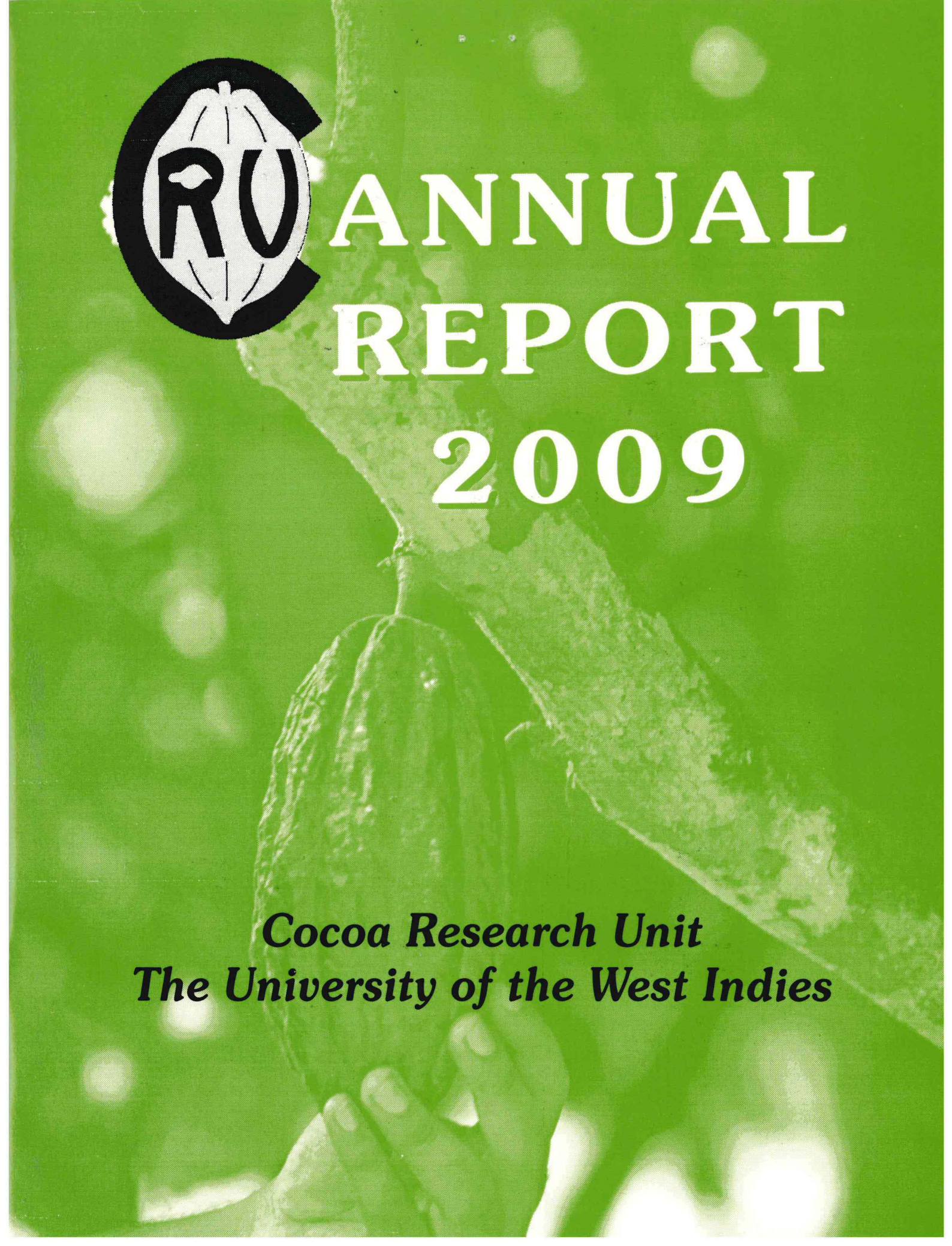

ANNUAL REPORT 2009

Cocoa Research Unit
The University of the West Indies

Annual Report 2009. St. Augustine, Trinidad and Tobago: Cocoa Research Unit, the University of the West Indies. 85 pp.

The work of CRU is made possible by support from

CRA Ltd. research towards sustainable cocoa — Cocoa Research Association, UK

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

Ministerie van Landouw, Natuur en Voedselkwaliteit, Holland

United Nations Common Fund for Commodities

Lindt MAÎTRE CHOCOLATIER — Lindt & Sprüngli (International) AG, Switzerland

Guittard Chocolate Company, Burlingame, USA

Cadbury — Cadbury Ltd., UK

Masterfoods — Masterfoods, UK

Bioversity International — Bioversity International

VALRHONA — Valrhona, France

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

USDA — United States Department of Agriculture, USA

WCF — World Cocoa Foundation, USA

The University of the West Indies, Trinidad and Tobago

University of Reading — The University of Reading, UK

International Cocoa Germplasm Database

UHH Universität Hamburg — The University of Hamburg, Germany

Queensland Government — Queensland Department of Primary Industries & Forestry, Australia

Towson University, Maryland, USA

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

Tel: +1 868 662 8788
+1 868 662 2002 Ext. 2115
Fax:+1 868 662 8788
E-mail: cru@sta.uw.edu

Cover photograph. Trinitario cocoa pod at the ICG,T being examined by Valmiki Singh

Annual Report 2009

Cocoa Research Unit
The University of the West Indies
St. Augustine, Trinidad and Tobago
2010

New insight of genetic diversity and genotype identification using SNP-gene based markers

M. Boccara[1,2], M. Allegre[1], O. Fouet[1], J-M. Thevenin[1], D. Brunel[3], A. Berard[3], D. Zhang[4], D.R. Butler[2], C. Lanaud[1]

Conservation of genetic resources is one of the priorities of the Cocoa Research Unit in Trinidad. Mislabelling, however, is a major problem in germplasm collections all around the world, and the ICG, T is no exception. Estimates of the size of the problem vary, and some estimates indicate that it could reach up to 40% of the trees (Motilal and Butler, 2003).

A joint USDA/CRU collaborative project that aims to fingerprint each original accession held in the ICG,T with microsatellite markers started in 2001.The results of the DNA profiles obtained with SSR markers are currently available for 1,400 accessions from UCRS and Marper Farm.

In recent years, a new DNA fingerprinting technique using SNP (single nucleotide polymorphism) markers has been developed allowing a rapid and accurate identification of an accession. The technique has been used in a collaborative project with CIRAD[1] and CNG[3] with a sample of the ICG,T collection. Although SSR is the marker of choice because they are well characterized with respect to the number of alleles, data collection and analysis are time consuming. On the other hand, because SNP markers typically are bi-allelic, analysis is easier. However being less informative, the number of SNP makers needed would be significantly higher.

Here, results obtained with the two markers are compared to examine their respective suitability and precision.

Materials and methods

DNA samples sent to the USDA-ARS Beltsville laboratory were analyzed with 15 selected SSR primers, following a recommended protocol and guide-lines (Saunders, 2000). Out of the 1,400 accessions analysed so far, 138 accessions were selected among 5 diverse groups: Trinitarios, Iquitos (IMC), French Guyana, Nanay and Parinaris. (Table 1).

DNA samples used by CIRAD/CNG were analyzed with 835 SNP markers; however for the purpose of this study, a selection of 100 of these markers spread over the genome, was used for data analysis.

[1] CIRAD (Centre de coopération Internationale en Recherche Agronomique pour le Développement), - URM DAP TA A96/03-34398 Montpellier cedex 5 – France
[2] CRU (Cocoa Research Unit) The University of the West Indies, St Augustine, Trinidad and Tobago
[3] INRA-EPGV, CEA/Institut de Génomique - Centre National de Génotypage, 2 rue Gaston Crémieux, CP 5724, 91057 Evry Cedex – France
[4] United States Department of Agriculture- Agriculture Research Service, PSI SPCL,10300 Baltimore Avenue, Bldg. 001, Rm. 223, BARC-W, Beltsville, MD 20705, USA

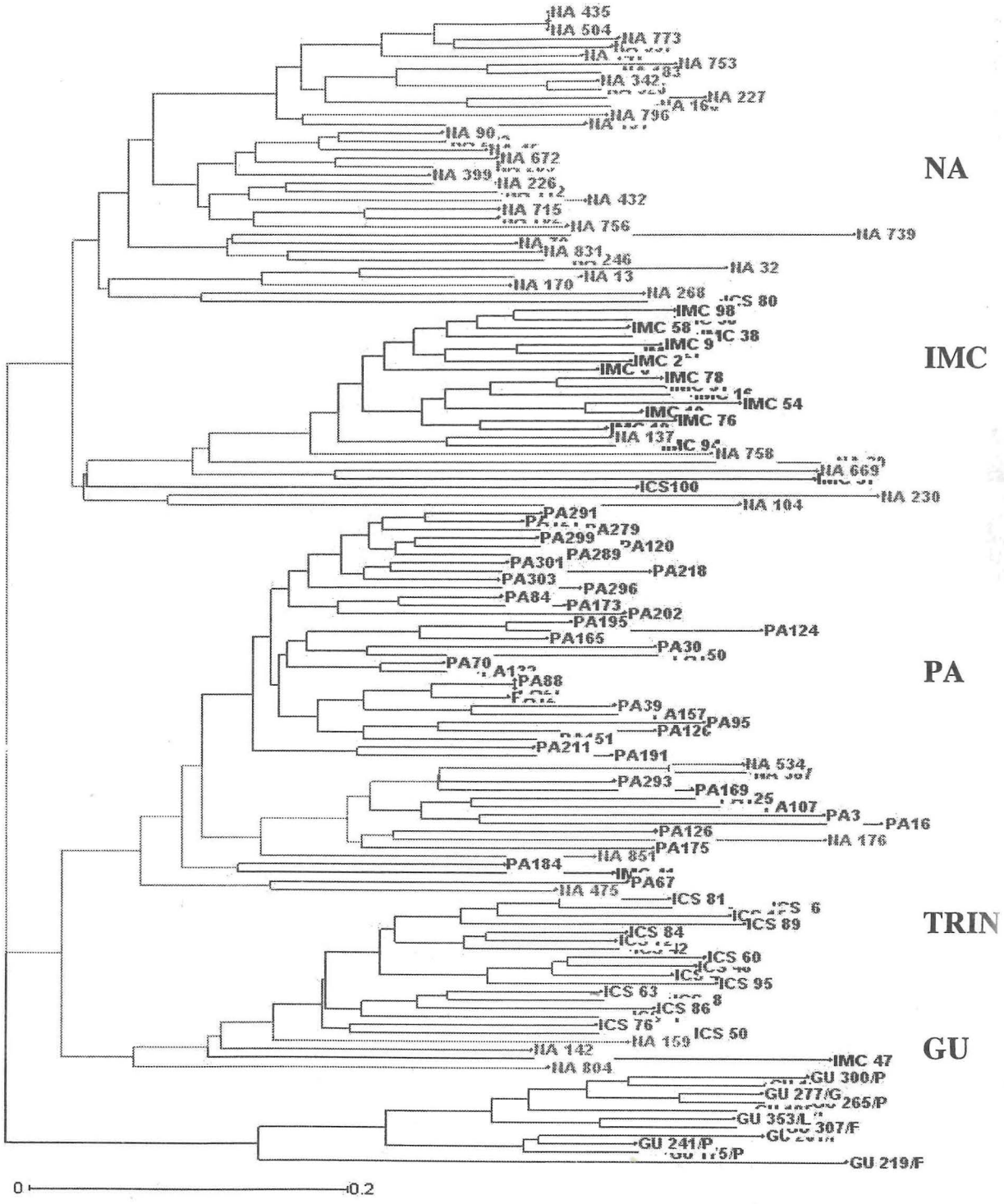

Figure 1. Dendrogram of dissimilarity analysis run with 15 SSR markers.

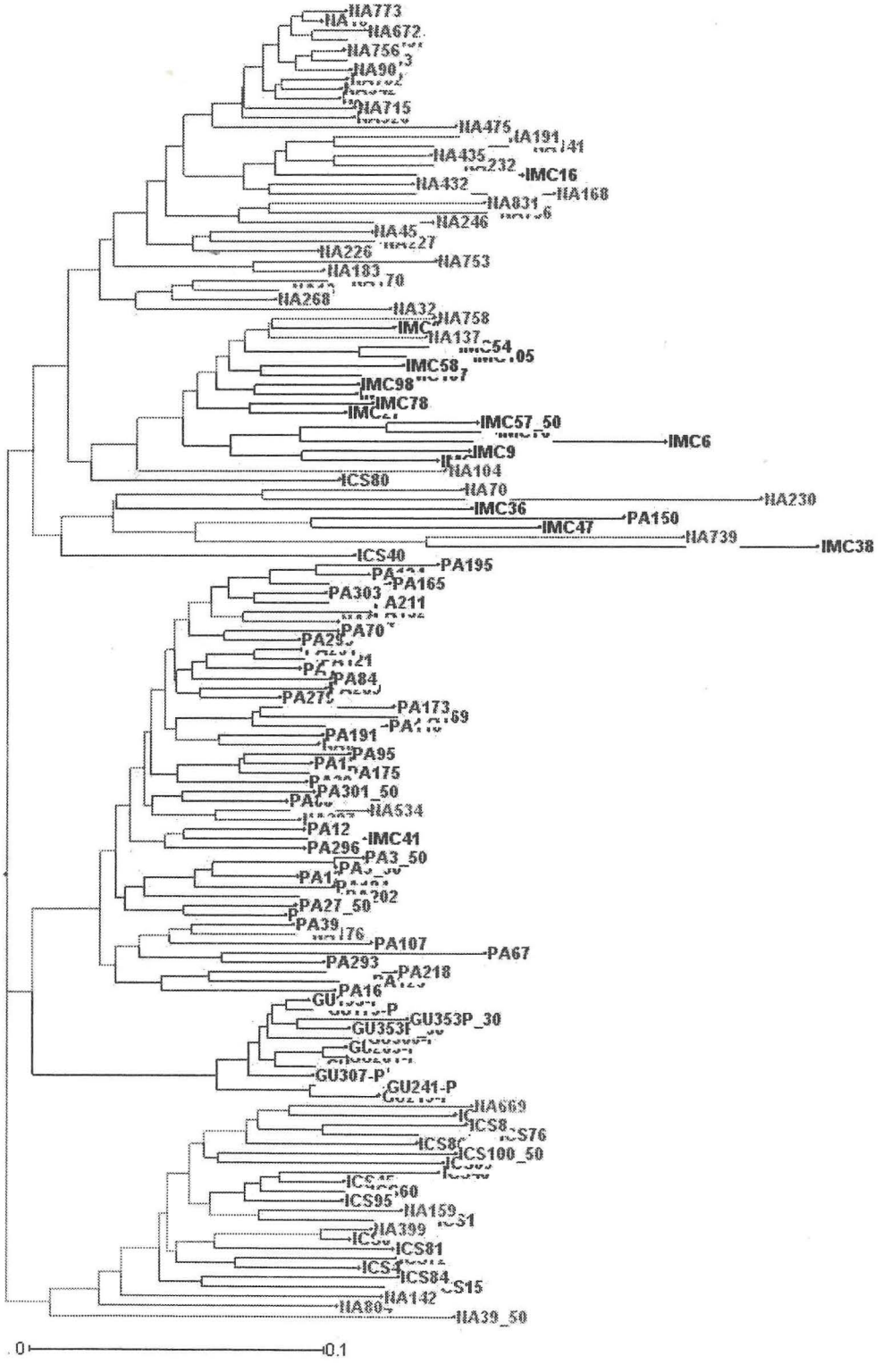

Figure 2. Dendrogram of dissimilarity analysis run with 100 SNP markers.

Data analysis

Results of DNA fingerprints obtained with SSR and SNP markers were compared to verify the accuracy of both techniques for:

- assessing genetic relationships
- detecting mislabelling

Fingerprints generated with 15 SSR markers and 100 SNP markers were run with DARwin software (DARwin5, version 5.0.158) for cluster analysis (displayed as dendrograms) and off-type detection, by generating Weighted Neighbour-Joining trees. The program Structure v2.3 was also used, with the number of clusters set to 5. All Structure runs used an admixture model with 100,000 iterations after a burn-in period of 100,000 and reiterated 10 times.

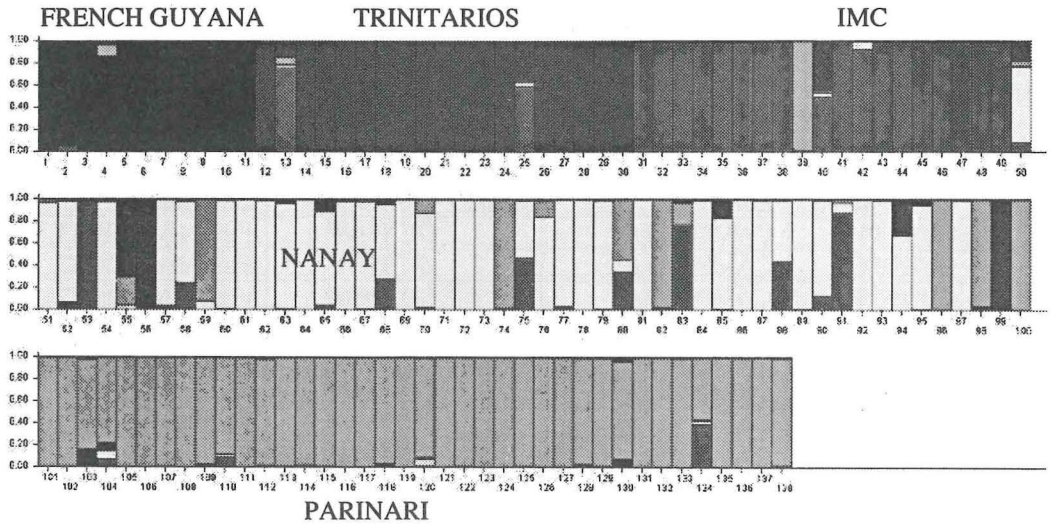

Figure 3. Representation of clusters using STRUCTURE with K=5 : 138 accessions run with 15 SSR markers.

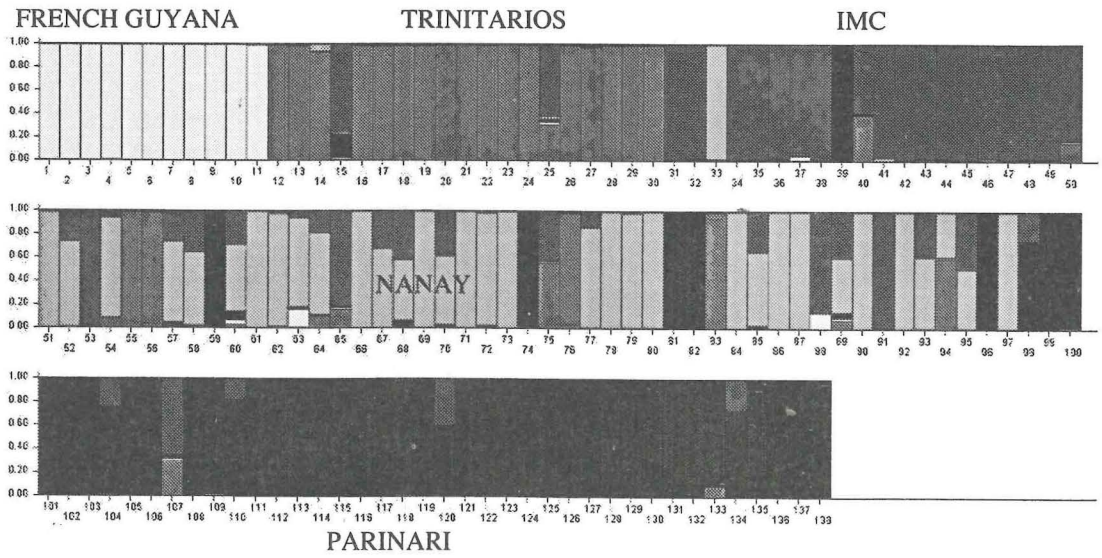

Figure 4. Representation of clusters using STRUCTURE with K=5 : 138 accessions run with 100 SNP markers.

Results

Analysis of diversity

The dendrogram of dissimilarity for 138 accessions analyzed with the 15 SSR markers showed the expected clustering into 5 groups according to their origin: Trinitario, Guyanese, IMC, Nanay and Parinari. (Figure 1). Very similar clustering was obtained from the results of fingerprints given by the selected 100 SNP markers (Figure 2).

The results of Bayesian clustering analysis (Structure software) confirmed the above results, with a similar outcome for SSR and SNP markers when K was set to 5 (Figures 3 and 4). Out of the 138 samples, 102 were assigned in expected group with a 90% confidence threshold, and off-type trees were detected with the remaining samples (Tables 2 & 3).

Table 2. List of true-to-type accessions identified with SSR markers and confirmed with SNP markers.

Accessions run with SSR markers			Accessions run with SNP markers		
Code	**Clone name**	**Location**	**Code**	**Clone name**	**Location**
1	GU 114/P	CAMPUS 1A	1	GU 114/P	Field 4A B195 T1
2	GU 175/P	CAMPUS 1A	2	GU 175/P	Field 4A B228 T2
3	GU 195/P	CAMPUS 1A	3	GU 195/P	Field 4A B229 T2
4	GU 219/F	CAMPUS 1A	4	GU 219/F	Field 4A B237 T1
5	GU 241/P	CAMPUS 1A	5	GU 241/P	Field 4A B258 T2
6	GU 261/P	CAMPUS 1A	6	GU 261/P	Field 4A B231 T2
7	GU 265/P	CAMPUS 1A	7	GU 265/P	Field 4A B230 T2
8	GU 277/G	CAMPUS 1A	8	GU 277/G	Field 4A B259 T2
9	GU 300/P	CAMPUS 1A	9	GU 300/P	Field 4A B197 T2
10	GU 307/F	CAMPUS 1A	10	GU 307/F	Field 4A B232 T2
11	GU 353/L	CAMPUS 1A	11	GU 353/L	Field 4A B199 T1
12	ICS 1	S.J.E	12	ICS 1	Field 6B B122 T9
14	ICS 15	S.J.E	14	ICS 15	Field 4A C302 T1
16	ICS 42	S.J.E	16	ICS 45	Field 6B B113 T6
17	ICS 48	S.J.E	17	ICS 48	Field 6B E318 T6
18	ICS 50	S.J.E	18	ICS 49	Field 6B B121 T4
19	ICS 6	S.J.E	19	ICS 6	Field 6B E281 T15
20	ICS 60	S.J.E	20	ICS 60	Field 6B E332 T6
21	ICS 63	S.J.E	21	ICS 63	Field 6B E317 T2
22	ICS 72	S.J.E	22	ICS 72	Field 6A A72 T9
23	ICS 76	CAMPUS 11	23	ICS 76	Field 6B B109 T3
24	ICS 8	S.J.E	24	ICS 8	Field 6B B111 T2
26	ICS 81	S.J.E	26	ICS 81	Field 4A C281 T1
27	ICS 84	S.J.E	27	ICS 84	Field 6B E329 T3
28	ICS 86	S.J.E	28	ICS 86	Field 6B E330 T5
29	ICS 89	S.J.E	29	ICS 89	Field 6B E344 T14
30	ICS 95	S.J.E	30	ICS 95	Field 6B B84 T4
31	IMC 105	MARPER	31	IMC 105	Field 6B A24 T6
32	IMC 107	Field 6B A28 T3	32	IMC 107	Field 6B A28 T3
34	IMC 2	Field 6B A41 T8	34	IMC 2	Field 6B A41 T8
35	IMC 27	MARPER	35	IMC 27	Field 6B A20 T3
36	IMC 31	MARPER	36	IMC 31	Field 6B A32 T9
37	IMC 38	MARPER	37	IMC 36	Field 6B A62 T1

38	IMC 38	Field 6B A21 T15	38	IMC 38	Field 6B A21 T3
41	IMC 54	MARPER	41	IMC 54	Field 6B A9 T3
42	IMC 57	MARPER	42	IMC 57	Field 6B A22 T7
43	IMC 58	MARPER	43	IMC 58	Field 6B A47 T5
44	IMC 6	MARPER	44	IMC 6	Field 6B A1 T6
45	IMC 76	MARPER	45	IMC 76	Field 6B A45 T4
46	IMC 78	MARPER	46	IMC 78	Field 6B A14 T2
47	IMC 9	MARPER	47	IMC 9	Field 6B A40 T3
48	IMC 94	MARPER	48	IMC 94	Field 6B A17 T7
49	IMC 98	Field 6B A63 T3	49	IMC 98	Field 6B A63 T3
51	NA 112	Field 5B F491 T1	51	NA 112	Field 5B F491 T1
52	NA 13	MARPER	52	NA 13	Field 6B C174 T2
54	NA 141	Field 5B G620 T4	54	NA 141	Field 5B G620 T4
60	NA 183	Field 5B G603 T2	60	NA 183	Field 5B G603 T2
61	NA 184	Field 5B G612 T1	61	NA 184	Field 5B G612 T8
62	NA 191	Field 5B F433 T3	62	NA 191	Field 5B F433 T3
63	NA 226	Field 6B E292 T6	63	NA 226	Field 6B E292 T6
64	NA 227	Field 5A D312 T1	64	NA 227	Field 5A D312 T1
66	NA 232	Field 5B G628 T7	66	NA 232	Field 5B G628 T7
69	NA 283	Field 5B G618 T3	69	NA 283	Field 5B G618 T3
71	NA 326	MARPER	71	NA 326	Field 5B E416 T1
72	NA 337	MARPER	72	NA 337	Field 5B G617 T2
73	NA 342	Field 6B C168 T6	73	NA 342	Field 6B C168 T6
77	NA 432	MARPER	77	NA 432	Field 6B E293 T9
78	NA 435	Field 5B F531 T3	78	NA 435	Field 5B F531 T3
79	NA 45	MARPER	79	NA 45	Field 5B F510 T10
84	NA 672	MARPER	84	NA 672	Field 5B F477 T3
86	NA 702	MARPER	86	NA 702	Field 5B G631 T3
87	NA 715	MARPER	87	NA 715	Field 5A D338 T5
90	NA 756	MARPER	90	NA 756	Field 6A B101 T9
92	NA 773	Field 5B F547 T3	92	NA 773	Field 5B F547 T3
97	NA 90	MARPER	97	NA 90	Field 5B F550 T12
98	PA 107 [PER]	MARPER	98	PA 107 [PER]	Field 5A D247 T3
100	PA 12 [PER]	MARPER	100	PA 12 [PER]	Field 6B D200 T2
101	PA 120 [PER]	Field 6B D188 T2	101	PA 120 [PER]	Field 6B D188 T13
102	PA 121 [PER]	MARPER	102	PA 121 [PER]	Field 6B C166 T10
103	PA 124 [PER]	MARPER	103	PA 124 [PER]	Field 6B D192 T8
104	PA 125 [PER]	MARPER	104	PA 125 [PER]	Field 5B F527 T8
105	PA 126 [PER]	Field 6B D198 T14	105	PA 126 [PER]	Field 6B D198 T4
106	PA 132 [PER]	MARPER	106	PA 132 [PER]	Field 5B D275 T3
108	PA 151 [PER]	MARPER	108	PA 151 [PER]	Field 5B F437 T1
109	PA 157 [PER]	Field 5B F466 T3	109	PA 157 [PER]	Field 5B F466 T11
110	PA 16 [PER]	Field 6B D186 T1	110	PA 16 [PER]	Field 6B D186 T13
111	PA 165 [PER]	Field 5B F451 T1	111	PA 165 [PER]	Field 5B F451 T1
112	PA 169 [PER]	MARPER	112	PA 169 [PER]	Field 6B C180 T3
113	PA 173 [PER]	Field 5B F480 T8	113	PA 173 [PER]	Field 5B F480 T3
114	PA 175 [PER]	MARPER	114	PA 175 [PER]	Field 5B F473 T6
115	PA 184 [PER]	MARPER	115	PA 184 [PER]	Field 5B F490 T8
116	PA 191 [PER]	MARPER	116	PA 191 [PER]	Field 5B F536 T4
117	PA 195 [PER]	Field 6B C165 T1	117	PA 195 [PER]	Field 6B C165 T1
118	PA 202 [PER]	Field 5A D309 T1	118	PA 202 [PER]	Field 5A D309 T1
119	PA 211 [PER]	MARPER	119	PA 211 [PER]	Field 5B F528 T2
121	PA 27 [PER]	Field 5B E423 T4	121	PA 27 [PER]	Field 5B E423 T3

122	PA 279 [PER]	MARPER	122	PA 279 [PER]	Field 6B D197 T2
123	PA 289 [PER]	Field 5B F535 T12	123	PA 289 [PER]	Field 5B F535 T1
124	PA 291 [PER]	Field 6B C167 T6	124	PA 291 [PER]	Field 6B C167 T13
125	PA 293 [PER]	MARPER	125	PA 293 [PER]	Field 5A D308 T7
126	PA 296 [PER]	Field 6B D207 T1	126	PA 296 [PER]	Field 6B D207 T6
127	PA 299 [PER]	Field 5B E398 T6	127	PA 299 [PER]	Field 5B E398 T2
128	PA 3 [PER]	MARPER	128	PA 3 [PER]	Field 5B E355 T2
129	PA 30 [PER]	Field 6B C144 T1	129	PA 30 [PER]	Field 6B C144 T7
130	PA 300 [PER]	Field 5B E407 T14	130	PA 300 [PER]	Field 5B E407 T14
131	PA 301 [PER]	MARPER	131	PA 301 [PER]	Field 5A D320 T7
132	PA 303 [PER]	MARPER	132	PA 303 [PER]	Field 6B D211 T3
133	PA 39 [PER]	Field 5A D264 T1	133	PA 39 [PER]	Field 5A D264 T3
135	PA 70 [PER]	Field 5B F489 T14	135	PA 70 [PER]	Field 5B F489 T10
136	PA 84 [PER]	Field 5B E388 T2	136	PA 84 [PER]	Field 5B E388 T7
137	PA 88 [PER]	MARPER	137	PA 88 [PER]	Field 5B F443 T1
138	PA 95 [PER]	MARPER	138	PA 95 [PER]	Field 5B F460 T11

Table 3. List of off-type accessions identified with SSR markers and confirmed with SNP markers.

Accessions run with SSR markers			Accessions run with SNP markers		
Code	**Clone name**	**Location**	**Code**	**Clone name**	**Location**
53	NA 137	Field 6B C155 T15	53	NA 137	Field 6B C155 T10
55	NA 142	MARPER D682	55	NA 142	Field 6A B89 T3
56	NA 159	MARPER D650	56	NA 159	Field 5B G635 T14
59	NA 176	Field 4A D389 T4	59	NA 176	Field 5B E403 T2
74	NA 387	Field 5A D251 T2	74	NA 387	Field 5A D251 T2
75	NA 39	MARPER D138	75	NA 39	Field 4A D370 T1
76	NA 399	MARPER D456	76	NA 399	Field 4A D408 T3
82	NA 534	Field 5B G630 T1	82	NA 534	Field 5B G630 T2
83	NA 669	MARPER C733	83	NA 669	Field 4A D418 T2
39	IMC 41	Field 6B F418 T1	39	IMC 41	Field 6B F418 T15
40	IMC 47	Field 6B F401 T1	40	IMC 47	Field 6B F401 T9
13	ICS 100	S.J.E	13	ICS 100	Field 6B B100 T1
15	ICS 40	S.J.E	15	ICS 40	Field 6B E287 T4
25	ICS 80	S.J.E	25	ICS 80	Field 6A A72 T9
107	PA 150 [PER]	MARPER D697	107	PA 150 [PER]	Field 6B C179 T1
134	PA 67 [PER]	Field 5B E346 T4	134	PA 67 [PER]	Field 5B E346 T11

Detection of mislabelling

Some accessions used in the sample had been previously identified as mislabelled or rootstock in the CRU/USDA Fingerprinting Project (Table 4, Boccara and Zhang, 2005, 2008).

Both marking techniques were able to detect mislabelling of trees.

Table 4. List of accessions previously identified as off-types.

Clustered with Trinitario accessions	Clustered with IMC accessions	Clustered with PA accessions
NA 159	NA 137	NA 176
NA 142	NA 758	NA534
NA 804	NA 39	NA387
PA 114 [PER]	NA 669	NA 851
IMC 47	NA 230	IMC 41
	ICS 80	
	ICS 100	

French Guyana

No mislabelled tree was recorded with SSR markers with samples collected from Campus Field 1B; SNP markers showed that all the trees tested were correctly propagated when they were established in Field 4A at UCRS.

Trinitario trees

The analysis with SNP confirmed that ICS 80, detected as an off-type at the San Juan Estate with SSRs, doesn't belong to the Trinitario group.

The SNP results suggest that ICS 100 (detected as non-Trinitario in the San Juan Estate with SSRs), could be correctly labelled in Field 6B, Plot B100 at UCRS. In the case of ICS 40, the tree sampled from the genebank (Field 6B, Plot E287, T4) doesn't match the original tree, confirming earlier morphological observations (Bekele et al., 2004). Structure analysis suggests with a probability of 80% assigning this accession to the IMC group (Figure 4).

IMC trees

The SNP markers confirm without ambiguity that IMC 41 belongs to the Parinari group. It had previously been suggested that the trees in Field 6B, Plot F418 were propagated from one of the neighbour trees in Marper Farm, PA 200 [PER] or PA 207 [PER] (Boccara, 2006; Bekele et al., 2005).

The analysis confirms that neither of the two IMC 47 samples (trees 1 and 9) from Field 6B, Plot F401, conform to the IMC group. Previous results (Boccara et al., 2004) suggested that trees 1, 3, 4, 11, 12 were identical, but were different to both the Marper tree and the Campus tree, and were similar to IMC 57 (IMC 57 not falling in the IMC group) .

The SSR profile IMC 16 in Marper Block D603 showed that it belongs to the IMC group; however the same technique revealed that Tree 2, in Field 6B, Plot A11 has a NA profile; SNP marker analysis of this tree gave the same information, inferring that budwood could have been taken from NA 105, the adjacent tree in Marper Farm (now recorded as dead).

Nanay trees

All the accessions already identified as off-type (Table 4), were confirmed with SNP markers, and furthermore each can be assigned to a specific accession group.

Case of NA 475
Tree 9 in Field 5B, Plot F534 was identified as an off-type with SSR markers; the SNP profile however, suggests that tree 3 in Field 4A, Plot D415 was correctly re-propagated from the original tree in Marper Farm.

Parinari trees

As for Nanay group, SNP markers confirmed the identity of the correct labelled accessions and identified the mislabelled ones.
Case of PA 150
Whereas the original tree in Marper Farm had been identified as correct, the SNP profile confirms that Tree 1 in Field 6B, Plot C179 is not identical. The result confirms earlier pod morphology observations and microsatellites markers results (Motilal et al., 2008).

Conclusion and future perspectives

The results of the analysis of the genetic diversity with SNP markers are in complete agreement with those obtained with SSR markers. Comparison with original reference trees and assignment tests demonstrated the efficiency and accuracy of the 100 selected SNP markers for the detection of mislabelling.

More verification of mislabelled trees will be needed to reduce the risks of erroneous duplication and distribution of trees from UCRS.

A technical problem of the use of SSR markers is that it is not easy to compare data produced by different laboratories: discrepancy in allele size calling mainly due to the large variety of automatic sequencing machines and software used; SNP markers would be a suitable tool for use at CRU for future identification work. Preliminary tests have shown that a subset of 65 selected markers was efficient for the unambiguous recognition of mislabelled trees.

Acknowledgements

We thank Antoinette Sankar for DNA sample preparation, Frances Bekele for sharing her knowledge on morphological traits, and L.A. Motilal for advice. USDA-ARS Beltsville, CIRAD and CNG were responsible for the efforts made to run the samples and collect the data.

References

Bekele, F.L., Bidaisee, G.G., Persad, N., Bhola, J. (2005) Examining phenotypic relationships among Upper Amazon Forastero clones. Pages 27-42, in: Annual Report 2004. St Augustine, Trinidad and Tobago: Cocoa Research Unit, the University of the West Indies.

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

Boccara, M., Zhang, D., (2006) Progress in resolving identity issues among Parinari accessions held in Trinidad: the contribution of the collaborative USDA/CRU project. Pages 25-37, in: Annual Report 2005. St Augustine, Trinidad and Tobago: Cocoa Research Unit, the University of the West Indies.

Boccara, M., Zhang, D., (2009) Progress in resolving identity issues among the Nanay accessions held in Trinidad: the contribution of the collaborative USDA/CRU project. Pages 37-45, in: Annual Report 2008. St Augustine, Trinidad and Tobago: Cocoa Research Unit, the University of the West Indies.

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

Motilal, L.A., Umaharan, P., Zhang, D., Boccara, M. (2008) Fingerprinting cacao trees in the International Genebank, Trinidad with microsatellites. Pages 22-29, in: Annual Report 2007. St Augustine, Trinidad and Tobago: Cocoa Research Unit, the University of the West Indies

Perrier, X., Flori, A., Bonnot, F. (2003). Data analysis methods. Pages 43-76 in: Genetic diversity of cultivated tropical plants (P. Hamon, M. Seguin, X. Perrier, J.C. Glaszmann Eds). Montpellier. Enfield Science.

Perrier, X., Jacquemoud-Collet, J.P. (2006). DARwin software http://darwin.cirad.fr/darwin

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000a). Inference of population structure using multilocus genotype data. Genetics, 155:945–959.

Structure software: Version 2.3 http://pritch.bsd.uchicago.edu/structure.html April 21, 2009

Saunders, J.A. (2000) USDA DNA Fingerprinting Programme for Identification of Theobroma Cacao Accessions. Pages 108-114 in: Proceedings of the International Workshop on New Technologies and Cocoa Breeding. 16th-17th October 2000, Malaysia.