The relic Criollo cacao in Belize – genetic diversity and relationship with Trinitario and other cacao clones held in the International Cocoa Genebank, Trinidad

Lambert A. Motilal[1]*, Dapeng Zhang[2], Pathmanathan Umaharan[3], Sue Mischke[2], Vishnarayan Mooleedhar[4†] and Lyndel W. Meinhardt[2]

[1]*Cocoa Research Unit, The University of the West Indies, St Augustine, Trinidad, Republic of Trinidad and Tobago, West Indies,* [2]*USDA/ARS, Beltsville Agricultural Research Center, PSI, SPCL 10300, Baltimore Avenue, Bldg. 001, Rm. 223, BARC-W, Beltsville, MD 20705, USA,* [3]*School of Plant Sciences, Faculty of Science and Agriculture, The University of the West Indies, St Augustine, Trinidad, Republic of Trinidad and Tobago, West Indies and* [4]*The University of Trinidad and Tobago, Valsayn, Trinidad, Republic of Trinidad and Tobago, West Indies*

Abstract

Cacao (*Theobroma cacao* L.) was domesticated in Mesoamerica and is native to the South American rainforest. Belizean Criollo is a group of relic landraces that are thought to be similar to those used by the Olmecs and Mayans during early domestication. Knowledge of genetic diversity is essential for efficient conservation and use of these relic landraces. Using 30 microsatellite markers, we characterized genetic diversity in 77 Belize Criollo accessions collected from the Maya Mountains in Belize, and assessed their relationship with 62 cacao accessions including 25 Trinitario accessions. Genetic diversity and heterozygosity were low in Belizean Criollo germplasm. Eleven distinctive genotypes were identified among the Belizean germplasm. Results of ordination and cluster analysis supported their putative ancestral contribution to the Trinitario cacao. However, results of Bayesian assignment and parentage analysis both suggested that the contribution of Criollo cacao to the Imperial College Selections Trinitario is small. Our preferred hypothesis for the genesis of Trinitario cacao is that a limited population of Criollo × Forastero hybrids emanated from the introduced Forastero population of Trinidad. The present study provides new insights into the origin of the Trinitario cacao, which will be useful in the *ex situ* and *in situ* conservation of cacao landraces from Mesoamerica.

Keywords: cacao; Criollo; genebank; genetic contribution; genetic uniformity; microsatellite; population structure; synonymous groups; Trinitario

Introduction

Cacao (*Theobroma cacao* L.) is a neotropical allogamous fruit tree crop. Its seeds are used in a range of products resulting in a worldwide commodity expected to exceed 4 million tonnes in 2010 (Eyre, 2007). Cacao was domesticated and utilized by Amerindian civilizations including the Mayas (300–900 AD) and the

---

* Corresponding author. E-mail: lamotilal@yahoo.com
† Formerly at Cocoa Research Unit, The University of the West Indies, St. Augustine, Trinidad, Republic of Trinidad and Tobago.

Olmecs (400–1200 BC; Henderson *et al.*, 2007). These early civilizations utilized Criollo cacao, which is distinct from the traditional Forastero and Trinitario groups (Cheesman, 1944). Other groups include Refractario accessions collected in Ecuador (1920s–1930s) for field resistance to witches' broom disease (Pound, 1938; Bartley, 2001) and Nacional and Amelonado (Motamayor *et al.*, 2003). Recently, Motamayor *et al.* (2008) have proposed that *T. cacao* had ten phylogenetic clusters, one of which was Criollo.

Although Criollo cacao is of historical and cultural significance, the group is underrepresented in most field collections. For instance, the International Cocoa Genebank, Trinidad (ICG, T) is one of the two universal collection depositories (IBPGR, 1981) but contained only five accessions (CRIOLLO 22, MXC 67, PENTAGONA 1, PENTAGONA 2 and STAHEL) with Criollo background, representing 0.25% of the approximately 2000 accessions within the ICG, T (see Iwaro *et al.*, (2003) and Bekele *et al.*, (2006) for groupings and Turnbull *et al.* (2004) for confirmation and historical notes).

An expedition was therefore undertaken in 1994 to locate wild or historic stands of Criollo cacao trees within the Columbia River Forest Reserve (Cayo district of northwestern Belize) and the Bladen River Nature Reserve (Toledo district of southwestern Belize) in collaboration with the Ministry of Agriculture of the Government of Belize and in conjunction with The Maya Mountain Archaeological Project of Cleveland State University (Mooleedhar *et al.*, 1995; Mooleedhar, 1998).

It is believed that Criollo cacao grown in Trinidad up until the 16th century was devastated in 1727 by an unknown cause. Thereafter, Forastero cacao from the South American mainland was brought to the island, and the resultant hybridization between the survivors of 1727 and the introduced material generated Trinidad Trinitario (Wood and Lass, 1985). This population was extensively surveyed for 100 most productive clonal trees (Imperial College Selections, ICS; Pound, 1934, 1935, 1936). These reference ICS accessions are maintained as clonal replicates over five blocks at the San Juan Estate, Trinidad. Further information on ICS clones is available (Johnson *et al.*, 2009).

Elucidation of genetic variability among and within cacao accessions is reliant on clone identification (Saunders *et al.*, 2004; Cryer *et al.*, 2006; Zhang *et al.*, 2006) and minimizing mislabelling (Takrama *et al.*, 2005; Motamayor *et al.*, 2008; Zhang *et al.*, 2008, 2009). Mislabelling is a known issue in cacao germplasm collections (Motilal and Butler, 2003). The ICS material was assessed with microsatellites for parentage (Motamayor *et al.*, 2003) and diversity (Johnson *et al.*, 2004). Model-based clustering showed that ICS formed a homogenous

group with other Trinidad Trinitario material and clustered with Criollo accessions suggesting a significant role for Criollo in the origin of the Trinitario group (Johnson *et al.*, 2009). However, there were only two Criollo accessions in the latter study, which may not have permitted the analysis of the inter-population relationship between Criollo and ICS Trinitario. The relic Criollo collection from Belize therefore presented an opportunity to (1) understand the genetic diversity within this relic Criollo population, (2) examine the genetic relationship of the above population with other Criollo accessions within the ICG, T and (3) reassess the genetic contribution of the Criollo to the parentage of the ICS accessions.

Materials and methods

Plant material

Seventy-seven grafted Belizean Criollo plants held at the University of the West Indies, Trinidad and 20 ICS accessions from San Juan Estate, Trinidad were sampled. There were 42 reference samples that included: (1) five accessions reportedly with Criollo make-up from the ICG, T (six trees) and eight Criollo references from Costa Rica, Honduras and Venezuela; (2) six Upper Amazon Forastero accessions; (3) nine Lower Amazon Forastero accessions; (4) seven Refractarios; (5) four French Guiana accessions; and (6) three accessions of uncertain grouping. Groupings were based on Iwaro *et al.* (2003), Bekele *et al.* (2006) and Turnbull *et al.* (2004). Accession details are presented in Supplementary Table S1, available online only at http://journals.cambridge.org. Tree locations in the ICG, T are available upon request.

DNA extraction, amplification and fragment analysis

Cacao DNA was extracted from leaf tissue, either with a modified Kobayashi protocol (Kobayashi *et al.*, 1998) described earlier (Motilal *et al.*, 2008), or the DNeasy Plant System (Qiagen Inc., Valencia, CA, USA) according to Saunders *et al.* (2004).

Cacao microsatellite markers have been previously described (Lanaud *et al.*, 1999; Risterucci *et al.*, 2000; Saunders *et al.*, 2004). Primers were synthesized by Proligo (Boulder, CO, USA) or Sigma-Aldrich Corp. (St Louis, MO, USA). Forward primers were 5′-labelled using WellRED fluorescent dyes (Beckman Coulter, Inc., Fullerton, CA, USA). DNA was separately amplified (annealing at 51°C) and scored for 30 microsatellite primer pairs (MPPs; mTcCIR1, 3, 6, 7, 8, 9, 10, 11, 12, 15, 18, 22, 24, 26, 29, 33, 37, 40, 42, 43, 57, 58, 60, 184, 210, 229, 243, 244, 274 and SHRSTc016) similar to Motilal *et al.* (2008).

The reaction mix contained, in final volume, 2.5 mM magnesium, 0.35 μM MPP, 0.2 mM total dNTP, 0.25 U DNA polymerase and 0.3–0.7 ng DNA. Amplified sequences were separated on a CEQ™ 8000 or 8800 (Beckmann Coulter Inc.) sequencer.

Dataset preparation for population analysis

The allelic dataset was assessed for duplicate accessions with GenAlEx v6.1 (Peakall and Smouse, 2006). Missing data were treated as null alleles and were not interpolated. Accessions that were fully matched based on pairwise comparisons across the 30 loci were deemed synonymous. In such instances, 'type' accessions were selected to represent the synonymous groups before analysing for genetic diversity and population structure. The 140 accessions were sorted into 74 groups in GIMLET v1.3.3 (Valière, 2002), primarily as a result of synonymous Belizean and Criollo material. Eleven samples (61B_2, 61B_6, AC 2 T1, AGUACARTE 3, BANANA CREEK 2, C 61, CC 7, CC 8, HF 3_1, HF 8 and ST 4_1) typified the Belizean material. The probability of identity values (Waits *et al.*, 2001) per microsatellite locus is presented in Supplementary Table S2, available online only at http://journals.cambridge.org.

Microsatellite diversity and population analysis

Primary summary statistics were computed for each marker locus with GenAlEx v6.1 (Peakall and Smouse, 2006). Pairwise estimates of Nei's genetic distance (Nei *et al.*, 1983) among individual accessions ($n = 74$) with 100 bootstrap resamplings across loci were calculated with Microsatellite Analyser (Dieringer and Schlötterer, 2002). Trees were generated from the resulting distance matrix with the neighbour-joining algorithm (Saitou and Nei, 1987) available in PHYLIP (Felsenstein, 1989). A consensus tree (50% majority rule) based on 100 bootstrapped trees was created using the CONSENSE procedure implemented in PHYLIP (Felsenstein, 1989). The dendrogram was visualized with TreeView version 1.6.6 (Page, 1996). Principal coordinate analysis (PCoA), as implemented in GenAlEx (Peakall and Smouse, 2006), was performed and the results were presented as a two-axis PCoA plot.

Population structure and inference of admixed ancestry (hybrids or ancestral forms) were assessed with a model-based clustering method, implemented in STRUCTURE (Pritchard *et al.*, 2000). The *K* value was set from 2 to 10, and the analysis was carried out (200,000 iterations after a burn-in period of 100,000) without assuming any prior information about the genetic group or geographic origin of the samples. Fifty independent runs were assessed for each *K* value. Results of the five iterations with highest score in each replicated run were matched by permutation in CLUMPP to generate the optimum alignment over multiple runs (Jakobsson and Rosenberg, 2007).

To test whether the Criollo and Lower Amazon Forastero accessions are the hypothesized parents of the ICS Trinitario, we used a likelihood-based method implemented in the program CERVUS 3.0 (Marshall *et al.*, 1998; Kalinowski *et al.*, 2007). All the ICS Trinitario clones plus MXC 67 and OC 61 (22 genotypes) were treated as putative offspring, whereas all the non-Trinitario clones plus PENTAGONA and STAHEL (52 genotypes) were treated as putative mothers. For each mother–offspring pair, the natural logarithm of the likelihood ratio (LOD score) was calculated. This score is the likelihood of maternity of a particular candidate parent, relative to an arbitrary individual. Critical LOD scores were determined for the assignment of maternity to all the ICS clones. The most probable single mother for each produced offspring was identified on the basis of the LOD scores for assignment at greater than 95% confidence.

Results

Population genetic analysis

Summary population statistics from a reduced subset of 74 accessions are presented in Supplementary Table S3, available online only at http://journals.cambridge.org. Gene diversity ranged from 0.447 (SHRSTc016) to 0.824 (mTcCIR58).

Microsatellite loci that detected differences in the Belizean material were mTcCIR10, 15, 26, 33, 37, 42 and 60, whereas mTcCIR6, 8, 11, 33, 37 and 244 were useful in discriminating among the reference Criollo material. The genetic relationships among the 74 accessions are represented in a distance-based dendrogram (Fig. 1). Accessions clustered together according to traditional genetic groups of Criollo, French Guiana, Lower Amazon Forastero, Upper Amazon Forastero, Refractario and Trinitario. The PCoA (Fig. 2) accounted for 67.5% of the variation with 51.7% on the first axis and 15.8% on the second axis. Criollo, Trinitario, Lower Amazon Forastero and Upper Amazon Forastero were clearly separated. The Refractario accessions fell in between Upper Amazon Forastero and Trinitario.

The results of Bayesian clustering analysis were compatible with the distance-based multivariate analysis. At $K = 6$–9, the assigned population only changed in the Upper Amazon Forasteros, whereas the rest of the populations remained largely the same (Fig. 3). At all different
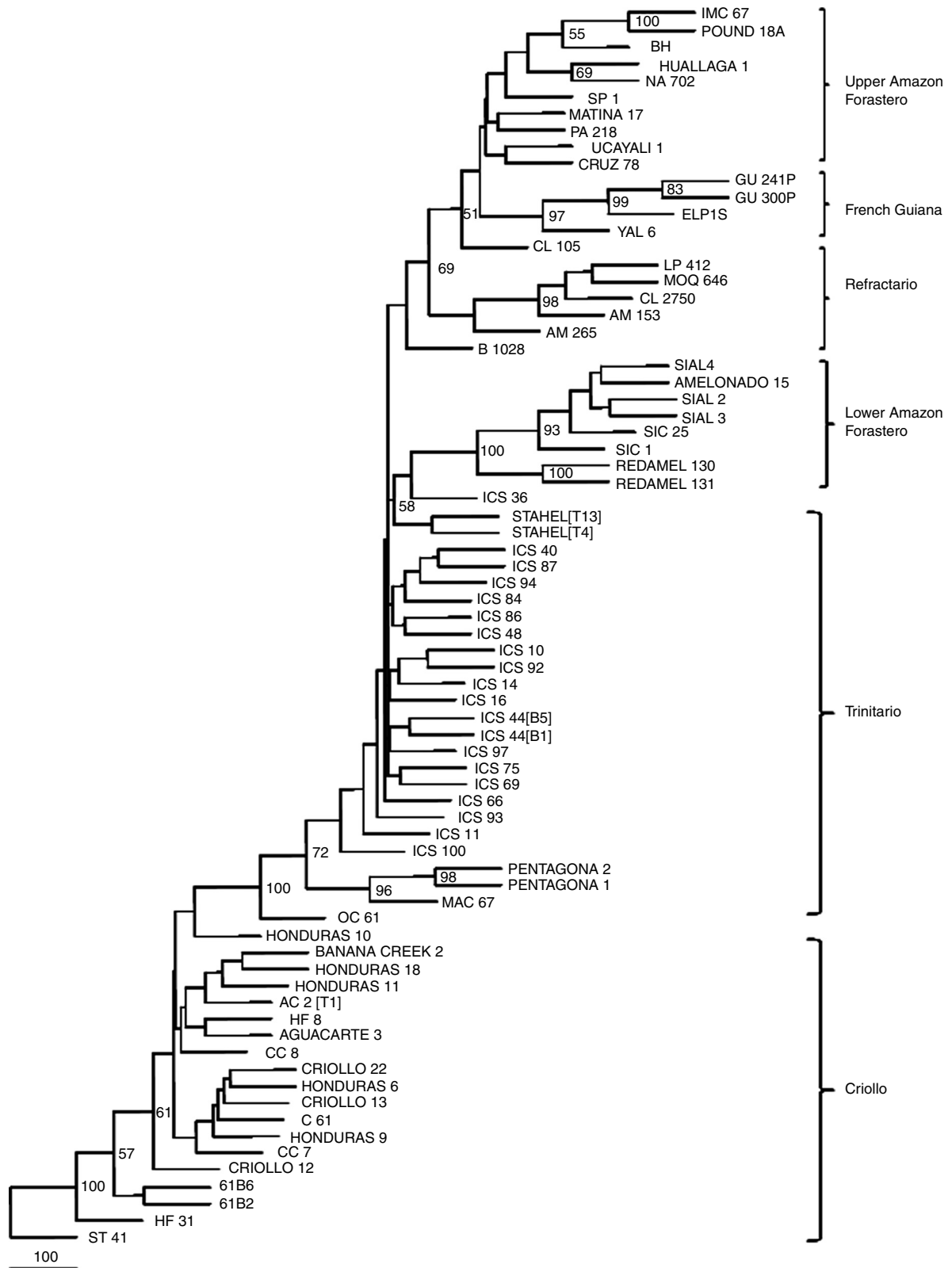
Fig. 1. Consensus dendrogram depicting the relationship between relic Belizean Criollo material and other reference accessions. Nei *et al.* (1983) genetic distance with 100 bootstrapping and a 50% consensus majority rule was used.
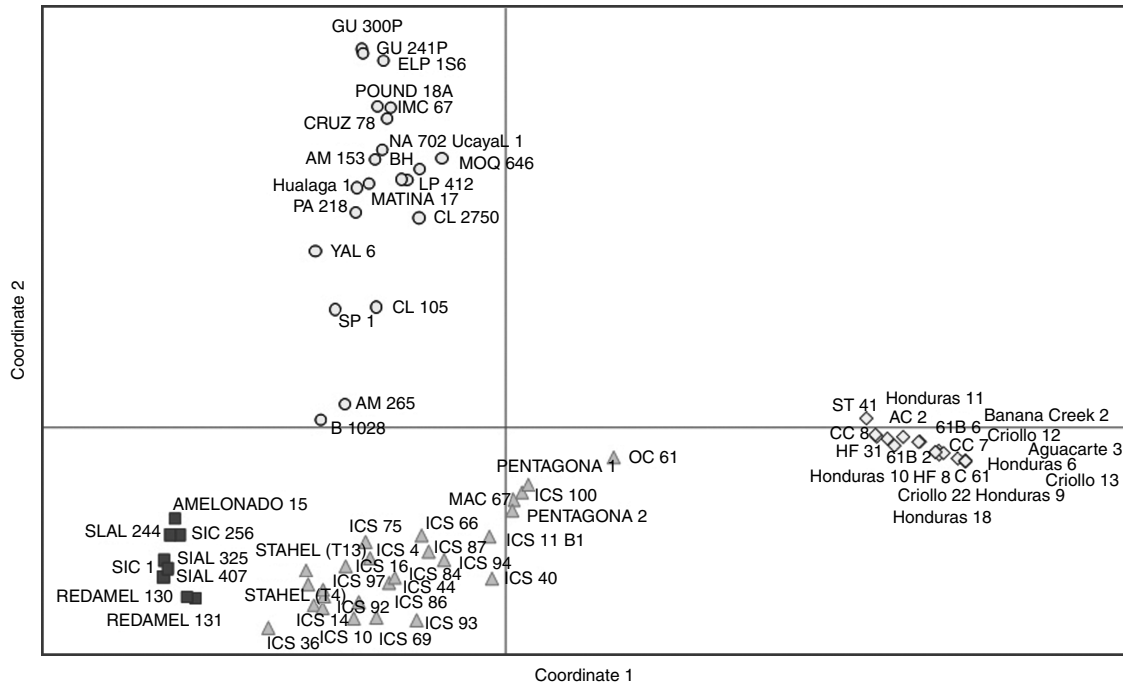
Fig. 2. Multivariate analysis of genetic distances among 74 cacao accessions. Groupings are: ■, Lower Amazon Forastero; △, Trinitario; ◇, Criollo; ○, others (Upper Amazon Forastero, Refractario and French Guiana).

K values, only a small fraction of the 20 ICS Trinitario genotypes showed significant levels of admixture. Four accessions originally at San Juan Estate (ICS 11 Block 1, ICS 40 Block 2; ICS 87 Block 2; ICS 100 Block 2) had at least 10% Criollo ancestry. The other 16 accessions had minimal Criollo ancestry (Fig. 3).

Resolved genetic grouping

Several accessions based on the clustering, ordination and STRUCTURE output had their genetic grouping clarified. OC 61 originating from Venezuela appeared to be a mixture of Criollo, Lower Amazon Forastero and Trinitario, and grouped with the latter. CRUZ 7/8, initially designated as Lower Amazon Forastero, was better placed on its own or constrained into the Upper Amazon Forastero group. MATINA 1/7 and SP 1, both previously of undetermined genetic origin, fell into the Upper Amazon Forastero group. The accession BH, originating from Ecuador but of undetermined genetic group, was placed into the Upper Amazon Forastero group. The two trees of the STAHEL accession previously designated as Criollo types fell within the Trinitario cluster. MXC 67, PENTAGONA 1 and PENTAGONA 2 were previously designated as selected Criollo material from cultivated areas in Mexico (MXC 67) or from the wild in Costa Rica (PENTAGONA). However, the data analysis indicated that these accessions were more Trinitario in nature, although they contained approximately 30% Criollo ancestry.

Parentage analysis

Of the 48 genotypes used as candidate parents, only 16 of them were identified at 95% confidence level as probable mothers of five Trinitario clones (ICS 44, ICS 75, ICS 97, MXC 67 and OC 61) by parentage analysis (Table 1). The identified probable mothers include RED AMELONADO 1/30 and 1/31 from the Lower Amazon Forastero group and STAHEL and PENTAGONA from the Trinitario group. None of the Criollo clones were identified as probable parents for the ICS Trinitario clones. In contrast, the accessions OC 61 could have been derived from any one of the 11 Belizean Criollo accessions (Table 1).

Discussion

Identity and germplasm management

The present study confirmed that the Belizean material sampled and proposed as relic Criollo (Mooleedhar *et al.*, 1995; Mooleedhar, 1998, 2000) was authentic Criollo material. A low genetic diversity and heterozygosity were found in the Criollo samples in accordance with other authors (Mooleedhar, 2000; Motamayor *et al.*, 2002).
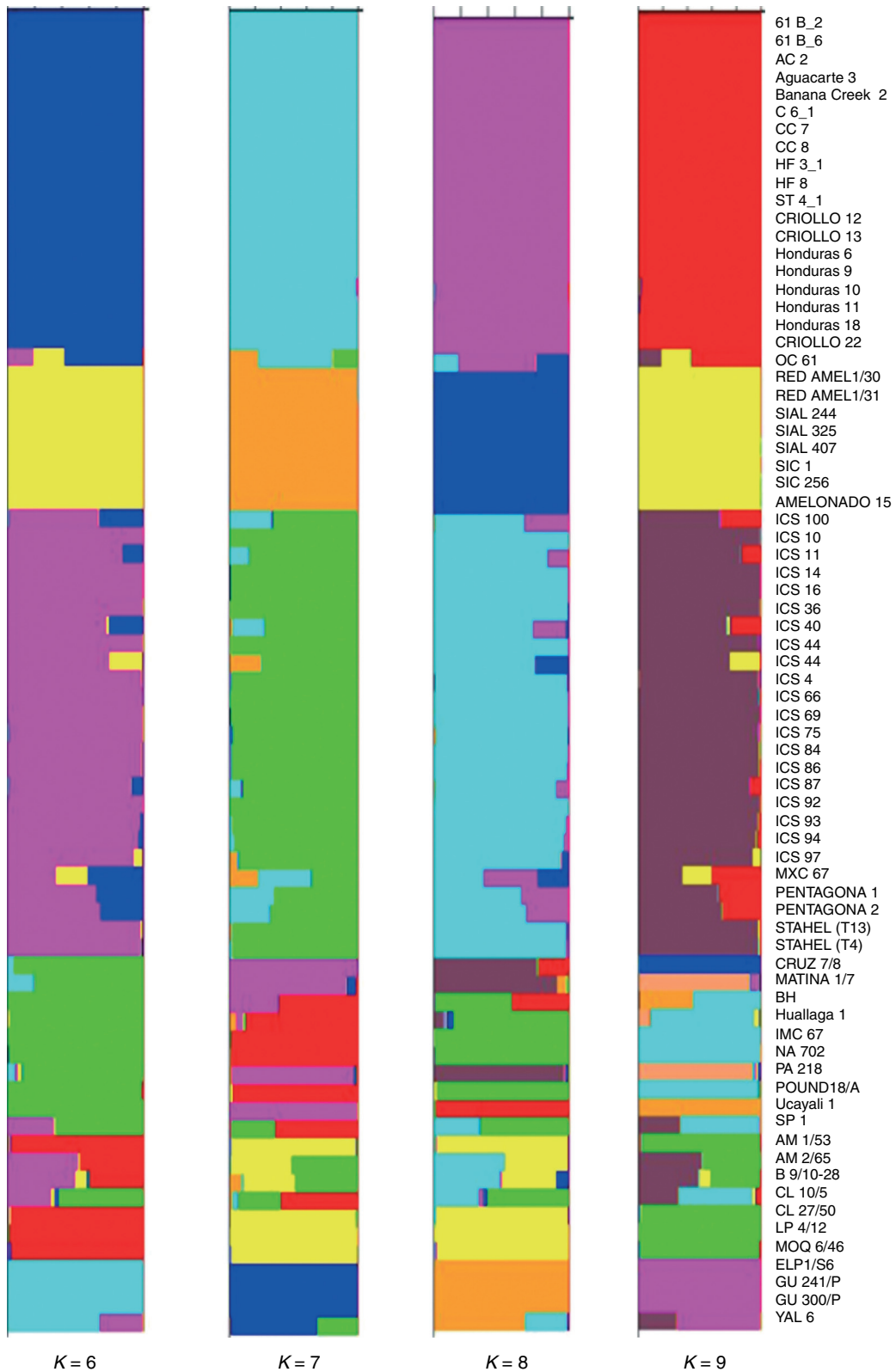
Fig. 3. Population stratification (for *K* groups 6–9) of 74 cacao accessions from a microsatellite dataset of 30 primers based on CLUMPP (Jakobsson and Rosenberg, 2007) alignment of STRUCTURE (Pritchard *et al.*, 2000) output.

Table 1. Likelihood assignment for five Trinitario accessions (18 mother–offspring pairs) based on 13 candidate mother genotypes

Offspring	Candidate mother[a]	LOD[b]
ICS 44	RED AMELONADO 1/30	7.40
ICS 44	RED AMELONADO 1/31	6.73
ICS 97	RED AMELONADO 1/30	5.05
ICS 97	RED AMELONADO 1/31	4.36
ICS 75	STAHEL (T4)	0.36
MXC 67	PENTAGONA 2	6.80
MXC 67	PENTAGONA 1	3.68
OC 61	CC 8	4.24
OC 61	Honduras 6	4.18
OC 61	CRIOLLO 13	4.18
OC 61	Honduras 18	4.18
OC 61	Honduras 9	4.18
OC 61	C 61	4.18
OC 61	HF 8	4.18
OC 61	Banana Creek 2	4.18
OC 61	CC 7	3.50
OC 61	AGUACARTE 3	3.50
OC 61	CRIOLLO 12	3.49

[a] Putative mother trees in the assessed Criollo and Lower Amazon Forastero clones. [b] Critical natural logarithm of the likelihood (LOD) ratios for assignment of maternity are 1.46 at >95% confidence and 0.19 at >80% confidence.

Furthermore, the ability to compress the genetic diversity of this Belize group through a substantial reduction of 85.7% and the tight cluster of Criollo individuals regardless of geographic origin suggested that the Criollo group can be adequately represented by a few individuals. Indeed, our results indicated that 11 accessions would be sufficient to represent the Belizean germplasm that was collected.

The present study found that of the five Criollo-allocated accessions in the ICG, T, three of these (MXC 67, PENTAGONA 1 and PENTAGONA 2) were introgressed material. These introgressed Criollos were better grouped with the Trinitarios confirming the results of Motamayor *et al.* (2002). The true Criollo at the ICG, T (CRIOLLO 22) has since died and the Belizean material therefore represents an important pool for conservation. It is therefore recommended that the 11 type accessions be prioritised for preservation and safeguarding.

The sampled STAHEL trees had very little Criollo ancestry and may either be cases of mislabelling or reflect the true nature of this accession. Prior published work listed STAHEL as a Forastero type (Mooleedhar *et al.*, 1991) and was described as collected from cultivated areas in Suriname but probably originated from Venezuela (Lockwood and Gyamfi, 1979; Turnbull *et al.*, 2004). However, our results suggest that this accession may be Trinitario in nature, which agrees with the hybridization thought to have occurred in Venezuela (Bartley, 2005).

The accession SP 1 (held in the ICG, T) grouped with the Upper Amazon Forastero group rather than the Criollo group. This suggests that the SP 1 used here may be mislabelled or another clone with the same name exists in Venezuela and was utilized by Motamayor *et al.* (2003).

Bartley (2005) indicated that the first varieties grown in Martinique were Amazonian and were exported to Costa Rica being renamed as MATINA. The seeds first introduced to Martinique may have come from Brazil (Bartley, 2005). However, the present study assigned the accession MATINA 1/7 as an Upper Amazon rather than Lower Amazon.

The accession OC 61 had moderately high Criollo ancestry in agreement with its putative history (Turnbull *et al.*, 2004). Furthermore, the population assignment (Fig. 3) indicated that OC 61 was derived from both Lower Amazon Amelonado Forastero and Trinitario, which agreed with the hypothesis that material from Trinidad, introduced into Ocumare de la Costa in Venezuela (Turnbull *et al.*, 2004), could have hybridized with the remnant Criollo material in this country (Bartley, 2005).

The correction of mislabelling in genebanks, proper passport data and consistent nomenclature applied globally would be invaluable to resolve these conflicts and allow for better comparative research. This study has clarified the uncertainties surrounding the historical documentation for these accessions.

Trinitario ancestry

The origin of the Trinidad Trinitarios, from which the ICS material was selected, is uncertain (Wood and Lass, 1985). It is believed that Amelonado seed, either from Guyana or from Orinoco estuary in Venezuela, may have been introduced into Trinidad by the cocoa planters, which then crossed naturally with the remaining Criollo following the devastation of cacao plantations in 1727, to form the Trinitario population (Cheesman, 1944; Wood and Lass, 1985). Bartley (2005) mentioned that Capuchin priests introduced seed in 1757 after arrival in Trinidad and suggested that this seed was from the Rio Amazon. Another theory is that hybrids of Venezuelan Criollos and Lower Amazon Forasteros that occurred in the Orinoco Estuary in Venezuela (Wood and Lass, 1985; Figueira and Cascardo, 2001) were subsequently transported to Trinidad.

The present study confirmed a Lower Amazon Amelonado ancestry for a limited number of accessions including OC 61, a few ICS (ICS 4, 40 and 97) accessions and MXC 67, indicating that Amelonado had a minor contribution to the Trinitario population. This was in contrast to Motamayor *et al.* (2003) but in agreement with Johnson *et al.* (2009). The present study

suggested that the Trinitario samples contained Upper Amazon ancestry, which agreed in part with Johnson *et al.* (2009). However, the Refractario accessions SP 1, AM 2/65, B 9/10-28 and CL 10/5 had Trinitario composition and may represent the early movement of Trinitario material to Ecuador from Trinidad (Pound, 1938) or Venezuela (Wood and Lass, 1985; Bartley, 2005) or indicate that a hybridization process occurred in Ecuador similar to that of Trinidad.

In this study, the overall Trinitario group of ICS accessions had little Criollo background. The present finding, of little Criollo ancestry in the ICS accessions, is not in agreement with the result of Johnson *et al.* (2009), which showed that Criollo and ICS Trinitario belong to same Bayesian cluster. This may be due to several factors: (1) the ICS group is heterogeneous and the subset used here captured more of the low Criollo ancestry accessions; (2) the composition of the dataset influences the STRUCTURE output; (3) if (2) is true, then the few (two) Criollo accessions used by Johnson *et al.* (2009) could adversely influence the final output; and (4) the above authors used a low K value, which may have constrained the accessions into a smaller number of groups.

The existing body of literature defines Trinitario as a hybrid between Criollo and Forastero. However, the present finding casts some doubts on this. Johnson *et al.* (2009) cited Ciferri and Ciferri (1957):

the distinction between what was called Forastero cacao of Trinidad and Criollo cacao of Venezuela was really based on the commercial bean grading system in Venezuela and that the two populations actually shared a common genealogy.

Furthermore, Criollo may be translated as 'native' or 'first grown' in Spanish-speaking countries of South America (Bartley, 2005). It is therefore highly probable that the cacao being cultivated in Trinidad prior to 1727 was an introgressed Criollo material rather than what is now known as pure Criollo. We concur with Bartley (2005) that the term Criollo was a convenient naming system for a miscellaneous genetic group of plants that may have been introduced multiple times into Trinidad. It is stated that the cacao industry was decimated by the blast of 1727; therefore, we can expect that the surviving plants would have been of hardier stock or lucky escapees. However, it is well known that Forastero hybrids are hardier and more disease resistant, which would then suggest that the remaining individuals were introgressions of predominantly Forastero ancestry. The introduction of further Forastero material and the ensuing hybridizations over the 200 years until the selection by Pound in 1932–35 would mean that there would have been an increased probability of obtaining more Forastero × Forastero hybrids than that of Criollo × Forastero introgressions. Since Pound collected over the entire island of Trinidad, the chance that he obtained a diverse group is increased. Additionally, it seems more probable that individuals from parents with high Forastero ancestry would be selected over other individuals, due to their higher vigour or heavier yields. In this regard, the 'limited Criollo × Forastero hybrids within a larger introduced Forastero population' hypothesis proposed by Mooleedhar (2000) for the genesis of the Trinitario accessions is supported. A similar hypothesis accounted for 40% of the modern Nacional genotypes in Ecuador (Loor *et al.*, 2009).

The relic Belizean Criollo material was employed in this study to review the origin of the Trinitario ICS accessions that probably arose sub-spontaneously in Trinidad. The study showed that the ICS material possessed a more varied parentage than had been previously envisaged. It appeared that both Upper and Lower Amazon material, in addition to Criollo, contributed to the ICS group. It is proposed that repeated introgressions of Amazonian material and early farmer selections of vigorous material may have led to a limited Criollo ancestry in the ICS accessions.

Acknowledgements

References

Bartley BGD (2001) Refractario – an explanation of the meaning of the term and its relationship to the introductions from Ecuador in 1937. *INGENIC Newsletter* 6: 10–15.

Bartley BGD (2005) *The Genetic Diversity of Cacao and its Utilization*. Wallingford: CABI Publishing, 400 p.

Bekele FL, Bekele I, Butler DR and Bidaisee GG (2006) Patterns of morphological variation in a sample of cacao (*Theobroma cacao* L.) germplasm from the International

Cocoa Genebank, Trinidad. *Genetic Resources and Crop Evolution* 53: 933–948.

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

Ciferri R and Ciferri F (1957) The evolution of cultivated cacao. *Evolution International Journal of Organic Evolution* 9: 381–397.

Cryer NC, Fenn MGE, Turnbull CJ and Wilkinson MJ (2006) Allelic size standards and reference genotypes to unify international cacao (*Theobroma cacao* L.) microsatellite data. *Genetic Resources and Crop Evolution* 53: 1643–1652. doi:10.1007/s10722-005-1286-9.

Dieringer D and Schlötterer C (2002) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* 3: 167–169.

Eyre C (2007) Cocoa demand on the up. Online documentation. Retrieved 2 August 2007. http://confectionerynews.com

Felsenstein J (1989) PHYLIP – phylogeny inference package (Version 3.2). Cladistics 5: 164–166. Software available at http://evolution.genetics.washington.edu/phylip/getme.html

Figueira AVO and Cascardo JCM (2001) Molecular markers in breeding. In: Dias LAS (ed.) *Genetic Improvement of Cacao*. Translated by Abreu-Reichart CE, Viçosa MG, Dias LAS and Griffee P. Viçosa: FUNAPE, Brazil. EcoPort version, http://ecoport.org/ep?Search Type=earticleView &earticleId=197&page=-1: FAO, pp. 120–135.

Henderson JS, Joyce RA, Hall GR, Hurst WJ and McGovern PE (2007) Chemical and archaeological evidence for the earliest cacao beverages. *Proceedings of the National Academy of Sciences USA* 104: 18937–18940.

IBPGR (1981) *Genetic Resources of Cocoa (AGP:IBPGR/80/56)*. Rome: IBPGR Working Group on Genetic Resources of Cocoa, IBPGR Secretariat, 25 p.

Iwaro AD, Bekele FL and Butler DR (2003) Evaluation and utilisation of cacao (*Theobroma cacao* L.) germplasm at the International Cocoa Genebank, Trinidad. *Euphytica* 130: 207–221.

Jakobsson M and Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801–1806. doi:10.1093/bioinformatics/btm233.

Johnson E, Bekele F and Schnell R (2004) *Field Guide to The ICS Clones of Trinidad*. Serie Técnica, Manual Técnico No. 54, 32 p. Turrialba: CATIE.

Johnson ES, Bekele FB, Brown SJ, Song Q, Zhang D, Meinhardt LW and Schnell RJ (2009) Population structure and genetic diversity of the Trinitario cacao (*Theobroma cacao* L.) from Trinidad and Tobago. *Crop Science* 49: 564–572. doi:10.2135/cropsci2008.03.0128.

Kalinowski ST, Taper ML and Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099–1106.

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 and Biotechnology* 4: 76–80.

Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A and Lagoda PJL (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. *Molecular Ecology* 8: 2141–2152.

Lockwood G and Gyamfi MMO (1979) The CRIG cocoa germplasm collection with notes on codes used in the breeding programme at Tafo and elsewhere. Technical Bulletin No. 10. Tafo: Cocoa Research Institute.

Loor RG, Risterucci AM, Courtois B, Fouet O, Jeanneau M, Rosenquist E, Amores F, Vasco A, Medina M and Lanaud C (2009) Tracing the native ancestors of the modern *Theobroma cacao* L. population in Ecuador. *Tree Genetics and Genomes* 5: 421–433.

Marshall TC, Slate J, Kruuk LEB and Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639–655.

Mooleedhar V (1998) A study of the morphological variation in a relic Criollo cacao population from Belize. *Report for 1997*. St. Augustine: Cocoa Research Unit, The University of the West Indies, pp. 5–14.

Mooleedhar V (2000) Morphological characterisation and genetic evaluation of a relic Criollo cacao (*Theobroma cacao* L.) population in Belize. PhD thesis, The University of the West Indies, 272 p.

Mooleedhar V, Bekele F, Cassie T, Warren J and Spence J (1991) Accessions held in the International Cocoa Genebank, Trinidad (ICG, T). *Report for 1990*. St Augustine: Cocoa Research Unit, The University of the West Indies, pp. 31–33.

Mooleedhar V, Maharaj W and O'Brien HO (1995) The collection of Criollo cocoa germplasm in Belize. *Cocoa Growers' Bulletin* 49: 26–40.

Motamayor JC, Lachneaud P, da Silva e Mota JW, Loor R, Kuhn DN, Brown JS and Schnell RJ (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PLoS ONE* 3: e3311. doi:10.1371/journal.pone.0003311.

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

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

Motilal L and Butler D (2003) Verification of identities in global cacao germplasm collections. *Genetic Resources and Crop Evolution* 50: 799–807.

Motilal LA, Zhang D, Umaharan P, Mischke S, Boccara M and Pinney S (2008) Increasing accuracy and throughput in large-scale microsatellite fingerprinting of cacao field germplasm collections. *Tropical Plant Biology* 2: 23–37. doi:10.1007/s12042-008-9016-z.

Nei M, Tajima F and Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data II. Gene frequency data. *Journal of Molecular Evolution* 19: 153–170.

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12: 357–358. TREEVIEW v1.6.6 available at http://taxonomy.zoology.gla.ac.uk/rod/treeview.html.

Peakall R and Smouse PE (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

Pound FJ (1934) The progress of selection. In: *Third Annual Report on Cacao Research 1933*. Port-of-Spain: Trinidad Government Printing Office, pp. 25–28.

Pound FJ (1935) The progress of selection. In: *Fourth Annual Report on Cacao Research 1934*. Port-of-Spain: Trinidad Government Printing Office, pp. 7–11.

Pound FJ (1936) The completion of selection. In: *Fifth Annual Report on Cacao Research 1935*. Port-of-Spain: Trinidad Government Printing Office, pp. 7–16.

Pound FJ (1938) *Cacao and Witches' Broom Disease (*Marasmius perniciosus*) of South America with Notes on Other Species of* Theobroma. Port-of-Spain: Yuille's Printery.

Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure from multilocus genotype data. *Genetics* 155: 945–959.

Risterucci AM, Grivet L, N'Goran JAK, Pieretti I, Flament MH and Lanaud C (2000) A high-density linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics* 101: 948–955.

Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406–425.

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

Takrama JF, Cervantes-Martinez C, Philips-Mora W, Brown JS, Motamayor JC and Schnell RJ (2005) Determination of off-types in a cacao breeding program using microsatellites. *INGENIC Newsletter* 10: 2–7.

Turnbull CJ, Cryer NC, Ford CS, Daymond AJ, Wadsworth RM and Hadley P (2004) ICGD. Available at www.icgd.rdg.ac.uk. Euronext.liffe/University of Reading.

Valière N (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes* 2: 377–379. doi:10.1046/j.1471-8286.2002.00228.x.

Waits LP, Luikart G and Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology* 10: 249–256.

Wood GAR and Lass RA (1985) *Cocoa.* 4th edn., Essex: Longman Group Ltd, 620 p.

Zhang D, Boccara M, Motilal L, Butler DR, Umaharan P, Mischke S and Meinhardt L (2008) Microsatellite variation and population structure in the 'Refractario' cacao of Ecuador. *Conservation Genetics* 9: 327–337. doi:10.1007/s10592-007-9345-8.

Zhang D, Boccara M, Motilal L, Mischke S, Johnson ES, Butler DR, Bailey B and Meinhardt L (2009) Molecular characterization of an earliest cacao (*Theobroma cacao* L.) collection from Upper Amazon using microsatellite DNA markers. *Tree Genetics and Genomes* 5: 595–607. doi:10.1007/s11295-009-0212-2.

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