCR) amplification and microsatellite marker electrophoresis, allele calling, and binning were as previously described (Johnson et al., 2007). The PCR amplifications were performed with 35 fluorescent-labeled microsatellite primers (Table 1). Twelve of the primers are being used routinely for DNA fingerprinting of cacao (Saunders et al., 2004; Zhang et al., 2006). The primers, developed by CIRAD (Lanaud et al., 1999; Pugh et al., 2004), were used in 10- or 25-μL PCR reactions, for single and multiplexed reactions, respectively, containing at final concentration, 1× PCR buffer with 15 mM of $MgCl_2$ (Applied Biosystems, Inc., Foster City, CA), 200 μM of dNTP, 1 mg of bovine serum albumin, 200 nM of each forward and reverse primer, 0.5 U of Ampli*Taq* DNA polymerase (Applied Biosystems, Inc.), and 2.5 ng of DNA template.

The PCR amplifications were performed according to the following cycling protocol: 1 cycle 94°C for 4 min, 32 cycles 94°C for 30 sec, primer annealing temperature (Table 1) 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 a final 4°C holding cycle. For electrophoretic separation of each sample, 1 μL of amplification product, 12 μL of Hi-Di Formamide, and 0.2 μL of Rox 500 size standard (Applied Biosystems, Inc.) was denatured at 95°C for 5 min and placed immediately on ice. Capillary electrophoresis was performed on the ABI 310 automated sequencer (Applied Biosystems, Inc.) in a 36-cm capillary using POP4 polymer at 60°C, injection parameters of 15 kV for 5 sec, and run parameters of 15 kV.

Population Structure and Genetic Diversity Analyses

Microsatellite allele calling and sizing were accomplished using GeneScan and Genotyper 3.7 software (Applied Biosystems, Inc.) as previously described (Johnson et al., 2007). Population structure was inferred from the 35-microsatellite multilocus profiles of all 154 individuals with the software Structure (Pritchard et al., 2000; Pritchard and Wen, 2003). An admixture ancestry model in Structure was used that best deals with individuals of mixed ancestry, a common occurrence in cacao germplasm. The model assumes that any individual could have inherited some fraction of its genome from any of the K populations. Three independent runs were performed for each value of K between 2 and 7, using a burn-in period of 50,000 for 1,000,000 replications. The proportion of membership for each accession (Supplemental Table 2) to a predefined population was used for the calculation of distance among accessions. The distance matrix was estimated as $d_{ij} = 1 - (r_{ij} + 1)/2$, where r_{ij} is Pearson's correlation coefficient of the ith and jth accessions based on the proportion of membership. The consensus tree (Fig. 1) depicting the interpopulation genetic relationship was constructed with the unweighted pair group method with arithmetic mean, using the software Mega V3.1 (Kumar et al., 2004).

Table 2. Intrapopulation genetic diversity estimates for the 154 cacao individuals analyzed by 35 simple sequence repeat primers. Friedman's test shows the statistical significance of differences measured for the Nei's genetic diversity index between the four populations inferred by Structure software. Standard errors estimated by Bootstrap for 1000 simulations.

Population	*N*[†]	Nei's unbiased gene diversity	Friedman's test[‡]	Average no. of alleles	Effective no. of alleles	Percent polymorphic bands
2	15	0.462 ± 0.005	A	2.988 ± 0.044	1.860 ± 0.024	78
1	19	0.452 ± 0.006	B	2.912 ± 0.05	1.850 ± 0.025	69
4	107	0.432 ± 0.006	B	2.808 ± 0.058	1.846 ± 0.028	50
3	13	0.421 ± 0.008	C	2.419 ± 0.046	1.748 ± 0.028	29

[†] Number of individuals in each population.

[‡] Letters indicate significance at (P < = 0.05). Friedman's P = 0.0000.

Intrapopulation genetic diversity measures (Nei, 1973) for the populations identified by Structure were estimated with the software Info-Gen Pro (INFO-GEN, 2003). The 10 most informative microsatellite loci for fingerprinting and distinguishing between populations were derived from the highest polymorphism information content (PIC) values, estimated according to Botstein et al. (1980) for each of the 35 loci with Info-Gen Pro.

Genetic Diversity Management in Cacao Germplasm Collections

The occurrence of full-sib families among the ICS and TRD clones in Population 4 was assessed with the software Colony (Wang, 2004). A total of 97 genotypes, with one representative of identical genotypes included, were analyzed, setting parents as unknown for three independent runs to ensure repeatability of results.

The identity of ICS-83, -87, -95, and -98 were verified at multiple locations within the ICGT collection by comparing their multilocus marker profiles with Info-Gen Pro (Balzarini and Di Rienzo, 2003; INFO-GEN, 2003) and proportion of membership from Colony.

RESULTS

Population Structure

A total of 260 alleles were identified from the 35 microsatellite loci in the 154 genotypes. The parameters used in Structure software for this data produced the highest log-likelihood scores when K was set to 4, or assigned the 154 genotypes to four populations based on allele frequencies at each of the 35 microsatellite loci. The four populations (labeled 1–4) in this study were composed of 19, 15, 13, and 107 individuals, respectively, with Population 4 containing 84%, Population 1 holding 10%, and Population 2 the rest of the ICS and TRD clones. Population 3 was composed solely of individuals of Lower Amazon origin. Populations 1 and 2 were composed principally of individuals of Upper Amazon and Ecuadorian ancestry,

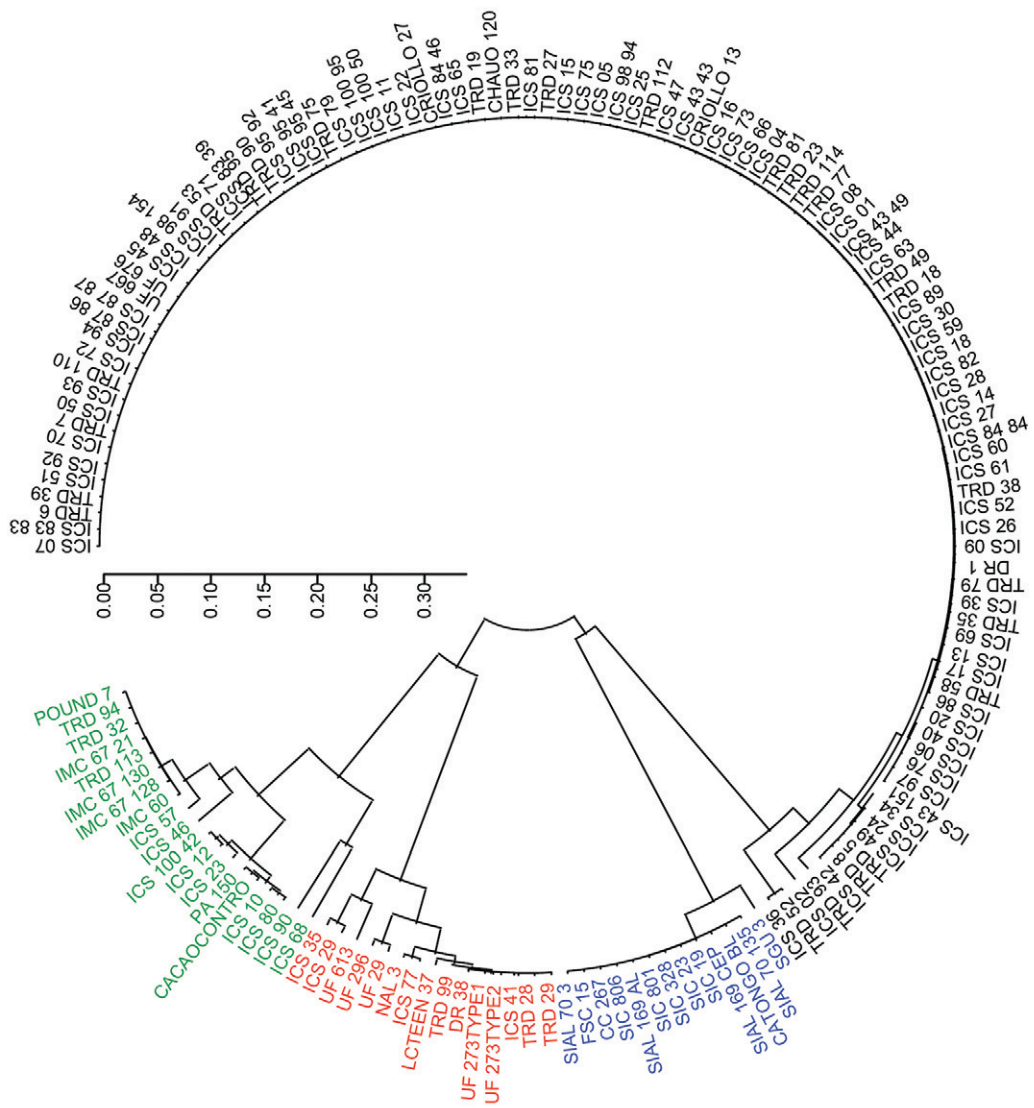

Figure 1. Consensus tree depicting membership of 154 cacao genotypes to the four populations derived from analyses with Structure software using 35-microsatellite multilocus profiles. Population 1 in green—Upper Amazon ancestry, Population 2 in red—Ecuadorian ancestry, Population 3 in blue—Lower Amazon ancestry, and Population 4 in black—Criollo and Trinitario ancestry.

respectively. Individuals of Criollo, Trinitario, and Venezuelan ancestry all clustered in Population 4.

The proportion of membership in each of the four populations, for each accession, is presented in Supplemental Table 1. Individuals sharing a common ancestry show high membership to a population, while more admixed individuals have a more distributed membership among populations. The majority (>90%) of the individuals in this study were clearly assigned to a single population, with >70% of their inferred ancestry derived from a recognized cacao population (Cheesman, 1944). There were only six individuals of admixed ancestry (ICS-10, ICS-46, ICS-90, ICS-35, SGU-3, and UF-613) assigned to the population to which they possess the highest membership for further analyses. The proportion of membership was used to generate the distance matrix for the consensus tree in Fig. 1. This is the first report to graphically illustrate the degree of relatedness among individuals within populations by the length of the branches in a tree format, instead of the normal color bar graph output of Structure.

Intrapopulation Genetic Diversity

Intrapopulation genetic diversity was estimated by the average number of alleles, the effective number of alleles, and the unbiased genetic diversity (Nei, 1973). The genetic diversity estimates, obtained from the 35 simple sequence repeat (SSR) loci, is not significantly different for Populations 1 and 4 as shown by Friedman's test (Friedman, 1937) in Table 2. The statistical significance of the difference in genetic variation among Populations 1 and 4 to Population 3 is further emphasized by the percentage of polymorphic loci in Population 3 being approximately half of that in Populations 1 and 4.

The PIC value is a measure of the ability of the alleles of a marker to detect differences. The 10 most informative SSR markers with the highest potential for intrapopulation

Table 3. List of full-sib families and identical sets obtained over three independent runs using COLONY software and individuals possessing the same simple sequence repeat (SSR) profile (grouped by profile in each identical set) determined for 35 SSR loci primers.†

Full-sib Family 1	Full-sib Family 2	Full-sib Family 3
ICS-39[TT_ICG]	TRD-5[TT_UCRS]	TRD-6[TT_UCRS]
ICS-40[TT_SJE]	TRD-8[TT_UCRS]	TRD-18[TT_UCRS]
ICS-45[TT_ICG]	TRD-53[TT_UCRS]	TRD-110[TT_UCRS]
ICS-47[TT_ICG]	TRD-58[TT_UCRS]	TRD-112[TT_UCRS]
ICS-48[TT_SJE]		TRD-114[TT_UCRS]
ICS-60[TT_SJE]		
ICS-61[TT_UWI]		
ICS-84[CR_LotE]		
TRD-35[TT_UCRS]		
Full-sib Family 4	**Full-sib Family 5**	**Full-sib Family 6**
ICS-43[CR_LotE]	ICS-52[TT_ICG]	ICS-43[CR_Cab.2]
ICS-44[TT_SJE]	ICS-76[TT_ICG]	ICS-95[CR_LotE]
ICS-63[TT_SJE]	TRD-38[TT_UCRS]	
Full-sib Family 7	**Full-sib Family 8**	**Full-sib Family 9**
ICS-14[TT_SJE]	TRD-77[TT_UCRS]	ICS-07[TT_SJE]
ICS-27[TT_ICG]	TRD-79[TT_UCRS]	ICS-97[TT_SJE]
Full-sib Family 10	**Full-sib Family 11**	**Full-sib Family 12**
ICS-18[TT_SJE]	ICS-91[TT_UWI]	ICS-20[TT_SJE]
ICS-93[TT_SJE]	ICS-98[TT_SJE]	ICS-43[TT_SJE]

Identical Set 1	Identical Set 2	Identical Set 3
ICS-87[TT_SJE]	ICS-100[CR]	ICS-83[TT_SJE]
ICS-87[TT_UWI]	ICS-79[TT_UWI]	ICS-95[CR_LotE]
Identical Set 4	ICS-100[TT_UWI]	ICS-95[TT_SJE]
UF-667[CR]	**Identical Set 5**	ICS-95[TT_UWI]
UF-676[CR]	TRD-28[TT_UCRS]	TRD-75[TT_UCRS]
Identical Set 6	TRD-29[TT_UCRS]	TRD-90[TT_UCRS]
TRD-32[TT_UCRS]	**Identical Set 7**	TRD-95[TT_UCRS]
TRD-94[TT_UCRS]	IMC-67[CR_Turri 1:plotB]	
	IMC-67[CR_Turri 1:plotA]	

†CR, Costa Rica; IMC, Iquitos Mixed Calabacillo; ICS, Imperial College selections; TRD, relict clones from abandoned cocoa estates in Trinidad; TT_ICG, Trinidad and Tobago_Int. Cocoa Genebank; TT_SJE, Trinidad and Tobago_San Juan Estate; TT_UCRS, Trinidad and Tobago_Univ. Cocoa Res. Station; TT_UWI, Trinidad and Tobago_Univ. West Indies, St. Augustine campus; UF, United Fruit.

fingerprinting of individuals are not the same for each population (Table 1). In Population 3, the PIC value for 40% of the SSR primers used in this study were extremely low (0.07) and primers mTcCir24, 35, 157, 160, and 239 were monomorphic.

Genetic Diversity Management in Cacao Germplasm Collections

Based on previous classification (Supplemental Table 1), the proportion of membership to a particular population (Supplemental Table 2), and nomenclature, the following individuals should be assessed for mislabeling in the relevant collections. ICS-41 (Trinidad and Tobago_University Cocoa Research Station [TT_UCRS]) is of 98% Ecuadorian ancestry; PA-150 [Costa Rica_Cabiria collection] appears to be an F_1 hybrid of Upper Amazon and Ecuadorian parents; TRD-32, -94, and -113 all in TT_UCRS are of 99% Upper Amazon origin; TRD-28, -29, and -99 in TT_UCRS are of 99 to 84% Ecuadorian heritage.

The result of sibship analyses on the ICS and TRD clones in Population 4 revealed 12 full-sib families, one of which is located in Costa Rica (Table 3). The largest full-sib family comprises one TRD, seven ICS clones from the ICGT, and one ICS clone from Costa Rica. The second and third largest full-sib families comprise only TRD clones, with the other 75% of full-sib families comprising two or three genotypes. The latter part of Table 3 lists seven identical sets of genotypes that share the same multilocus SSR profile for all 35 microsatellite loci within a set. Identical Set 2 is of particular interest, as it contains ICS-79 and ICS-100 from the University of the West Indies, St. Augustine campus (UWI) (ICGT) and ICS-100 from Costa Rica. However, neither the ICS-79 nor ICS-100 at the UWI campus matched Pound's description for these original clones (F.L. Bekele, unpublished data, 2006), indicating the discovery of a misidentified clone in the ICGT and IC3.

In the verification of identity of ICS-83, -87, -95, and -98 at multiple locations within the ICGT, interesting discoveries were made. The ICS-83 in Block 5 at the San Juan Estate (SJE) is a mislabeled genotype, as it is identical to ICS-95 in Lot E in Costa Rica and ICS-95 in Block 1 at the SJE, ICS-95 in Campus 4 (13,8) at the UWI, and TRD 75 at the UCRS in the ICGT. The ICS-83 in Field 6B (E384, Tree4) has a 98% Trinitario ancestry and should be assessed with Pound's description of the original clone as a reference. The 35-microsatellite multilocus profile for ICS-87 at the SJE and UWI campus were found to be identical (Table 3). The ICS-98 in the SJE is not the same as the ICS-98 at the UWI campus, with the clone at the SJE possessing 0.004% more Trinitario ancestry than the one at the UWI campus. Because the distribution of the rest of their genetic makeup is different, it is more likely that the ICS-98 at the UWI campus is a mislabeled clone.

DISCUSSION

Population Structure and Genetic Diversity of the Old Trinitario Cacao of Trinidad

The 35 microsatellite loci used in this study show that the ICS and TRD clones, or old Trinitario, form a single genetically homogeneous population that could be considered a landrace. This conclusion is further supported by morphological and agronomic characterization, as no significant differences were found among these clones in the ICGT (Bekele et al., 2007). Both molecular and agronomic characterization showed clones TRD-94 and -113 to be genetically different from the rest of the population or mislabeled accessions. The majority (84%) of the old Trinitario share a common ancestry with the cacaos of Venezuelan Criollo origin used in this study and is in agreement with the findings of previous studies (Laurent et al., 1994; Lerceteau et al., 1997). This result is also congruent with the proposed genealogy of what was being called the Trinidad 'Forastero' (Ciferri, 1949; Ciferri and Ciferri, 1957) and later supported by Motamayor et al. (2002), who showed that Criollo cacao originated in Venezuela. A common ancestry for the cacaos of Trinidad and Venezuela is not so farfetched, as Trinidad was the first island in the Caribbean settled by preagricultural societies >7000 yr ago and later by ceramic-using agriculturalists around 250 BCE, both originating from South America.

The new Trinitario population, comprising mainly cloned TSH cultivars, was derived by introgressing resistance to witches' broom disease from selected Peruvian Upper Amazon cacaos through an intensive backcross breeding program involving a few of the best ICS clones (Montserin et al., 1957; Kennedy et al., 1987). Testament to this hybridization is seen in the 10% of the ICS and TRD clones that share a common ancestor with cacao from Peru but not Ecuador.

The results of this study substantiate a genetic base for the Trinitario flavor. Both the new (TSH cultivars) and old (ICS and TRD clones) Trinitario populations possess the same industry-recognized, distinctive Trinitario fine flavor recently characterized by organoleptics (Sukha and Butler, 2005). This Trinitario flavor is distinct from the fine Nacional flavor for which Ecuadorian cacao is known, which is described as nutty with a unique floral attribute that is a combination of herbal, forest green, and moderate fresh flower aromas (Sukha and Butler, 2005). It was also shown that the Nacional and Trinitario populations could be distinguished by the microsatellite loci used in this study, sharing only three out of 10 most informative loci in common. This represents the first formal report of a marker-supported genetic basis for the Trinitario flavor in cacao. A genetic basis for Nacional flavor was previously reported (Lerceteau et al., 1997).

The development of organoleptic characterization in the International Cocoa Organization/Common Fund for Commodities fine flavor project (Sukha and Butler, 2005) and the highly inbred TSH cultivars, a model Trinitario population, makes marker-assisted selection of particular flavor traits in cacao a possibility for the first time. But before this can be a reality, the myriad of steps involved in the processing of cacao beans must be standardized as was done for organoleptic characterization.

The Introgression of Frosty Pod Rot Resistance into the New Trinitario Population

The superior attributes of ICS-95 were manifested early in the selection program, as it was described as showing outstanding resistance to witches' broom disease among the ICS clones (Toxopeus, 1969), a good yielder (Cope, 1953), and an early bearer, being rated second among the top 25 ICS clones identified at the SJE (Jolly, 1953). Recently, ICS-95 showed good resistance to seven genetic groups of the frosty pod rot pathogen (Phillips-Mora et al., 2005) and is the source of resistance deployed in the TSH clones. This implies that resistance to frosty pod rot is inherent in the Trinitario population, and new sources of resistance should be sought in the Venezuelan Criollo population from which Chauo-120 and Criollo-13 were derived. Ciferri and Ciferri (1957) identified the cacao populations of the Paria Peninsula of Venezuela extending southward through Caños de San Juan and the delta of the Orinoco River (Delta Amacuro) as possible origins of the Criollo cacao of Trinidad.

Ancestry in the Conservation of Families as Cacao Germplasm

One of the aims in the conservation of cacao is to maximize genetic diversity in a minimum number of genotypes for cost-effective management of germplasm collections. The PIC values for the microsatellite loci demonstrate a high power of discrimination for 60% of the loci used in both the Upper Amazon and Trinitario populations. This substantiates the use of these microsatellite loci in the assessment of genetic diversity among the members of each population. Although these clones are very closely related, as demonstrated by strong membership (common ancestry) to their respective population, only 12 full-sib families were identified, the largest of which contained eight individuals. However, the limited gene diversity represented in the 107 individuals of the Trinitario population suggests that a much smaller number, selection of which can be facilitated by the proportion of membership, will capture the range of genetic diversity, making conservation more efficient.

This ancestry-based approach in the selection of individuals within known cacao families for conservation can be applied to other groups of sibs currently maintained as cacao germplasm in national and international collections. The prime example is the accessions introduced from Ecuador and Peru in the search for resistance to witches' broom (Pound, 1938, 1943). Pods were collected from five populations in Peru and named after their collection sites: Iquitos Mixed Calabacillo, Morona, Nanay, Parinari, and Scavina. In Ecuador, some 320 pods were collected from 80 witches' broom–resistant *refractario* trees from commercial farms (Kennedy et al., 1987). There are many such instances where sibs are maintained as cacao germplasm in both national and international collections.

These families can be rationalized, as was done for the ICS and TRD clones in this study, by the identification of a set of microsatellite loci to determine the ancestry of each individual. The ancestry values for individuals can be compared and used to identify those that are closely related for morphological and agronomic characterization. Curators can then make the decision to keep or reduce the number of individuals from the same origin of extremely close ancestry and characteristics, thereby reducing genetic redundancy in the cacao germplasm collections. At the plot or genotype level, a method for the identification of off-type trees within a row representing an accession was previously described (Johnson et al., 2007).

Population Structure Analyses

This is the first study to represent the proportion of membership of individuals to each population as a phylogenetic tree rather than the usual color bar graph generated by the Structure software (Pritchard et al., 2000). In the color bar graph, each individual is represented by a vertical line divided into colored segments corresponding to different populations. The length of each colored segment or bar is equal to the estimated proportion of the individual's membership in the population of corresponding color and is frequently used as the basis of clustering genotypes (Parker et al., 2004; Garris et al., 2005; Hyten et al., 2007). However, associating a bar to a particular genotype in the population and assessing relatedness among specific individuals in a population is not obvious in the bar graph format.

We constructed a phylogeny tree after converting the membership of individuals to a distance matrix for a clearer visualization of the relationship among genotypes and their assignment to a population. The method used to calculate the distance matrix, unlike commonly used methods, can detect similar differences caused by different variables. The consistency of assignment of individuals to populations, based on our measurement-derived tree, was comparable to the individual membership matrix data, indicating that the measurement is reliable.

In this, as in numerous previous studies, the value of the information gained by the use of microsatellite analyses in cacao is undeniable. However, the cost of these analyses in terms of man-hours and consumables quickly becomes prohibitive when considering large numbers of individuals and loci. As such, it is desirable in cacao, as in other crops, to find the lowest number of microsatellite loci required to obtain the desired information on which selection as a parent or conservation can be based.

Another application of microsatellite loci involves the identification of species or cultivars in highly processed foods for the protection of high-end niche markets. Melchiade et al. (2007) identified four microsatellite loci that could detect 17 accessions of the highly prized 'Annurca' apple (*Malus pumila* Mill. cv. Annurca) of southern Italy, with superior organoleptic characteristics in nectar and purée. In this study, the two sets of 10 loci identified by PIC values for Populations 1 and 4 were as informative as all 35 loci in differentiating among all individuals in the ICS and TRD clones within the respective populations. However, it is important to note that the 10 most informative microsatellite loci were different for each population, with the exception of two that were applicable to three of the four populations identified in this study. It was also demonstrated that a different subset of microsatellite loci would be needed to discriminate among individuals of Lower Amazon ancestry. These findings have significant implications to the approach of fingerprinting cacao germplasm.

Acknowledgments

The authors wish to thank Allan Meneses and Clare Bowen-O'Connor for technical assistance and Drs. Ainong Shi of the Institute of Genomic Diversity, Cornell Univ., and Sue Mischke of the USDA for reviewing this manuscript. The advice of Dr. Fernando Casanoves of CATIE in the precise application of INFO-GEN for data analyses is greatly appreciated. This work was funded by the USDA.

References

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

Balzarini, M., and J. Di Rienzo. 2003. Info-Gen: Software for the statistical analysis of genetic data. Available at www.info-gen.com.ar/ (verified 15 Dec. 2008). Facultad de Cienc. Agropecuarias, Univ. Nacional de Córdoba, Córdoba, Argentina.

Bekele, F. 2004. The history of cocoa production in Trinidad and Tobago. Paper presented at the APASTT Seminar Series on the re-vitalization of the Trinidad & Tobago cocoa industry, Univ. West Indies, St. Augustine, Trinidad. 20 Sept. 2004.

Bekele, F.L., G.G. Bidaisee, and J. Bhola. 2007. A comparative morphological study of two Trinitario groups from the International Cocoa Genebank, Trinidad. Annu. Rep. Cocoa Res. Unit 2006:34–42.

Botstein, D., R.L. White, M.Y. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314–331.

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

Cheesman, E.E. 1947. Results of cacao experiments in 1944–45. Trop. Agric. 23:63–65.

Cope, F.W. 1953. Some results of the cocoa clonal trials at River Estate. p. 12–23. *In* Report on cocoa research 1945–51. Imperial College of Trop. Agric., Univ. West Indies, St. Augustine, Trinidad.

Cope, F.W., and B.G.D. Bartley. 1954. Some aspects of the plant improvement programme. Imperial College of Trop. Agric. Annu. Rep. Cocoa Res. 1954:27–31.

Ciferri, R. 1949. Early criollo in Surinam and the origin of 'Forateros' of Trinidad and Venezuela. Nature 163:953.

Ciferri, R., and F. Ciferri. 1957. The evolution of cultivated cacao. Evolution Int. J. Org. Evolution 9:381–397.

Friedman, M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc. 32:675–701.

Garris, A.J., H.T. Thomas, J. Coburn, S. Kresovich, and S. McCouch. 2005. Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638.

Hyten, D.L., I.Y. Choi, Q.J. Song, R.C. Shoemaker, R.L. Nelson, J.M. Costa, J.E. Specht, and P.B. Cregan. 2007. Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics 175:1937–1944.

INFO-GEN. 2003. Info-Gen statistical software version 1.0. Univ. Nacional de Córdoba, Córdoba, Argentina.

Johnson, E.S., A. Mora, and R.J. Schnell. 2007. Field guide efficacy in the identification of reallocated clonally propagated accessions of cacao (*Theobroma cacao* L.). Genet. Resour. Crop Evol. 54:1301–1313.

Jolly, A.L. 1953. Notes on the performance of ICS clones at the San Juan Estate. p. 26–28. *In* Report on Cocoa Research 1945–51. Imperial College of Trop. Agric, Univ. West Indies, St. Augustine, Trinidad.

Kennedy, A.J., G. Lockwood, G. Mossu, N.W. Simmonds, and G.Y. Tan. 1987. Cocoa breeding: Past, present, and future. Cocoa Growers' Bull. 38:5–22.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings Bioinformatics 5:150–163.

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

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

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

Melchiade, D., I. Foroni, G. Corrado, I. Santangelo, and R. Rao. 2007. Authentication of the 'Annurca' apple in agro-food chain by amplification of microsatellite loci. Food Biotechnol. 21:33–43.

Montserin, B.G., L.L. de Verteuil, and W.E. Freeman. 1957. A note on cacao hybridization in Trinidad with reference to clonal selection and hybrid seed. Paper 8. Caribbean Commission Public Exchange Serv. 33, Port-of-Spain, Trinidad.

Motamayor, J.C., A.M. Risterucci, P.A. Lopez, C.F. Ortiz, A. Moreno, and C. Lanaud. 2002. Cacao domestication: I. The origin of the cacao cultivated by the Mayas. Heredity 89:380–386.

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

Parker, H.G., L.V. Kim, N.B. Sutter, S. Carlson, T.D. Lorentzen, T.B. Malek, G.S. Johnson, H.B. DeFrance, E.A. Ostrander, and L. Kruglyak. 2004. Genetic structure of the purebred domestic dog. Science 304:1160–1164.

Phillips-Mora, W., J. Castillo, U. Krauss, E. Rodríguez, and M.J. Wilkinson. 2005. Evaluation of cacao (*Theobroma cacao*) clones against seven Colombian isolates of *Moniliophthora roreri* from four pathogen genetic groups. Plant Pathol. 54:483–490.

Phillips-Mora, W., J. Cawich, W. Garnett, and M.C. Aime. 2006a. First report of frosty pod rot (moniliasis disease) caused by *Moniliophthora roreri* on cocoa in Belize. Plant Pathol. 55:584.

Phillips-Mora, W., J. Coutiño, C.F. Ortiz, A.P. López, J. Hernández, and M.C. Aime. 2006b. First report on *Moniliophthora roreri* causing frosty pod rot (moniliasis disease) of cocoa in Mexico. Plant Pathol. 55:584.

Posnette, A.F. 1986. Fifty years of cocoa research in Trinidad and Tobago. Cocoa Res. Unit, Univ. West Indies, St. Augustine, Trinidad.

Pound, F.J. 1931. The genetic constitution of the cacao crop. p. 10–24. *In* First Annu. Rep. on Cacao Res. 1931. Trinidad Gov. Printing Office, Port-of-Spain.

Pound, F.J. 1934. The progress of selection. p. 25–28. *In* Third Annu. Rep. on Cacao Res. 1933. Trinidad Gov. Printing Office, Port-of-Spain.

Pound, F.J. 1935. The progress of selection. p. 7–11. *In* Fourth Annu. Rep. on Cacao Res. 1934. Trinidad Gov. Printing Office, Port-of-Spain.

Pound, F.J. 1936. The completion of selection. p. 7–16. *In* Fifth Annu. Rep. on Cacao Res. 1935. Trinidad Gov. Printing Office, Port-of-Spain.

Pound, F.J. 1938. Cacao and witchbroom (*Marasmius perniciosus*) of South America with notes on other species of *Theobroma*. Report on a visit to Ecuador, the Amazon Valley and Colombia, April 1937–April 1938. Yuille's Printery, Port-of-Spain, Trinidad and Tobago.

Pound, F.J. 1943. Cacao and WitchBroom disease. Report on a recent visit to the Amazon territory of Peru, September 1942–February, 1943. Yuille's Printery, Port-of-Spain, Trinidad and Tobago.

Pritchard, J.K., M. Stephens, and P.J. Donnell. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Pritchard, J.K., and W. Wen. 2003. Structure version 2.2.2. Available at http://pritch.bsd.uchicago.edu/software.html (verified 15 Dec. 2008). Univ. of Chicago, Chicago, IL.

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

Saunders, J.A., S. Mischke, E.A. Leamy, and A.A. Hemeida. 2004. Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. Theor. Appl. Genet. 110:41–47.

Sukha, D.A., and D.R. Butler. 2005. The CFC/ICCO/INIAP cocoa flavor project: Investigating the spectrum of fine flavor within genotypes and between origins. INGENIC Newsl. 10:22–25.

Toxopeus, H. 1969. Cacao (*Theobroma cacao* L.). p. 98. *In* F.P. Ferwerda and F. Wit (ed.) Outlines of perennial crop breeding in the tropics. Misc. Papers 4. Landbouwhogeschool Agric. Univ., Wageningen, the Netherlands.

Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. Genetics 166:1963–1979.

Zhang, D., S. Mischke, R. Goenaga, A.A. Hemeida, and J.A. Saunders. 2006. Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. Crop Sci. 46:2084–2092.