Pak. J. Bot., 49(1): 133-142, 2017.

PHYLOGENETIC RELATIONSHIPS AMONG VIETNAMESE COCOA ACCESSIONS USING A NON-CODING REGION OF THE CHLOROPLAST DNA

LAM THI VIET HA[1,3,5*], DO TAN KHANG[2,3], HELENA EVERAERT[1,5], TRAN NHAN DUNG[3], PHAM HONG DUC PHUOC[4], HA THANH TOAN[3], KOEN DEWETTINCK[1] AND KATHY MESSENS[5]

[1] *Laboratory of Food Technology and Engineering, Department of Food Safety and Food Quality, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000 Gent, Belgium*
[2] *Graduate School of International Development and Cooperation, Hiroshima University, Kagamiyama, Higashi-Hiroshima, 739-8529, Japan*
[3] *Biotechnology Research and Development Institute, Cantho University, Campus II, 3/2 Street, Cantho city, Vietnam*
[4] *Cocoa Research Institute, Nonglam University, Kha Van Can street, Hochiminh, Vietnam*
[5] *Laboratory Agri Fing, Department of Applied Biosciences, Faculty of Bioscience Engineering, Ghent University, Valentin Vaerwyckweg 1, 9000 Gent, Belgium*
[*] *Corresponding author's email: vietha.lamthi@ugent.be; Ph: 0032-9 264 62 65; Fax: 0032-9 264 62 22*

Abstract

Cocoa cultivation has increased in tropical areas around the world, including Vietnam, due to the high demand of cocoa beans for chocolate production. The genetic diversity of cocoa genotypes is recognized to be complex, however, their phylogenetic relationships need to be clarified. The present study aimed to classify the cocoa genotypes, that are imported and cultivated in Vietnam, based on a chloroplast DNA region. Sixty-three Vietnamese Cocoa accessions were collected from different regions in Southern Vietnam. Their phylogenetic relationships were identified using the universal primers c-B49317 and d-A49855 from the chloroplast DNA region. The sequences were situated in the *trnL* intron genes which are identify the closest terrestrial plant species of the chloroplast genome. DNA sequences were determined and subjected to an analysis of the phylogenetic relationship using the maximum evolution method. The genetic analysis showed clustering of 63 cocoa accessions in three groups: the domestically cultivated Trinitario group, the Indigenous cultivars, and the cultivations from Peru. The analyzed sequencing data also illustrated that the TD accessions and CT accessions were related genetically closed. Based on those results the genetic relation between PA and NA accessions was established as the hybrid origins of the TD and CT accessions. Some foreign accessions, including UIT, SCA and IMC accessions were confirmed of their genetic relationship. The present study is the first report of phylogenetic relationships of Vietnamese cocoa collections. The cocoa program in Vietnam has been in development for thirty years.

Key words: Cocoa, Chloroplast, Non-coding region, Phylogeny, *Theobroma cacao* L.

Introduction

Cocoa (*Theobroma cacao* L.) has recently been cultivated in some areas of Vietnam and has shown promise as a crop for further economic development. In fact, cocoa trees are well-adapted to Southern Vietnam due to its suitable climate and soil conditions, especially in the Highland areas and the Mekong Delta provinces (Phuoc, 2009). Cocoa is pollinated by a species of the genus *Forcipomyia*, therefore it is believed that cocoa has a high chance of outcrossing (Yang *et al*., 2013). This advantage has produced a number of cocoa phenotypes with good characteristics through breeding. However, the diversity of hybrids also causes some difficulties in selecting parental genotypes. Thus far, cultivated cocoas have been categorized in three main groups. The first group is Criollo, which has a nicely flavored bean, but is susceptible to diseases. The second group is Forastero, which has a high yield and is highly tolerant to diseases, although the bean aroma is weak. The third group is Trinitario, which is the hybrid group of both Criollo and Forastero (Cuatrecasae, 1964; Laurent *et al*., 1994; Sounigo *et al.*, 2005; Sereno *et al.*, 2006; Bekele *et al.*, 2006; Wood & Lass, 2008; Motamayor *et al.*, 2008; Shri *et al*., 2009).

In the previous studies, there was limited information about the genetic structure of natural *T. cacao* populations. The pre-existing research was based on non-natural populations, which were represented either by the morphology of their geographic groups – their country of origin (Lanaud *et al.*, 1999; Ronning & Schnell, 1994; Bekele & Bekele, 1996; Whitkus *et al.*, 1998; Lachenaud & Oliver, 2005; Marcano *et al*., 2009). The authors of these studies lacked the knowledge of the origins and the relationships of the genotypes under molecular analysis. Throughout the past decade, modern molecular techniques, such as restriction fragment length polymorphic DNA, random amplified polymorphic DNA, single nucleotide polymorphism (SNP) markers, and microsatellite markers have been applied to determine the genetic diversity of cocoa genotypes. A summary of the results obtained from former studies using molecular markers is described by N'goran *et al.* (1994), Motamayor *et al.* (2002), Pugh *et al.* (2004), Iwaro *et al.* (2003), Sereno *et al.* (2006), Bekele *et al.* (2006), Lachenaud *et al.* (2008), Smulders *et al.* (2010), Aikpokpodion (2010), Maharaj *et al.* (2011), Motilal *et al.* (2012), Ji *et al.* (2013), and Yang *et al.* (2013). These authors have classified cocoa populations into a number of groups, such as the Criollo and Forastero groups (N'goran *et al.*, 2000); the Criollo, Amazonia region and Pentagona populations (Bartley, 2005); the Indigenous and Trinitario cultivations (Sounigo *et al.*, 2005); the lower and upper Amazon regions (Sereno *et al.*, 2006); the upper Amazon Forastero accessions (Zhang *et al.*, 2009); the two groups of Amelonado/Trinitario ancestry and the other of Nanay/Parinari ancestry (Aikpokpodion, 2010); the four subclusters OB of the Refractario accessions (Motilal *et al.*, 2012); the five genetic groups include ancient Criollo, Amelonado, Trinitario (including Nicaragua Trinitario and Honduras Trinitario) and Upper Amazon Forastero (Ji *et al.*, 2013); the Criollo, Upper Amazon Forastero, and Lower Amazon Forastero varietal groups (Yang *et al.*, 2013).

Chloroplast DNA (cpDNA) is a source of original DNA which can be used for phylogenetic research and population diversity. At a low taxonomic levels, the analysis of consensus sequences of amplified non-coding chloroplast DNA separation in much conserved regions is also highly useful (Petit *et al.*, 1998, Lanner *et al.*, 1998; Petit & Vendramin, 2007; Geleta *et al.*, 2010; De Andrés *et al.*, 2012; Kane *et al.*, 2012; Yang *et al.*, 2013). The *trn*L intron regions were already used to establish the phylogenetic relationships in Carpha (Zhang *et al.*, 2004) and *Psidium guajava* L. (Chen *et al.*, 2007). Moreover, Mulatu determined the phylogenetic and taxonomic structure of the Guizotia genus based on *trn*L region (Geleta *et al.*, 2010). With these molecular markers it is possible to clarify the discrimination of cultivars easier, faster and less expensively (Chen *et al.*, 2007).

Due to significant genetic diversity, this study focused on (1) clarifying the phylogenetic relations and (2) the gene constitution of the 63 cocoa accessions currently cultivated in Southern Vietnam by the *trn*L intron region. To our knowledge there is no published research about the genetic diversity of Vietnamese cocoa. Consequently, this result resolves the deficiency of information on original cocoa varieties. With the increased understanding of the genetic origin, there is higher potential for selective breeding activities and the development of cocoa cultivars in Vietnam. The resultant information is also potentially useful for increasing production yields, developing seed propagated cultivars, and genetic conservation.

Materials and Methods

Cocoa accessions collection: The leaves of sixty three cocoa accessions were sampled from large regions in the south of Vietnam, including Dak Lak province (18 accessions), Dong Nai province (28 accessions), Ben Tre province (9 accessions), Nong Lam University-Ho Chi Minh city (4 accessions), and Can Tho city (3 accessions) (Fig. 1). Detailed information of original cocoa population groups from Pound (1938), N'goran *et al.* (1994), Sounigo *et al.* (2005), Bartley (2005), Phuoc (2009), and Ha *et al.* (2015a,b) is presented in Tables 1 and 2 presents the overview of analyzed cocoa collection by their name code, institute code, and sampling sites.

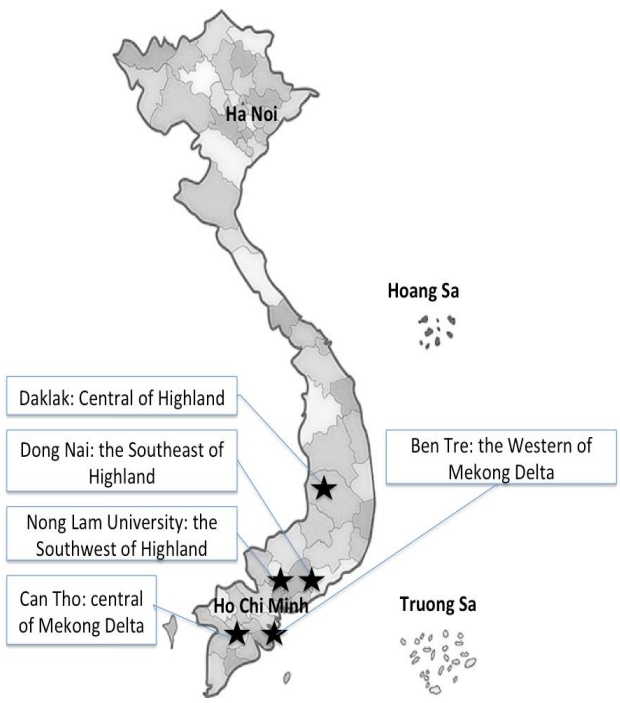

Fig. 1. The location of cocoa plantations in the South of Vietnam where the 63 cocoa varieties were collected for examination. Collected cocoa regions have been marked with black stars.

Table 1. Detail information of *Vietnamese cocoa* collection by accession name, origin name, country or origin, characteristics of fruit, pods, flowers, seeds and cultivation types

Accession name	Origin name	Country of origin	Characterics of fruit, shape of pod, flower, seed	Type of cultivation	References
SCA	Scavina	Ecuador	Elongated shape dark green fruit,	Indigenous	Bartley (2005), N'goran (1994)
MO	Morona	Peru	Short oval fruit	Indigenous	Bartley (2005), Pound (1938)
PA	Parinari or Lagarta	Peru	Elongated shape, dark green fruit	Indigenous	Bartley (2005) Sounigo (2005), Pound (1938)
IMC	Iquitos Mixed	Peru	Oval shape (Calabacillo), high ovule numbers	Indigenous	Sounigo (2005), Bartley (2005) Pound (1938)
ICS	Imperial College Selection	Trinidad	Pentagonum fruit, broad and shorter in shape	Cultivated	Bartley (2005), Sounigo (2005) Pound (1938)
LCTEEN	London Cocoa trade Amazon project	Ecuador	Elongated shape, dark green fruit	Indigenous	Sounigo (2005), N'goran (1994)
AMAZ	Rio Amazonas	Peru	Elongated shape, dark green fruit	Indigenous	Sounigo (2005),
NA	Nanay	Peru	Elongated shape, small seeds with dark purple cotyledon	Indigenous	Bartley (2005), Sounigo (2005), Pound (1938)
MA		Careiro Island	Elongated shape, high ovule numbers		Bartley (2005)
Pound	Pound	Peru	Elongated shape, dark green fruit	Indigenous	Pound (1938), Bartley (2005)
IFC		Côte d'Ivoire	Criollo shape, dark green fruit	Indigenous	N'goran (1994), Bartley (2005)
SIAL		Brazil	Criollo shape, dark green fruit	Indigenous	N'Goran (1994), Bartley (2005)
TD1	PA35 x NA32	Malaysia	Criollo shape (Angoleta), rough, oval seed with dark purple colytedon	Cultivated	Phuoc (2009), Ha (2015)
TD9	NA31 x PA15	Malaysia	Elongated shape (Amelonado), rough, dark green fruit, oval seed with dark purple colytedon	Cultivated	Phuoc (2009), Ha (2015)
TD12	(PA76 x SCA20) x (UIT1 x SCA6)	Malaysia	Criollo shape (Angoleta-long shape), rough, dark green fruit, oval seed with dark purple colytedon	Cultivated	Phuoc (2009), Ha (2015)
TD13	UIT1 x NA33	Malaysia	Elongated shape (Amelonado), dark green fruit, rough, oval seed with dark purple colytedon	Cultivated	Phuoc (2009), Ha (2015)
TD14	PA173 x SCA9	Malaysia	Criollo shape, dark green fruit, rough, oval seed with dark purple colytedon	Cultivated	Phuoc (2009), Ha (2015)

Table 2. Summary of the sixty three analyzed cocoa accessions.

Collection code	Institute accession code	Genebank accession number	Imported country	Sampling location
DK1	CT3	KR864819	Columbia	EaKar-Daklak province (Highland)
DK2	CT5	KR864820	Columbia	EaKar-Daklak province (Highland)
DK3	CT6	KR864758	Columbia	EaKar-Daklak province (Highland)
DK4	CT7	KR864815	Columbia	EaKar-Daklak province (Highland)
CT1	CT8	KR864816	Columbia	Phongdien-Cantho (Mekong Delta)
CT2	CT9	KR864806	Columbia	Phongdien-Cantho province (Mekong Delta)
CT3	CT21	KR864814	Columbia	Phongdien-Cantho province (Mekong Delta)
BT1	TD1	KR864779	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT2	TD2	KR864780	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT3	TD3	KR864781	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT4	TD5	KR864782	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT5	TD6	KR864783	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT6	TD7	KR864784	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT7	TD8	KR864785	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT8	TD9	KR864786	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT9	TD10	KR864787	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT10	TD11	KR864788	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT11	TD12	KR864789	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT12	TD13	KR864790	Malaysia	Chauthanh-Bentre province (Mekong Delta)
BT13	TD14	KR864791	Malaysia	Chauthanh-Bentre province (Mekong Delta)
DN1	TD15	KR864792	Vietnam	Trangbom-Dongnai province (Highland)
DN2	LCTEEN37/A	KR864759	Ecuador	Trangbom-Dongnai province (Highland)
DN3	LCTEEN62/S	KR864760	Ecuador	Trangbom-Dongnai province (Highland)
DN4	POUND16/B	KR864775	Peru	Trangbom-Dongnai province (Highland)
DN5	POUND16/A	KR864774	Peru	Trangbom-Dongnai province (Highland)
DN6	MAN15/6	KR864762	Trinidad	Trangbom-Dongnai province (Highland)
DN7	IMC105	KR864809	Trinidad	Trangbom-Dongnai province (Highland)
DN8	IMC67	KR864810	Trinidad	Trangbom-Dongnai province (Highland)
DN9	IFC5	KR864807	Amazon	Trangbom-Dongnai province (Highland)
DN10	UIT1	KR864805	Amazon	Trangbom-Dongnai province (Highland)
DN11	PA88	KR864767	Peru	Trangbom-Dongnai province (Highland)
DN12	IMC53	KR864808	Peru	Trangbom-Dongnai province (Highland)
DN13	PA137	KR864771	Peru	Trangbom-Dongnai province (Highland)
DN14	PA70	KR864768	Peru	Trangbom-Dongnai province (Highland)
DN15	PA120	KR864769	Peru	Trangbom-Dongnai province (Highland)
DN16	ICS1	KR864813	Trinidad	Trangbom-Dongnai province (Highland)
DN17	ICS43	KR864811	Trinidad	Trangbom-Dongnai province (Highland)
DN18	APA4	KR864818	Amazon	Trangbom-Dongnai province (Highland)
DN19	AMAZ1515	KR864817	Ecuador	Trangbom-Dongnai province (Highland)
DN20	SCA6	KR864776	Peru	Trangbom-Dongnai province (Highland)
DN21	SCA9	KR864777	Peru	Trangbom-Dongnai province (Highland)
DN22	PA169	KR864773	Peru	Trangbom-Dongnai province (Highland)
DN23	NA33	KR864765	Peru	Trangbom-Dongnai province (Highland)
DN24	MA12	KR864761	Ecuador	Trangbom-Dongnai province (Highland)
DN25	MO81	KR864763	Peru	Trangbom-Dongnai province (Highland)
DN26	TD32	KR864798	Vietnam	Trangbom-Dongnai province (Highland)
DN27	TD60	KR864802	Vietnam	Trangbom-Dongnai province (Highland)
DN28	TD57	KR864801	Vietnam	Trangbom-Dongnai province (Highland)
DK5	TD17	KR864793	Vietnam	EaKar-Daklak province (Highland)
DK6	TD18	KR864794	Vietnam	EaKar-Daklak province (Highland)
DK7	PA127	KR864770	Peru	EaKar-Daklak province (Highland)
DK8	TD24	KR864796	Vietnam	EaKar-Daklak province (Highland)
DK9	NA149	KR864766	Peru	EaKar-Daklak province (Highland)
DK10	TD77	KR864804	Vietnam	EaKar-Daklak province (Highland)
DK11	TD42	KR864799	Vietnam	EaKar-Daklak province (Highland)
DK12	TD31	KR864797	Vietnam	EaKar-Daklak province (Highland)
DK13	TD52	KR864800	Vietnam	EaKar-Daklak province (Highland)
DK14	TD70	KR864803	Vietnam	EaKar-Daklak province (Highland)
DK15	TD20	KR864795	Vietnam	EaKar-Daklak province (Highland)
DK16	SIAL339	KR864778	Amazon	EaKar- Daklak province (Highland)
DK17	NA32	KR864764	Amazon	EaKar-Daklak province (Highland)
DK18	EET376	KR864812	Amazon	EaKar-Daklak province (Highland)
DN14	PA156	KR864772	Amazon	Trangbom-Dongnai province (Highland)

DNA isolation: Cocoa leaves were collected from cocoa plantations in Southern Vietnam. The materials were packed into wet paper bags and put into a cold box before DNA extraction.

Genomic DNA was isolated by following the CTAB-SDS procedure (Ha *et al.*, 2015a) with some minor modifications: After cleaning the sample (50 mg) with ethanol 70%, the leaves were cut into small pieces and ground using the Retsch mixer mill model MM 200 with liquid Nitrogen. The milled sample was mixed with 1 mL of extraction buffer in a tube. 50 µL SDS (sodium dodecyl sulfate) 10% and 0.7 µL β-mercaptoethanol were added to the mixture. After incubation at 65ºC for 30 min, the suspension was centrifuged at 13,000*g* for 10 min. The upper phase was mixed with an equal volume isopropanol and the mixture was incubated at -20ºC for 2 hours. After centrifugation at 13,000*g* for 10 min, the supernatant was discarded and the pellet was resuspended in 400 µL 0.1X TE (Tris-EDTA) buffer and 400 µL CTAB. 10 µL RNase was added into the mixture and then incubated at 65ºC for 15 min. An amount of 800 µL chloroform: isoamylalcohol (24:1) was added to the lysate. After centrifugation for 5 min at 13,000*g*, the supernatant was mixed with 1.4 mL of ethanol solution (95% v/v) at -20ºC to precipitate the DNA. After incubation for 15 min at room temperature, the mixture was centrifuged at 13,000*g* for 10 min. The pellet was then washed twice with 700 µL ethanol 70% v/v. Afterwards, the pellet was dried at room temperature for 30min and the DNA pellet was then dissolved in 200 µL of Tris-EDTA buffer.

PCR amplification and sequencing of the non-coding region of the chloroplast genome

PCR amplification: The DNA barcodes were amplified by *trn*L intron primers: c-B49317 (5'-CGA AAT CGG TAG ACG CTA CG-3') and d-A49855 (5'-GGG GAT AGA GGG ACT TGA AC-3') which were used to detect a 500–600 bp fragment – a non-coding region of the chloroplast genome (Taberlet *et al.*, 1991). The total PCR reaction volume was 50 µL, containing 1 µL of extracted DNA product and 49 µL of master mix. The master mix consisted of 32.75 µL distilled water, 10 µL of 5X GoTaq PCR buffer (Fermentas), 2 µL of each primer (10 µM), 2 µL dNTPs solution (5 mM) and 0.25 µL GoTaq DNA polymerase (5 U). The negative control consisted of 2 µL of bi-distilled water.

The PCR amplifications were performed in a Gene Amp PCR System 9700 (Applied Biosystems, USA) using the following program for primer pair c/d: initial denaturation at 95ºC for 3 min, 35 amplification cycles of denaturizing at 95ºC for 20 s, primer annealing at 54ºC for 40 s, and extension at 72ºC for 30 s followed by a final extension step of 3 min at 72ºC.

Sequencing PCR products: PCR amplifications were purified by GE Healthcare. Purified fragments were directly sequenced with PCR primers using the ABI prism BigDyeTM Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) on an automated sequencer (ABI prism 3130xl, Applied Biosystems, USA).

Phylogenetic analyses: Sequences of 63 isolated accessions were analysed using BioEdit software version 7.1.11. All DNA sequences were compared to NCBI database for similarity based on the BLASTN program. The evolutionary history was inferred using the Minimum Evolution method (Rzhetsky and Nei, 1992). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al*., 2004). The ME tree was established using the Close-Neighbor-Interchange (CNI) algorithm (Nei and Kumar, 2000) at a search level of 1. The Neighbor-joining algorithm (Saitou and Nei, 1987) was used to generate the initial tree. Evolutionary analyses were conducted in MEGA6 (Tamura *et al*., 2013).

Results

The amplification of the *tnr*L (UAA) intron in cocoa yielded a size of 598 bp when using primers plant c and d. This result conformed to earlier results in tobacco (577 bp), rice (614 bp) and *Marchantia* (389 bp) (Taberlet *et al*., 1991). The sequence of the *tnr*L (UAA) intron was compared to the NCBI database, and the result showed that all 63 sequences had a 99% similarity to the chloroplast gene in *Theobroma cacao* L. (HQ244500.2).

The optimal tree with the sum of branch length = 0.04077851 is shown in Fig. 2. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The analysis involved 63 nucleotide sequences. The 1st + 2nd + 3rd + Noncoding codon positions were included. All positions containing gaps and missing data were eliminated. There were a total of 557 positions in the final dataset.

The result of parent-offspring design is mostly compatible with established clustering analysis (Fig. 3). The bootstrap test showed that all accessions had a strong relationship in the sequences of the *trn*L sequence 100% except for the two taxa PA156 and EET376. PA156 was used as a parental genotype for breeding because of its high resistance to Black Pod disease.

The genetic relationships among 63 Vietnamese cocoa accessions is illustrated by the phylogenetic tree in Fig. 2. As can be observed in the phylogenetic tree, the cocoa taxa can be separated into three groups: Domestic Trinitario Cultivars (38 accessions), Indigenous Cultivars (20 accessions), and Cultivars from Peru (5 accessions) (Fig. 3).

Discussion

In the study sequence data based on chloroplast region c and d of 63 Vietnamese cocoa accessions were illustrated, and their genetic evolution examined, using maximum parsimony analysis. The non-coding region of the chloroplast genome allows us to study genetic relationships. The study's results demonstrated the genetic relationships at a low taxonomic level. In addition, previous authors had successfully used this non coding region to establish phylogenetic relationships of plants (Zhang *et al.*, 2004; *Chen et al.*, 2007; Geleta *et al.*, 2010). In contrast to current studies the phylogenetic analysis (Zhang *et al.*, 2009; Smulders *et al.*, 2010; Motilal *et al.*, 2012) applied microsatellite markers. Our results also demonstrated the genetic relationships at a low taxonomic level.

The division of the Vietnamese cocoa groups is cleared by three different clusters (Fig. 3), as shown by chloroplast region c and d of the DNA analysed sequencing.

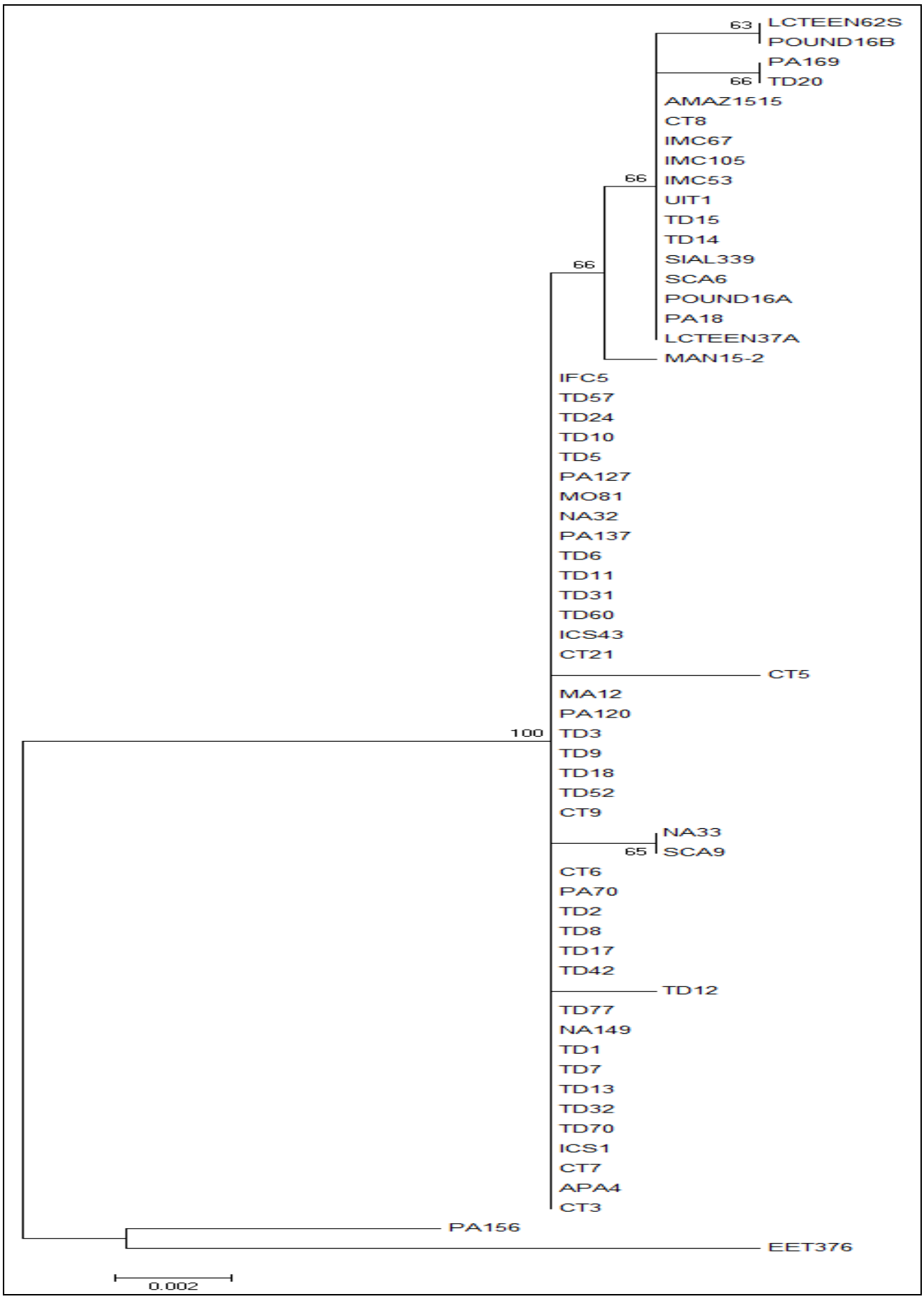

Fig. 2. Evolutionary relationships of 63 cocoa taxa with bootstrap values.

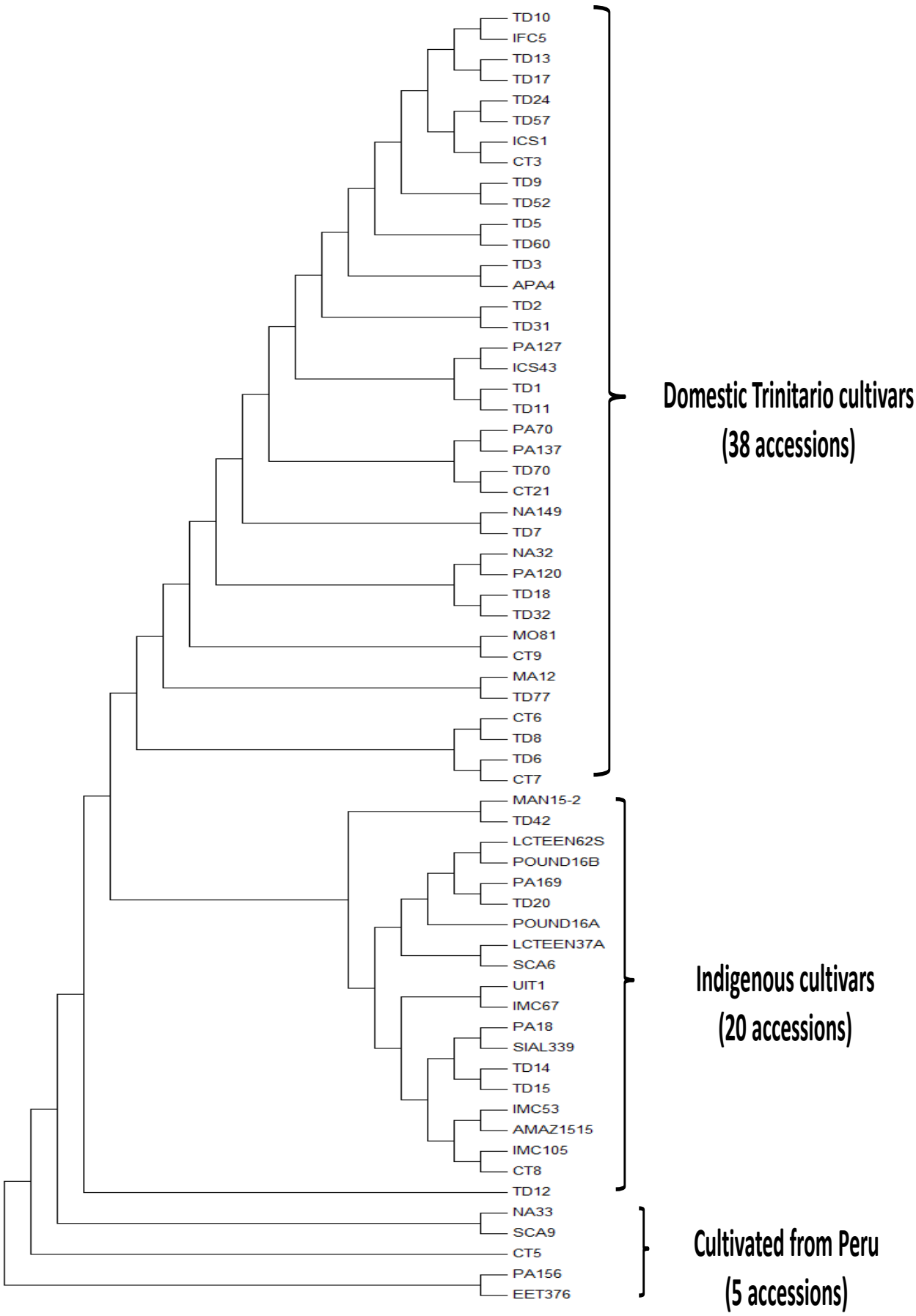

Fig. 3. Clustering of analyzed cocoa accessions based on Neighbor-Joining method.

The domestic trinitario cultivars group: The samples in this study were obtained from four different sources: the Amazon region of South America, Columbia, Malaysia, and Vietnam. There are 20 cocoa cultivars that are being grown in the Mekong-Delta area (TD1, TD2, TD3, TD5, TD6, TD7, TD8, TD9, TD10, TD11, TD12, TD13, TD14, CT3, CT5, CT6, CT7, CT8, CT9, and CT21), while the others are mostly located in highland regions of Vietnam. The hybrid and grafting of high quality parentage clones has biologically improved the original properties of plants in different ways. As a result, selective breeding to obtain certain characteristics comes with the good features of the original parents. Regarding the genetic diversity of cocoa, Sounigo analyzed the dissimilarity of 459 accessions collected in the International Cocoa Genebank in Trinidad using RAPD and isozyme methods (Sounigo *et al*., 2005). Both techniques distinguished three groups of cocoa by their genetic origin: the indigenous tree, the cultivated Trinitario and the cultivated trees from Ecuador. Their results also revealed that about 70% of diversity accessions were different from the actual populations. The results were obtained from various studies which lead to different conclusions on the origin of the cocoa (Sounigo *et al*., 2005; Bartley, 2005; Sereno *et al*., 2006; Zhang *et al*., 2009; Motilal *et al*., 2012; Yang *et al*., 2013). According to earlier studies (Kyndt *et al*., 2010, Smulders *et al*., 2010, Gelata *et al*., 2010) with 70% bootstraps, there is insufficient information for the identification of genetic relationships.

According to Phuoc (2009), some TD accessions have their origin in PA, NA, and SCA varieties which were imported from Malaysia (Table 1). The TD1 clone, for example, is a progeny of the PA35 and NA32 varieties. Furthermore, NA32 belongs to the Forastero group (Lachenaud & Oliver, 2005) and the morphological characteristics of the TD groups, e.g. Amelonado or Agoleta shape, belong to the Forastero group or Trinitario group (Uyen & Sum, 1996; Sounigo *et al*., 2003; Phuoc, 2009; Ha *et al*., 2015b). In contrast, the previous studies demonstrated that the PA and NA species showed some features of the Criollo group, including the dark green fruit feature and Cuademore shape (Bartley, 2005). Therefore, our analysed data confirmed a closed genetic relationship among the TD groups. There is no doubt that the parents of these accessions belong to the traditional Trinitario hybrid varieties. The result is similar to the microsatellite-based analysis which proved that the NA and PA clones have their ancestry from Upper Amazon's (Aikpokpodion *et al*., 2010).

Cocoa varieties were initially imported from cocoa countries in the Amazon area, Africa, and Asia with inaccurate origin data. Among them, ten TD cocoa cultivars (TD1, TD3, TD5, TD6, TD8, TD9, TD10, TD11, TD12, and TD14) are recognized as part of the National Cocoa Tree by the Ministry of Agriculture in Vietnam (Phuoc, 2009). Moreover, these cocoa cultivars are crossbred with domestic cultivars by the Nonglam University Cocoa Research Institute and the farmers of Cocoa Plantations in the South of Vietnam without variety origin information. Nong Lam Cocoa Institute - Vietnam will introduce more TD clones into the National Cocoa Clones registry over the next ten years (TD13, TD17, TD20, SIAL339, LCTEEN62/S, POUND16A, IMC105, PA8, PA70, AMAZ1515, SCA9, NA33, and PA137). The analysis of the genetic relationships revealed two groups among these TD clones: Domestic Trinitario and Indigenous Cultivars. Similar results were obtained by Ha (Ha *et al*., 2015a,b) using only morphological characteristics. The author proved that most of the cocoa clones in this group showed the Amelonado and Agoleta shapes (Forastero or Trinitario's characteristics). Therefore, these results will provide beneficial data for Vietnamese cocoa conservation and breeding activities.

It is important to mention that for almost all the domestic Trinitario cultivars in this group, especially TD and CT, farmers have grafted indigenous cultivars with imported cultivars without having any information on their origins. Lacking this information, farmers are evaluating the yield of different cultivars by planting them in southern regions. Consequently, the TD and CT cocoa accessions constituted the low variability in genetic relationships when farmers engaged in a mass production. For example, TD3 produced the fine flavor beans, but its fruits (pods) were small (Ha *et al*., 2015a,b). In contrast, TD9 or TD10 produced very big fruits (Ha *et al*., 2015a,b), so it would be useful to graft these accessions to create a new cultivar with good characteristics. In addition, these collected genotypes are absent in genetic resource centers and other institutions, where they can be multiplied for many regions. The possibility of growth of new cocoa hybrids in many regions in the South of Vietnam presents a great challenge. Although CT3 showed a high yield in Mekong Delta, it had a low production in Highland where TD accessions can be appropriately grown. Therefore, grafting between CT3 and TD9 could be applied to develop a new hybrid with high yield in Highland regions.

The CT domestic Trinitario accessions were grafted in Cantho University. Among them, CT3 is imported from Costa Rica while for other accessions there is a lack of knowledge regarding their genetic origins. The result showed that the CT and TD accessions proved to have a closed genetic relationship. Similar result were obtained by Dung *et al.* (2005), who verified that CT6, TD7 and TD14 showed a close relationship by applying microsatellite markers. In addition, the author also gave similar result with TD6, TD8, TD11 accesions.

The indigenous cultivars group: The result showed that the NA, SCA, and IMC accessions have parentage relationship based on the non-coding regions analysed. Our result was found to be compatible to the previous study who confirmed these clones had same parents from Upper Amazon Foratero when applying SNP markers to genetic diversity analysis (Ji *et al*., 2013). Although the bootstrap 66% is indicated for IMC group and SCA6, they belong to Indigenous cultivars group. Furthermore, Aikpokpodion (2010) indicated that PA and IMC clones had the closest genetic relationships. The two PA accessions and three IMC accessions also originated from the Indigenous cultivation group.

The Indigenous group of cocoa accessions was imported from the Amazon region, including Costa Rica and other South American countries (Uyen & Sum, 1996; Phuoc, 2009). These results were found to be compatible with previous results on the origins of these cocoa accessions (Sounigo *et al*., 2005; Sereno *et al*., 2006; Motamayor *et al*., 2008). In this group, CT5, TD14 and TD15 are domestic hybrid clones showing a genetic relation with others clones in bootstrap 62% value. The clustering of the molecular data confirmed that these three accessions were hybrids of these Amazon accessions. This group confirmed "Scavina"-PA accessions. Bartley (2005) already summarized the PA accessions which originate from the Amazona region. Most of the cocoa accesions in this group were confirmed as the Forastero type. For instance, it was found by Smulders that SCA6 and IMC67 belong to the Forastero type (Smulders *et al.*, 2010). In addition, the hybrid cultivars in Vietnam (CT5, TD14 and TD15) showed the Agoleta and Amelonado shape which belong to the Forastero group (Ha *et al.*, 2015a,b). Therefore, it was demonstrated that these Indigenous accessions have Forastero characteristics.

According to Phuoc (Phuoc, 2009), UIT means unidentified accession. The obtained data reveals a closed genetically relation between UIT1 and IMC67 accessions. As a result, UIT1 accession can be identified its origin by genetic data. This UIT1 accession is highly recommended for the further breeding program because of its Criollo character (Phuoc, 2009; Ha *et al.*, 2015a,b).

The cultivars from peru group: The cultivated populations of Peru (5 accessions NA33, SCA9, CT5, PA156, and EET376) clustered distinctly from the other groups because of their complex origin. The origins of these populations were confirmed as Peru and Costa Rica geographically (Phuoc, 2009). The CT5 clone was imported from Peru, and due to a lack of information on its origin, cocoa farmers labeled it as CT5. The previous authors demonstrated that NA, PA, and SCA have origins in the Amazon region (Sounigo *et al.*, 2005; Bartley, 2005). In addition, Sounigo classified the NA, SCA and PA varieties as belonging to the Indigenous group (Sounigo *et al.*, 2005). Consequently, these accessions showed the closed relationships based on the *tnr*L intron- data. This result distributed important information of genetic origin for the cocoa breeders. The NA33 accession is also introduced by Nong Lam Cocoa Institute-Vietnam which is supporting more hybrid clones into the National Cocoa Program over the next ten years.

Conclusion

The present study indicates that the CTAB-SDS procedure is a sufficient technique to extract DNA for analyzing genetic diversity in cocoa. This method clearly showed the high PCR amplification results. It is worth mentioning that the result proved based on chloroplast restriction site data we can access the genetic diversity, therefore sampling from the chloroplast region is an adequate technique to assess the genetic diversity in cacoa accessions.

In conclusion, the use of *trn*L intron region markers (c-B49317 and d-A49855) in the cpDNA has clarified the genetic relationships of Vietnamese cocoa populations, and our understanding of the genetic constitution of 63 cocoa accessions. The bootstrap test demonstrated that all accessions have a strong relationship in the sequences of the *trn*L sequence 100%. This study illustrates the genetic diversity of cocoa in Southern Vietnam. Three different groups have been deduced: Domestic Trinitario Cultivars (38 accessions), Indigenous Cultivars (20 accessions), and Cultivars from Peru (5 accessions). The present results provide a significant contribution to further breeding activities and to the conservation of Vietnamese cocoa accessions genetic.

Acknowledgement

We highly appreciate Vietnam's Ministry of Education and Training for their financial support. The authors thank colleagues and farmers of cocoa plantations in the Central Highlands region, the Mekong Delta area, and Nonglam University Cocoa Research Institute (Vietnam) for providing materials in this study.

References

Aikpokpodion, P.O. 2010. Variation in agro-morphological characteristics of cacao, *Theobroma cacao* L., in farmers' fields in Nigeria. *New Zealand J. of Crop and Horticultural Sci.*, 38(2): 157-170.

Bartley, B.G. 2005. The genetic diversity of cacao and its utilization. Cabi, pp. 98-310.

Bekele, F.L. and I.Bekele. 1996. A sampling of the phenetic diversity of cacao in the International Cocoa Gene Bank of Trinidad. *Crop Sci.*, 36(1): 57-64.

Bekele, F.L., I. Bekele, D.R. Butler and G.G. Bidaisee. 2006. Patterns of morphological variation in a sample of cacao (*Theobroma cacao* L.) germplasm from the International Cocoa Genebank, Trinidad. *Genet Resour and Crop Evol.*, 53(5): 933-948.

Chen, T., C.W.C. Ng and Y. Shyu. 2007. Molecular identification and analysis of *Psidium guajava* L. from indigenous tribes of Taiwan. *J. of Food and Drug Analys.*, 15(1):82.

Cuatrecasae, J. 1964. Cacao and its allies: a taxonomic revision of the genus *Theobroma. Contributions from the U.S National Herbarium*, 35: 379-614.

De Andrés, M.T., A. Benito, G. Perez‐Rivera, R. Ocete, M.A. Lopez, L. Gaforio and R. Arroyo-Garcia, 2012. Genetic diversity of wild grapevine populations in Spain and their genetic relationships with cultivated grapevines. *Mol. Ecol.*, 21(4): 800-816.

Dung N.T.P., P.H.D. Phuoc, B.M. Tri and J.C. Motamayor. 2005. Construction of 24 clones of *Theobroma cacao* L. Identification and analysis of genetic diversity using Microsatellites. Bachelor thesis. Nonglam Univ., Vietnam.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*., 39: 783-791.

Geleta, M., E. Bekele, K. Dagne, and T. Bryngelsson. 2010. Phylogenetics and taxonomic delimitation of the genus Guizotia (*Asteraceae*) based on sequences derived from various chloroplast DNA regions. *Plant syste. and Evol.*, 289(1-2): 77-89.

Ha, L.T.V, L. Vanlerberghe, H.T. Toan, K. Dewettinck and K. Messens. 2015a. Comparative Evaluation of Six Extraction Methods for DNA Quantification and PCR Detection in Cocoa and Cocoa-Derived Products. *Food Biotech*., 29(1): 1-19.

Ha, L.T.V, T.H. Phung, N.D. Tran, H.T. Toan, K. Dewettinck and K. Messens. 2015b. Characterisation of morphology and anatomy of the main cocoa varieties in the Southern Vietnam. *J. of Scientific Cantho Univ*,. 36(A): 47-56.

Iwaro, A.D., F.L. Bekele and D.R. Butler. 2003.. Evaluation and utilisation of cacao (*Theobroma cacao* L.) germplasm at the International Cocoa Genebank, Trinidad. *Euphytica*, 130(2): 207-221.

Ji, K., Zhang, D., L. A. Motilal, M. Boccara, , P. Lachenaud and L. W. Meinhardt. 2013. Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. *Gene. Resour. and Crop Evol.*, 60(2): 441-453.

Kane, N., S. Sveinsson, H. Dempewolf, J. Y. Yang, D. Zhang, J. M. Engels and Q. Cronk. 2012. Ultra-barcoding in cacao (Theobroma spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *Ameri. J. of Bot.*, 99(2): 320-329.

Kyndt, T., T. N. Dung, P. Goetghebeur, H. T. Toan and G. Gheysen. 2010. Analysis of ITS of the rDNA to infer phylogenetic relationships among Vietnamese Citrus accessions. *Genet. Resour. and Crop Evol.*, 57(2), 183-192.

Lachenaud, P. and D. Zhang. 2008. Genetic diversity and population structure in wild stands of cacao trees (*Theobroma cacao* L.) in French Guiana. *Anals. of Forest Sci.*, 65(3): 1.

Lachenaud, P. and G. Oliver. 2005. Variability and selection for morphological bean traits in wild cocoa trees (*Theobroma cacao* L.) from French Guiana. *Gene. Resour. and Crop Evol*., 52(3): 225-231.

Lanaud, C., A.M.Risterucci, I. Pieretti, M. Falque, A. Bouet and P.J.L. Lagoda. 1999. Isolation and characterization of microsatellites in *Theobroma cacao* L. *Mol. Ecol.*, 8(12): 2141-2143.

Lannér, C. 1998. Relationships of wild Brassica species with chromosome number 2 n= 18, based on comparison of the DNA sequence of the chloroplast intergenic region between trnL (UAA) and trnF (GAA). *Canadian J. of Bol.*, 76(2): 228-237.

Laurent, V., A.M. Risterucci and C. Lanaud. 1994. Genetic diversity in cocoa revealed by cDNA probes. *Theor. Appl. Genet.*, 88(2): 193-198.

Maharaj, K., P. Maharaj, F.L. Bekele, D. Ramnath, G.G. Bidaisee, I. Bekele and R. Sankar. 2011. Trinidad selected hybrids: An investigation of the phenotypic and agro-economic traits of 20 selected cacao cultivars. *Trop. Agric.*, 41(3216): 040175-11.

Marcano, M., S. Morales, M.T. Hoyer, B. Courtois, A.M. Risterucci, O. Fouet, C. Lanaud. 2009 .A genome wide admixture mapping study for yield factors and morphological traits in a cultivated cocoa (*Theobroma cacao* L.) population. *Tree Genet. & Geno.*, 5(2): 329-337.

Motamayor, J.C., A.M. Risterucci, P.A. Lopez, C.F. Ortiz, A. Moreno and C. Lanaud. 2002. Cacao domestication I: The origin of the cacao cultivated by the Mayas. *Heredity*, 89: 380-386.

Motamayor, J.C., P. Lachenaud, J.W.D.S. eMota, R.. Loor, D.N. Kuhn, J.S. Brown and R.J Schnell. 2008. Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PLoS One*, 3(10): e3311.

Motilal, L.A., D. Zhang, P. Umaharan, M. Boccara, S. Mischke, A. Sankar and L.W. Meinhardt. 2012. Elucidation of genetic identity and population structure of cacao germplasm within an international cacao genebank. *Plant Gene. Resour.*, 10(03): 232-241.

Nei, M. and S. Kumar. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

N'goran, J.A.K, V. Laurent, A.M. Risterucci and C. Lanaud. 2000. The genetic structure of cocoa populations (*Theobroma cacao.* L.) revealed by RFLP analysis. *Euphytica*, 115(2): 83-90.

N'goran, J.A.K., V. Laurent, A.M. Risterucci and C. Lanaud. 1994. Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity*, 73(6): 589-597.

Petit, R.J. and G.G. Vendramin. 2007. Plant phylogeography based on organelle genes: an introduction. In: (Eds.): Weiss, S. & N. Ferrand. *Phylogeography of southern European refugia*. *Springer Netherlands*, pp. 23-97.

Petit, R.J., B. Demesure and S. Dumolin. 1998. cpDNA and mtDNA primers in plants. In Molecular tools for screening biodiversity. *Springer Netherlands*, pp. 256-261.

Phuoc, P.H.D. 2009. Cocoa cultivated technique in Vietnam. Vietnam Agriculture Publisher, pp. 1-46.

Pound, F.J. 1938. Cacao and witchbroom disease (Marasmius perniciosus) of South America. With notes on other species of Theobroma. Report by Dr. FJ Pound on a visit to Ecuador, the Amazon Valley, and Colombia, April 1937-April 1938.

Pugh, T., O. Fouet, A.M. Risterucci, P. Brottier, M. Abouladze, C. Deletrez and C. Lanaud. 2004. A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. *Theor. and Appl. Gene.*, 108(6): 1151-1161.

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

Rzhetsky, A. and M. Nei. 1992. A simple method for estimating and testing minimum evolution trees. *Mol. Biol. and Evol.*, 9: 945-967.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. and Evol.*, 4: 406-425.

Sereno, M.L., P.S. Albuquerque, R.Vencovsky and A. Figueira. 2006. Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers. *Conse. Gene.*, 7(1): 13-24.

Shri, M.J. and P.M. Priyadarshan. 2009. Breeding Plantation Tree Crops. Springer, New York, USA, pp.589-591.

Smulders, M.J.M., D. Esselink, F. Amores, G. Ramos, D.A. Sukha, D.R. Butler, B.Vosman and LooEN van. 2010. Identification of cocoa (*Theobroma cacao* L.) varieties with different quality attributes and parentage analysis of their beans. *Ingenic Newsl*., 12: 1-13.

Sounigo, O., P. Lachenaud, P. Bastide, C. Cilas, J. N'goran and C. Lanaud. 2003. Assessment of the value of doubled haploids as progenitors in cocoa (*Theobroma cacao* L.) breeding. *J. of Appl. Gene.*, 44(3): 339-354.

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

Swofford, D.L. 2002. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sunderland, MA: Sinauer Associates.

Taberlet, P., L. Gielly, G. Pautou and J. Bouvet. 1991. Universal Primers for Amplification of 3 Noncoding Regions of Chloroplast DNA. *Plant Mol. Biol.*, 17: 1105-1109.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. and Evol.*, 30: 2725-2729.

Tamura, K., M. Nei and S. Kumar. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences (USA)*, 101: 11030-11035.

Uyen, N.V. and N.T Sum. 1996. Cocoa tree in the world and its potential in Vietnam. *Vietnam Agriculture Publisher*, pp. 8-30.

Whitkus, R., M. De la Cruz, L. Mota-Bravo and A. Gómez-Pompa. 1998. Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern Mexico. *Theo. and Appl. Gene.*, 96(5): 621-627.

Wood, G.A.R. and R.A. Lass. 2008. Cocoa. John Wiley & Sons, London, UK.

Yang, J.Y., M. Scascitelli, L.A. Motilal, S. Sveinsson, J.M. Engels, N.C. Kane and Q.C. Cronk. 2013. Complex origin of Trinitario-type Theobroma cacao (Malvaceae) from Trinidad and Tobago revealed using plastid genomics. *Tree Gene. & Geno.*, 9(3): 829-840.

Zhang, D., S. Mischke, E.S. Johnson, W. Phillips-Mora and L. Meinhardt. 2009. Molecular characterization of an international cacao collection using microsatellite markers. *Tree Gene. & Geno.*, 5: 1-10.

Zhang, X., A. Marchant, K.L. Wilson and J.J. Bruhl. 2004. Phylogenetic relationships of Carpha and its relatives (*Schoeneae, cyperaceae*) inferred from chloroplast trnL intron and *trn*L–*trn*F intergenic spacer sequences. *Mol. Phyl. And Evol.*, 31(2): 647-657.

(Received for publication 11 November 2015)