Genetic variation in Maturase K (matK) from cacao (Theobroma cacao L.) varieties in Indonesia

S Immanissaa, I Faizalb, A Salamah1,3, I A Saric and A W Susilod

1Department of Biology, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia
2Agency for the Assessment and Application of Technology, Jl. M.H. Thamrin 8, Jakarta 10340, Indonesia
3Research Group of Cellular and Molecular Mechanisms in Biological System, Department of Biology, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia
4Indonesian Coffee and Cocoa Research Institute, Jl. P.B. Sudirman 90, Jember 68118, Indonesia

Corresponding author’s email: salamah@sci.ui.ac.id

Abstract. Cacao is a major agricultural commodity in Indonesia, yet its development is hindered by limited germplasm collections. In this study, the maturase K gene (matK) was used as a marker to determine patterns of genetic variation in Indonesia’s Trinitario and Forastero cacao varieties, with the results showing that the matK sequence does differentiate the varieties. However, the origin of at least one sample is unclear, as it may have been derived from crosses between the Forastero and Trinitario varieties. Similarly, an additional sample appears to be the result of the introduction of Forastero varieties from England, highlighting the importance of careful germplasm collections and molecular studies to identify contaminants.

Keywords: Theobroma cacao, matK gene, genetic variation, trinitario, forastero

1. Introduction
Cacao is a major agricultural commodity in Indonesia, the third largest cacao producer in the world after the Ivory Coast and Ghana. However, problems exist in the cultivation of cacao that have not yet been resolved, such as poor development, pests and diseases, and low quality of cacao beans. Research continues to improve the enrichment of cacao germplasm through exploration and selection of varieties in Indonesia [1].

The continued exploration and selection of novel cacao varieties is expected to help improve germplasm collections and to develop cacao-based food products with superior tastes. In an effort to increase the genetic diversity of cacao germplasm collections, a number of cacao varieties were introduced into Indonesia between 1980 and 2010 [2, 3]. However, research has been limited to differences in morphological characteristics of two varieties—Forastero and Trinitario—and molecular studies have been performed to date still very few [4, 5]. The primary aims of this research are to determine patterns of genetic variation of Trinitario cacao from Lampung and West Sumatra, as well as patterns of Forastero’s cacao from Sulawesi.
2. Materials and method

2.1. Plant materials
Five varieties of cacao (*Theobroma cacao* L.) were used: (1) Sulawesi-1 (Sul-1); (2) Masamba Cocoa Clone-01 (MCC-01); (3) Harapan Jaya-5 (HJ-5); (4) Parinari-191 (Pa-191); and (5) Pasaman Barat-1 (PB-1). Sul-1, MCC-01, and Pa-191 are varieties of Forastero and were obtained from the Indonesian Coffee and Cocoa Research Institute (ICCRI), Jember. HJ-5 and PB-1 are Trinitario varieties and were obtained from explorations conducted by ICCRI and the Agency for the Assessment and Application of Technology in Lampung and West Sumatra.

2.2. Genomic DNA extraction
Genomic DNA Plant Mini Kits were used for DNA extraction from cacao leaves. Cacao leaves were cut to lengths weighing approximately 150 mg. The leaves were crushed in a mortar under liquid nitrogen. The procedures for extraction was followed the manufacturer’s protocols.

2.3. Polymerase chain reaction
*matK* from the *T. cacao* L. chloroplast genome was amplified using the mac02-F primers (MGGATGCCCGGCTGT and ATCGGCXGGGAGCAGT), which resulted in an 872-bp product, and the mac09 primers (SGGAATAGCCCAATC and SAGCTCTCTACTGO), which resulted in a 1153-bp fragment. Amplification started with denaturation at 98 °C for 10 s, annealing at 57.2 °C for mac02-F and 57 °C for mac09-F for 15 s, and extension at 68 °C for 1 min. The process was repeated for 35 cycles. Visualization of amplification was performed using 0.8 % agarose gels at 100 V for 30 min.

2.4. DNA sequencing and sequence analysis
Amplified DNA was sequenced by the 1st Base Sequence Services Company using mac02 F-primer and mac09 F-primers. Sequences were analyzed using Sequence Scanner 2.0 software to check the quality of the nucleotide bases. Nucleotide blast (blastn) was used to confirm the nucleotide sequences corresponding to those of *Theobroma cacao matK*. Multiple sequence alignment was performed using ClustalX. Phylogenetic analyses of the sequences were performed with the Unweighted Pair-Group Method with Arithmetic mean using ClustalX 2.1 and NJPLOT.

3. Results and discussion
DNA extractions consistently had purity ratios of approximately 1.9–2.0 (table 1). Ratios of 1.8–2.0 indicate purity, whereas ratio of > 2.0 indicates the presence of contaminants in the form of RNA [6]. Some impurities were found in the Sul-1, HJ-5, and PB-1 samples, which had purity ratios of > 2.0. DNA was diluted for these samples prior to PCR. The PCR amplification results from the *matK* gene

| Sample     | DNA concentration (ng/μL) | Purity (A260/A280) |
|------------|----------------------------|--------------------|
| Sul-1      | 244.3                      | 2.05               |
| MCC-01     | 174.1                      | 1.98               |
| HJ-5       | 165.6                      | 2.07               |
| Pa-191     | 113.5                      | 1.93               |
| PB-1       | 301.4                      | 2.03               |
Figure 1. Visualization of matK gene amplification results using two primers: (a) mac02-F; (b) mac09-F; M: 1 Kb DNA molecular marker, 1: DNA sample from Sul-1, 2: DNA sample from MCC-01, 3: DNA sample from HJ-5, 4: DNA sample from Pa-191, 5: DNA sample from PB-1.

are shown in figure 1. The results showed an 872-bp fragment for the mac02-F primer and 1153-bp fragment for the mac09-F primer. The DNA fragment results were specific enough, which is likely due to the RNase treatment performed on the DNA template [6].

The nucleotide sequences of matK region were confirmed using blastn (nucleotide blast) in NCBI with 98–99 % identity with the matK region from T. cacao L. genotype Scavina-6, ICS-39, Pentagonum, Stahel, and Amelonado in GenBank NCBI. This reveals that the nucleotide sequences we obtained from the sequencing are indeed the matK region of T. cacao L.

The Multiple alignments of the matK region amplified using the mac02-F primer showed that there are differences in the nucleotide bases of samples Sul-1 and Pa-191 (figure 2a). The sample Sul-1 showed Adenine (A) base at the 400th position, while the other four samples showed Thymine (T). The sample Pa-191 showed a Cytosine (C) base sequence at the 834th position and A base at 842nd position, while the other four samples showed T base at the 834th position and Guanine (G) base at 842nd position.

The sequences from both primers were then merged to obtain a more complete matK fragments (1320 bp). The first combination sequences (460 bp mac09-F primer and 872 bp mac02-F primer products) of the five samples were aligned with matK regions of five genotypes cacao from Genbank NCBI. The multiple alignments showed that the Sul-1 is identical to that of Scavina-6 in which also showed A base at 856th position, while the other four samples showed T base. The sample Pa-191 showed C base at 1290th position, whereas the other five genotypes showed T base (figure 2b). Scavina-6 is a variety of Forastero, whereas Amelonado, ICS-39, Pentagonum, and Stahel are Trinitario varieties [7, 8].

The Multiple alignments of the matK region amplified using the mac09-F primer showed differences in the nucleotide bases of sample PB-1, at sequences positions of 1045th–1047th (figure 3a). The sample PB-1 showed AGT, whereas samples MCC-01, HJ-5, and Pa-191 showed GTA. The differences occurred in nucleotide bases at the 1045th–1047th positions of forward primer sequences actually cannot be set differently, since according to the literature, nucleotide sequences after position 850 will be poor in quality.

The 1153 bp sequences of mac09-F primer was then merged with the 200 bp sequences of mac02-F primer to form second combination of complete matK fragments (1320 bp). The Multiple alignments of the matK region of the five samples and five genotypes of cacao from Genbank NCBI showed that the matK region of PB-1 is identical to five genotypes of cacao from Genbank NCBI that showed bases AGT at positions 1037–1039 (figure 3b).

The results of multiple alignments of the matK region were made up into a dendrogram based on sequence comparison among the five samples and the matK region of five genotypes in GenBank NCBI (figure 4). Phylogenetic analysis was performed after multiple alignments and sequence-editing process with the Unweighted Pair-Group Method with Arithmetic mean using ClustalX 2.1.
Figure 2. (a) Multiple alignments of the *matK* region amplified using the mac02-F primer, and (b) the first combination of mac02-F and mac09-F primer.

Figure 3. (a) Multiple alignments of the *matK* region amplified using the mac09-F primer and (b) the first combination of mac02-F and mac09-F primer.

The phylogenetic tree was displayed onto a dendrogram (figure 4). The dendrogram were made based on the results of multiple alignments and editing of the *matK* regions from both combination of primers product. The regions of first combination showed better quality sequences than the second one. Multiple alignments of the first combining has a higher level of confidence than the second one, in which the generated dendrogram can figure out the differences in each sample, so the dendrogram results could be more accurate.

The dendogram indicated that the samples formed into two clusters. The first cluster consisted of samples Sul-1 and Scavina-6 from the Forastero varieties. These samples have identical nucleotide
Figure 4. Dendogram based on sequence comparison of the matK region from five samples sequenced in this study and five genotypes of cacao in GenBank NCBI.

sequences. The second clade consisted of two subclusters: the first consists of samples Pa-191 and MCC-01, which are Forastero varieties; and the second consisted of samples HJ-5 and PB-1, which represent the Trinitario varieties. Thus, even though the samples from the Forastero variety do not all group together, they are separated from the samples of the Trinitario variety, revealing some genetic differences between the varieties.

Samples MCC-01 and Pa-191 did not cluster with Sul-1 due to differences in their sequences. These differences are supported by morphological differences between the samples. Sample Sul-1 has a round fruit with unclear grooves, blunt end, and is red when immature. However, sample MCC-01 has an oval fruit with an obvious groove, a pointed end, and is green when immature. The morphological characteristics of Sul-1 are characteristic of all Forastero varieties, but MCC-01 has morphological characteristics that are more similar to Trinitario varieties [9].

Sample Sul-1 was the first cacao material developed in Indonesia in 2008, whereas MCC-01 was more recently developed in 2014 [9]. Further research regarding the origin of MCC-01 is necessary to confirm whether the sample was a pure Forastero variety, or if it was derived from crosses between Forastero and Trinitario varieties. Sample Pa-191 is a result of the introduction of Forastero varieties from England, so it is possible to have different nucleotide sequences of the matK region compared with Sul-1, which is the Forastero’s cacao from Indonesia.

4. Conclusion
The nucleotide sequences of cacao from Indonesia, Sul-1, MCC-01, HJ-5 and PB-1, and cacao introduction Pa-191 were identified using matK as a genetic marker for phylogenetic analysis. Trinitario and Forastero varieties are genetically differentiated, and differences between samples Sul-1 and MCC-01 are supported by morphological characteristics as well.

References
[1] Iswanto A, Winarno H and Mawardi S 1999 Int. Workshop on the Contribution of Disease Resistance to Cocoa Variety Improvement (Salvador: INGENIC) pp 205-11
[2] Bartley B G D 2005 The Genetic Diversity of Cocoa and Its Utilization (Wallingford, Oxfordshire: CABI Publishing, CABI Int.)
[3] Mawardi S, Winarno H and Suhendy D 1994 Proc. of the Int. Workshop on Cocoa Breeding Strategies (Kuala Lumpur) (United Kingdom: INGENIC) pp 81-7
[4] Dinarti D et al. 2015 Breed. Sci. 65 438-46
[5] Lukman et al. 2014 Trop. Plant Biol. 7 133-43
[6] Sambrook J and Russell D W 2001 Molecular Cloning A Laboratory Manual 3rd edition (New York: Cold Spring Harbor Laboratory Press)
[7] Kane N et al. 2012 Am. J. Bot. 99 320-9
[8] Motamayor J C, Risterucci A M, Heath M and Lanaud C 2003 Heredity 91 322-30
[9] McMahon P et al. 2015 Crop. Prot. 70 28-39