2017 International Symposium of Cocoa Research (ISCR), Lima, Peru, 13-17 November, 2017

DNA PROFILING OF CACAO (*Theobroma cacao* L.) VARIETIES IN THE PHILIPPINES USING MICROSATELLITE MARKERS

R. R.C. Espino[1], K.A. Tonogbanua[2], M. R. M. Espino[1] and J. V. Ramos[3]

[1]Institute of Crop Science, College of Agriculture and Food Sciences, University of the Philippines, Los Baños, Laguna, Philippines

[2]University of San Carlos, Cebu City, Philippines.

[3]Organization for Partnerships, Teamworks and Initiatives on Opportunities for Nature Stewards, Inc. Sto. Tomas, Batangas, Philippines.

ABSTRACT

Fifteen (15) single sequence repeat (SSR) markers were used to construct a DNA profile of Philippine cacao varieties. This is as to assess the genetic similarity, genetic diversity and elucidate relationships between varieties/clones. A total of 50 accessions of cacao were collected from 10 provinces in the country composed of 20 varieties/clone. Of these, six (6) are National Seed Industry Council (NSIC) -registered. UPGMA similarity coefficients observed from the cluster analysis of the standard varieties ranged from 0.09 to 0.72, indicating that standard varieties is distinct from one another. The varieties with the highest similarity (72%) were between Criollo clones (Criollo Green and Criollo Red). At 0.22 similarity trinitarios composed majority in one group. The mixing of other criollos with the trinitario group could be due to their genetic relationship where trinitarios are hybrids between Criollo and Amellonado. All of the primers characterized were highly polymorphic with values ranging from 0.6341 to 0.8702. Markers used were repeatable and sufficient and effective in detecting genetic similarities and establishing genetic profiles of the collection.

INTRODUCTION

Since 1970, the global demand for cacao has tripled. According to the International Cocoa Organization (ICCO), Europe and US remains to be the greatest importers of cocoa with recorded net imports of 1,600,000MT and 780,000MT, respectively in 2004 to 2009. Recently, Asia is said to have an increasing demand for cacao. There was observed 12% increase in the number of consumers from 2003 to 2008 in the region. One country in Asia, Philippines, is now gaining popularity for its good chocolate quality. Philippines, is a country which lies along the same general area in the equator with the biggest cacao producers are the Ivory Coast, Ghana, Papua, New Guinea, Indonesia, and the South American countries. The country possesses the climatic conditions that is favorable for cacao production. However, with local consumption higher than the production, Philippines is still a net importer. In the past few years there has been an improvement in the growth rate of cacao production in the country. Many efforts have been made to boost the cacao industry in the country. The Cacao Industry Development Association of Mindanao, Inc. (2017) is an example. It is an organization which expressed their commitment to help achieve 100,000MT cocoa production by the year 2020 and onwards.

Cacao (*Theobroma cacao* L.) is a perennial crop famous as the raw material for making chocolates is cultivated between latitudes 20°N and 20°S in tropical regions around the world. Cacao trees have a long juvenile stage. It usually takes 4-5 years (seed-to-seed) before it can produce its first pod (Bartley, 2005). This long juvenile stage is a major challenge in authenticating varieties as morphological markers differentiating from one variety to another might only be observable only at mature stage. Another problem is that, there is occurrence of tree misidentification or mislabeling of clones in the germplasm collection, as well as replacement of grafted scions by the rootstock that may have occurred for various reasons. The economically important characteristics of the crop are genetically defined and most of the time, operators selling seedlings only characterize or identifies their plants based on the phenotype. This matter then raises confusions and doubt in terms of the accuracy of varietal labels of planting materials being released by nurseries and sold to the farmers and private sector investors in cacao production.

Other than morphological markers, there are other types of markers that has been developed, namely anatomical, biochemical and molecular. The last is also referred to as DNA marker technology, which is based on DNA polymorphisms. DNA markers are not dependent on the age of the plant thus even at vegetative stage DNA can be isolated and analyzed in the laboratory for varietal identification. Microsatellites also known as SSR primers are PCR-based DNA markers wherein they can amplify tandem sequences repeat. Microsatellite markers are numerous in plants. They can be applied in varietal identification, genetic diversity analysis, DNA fingerprinting and linkage map construction. Use of this type of DNA markers has been successful in crops like maize, rice, grapevines, brassicas and guava (Westman and Kresovich, 1999; Rai et al., 2013). There are also works on cacao characterization using this type of markers such that of Johnsiul (2016), Motilal et al (2012) and Schawe et al (2013).

There is no established DNA profile for the Philippine cacao collection that can serve as basis for varietal authentication of seedling distributed to local farmers. The goal of the study is to establish the genetic characteristics of cacao germplasm in the Philippines by performing molecular and field data analysis using SSR technology to aid in the conservation and varietal development.

MATERIALS AND METHODS

Site and Date of Study

The study was conducted at the Molecular Laboratory, Tissue Culture Facility and Molecular Plant Breeding Laboratory, Crop Science Cluster, College of Agriculture, UPLB, Laguna from November 2015 to December 2016. Plant samples were kept and maintained at the Fruit Crops Nursery, Crop Science Cluster, UPLB.

Collection and Microsatellite Markers

Cacao accessions were collected from 10 provinces around the country. Varieties used as standards for this study were from University of Southern Mindanao-Genebank and Bureau of Plant Industry, Davao. Complete list of cacao accession is presented in Table 1. Intenational standard single sequence repeat (SSR) markers for cacao characterization are listed in Table 2.

DNA Extraction

DNA was isolated from leaves by modifying the cethyltrimethyl ammonium bromide (CTAB) protocol. The extraction buffer contained 10% Tris, 16% sodium chloride (NaCl), 1.6% sodium-EDTA, 2% CTAB and 2% mercaptoethanol. Leaf samples were ground using liquid nitrogen and polyvinyl pyrolidine (PVP) .Samples were then placed in falcon tubes containing the buffer and incubated for 1 hour at 65°C, inverted every 15mins. Centrifuged after incubation at 10 000rpm for 5min. Volume of 450µL of the liquid were pipetted out and placed in new eppy tubes. 24:1 chloroform isoamyl (450µL) and 100µL 2M NaCl + 4% polyetheylene glycol (PEG) were added. Samples were chilled for 15°C then centrifuged at 10 000rpm for 10min. The upper component of the mixture (450µL) was then transferred to new tubes and cold 2-isopropanol (380µL) was added to them. Samples were then incubated overnight at -20 ºC then centrifuged at 10 000 rpm for 5min to pellet out the DNA. Pellets were then washed twice using 70% ethyl alcohol (300µL) and dried. TE buffer (30µL) with RNase (0.4µL) was added.

Single Sequence Repeat (SSR) Characterization

Band products were scored as either present (1) or absent (0). Observed number of bands, band size range (bp), number of polymorphic alleles, polymorphism rate (%) and polymorphism information content (PIC). Polymorphism rate was computed by dividing the number of bands observed with the number of polymorphic alleles and multiplying the result by 100. The allelic diversity at a locus was also measured using polymorphism information content described by Saal and Wrickle (1999):

$$\mathbf{PIC = 1- \Sigma} p_i^{\,2}$$

where p_i is the frequency of the i th allele out of the total number of alleles at an SSR locus. PIC is the probability that an individual is informative with respect to the segregation of its inherited alleles.

Cluster Analysis

Jaccard similarity coefficients generated were used to construct dendrograms. Unweighted pair group arithmetic mean (UPGMA) was used to form clusters using NTSYSpc Version 2.10.

DNA Profiling

A haplotype table was generated to present the resulting bands in a simpler manner for easy visualization of differences in banding patterns. Allele presence was represented as shaded cells on the table. The alleles were presented accordingly with the loci (markers) where they were observed. Alleles per locus were arranged according to their sizes (from higher bp to lower bp). This way, the map appears similar with the PAGE gels. Larger amplicons (higher bp) migrate slower and appear on the top part of the gel while the smaller ones migrate faster and appears at the bottom. The gel serves as a molecular sifter where amplified fragments of different size can be visualized.

The haplotype map provided information on the number of alleles that was observed per locus, sizes of these alleles and the diverse banding patterns.

RESULTS AND DISCUSSION

Collection of Cacao Clones

A total of 50 accessions of cacao were collected from 10 provinces in the country composed of 13 varieties and 7 Criollo clones. Of the 14 varieties, 6 are NSIC registered namely, BR25 (NSIC 2000 Cc05), K1 (NSIC 2001 Cc 06), K2 (NSIC 2001 Cc07), ICS 40 (NSIC 2000 Cc01), UF18 (NSIC 2008 Cc 08) and UIT 1(NSIC 2000 Cc02). A complete list of the accessions collected and used in the study are presented in Table 2.

DNA Extraction and Quantification

DNA yields ranging from ~84.90 to 576.11 ng/µl were obtained using the modified CTAB (Doyle and Doyle, 1987) protocol. Since only 1µL with 30ng concentration of DNA template is needed per reaction, the quantity of DNA isolated in this study was sufficient for the necessary PCR runs for PCR condition optimization and DNA profile construction.

Most of isolated DNA were also within the 1.8-2.0 absorbance ratio (260/280). Those absorbance ratio below 1.8 contains proteins such and/or other secondary metabolites. This is likely to be present in cacao as its ground leaf produces a thick consistency when mixed with the CTAB buffer during extraction. During transfers of supernatant the thick consistency may promote inclusions of other proteins. A double extraction was done but the DNA yield after is very low with little adjustments on the absorbance ratio. Samples with absorbance ratio below 1.8 were tested to check if it can produce reliable amplicons. The DNA isolated was of good quality. Running it in 2% agarose gel resulted in clear bands with little or no smears, which meant that the genomic DNA isolated was pure and intact and can be used in PCR runs. SSR markers do not require high DNA quality. But a pure and intact template is preferred as it results in highly repeatable amplicons.

Genetic Diversity

To assess the genetic similarity, genetic diversity and elucidate relationships between the clones, analysis at the genetic level was done. Two dendrogram were constructed. First tree plot (Figure 1) includes the authentic varieties from BPI Davao, such as BR25, KI, K2, UF18, PBC123, UIT1, ICS40, K9, S5 and two Criollo types (Criollo Red and Criollo Green). Other standard varieties such as USMCH1, USMCH2, P7, DR1 and Criollos (Criollo 21 and Criollo 22) were from USM, Kabacan repository. The second tree plot (Figure 2) includes all of the clones collected.

Similarity coefficients observed from the cluster analysis of the standard varieties (Figure 1) using UPGMA ranged from 0.09 to 0.72. The varieties with the highest similarity (72%) were Criollo Green and Criollo Red. Closely related to these two were Criollo 22 with 65.50% similarity. Criollo 21 and Criollo (New Leyte) are closely related with 69.30% similarity. Remaining Criollos (Criollo-Quezon and Criollo-Maragusan) were observed to group closer to other varieties. At 0.10 cut of the dendrogram, 2 clusters are formed. In cluster I, 13 out of 17 varieties included were of Trinitario type. Cluster II were composed only of Criollo types. These observed clustering is due to the inherent difference between these two cacao types. The mixing of other Criollos with the Trinitario group could be due to their genetic relationship where trinitario clones are hybrids between Criollo and Amellonado.

Cluster analysis for the whole collection (Figure 2) obtained similarity coefficient range of 0.14-0.98. The highest similarity was between clones of BR25, one from Kabacan and the other was from Batangas. At 0.78, two other BR25 (Zamboanga Norte and Quezon) and the standard BR25 from Davao were grouped with those two. All K9 are 75.40% similar with clones from Davao and Kabacan obtaining highest similarity of 90.80% among their group. Similar to K9, all K1 clones are grouped together. They were observed to be similar by 61.80% with clones from Davao and Kabacan getting the highest similarity coefficient (0.87) among the group. Clones of other varieties such as K2, UF18, PBC123, ICS40 and S5 were also clustered within their varieties at coefficients 0.70, 0.75, 0.67, 0.79 and 0.83 respectively. Three of the four UIT 1 were clustered at 0.68. USMCH1, USMCH2, P7 and DR1 are grouped at 0.36, a value that is not as high with that of clones of a variety.

A cut was made at 0.13(A) and 0.66(B) to the phenogram including all the clones. Two clusters are formed for cut A, cluster I includes varieties that were mostly Trinitarios while cluster II are Criollos. At cut B, distinct groups according to varieties were formed. If we are then to use this as reference in authenticating clones, threshold value can be set to 0.66.

Some clones that did not group with their respective varieties may have been misidentified and/or mislabelled. These collections are UIT1 (Camarines Sur), BR25 (Palawan), BR25 (Camarines Sur). Criollo (Maragusan) and Criollo (Quezon) being genetically distant with the other Criollos might also have been due to misidentification but since there is no thorough study on the genetic aspect of Criollo population yet in the country, these two could've only possessed an inherent genetic variation brought by selection pressures.

SSR Characterization

Fifteen (15) SSRs designed for cacao characterized as polymorphic were assessed by identifying the number of bands observed, band sizes (bp), number of polymorphic alleles and PIC presented in Table 3.

No. of bands observed ranged from 3 to 9 and observed polymorphic alleles ranged from 3 to 9 as well. The range obtained were lower compared to that from the work of Junsiul (2016), with range of 16-23 bands observed. He made use of capillary electrophoresis instead of PAGE. The bands obtained in this study is lower but repeatable and sufficient in detecting genetic similarities. Rate of polymorphism was consistent to be 100% for all the markers.

Botstein et al. (1980) declared SSR PIC values above 0.5 to be highly polymorphic, from to 0.25 to 0.49 moderately polymorphic, and less than 0.25 lowly polymorphic. All of the primers characterized were highly polymorphic with values ranging from 0.6341 to 0.8702.

DNA Profile

A haplotype map (Figure 3) was constructed where alleles were presented accordingly with the loci that were observed. A total of 84 alleles were observed for 15 polymorphic primers designed for cacao fingerprinting. The haplotype map provided information on the number of alleles that was observed per locus, sizes of these alleles and their banding patterns. Since it is a visual presentation of the diverse banding patterns, the map will serve as a guide in variety authentication experiments using same set of markers.

CONCLUSION

Standard cacao microsatellites used in this study were sufficient in constructing the DNA profile of the cacao Philippine collection. The success in using these markers could be due to their neutral nature, high repeatability and high degree of polymorphism. Clones were distinct from each other, meaning there is a quantified degree of genetic difference between the accessions that can be useful for identification. Most of the clones of same type clustered with few exceptions. This could be due to common problem in mislabeling in repositories. Established DNA profile of the cacao standard varieties in this study will serve as a reference in varietal authentication. The knowledge of the genetic relationships between these varieties/clones will also accelerate the breeding and conservation efforts for this crop.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Agriculture-Bureau of Agricultural Research (DA-BAR) and Department of Agriculture-High Value Crops Development Program (DA-HVCDP) for funding the project, and Philippine Agriculture Resources Research Foundation, Inc. (PARRFI) acting as the implementing agency.

REFERENCES

Aikpokpodion, P., Motamayor J., Adetimirin, V., Adu-Ampomah, Y., Ingelbrecht, I., Eskes, A., Schnell, R., and Kolesnikove-Allen M. 2009. Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. Tree Genet Genomes 5:699–711.

Bartley, B.G.D. 2005. The genetic diversity of cacao and its utilization. CABI Publishing, Wallingford.

Bhat, K.V. 2001. Molecular Characterization of Germplasm. In: Dhillon BS, Vraprasad KS, Srinivasan K, Singh M, Archak S, Srivastava U and Sharma GD(eds). National Bureau of Plant Genetic Resources: A compendium of Achievements. National Bureau of Plant Genetic Resources, New Delhi, India. 266-182.

Bostein, D.R., White R.L., Skolnick M., and Davis, R.W. 1980. Construction of genetic linkage map in man using restriction fragment length polymorphisms. Amer. J. Human Gen. 32:314-331.

Booza, EJ., Irish, B., Meerow, A., Tondo, C., Rodriguez, Lopez, M., Gomez, J., Moore J.M., Zhang, D., Motamayor J.C. and Schnellet, R. 2013. Genetic diversity, conservation, and utilization of *Theobroma cacao* L.: genetic resources in the Dominican Republic. Genet Resour Crop Evol 60:605–619.

Brown, J.S., and Sautte, R.T. 2008. A composite linkage map from three crosses between commercial clones of cacao, *Theobroma cacao* L. Trop Plant Biol 1:120–130.

CacaoA Industry Development Association of Mindanao Inc. 2017. The Philippine 2020 Cacao Challenge. Retrieved from http://www.cidami.org/philippine-2020-challenge/

Doyle, J.J. aND Doyle, J.L. 1987. Isolation of plant DNA from fresh tissue. Focus 2(1)1:13-15.

Junsuil, l. 2016. Utilization of molecular markers to detect the authenticity of cocoa clones. International Journal of Agriculture, Forestry and Plantation 3:101-104.

Motilal, L., Zhang, D., Umaharan, P., Bocarra, M., Maschke, S., Sankar, A., and Meinhart, L. 2012. Elucidation of genetic identity and population structure of cacao germplasm within an international cacao genebank. Plant Genetic Resources, 10(3): 232-241.

Rai, M.K., Phulwaria, M., and Shekthwat, N.S. 2013. Transferability of simple sequence repeat (SSR) markers developed in guava (*Psidium guajava* L.) to four Myrtaceae species. Molecular Biol. Reps. 40:067-5071.

Saal, B., and WrinckleR, G. 1999. Development of simple sequence repeat marker in Rye (*Secale cereal* L.). Genome 42:964-972.

Saunders, J.A., Mischke, S., Leamy, E.A. and Hemeida, A.A. 2004. Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. Theoretical and Applied Genetics 110:41-47.

Westman, A.L., & Kresovich, S. 1999) Simple sequence repeat (SSR)-based marker variation in Brassica nigra genebank accessions and weed population. Euphytica 109(2):85-92.

Table 1. List of cacao varieties and clones profiled using single sequence repeat (SSR) markers.

	Variety	Collection Area		Variety	Collection Area
1	BR25	Davao	**28**	UIT1	Davao
2		Kabacan	**29**		Kabacan
3		Quezon	**30**		Palawan
4		Batangas	**31**		Camarines Sur
5		Palawan	**32**	ICS40	Davao
6		Zamboanga Norte	**33**		Kabacan
7		Camarines Sur	**34**		Palawan
8	K1	Davao	**35**	K9	Davao
9		Kabacan	**36**		Kabacan
10		Quezon	**37**		Quezon
11		Batangas	**38**	S5	Davao
12		Palawan	**39**		Kabacan
13		Camarines Norte	**40**	Criollo Red	Davao
14	K2	Davao	**41**	Criollo Green	Davao
15		Kabacan	**42**	Criollo 21	Kabacan
16		Quezon	**43**	Criollo 22	Kabacan
17		Batangas	**44**	Criollo	New Leyte
18		Palawan	**45**	Criollo	Maragusan
19	UF18	Davao	**46**	Criollo	Quezon
20		Kabacan	**47**	USMCH1	Kabacan
21		Quezon	**48**	USMCH2	Kabacan
22		Batangas	**49**	DR1	Kabacan
23		Camarines Sur	**50**	P7	Kabacan
24	PBC123	Davao			
25		Kabacan			
26		Batangas			
27		Palawan			

Table 2. List of international standard single sequence repeat (SSR) primers and their respective sequences, repeat motifs and optimized annealing temperatures used to characterize the cacao collection.

SSR Primer	Forward Sequence	Reverse Sequence	Repeat Motif	A°C[1]
MTcCIR1	GCAGGGCAGGCTCAGTGAAGCA	TGGGCAACCAGAAAACGAT	$(CT)_{14}$	51
MTcCIR6	TTCCCTCTAAACTACCCTAAAT	TAAAGCAAAGCAATCTAACATA	$(TG)_2(GA)_{13}$	50
MTcCIR7	ATGCGAATGACAAACTGGT	GCTTTCAGTCCTTTGCTT	$(GA)_{11}$	49
MTcCIR8	CTAGTTTCCCATTTACCA	TCCTCAGCATTTTCTTTC	$(TC)_5TT(TC)_{17}TTT(CT)_4$	46
MTcCIR11	TTTGGTGATTATTAGCAG	GATTCGATTTGATGTGAG	$(TC)_{13}$	49
MTcCIR12	TCTGACCCCAAACCTGTA	ATTCCAGTTAAAGCACAT	$(CATA)_4N_{18}(TG)_6$	46
MTcCIR15	CAGCCGCCTCTTGTTAG	TATTTGGGATTCTTGATG	$(TC)_{19}$	48
MTcCIR18	GATAGCTAAGGGGATTGAGGA	GGTAATTCAATCATTTGAGGATA	$(GA)_{12}(TC)_{19}$	51
MTcCIR22	ATTCTCGCAAAAACTTAG	GATGGAAGGAGTGTAAATAG	$(TC)_{12}N_{146}(CT)_{10}$	50
MTcCIR24	TTTGGGGTGATTTCTTCTGA	TCTGTCTCGTCTTTTGTGA	$(AG)_{13}$	48
MTcCIR26	GCATTCATCAATACATTC	GCACTCAAAGTTCATACTAC	$(TC)_9C(CT)_4TT(CT)_{11}$	49
MTcCIR33	TGGGTTGAAGATTTGGT	CAACAATGAAAATAGGCA	$(TG)_{13}$	51
MTcCIR37	CTGGGTGCTGATAGATAA	AATACCCTCCACACAAAT	$(GT)_{15}$	49
MTcCIR40	AATCCGACAGTCTAATC	CCTAGGCCAGAGAATTGA	$(AC)_{15}$	51
MTcCIR60	CGCTACTAACAAACATCAAA	AGAGCAACCATCACTAATCA	$(CT)_2(CA)_{20}$	50

[1]A^0C- Annealing temperature

Table 3. Observed number of bands, size range of amplicons, number of polymorphic alleles and computed polymorphism information content (PIC) values of 15 single sequence repeats (SSR) markers.

Primer	No. of bands	Size range (bp)	No. of polymorphic alleles	PIC
MTcCIR1	3	160-200	3	0.6491
MTcCIR6	5	160-290	5	0.7005
MTcCIR7	4	160-220	4	0.6835
MTcCIR8	6	210-350	6	0.7844
MTcCIR11	4	180-400	4	0.7485
MTcCIR12	9	190-310	9	0.8679
MTcCIR15	4	240-290	4	0.7424
MTcCIR18	6	240-420	6	0.7887
MTcCIR22	6	280-420	6	0.8249
MTcCIR24	4	185-280	4	0.7293
MTcCIR26	4	280-320	4	0.6341
MTcCIR33	9	260-530	9	0.8702
MTcCIR37	6	140-280	6	0.8154
MTcCIR40	9	200-400	9	0.8308
MTcCIR60	5	190-300	5	0.7557

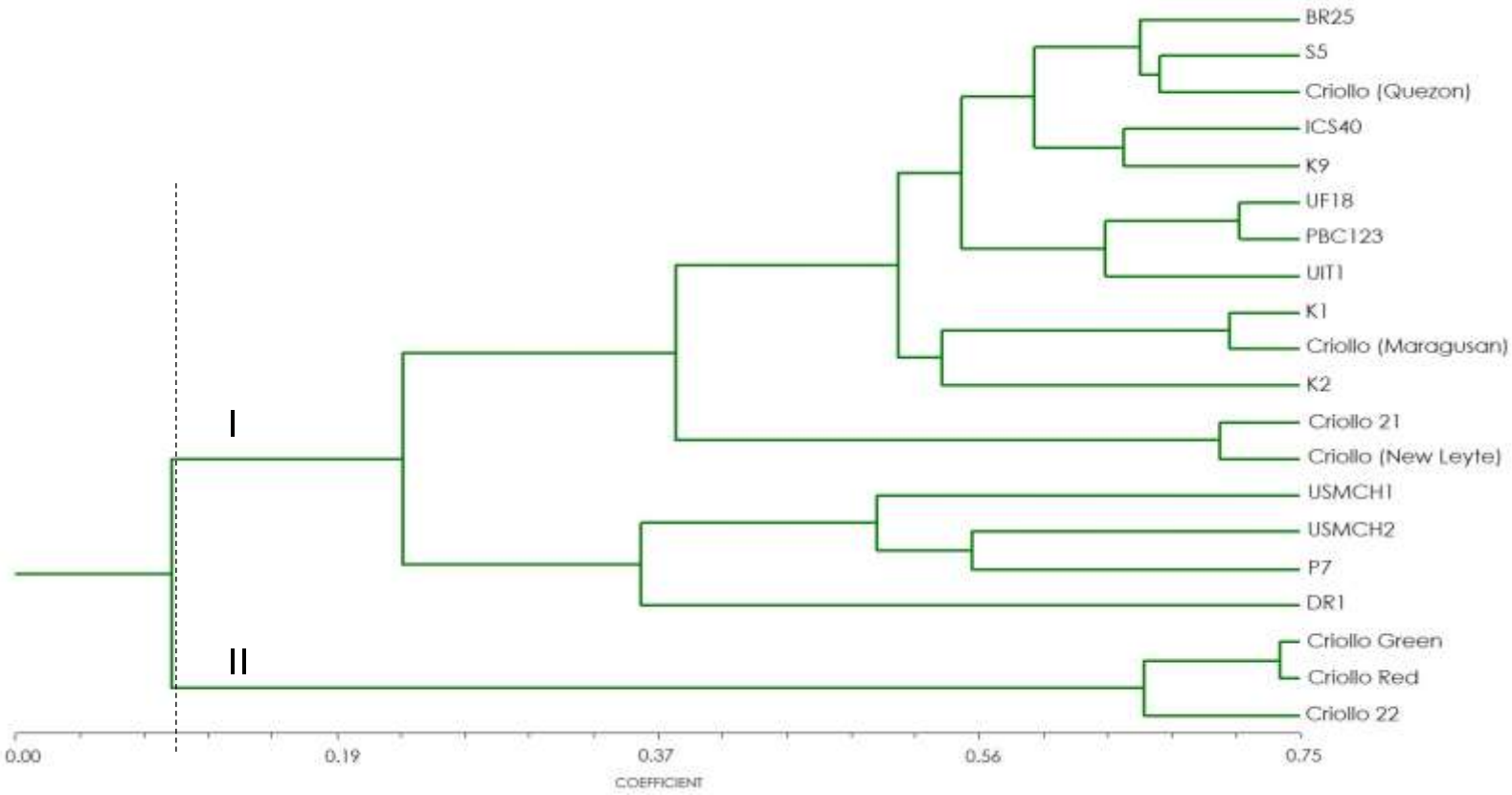

Figure 1. Dendrogram of 13 standard cacao varieties and 7 Criollo clones generated from 15 single sequence repeat (SSR) markers by unweighted pair-group mean avarage (UPGMA) cluster analysis.

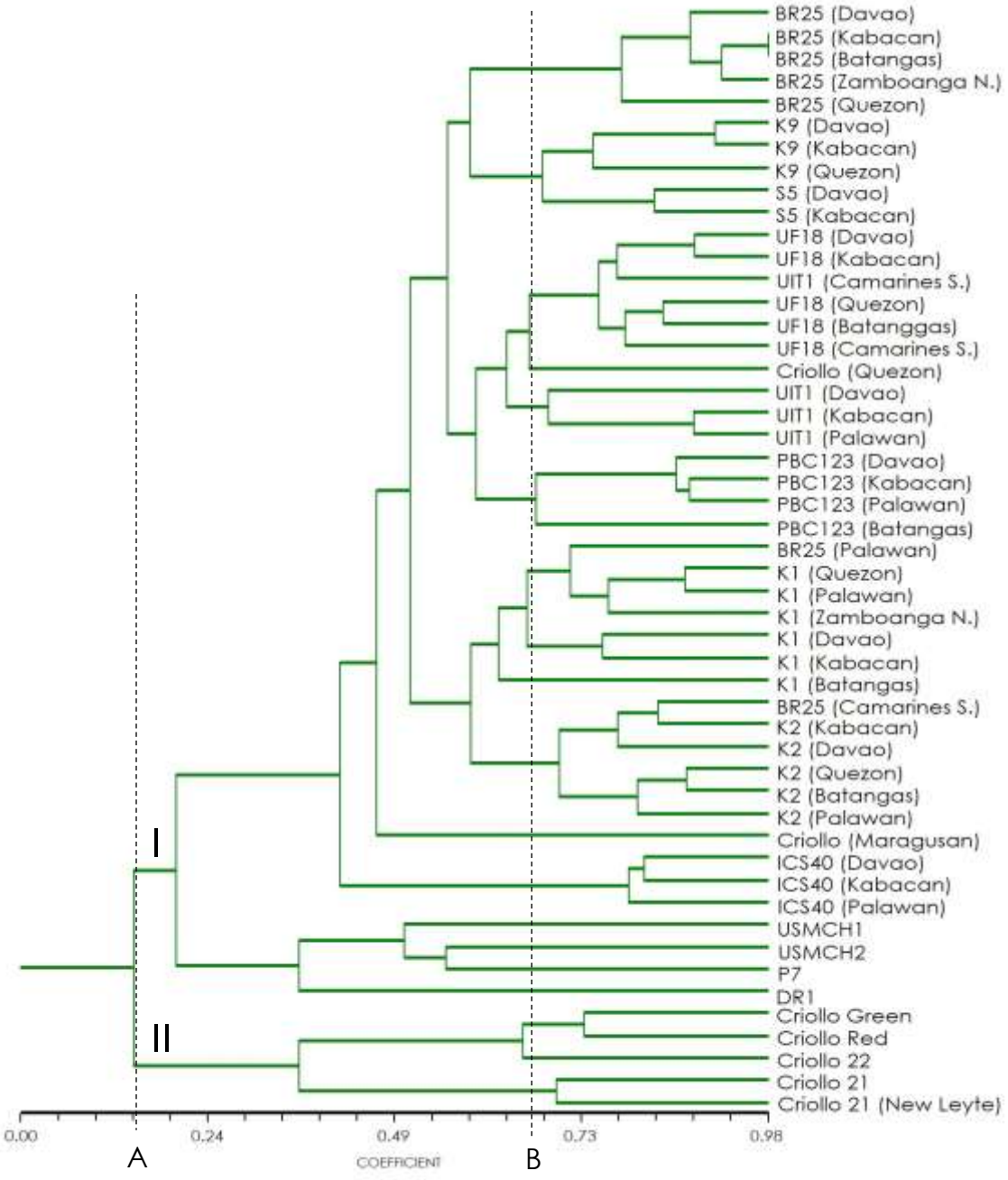

Figure 2. Dendrogram of 50 cacao clones generated from 15 single sequence repeat (SSR) markers using unweighted pair-group mean average (UPGMA) cluster analysis.

Figure 3. Haplotype map of cacao varieties in the Philippines using 15 single sequence repeat (SSR) markers.