Upper Amazon Forastero cacao (*Theobroma cacao* L.) 1: An assessment of phenotypic relationships in the International Cocoa Genebank, Trinidad

F.L. Bekele*, D.R. Butler and G.G. Bidaisee

Cocoa Research Unit, The University of The West Indies, St. Augustine, Trinidad and Tobago;
**(e-mail:Frances.Bekele@sta.uwi.edu)*

Upper Amazon Forastero (UAF) cacao clones account for approximately 60% of cacao cultivation globally, and are widely used in breeding. This study was undertaken at the International Cocoa Genebank, Trinidad to examine the phenotypic relationships among several UAF accession groups, *viz.*, Iquitos Mixed Calabacillo (IMC), Morona (MO), Nanay (NA), Parinari (PA) and Scavina (SCA) and some of their known progenitors, POUND accessions using Cluster Analysis (CA) and Principal Component Analysis (PCA). Some of the IMC, NA and SCA clones clustered with their putative mother trees. The grouping of the IMC clones in a few almost homogeneous clusters suggested that this group was the least diverse phenotypically. More diversity was expressed within the POUND, MO, NA, PA, and SCA groups. The considerable phenotypic diversity in the SCA and MO groups, expressed for traits such as pod index, cotyledon weight and width and pod apex form, seems not to be explained by a single mother tree source for each group. Significant phenotypic diversity in this UAF germplasm sample, measured by the Shannon Weaver Diversity Index (values ≥ 0.5) and coefficients of variation (>10%), was observed for some of the traits studied including stamen filament pigmentation, pod apex form, pod index, cotyledon weight and style length. IMC 3 and 63 were among the most phenotypically distinct accessions based on the results of CA and PCA. These results, in conjunction with those from genetic diversity studies, should prove useful for future cacao breeding programmes and will facilitate the formulation of strategies to effectively manage cacao genetic resources - select core and working collections and plan future collections in the wild to increase the genetic diversity of cacao conserved *ex situ*.

Keywords: Cacao; Cocoa; Cluster analysis; Conservation; Genetic improvement; Phenotypic diversity; Principal Component analysis; *Theobroma cacao* L.; Utilisation

Cacao, *Theobroma cacao* L., Malvaceae *sensu lato* (Alverson *et al.*, 1999) is a Neotropical perennial crop, on which the thriving global chocolate industry is based. It is an important agricultural commodity in many developing countries in West Africa, South-East Asia, Latin America and the Caribbean (Becker, 1999). More than 15 million people in African, Caribbean and Pacific countries are now directly involved in cacao cultivation (Mossu, 1992), and there are 2 million cocoa farmers in over 50 countries (Sereno *et al.*, 2005). The annual consumption of cocoa beans is approximately 3 million tonnes (ICCO, 2006), and is valued at USD 5 billion.

T. cacao L. is a predominantly outbreeding, diploid (2n = 20), under-storey plant, which is indigenous to the Amazon and Orinoco rainforests of South America, and has its putative centre of genetic diversity at the headwaters of the Amazon River (Cheesman, 1944). The greatest known variation in morphological and physiological characters of *T. cacao* L. is found there.

Under domestication, three major classes of cacao are recognised: Forastero; Upper and Lower Amazon and Guianese (Wood & Lass, 1985; Lachenaud *et al.*, 2001), Criollo, and Trinitario (Cheesman, 1944). The classification is based on morphological traits, particularly those of the fruit (pod) and seed (bean).

0041-3216/2008/010001-015

Approximately 2,300 cacao accessions of diverse origin, many of which are primary germplasm from South America, are conserved in the International Cocoa Genebank, Trinidad (ICG,T), managed by the Cocoa Research Unit (CRU). The accessions in the ICG,T are assigned to almost 90 accession groups (Iwaro *et al.*, 2003; Bekele *et al.*, 2006). The geographical area over which the individual accession groups were collected varies from limited to vast (Lockwood & End, 1993).

One of CRU's missions is to generate information on the germplasm conserved in the ICG,T to facilitate its efficient management, utilisation and improvement will be facilitated (Bekele, 1999; Bekele and Bekele, 1996; Bekele *et al.*, 1994; 2003, 2006; Iwaro *et al.*, 2003; Sounigo *et al.*, 2005b). Accordingly, the objective of this study was to examine phenotypic relationships among several Upper Amazon Forastero (UAF) accession groups, *viz.*, Iquitos Mixed Calabacillo (IMC), Morona (MO), Nanay (NA), Parinari (PA) and Scavina (SCA) in relation to some of their known progenitors, POUND accessions (Pound, 1943a). UAF clones were selected because they account for close to 60% of cocoa cultivation globally, and have been widely used in breeding in Trinidad (Bekele, 1999; Abdul-Karimu *et al.*, 2003) and elsewhere (Eskes and Lanaud, 2001; Wood and Lass, 1985). In particular, IMC 67, SCA 6, POUND 7 [POU], and selections from the PA accession group have been used in numerous breeding programmes. UAF clones possess many desirable traits, which justify their continued prominence in germplasm enhancement programmes such as those described by Iwaro *et al.* (2005). CRU is committed to undertaking and facilitating research that is geared towards genetic improvement of cacao, such as that in the Common Fund for Commodities/International Cocoa Organisation/International Plant Genetic Resources Institute (now Bioversity International) (CFC/ICCO/IPGRI) Project entitled *Cocoa germplasm utilisation and conservation: a global approach* (April 1998-June 2004).

To maximise the value of cacao genetic resources conserved *ex situ*, comprehensive diversity studies are necessary. Significant progress has already been made by researchers studying cacao genetic diversity including Laurent *et al.* (1993); Lerceteau *et al.* (1997); N'Goran *et al.* 2000; Motamayor *et al* 2002, 2003; Ronning and Schnell (1994); Russell *et al.* (1993); Sounigo *et al.* (2005b) and Sereno *et al.* (2006) among others. Reports on phenotypic diversity of germplasm conserved in the *universal depositories* (Anon, 1981) in Costa Rica and Trinidad (ICG,T) have been published by Engels (1986), and Bekele and Bekele (1996) and Bekele *et al.* (1994; 2006), respectively.

Materials and Methods

Plant material

In 1937/38 and 1942/43, Dr. F.J. Pound, of the Department of Agriculture, Trinidad, conducted two expeditions to South America (Ecuador and Peru) in search of cacao resistant to Witches' Broom disease (WB) caused by the fungal pathogen, *Moniliopthora perniciosa* (Aime and Phillips, 2006), previously named *Crinipellis perniciosa* (Stahel).

The POUND clones

Dr. Pound collected the POUND clones as budwood from the Loreto Region, Peru in 1942 (Pound, 1943). Collecting from trees, with no obvious signs of Witches' Broom (WB) disease, was usually restricted to 2-3 hundred yards inland from the respective river (Table 1). Three of four plants, propagated at Iquitos, Peru were later shipped to Barbados for quarantine purposes. Originally, each collection was assigned an alpha-numeric code (*e.g.* P 1, later written in full as POUND 1 [POU]), but suffixes were subsequently appended in Trinidad to produce designations such as POUND 1/B [POU] since some variants of the replicated original tree were observed over time.

Morphological data were collated at CRU for POUND 1/B, 2/B, 4/A, 5/B, 5/C, 7/A, 7/B, 7/C, 8/C, 9/B, 10/B, 10/C, 12/A, 15/A, 16/B, 18, 18/A, 25/A, 26/C, 27/C, 31/A, 31/C, 32/A [POU].

The POUND mother trees of the other UAF groups under study were identified by Pound (1943a) as:

IMC – POUND 18 [POU] and POUND 21 [POU];

[1]NA – POUND 1*, 2*, 3*, 4*, 5, 6, 7, 8, 9, 10, 11, 12[‡], 15, 19, 20, 30 and 32* [POU];

SCAVINA – POUND 31 [POU];

PA and MO – unspecified: approximately 20 trees and 1 tree, respectively.

Table 1 Collecting sites of some of the POUND clones featured in this study

POUND Clone	Collection site
POUND I – 8, 15	Río Nanay, river bank
POUND 9 -14, 16, 17, 19, 20, 30	Río Nanay
POUND 18	Island south of Iquitos
POUND 21	Island river bank
POUND 31, 32	Río Ucayali, Contamana

IMC clones

The IMC clones were collected directly opposite the town of Iquitos in the Loreto Region, Peru, on a large island "in front of Iquitos", 73.12 W, 3.5S (Pound, 1938). In 1937, pods were collected from at least two groups of trees free of WB disease. They were probably progenies from POUND 18 and 21, two trees later collected and propagated as clones in 1942 (Pound, 1943a). One hundred and twelve IMC clones were planted at Marper Farm, Manzanilla, East Trinidad in 1939 (Pound, 1943b). Three main pod morpho-types for IMC clones were recognised by Pound; (1) a very large 'lagarta calabacillo', oval- shaped and smooth, pale green (*normal blanco*) type; (2) a large, somewhat warty at stem end, but smooth and round at the apex type and (3) a small, oval and warty type.

MO clones

These clones were collected in the Loreto Region, Peru, on the Río Morona, 77.17 W, 4.12 S (300 miles from Iquitos). The pods are thought to have been collected from one tree free of WB from among a group of 25 trees. Eighty MO clones were subsequently planted at Marper Farm (Pound 1943b). Accessions with large numbers (>100) were surmised to be MOQ clones by Bartley (personal communication in 1993, cited in the ICGD by Wadsworth and Harwood, 2000), but this does not agree with records at CRU dating from 1943. The pods are typically short, oval, and warty. Some have a red tint such as MO 4, 9, 14, 99 and 121.

NA clones

The collection site for the NA clones is given as the Loreto Region, Peru, along the Río Nanay at 73.17 W, 3.38 S, about 15km from Iquitos. The pods were probably collected from 14 (Pound, 1938) or 17 trees (Pound, 1943a, p. 9 and 10) free of WB and laden with fruit. Nine hundred and eight NA clones were originally established at Marper Farm between 1939 and 1941 (Pound 1943b). The fruits are typically completely unpigmented,

pale green, long oval, and slightly warty with no conspicuous point or bottleneck. Some NA trees currently appear to have several different characteristics from the original descriptions that are available. This is postulated to be a result of the generation of seedling progenies of the original collections (Bartley, personal communication in 1995,

[1] *Noted as being disease-free in 1937 (Pound, 1938).
[‡] The genotype collected on Río Nanay as POUND 12 grew not far from the area where the IMC trees were collected (Pound, 1943a).

cited in the ICGD by Wadsworth and Harwood, 2000).

PA clones

Pound (1938) reported the collection site of the PA clones to be the Loreto Region, Peru, lower Río Marañon at 74.60 W, 4.6S, Parinari. Pods were collected from 7-20 trees free of WB. Two hundred and seventy-seven PA clones were transferred to Marper Farm (Pound 1943b). Pound described the pods as being long, generally warty, with a pronounced bottleneck and conspicuous point. However, some pods were almost smooth with five, shallow furrows. The pods varied in appearance from *blanco* (pale) to pure green. Six morpho-types were distinguished by Pound (1938).

SCA clones

These clones were collected in the Loreto Region, Peru, along the Río Ucayali at 75.00 W, 7.21S. The pods were reputedly collected from one tree at Contamana, and 27 plants were generated. Twenty-two clones were eventually planted at Marper Farm (Pound 1943b). Pound (1938) described the pods as green or slightly "blanco", with a central type that was mid-green (such as SCA 6), only moderately warty, with a definite but blunt point and lacking a bottleneck. Pound also recorded a *lagarta* type, which was mostly warty with a less pronounced point and generally no bottleneck constriction. Variants were also observed; more warty or of the smooth, five-furrowed type. The latter were the only ones slightly susceptible to WB.

Phenotypic observations

Morphological or phenotypic data have been collated at the ICG, T over several years between 1990 and 2006 for the UAF clones included in this study.

Details of the methodology and collection site are provided by Bekele and Butler (2000); Bekele and Bidaisee (2006), Bekele *et al.* (2003, 2006) and Iwaro *et al.* (2003a). Observations were made using 25 descriptors (Table 2) from the International Board for Plant Genetic Resources (now Bioversity International) descriptor list for cacao (Anon., 1981) for 111 accessions - 20*[2] IMC, 20* NA, 17 MO, 20* PA, 11 SCA and 23 POUND.

Statistical analyses

The relationships within and between the aforementioned accessions groups were examined using the Cluster Analysis (CA) module of NTSYSpc Ver. 2.10b (NTSYS, 2000; Rohlf, 2001). The data were first standardised to eliminate the effects of different scales of measurement. Similarity matrices were generated prior to clustering. The group average method (UPGMA-unweighted pair-group method using arithmetic means) was used to perform Cluster Analysis as recommended for this type of dataset (mixed continuous and categorical) (Bekele and Bekele, 1996; Mardia *et al.*, 1979; Sneath and Sokal, 1973). Dendrograms were generated to depict the inter-relationships among the accessions studied.

In order to test whether the clusters obtained from the analysis were valid, a cophenetic value matrix of the tree matrix, produced after clustering using the Sequential Agglomerative Hierarchical Nested (SAHN) clustering procedure of NTSYSpc, was computed using COPH (cophenetic). The MXCOMP (matrix comparison) module was then used to compare the cophenetic value matrix with the original matrix that was clustered.

[2] * Randomly selected from the list of accessions from the ICG,T with phenotypic data available

Table 2 Descriptors used for morphological characterisation

Descriptor	State [sample size]
Flower, anthocyanin intensity in column of pedicel	1=green, 2=reddish, 3=red [n=10].
Flower, sepal length (mm) [n=10]	
Flower, anthocyanin intensity on ligule	0=absent, 3=slight, 5=intermediate, 7=intense [n=10]
Flower, ligule width (mm) [n=10]	
Flower, anthocyanin intensity in filament	0=absent, 3=slight, 5=intermediate, 7=intense [n=10]
Flower, style length (mm) [n=10]	
Flower, ovule number [n=10]	
Fruit, shape	1= oblong, 2= elliptic, 3=obovate, 4= orbicular [n=10], 5= other.
Fruit, basal constriction	0=absent, 1=slight, 2=intermediate, 3=strong, 4=wide shoulder [n=10]
Fruit, apex form	1=attenuate, 2=acute, 3=obtuse, 4=rounded, 5=mammillate, 6=indented [n=10]
Fruit, surface texture (rugosity or degree of wartiness)	0=absent, 3=slight, 5=intermediate, 7=intense [n=10]
Fruit, anthocyanin intensity in mature ridges	0=absent, 3=slight, 5=intermediate, 7=intense [n=10]
Fruit, ridge disposition	1=equidistant, 2=paired [n=10]
Fruit, primary ridge separation	1=slight, 2=intermediate, 3=wide [n=10]

Table 2 continued Descriptors used for morphological characterisation

Descriptor	State [sample size]
Fruit, pod wall hardness [n=10]	3= ≤ 1.6 MPa, 5 = > 1.6 MPa ≤ 2.0 MPa, 7= > 2.0 MPa
Fruit, length (cm) [n=10]	
Fruit, width (cm) [n=10]	
Seed, number [n=10]	
Seed, shape	1=oblong 2=elliptic 3=ovate
Seed, cotyledon colour	1=white, 2=grey, 3=light purple, 4=medium purple, 5=dark purple, 6=mottled [n=40]
Wet bean weight (total) (g) [n=10]	
Cotyledon length (cm) [n=20].	
Cotyledon width (cm) [n=20].	
Cotyledon weight (g) [n=20]	
Pod index (the number of pods required to produce 1 kg of dried cocoa) [n=10]	

Principal Components Analysis (PCA) was also performed using NTSYSpc, and the pattern of grouping of the accessions in the three-dimensional PCA plot was compared with that revealed by CA for the purpose of validation. A two-dimensional PCA plot was also generated, using MINITAB Ver 14.1, to examine the distribution of the clones in relation to their respective accession groups.

Descriptive statistics for each descriptor studied were generated using MINITAB Ver. 14.1 (Minitab Inc., 1997). Variation in the expression of quantitative traits was assessed from the coefficients of variation (COV). The

significance of the between accession group variation in pod index (PI), which is of economic interest as a component of yield, was tested using the One-Way Analysis of Variance (ANOVA) module of MINITAB (Minitab Inc., 1997). Normality of the distribution for PI was first tested using the Kolmogorov-Smirnov test (Minitab Inc., 1997).

Shannon Weaver Diversity Index (SWDI) values, $H = -\Sigma_k \ln p(k)$, where $p(k)$ = frequency of the class "k", (Shannon and Weaver, 1949) were calculated for the 13 qualitative (categorical) descriptors recorded for each of the UAF accession groups studied. This was done in order to assess the relative phenotypic diversity of the groups as well as the evenness in the distribution of frequency classes for each descriptor. High SWDI values (closer to 1) correspond to higher levels of diversity and evenness in the distribution of descriptor categories.

Results

Considerable phenotypic diversity was observed among the UAF clones studied (Tables 3 and 4), and this coincided with observations of other samples of germplasm from the ICG,T (Bekele *et al.*, 1994; Bekele and Bekele, 1996; Bekele *et al.*, 2006; Iwaro *et al.*, 2003; Sounigo *et al.*, 2005b). PI, a component of yield of particular interest to breeders, was normally distributed ($P > 0.05$), and ANOVA revealed that the between accession group variation in this trait was highly significant ($F = 6.01$, $P < 0.0001$). IMC was the most promising accession group in terms of PI. A similar result was also found in a study by Bekele *et al.* (2006), which included Trinitario and Refractario germplasm. In addition, all of the IMC clones studied were classified according to Pound's (1938) convention (three categories described in the introduction).

Table 3 Descriptive statistics of the 111 UAF clones studied using 25 morphological descriptors

Descriptor (units)	Mean ± SE	Minimum value	Maximum value	Variance
LIGULE COLOUR		0.00	7.00	
FILAMENT COLOUR		0.00	7.00	
PEDICEL COLOUR		1.00	3.00	
SEPAL LENGTH (mm)	7.1 ± 0.08	5.29	9.12	0.62
LIGULE WIDTH (mm)	2.4 ± 0.03	1.83	3.83	0.09
OVULE NUMBER	46.9 ± 0.6	33.0	62.0	41.0
STYLE LENGTH (mm)	2.2 ± 0.03	1.33	3.56	0.097
MATURE POD RIDGE COLOUR		0.00	5.00	
POD SHAPE		1.00	3.00	
POD BASAL CONSTRICTION		0.00	3.00	
POD APEX FORM		1.00	6.00	
POD SURFACE TEXTURE		0.00	7.00	
POD RIDGE DISPOSITION		1.00	2.00	
POD RIDGE (PAIR) SEPARATION		1.00	3.00	
HUSK HARDNESS		3.00	7.00	
BEAN COLOUR		2.00	5.00	
BEAN SHAPE		1.00	3.00	
POD LENGTH (cm)	15.8 ± 0.2	11.1	20.0	2.74
POD WIDTH (cm)	7.8 ± 0.1	6.0	11.1	0.67
WET BEAN WEIGHT (excluding mucilage (g))	50.2 ± 1.1	26.0	82.2	130.2
BEAN NUMBER	41.9 ± 0.7	26.6	59.3	47.4
COTYLEDON WEIGHT (g)	0.84 ±0.02	0.50	1.19 **(PA 71)***	0.02
COTYLEDON LENGTH(g)	2.07 ±0.02	1.64	2.68	0.04
COTYLEDON WIDTH (cm)	1.13 ±0.01	0.86	1.38	0.01
POD INDEX	30.3 ± 0.7	17.0 **(IMC 96)***	65.4	59.3

* superior accessions (low pod index, large cotyledon weight)

Table 4 Descriptive statistics for some of the quantitative traits accounting for most of the phenotypic variation expressed in the UAF groups studied

Accession Group	Mean Pod Index	Standard Error	Coefficient of variation (%)	Minimum value	Maximum value
IMC	23.6	0.8	15.2	17.0	29.5
NA	29.6	1.1	15.7	24.0	42.8
POUND	30.4	1.7	26.4	17.8	46.7
PA	31.3	1.4	20.5	19.7	48.0
MO	34.2	1.5	17.8	25.4	46.0
SCA	35.1	3.8	36.2	20.2	65.4

Accession Group	Cotyledon weight (g)	Standard Error	Coefficient of variation (%)	Minimum value	Maximum value
IMC	0.89	0.03	16.5	0.62	1.19
NA	0.86	0.03	16.3	0.57	1.11
POUND	0.85	0.02	13.7	0.55	1.05
PA	0.86	0.04	18.3	0.5	1.19
MO	0.75	0.03	18.3	0.53	1.03
SCA	0.78	0.06	26.6	0.51	1.11

Accession Group	Cotyledon width (cm)	Standard Error	Coefficient of variation (%)	Minimum value	Maximum value
IMC	1.15	0.02	6.5	0.99	1.3
NA	1.19	0.03	9.2	0.97	1.4
POUND	1.15	0.02	9.3	0.86	1.3
PA	1.09	0.02	6.4	0.92	1.2
MO	1.09	0.03	9.7	0.95	1.3
SCA	1.08	0.05	15.1	0.88	1.3

Accession Group	Pod length (cm)	Standard Error	Coefficient of variation (%)	Minimum value	Maximum value
IMC	16.7	0.24	6.4	14.8	19.9
NA	14.9	0.4	12.7	11.1	18.4
POUND	15.9	0.28	8.4	13.4	18.8
PA	16.7	0.38	10.1	14.0	20.0
MO	15.0	0.34	9.3	12.5	17.3
SCA	15.3	0.45	9.8	12.1	17.1

Accession Group	Style length (cm)	Standard Error	Coefficient of variation (%)	Minimum value	Maximum value
IMC	2.11	0.05	9.9	1.57	2.49
NA	2.20	0.09	17.9	1.33	3.02
POUND	2.14	0.05	10.8	1.74	2.70
PA	2.27	0.09	18.0	1.79	3.56
MO	2.34	0.05	8.9	1.70	2.63
SCA	2.19	0.10	15.3	1.52	2.70

Cluster Analysis of the UAF accession groups studied

The phenotypic relationships among the UAF groups analysed using Cluster Analysis (CA) are depicted in Figure 1. The cophenetic correlation, calculated as a measure of goodness of fit for the CA, provided evidence of a good fit (r = 0.805) (Rohlf 2001). All of the clones formed one cluster at Dissimilarity Coefficient = 2.1. Details based on observations of the relationships depicted in Figure 1 are as follows:

IMC clones tended to cluster together;

IMC 39 and its putative progenitor, POUND 18/A [POU], were very closely linked;

IMC 59 and 67 were fairly closely linked;

NA clones formed several small clusters together, interspersed with SCA, PA and MO clones;

NA 824 and MO 17 were phenotypically almost identical (Dissimilarity Coefficient = 0.5)[3];

Some PA clones are grouped in small homogenous clusters, *e.g.* PA 165, 90 and 200;

NA 406 and SCA 23 were very closely related;

NA 3 and PA 68 were very closely related;

PA 44 and MO 83 were closely linked. The latter was distinct from other MO clones;

PA 303 and IMC 41, and PA 171 and SCA 5 were closely linked;

MO 4 and 9, and MO 99 and 121 (all slightly pigmented unlike the majority of MO clones) were closely linked;

SCA 6 and SCA 11 were fairly closely linked; and were distinct from other the SCA clones studied;

[3] The smaller the value of the Dissimilarity Coefficient, the more similar two clusters are.

POUND 7/A [POU], POUND 7/C [POU] and POUND 7/B [POU] were closely linked;

POUND 10/B [POU] and POUND 10/C [POU] were fairly closely linked;

POUND 31/A and POUND 31/C [POU] were fairly closely linked;

POUND 18 [POU] was fairly closely linked to PA 71 and SCA 9;

POUND 4/A [POU] and SCA 20 were closely linked;

MO 3, PA 143, IMC 63 and IMC 3 were the most distinct clones; ungrouped until a Dissimilarity Coefficient > 1.9 was applied. Sounigo *et al.* (2001a) also found MO 3 to be distinct from the MO clones studied using RAPD markers.

PCA of the UAF accession groups studied

The first two principal components (PC) accounted for 31.4 % of the total variation expressed by the phenotypic traits in Figure 2 (PC 1 – 21.9 %, PC 2 – 9.5 %). The first four principal components accounted for 47.5% of the total variation expressed. The descriptors (Table 2) responsible for most of the variation expressed by the first four PC were as follows:

PC 1 – total bean weight (wet), pod index, cotyledon length and pod width;

PC 2 – pod length, cotyledon width, and pod apex form;

PC 3 – bean number, ovule number, cotyledon colour and pedicel colour;

PC 4 – sepal length and ligule width.

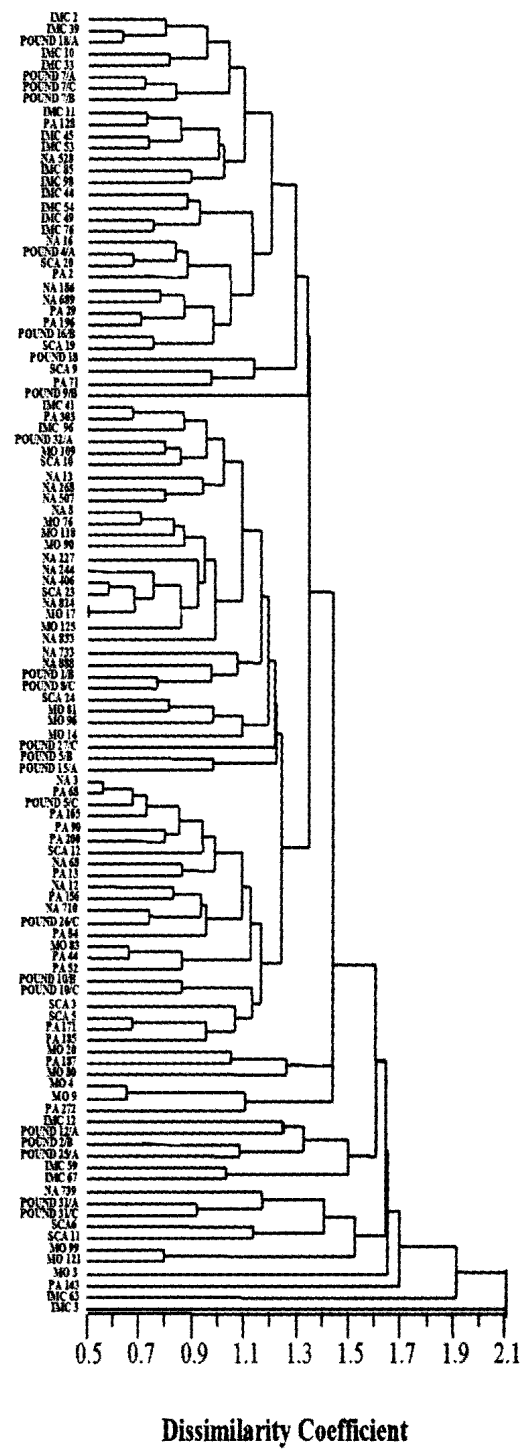

Figure 1 Dendrogram depicting phenotypic inter-relationships among the 111 UAF clones studied

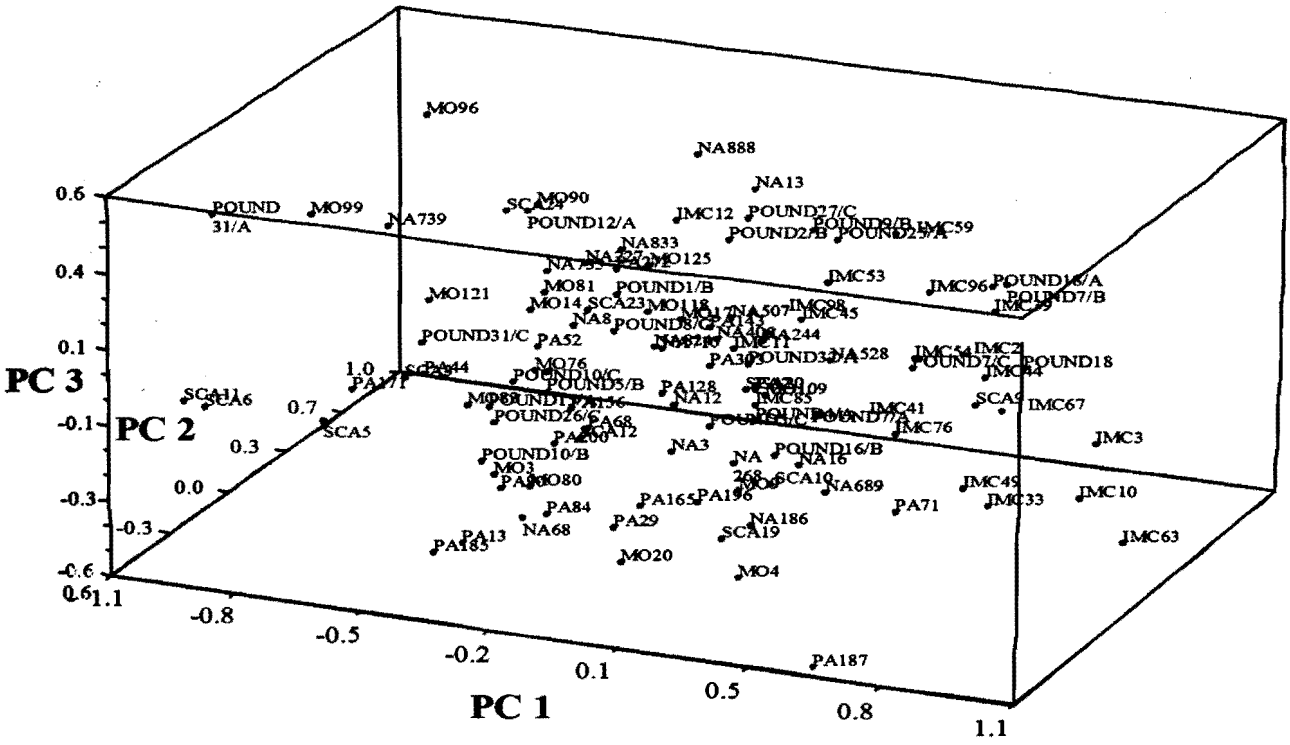

Figure 2 Principal Component plot of the 111 UAF accessions studied

Principal Component (PC) 1 accounted for 21.9% of the total variation expressed; PC1 and 2 accounted for 31.4 % of the total variation expressed; PC1, 2 and 3 accounted for 39.6% of the total variation expressed.

Comparison of the results of CA and PCA of the UAF clones studied

Congruent groupings for both analyses:
The following were grouped rather closely in Figures 1 and 2:

POUND 7/A [POU], POUND 7/B [POU] and POUND 7/C [POU];

POUND 10/B [POU] and POUND 10/C [POU];

SCA 6 and SCA 11;

POUND 4/A [POU] and SCA 20;

POUND 18 [POU], SCA 9 and PA 71;

IMC clones (IMC 49, 33, 10; and IMC 44, 67 and 59); and PA clones.

IMC 3 and IMC 63 were among the most disparate clones in Figures 1 and 2.

Observations exclusive to PCA (Figure 2)
MO 96, PA 187, SCA 6, SCA 11, and IMC 10 were among the most disparate clones in the PCA plot (IMC 3 and 63 were distinct in both Figures 1 and 2);

MO 90 and POUND 12/A [POU] grouped together closely as did:

POUND 9/B [POU] and POUND 27/C [POU];

NA 733 and MO 81;

SCA 23 and NA 8;

MO 3, 80 and POUND 10/B [POU];

MO 9 and SCA 10;

POUND 7/C [POU] and IMC 54;

POUND 16/B [POU] and NA 16;

POUND 4/A [POU] and POUND 5/C [POU];

IMC 41 and MO 109;

POUND 5/B [POU] and MO 76;

PA 84 and NA 68;

SCA 19 and NA 186;

POUND 18/A [POU] and POUND 7/B [POU];

POUND 7/A [POU] and POUND 32/A [POU];

MO 109 and IMC 85;

PA 68 and SCA 12;

POUND 31/A [POU] and MO 99;

MO 14, MO 81 and MO 121 (MO 14 and 121 have slightly pigmented pods); MO 118 and MO 17.

POUND 31/C [POU] and 31/A [POU] were in the vicinity of SCA 6, 11, 3 and 5 in Figure 2; (the other SCA clones were dispersed);

POUND 18 [POU] and POUND 18/A [POU] were in fairly close juxtaposition unlike in Figure 1;

POUND 5/B [POU] and POUND 5/C [POU] were in the same general vicinity in Figure 2.

Relationship between putative Mother trees and their 'progenies'

- POUND 18 [POU] appears to be phenotypically related to its putative progenies, IMC 67 and 59 (Figures 1 and 2). IMC 44, 2, 54, 67, 39, 96, 3, 76, 41, 49, 53, 45, 10, 33 and POUND 18 [POU] and POUND 18/A [POU] were all in close association in Figure 2. No data were available for POUND 21 and therefore no comparison could be made for this other putative mother tree of the IMC clones.
- NA 888 and 733 were fairly closely related to POUND 1/B [POU] and 8/C, two putative NA mother trees, as indicated in Figure 1. NA 733, 227 and 833 were in close proximity to POUND 1/B [POU] in Figure 2. No other NA clones grouped closely with their stated POUND mother trees, and no data were available for other putative mother trees of NA, *viz.*, POUND 19 [POU], POUND 20 [POU], or POUND 30 [POU].
- Some SCA clones, such as SCA 6, 5, 3 and 11 grouped fairly closely with POUND 31/A [POU] and 31/C, the putative mother tree(s) of SCA clones (Figures 1 and 2).

Summary of findings

The IMC clones tended to cluster together and separately from the SCA, MO and NA clones (Figures 1, 2 and 3). However, IMC 3, 63 and 12 were distant from the other IMC clones in Figures 1 and 2.

Frequent grouping together of NA clones was observed in Figures 1, 2 and 3.

SCA 6, 11, 3 and 5 were grouped fairly closely to POUND 31/C [POU] in Figure 2, but were further removed from POUND 31/A [POU]. It must be noted that Motamayor (personal communication) found that POUND 31/A [POU] and 31/C were not very closely linked when analysed using SSR markers.

MO 3, **IMC 3, IMC 63** and PA 37 [PER], and PA 143 [PER] were the most divergent clones in their respective accession groups (Figure 1). MO 96, PA 187 [PER], **IMC 63**, SCA 6, SCA 11, **IMC 3** and IMC 10 were phenotypically distinct in Figure 2.

POUND 18 and 18/A [POU] were well separated in Figure 1 and to a lesser extent in Figure 2. This suggests that there may be a valid basis for distinguishing them as accessions (with suffixes). The relationships among all POUND clones and their variants may be investigated using molecular markers.

According to Pound (1943a), the genotypes collected on the Nanay River as POUND 12

[POU] grew not far from the area where the IMC trees were collected. However, our results suggest that Pound 12/A [POU] is not phenotypically very closely related to the IMC clones.

The SWDI values obtained for the qualitative traits implied that the POUND and MO groups were more diverse than the NA, PA, IMC and SCA groups (Table 5). It is noteworthy that the POUND clones, putative mother trees of the other UAF clones studied, covered the full range of phenotypic diversity expressed in Figure 3.

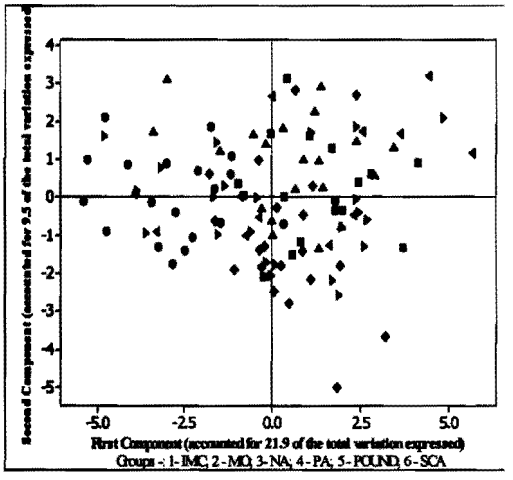

Figure 3 PCA plot depicting phenotypic inter-relationships among the accession groups studied: Group: 1●; 2■; 3♦; 4▲; 5►; 6◄

Discussion and Conclusion

The IMC group was generally phenotypically distinct, particularly from MO accessions (Figure 3). In contrast, there was overlap among the other accession groups. There was slight overlap between the IMC group and 8 POUND, 3 PA, 1 SCA and 2 NA clones (Figure 3). It is noteworthy that Sounigo *et al.* (2001b) found the genetic distance between NA and IMC clones to be relatively small. In comparison to the IMC group, more phenotypic diversity was expressed within the NA, PA, MO and SCA groups, as evident from the many small homogeneous clusters that were widely dispersed in Figure 1, and from the horizontal (revealed by PC1) and vertical (revealed by PC2) spread of the latter accessions in Figure 3. Conversely, Sounigo *et al.* (2005a), who obtained Shannon Weaver diversity index values based on RAPDs for several UAF groups (Table 6), found that the IMC group was just as genetically diverse as the MO, POUND and SCA groups. It must be noted that only 20 accessions were randomly selected to represent the IMC group in this study, and these may differ from those studied by Sounigo *et al.* (2005a).

The pattern of phenotypic diversity, based on the SWDI values obtained for qualitative traits included in this study, was similar to the genetic diversity assessed by Sounigo *et al.* (2005a) in some respects. The POUND and MO groups were the most diverse (Table 5; Figure 3). This was unexpected for the MO group with its one putative mother tree in contrast to the NA and PA groups, which have many putative mother trees.

The relative phenotypic diversity for the accession groups studied, assessed using COV of the quantitative traits (Table 4), showed no clear pattern. However, the considerable phenotypic diversity in the SCA and MO groups, expressed for traits such as PI, cotyledon weight and width, filament pigmentation and pod apex form (Tables 4 and 5), seems not to be explained by a single mother tree source.

Sounigo *et al.*, (2005a) obtained a SWDI value of 0.31 (Table 6) for the NA group. The number of putative mother trees also does not explain this measure of genetic diversity.

The NA group has the most mother trees among the groups studied (approximately 17), but was found to be least diverse genetically. Conversely, the IMC, MO and SCA accession groups had larger SWDI values (0.36) (Table 6) although they only had two and one putative mother tree(s), respectively. The significant yet unexpected genetic diversity in SCA accessions, found by Sounigo *et al.* (2001a; 2005b), was also observed by Mooleedhar (1986). The latter identified eight compatibility alleles among the 11 SCA clones studied, and proposed that at least four parents were involved.

Table 5 Shannon Weaver Diversity Index Values for the categorical descriptors used to characterise the UAF groups

Descriptor	SWDI for Accession Groups					
	IMC	MO	NA	PA	POUND	SCA
Ligule pigmentation	0.54	0.45	0.43	0.36	0.45	0.40
Filament pigmentation	0.53	0.59	0.51	0.46	0.51	0.43
Pedicel pigmentation	0.32	0.25	0.32	0.22	0.45	0.39
Colour of mature pod ridges	0.00	0.26	0.00	0.09	0.00	0.00
Pod shape	0.30	0.26	0.27	0.23	0.40	0.13
Pod basal constriction	0.37	0.45	0.51	0.37	0.47	0.26
Pod apex form	0.21	0.65	0.58	0.52	0.63	0.61
Pod surface texture	0.27	0.31	0.40	0.43	0.43	0.40
Pod furrow disposition	0.18	0.00	0.00	0.00	0.17	0.00
Pod furrow separation	0.29	0.44	0.27	0.32	0.39	0.26
Pod wall hardness	0.22	0.44	0.30	0.35	0.37	0.37
Cotyledon colour	0.37	0.29	0.30	0.30	0.28	0.29
Cotyledon shape	0.45	0.26	0.14	0.37	0.33	0.39
Mean	0.31	**0.36**	0.31	0.31	**0.38**	0.30

Table 6 Shannon Weaver Diversity Index values based on RAPD marker data for UAF groups

Accession Group	Shannon Weaver Diversity Index
IMC (5 clones)	0.36
MO (4 clones)	0.36
NA (8 clones)	**0.31**
PA (11 clones)	0.32
POUND (4 clones)	0.36
SCA (3 clones)	0.36

Source: **Sounigo *et al.* (2005a)**

Bartley (1968) concluded that the SCA accessions were of "mixed origin" based on varied WB resistance levels and compatibility tests.

The aforementioned anomalies may be explained by the widespread self-incompatibility in UAF clones and the possibility that mixed pollinations provided opportunities for a large number of pollen parents to contribute to the genepool. Therein lies the possibility of large diversity among half-sib progenies, as was also observed by Lockwood and End (1993).

These results, in conjunction with those from genetic diversity studies, should prove useful for future cacao breeding programmes and will facilitate the formulation of strategies to effectively manage cacao genetic resources. Detailed consideration of the phenotypic relatedness among the clones studied along with information on their allelic richness will be useful for selecting core and working collections, such as that selected by Sounigo *et al.*, (2005a), where the objective was to preserve as much diversity as possible relative to the larger collection as described by Bekele *et al.* (2004). Information on diversity patterns is also important when formulating strategies for future collections in the wild aimed at increasing the genetic diversity of cacao conserved *ex situ*. A tentative link was established between geographic origin and genetic grouping in cacao by Bekele *et al.* (2006); Laurent *et al.* (1993); Motamayor *et al.* (2002); and Russell *et al.* (1993); Sounigo *et al.* (2005b) among others. This suggests that future collection expeditions in the wild should be conducted in yet unexplored or under-represented areas. Sounigo *et al.* (2005b) recommended collecting in diverse areas to capture more genetic diversity. The apparent uniqueness of Guianese clones (Lachenaud *et al.*, 2001) attests to the usefulness of these strategies.

Research at CRU continues to demonstrate that the ICG,T is an invaluable resource for cacao genetic improvement. It's rich genetic diversity and the presence of superior genotypes such as those identified by Iwaro *et al.* (2003) will facilitate the identification of potentially heterotic groups that may produce transgressive segregants[4], useful when selecting candidate parents for further germplasm enhancement. The ultimate goal is the development of high-yielding, disease-resistant genotypes that will lead to a reduction in the cost of cocoa production, and ideotypes (ideal plant types)[5] to safeguard against yet unknown pests, diseases and challenges to successful cocoa cultivation.

[4] Transgressive segregation may be more likely to occur when parents in a cross are less similar, allowing different favourable alleles to be combined in the off-spring.

[5] An idealised multi-trait characterisation; for example, a crown that is long, narrow and dense with high branch angles and waxy leaves.

Acknowledgements

We wish to acknowledge the Government of Trinidad and Tobago and Cocoa Research Association, U.K. for financial support; Profs. J.A. Spence and L.A. Wilson, Drs. I. Bekele, A.J. Kennedy, T.N. Sreenivasan, A.D. Iwaro, A.B. Eskes and E.S. Johnson for advice, and W. Mollineau, V. Badall, J. Bhola, A. Richardson-Drakes, S. Samnarine, C. Jagroop and others for technical assistance.

References

Abdul-Karimu, A., Butler, D.R., Iwaro, A.D., Sukha, D.A., Bekele, F., Mooleedhar, V. and Shripat, C. (2003) Farmers' perceptions of cocoa planting material and factors affecting the cocoa industry in Trinidad and Tobago, *Tropical Agriculture (Trinidad)* Vol. **80** 261-266

Aime, M.C. and Phillips-Mora, W. (2006) The causal agents of witches broom and frosty pod rot of cocoa (chocolate, *Theobroma cacao*) form a new lineage of Marasmiaceae, *Mycologia* **97**(5) 1012-1022

Alverson, W.S., Whitlock, B.A., Nyffeler, R., Bayer, C. and Baum, D.A. (1999) Phylogeny of the core Malvales: evidence from *ndhF* sequence data, *Am. Journal of Botany* **86** 1474-1486

Anon (1981) *Genetic resources of cocoa* (AGP: IBPGR/80/56) IBPGR Working Group on genetic resources of cocoa, Rome, Italy, IBPGR Secretariat, 25pp.

Bartley, B.G.D (1968) Witches' Broom, *Annual Report of Cacao Research, 1967*, Regional Research Centre, Imperial College of Tropical Agriculture, University of the West Indies, pp. 28-32

Becker, H. (1999) Cacao-good for agriculture and the environment, *Agricultural Research Magazine* **7 (11)** 1

Bekele, F.L. (1999) The role of characterisation and preliminary evaluation data in the genetic improvement of cacao, *Proceedings of Agriculture in the Caribbean: Issues and Challenges,* Volume **11**, August 1998, St. Augustine, Trinidad, The University of the West Indies, Faculty of Agriculture and Natural Sciences-CEPAT, pp. 18-26

Bekele, F.L. and Bekele, I. (1996) A sampling of the phenetic diversity in the International Cocoa Genebank of Trinidad, *Crop Science* **36** (1) 57-64

Bekele, F.L. and G.G. Bidaisee (2006) Characterization of germplasm in the International Cocoa Genebank, Trinidad. Global approaches to cocoa germplasm utilization and conservation, *Final report of the CFC /ICCO/IPGRI project on "Cocoa germplasm utilization and Conservation: a global approach" (1998-2004)* (eds Eskes, A.B. and Efron, Y.), CFC, Amsterdam, The Netherlands/ICCO, London, UK/IPGRI, Rome, Italy, pp. 58-66

Bekele, F. and Butler, D.R. (2000) Proposed list of cocoa descriptors for characterisation. Working procedures for cocoa germplasm evaluation and selection, *Proceedings of the CFC/ICCO/IPGRI Project Workshop* (eds Eskes, A.B., Engels, J.M.M. and Lass, R.A.), February 1998, Montpellier, France, IPGRI, pp. 41-48

Bekele, F.L., Kennedy, A.J., McDavid, C., Lauckner, B. and Bekele, I. (1994) Numerical taxonomic studies on cacao (*Theobroma cacao* L. in Trinidad, *Euphytica* (Kluwer Academic Publishers, Netherlands) **75** 231-240

Bekele, F.L., Iwaro, A.D. and Butler, D.R. (2003) Potential value of cacao germplasm at the International Cocoa Genebank, Trinidad (ICG,T), *Proceedings of the 13th International Cocoa Research Conference*, Kota Kinabalu, Malaysia, October 2000, pp. 219-227

Bekele, F., Bidaisee, G. and Persad, N. (2004) A comparison of the phenotypic diversity of two samples from the International Cocoa Genebank, Trinidad. Annual Report of the Cocoa Research Unit for 2003, St. Augustine, Trinidad, Cocoa Research Unit, The University of the West Indies, p 27-35

Bekele, F.L., Bekele, I., Butler, D.R. and Bidaisee, G.G. (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* (Kluwer Academic Publishers, Netherlands) **53** (5) 933-948

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

Engels, J.M.M. (1986) The systematic description of cacao clones and its significance for taxonomy and plant breeding, Ph.D. thesis, Agricultural University, Wageningen, The Netherlands, 125 pp.

Eskes, A.B. and Lanaud, C. (2001) Cocoa, in: *Tropical Plant Breeding* (eds Charrier, A., Jacquot, M., Hamon, S. and Nicolas, D.), (Translation editor Razdan, M.K.), Plymouth, United Kingdom, Science Publishers Inc. and Montpellier, France, CIRAD, pp. 78-105

ICCO (2006) *Review of the cocoa market situation* – December 2006, Document-**EX/131/3, 7,** November 2006, ICCO Executive Committee - 131st Meeting, December 5-6 2006, EBRD offices, London, http://www.icco.org/economics/cost.aspx (accessed 01/24/2007)

Iwaro, A. D., Bekele, F.L. and Butler, D.R. (2003) Evaluation and utilisation of cacao (*Theobroma cacao* L.) germplasm at the International Cocoa Genebank, Trinidad, *Euphytica* (Kluwer Academic Publishers, Netherlands) **130** 207-221

Iwaro, A.D. Bharath, S.M., Bekele, F.L and Butler, D.R. (2005) Germplasm enhancement for resistance to black pod disease: strategy and prospects. *Proceedings of the 14th International Cocoa Research Conference*, October 13-18 2003, Accra, Ghana, Nigeria, Cocoa Producers Alliance, pp 75-84

Lachenaud, P., Oliver, G. and Letourmy, P. (2001) Agronomic assessment of wild cocoa (*Theobroma cacao* L.) trees from the Camopi and Tanpok Basins (French Guiana). *INGENIC Newsletter* **6**: 24-26

Laurent, V., Risterucci, A-M., and Lanaud, C. (1993) Genetic diversity in cocoa as revealed by cDNA probes, *Theoretical & Applied Genetics* **88** 193-198

Lerceteau, E., Robert,T., Pétiard, V. and Crouzillat, D. (1997) Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RAPD and RFLP markers, *Theor. Appl. Genet.* **95 (1/2)** 10-19

Lockwood, G. and End, M. (1993) History, Technique and future needs for cacao collection. *Proceedings of the International Workshop on Conservation, Characterisation and Utilisation of Cocoa Genetic Resources in the 21st Century,* September 13-17 1992, Port-of-Spain, Trinidad, The Cocoa Research Unit, pp. 1-14

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) *Multivariate Analysis*, London, Academic Press, pp. 213-228

Minitab Inc. (1997) *MINITAB User's Guide 2: Data analysis and quality tools.* Release 12 for Windows. USA, Minitab Inc.

Mooleedhar, V. (1986) An investigation of the alleles of incompatibility in the Scavina population of *Theobroma cacao* L., M.Phil Thesis, University of the West Indies, St. Augustine, Trinidad

Motamayor, J.C., Risterucci, A.M., Heath, M., and Lanaud, C. (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas, *Heredity* **89** 380-386

Motamayor, J.C., Risterucci, A.M., Heath, M. and Lanaud, C. 2003. Cacao domestication II: progenitor germplasm of the Trinitario cacao cultivar, *Heredity* **91** 322-330

N'Goran, J.A.K., Laurent, V., Risterucci, A.M. and Lanaud, C. (2000) The genetic structure of cocoa populations (*Theobroma cacao* L.) revealed by RFLP analysis, *Euphytica* **115** 83-90

NTSYS (2000) *NTSYSpc ver 2.10b users guide*, U.S.A., Applied Biostatistics Inc.

Pound, F.J. (1938) *Cacao and Witches' Broom disease (Marasmius perniciosus) of South America with notes on other species of Theobroma*, Port-of- Spain, Yuille's Printery

Pound, F.J. (1943a) Cacao and Witches' Broom Disease (*Marasmius perniciosus*), Report on a recent visit to the Amazon Territory of Peru, September, 1942 – February 1943, Trinidad and Tobago, Government Printer

Pound, F.J. (1943b) First Report on the selection of cacao trees for resistance to Witches' Broom Disease, Unpublished report, Ministry of Agriculture, Trinidad

Rohlf, F.J (2001) *NTSYS-pc numerical taxonomy and multivariate analysis system.* Version 2.10q, Setauket, N.Y., Exeter Publishing Ltd

Ronnig, C.M. and Schnell, R.J. (1994) Allozyme diversity in a germplasm collection of *Theobroma cacao* L, *Heredity* **85** (4) 291-295

Russell, J.R., Hosein, F., Johnson, E., Waugh, R. and Powell, W. (1994) Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis, *Molecular Ecology* **2** 989-97

Sereno, M.L., Albuquerque, S.B., Vencovsky, R. and Figueira, A. (2005) Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers, *Conservation Genetics* **7** 13-24

Shannon, C.E. and Weaver, W. (1949) *The mathematical theory of communication,* **Urbana**, University Illinois Press

Sneath, P. H. A. and Sokal, R.R. (1973) *Numerical Taxonomy*, San Francisco, U.S.A., Freeman, 573 pp.

Sounigo, O., Ramdahin, S. and Sankar, A. (2001a) Genetic diversity assessment using RAPD, *Annual Report for 2000,* St. Augustine, Trinidad, Cocoa Research Unit, The University of the West Indies, pp. 22-27

Sounigo, O., Christopher, Y., Ramdahin, S., Umaharan, R., and Sankar, A. (2001b) Evaluation and use of the genetic diversity present in the International Cocoa Genebank (ICG,T) in Trinidad, *Proceedings of the International Workshop on New Technologies and Cocoa Breeding,* October 2000, Kota Kinabalu, Sabah, Malaysia, London, INGENIC, pp. 65-71

Sounigo, O., Bekele, F., Iwaro, D., Thévenin, J-M., Bidaisee, G., Umaharan, R., Sankar, A., Sukha, D., Boccara, M., Butler, D.R., Eskes, A.B. (2005a) Description of the CFC/ICCO/IPGRI project collection, *Proceedings of the 14th International Cocoa Research Conference*, October 13-18 2003, Accra, Ghana , Nigeria, Cocoa Producers Alliance, pp. 21-32

Sounigo, O., Umaharan, R., Christopher, Y., Sankar, A. and Ramdahin, S. (2005b) Assessing the genetic diversity in the International Cocoa Genebank, Trinidad (ICG,T) using isozyme electrophoresis and RAPD, *Genetic Resources and Crop Evolution* **52 (8)** 1111-1120

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

Wood, G.A.R. and Lass, R.A. (1985) *Cocoa,* 4th Edn, London, U.K., Longmans