EVALUATION OF CLONAL UNIFORMITY IN CLASS ONE MALAYSIAN COMMERCIAL COCOA CLONES BASED ON SSR MARKERS

Lea Johnsiul
Centre for Cocoa Biotechnology Researches,
Malaysian Cocoa Board, Locked Bag 211, 88999 Kota Kinabalu, Sabah, Malaysia.
Email:lea@koko.gov.my

Azwan Awang
Faculty of Sustainable Agriculture,
Universiti Malaysia Sabah,
Sandakan Campus, Locked Bag No. 3, 90509 Sandakan, Sabah, Malaysia.

ABSTRACT

The Malaysian Cocoa Board has shortlisted 54 cocoa clones to be the recommended commercial clones distributed for farmers' planting. These 54 cocoa clones are divided into 4 classes, Class I, Class II, Class III and Class IV according to their adaptability to a wide range of Malaysia agro-climatic condition, good agronomic traits, tolerant to major pests and diseases and high butter fat content and good flavour. Propagation of cocoa plants is usually carried out vegetatively and therefore should be genetically uniform as genetic uniformity in cocoa clones is a very crucial information for germplasm conservation, in obtaining correct parental crosses in breeding program and plant materials generation for farmers' fields' distribution. The evaluation was done using microsatellite markers aimed to determine the genetic uniformity of five Class I Malaysian Commercial Cocoa clones using the International Cocoa Molecular Markers Standard. Comparison were done between samples taken from one cocoa germplasm collection in Kota Kinabalu, Sabah. Two samples were taken randomly to represent each of the clones. DNA extraction was done using Qiagen DNeasy Plant DNA kit and amplification was done using 14 SSR markers. The polymorphic information content (PIC) analysis was carried out and was very informative in identifying the genetic uniformity of the evaluated cocoa samples. Result showed that there are differences between the same clone name but obtained from the same location indicating there are off-type plants in the collection.

Key words: Genetic uniformity, microsatellite markers, SSR markers, Theobroma cacao L., Malaysian commercial cocoa clones

INTRODUCTION

Cocoa is a commercial crop of great importance worldwide, with the production of 4,739 thousand tonnes of beans in 2016/2017, forecasted to reach 4,799 thousand tonnes in 2018/2019, on 9.9 million hectares and affecting the livelihoods of 40–50 million people (ICCO, 2019).

Malaysia currently has 16,871 hectares cocoa trees, with about 6,760 hectares in Sabah, 6,819 hectares in Sarawak and the remaining in rest of the peninsula. Smallholders dominate the cocoa plantation with 14,776 hectares managed by 15,045 small holders throughout the country (Malaysian Cocoa Board, 2019). Malaysia is the second largest cocoa grinder in Asia and the eighth largest globally, forecasted to move up to the seventh in 2018/2019 (ICCO, 2019).

The forecasted increases in demand for cocoa beans in the recent years create demands for the increase in the national production. One of the approaches that can be applied is to improve the national production by ensuring only recommended cocoa clones were planted by the smallholders (ICCO, 2019). Another approach is production of new superior varieties through breeding activities that utilize genetic materials with superior traits such as high yield, big bean size, as well as resistance to biotic and abiotic stresses.

Due to the long breeding cycle, the choice of varieties used in the breeding program is also crucial to generate progenies that are superior varieties with superior traits. Cocoa breeding programme usually done in several steps, firstly collection of cocoa genetic materials and germplasm through collaboration with various germplasm collections in the world, secondly evaluation of genetic materials and thirdly genetic materials utilization through selection, hybridization, mutation and genetic engineering. Hybridization of two parental clones is predicted to produce progenies with combined characteristics from the parents. According to Dias and Kageyama (1997), cocoa exhibits strong heterosis for yield and yield-contributing characters and the success of plant hybridization is determined by the genetic distance between both parental materials. The farther the parental genetic distance, the higher the probability of heterosis to occur. In order to exploit the heterosis in cocoa, information on the parental clones' genetic uniformity must be accurate.

Inconsistency in performance of known clones when used as parents in field trials is partly due to incorrect genetic identity. The improvement of cocoa planting materials can greatly be facilitated by correctly identifying the clones within the existing cocoa collection (Young, 1994). A significant number of mislabelled clones have been reported in several cocoa germplasms around the world (Motilal and Butler, 2003, Turnbull *et al.*, 2004, Motilal *et al.*, 2012, Irish *et al.*, 2010, Boza *et al.*, 2014, DuVal *et al.*, 2017). Thus, efficient and effective strategy to identify these non-uniform (mislabelled and/or duplicate) trees in the cocoa

germplasm collections is needed for efficient and accurate management of genetic resources (Wickramasuriya and Dunwell, 2018).

The plant genetic uniformity can be evaluated based on morphology or molecular markers. However, the use of molecular markers gives some advantages in breeding programme as they can be used in plant selection at seedling stage and not affected by environmental effects. One of the markers that frequently used to analyze genetic uniformity is microsatellite markers or SSR (simple sequence repeats). Microsatellite markers is frequently used due to its relatively high number of allelic polymorphisms generated at each locus and across the chromosomes and can be analysed using PCR (Polymerase Chain Reaction) method (Parida *et al*., 2010). The use of microsatellite markers to analyse the genetic variability of cocoa had been done by Zhang *et al*. (2009), Irish *et al*. (2010) and Boza *et al*. (2014).

The Malaysian Cocoa Board Class I Commercial cocoa clones are superior clones generally suitable for planting throughout Malaysia with good yield potential (>2.5 tonne/year) and tolerant to major pest and diseases (Malaysian Cocoa Board, 2012). These Class I cocoa clones have potentials to be used in the hybridization programme to produce new planting materials.
The paper aims to evaluate the genetic uniformity of the Class I Malaysian Commercial Cocoa Clones taken from KKIP Cocoa Germplasm Collection in Kota Kinabalu, Sabah.

MATERIALS AND METHODS

PLANT MATERIALS

Leaf samples were collected from KKIP Cocoa Germplasm Collection in Kota Kinabalu, Sabah. A total of 10 plant samples (2 samples from each clone), consisting of five Class I Cocoa Commercial Clones were used to access genetic variability in this study (Table I). These five clones are Class I Cocoa Clones selected by the Malaysian Cocoa Board (2012).

Table 1: Malaysian Cocoa Board Class I Cocoa Clones Technical Information (Malaysian Cocoa Board, 2012)

Clone Name	Potential Yield		Pod and bean characteristics	
	Pod yield per tree	Dry bean yield per hectare (kg/ha/yr)	Bean number per pod	Average dry bean weight (g)
KKM 22	58.1	2420	38	1.09
MCBC 1	68.9	3590	39	1.30
PBC 123	58.7	2550	42	1.04
QH 1003	50.5	2030	31	1.36
MCBC 8	59.9	3510	40	1.50

DNA ISOLATION, SSR MARKERS AND POLYMERASE CHAIN REACTION (PCR) AMPLIFICATION

Genomic DNA was extracted from 100mg of leaf sample using DNeasy Plant Mini Kit (Qiagen, Hilden, DE) according to manufacturer's instruction. The study used 15 SSR loci (Table 1) identified as the International Standard set for cocoa germplasm characterization as reported by Saunders *et al.* (2004). Primers were synthesized by First Base (Malaysia) and forward primers were 5' labelled using HEX or 6-FAM dyes. PCR reaction mixtures consisted of 1µl containing 10-200ng of genomic DNA template, 1.6µl of a stock solution containing both forward and reverse primers (10µM each), 2µl of 2.5mM dNTP mix, 1x PCR buffer, 1.6µL of 2.5mM $MgCl_2$, 0.5µL of 0.5unit/µl Taq DNA Polymerase and sterile distilled water for a total volume of 20µL.

PCR amplification were carried out in a GeneAmp PCR System 9700 (PE Applied Biosystems) and MyIQ Single Color Real-Time PCR Detection System iCycler (BioRad) with following profile: 94°C for 4 minutes, followed by 35 cycles of 94°C 30 seconds, 46°C to 60°C (depending upon the annealing temperature of the primer pair) for 1 minute; followed by a hold at 72°C for 5 minutes.

The amplified microsatellite loci were separated by capillary electrophoresis and analysed on an ABI Prism 3730 Genetic Analyzer (fragment analysis service outsourced to First Base Sdn Bhd, Malaysia). A dataset of multi-locus genotype was generated as a result.

Table 2: Description for the 14 SSR loci genotyped in the cocoa germplasm collection.

Locus Name	Linkage group	Tm (^{o}C)	Estimated size range (bp)
mTcCIR1	8	56	120-140
mTcCIR6	6	52	220-260
mTcCIR7	7	56	150-170
mTcCIR8	9	52	300-320
mTcCIR11	2	50	280-320
mTcCIR12	4	56	170-260
mTcCIR15	1	56	230-270
mTcCIR22	1	50	280-300
mTcCIR24	9	60	180-210

mTcCIR26	8	46	280-310
mTcCIR33	4	52	270-350
mTcCIR37	10	50	140-190
mTcCIR40	3	52	270-300
mTcCIR60	2	52	190-220

DATA ANALYSIS

Duplicates were identified by using pair-wise comparisons among all the 10 individuals (5 clones x 2 replicates) based on their multilocus SSR profile. The program GenAlEX 6.5 (Peakall and Smouse, 2006) was used for genotype matching.

After exclusion of duplicates, summary descriptive statistics were computed for the collection. The descriptive statistics were computed using POPGENE version 1.32 (Yeh *et al.*, 1999). Polymorphic information content (PIC) was calculated using the formula

PIC = $1 - \sum p_i^2$ where p_i is the frequency of the allele.

The genetic distance were calculated using POPGENE software and matrices used to build dendrogram using software NTSYS pc version 2.2 (Exeter Software, Setauket, NT) using the unweighted pair group method with the arithmetic mean (UPGMA) mathematical averaging function.

RESULTS AND DISCUSSION

VARIABILITY OF ALLELES GENERATED FROM MICROSATELLITE MARKERS GENOTYPING.

The amplification process using PCR methods in amplifying 15 DNA samples from five clones (3 samples from each clones) using 14 microsatellite markers. All amplified markers were scored and used for further analysis.
Allelic variation of the polymorphic microsatellite markers was determined using POPGENE (Yeh *et al*., 1999) analysis. The result showed that polymorphic microsatellite markers used possessed diverse alleles. All 14 loci generated 264 alleles with average of 18.85 alleles per locus. Most of the primers had high allele numbers except mTcCIR1, which had slightly less number of alleles (Table 3). High level of genetic variability is important in accessing the collection genetic diversity. According to Saunders *et al*., 2004), small numbers of alleles is known to be a bottleneck in accessing genetic diversity in a population and their ability to discriminate between cocoa clones. High level of heterozygosity also indicates that there are considerable levels of admixture present in the studied collection.
The PIC value is between 0.5056 to 0.9937 throughout the primers that indicates the loci are informative loci, in accordance with Borstein *et al*. (1980) who stated that locus with PIC value > 0.5 can be considered informative. The high variation in the PIC values indicated the high diversity of the Class I Malaysian Commercial Cocoa clones and can be utilized as genetic materials in cocoa breeding programmes.

GENETIC UNIFORMITY AND RELATIONSHIP OF FIVE CLASS I MALAYSIAN COMMERCIAL COCOA CLONES

The genetic variability and relationship of five Class I Malaysian Commercial Cocoa Clones were analyzed using genetic similarity values using NTSYS programme. A dendrogram of five clones were successfully developed which formed 3 groups at coefficient 0.91. Group 1 held six samples comprises of KKM22 clone, MCBC1, PBC123 and QH1003(3), Group 2 held MCBC8 and Group 3 held one sample QH1003(5). Interestingly, Group 1 comprise of MCBC1, which is newer released clones grouped with KKM22, and PBC123 that are older clones which indicates that KKM22 and PBC123 probably share the same common ancestors to the new clones MCB1.

Table 3: Allele variations of 14 polymorphic SSR markers used to evaluate the variability of five cocoa clones

Locus	Allele Number	Exp_Het*	Nei**	Ave_Het	PIC
mTcCIR1	12	0.6818	0.625	0.15	0.9895
mTcCIR6	20	0.5053	0.48	0	0.9119
mTcCIR7	16	0.6417	0.6016	0.25	0.9172
mTcCIR8	20	0.6684	0.635	0.5	0.8957
mTcCIR11	20	0.7684	0.73	0.45	0.8593
mTcCIR12	20	0.6368	0.605	0.4	0.9528
mTcCIR15	20	0.7263	0.69	0.5	0.8448
mTcCIR22	20	0.5632	0.535	0.35	0.6950
mTcCIR24	18	0.2092	0.1975	0.1	0.5056
mTcCIR26	20	0.6895	0.655	0.4	0.9653
mTcCIR33	20	0.8053	0.765	0.45	0.9937

mTcCIR37	20	0.7842	0.745	0.5	0.8696
mTcCIR40	20	0.5684	0.54	0.4	0.9506
mTcCIR60	18	0.8562	0.8086	0.45	0.9599
Mean	18.85	0.6503	0.6152	0.35	0.8793

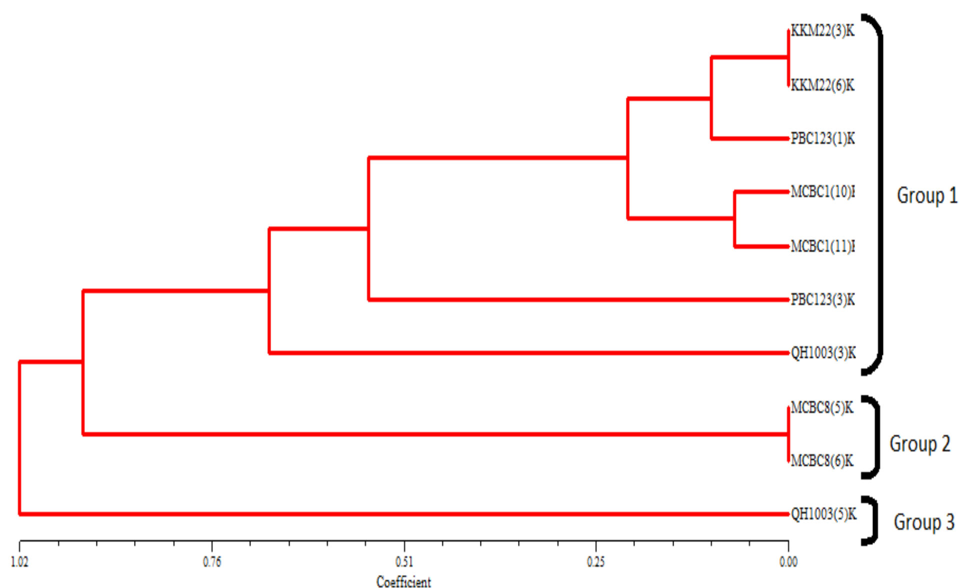
* Expected heterozygosity were computed using Levene (1949)
** Nei's (1973) expected heterozygosity

Based on the dendrogram in Figure 1, it indicates that there are mislabeled trees in the collection. In the clone PBC123 collection, sample PBC123(1) and PBC123(3) were far separated in the tree, which indicates they were not uniform genetically even though they were labeled the same clone name. Similar occurrence can be seen at clone QH1003, where the two samples were separated far to different groups, this also indicates that the clone QH1003 in this collection is not uniform genetically. There were no off-types observed in the two clone KKM22 and MCBC8 based on the comparison between two samples used in this study.

There are many reasons for non-uniformity in the cocoa germplasm collection. According to Turnbull at al. (2004), there are a number of ways in which mislabeling can occur. Examples are (i) plant lose their labels or the label become illegible, (ii) plant moved before properly labeled, (iii) mixed up of labels during vegetative propagation (iv) chupons grew from the rootstock confused as the scion, (v) plant mislabeled in the greenhouse (human error) or confused with the original tree and (vi) introduction of synonymous germplasm (with different name) from abroad.

In addition to grouping of cocoa clones, microsatellites markers may also help breeders in selecting promising parental clones candidates for new varieties. Cross breading two parental clones is predicted to produce progenies with combination of both parents' traits. According to Dias and Kageyama (1997), cocoa exhibits strong heterosis for yield and yield-contributing traits and the success of heterosis in cross breeding is determined by the genetic distance between the parental clones. The further the genetic distance between two parents, the higher probability of heterosis to occur. The clones located in different groups can assumed to have high genetic distance values, thus can be selected as parental clones or candidate clones in breeding programme of the release of new superior varieties.

Figure 1: The grouping of MCB Class I Clones based on 14 polymorphic SSR markers.



The promising parents can also be predicted based on the yield data on Table 1 and the genetic distance values as shown in Table 4. Based on Table 4, two promising parental clones combinations can be deduced, KKM22 with QH1003(5) and MCBC1(10) with MCBC8 with the highest genetic distances of 1.1935 and 1.1453 respectively. However, since Dias and Kageyama (1997) reported that cocoa exhibit strong heterosis for yield, based on Table 1, it showed that MCBC1 and MCBC8 have the highest yield and yield contributing traits in Class I commercial clones, thus crossing MCBC1(10) with MCBC8 would be predicted to produce higher probability of positive heterosis on yield.

CONCLUSION

The study using 14 microsatellite markers had successfully generated a dendrogram based on five Class I Malaysian Commercial Cocoa Clones. The markers used in this study proven capable to analyse the uniformity of the Class I Malaysian Commercial

Cocoa Clones with high polymorphism level. The clones can be grouped into three whereby Group 1 held six samples comprises of KKM22 clone, MCBC1, PBC123 and QH1003(3), Group 2 held MCBC8 and Group 3 held one sample QH1003(5). Nevertheless, the QH1003(3) in Group 3 is suspected to be an off-type tree. Additionally, based on the dendrogram and genetic distance values, we could also predict that the clones in the same group might share the same common ancestors.

Table 4: Genetic Distance values calculated from MCB Class I clones based on SSR Markers.

Clone	KKM22(3)	KKM22(6)	MCBC1(10)	MCBC1(11)	MCBC8(5)	MCBC8(6)	PBC123(1)	PBC123(3)	QH1003(3)	QH1003(5)
KKM22(3)	0.0000									
KKM22(6)	0	0.0000								
MCBC1(10)	0.2709	0.2709	0.0000							
MCBC1(11)	0.1754	0.1754	0.0716	0.0000						
MCBC8(5)	0.9126	0.9126	**1.1453**	0.8803	0.0000					
MCBC8(6)	0.9126	0.9126	**1.1453**	0.8803	0	0.0000				
PBC123(1)	0.1013	0.1013	0.2386	0.1431	0.9544	0.9544	0.0000			
PBC123(3)	0.4336	0.4336	0.8559	0.6591	0.6148	0.6148	0.4013	0.0000		
QH1003(3)	0.752	0.752	0.5876	0.5479	1.1256	1.1256	0.5479	0.9445	0.0000	
QH1003(5)	1.1935	1.1935	1.0896	0.9789	0.9429	0.9429	0.9789	1.07	0.7824	0.0000

Based on this finding, it is suggested that there is a need for a coordinated policy on identifying mislabeled or/and duplicate in the cocoa germplasm collection in all locations in Malaysia. Identified mislabeled plants should be assigned a new name unique to the clone which have meaning and could assist in documenting the origin of off-types. It is also possible that this new unique name might just be temporary once the correct clone name is identified.

Based on the genetic distance values, the findings obtained from this study may also provide valuable information for future cocoa breeding programme. Clones in different groups or with high genetic distances can be chosen as parental materials for cross breeding works to produce superior planting materials based on heterosis occurrence.

One limitation of this study is samples were obtained from just one cocoa germplasm location. It is thus suggested that the study should be extended to all the other cocoa germplasm collections that housed the Class I Malaysian Commercial Cocoa Clones. Furthermore, to ensure all the non-uniform trees are accurately identified, all the trees with the Class I clones' labeled should be analyzed for their genetic uniformity using the microsatellite markers in order for mislabeled trees to be identified and marked properly for future reference and efficient management of germplasm collection.

The study also suggested that the use of microsatellite markers could be a powerful tool to tap the genetic variability and relationships, to select new planting materials and to predict promising combination of parental clones candidates for breeding programmes. and it is recommended that the same study should be done with the other classes of Malaysian Commercial Clones.

REFERENCES

Borstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980) Construction of A Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *American Journal of Human Genetics* 32:314 – 331

Boza, E., Irish, B. M., Meerow, A. W., Tondo, C. L., Rodriquez, O. A., Ventura-Lopez, M., Gomez, J. A., Moore, J. M., Zhang, D., Motamayor, J. C., and Schnell, R. (2013). Genetic Diversity, Conservation And Utilization Of *Theobroma Cacao* L.: Genetic Resources In The Dominican Republic. *Genetic Resources Crop Evolution* 60:605 - 619.

Dias, L. A. and Kageyama, P. Y. (1997) Temporal Stability of Multivariate Genetic Divergence in Cacao (*Theobroma Cacao* L.) In Southern Bahia Conditions. *Euphytica* 93: 181 – 187

Duval, A., Gezan, S. A., Mustiga, G., Stack, C., Marelli, J-P., Chaparro, J., Livingstone, D. III., Royaert, S. and Motamayor, J. C. (2017). Genetic Parameters and The Impact of Off-Types of *Theobroma Cacao* L. In A Breeding Program in Brazil. *Front. Plant Sci*. 8:2059.

ICCO (2019) International Cocoa Organization. *ICCO Quarterly Bulletin of Cocoa Statistics, Vol. XLV, No.1, Cocoa Year 2018/19.*

Irish, B. M., Goenaga, R., Zhang, D., Schnell, R. J., and Brown, J. S. (2010) Microsatellite Fingerprint of the USDA-ARS Tropical Agricultural Research Station Cacao (*Theobroma Cacao* L.) Germplasm Collection. *Crop Science* 50:656 - 668

Malaysian Cocoa Board (2012) Malaysia Cocoa Clones. *Malaysian Cocoa Board*.

Malaysian Cocoa Board (2019) *Statistics Cocoa Cultivated Area by Region and Sector (Ha).* https://www.koko.gov.my/lkm/industry/statistic/cocoacultivated.cfm

Motilal, L., And Butler, D. (2003). Verification of Identities in Global Cacao Germplasm Collections. *Genet. Resour. Crop Evol*. 50:799-807.

Motilal, L. A., Zhang, D., Umaharan, P., Boccara, M., Mischke, S., Sankar, A., and Meinhardt, L. W. (2012). Elucidation of Genetic Identity and Population Structure of Cacao Germplasm Within an International Cacao Genebank. *Plant Genet. Resour.* 10:232-241.

Parida, S. K., Yadava, D. K., and Mohapatra, T. (2010). Microsatellites in Brassica Unigenes: Relative Abundance, Marker Design, And Use in Comparative Physical Mapping and Genome Analysis. *Genome* 53:55- 67.

Peakall, R. and Smouse, P. E. (2006) GENAlEX 6: Genetic Analysis in Excel: Population Genetic Software for Teaching and Research. *Mol. Ecol. Notes* 6: 288 – 295.

Saunders, J., Mischke, S., Leamy, E., and Hemeida, A. (2004) Selection of International Molecular Standards for DNA Fingerprinting of Theobroma Cacao L. *Theor. Appl. Genet*. 100:41 – 47.

Turnbull, C. J, Butler, D. R., Cryer, N. C., Zhang, D., Lanaud, C., Daymond, A. J., Ford, C.S., Wilkinson, M. J. and Hardly, P. (2004) Tackling Mislabelling in Cocoa Germplasm Collection. *INGENIC Newsletter* 9:8-11.

Wickramasuriya, A. M., and Dunwell, J. M. (2018). Cacao Biotechnology: Current Status and Future Prospects. *Plant Biotechnology Journal* 16:4-17.

Yeh, F. C., Yang, R. C., And Boyle, T. (1999). POPGENE Version 1.32: Microsoft Window-Based Freeware for Population Genetics Analysis. *University of Alberta, Edmonton.*

Young A, and Severson, D. (1994) Comparative Analysis of Steam Distilled Floral Oils of Cacao Cultibars (*Theobroma Cacao* L.) And Attraction of Flying Inserts: Implication for A *Theobroma* Pollination Syndrome. *J. Chemical Ecol*. 20:2687-2703

Zhang, D., Mischke, S., Johnson, E. S., Philips-Mora, W., and Meinhardt, L. (2009) Molecular Characterization of An International Cacao Collection Using Microsatellite Markers. *Tree Genet Genomes* 5:1 – 10.

ACKNOWLEDGEMENTS

The project was supported by Malaysian Cocoa Board Development Fund. The authors would like to thank the Director General of the Malaysian Cocoa Board for permission to publish these results. We also thank Mr Haya Ramba, Ms. Sairan Asim, Ms. Rafiah Abdul Karim, Ms. Nuraziawati Mohd Yazik and Hj Zailaini Md. Jamil, for their excellent experience in cocoa breeding, cocoa germplasm, technical and samples collection assistance.