Theor Appl Genet (2004) 110: 41–47
DOI 10.1007/s00122-004-1762-1

ORIGINAL PAPER

James A. Saunders · Sue Mischke · Emily A. Leamy
Alaa A. Hemeida

Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*

Received: 2 January 2003 / Accepted: 16 June 2004 / Published online: 12 November 2004

Abstract A collaborative international program was initiated to identify and describe the genetic diversity of living germplasm collections of *Theobroma cacao* genotypes that are maintained in several international collections scattered throughout tropical cacao growing countries of the world. Simple sequence repeat (SSR) DNA analysis was identified as the most appropriate molecular tool for DNA fingerprinting these collections during an international forum representing academic, government and industry scientists in the cacao community. Twenty-five SSR primers, which had been previously described, were evaluated as potential candidates to define an efficient, standardized, molecular fingerprinting protocol for *T. cacao* accessions. These primers have been evaluated for reliability, widespread distribution across the cacao genome, number of alleles produced by the SSR primers in cacao and their ability to discriminate between cacao accessions. Approximately 690 cacao accessions were used to evaluate the utility of these SSR primers as international molecular standards, and a small number of test samples of *T. cacao* were sent to two other independent laboratories for verification. DNA fragments were selectively amplified by PCR, using the SSR primers labeled with fluorescent dyes, and separated by capillary electrophoresis. Based on this study, the 15 SSR primers that had the highest reproducibility and consistency within a common genotype, while allowing the differentiation of separate divergent genotypes, were selected as international molecular standards for DNA fingerprinting of *T. cacao*.

Communicated by D.B. Neale

J. A. Saunders · S. Mischke · E. A. Leamy
Alternate Crops and Systems Lab,
Plant Sciences Institute, Beltsville Agricultural
Research Center, USDA/ARS, Building 50,
Rm. 100, Beltsville, MD 20705, USA

A. A. Hemeida
Genetic Engineering and Biotechnology
Research Institute, Minufiya University,
Sadat City, Egypt

J. A. Saunders (✉)
Molecular Biology,
Biochemistry and Bioinformatics Program,
Towson University, Smith Hall Rm. 360,
8,000 York Rd., Towson, MD 21252, USA
E-mail: jsaunders@towson.en
Tel.: +1-410-7043491
Fax: +1-410-7043490

Introduction

Theobroma cacao, the tree that produces the fruit from which chocolate is derived, is endemic to tropical regions of Central and South America. Currently, cocoa is commercially produced throughout warmer climates in South and Central America, Africa, South East Asia and numerous tropical islands. Production of cocoa has a major economic impact on selected regions of the world economy and generates an $8.6 billion chocolate industry in the US (McBride 2002). Although a cacao pod may produce from 25 to 55 seeds per pod, the seeds do not remain viable for much longer than a week once the pod has been harvested (Coe and Coe 1996). This requires that germplasm collections of *T. cacao* be maintained as living collections in the tropical environments to which they are adapted. In cocoa producing countries, some of these cacao tree collections date back more than 180 years and are propagated via rooted cuttings or are grafted onto common rootstocks (Wright 1999). Accurate records describing the site of collection of the cacao from wild sources and documents regarding the diversity of the original collections are typically limited. In addition, the logistics of maintaining a living tree collection in a tropical rain forest environment face numerous challenges. Physical labels that identify individual trees have often been destroyed or lost and misidentification of the germplasm is common. Frequently, when individual trees die, they are replaced with different genotypes of cacao accessions without amending the

collection records. It has been estimated that in some major international collections, the misidentification of cacao accessions could be as high as 20–30% of the trees in the germplasm collection (Christopher et al. 1999; Motilal et al. 2002; Saunders et al. 2001a).

Some selective improvement of cacao has been accomplished through genetic crosses made by plant breeders since the 1930s. However, organized crop improvement, through scientific crop enhancement programs that are common in annual crops, is certainly not advanced in this tree crop. To facilitate the improvement of *T. cacao* using modern breeding techniques, the correct identification of accessions within existing cacao collections is an essential initial step (Young 1994). Typically, it takes 3–6 years to produce a sexually mature cacao tree suitable for use in a breeding program. Partly because of this prolonged breeding cycle, cacao has been maintained in a state of wild cultivation for almost 2,000 years with only modest crop improvement characteristic of other modern cultivated crops. In addition, attempts to characterize the genetic diversity of the international cocoa germplasm collections using molecular techniques have been complicated by a lack of consistency in the techniques that have been used, as well as the high ratio of mislabeled accessions in these collections (Motamayor and Lanaud 2002).

DNA fingerprinting in plants can be adapted to numerous applications and uses, including characterizing individual plants to clarify errors in the identification of accessions and cultivars (Saunders et al. 2001b; Chavarriaga-Aguirre et al. 1998). It is also a useful technique to study genetic diversity within breeding populations of a crop, once the identity of individuals has been established (Degani et al. 2001; Saunders et al. 1999). Several molecular techniques and procedures can be employed for DNA fingerprinting in plants, each of which has its strengths and weaknesses. Simple sequence repeat (SSR) analysis is particularly useful due its ability to detect heterozygotes and because of the ease of scoring and analyzing data from large numbers of samples (Bredemeijer et al. 1998). A number of researchers have suggested the need for a core set of SSR DNA primers to fingerprint cocoa to cultivar type (Charters and Wilkinson 2000). We have analyzed 691 cacao accessions held at either the USDA cacao germplasm collection in Puerto Rico, the CATIE cacao germplasm collection in Costa Rica or the International Cacao Germplasm collection held by the Cocoa Research Unit in Trinidad and Tobago. Twenty-five SSR primers previously described by Lanaud et al. (1999, personal communication) were suggested by these researchers to be used to evaluate the utility of this technique as a high-throughput DNA fingerprinting protocol. This report describes the results based on these analyses, and 15 SSR primers described in this report were selected to serve as proposed international molecular standards for DNA fingerprinting *T. cacao*.

Materials and methods

Plant material

The cacao samples used for these DNA fingerprinting profiles included 691 individual cacao accessions from three separate international germplasm collections, to represent the widest diversity possible in the analysis profiles. These germplasm centers consisted of the USDA cacao germplasm collection held in Mayaguez, Puerto Rico; the CATIE cacao germplasm collection held in Turrialba, Costa Rico; and the International Cacao Germplasm collection held by the Cocoa Research Unit in Trinidad and Tobago. Fresh young leaves (4–5 per tree) were collected from each tree that could be identified from the cacao germplasm database records and stored at room temperature in stapled paper bags to prevent moisture accumulation. The plant material was shipped to the US USDA labs within 7 days and then frozen for subsequent analysis. The location of each tree sampled was recorded and a permanent label attached to the tree to indicate a typed specimen.

DNA isolation

DNA was isolated from 50 mg samples of *T. cacao* young leaf material using either the DNA X-tract Plus kit (D^2 BioTechnologies , Atlanta, Ga., USA) or the DNeasy Plant System (Qiagen, Valencia, Calif., USA). Both commercial DNA isolation procedures yielded DNA from *T. cacao* that was suitable for microsatellite analysis. For either method, the air-dried or frozen leaf samples were first cut into small pieces with scissors and placed in a 2-ml tube, sandwiched between two 6 mm ceramic spheres with a garnet matrix (Qbiogene, Carlsbad, Calif., USA). Lysis solution was added following the manufacturers' recommendations and samples were homogenized in a shaker-basher (Bio101 Fast Prep, Qbiogene) at oscillation speed 5.0 for 40–45 s as described previously (Saunders et al. 2001a).

For extraction with the DNA X-tract Plus kit, lysis was achieved using 600–700 µl solution 1 during homogenization, the samples were centrifuged 30 s at low speed (3,000*g*), and 175 µl of solution 2 was added to the homogenate. Tubes were vigorously mixed by inversion, incubated at room temperature for 10 min, centrifuged at 5,000*g* for 2 min, and the supernatants were transferred to 1.5 ml microfuge tubes. An equal volume of chloroform was added to the supernatant and the tube was vigorously inverted until homogenous. Samples were then centrifuged at 12,000*g* for 5 min to separate the phases, the aqueous (upper) layer was transferred to a new tube, and the chloroform extraction step was repeated. The final upper aqueous phase (about 375–500 µl) was transferred to a clean 1.5 ml tube, 600 µl of DNA precipitation solution was added and samples were incubated on ice for at least 30 min. DNA

was precipitated by centrifugation at 12,000*g*. The DNA pellet was washed in 1 ml 70% ethanol, centrifuged at 12,000*g* for 5 min, dried and resuspended in 200 µl sterile water.

DNA extraction was modified from the DNeasy Plant system, which required 700 µl warm (up to 65°C) buffer AP1 for lysis during homogenization. Samples were centrifuged for 30 s at low speed (4,000*g*), and 7 µl of RNase A stock solution (100 mg/ml) was mixed into each tube until no tissue clumps were visible. The mixture was incubated for 20 min at 65°C, and the tubes were inverted twice during incubation. Addition of 228 µl buffer AP2 to the lysate and incubation on ice for 5 min precipitated detergent, proteins and polysaccharides. The slurry was decanted and scraped into a QIA shredder spin column within a 2-ml collection tube. The column–tube assembly was centrifuged at 12,000*g* for 2 min in a microcentrifuge. The clear filtrate (500 µl) was transferred to a 1.5-ml microfuge tube. Buffer AP3 (0.5 volume) and 100% ethanol (one volume) were added to the lysate and the solution gently mixed by pipetting. Up to 650 µl sample mixture (including any precipitate that formed), was applied to the DNeasy spin column, which was placed in a 2-ml collection tube. The assembly was centrifuged 1 min at 12,000*g*, to bind the DNA to the column membrane, and the filtrate discarded. The remaining sample mixture was applied to the same column, and the procedure was repeated. The collection tube was replaced by a clean 2-ml tube, the column was washed two to three times by adding 500 µl buffer AW (containing ethanol) onto the DNeasy column, centrifuging 1 min at 12,000*g*, and discarding the filtrate. Following the final wash, the column–tube assembly was centrifuged for 2 min at 12,000*g* to dry the column membrane and the collection tube was discarded. The DNeasy column was transferred to a clean 1.5 ml microcentrifuge tube and DNA was eluted from the membrane by pipetting 100 µl of preheated (65°C) Buffer AE directly onto the DNeasy column membrane, followed by incubation for 5 min at room temp, then centrifugation for 1 min at 12,000*g*. The elution step was repeated yielding a final volume of 200 µl DNA solution.

The presence of DNA was verified by quantitation of double-stranded DNA concentration with PicoGreen (Molecular Probes, Eugene, Ore., USA) using a Fluoroskan Ascent microplate reader equipped with 485/538 nm excitation/emission filter settings (Labsystems, Helsinki, Finland).

SSR analysis

Primer sequences for amplification of microsatellite loci in *T. cacao* have been described previously (Lanaud et al. 1999; Risterucci et al. 2000). Primer sets were synthesized by Research Genetics (Huntsville, Ala., USA). Forward oligonucleotides were 5′-labeled with three separate WellRED fluorescent dyes [black (D2), green (D3) and blue (D4)]. PCR reaction mixtures consisted of 3 µl containing 10–150 ng of genomic DNA template, 1.0 µl of a stock solution containing both forward and reverse primers (10 µM each) and 20 µl Platinum PCR SuperMIX (Life Technologies, Invitrogen, Carlsbad, Calif., USA), which contained an additional 0.6 U of Platinum *Taq* (Life Technologies) for a total volume of 24 µl. PCR amplifications were performed with a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, Calif., USA) with the following profile: 95°C for 5 min; followed by 35 cycles of 94°C for 30 s, either 46°C or 51°C for 1 min (depending upon the annealing temperature of the primer pair), 72°C for 1 min; followed by a hold at 60°C for 15 min.

The amplified microsatellite loci were separated by capillary electrophoresis and analyzed on a CEQ 8000XL eight-channel capillary DNA Analysis System (Beckman Coulter, Fullerton, Calif., USA) to determine the size of the amplified microsatellites. The capillary injection consisted of a 30 s electrophoresis at 2.0 kV from a mixture of 0.3 µl CEQ DNA size standard-400 (Beckman Coulter), 1.0 µl of PCR-amplified DNA fragments and 29.7 µl sample loading solution containing deionized formamide. The CEQ 8000XL Frag-3 profile was used with running conditions of: capillary temperature 50°C, denaturation temperature 90°C for 120 s, and separation voltage 6.0 kV for a run time of 35 min. Data analysis was performed using the CEQ 8000XL Fragment Analysis software version 5.0 according to manufacturers' recommendations (Beckman Coulter).

Results

A total of 691 accessions of *T. cacao* were used to evaluate 25 SSR primers flanking dinucleotide repeating motifs described previously by Lanaud et al. (1999). The salient characteristics of the 15 primers that were selected are listed in Table 1.

A number of criteria were used for the selection of the specific SSR primers to serve as the international standard molecular markers for cacao. The first criterion was the reliability and consistency of the primer's ability to amplify its respective locus during the PCR amplification reactions. Table 1 shows the optimal annealing temperatures empirically determined for each primer PCR reaction (46 or 51°C), which was used to ensure the reliability of each primer. Although many of the SSR primers that were tested were reliable, the selection of the best primers to serve as international standards for cacao DNA fingerprinting was also based on several additional factors. One other major criterion was to select the primers encompassing loci that showed the largest number of polymorphic alleles, since these were the most discriminating loci, using the samples tested in this study. Primers were also selected for their ability to amplify non-overlapping DNA fragments in order to

Table 1 SSR primers selected as international molecular standards for DNA fingerprinting of *T. cacao*. The table lists the chromosome on which the locus is mapped, the optimal annealing temperature for the PCR amplification of that primer pair, and the size range of the alleles found for that locus

Microsatellite EMBL no.	Chromosome no.	5′ Forward primer	3′ Reverse primer	Annealing temperature (°C)	Size (bp)	Motif
MTcCIR7-Y16981	7	ATGCGAATGACAACTGGT	GCTTTCAGTCCTTTGCTT	51	150–167	$(GA)_{11}$
MTcCIR18-Y16991	4	GATAGCTAAGGGGATTGAGGA	GGTAATTCAATCATTTGAGGATA	51	333–357	$(GA)_{12}$
MTcCIR40-AJ271943	3	AATCCGACAGTCTTTAATC	CCTAGGCCAGAGAATTGA	51	262–288	$(AC)_{15}$
MTcCIR22-Y16995	1	ATTCTCGCAAAAACTTAG	GATGGAAGGAGTGTAAATAG	46	276–301	$(TC)_{12}N_{146}(CT)_{10}$
MTcCIR24-Y16996	9	TTTGGGGTGATTTCTTCTGA	TCTGTCTCGTCTTTTGGTGA	46	186–207	$(AG)_{13}$
MTcCIR15-Y16988	1	CAGCCGCCTCTTGTTAG	TATTTGGGATTCTTGATG	46	234–263	$(TC)_{19}$
MTcCIR33-AJ271826	4	TGGGTTGAAGATTTGGT	CAACAATGAAAATAGGCA	51	265–348	$(TG)_{11}$
MTcCIR1-Y16883	8	GCAGGGCAGGCTCAGTGAAGCA	TGGGCAACCAGAAAACGAT	51	128–146	$(CT)_{14}$
MTcCIR60-AJ271958	2	CGCTACTAACAAACATCAAA	AGAGCAACCATCACTAATCA	51	190–218	$(CT)_7(CA)_{20}$
MTcCIR11-Y16985	2	TTTGGTGATTATTAGCAG	GATTCGATTTGATGTGAG	46	286–321	$(TC)_{13}$
MTcCIR12-Y16986	4	TCTGACCCCAAACCTGTA	ATTCCAGTTAAAGCACAT	46	165–256	$(CATA)_4N_{18}(TG)_6$
MTcCIR26-Y16998	8	GCATTCATCAATACATTC	GCACTCAAAGTTCATACTAC	46	285–310	$(TC)_9C(CT)_4TT(CT)_{11}$
MTcCIR37-AJ271942	10	CTGGGTGCTGATAGATAA	AATACCCTCCACACAAAT	46	136–187	$(GT)_{15}$
MTcCIR6-Y16980	6	TTCCCTCTAAACTACCCTAAAT	TAAAGCAAAGCAATCTAACATA	46	224–253	$(TG)_7(GA)_{13}$
MTcCIR8-Y16982	9	CTAGTTTCCCATTTACCA	TCCTCAGCATTTTCTTTC	46	290–307	$(TC)_5TT(TC)_{17}TTT(CT)_4$

facilitate multiplexing with standard capillary electrophoretic separation profiles. Even with the use of three differential dyes for sample identification (in addition to the red dye reserved for the internal standards within each capillary run), overlapping peaks tended to interfere with accurate analysis of the DNA fragment patterns. By selecting SSR primers that produced non-overlapping DNA fragments of alleles between different loci, this problem could be avoided. Finally, primers were selected that encompassed loci on the greatest number of chromosomal regions to represent the widest diversity in the physical genome of cacao. Within the 15 loci chosen for these SSR DNA profiles, nine of the ten cacao chromosomes of *T. cacao* were represented. Using these criteria, the optimal 15 SSR primers were selected to serve as international molecular standards for *T. cacao* as shown in Table 1. These represent the optimal SSR primers based on quality evaluation of over 10,000 SSR DNA profiles (excluding data for primers not shown).

By using multiplexed primers with different dyes, three SSR DNA fragment analysis profiles could be separated and analyzed in each column during a single run of the capillary electrophoresis analysis system. Figure 1 shows a typical electrophoretogram of one such analysis depicting DNA fragment patterns from three pairs of SSR primers (Y16986, Y16988, Y16985). SSR analysis produced allelic bands of variable sizes for each primer and the template DNA under investigation. Individual samples that are heterozygous for a single locus defined by a SSR primer pair will produce two DNA fragment sizes in the DNA profile while samples from parental lines that are homozygous for a loci will show a single peak. As seen in Figure 1, PMCT-45 (CATIE accession 19576), a sample of *T. cacao* taken from the international germplasm collections held in CATIE, Costa Rica, is heterozygous for two of the primer pairs (Y16986 and Y16985) and homozygous for a locus of the third primer pair (Y16988). The CATIE collection has two living plants labeled with this accession number, and the germplasm was purportedly acquired from Nicaragua. We have thus far identified 17 alleles of locus Y16986, 14 alleles of Y16988 and 16 alleles of locus Y16985.

Simple sequence repeat regions of the genomic DNA are typically located in non-coding regions of the genome and thereby are prone to be highly variable. This variability in the repeating unit, or occasionally in its flanking regions, typically produces several different allelic DNA fragment sizes that can result from the SSR primers pairs designed to amplify across the repeating region. By analyzing 691 different cacao accessions from the three different germplasm collections, a composite allele list for each of the 15 SSR primer loci was generated as shown in Table 2. Variation among this allele list of 176 DNA fragment categories provided the basis for differentiating between accessions in the diverse collection of cacao samples assayed in this report. The specific DNA fragment size categories depicted in the allele list

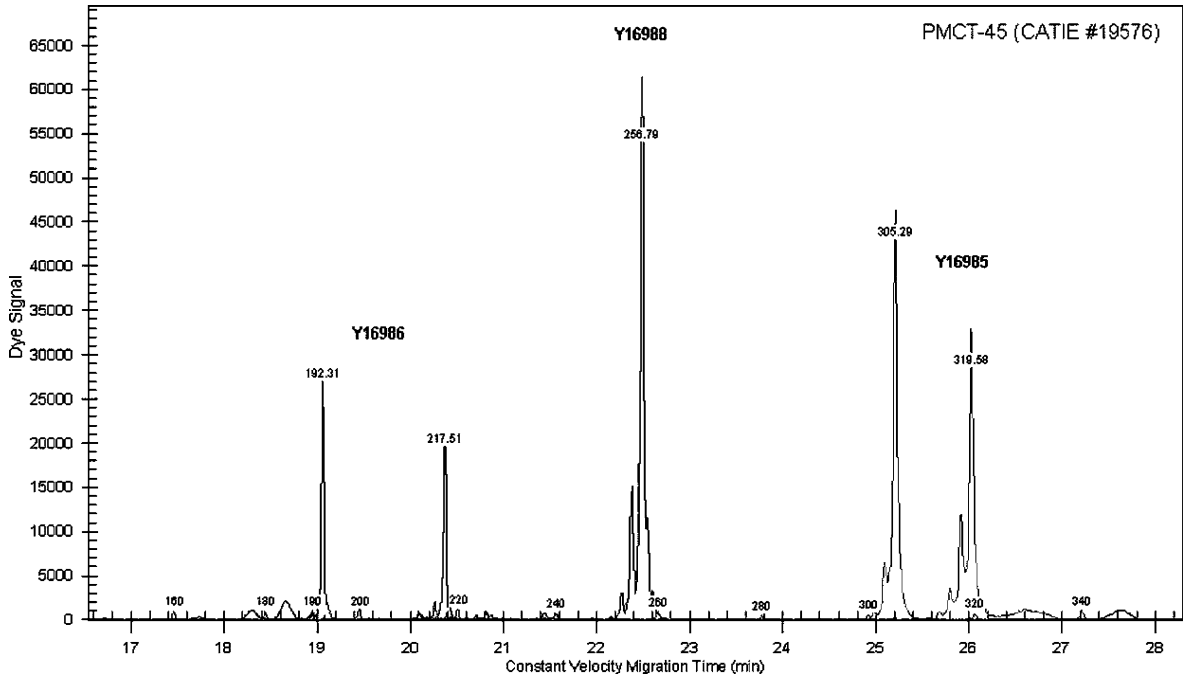

Fig. 1 This electrophoretogram shows the DNA fragment profile characteristics of three SSR primers for different loci. The amplified products, labeled with three different dyes, were multiplexed and run concurrently in the CEQ 8000 for separation by capillary electrophoresis. The DNA was extracted from the *T. cacao* accession no. 19576, which is part of the germplasm collection maintained in Costa Rica (CATIE). Alleles from three loci were amplified as shown in this sample: The locus Y16986 (mTcCIR12) is located on chromosome 4, the locus Y16988 (mTcCIR15) on chromosome 1, and locus Y16985 (mTcCIR11) on chromosome 2. The accession is homozygous for the 257 bp allele of the mTcCIR12 locus, as evidenced by the presence of only a single *peak*, indicating that both parental lines contributed the same allele fragment (257 bp). The accession is heterozygous at the other two loci, possessing both the 192 bp allele and the 218 bp allele at locus mTcCIR12, and alleles sizes of 305 bp and 320 bp at the locus on chromosome 2 (mTcCIR11, two *peaks*). The *peaks* in the graph represent the internal size standards

reflect the ranges of peaks that have been observed using the SSR primer pairs and cacao samples in this study. In evaluating the population of *T. cacao* used for this study, efforts were made to include as diverse a group as possible. Table 3 lists the number of accessions that were evaluated based on the different general types of cacao included in this survey. The major types of cacao identified in this list include: upper, lower and unidentified Forastero cacao, Criollo, Trinitario, hybrids, and cacao of mixed or unknown origin.

Discussion

Theobroma cacao is an important economic crop plant in many tropical countries, however, the genetic diversity of the available germplasm has never been fully and systematically characterized. Figueira et al. (1994) reported the classification of *T. cacao* using molecular markers but were not able to clearly distinguish between *Herrania* and several other *Theobroma* species. This report attempts to distinguish between the use of molecular markers for identification of genotypes of *T. cacao* by DNA profiling with SSR markers located in non-coding regions of the genome. The International Cocoa Germplasm Database published by the University of Reading, England, lists over 14,000 accessions of *T. cacao* that are held in world germplasm collections. It was the intent of this study to evaluate SSR primers of cacao to select specific loci to be utilized as standards by scientists within the cacao global community. These molecular markers will help to avoid duplication of names, mislabeling, and to assess a preliminary genetic diversity of the major cacao germplasm collections using DNA fingerprinting techniques. To achieve this task, a balance between the number of SSR primers used for the study must be made in view of the large number of samples to be processed. To address this issue, a consortium of industry, academic, and government scientists agreed to establish a common molecular probe-based system using 15 SSR DNA primers that would be accepted as international standards for *T. cacao* DNA fingerprinting. A total of 60 potential SSR primers that had been previously identified and described for cacao (Lanaud et al. 1999) were pared down to 25 suggested potential primers by the Centre de Cooperation Internationale en Recherché Agronomique pour le Développement (CIRAD) research team using the criteria of fragment size, reliability, and distribution on the cacao genome (C. Lanaud, personal communication). This study describes how these 25 primers were further evaluated as candidates for the final selection of 15 loci

Table 2 Allele list for DNA fragment patterns from 15 SSR primers applied to *T. cacao*

Primer locus [Allele range (bp)]																
Y16981 mTcCIR7	Y16980 mTcCIR6	Y16995 MTcCIR22	Y16996 mTcCIR24	Y16982 mTcCIR8	Y16883 mTcCIR1	Y16985 mTcCIR11	Y16986 mTcCIR12	Y16988 mTcCIR15	AJ271942 mTcCIR37	AJ271826 MTcCIR33	Y16991 mTcCIR18	Y16998 mTcCIR26	AJ271943 mTcCIR40	AJ271958 mTcCIR60
149.7–150.4	223.9–224.5	276.2–276.8	186–187.5	289.7–291.1	127.6–128.0	286.2–286.7	165.1–165.5	234.3–235.1	135.7–136.4	265.4–266.1	333.1–333.8	285.6–286.0	261.7–262.4	190.2–190.6
152.6–152.8	225.7–225.9	282.0–283.0	188.0–189.0	291.5–292.6	128.8–130.7	290.2–291.8	177.7–178.0	235.7–237.2	137.3–138.0	273.5–274.3	335.0–335.8	287.6–287.9	269.8–270.8	192.0–192.6
155.2–155.8	227.5–229.5	284.6–285.7	192.3–193.1	293.9–294.6	131.9–132.2	293.1–294.8	186.4–187.0	237.6–239.0	139.8–140.5	275.5–276.3	336.9–337.7	291.6–292.0	274.5–275.1	194.1–194.6
156.0–157.4	230.8–232.6	286.0–287.0	194.3–195.0	296.2–296.6	133.9–134.9	295.2–296.7	189.5–190.9	240.4–240.8	142.1–142.8	280.3–280.8	338.9–339.1	297.6–298.8	276.1–277.8	196.0–196.7
158.0–159.0	233.2–234.7	288.1–288.6	198.0–199.4	298.0–298.5	135.7–136.2	297.1–298.4	198.5–198.8	241.8–242.9	144.2–144.6	282.6–284.3	344.7–345.1	299.5–300.4	278.7–279.8	197.8–198.7
160.5–161.0	236.1–236.5	290.0–291.1	200.4–200.9	303.9–304.5	140.3–142.0	299.2–300.2	200.9–201.5	244.5–245.0	146.3–146.9	284.8–285.2	346.6–347.3	301.6–301.9	281.3–282.5	200.1–200.4
162.3–163.0	237.5–238.9	292.0–294.0	202.1–203.8	305.6–307.1	145.9–146.4	301.1–301.6	202.4–203.6	246.6–247.0	148.3–149.4	286.0–286.5	348.6–349.1	303.7–304.2	283.0–284.1	211.4–212.1
164.4–164.7	239.6–240.5	295.8–296.2	204.1–205.6			303.1–304.1	204.5–205.5	250.2–251.1	150.1–150.5	287.5–288.4	350.7–351.6	305.6–306.8	285.0–285.9	213.5–214.0
166.3–166.9	241.5–242.3	300.5–300.8	206.0–206.7			305.2–306.5	206.5–207.3	251.8–252.9	152.4–152.8	291.0–291.8	356.6–357.0	307.7–308.9	287.1–287.7	215.0–216.0
	245.6–246.0					307.8–308.2	208.7–209.3	253.5–255.2	154.6–155.3	298.4–299.0		309.6–310.2		217.4–217.6
	247.2–247.8					309.1–310.0	212.8–213.9	255.9–257.2	159.0–159.9	302.1–303.3				
	249.8–251					311.2–311.9	214.7–216.4	258.6–259.5	160.9–161.9	304.5–304.8				
	252.1–252.6					313.3–313.7	217.1–217.8	260.9–261.2	164.8–166.3	305.9–306.7				
						317.3–318.6	219.8–220.2	262.8–263.1	166.9–167.1	308.6–309.2				
						319.1–319.9	223.6–224.3		168.1–168.8	309.8–311.1				
						321.0–321.5	253.7–254.8		176.0–176.3	342.5–343.2				
							255.9–256.5		177.2–178.9	345.3–346.5				
									180.2–180.5	347.0–347.8				
									186.4–187.2					

for international standards for cacao, based on the results we now report after screening of 691 cacao accessions taken from three diverse germplasm collections.

In our hands, this DNA analysis procedure, coupled with either capillary or gel electrophoresis, is a reliable genetic analysis system for clarification of the correct identity of cacao accessions. The list of loci and their respective 176 alleles shown in Table 2 should be regarded as the basis for the identification of cacao accessions. Although an attempt was made to select a highly diverse population of cacao accessions to generate this allele list, we expect that additional alleles will be identified as divergent cacao populations are evaluated. Allelic variation at these 15 loci can be used to assess the genetic diversity of *T. cacao* collections from world populations. It is understood that detailed molecular characterizations of the genetic nature of any single accession or group of accessions may require the use of additional molecular markers, however, this study demonstrates that these 15 loci are capable of successfully identifying the correct genotype of cacao. Intensive genetic analysis of siblings, some of which were done in this study, may be difficult to fully characterize with only 15 loci regardless of which primers are utilized. In many cases within this study, samples thought to be closely related, or even duplicates, were not shown to be closely linked by the genetic analysis. Unexpected results of this type are not indicative of the poor performance of the DNA fingerprinting technique utilized in the study, but were recognized as mislabeled accessions. This factor alone strengthens the justification for a core set of DNA probes to be utilized for the genetic characterization of *T. cacao* germplasm collections.

To verify the utility of these primers in other laboratories using different DNA analyzers operated by other researchers, the Trinidad Cocoa Research Unit provided eight samples of *T. cacao* cultivars for verification testing to several independent cocoa researchers. The Pennsylvania State University Cocoa Research Unit under the direction of Mark Guiltinan recently published this verification testing (Swanson et al. 2003). They found that 11 of the primers were more than sufficient to characterize each of the test samples to type, however four of the SSR primers (mTcCIR40, mTcCIR33, mTcCIR12, and mTcCIR6) were difficult to amplify by common PCR techniques in use in their laboratory. The same samples of *T. cacao* test cultivars were also provided to the USDA, Subtropical Horticulture Research Station in Miami, Florida, where the samples were subjected to a third set of independent SSR analysis and all 15 primers were successfully amplified and utilized for DNA fingerprinting (R. Schnell, personal communication).

Due to the large numbers of *T. cacao* accessions that are available in global germplasm collections, the use of high-throughput techniques for DNA fingerprinting greatly facilitated analysis of samples. Multiplexing of multiple samples in the same capillary electrophoresis DNA fragment separation profile is an essential element

Table 3 Major types of *T. cacao* accessions that were used for evaluation of loci for DNA profiling SSR primer standards

Cacao type	Forastero upper Amazon	Forastero lower Amazon	Unknown Forastero	Criollo	Trinitario	Hybrid	Mixed unknown	Total
Accessions analyzed	65	28	59	94	262	105	78	691

of high-throughput analysis. To avoid any confusion or overlapping peaks within a common run, only three loci, each labeled with a different dye were used in each common analysis within our study. Further, to avoid potential cross dye contamination of data, the PCR primers that were used to flank the loci were designed to amplify alleles of different sizes with little or no overlap of DNA fragment sizes. As shown in Fig. 1, this allowed rapid identification of the DNA fragments sizes and minimized miscalls in the scoring of allelic peaks. All of the allelic DNA fragment sizes listed in Table 2 were derived from analyses of *T. cacao* alone. Analysis of germplasm in other species of *Theobroma* or other related genera should result in additional alleles not listed in Table 2 if the primers were functional in those divergent genetic lines.

Acknowledgements The authors gratefully thank Dr. Ulrike Krauss from CATIE, Costa Rica, Dr. Ricardo Goenaga from the USDA in Puerto Rico, and Dr. David Butler, Dr. Elizabeth Johnson, and Mr. Oliver Sounigo from CRU, Trinidad and Tobago for providing the cacao plant accessions used in this study. They also thank Dr. Claire Lanaud for recommendations on SSR primers and review of the manuscript and Dr. Mark Guiltinan and Dr. Ray Schnell for the independent verification analysis of the DNA primers.

References

Bredemeijer GMM, Arens P, Wouters D, Visser D, Vosman B (1998) The use of semi-automated fluorescent microsatellite analysis for tomato cultivar identification. Theor Appl Genet 97:584–590

Charters MY, Wilkinson MJ (2000) The use of self-pollinated progenies as ''in-groups'' for the genetic characterization of cocoa germplasm. Theor Appl Genet 100:160–166

Chavarriaga-Aguirre P, Maya MM, Bonierbale MW, Kresovich S, Fregene MA, Tohme J, Kochert G (1998) Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. Theor Appl Genet 97:493–501

Christopher Y, Mooleedhar V, Bekele F, Hosein F (1999) Verification of accession in the ICG, T using botanical descriptors and RAPD analysis. In: Annual report 1998. St. Augustine, Trinidad and Tobago: Cocoa Research Unit, The University of the West Indies, pp 15–18

Coe SD, Coe MD (1996) The true history of chocolate. Thames and Hudson, London, pp 16–34

Degani C, Rowland LJ, Saunders JA, Hokanson SC, Ogden EL, Golan-Goldhirsh A, Galletta GJ (2001) A comparison of genetic relationship measures in strawberry (*Fragaria ananassa*Duch.) based on AFLPs, RAPDs, and pedigree data. Euphytica 117:1–12

Figueira A, Janick J, Levy M, Goldsbrough PB (1994) Re-examining the classification of *Theobroma cacao* L. using molecular markers. J Am Soc Hort Sci 119:1073–1082

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

McBride J (2002) Tropical agriculture gets attention. Agric Res 50:10–11

Motamayor JC, Lanaud C (2002) Molecular analysis of the origin and domestication of *Theobroma cacao* L., In: Engles JMM, Ramanatha RV, Brown AHD, Jackson MT (eds) Managing plant genetic diversity. IPGRI, pp 77–87

Motilal LA, Butler DR, Mooleedhar V (2002) Verification in global cacao germplasm collections. Ingenic Newslett 7:4–8

Risterucci AM, Grivet L, N'Goran JAK, Pierevtti I, Flament MH, Lanaud C (2000) A high density linkage map of *Theobroma cacao* L. Theor Appl Genet 101:948–955

Saunders JA, Pedroni MJ, Daughtry CS (1999) DNA fingerprinting of marijuana by the AFLP technique. Focus 20:10–11

Saunders JA, Hemeida AA, Mischke S (2001a) USDA DNA fingerprinting programme for identification of *Theobroma cacao* accessions. In: Proceedings of the international workshop on new technologies and cocoa breeding, 16–17 October 2000, Kota Kinabalu, Malaysia, pp 108–114

Saunders JA, Pedroni MJ, Penrose LDJ, Fist AJ (2001b) AFLP DNA analysis of opium poppy. Crop Sci 41:1596–1601

Swanson J-D, Lee AC, Guiltinan MJ (2003) Ring test: results from Penn State University. Ingenic Newslett 8:22–27

Wright H (1999) Cocoa, its botany, cultivation, chemistry and diseases. Biotech Books, Delhi, pp 1–22

Young AM (1994) The chocolate tree, a natural history of cacao. Smithsonian Institution, Washington, pp 65–79