Identification of Cocoa (*Theobroma cacao* L.) Varieties with Different Quality Attributes and Parentage Analysis of Their Beans

M.J.M. Smulders,[1] D. Esselink,[1] F. Amores,[2] G. Ramos,[3] D.A. Sukha,[4] D.R. Butler,[4] B. Vosman,[1] and E.N. van Loo[1]

[1] Plant Research International, Wageningen UR, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands
[2] Estaciòn Experimental Tropical Pichilingue, INIAP, P.O. Box 24, Quevedo, Ecuador
[3] Instituto de Investigaciones Agricola (INIA-Mérida), P.O. Box 425, Mérida, Venezuela
[4] Cocoa Research Unit, The University of the West Indies, St. Augustine, Trinidad, Rep. of Trinidad and Tobago, West Indies

Corresponding author:
M.J.M. Smulders, Plant Research International, Wageningen UR, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands, Tel. +31 317 480840
Fax +31 317 418095
Email rene.smulders@wur.nl

Abstract

We applied fifteen microsatellite markers for molecular identification of cocoa (*Theobroma cacao* L.) varieties and their beans in the production chain, and for distinguishing 'fine or flavour' cocoa from bulk varieties. Thirteen markers were from Lanaud *et al. (*1999) and listed in the International Cocoa Germplasm Database. All of these mTcCIR markers are dinucleotide repeat markers, which means that the amplification of stutter bands alone may make the genotyping less reliable. We added two markers developed on a tri- and a pentanucleotide repeat. The advantage of these is that the alleles are at least three or five nucleotides apart, which facilitates scoring. In our set of varieties tested, the most polymorphic markers were mTcCIR6, mTcCIR12, mTcCIR15, mTcCIR18, mTcCIR33 and mTcCIR37, which had an effective number of alleles of 4 or more. Among representative varieties of the Nacional, Trinitario and Criollo types, most varieties were clearly differentiated, while duplicate samples from different countries had identical patterns, although some mislabelling occurred. Bulk and 'fine or flavour' varieties were clearly distinguished, but they did not form two genetically distinct groups. The genotypes of ICS 1 beans from different plantations indicated various numbers of pollinators and proportions of self-pollination. The bean genotypes enabled reconstruction of the genotype of the maternal parent, and they contained information on the number and type of pollen donors of each plantation. If plantations have unique genotypes or combinations of genotypes, it may be possible to distinguish them based on the combined genotypes of a set of beans.

Introduction

In cocoa (*Theobroma cacao* L.), originating from tropical rain forests of South America, three domesticated groups are distinguished: Criollo, Forastero and a hybrid group, Trinitario (Cheesman 1944; Motamayor and Lanaud 2002). Generally, 'fine or flavour' cocoa beans are produced from Criollo or Trinitario varieties, while 'bulk' cocoa beans come from Forastero trees, but there are exceptions. Nacional trees in Ecuador (considered to be Forastero by some, but with traits distinguishing them from all other groups (Enríquez 1993) produce fine or flavour cocoa, while Cameroon cocoa beans, which are produced by Trinitario trees and whose cocoa powder has a distinct and sought-after red colour, are classified as bulk cocoa beans (www.ICCO.org).

In the cocoa trade, fine or flavour beans are distinguished by organoleptic tests, which are expensive, difficult to standardise, and therefore not easy to use in the cocoa production chain. Better, more objective, or faster methods of identification of fine or flavour cocoa need to be developed, which would benefit (i) smallholder farmers, who would receive higher prices for their beans; (ii) cocoa traders and exporters, who would experience fewer problems in standardising, grading and certifying batches of 'fine or flavour' cocoa beans at the source of origin; (iii) manufacturers, who would see a reduction in the risks associated with buying products from sources where supplies are currently unreliable, with regard to both quality and quantity of fine or flavour beans. Since poor quality beans directly affect the quality and cost of the end-product, consumers would benefit from more consistent, high quality chocolate as well.

One possibility for a faster method to identify fine or flavour cocoa would be to genetically identify the varieties, and in this way distinguish varieties that produce 'fine or flavour' cocoa from those that produce bulk cocoa using DNA fingerprinting techniques. Among the molecular marker techniques available, microsatellite markers have been developed for a large number of species, and have been applied for routine testing of varieties and the design of databases of profiles of hundreds of varieties, including selfing species such as tomato (Bredemeijer *et al*. 2002) and wheat (Röder *et al*. 2002), and outbreeding species such as rose (Esselink *et al*. 2003) and apple (Liebhard *et al*. 2002). A number of microsatellite markers have been developed for cocoa (Lanaud *et al.* 1999; Risterucci *et al.* 2000; Pugh *et al*. 2004). Saunders *et al.* (2004) proposed a selection of 15 microsatellite markers to be used for identification of clones in cocoa. The International Cocoa Germplasm Database (Wadsworth and Harwood 2000) contains DNA profiles of a small number of cocoa varieties, for which the DNA was extracted from leaf samples.

Since it is cocoa beans that are traded, a system to trace the genetic character of material in the production chain should allow the source of beans to be identified. *T. cacao* has a unique incompatibility system, in which the incompatibility reaction does not occur at the style and stigma or by inhibition of pollen tube growth, but by failure of gamete nuclei fusion in the embryo sac (Cope 1962; Alves *et al*. 2003). Nevertheless, nearly homozygous lines do exist (Lanaud *et al*. 2001), and a number of Criollo and Trinitario varieties are compatible with ICS 1, which is self-compatible. After cross-hybridisation offspring will inherit half the alleles from the maternal parent, and the other half from another plant on the plantation that acted as pollen donor. The quality of the beans is predominantly influenced by the maternal parent (Clapperton *et al.* 1994; Sukha 2008), so it would be valuable to identify this through genotyping.

We did not find reports of cocoa bean genotyping in the literature. We therefore have tested a set of microsatellite primer pairs on representative varieties of the Nacional, Trinitario and Criollo groups, using material from four different countries (Ecuador, Venezuela, Trinidad, and Papua New Guinea). It was found that the markers could be used for variety identification, including the distinction of bulk and fine varieties. Beans can be genotyped during the different stages of cocoa production and processing, a process that yields information on both parents. In addition, chocolate can be genotyped.

Material and methods

Collection of material

Leaf material was collected on plantations from Trinidad (8 varieties), Ecuador (18), Papua New Guinea (9) and Venezuela (7) (Table 1). Leaves were dried at ambient temperature, stored in paper bags and sent to Plant Research International, The Netherlands.

Beans harvested from ICS 1 were taken directly from pods of ICS 1 plants grown in field locations in Papua New Guinea, Venezuela, and Trinidad. For Trinidad, additional information was collected on the other varieties grown at the four sample locations, allowing the identification of additional possible pollen donors. The first field was located at the

University of the West Indies (UWI) St. Augustine Campus, in a field designated UWI Campus 4. The ICS 1 trees located in this field were in close proximity to IMC 67 and ICS 95 trees as well as the following other ICS accessions: ICS 6, 8, 16, 40, 43, 53, 60, 61 and 84. Therefore the ICS 1 flowers could have been pollinated with pollen from ICS 1 (which is self-compatible), IMC 67, ICS 95 or any of the other ICS accessions in close proximity. The second field was located at the La Reunion Estate, Centeno, which is part of the Ministry of Agriculture, Land and Marine Resources Central Experimental Station. This field is designated La Reunion Field 13, and a commercial clone, CCL 200, is located in another field approximately 60 metres away and up wind from Field 13. CCL 200 is also grown with other commercial clones in a field adjacent to Field 13 and down wind. These two fields are separated by a 1.5 m grass trace. It is therefore possible that pollen and pollinators could be carried between these fields either by wind or vectors. The two other fields, from which ICS 1 pods were harvested, were located at UWI Campus Field 14, in close proximity to West African Amelonado (WAA), ICS 95 and ICS 60 trees, and at the International Cocoa Genebank, Trinidad (ICG,T), where the ICS 1 trees were surrounded by ICS 2, 4, 5, 13, and 49.

Five small bars of chocolate made from beans of plants of ICS 1, ICS 95, IMC 67, EET 400, or CCL 200 were provided by the Cocoa Research Unit, The University of the West Indies.

DNA extraction

DNA from leaves, beans (using only the germ or cotyledon), as well as chocolate bars (1 gramme, as in Gryson *et al*. 2004) was extracted using a QIAamp DNA Stool Mini Kit (Qiagen), originally designed for the isolation of total DNA from stool samples. Stool samples typically contain many compounds that can degrade DNA and inhibit enzymatic reactions. The beans were at various stages of fermentation. With this kit, the DNA obtained from leaves and beans was of homogeneous quality and reliable in the PCR assays, and all samples could be genotyped without drop-outs. Chocolate bars yielded only traces of DNA.

Microsatellite markers

We used 15 dinucleotide repeat microsatellite markers developed by Lanaud *et al.* (1999) and included in the International Cocoa Germplasm Database: mTcCIR1, TcCIR6, mTcCIR7, mTcCIR8, mTcCIR11, mTcCIR12, mTcCIR15, mTcCIR18, mTcCIR22, mTcCIR24, mTcCIR26, mTcCIR33, mTcCIR37, mTcCIR40, and mTcCIR60. In addition, we included the trinucleotide repeat microsatellite marker SHRSTc12 (AY389500) and the pentanucleotide repeat microsatellite marker SHRSTc16 (AY389503), both also used by Schnell *et al.* (2004). Reverse primers were PIG-tailed (Brownstein *et al*. 1996) to increase the scorabilty of the profiles (Bredemeijer *et al.* 1998). Markers were tested with eight Trinidad genotypes for pattern quality and polymorphism in order to select quality 1 or 2 markers (no or few stutter bands) according to Smulders *et al.* (1997). Amplification was done on a MJ PTC200 using the conditions of Esselink *et al.* (2003) for rose microsatellite markers, including a uniform Tm of 50 ^{o}C. Genotyping was done on an ABI 3700, and the results were analysed using Genotyper (Applied Biosystems). Some additional microsatellite profiles were obtained from the International Cocoa Germplasm Database (http://www.icgd.reading.ac.uk/public_html/ms-index.htm).

Data analysis

Observed and expected heterozygosity, observed and effective number of alleles, and Nei's 1972 genetic distance between each pair of samples were calculated using POPGENE (Yeh and Boyle 1997). A dendrogram was generated with SAHN (Sequential agglomerative hierarchical nested cluster analysis) in NTSYSpc 2.10j (Applied Biostatistics). Maternal and

paternal parent contributions in beans were analysed manually, using data from the following eight markers: mTcCIR11, mTcCIR12, mTcCIR15, mTcCIR18, mTcCIR26, mTcCIR33, mTcCIR37, and mTcCIR60.

Results

DNA extraction of field-collected samples

Leaf material was collected from plantations in four countries (Table 1). It was dried at ambient temperature. As a result, some dried leaf samples were still green, while others were brown to dark-brown. This caused variation in the quality of DNA when extracted with the protocols of the DNeasy Plant Kit, as was used by Lanaud *et al*. (1999) and Saunders *et al*. (2004). As a consequence, a large number of samples from some countries gave no amplification with most or all microsatellite markers. Also, other DNA isolation protocols (Fulton *et al.* 1995; Shure *et al*. 1983) failed. However, using a QIAamp Stool kit, originally designed for DNA extractions from fresh or frozen human stool or other sample types with high concentrations of PCR inhibitors, and used by Tengel *et al*. (2001) for the isolation from DNA from highly processed foodstuffs including those containing processed cocoa, resulted in DNA samples of homogeneous quality. In this way, we could generate microsatellite patterns without drop-out of loci, regardless of the leaf colour. Also the fermentation stage of the beans (0-8 days fermentation tested) had no effect on the microsatellite amplifications.

Microsatellite markers

If the long-term aim is to set up a database of microsatellite patterns for a large number of cultivars, it is important to establish protocols for widespread use so that a network of different laboratories can feed data into it (Bredemeijer *et al.* 2002). Hence, selected markers should be easily reproducible in any laboratory. When we tested the set of 15 microsatellites as recommended by Saunders *et al.* (2004), we found that two markers were not satisfactory for generating a database of variety profiles, either because of low level of polymorphism (mTcCIR1) or because the patterns were too complex for unambiguously scoring (mTcCIR7) as assessed with silver staining. Therefore these two markers were discarded. The other markers revealed patterns that could be scored unambiguously (quality 1 or 2 of Smulders *et al.* 1997), although mTcCIR22 amplified a-specific bands (quality 4), and therefore did not generate a clear pattern. New primer design or even another batch of primers might be able to solve this. For marker mTcCIR40, it was not clear whether 1 or 2 alleles should be assigned even after repeated amplifications of samples KEE 43 and clone 16-2/3.

We added two markers developed on a tri- and a pentanucleotide repeat (Schnell *et al*. 2004). The advantage of these is that the alleles are at least three or five nucleotides apart, which facilitates scoring. All mTcCIR markers from CIRAD (Lanaud *et al.* 1999; Saunders *et al.* 2004) are dinucleotide repeat markers, which means that the amplification of stutter bands alone may make the genotyping less reliable.

In our set of varieties tested, the most polymorphic markers were mTcCIR6, -12, -15, -18, -33 and -37, which had an effective number of alleles > 4 (Table 2).

Identification of varieties

Using the revised set of 15 microsatellite markers (Table 2), all samples could be distinguished by a unique pattern with as few as 3 microsatellite markers (*e.g*., the combination of mTcCIR26, -33, and -37), or they were completely identical (Figure 1). The only exception was a couple of highly related samples (EET 48 and the nearly identical samples EET 62 and EET 96), which differed only with markers mTcCIR8 and -26. Identical samples of ICS 1 and IMC 67 were obtained from three different countries each, and identical samples of EET 400, ICS 95, and SCA 6 from two different countries each. Our

results confirm that these plants have been vegetatively propagated from one clone each, and that no mislabelling had taken place. In addition, Porcelana “estufa” and Guasare from Venezuela also had identical profiles. Several EET and CCAT lines sampled in Ecuador fell into two groups of 5 and 3 identical genotypes. Here, mislabelling or exchange of samples may have taken place.

Some markers contained alleles that are present in only one genotype. The loci mTcCIR26, -37, and -60 contained a unique allele for SCA 6, mTcCIR40 for IMC67, and mTcCIR40 for CCN 51. There are no alleles unique for the group of ‘bulk’ cocoas (SCA 6 and IMC 67 (Forastero types), as well as CCN 51).

Parentage analysis of beans

The three sets of 12 beans of ICS 1 that we genotyped always contained at least one allele per locus that is present in the ICS 1 variety, thereby confirming the mother plant (Table 3). The other alleles found do not allow us to reconstruct the genotypes of the pollen donors, since we do not know the number of different pollen donors involved. We can estimate that the minimum number of different pollen donors is 2 (Venezuela beans), 4 (Papua New Guinea), and 5 (Trinidad).

Other varieties from the same country as the beans we examined and genotyped were potential pollen donors for Venezuela and Trinidad, but not for the beans from Papua New Guinea. This indicates that the combination of varieties on a plantation may lead to distinct patterns of pollen donor alleles.

Putative pollination of ICS 1 by ICS 1 was observed in beans from Venezuela and Trinidad, but not from Papua New Guinea. These beans had an increased level of homozygosity (with as much as 6 out of 7 loci in homozygous state in one bean), which is expected, as it is essentially self-fertilisation. The difference in frequency of self-fertilisation may be due to differences in the range of varieties planted on the plantations, in the relative number of plants per variety, or in the local conditions of the mother plant sampled here.

Since we had detailed information on the other varieties planted in nearby fields in Trinidad, we also determined whether there could be potential pollen donors among those varieties, based on information on the genotypes of these varieties at 5-7 common microsatellite loci that is present in the International Cocoa Germplasm Database. This produced many possible pollen donors for plants that also had ICS 1 as potential male parent, reflecting the fact that many ICS lines selected by Pound (1934) share some alleles, and that a large number of hybrid types, partly derived from ICS 1, exist (Wadsworth and Harwood 2000).

Analysis of DNA isolated from chocolate bars

We were able to extract trace amounts of DNA from chocolate bars, and amplify the microsatellite markers using this DNA. We amplified the same alleles as were identified using leaf DNA from the set of cocoa varieties. The marker profiles of the chocolate bars, which had been made from beans of one variety, sometimes contained more than two alleles, with variable peak heights, which is consistent with the contribution of multiple pollen donors to the beans. However, often only one of the two alleles from the maternal plant was present, and sometimes none. This is most likely due to failure to amplify microsatellite alleles from very small amounts of DNA, such as was also observed for *e.g.* DNA from hairs (Taberlet *et al.* 1997). Clearly, the DNA extraction of chocolate bars needs to be improved further, in combination with taking measures to avoid or compensate for allele drop-out.

Discussion

We have tested a set of microsatellite primer pairs on representative varieties of the Nacional, Trinitario and Criollo types, using leaf samples from four different countries

(Ecuador, Venezuela, Trinidad, and Papua New Guinea). These markers can be used for variety identification. Duplicate samples from different countries had identical patterns, while most varieties were clearly differentiated. The only exceptions were groups of EET and CCAT samples from Ecuador, confirming that mislabelling is a problem (Motilal and Butler 2003; Turnbull *et al*. 2004).

We were unable to genotype reliably using some of the markers from the set advised by Saunders *et al.* (2004). In our experience, tri- and tetranucleotide are easier to score unambiguously and to be reproducibly used across laboratories, as the alleles are more base pairs apart and they suffer less from stutter bands generated by the PCR amplification (Smulders *et al.* 2002). Saunders *et al.* (2004) and Motilal and Boccara (2004) also reported problems with amplification of the set of cocoa markers across laboratories, so it appears that a final set of robust cocoa microsatellite markers has not yet been established. Indeed, Schnell *et al.* (2004) have tested as many as 34 markers, and Motilal and Boccara (2004) proposed an alternative set of markers.

Bulk and 'fine or flavour' varieties were clearly distinguished based on the fingerprint data (Figure 1), but they did not form two genetically distinct groups. This may be expected because of the history of hybridisations between types, and because of the fact that the distinction between bulk and fine or flavour is based on appreciated characteristics for certain types of use. Using larger sets of samples, it may be that multiple genetically related groups of bulk and fine or flavour cocoa can be distinguished. In addition, it may be possible to distinguish these groups based on the metabolites present.

For use in quality control, genotyping has to be performed on beans of various degrees of fermentation. We found that the Qiagen stool kit reproducibly produced DNA of a suitable quality for genotyping, and for all degrees of fermentation of the beans. The genotypes of the beans allowed us to reconstruct the maternal variety, and to obtain information on the number and type of pollen donors in the plantation. We observed different numbers of pollinators and various proportions of ICS 1 × ICS 1 self-pollinations across plantations. This is consistent with findings by Lanaud *et al.* (1987), who observed that the percentage of self-fertilisation depends on ecological factors, and who cited other reports, which observed seasonal effects as well, and Sereno *et al.* (2006), who found lower than expected levels of heterozygosity in natural populations of *T. cacao*. It is in contrast to the wild species *T. grandiflorum*, which appears to be an obligatory outbreeder (Alves *et al.* 2003).

Is it possible to infer the variety from fingerprints of the beans? In cases where mixtures of beans are studied with pods from one or several trees of the same variety in one field, then in general more than one pollen donor will be involved. As a result, the DNA profile of the maternal plant will be present at a higher frequency than those of the male parents. For example, if we take the results from the Trinidad beans, which include possibly five ICS 1-ICS 1 pollinations, as representative of an average plantation on the island, then we would obtain a ratio of 17 ICS 1 to 7 other DNA profiles. The other DNA profiles were a combination of at least three different genotypes, with different allelic composition. Hence, we would essentially obtain an ICS 1 genotype when mixtures of entire beans from one tree are being genotyped, with minor peaks for other alleles from various pollen donors. This implies that in general it will be possible to determine the variety from the combination of genotypes of several beans from one or more pods on one plant. It may even be possible to do this using DNA extraction from pooled beans, but we have not tested this. In normal practice, the beans derived from any one plantation are mixed before being sold. If plantations have unique combinations of varieties, we think that it may be possible to distinguish plantations based on the combined genotypes of a set of beans. However, further down the commodity chain the products from several plantations may be mixed, as occurs with commercial samples for export. Such pooling would obscure precise reconstruction to the level of a plantation. At most, we would be able to infer the percentages of the major varieties used.

The fact that we can differentiate the pollen donor contribution will allow us, for the first time, to question whether beans derived from different pollen donors have a different intrinsic quality. Using the molecular marker set, it is now possible to answer such a

question, since we can genotype beans from one variety, assign them to various pollen donors, and then assess the quality of the cocoa of each group. However, the quality of the cocoa is determined not only by the (intrinsic) quality of the variety, but also by growing conditions, ripeness at harvest, proper fermentation and drying, and good storage conditions during transport. To incorporate this, direct measurements of the organoleptic characteristics will have to be done as well.

Acknowledgements

The project was funded by the Common Fund for Commodities, the Netherlands (CFC) as part of the CFC/ICCO/INIAP 'Project to establish physical, chemical and organoleptic parameters to differentiate between bulk and fine or flavour cocoa' (http://www.icco.org/projects/fine.htm). We thank the technical staff in each project member country for collecting and preparing leaf samples, Neil Hollywood (PNG Cocoa & Coconut Research Institute CCRI) for co-ordination of the sampling in Papua New Guinea, and Claire Lanaud for information on the microsatellite markers used.

References

Alves, R.M., A.S. Artero, A.M. Sebbenn, and A. Figueira. 2003. Mating system in a natural population of *Theobroma grandiflorum* (Willd. ex Spreng.) Schum., by microsatellite markers. Genet. Mol. Biol. **26**: 373-379.
http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572003000300025&lng=es&nrm=iso

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

Bredemeijer, G.M.M., R.J. Cooke, M.W. Ganal, R. Peeters, P. Isaac, Y. Noordijk, S. Rendell, J. Jackson, M.S. Röder, K. Wendehake, M. Dijcks, M. Amelaine, V. Wickaert, L. Bertrand and B. Vosman. 2002. Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theor. Appl. Genet. **105**: 1019-1026.

Brownstein, M.L., J.D. Carpten, and J.R. Smith. 1996. Modulation of non-templated addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. BioTechniques **20**: 1004-1010.

Cheesman, E.E. 1944. Notes on the nomenclature, classification and possible relationships of cacao populations. Tropical Agriculture (Trinidad) **21:** 144-159.

Clapperton, J., S. Yow, J. Chan, D.J. Lim, R.Lockwood, L. Romanczyk and J. Hammerstone. 1994. The contribution of genotype to cocoa (*Theobroma cacao* L.) flavour. Tropical Agriculture (Trinidad) **71**: 303-308.

Cope, R.W. 1962. The mechanism of pollen incompatibility in *Theobroma cacao* L. Heredity **17**: 157-182.

Esselink, D., M.J.M. Smulders and B. Vosman. 2003. Identification of cut-rose (*Rosa hybrida*) and rootstock varieties using robust Sequence Tagged Microsatellite markers. Theor. Appl. Genet. **106**: 277-286.

Enríquez, G.A. 1993. Characteristics of cacao "nacional" of Ecuador. In: *Proceedings of the International Workshop on the Conservation, Characterisation and Utilisation of Cocoa Genetic Resources in the 21st Century*, *Port-of-Spain, Trinidad, September 13-17, 1992*. Cocoa Research Unit, Port of Spain, Trinidad, pp. 269-278.

Fulton, E., J. Chunwongse, and S.D. Tanksley. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol. Biol. Rep. **13**: 207–209.

Gryson, N., K. Messens and K. Dewettinck. 2004. Evaluation and optimisation of five different extraction methods for soy DNA in chocolate and biscuits. Extraction of DNA as a first step in GMO analysis. J. Sci. Food Agric. **84**: 1357-1363.

Lanaud, C., J.C. Motamayor and A.M. Risterucci. 2001. Implications of new insight into the genetic structure of *Theobroma cacao* L. for breeding strategies. In: Proceedings of the International Workshop on New Technologies and Cocoa Breeding. INGENIC, London, pp 89–107. http://www.koko.gov.my/CocoaBioTech/ING_Workshop(89-97).html

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**: 2141–2152.

Lanaud, C., O. Sounigo, Y.K. Amefia, D. Paulin, Ph. Lachenaud and D. Clément. 1987. Nouvelles données sur le fonctionnement du systeme d'incompatibilité du cacaoyer et ses conséquences pour la sélection. Café Cacao Thé (Paris) **31**(4): 267-277 and translation in English **31**(4): 278-282.

Liebhard, R., L. Gianfranceschi, B. Koller, C.D. Ryder, R. Tarchini, E. Van de Weg and C.C. Gessler. 2002. Development and characterisation of 140 new microsatellites in apple (*Malus x domestica* Borkh.). Mol. Breeding **10**: 217-241.

Motamayor, J.C., and C. Lanaud. 2002. Molecular analysis of the origin and domestication of *Theobroma cacao* L. In: Managing Plant Genetic Diversity (J. Engels, V. Ramanatha Rao, A.H.D. Brown, M. Jackson, Eds) CABI Publishing.

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

Motilal, L.A., and M. Boccara. 2004. Screening and Evaluation of SSR Primers in Gel Systems for the Detection of Off-types in Cocoa Field Genebanks. INGENIC Newsletter **9**: 21-26.

Pound, F.J. 1934. The progress of selection, 1933. In: Annual Report on Cacao Research 1933. Imperial College of Tropical Agriculture, pp 25-28.

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.

Risterucci, A.M., L. Grivet, J.A.K. N'Goran, I. Pieretti, M.H. Flament and C. Lanaud. 2000. A high density linkage map of *Theobroma cacao* L. Theor. Appl. Genet. **101**: 948–955.

Röder, M.S., K. Wendehake, V. Korzun, G. Bredemeijer, D. Laborie, L. Bertrand, P. Isaac, S. Rendell, J. Jackson, R.J. Cooke, B. Vosman and M.W. Ganal. 2002, Construction and analysis of a microsatellite-based database of European wheat varieties. Theor. Appl. Genet. **106**: 67-73.

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

Schnell, R.J., M.A. Heath, E.S. Johnson, J.S. Brown, C.T. Olano and J.C. Motamayor. 2004. Frequency of Off-type Progeny among the Original ICS 1 x SCA 6 Reciprocal Families and Parental Clones used for Disease Resistance Selection in Trinidad. INGENIC Newsletter **8**: 34-39.

Sereno, M.L., P.S.B. Albuquerque, R. Vencovsky and A. Figueira. 2006. Genetic diversity and natural population structure of Cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers. Cons. Genet. **7**: 13-24.

Shure, M., S. Wessler and N. Fedoroff. 1983. Molecular identification and isolation of the Waxy locus in maize. Cell **35**: 225–233.

Smulders, M.J.M., G. Bredemeijer, W. Rus-Kortekaas, P. Arens, B. Vosman. 1997. Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor. Appl. Genet. **94**: 264-272.

Smulders, M.J.M., J. van der Schoot, B. Ivens, V. Storme, V., S. Castiglione, F. Grassi, T. Fossati, J. Bovenschen, B.C. van Dam and B. Vosman. 2002. Clonal propagation in Black Poplar (*Populus nigra* L.). In: (B.C. van Dam, S. Bordács, Eds) Genetic diversity in river populations of European Black poplar - implications for riparian ecosystem management. In: Proceedings of an international symposium held in Szekszárd, Hungary, from 16-20 May, 2001. pp 39-52.

Sukha, D.A. 2008. *The influence of processing location, growing environment and pollen donor on the flavour and quality of selected cacao (Theobroma cacao L.) genotypes.* Ph.D Thesis, Dept. Chemical Engineering. The University of the West Indies, St. Augustine, Trinidad and Tobago. 248 pp.

Taberlet, P., J.J. Camarra, S. Griffin, E. Uhrès, O. Hanotte, L.P. Waits, C. Dubois-Paganon, T. Burke and J. Bouvet. 1997. Non-invasive genetic tracking of the endangered *Pyrenean* brown bear population. Mol. Ecol. **6**: 869-876.

Tengel, C., P. Schüßler, E. Setzke, J. Balles, and M. Sprenger-Haußels. 2001. PCR-based detection of genetically modified Soybean and Maize in raw and highly processed foodstuffs. BioTechniques **31**: 426-429.

Turnbull, C.J., D.R. Butler, N.C. Cryer, D. Zhang, C. Lanaud, A.J. Daymond, C.S. Ford, M.J. Wilkinson and P. Hadley. 2004 Tackling Mislabelling in Cacao Germplasm Collections. INGENIC Newsletter **9**: 8-11.

Wadsworth, R.M. and T. Harwood. 2000. International Cocoa Germplasm Database (ICGD) version 4.1. London, Reading: The London International Financial Futures and Options Exchange and the University of Reading.

Yeh, F.C., and T.J.B. Boyle. 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg. J. Bot. **129**: 157.

Table 1: Cocoa genotypes used in this study

Country	Genotype
Trinidad	CCL 200
	CCL 202
	CCL 217
	EET 400
	ICS 1
	ICS 95
	IMC 67[a]
	SCA 6[a]
Ecuador	CCAT 1119
	CCAT 1858
	CCAT 2143
	CCAT 2664
	CCAT 3345
	CCAT 4675
	CCAT 4688
	CCAT 4998
	CCN 51[a]
	CLONE 37 13/1
	EET 103
	EET 19
	EET 400
	EET 48
	EET 62
	EET 95
	EET 96
	ICS 95
	IMC 67[a]
Papua New Guinea	CLONE 16 2/3
	CLONE 73 14/1
	ICS 1
	K 82
	KA2 106
	KEE 43
	NAB 11
	SCA 6[a]
Venezuela	CRIOLLO
	MERIDEÑO
	GUASARE
	ICS 1
	ICS 6
	ICS 8
	IMC 67[a]
	PORCELANA
	ESTUFA

[a] "Bulk cocoa" genotypes

Table 2: Characteristics of the selected 15 cocoa microsatellite markers

Locus	Label	PCR cycli	Number of alleles		Heterozygosity	
			observed	effective	observed	expected
mTcCIR06	FAM	40	6	4.73	0.88	0.80
mTcCIR08	NED	40	5	3.01	0.73	0.68
mTcCIR11	HEX	40	6	3.81	0.88	0.75
mTcCIR12	FAM	45	7	4.33	0.85	0.78
mTcCIR15	NED	35	9	4.66	0.77	0.80
mTcCIR18	HEX	35	5	4.40	0.73	0.79
mTcCIR22	HEX	40	3	2.05	0.46	0.52
mTcCIR24	NED	40	4	2.24	0.54	0.56
mTcCIR26	FAM	35	7	3.43	0.77	0.72
mTcCIR33	NED	35	8	4.24	0.69	0.78
mTcCIR37	FAM	45	10	4.68	0.73	0.80
mTcCIR40	HEX	40	6	3.54	0.65	0.73
mTcCIR60	NED	35	7	3.14	0.58	0.69
SHRSTc16	FAM	40	5	3.00	0.58	0.68
SHRSTc12	HEX	40	4	1.77	0.35	0.44
		Mean	6.13	3.54	0.68	0.70
		St. Dev	1.92	0.99	0.16	0.11

Table 3: Analysis of number of possible pollen donors of cocoa beans.

Bean nr	Mother plant from			Additional possible pollen donors for Trinidad beans
	Papua New Guinea	Venezuela	Trinidad	
1	unknown	ICS 1, ICS 8	ICS 1, ICS 95	ICS 6 8, 40, 60, 4, 5, 13
2	unknown	Unknown	CCL 200	
3	unknown	unknown	ICS 1, ICS 95	ICS 6, 8, 40, 60, 4, 5, 13
4	unknown	unknown	ICS 1	ICS 6, 8, 40, 60
5	unknown	ICS 1, ICS 8	Unknown	
6	unknown	Unknown	IMC 67	
7	unknown	Unknown	IMC 67	
8	unknown	Unknown	ICS 1	ICS 6, 8, 40, 60
9	unknown	Unknown	unknown	
10	unknown	ICS 1, ICS 8	ICS 1, ICS 95	ICS 6, 8, 40, 60, 4, 5, 13
11	unknown	ICS 1	unknown	
12	unknown	Unknown	unknown	

Note: We genotyped 12 beans of ICS 1 plants in each of three different countries, using 8 microsatellite markers. Possible pollen donors have a matching genotype at these 8 loci among, arbitrarily, those cultivars in this study that were from the same country as the mother plant of the beans. In the absence of information on the cultivars present on the plantations, this merely indicates the differences in paternal contribution among beans and among plantations. For Trinidad, this information was available, and using the database genotypes of most of these varieties on 6-8 loci produced a list of additional possible pollen donors. The Trinidad samples with unknown pollen donors had alleles not present in any of the possible pollen donors listed.

Figure 1: Dendrogram of cocoa accessions, based on pairwise genetic distances. E=Ecuador; P=Papua New Guinea; T=Trinidad; V=Venezuela