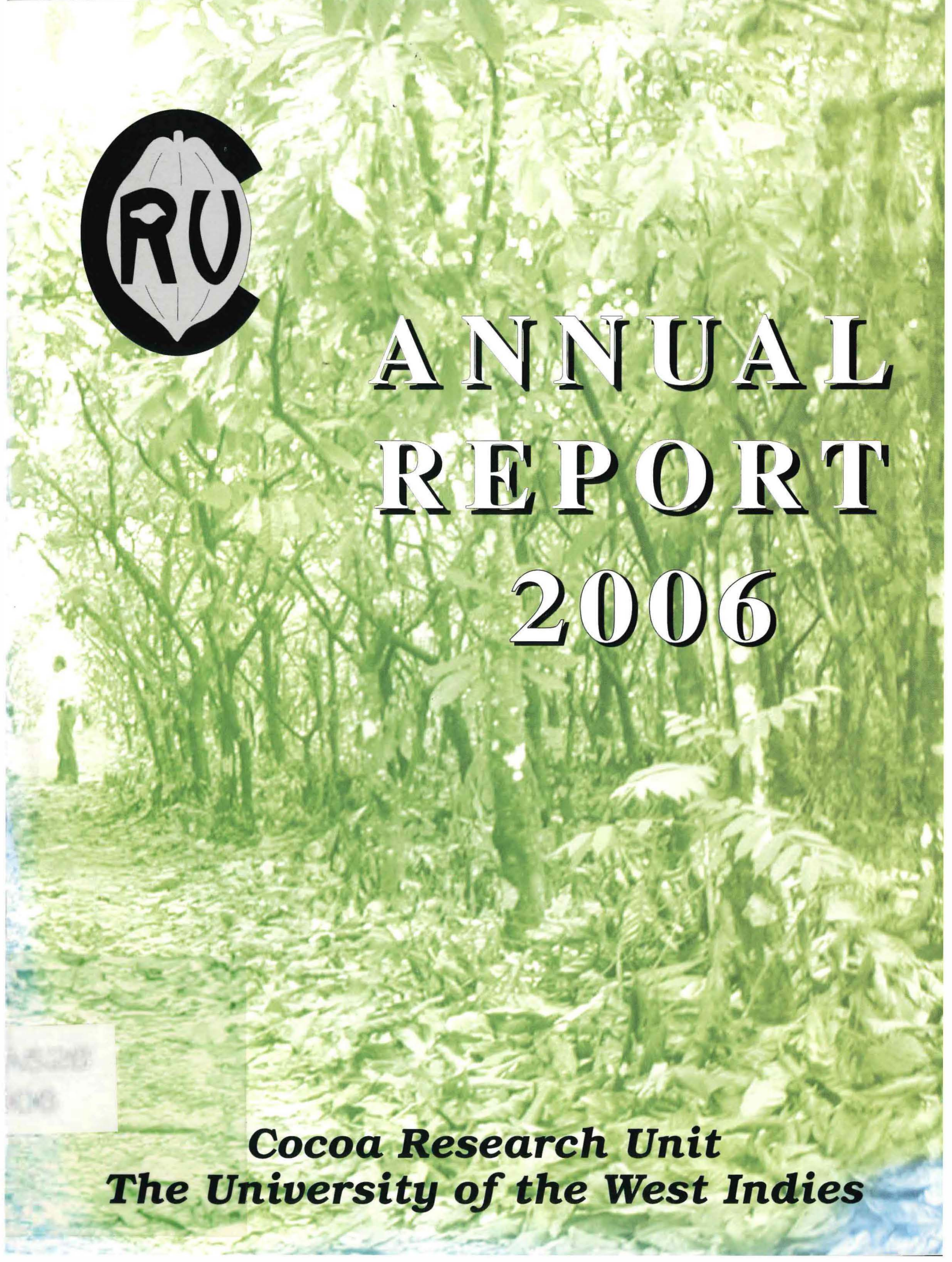


# ANNUAL REPORT 2006

**Cocoa Research Unit**
**The University of the West Indies**

*Annual Report 2006*. St. Augustine, Trinidad and Tobago: Cocoa Research Unit, the University of the West Indies. 79 pp.

**The work of CRU is made possible by support from**


The Biscuit, Cake, Chocolate & Confectionary Association (BCCCA), UK

United Nations Common Fund for Commodities (CFC)

Centre de coopération internationale en recherche agronomique pour le développement (CIRAD), France

Ministerie van Landouw, Natuur en Voedselkwaliteit, Holland

Cocoa Research Association, UK

Lindt & Sprüngli (International) AG, Switzerland

Guittard Chocolate Company, Burlingame, USA

Masterfoods, UK

Cadbury Ltd., UK

Bioversity International

Ministry of Agriculture, Land and Marine Resources (MALMR), Government of the Republic of Trinidad and Tobago

United States Department of Agriculture, USA

World Cocoa Foundation, USA

The University of the West Indies (UWI), Trinidad and Tobago

The University of Reading, UK

International Cocoa Germplasm Database (ICGD)

The University of Hamburg, Germany

Forschungskreis der Ernährungsindustrie e.V.

Arbeitsgemeinschaft industrieller Forschungsvereinigungen "Otto-von-Guericke" e.V.

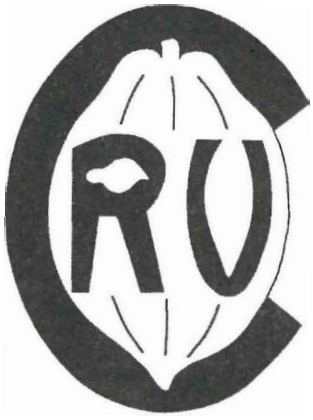


Cocoa Research Unit
The University of the West Indies
St. Augustine, Trinidad and Tobago

Tel. +1 868 662 8788
+1 868 662 2002 Ext. 2115
Fax +1 868 662 8788
E-mail cru@cablenett.net


**Cover photograph**. Intensively planted cocoa before (back cover) and after (front cover) pruning.

# Annual Report 2006




**Cocoa Research Unit**
**The University of the West Indies**
**St. Augustine, Trinidad and Tobago**
**2007**

# Evaluation of microsatellites for verification of identities in cacao field genebanks

L.A. Motilal, P. Umaharan, M. Boccara and D. Zhang

## Introduction

The ICG,T is one of the largest cacao germplasm collections containing over 2,000 accessions, each represented by replicated trees in plots in a field genebank. Formally planned in 1982, the genebank was assembled from germplasm, mostly originating from multiple collecting expeditions (1930 onwards) from Amazonian South America, Central America and the West Indies (Kennedy and Mooleedhar, 1993). Mislabelled plants have been identified as a serious problem in germplasm collections (Hurka *et al.*, 2004) and have been reported for *Cicer* (Shan *et al.*, 2005), French olive (*Olea europaea* L.) (Khadari *et al.*, 2003), persimmon (*Diospyros kaki* Thunb.) (Badenes *et al.*, 2003) and cacao (Figueira, 1998; Risterucci *et al.*, 2001; Motilal and Butler, 2003) among others. Mislabelling within and among accession plots in the ICG,T may have arisen from a combination of factors such as the multiplicity of introductions and plant movements; field mapping and plot demarcation; and the discontinuity in personnel involved in the flow of germplasm material from the original collections and in genebank maintenance.

Mislabelling issues can be resolved by multilocus fingerprinting. A variety of molecular markers are available, but microsatellite markers are well suited being co-dominant (thus allowing the detection of heterozygotes), found throughout the genome, have high allelic variability (Powell *et al.*, 1996) and are relatively fast and easy to analyse compared to other DNA markers (Morgante and Olivieri, 1993). Cacao microsatellites developed by Lanaud *et al.* (1999) have been utilised for cacao clone identification (e.g. Figueira, 1998; Risterucci *et al.*, 2001; Saunders *et al.*, 2004; Cryer *et al.*, 2006; Zhang *et al.*, 2006).

Unambiguous identification of individuals within germplasm collections is, however, a large-scale project requiring substantial resource and time allocations. Saunders *et al.* (2004) suggested that identity issues in cacao germplasm could be resolved with 15 microsatellite primer pairs (MPPs) which have since been utilised by cacao molecular biologists. However, there are approximately 12,000 trees (2,000 accessions × 6 propagated trees) in the ICG,T to be evaluated over 15 loci, and any effort to resolve identity issues in a timely and cost-effective manner would therefore be welcomed.

One way of making the task more efficient is to reduce the number of SSR loci required for detecting mislabelling in the germplasm collection. Furthermore, many other cacao SSRs have been identified since Saunders *et al.* (2004) made their recommendation and it is quite possible that other MPPs may be more useful than the proposed set. This study was therefore undertaken to determine the composition and minimum number of microsatellite loci required for an accelerated yet reliable protocol for fingerprinting individuals in a cacao germplasm collection.

## Materials and methods

### Plant material, DNA extraction and quantification

DNA from cacao leaf tissue collected from accessions in the ICG,T (Table 1) was extracted using

the Kobayashi protocol (Kobayashi *et al.*, 1998 ) using a FastPrep 120V machine (Qbiogene, Inc., California, USA) for maceration. Precipitated DNA was re-suspended in sterile deionised water (SDW) and kept as stock solutions. Dilutions (×100 or as required) of the stock DNA solutions were prepared in SDW and assayed with PicoGreen® (Molecular Probes, Eugene, Oregon, USA) in a Fluroskan Ascent (Labsystems, Finland) system. Final dilutions for experimental manipulations were prepared at 0.2 ng/µL in SDW. Two additional DNA samples (H1 and U1), extracted using a Qiagen DNeasy plant minikit (Qiagen GmbH, 2000), were obtained from a Peruvian collection held by USDA-ARS (Agriculture Research Service), Beltsville.

**Table 1. Details of accessions used in this study.**

| Accession | Position | Group | Country of Origin (Status) |
|---|---|---|---|
| AC 2 [BLZ] | T1 | Criollo | Belize (wild) |
| AC 20 [BLZ] | T1 (fp1032) | Criollo | Belize (wild) |
| B 9/10-25 [POU] | Marper Farm, C1078 | Refractario | Ecuador (cultivated) |
| BC 3 [BLZ] | T1 (fp1019) | Criollo | Belize (wild) |
| COCA 3348/44 [CHA] | F6B E374 T2 (fp1047) | Forastero | Ecuador (wild) |
| CRIOLLO 22 [CRI] | F4A T1 | Criollo | Costa Rica (cultivated) |
| EET 400 [ECU]* | F6B F455 T1 | Forastero | Ecuador (cultivated) |
| ELP 1 | T6 (fp950) | Forastero | French Guiana (wild) |
| GU 241/P | UWI, Campus 1a, x2y33 (fp500) | Forastero | French Guiana (wild) |
| H 1 | | Forastero | Peru (cultivated) |
| HF 8 [BLZ] | T1 (fp987) | Criollo | Belize (wild) |
| IB 2 [BLZ]* | T1 (fp1020) | Criollo | Belize (wild) |
| IB 9 [BLZ] | T1 (fp996) | Criollo | Belize (wild) |
| ICS 75 | San Juan Estate, Block 2 | Trinitario | Trinidad (cultivated) |
| ICS 97* | San Juan Estate, Block 1 | Trinitario | Trinidad (cultivated) |
| ICS 100* | San Juan Estate, Block 2 | Trinitario | Trinidad (cultivated) |
| IMC 3* | UWI, Campus 3 | Forastero | Peru (wild) |
| IMC 12* | Marper Farm, C1056 | Forastero | Peru (wild) |
| IMC 16 | Marper Farm, D603 | Forastero | Peru (wild) |
| IMC 67 | La Reunion Estate | Forastero | Peru (wild) |
| JA 5/4 [POU] | Marper Farm, C526 (fp2307) | Refractario | Ecuador (cultivated) |
| JA 5/5 [POU] | Marper Farm, C324 (fp1351) | Refractario | Ecuador (cultivated) |
| LCT EEN 31 | F6A A6 T3 (fp450) | Forastero | Ecuador (wild) |
| MO 9 | Marper Farm, D835 (fp253) | Forastero | Peru (wild) |
| MO 20* | Marper Farm, D809 (fp254) | Forastero | Peru (wild) |
| MOQ 6/95* | Marper Farm, C1 (fp582) | Refractario | Ecuador (cultivated) |
| MXC 67 | UWI, Campus 12, x3y6 | Criollo | Mexico (cultivated) |
| NA 702* | Marper Farm, D104 (fp819) | Forastero | Peru (wild) |
| NAPO 2 [CHA]* | UWI, Campus 7, x8y9 (fp1922) | Forastero | Ecuador (wild) |
| PA 279 [PER] | Marper Farm, D59 (fp426) | Forastero | Peru (wild) |
| PA 299 [PER] | Marper Farm, C936 (fp571) | Forastero | Peru (wild) |
| POR 1 [TTO]* | UWI, Campus 2, x2y12 (fp1897) | Criollo | Venezuela |
| POUND 7/B [POU] | F6B F407 T3 (fp521) | Forastero | Peru (wild) |
| SCA 12* | Marper Farm, D205 | Forastero | Peru (wild) |
| SCA 24* | Marper Farm, D569 | Forastero | Peru (wild) |
| SPA 5 [COL] | UWI, Campus 2, x1y15 (fp1817) | Forastero | Colombia or Peru |
| U 1 | | Forastero | Peru (cultivated) |
| UF 613 | F4A A93 T2 (fp1237) | Trinitario | Costa Rica (cultivated) |
| YAL 6 | | Forastero | French Guiana (wild) |

*Accessions with missing allelic data that were not used for microsatellite statistical analyses.

SSR Amplification

The polymerase chain reaction (PCR) mix was composed of 4 □L Eppendorf HotMaster Mix (Brinkmann Instruments Inc., New York, USA) giving 2.5 mM $Mg^{2+}$, 2 mM total dNTP[1], 0.2 units of Taq polymerase activity at final composition in reaction mix; 0.5 □L of a MPP solution in TE[2] buffer (10 □M each primer; forward primers from Operon Technologies, Inc., Alabama, USA; reverse primers from Proligo Japan KK, Kyoto, Japan); and 5.5 □L of appropriate DNA solution. Each MPP was amplified from separate reaction mixes. Cycling was carried out in GeneAmp PCR System 9700 thermal cyclers (Applied Biosystems, California, USA). A touchdown protocol was utilised. Initial denaturation at 94°C for 5 mins was followed by eight cycles with denaturation at 95°C for 30 secs, annealing at 55°C for 60 secs with reduction by 0.5°C after every cycle and extension at 72°C for one minute. Then 25 cycles with denaturation at 94°C for 30 secs, annealing at 51°C for one minute and extension at 72°C for one minute were performed. A final extension step at 60°C for 15 mins was included to ensure complete adenine addition and the products were held at 4°C until recovery. Each combination of DNA-MPP was amplified at least once.

Capillary Electrophoresis (CE)

Post-PCR bulking was conducted by pooling 1.5 □L of each MPP-PCR product. Sample loading buffer containing 29.8 □L of Hi-Di formamide (Applied Biosystems, Warrington, UK) and 0.2 □L of GenomeLab™ DNA size standard-400 (Beckman Coulter Inc., California, USA) was added to each well. Samples were overlaid with one drop of mineral oil (Beckman Coulter Inc.). Fragments were separated on an 8-capillary CEQ 8000 or 8800 (Beckman Coulter Inc.) according to the manufacturer's recommendations. Products with poor standard profiles (missing bands; improper sizing) were discarded and the appropriate PCR product pools were recomposed and run again to ensure that fragment profiling was suitable for allele sizing. Raw fragment sizes were converted into alleles classes (binning) with the bundled fragment analysis software (Beckman Coulter Inc.).

MPP assessment

Twelve of the 15 recommended MPPs (Saunders *et al.*, 2004) and 22 additional MPPs (Table 2) were assessed on a set of 39 accessions (Table 1) from the Criollo, Forastero and Trinitario groups. PCR and CE were repeated once as described above. Allele binning was as described earlier. Allele data for the 34 MPPs over the 39 accessions were cleaned and sizing discrepancies were manually resolved. Eleven accessions had missing data for at least three loci and were removed from the main analyses; SCA 12 with one missing locus and SCA 24 with two missing loci were also excluded yielding a reduced database of 26 accessions in which two samples lacked allele data at two loci only. Summary statistics were obtained with PowerMarker v3.25 (Liu and Muse, 2005). Probability of identity (PID) values were obtained with GIMLET v.1.3.2 (Valière, 2002). Five groups of MPPs (Saunder's 12, GIMLET's top 12, Most Alleles (6 MPPs), Random (7 MPPs) and Minimal (4 MPPs)) were assessed for their capacity to differentiate by comparison of cumulative probability of identity ($PID_{com}$) values.

[1] Deoxyribonucleotide triphosphate
[2] Tri-EDTA (ethylenediamine tetraacetic acid)

**Table 2. Microsatellite loci information on 26 cacao accessions.**

| Loci[1] | Number of accessions[2] | Groups separated | Number of alleles | Allele Range (bp[3]) | PIC[4] | PID[5] |
|---|---|---|---|---|---|---|
| mTcCIR1 | 24 | 6 | 5 | 127-157 | 0.561 | 0.185 |
| mTcCIR3 | 26 | 12 | 8 | 212-277 | 0.775 | 0.047 |
| mTcCIR6 | 24 | 9 | 6 | 230-251 | 0.677 | 0.095 |
| mTcCIR7 | 26 | 9 | 5 | 155-163 | 0.620 | 0.141 |
| mTcCIR8 | 22 | 10 | 7 | 290-308 | 0.675 | 0.098 |
| mTcCIR9 | 26 | 9 | 8 | 260-298 | 0.706 | 0.085 |
| mTcCIR10 | 26 | 11 | 6 | 207-218 | 0.704 | 0.085 |
| mTcCIR11 | 26 | 11 | 8 | 286-319 | 0.711 | 0.079 |
| mTcCIR12 | 26 | 12 | 8 | 188-219 | 0.762 | 0.054 |
| mTcCIR15 | 26 | 13 | 10 | 239-264 | 0.769 | 0.039 |
| mTcCIR17 | 26 | 5 | 3 | 274-284 | 0.472 | 0.258 |
| mTcCIR18 | 24 | 11 | 8 | 333-357 | 0.756 | 0.056 |
| mTcCIR26 | 24 | 10 | 8 | 276-310 | 0.760 | 0.058 |
| mTcCIR29 | 26 | 10 | 6 | 160-176 | 0.726 | 0.075 |
| mTcCIR30 | 26 | 8 | 4 | 176-186 | 0.637 | 0.128 |
| mTcCIR33 | 25 | 15 | 10 | 275-347 | 0.832 | 0.025 |
| mTcCIR37 | 26 | 16 | 12 | 136-181 | 0.840 | 0.020 |
| mTcCIR42 | 26 | 12 | 8 | 205-240 | 0.728 | 0.062 |
| mTcCIR43 | 26 | 11 | 7 | 203-215 | 0.667 | 0.085 |
| mTcCIR45 | 26 | 7 | 4 | 288-294 | 0.531 | 0.201 |
| mTcCIR55 | 26 | 5 | 3 | 238-250 | 0.386 | 0.348 |
| mTcCIR56 | 26 | 9 | 6 | 317-368 | 0.702 | 0.087 |
| mTcCIR57 | 26 | 8 | 4 | 248-256 | 0.585 | 0.024 |
| mTcCIR58 | 26 | 16 | 12 | 209-324 | 0.829 | 0.066 |
| mTcCIR60 | 23 | 12 | 9 | 189-215 | 0.721 | 0.059 |
| mTcCIR184 | 26 | 12 | 7 | 118-147 | 0.748 | 0.059 |
| mTcCIR210 | 26 | 6 | 4 | 139-151 | 0.662 | 0.116 |
| mTcCIR229 | 26 | 10 | 7 | 311-327 | 0.669 | 0.103 |
| mTcCIR243 | 26 | 9 | 5 | 128-143 | 0.695 | 0.088 |
| mTcCIR244 | 26 | 13 | 9 | 241-272 | 0.754 | 0.055 |
| mTcCIR274 | 26 | 11 | 9 | 188-277 | 0.696 | 0.089 |
| mTcCIR278 | 26 | 3 | 2 | 99-101 | 0.361 | 0.377 |
| S012 | 26 | 5 | 5 | 264-284 | 0.578 | 0.173 |
| S016 | 26 | 7 | 5 | 202-222 | 0.581 | 0.161 |

[1]Microsatellite code; [2]Number of accessions with allele data; [3]Base pair
[4]Polymorphic information content; [5]Probability of identity
Number of alleles, range and PIC obtained from PowerMarker v3.25 (Liu and Muse, 2005).
PID obtained from GIMLET v1.3.2 (Valière, 2002).

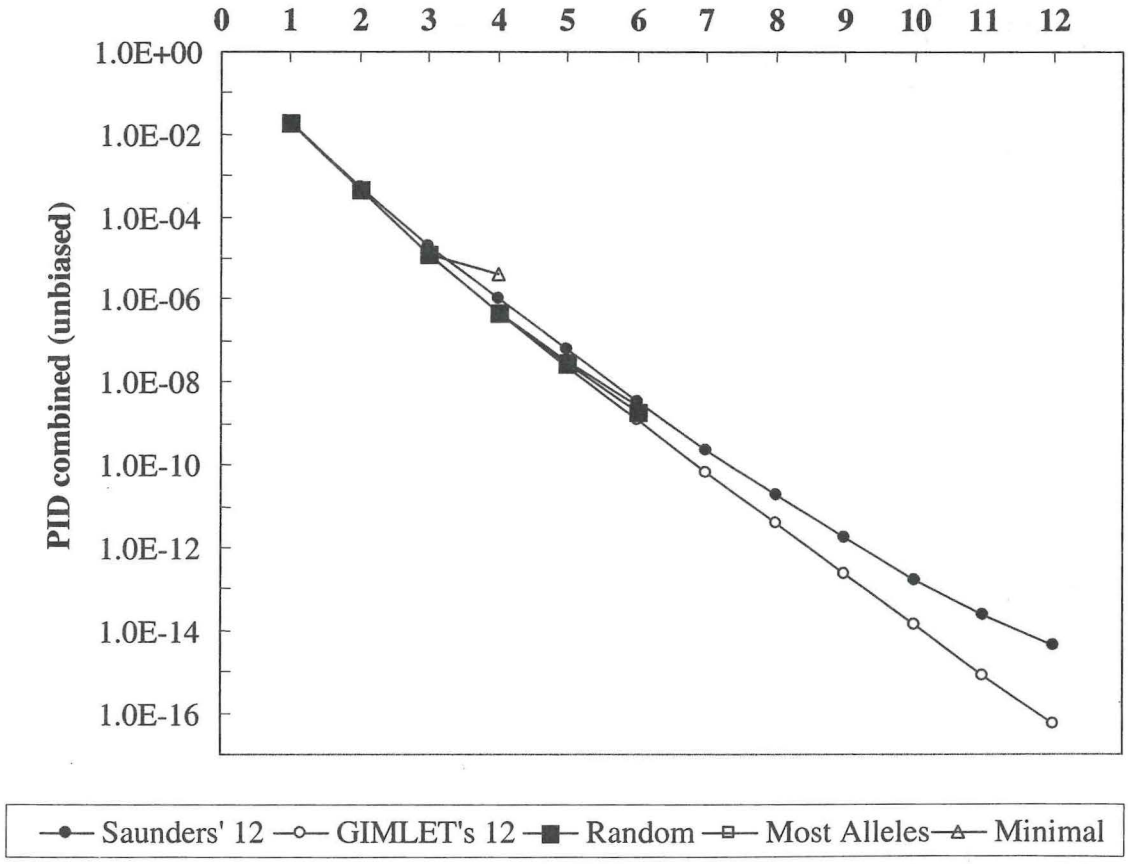
## Results

Characteristics of the individual MPPs based on 26 accessions are provided in Table 2. The MPPs, mTcCIR37 and mTcCIR58 generated the most alleles (12) followed by mTcCIR15 and mTcCIR33 (10) and then mTcCIR60, mTcCIR244 and mTcCIR274 with nine alleles. The separation ability of the MPPs was significantly correlated (large, positive; $r = 0.93$; $P<0.001$) with the number of alleles. The PID values from GIMLET v1.3.2 (Valière, 2002) enabled ranking of the differentiating power of the MPPs. The twelve primers recommended by Saunders *et al.* (2004) were ranked as mTcCIR37, 33, 15, 12, 18, 26, 60, 11, 6, 8, 7 and 1 from most to least

differentiating. The twelve top ranked from the full complement of 34 MPPs were mTcCIR 37, 58, 33, 15, 3, 12, 244, 18, 26, 184, 42 and 60 and included seven of the MPPs recommended by Saunders *et al.* (2004).

As the number of loci increased, the probability of grouping together more than one different accession was decreased (Figure 1), with MPPs from Random, Most Alleles and GIMLET's 12 achieving lower $PID_{com}$ values at a faster rate. The latter gave the most confident result (lowest probability of having two accessions that match each other). Two groups of accessions, each with two samples (BC 3 and HF 8; AC 20 and IB 9) could not be resolved with the primers recommended by Saunders *et al.* (2004) or with the full complement of 34 primers applied to the dataset of 26 accessions (Table 3). The random set of MPPs was unable to resolve the accessions to the same extent as the other MPP groupings even though the number of loci used was greater than that of two other groups. Three MPPs (mTcCIR 33, 37 and 58) were sufficient to assign the accessions to the same identity groupings as the full complement of MPPs used on the same 26 accessions.

**Figure 1. Probability of identity (PID) combined over increasing loci number for different sets of microsatellite loci.**




Lower values indicate an increased probability that the accessions are different; output from GIMLET v1.3.2 (Valière, 2002). Primer sets were: GIMLET's 12 = mTcCIR 3,12,15,18,26,33,37,42,58,60,184,244; Minimal = mTcCIR 33,37,55,58; Most Alleles = mTcCIR 15,33,37,58,60,184; Random = mTcCIR 6,9,10,45,57,229,243; and Saunder's 12 = mTcCIR 1,6,7,8,11,12,15,18,26,33,37,60;

**Table 3. Differentiation ability of select microsatellite loci groupings on 26 accessions.**

| MPP group[1] | Number separated | % separation | $PID_{com}$[2] | Unresolved accessions |
|---|---|---|---|---|
| All 34 loci | 24 | 92.3 | $1.49 \times 10^{-36}$ | (AC 20 T1, IB 9 T1); (BC 3 T1; HF 8 T1) |
| Saunder's 12 | 24 | 92.3 | $4.42 \times 10^{-15}$ | (AC 20 T1, IB 9 T1); (BC 3 T1; HF 8 T1) |
| GIMLET's 12 | 24 | 92.3 | $5.31 \times 10^{-17}$ | (AC 20 T1, IB 9 T1); (BC 3 T1; HF 8 T1) |
| Random | 20 | 76.9 | $2.02 \times 10^{-7}$ | (AC 2 T1, AC 20 T1, BC 3 T1, CRIOLLO 22 F4AT1, HF 8 T1, IB 9 T1); (ELP 1 T6, GU 241/P) |
| Most Alleles | 24 | 92.3 | $2.80 \times 10^{-9}$ | (AC 20 T1, IB 9 T1); (BC 3 T1; HF 8 T1) |
| Minimal | 24 | 92.3 | $4.22 \times 10^{-6}$ | (AC 20 T1, IB 9 T1); (BC 3 T1; HF 8 T1) |

[1]Microsatellite primer pairs; Saunder's 12 = mTcCIR 1,6,7,8,11,12,15,18,26,33,37,60;
GIMLET's 12 = mTcCIR 3,12,15,18,26,33,37,42,58,60,184,244; Random = mTcCIR 6,9,10,45,57,229,243;
Most Alleles = mTcCIR 15,33,37,58,60,184; Minimal = mTcCIR 33,37,55,58.
[2]Combined probability of identity obtained from GIMLET v1.3.2 (Valière, 2002).

When all 39 accessions were utilised a similar result occurred, even though missing values (17 accessions/102 missing alleles) were present in the dataset. The accession IB 2 T1 could not be clearly differentiated in the primer groups Saunder's 12, GIMLET's top 12, Most Alleles, Random and Minimal due to the presence of missing data for the loci mTcCIR 1, 11, 12, 26 and 33. The full complement of primers allocated this accession to a unique group based on the exclusive allele pattern obtained from mTcCIR30. Excluding the identical accessions (HF 8/BC 3; IB 9/AC 20), there were 14 accession pairs that had more than 25 identical loci. Three and eight pairs of Criollo material exhibited 31 and 32 matching loci respectively and the accession pair JA 5/4 vs. JA 5/5 had 27 matching loci. Resolution of the Criollo pairs was achieved with the primers mTcCIR26 (5/11 pairs), mTcCIR33 (10/11 pairs), mTcCIR37 (6/11 pairs) and mTcCIR55 (4/11 pairs). JA 5/4 was resolved from JA 5/5 by seven primers (mTcCIR 7, 8, 29, 45, 58, 229 and 274).

The accession NA 702 was homozygous at 26 of 31 loci (mTcCIR11, 12 and 26 with missing data). In-house data from the USDA/CRU fingerprinting project confirmed the high homozygosity of this accession and provided the missing data to reveal 28 of 34 loci (82.4%) to be homozygous. The MPPs mTcCIR 10, 12, 29, 42, 60 and 274 were heterozygous in this accession.

## Discussion

This study examined the possibility of accelerating verification of identities in genebanks by decreasing the number of genotyped loci. Subsumed within this objective was the identification of loci which were most suited for differentiating cacao accessions. Successful and efficient verification of identities in a germplasm collection relies on the judicious use of microsatellite loci. The loci chosen must be able to differentiate among existing and future accession holdings. Furthermore, loci should be used that would maximise differences among accessions. The latter is especially important when highly homozygous material need to be surveyed for mislabelled plants. Results of the primer survey suggested that the primers recommended by Saunders *et al.* (2004) were good discriminatory loci and that seven of these (mTcCIR 12, 15, 18, 26, 33, 37 and 60) were the most useful. Interestingly, three loci (mTcCIR33, 37 and 58) resolved the 26 accessions into the identical groupings generated by a set of 12 of the fifteen primers recommended by Saunders *et al.* (2004).

Furthermore the random set of seven primers underperformed in comparison to other primer

combinations, suggesting that the composition of the MPP set is more important than the number of loci that is used. The latter is important when a low PID is required as increasing the number of loci reduces the match probability. Zhang *et al.* (2006) demonstrated that seven loci could have sufficient differentiation power for cacao accessions. In-house CRU/UWI fingerprinting data revealed that the accessions NA 228 (fp1) and NA 266 (fp25) could not be resolved from each other with the 15 primers recommended by Saunders *et al.* (2004) although the MPPs mTcCIR6 and mTcCIR8 were useful in separating NA accessions from one another. Preliminary work (data not shown) indicated that mTcCIR57, 229 and 243 separated NA 228 from NA 266. This suggests that Type II errors (accessions declared similar when they are really different) may be overlooked in verification work and that this error is due not only to the number of loci used but, more importantly, the types of loci utilised for differentiation purposes.

**Table 4. Recommended MPPs for verification work.**

| Loci | Allele range | Suggested dye | Loci note and advantage |
|---|---|---|---|
| **Set L1** | | | |
| mTcCIR26 | 276-310 | Green | 9$^{th}$ ranked locus; resolves Criollo material |
| mTcCIR33 | 275-347 | Black | 3$^{rd}$ ranked locus; resolves Criollo material |
| mTcCIR37 | 136-181 | Black | 1$^{st}$ ranked locus; resolves Criollo material |
| mTcCIR55 | 238-250 | Blue | 33$^{rd}$ ranked locus; resolves Criollo material |
| **Set L2** | | | |
| mTcCIR8 | 290-308 | Green | 22$^{nd}$ ranked locus; differentiates JA 5/4 from JA 5/5; resolves NA material |
| mTcCIR57 | 248-256 | Blue | 27$^{th}$ ranked locus; resolves NA accessions |
| mTcCIR58 | 209-324 | Black | 2$^{nd}$ ranked locus; differentiates JA 5/4 from JA 5/5 |
| mTcCIR229 | 311-327 | Blue | 23$^{rd}$ ranked locus; differentiates JA 5/4 from JA 5/5; resolves NA accessions |

Loci ranking obtained from GIMLET v1.3.2 (Valière, 2002).

It is suggested that a set of eight MPPs could be used for verification purposes in the ICG,T and that electrophoresis of amplified fragments could be performed by pooling products from four MPPs. In this way, both the number of PCR and CE runs can be substantially reduced. The recommended primers are presented in Table 4. The inclusion of low ranked loci (based on PID values) was justified by their utility in discriminating amongst homozygous material. With this recommended set of primers used on the dataset of 26 accessions, a combined PID of $1.64 \times 10^{-10}$ would be obtained providing sufficient confidence in match declarations. The adoption of these primers by the cacao community would allow for more rapid and definitive resolution of identities in other cacao germplasm collections. Accessions declared as similar under these circumstances may be considered to be genetically similar (e.g. BC 3/HF 8 and AC 20/ IB 9) and unless morphological or agronomic evidence indicates otherwise may be grouped together as the same accession. However, since the possibility always exists that some other primer (existing or future) may separate these accessions, genebank curators may prefer to keep the accessions separate but flag these as being synonymies.

**Acknowledgements**

Stephen Pinney, Sue Mischke and Eric Tillson are thanked for their assistance in the laboratory and for suggestions during this work.

## References

Badenes, M., Garcés, A., Romero, C., Romero, M., Clavé, J., Rovira, M. and Llácer, G. (2003) Genetic diversity of introduced and local Spanish persimmon cultivars revealed by RAPD markers. *Genet. Resour. Crop Evol.* **50**: 579-585.

Cryer, N.C., Fenn, M.G.E., Turnbull, C.J. and Wilkinson, M.J. (2006) Allelic size standards and reference genotypes to unify international cocoa (*Theobroma cacao* L.) microsatellite data. *Genet. Resour. Crop Evol.* **53**: 1643-1652.

Figueira, A. (1998) Homononymous genotypes and misidentification in germplasm collections of Brazil and Malaysia. *INGENIC Newsletter* **4**: 4-8.

Hurka, H., Neuffer, B. and Friesen, N. (2004) Plant genetic resources in botanical gardens. Pages 35-44 in: *Proc. 21st International Symposium on Breeding Ornamentals, Part II* (G. Forkmann, S. Michaelis Eds) *Acta. Hort.* **651.**

Kennedy, A.J. and Mooleedhar, V. (1993) Conservation of cocoa in field genebanks – the International Cocoa Genebank, Trinidad. In: *Proc. Intl. Workshop on Conservation,Characterisation and Utilisation of Cocoa Genetic Resources in the 21st Century*, Port of Spain, Trinidad, September 13-17, 1992, The Cocoa Research Unit, Port of Spain, Trinidad, pp. 21-23.

Khadari, B., Breton, C., Moutier, N., Roger, J.P., Besnard, G., Bervillé, A. and Dosba, F. (2003) The use of molecular markers for germplasm management in a French olive collection *Theor. Appl. Genet.* **106**: 521-529.

Kobayashi, N., Horikoshi, T., Katsuyama, H., Handa, T. and Takayanagi, K. (1998) A simple and efficient DNA extraction method for plants, especially woody plants. *Plant Tissue Culture Biotech.* **4(2)**:76-80.

Lanaud, C., Risterucci, A.M., Pieretti, I., Falque, M., Bouet, A. and Lagoda, P.J.L. (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. *Mol. Ecol.* **8**: 2141-2152.

Liu, K. and Muse, S.V. (2005) PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**: 2128-2129.

Morgante, M. and Olivieri, A.M. (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant J.* **3**: 175-182.

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

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. and Rafalski, A. (1996) The comparison of RFPL, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* **2**: 225-238.

Qiagen GmbH. (2000) DNeasy Plant Mini Kit and DNeasy MaxiKit Handbook 08/2000. Germany: Qiagen GmbH.

Risterucci, A.M., Eskes, B., Fargeas, D., Motamayor, J.C. and Lanaud, C. (2001) Use of microsatellite markers for germplasm identity analysis in cocoa. Pages 25-33 in: *Proceedings of the 3rd International Group for Genetic Improvement of Cocoa (INGENIC) International Workshop on the New Technologies and Cocoa Breeding.* 16th-17th October 2000, Kota Kinabalu, Malaysia: INGENIC, UK.

Saunders, J.A., Hemeida, A.A. and Mischke, S. (2004) Selection of international molecular standard for DNA fingerprinting. *Theor. Appl. Genet.* **110**: 41-47.

Shan, F., Clarke, H.C., Plummer, J.A., Yan, G. and Siddique, K.H.M. (2005) Geographical patterns of genetic variation in the world collections of wild annual *Cicer* characterized by amplified fragment length polymorphisms. *Theor. Appl. Genet.* **110**: 381-391.

Valière, N. (2002) GIMLET: A computer program for analyzing genetic individual identification data. *Mol. Ecol. Notes* **2**: 377-379.

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