Identification of maturase K (matK) gene in trinitario cocoa 
(Theobroma cacao L.) from Lampung Province and Central Java, Indonesia

L P Dermawan¹, I Faizal², A Salamah¹, I A Sari³ and A W Susilo³

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
³Indonesian Coffee and Cocoa Research Institute, Jl. P.B. Sudirman 90, Jember 68118, Indonesia

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

Abstract. This research was conducted to determine the genetic variation in Trinitario cocoa (Theobroma cacao L.) accessions obtained from exploration and selection in Lampung Province, named HJ1, HJ2, HJ3, and HJ4, and the earlier Trinitario cocoa accession, named DR2. Identification was performed by comparing the DNA sequence of the matK region in the chloroplast genome. Genomic DNA isolation was performed with samples of old and fresh cocoa leaves. Amplification was conducted using two different primers Mac 02 (872 bp) and Mac 09 (1,153 bp). The sequencing results obtained indicated genetic variation in accession HJ3 amplified using the Mac 09 primer. A dendrogram constructed using the UPGMA method showed that the cocoa accessions from the exploration and selection group (HJ1, HJ2, HJ3, and HJ4) were grouped into one cluster and that DR2 was more ancient than the other four cocoa accessions.

Keywords: Theobroma cacao, matK gene, trinitario, genetic diversity, Lampung

1. Introduction
Theobroma cacao L., commonly known as cocoa, is usually grown in tropical forests with high rainfall and humidity. This species came from the Amazon rainforest and was initially brought into Indonesia at Minahasa, North Sulawesi, in 1560 [1]. To date, the cocoa plant in Indonesia continues to be developed and cultivated on both large and small scales to improve the country’s foreign exchange resources from the non-oil sector.

Cocoa plants can be grouped into three varieties: Criollo, Forastero, and Trinitario. Cocoa Criollo and Trinitario are part of the specialty cocoa ("edel cocoa"), characterized by the white color of the seeds [2]. The production of specialty cocoa in Indonesia is currently very low (< 1 % of the volume of national production), but it is one of the most high-value products and is highly sought after because it is rare and has a distinctive flavor. Indonesia itself has only a few of these specialty cocoa plants that are listed as being producers of cocoa with white seeds. One of the specialty cocoa accessions that continues to be developed and cultivated is the cocoa DR plant. Cocoa DR consists of several series,
such as DR1, DR2, and DR38. The cocoa DR plant is a Trinitario cocoa (Java Trinitario) first-generation hybrid, which arose in Djati Roenggo, Unggaran, Central Java, in 1912 [3].

The sustainability of specialty cocoa production in Indonesia is now facing serious problems due to disease incidence, due to which the production of this cocoa has been declining [4]. Therefore, it is vital to carry out genetic enrichment through the exploration and selection process to obtain disease-resistant fine cocoa seeds with a white bean percentage of > 80%. Exploration and selection have been carried out in Pesawaran and East Lampung District, Lampung Province. As a result, four accessions of the specialty cocoa plant were collected, including HJ1, HJ2, HJ3, and HJ4. The fourth variant of the cocoa is believed to come from the earlier, possibly parental generation of specialty cocoa, i.e., DR1 and DR2, and is expected to have inherited traits from both the parental plants [3].

Through molecular characterization using maturase K (matK) gene as a molecular marker, we aimed to determine whether there is genetic variation among the four cocoa plants obtained from exploration and selection in Lampung Province, with the DR2 plant suspected to be one of their parents. The aim of this research was to determine whether matK is a marker suitable for determining the genetic diversity in the cocoa plant (T. cacao L.) and whether there is genetic variation among the four accessions of the cocoa plant (HJ1, HJ2, HJ3 and HJ4) from Lampung Province, which can be traced from the first generation of the Trinitario cocoa plant (DR2) from Djati Roenggo, Central Java.

2. Materials and method

2.1. Plant samples
The material used in this research included fresh old cocoa leaves tissue, which were obtained from the exploration and selection process in Lampung Province, named HJ1, HJ2, HJ3 and HJ4, as well as an earlier accession from Central Java, named DR2.

2.2. DNA extraction
DNA was extracted from cocoa leaf tissue (150 mg/0.5 g). Liquid nitrogen was added to the leaf tissue, which was then crushed using a mortar and pestle. The next procedures following the protocols from the Geneaid genomic DNA mini kit (plant).

2.3. Amplification
Cocoa matK was amplified using two primers, Mac 02 and Mac 09. Mac 02 F-primer (MGGATGCCCCBGTGGT) and Mac 02 R-primer (ATCGGCGGGAFGCAGT) produced an 872-bp DNA fragment. The Mac 09 F-primer (SGGAATAAGCGCAACTG) and Mac 09 R-primer (SAGCTCTSCTACTG) produced a 1,153-bp DNA fragment. Amplification started with pre-denaturation at 98 °C for 2 min, denaturation at 98 °C for 10 s, annealing at 57 °C for 15 s, extension at 68 °C for 1 min, and final extension at 68 °C for 1 min. The process was carried out for 35 cycles. Visualization of the amplification product was performed using 0.8 % agarose gel at 100 V for 35 min.

2.4. Sequencing and data analysis
The products of PCR amplification were sequenced by the 1st Base Company. The sequence was analyzed using Sequence Scanner 2.0 software and then aligned using ClustalX2 (www.clustal.org) [5]. The result of the multiple alignments was then used as a base for drawing a phylogenetic tree as a dendrogram (with the UPGMA method) using the NJplot program.

3. Results and discussion
The results of DNA extraction are shown in table 1. The range of purity was 1.91–1.99, indicating good genomic DNA purity [6], with the DNA concentration ranging from 101.7 to 276.8 ng/μL. Variations in the genomic DNA concentrations were associated with differences in leaf colors among the samples,
reflecting different frequencies of chloroplasts within each leaf. All the samples had their genomic DNA concentration adjusted to 101.7 ng/μL, corresponding to the lowest concentration in DR2.

The *matK* was amplified from all the five samples using two pairs of primers, Mac 02 and Mac 09. The optimal annealing temperature was determined using the gradient temperature technique. The optimum temperature for Mac 02 was 57.2 °C, and for Mac 09 was 57 °C. The amplification results for *matK* are shown in figures 1 and figure 2. Based on the results of electrophoresis gel visualization, the band obtained from the Mac 02 primer was 872 bp, and that from the Mac 09 primer was 1153 bp. The bands obtained matched with the length of the targeted sequence from each primer.

Sequencing of the amplified fragment showed that the *matK* sequence was divided into two regions: Mac 02, with a sequence length of 872 bp, and Mac 09, with a sequence length of 1,153 bp. The sequence obtained was presented in the form of an electropherogram, and the quality of the sequence was analyzed with the help of Sequence Scanner application 2.

Sequence homology was carried out with the Basic Local Alignment Search Tool (BLAST; http://www.ncbi.nlm.nih.gov/). The results of BLASTn are shown in figure 3. The sequences from the five cocoa samples were 99% identical with the *T. cacao matK* sequence from the partial chloroplast genome from the NCBI database.

The nucleotide sequences were aligned with each other using the multiple alignment program on the ClustalX2 application. Based on the results from the alignment between the samples amplified with Mac 02 primers, there was no difference between the five samples. The alignment was also performed between the samples amplified using Mac 09 primers. Based on the results obtained, there were differences in the bases at the end of the base sequence 908–1180. The Mac 09 alignment results can be seen in figure 4, where the base differences are highlighted with a red square.

### Table 1. Genomic DNA concentrations from five samples of cocoa leaves.

| Sample | DNA Concentration (ng/μL) | Purity (A260/A280) |
|--------|---------------------------|---------------------|
| HJ1    | 154.3                     | 1.91                |
| HJ2    | 142.1                     | 1.99                |
| HJ3    | 276.8                     | 1.94                |
| HJ4    | 254.6                     | 1.98                |
| DR2    | 101.7                     | 1.96                |

![Figure 1](image_url)  
Figure 1. Amplification result of the *matK* gene fragment using the Mac 02 primer; M: 1 kb DNA molecular marker, 1: DNA sample from HJ1, 2: DNA sample from HJ2, 3: DNA sample from HJ3, 4: DNA sample from HJ4 and 5: DNA sample from DR2.
Figure 2. Amplification result of the \textit{matK} gene fragment using the Mac 09 primer; M: 1 kb DNA molecular marker, 1: DNA sample from HJ1, 2: DNA sample from HJ2, 3: DNA sample from HJ3, 4: DNA sample from HJ4 and 5: DNA sample from DR2.

Figure 3. NCBI BLASTn results of \textit{matK} gene sequences.

Alignment results from the samples amplified with Mac 09 primers indicated the presence of a number of base differences between the samples. The alignment results were verified using data quality analysis with the Sequence Scanner 2 software for only high-quality sequences (figure 5). Based on the sequence alignment results, the sample which differed most from the other samples was found to be HJ3, which differed from other samples for one base at base sequence number 908. On the other hand, other samples showed no sequence differences.

The merger between the sequences from Mac 02 and 09 was performed to obtain the overall \textit{matK} sequence. The overall length of the sequence obtained was 1,320 bp. The alignment of \textit{matK} sequences was then performed between the samples along with the reference cocoa sequence that was registered in GenBank (NCBI; Criollo cocoa-22, ICS-01, ICS-39, Pentagonum, and Stahel) [7]. This fifth reference was used as a comparative sequence [8]. Based on the results of alignment, the variation in nucleotide sequences between the samples was used as a basis for preparing a phylogenetic tree. The dendrogram is presented in figure 6.
Figure 4. Alignment result of the samples amplified with Mac 09 primers.

![Alignment result of the samples amplified with Mac 09 primers.](image)

*differences between the bases

Figure 5. *MatK* sequence differences among the five cocoa samples.

![MatK sequence differences among the five cocoa samples.](image)

Figure 6. Dendrogram using the UPGMA method.

![Dendrogram using the UPGMA method.](image)
Based on the results of the dendrogram, samples HJ1, HJ2, HJ3 and HJ4 were grouped into one cluster, having several subclusters that were increasingly distant from the initial cluster. The position of the sample DR2 on the dendrogram was distinct from the subcluster containing the exploration and selection samples. The position of DR2 indicated that the four samples from the exploration and selection population originated from DR2; thus, DR2 could be considered as a parental line of these four accessions. This finding is consistent with that in the literature, stating that cocoa accession DR2 is a Trinitario first-generation parent, and HJ1, HJ2, HJ3 and HJ4 are the offspring of cocoa DR2 from Jati Roenggo [3].

4. Conclusion
The \textit{MatK} can be used as a marker for determining genetic diversity in cocoa plants (\textit{T. cacao}). The sequencing results showed that there is genetic variation in HJ3 compared with HJ1, HJ2, HJ4, and DR2 following \textit{matK} amplification using the primer Mac 09. These differences occurred at base sequence 908. Dendrogram results showed that DR2 is a parent of the four cocoa plants obtained from the exploration and selection program (HJ1, HJ2, HJ3 and HJ4).

References
[1] Motamayor J C et al. 2008 \textit{PlosONE} \textit{3}, e3311
[2] Bartley B G D 2005 \textit{The Genetic Diversity of Cocoa and Its Utilization} (Wallingford, Oxfordshire: CABI Publishing, CABI Int.)
[3] Mawardi S, Winarno H and Suhendy D 1994 \textit{Proc. of the Int. Workshop on Cocoa Breeding Strategies} (Kuala Lumpur) (United Kingdom: INGENIC) pp 81-7
[4] Iswanto A, Winarno H and Mawardi S 1999 \textit{Int. Workshop on the Contribution of Disease Resistance to Cocoa Variety Improvement} (Salvador: INGENIC) pp 205-11
[5] Clustal: Multiple Sequence Alignment 2017 available at www.clustal.org
[6] Sambrook J and Russell D W 2001 \textit{Molecular Cloning A Laboratory Manual} 3rd edition (New York: Cold Spring Harbor Laboratory Press)
[7] National Center for Biotechnology Information (NCBI) 2020 Nucleotide BLAST (United States: National Library of Medicine) available at https://blast.ncbi.nlm.nih.gov/Blast.cgi
[8] Kane N et al. 2012 \textit{Am. J. Bot.} \textit{99} 320-9