RESEARCH ARTICLE

SABRAO Journal
of Breeding and Genetics
52 (4) 493-505, 2020

IDENTIFICATION OF SINGLE NUCLEOTIDE POLYMORPHISM IN *FATA* GENE ENCODING FOR ACYL-ACP THIOESTERASE TYPE-A OF OIL PALM

M.D. PRASETYO[1], SUHARSONO[1,2,3*], T. LIWANG[4] and ROBERDI[4]

[1] Study Program of Biotechnology, Graduate School, IPB University, Indonesia
[2] Research Center for Bioresources and Biotechnology, IPB University, Indonesia
[3] Department of Biology, IPB University, Indonesia
[4] Department of Biotechnology, PT SMART Tbk, Indonesia
*Corresponding author email: sony-sh@apps.ipb.ac.id
Email addresses of coauthors: prasetyo_tyo@apps.ipb.ac.id, tony.liwang@sinarmas-agri.com, roberdi@sinarmas-agri.com

SUMMARY

The saturated and unsaturated fatty acid compositions in commercial oil palm varieties are almost equal. The major unsaturated fatty acid component in palm oil is oleic acid. Acyl-ACP-thioesterase Type A (FATA) is 1 of the enzymes responsible for oleic acid synthesis. The discovery of SNP (single nucleotide polymorphism) in FATA gene using genotypes with different oleic acid contents is important for marker development. The objective of this study was to obtain SNP in the FATA gene to develop a SNAP (single nucleotide amplified polymorphism) markers. The FATA gene sequences from the National Center for Biotechnology Institute (NCBI) located in chromosomes 7 and 8 was used to design primers. Those primers were used to amplify the exon region in both chromosomes. The number of exons of the FATA gene in chromosome 7 and 8 were 7 and 8 exons, respectively. Eleven primer pairs were designed for the amplification of all these exons. The DNA of 13 plants composed of 5 *Elaeis guineensis*, 3 *Elaeis oleifera*, and 5 hybrids which were predicted producing different oleic acid content were amplified with the primers. Based on the sequence alignment analysis, 103 SNPs were found. Ten SNPs were selected consisted of 9 nonsynonymous and 1 synonymous SNP. All of 10 SNPs were used as a template for SNAP primer design. In total 9 SNAPs primers were chosen for analysis. Four (44.4%) of the 9 SNAPs were able to distinguish between *E. oleifera* and *E. guineensis*. All of 4 selected SNAP primers were used for the analysis with a larger population. PIC values of the primers were ranged from 0.414 to 0.482. Phylogenetic analysis based on FATA gene showed that 13 accessions of *Elaeis guinensis*, *E. oleifera* and its hybrids were separated into 3 groups.

Keywords: FATA gene, oil palm, oleic acid, PIC, SNP, SNAP

Key findings: The 4 SNAPs primers were able to distinguish between *E*. *oleifera* predicted containing high oleic acid and *E. guineensis* as normal oleic acid content with informative value.

Manuscript received: August 10, 2020; Decision on manuscript: October 18, 2020; Accepted: November 22, 2020.

Communicating Editor: Dr. Naqib Ullah Khan

INTRODUCTION

Oleic acid is an omega-9 unsaturated fatty acid with 18 carbon atoms with a double bond position on the 9th carbon. Vegetable oils with a high content of oleic acid (18:1) are of interest for both food and industrial purposes. The need for high-oleic crops is increasing as the food market and the agroindustry demand for oils that are more resistant to oxidation. It means that palm oil with a higher proportion of unsaturated fatty acids could mean new market opportunities for oil palm planters. Oleic acid can reduce blood pressure by decreasing the amount of low-density lipoprotein concentration which is responsible to trigger the risk of heart disease. Oleic acid is often found as 1 of the main components of vegetable oil. One important source of vegetable oil with high oleic acid is palm oil.

The oil palm consists of two species named African oil palm (*Elaeis guineensis* Jacq.) and American oil palm (*Elaeis oleifera*). *E. guineensis* is preferred for commercialization due to its high oil productivity. Oil palms were planted in South East Asia, Africa, and Latin America. Indonesia and Malaysia are the two main producing countries. According to the USDA (2019), until November 2019, Indonesia have exported 30.3 tons of palm oil and accounted for 55% of the world's palm oil production. Nevertheless*, E. oleifera* has an advantage in better oil quality with high unsaturated fatty acids composition.

The average of oleic acid in *E. oleifera*, *E. guineensis*, and the *Eo*/*Eg* hybrids is 61.7% (Lieb *et al.,* 2017), 41.4%, and 50% (Mozzon *et al.,* 2013), respectively. High unsaturated fatty acid content is positively correlated with Iodine Value (IV). IV is the size of unsaturated fats in oil (Rajanaidu *et al.,* 2000). Until recently, the oleic acid content of *Eo*/*Eg* hybrids has never exceeded its parents. This problem can be overcomed by implementing Marker-Assisted Selection (MAS) approach. The composition of the oil must be adapted to the growing needs of food, oleochemical, and fuel industries (Barcelos *et al*., 2015).

The application of MAS in oil palm breeding is greatly supported by the availability of the oil palm genome sequences. The oil palm genome is predicted to have at least 45,000 genes (Singh *et al.,* 2013). The rapid developments in genotyping technology make the detection of Single nucleotide polymorphism (SNP) cheaper and more accurate. This makes SNP very useful in plant breeding programs. SNPs are widely found in plant genome, including oil palm. A total of 200,000 SNPs were found in *E. oleifera* (Kwong *et al.,* 2016). SNP plays a major role in the analysis of phenotypic variations (Buckler *et al.,* 2002; Rafalski, 2002).

SNP is a variation of 1 base in a DNA sequence when compared between individuals. Point mutations that alter gene expression are called as non-synonymous mutations. While mutations that do not cause change in gene expression are called as silent mutations or synonymous mutations. Single base variations can be used as a reference for estimating kinship between offspring and its parents. Single nucleotide base differences can be used to develop PCR-based markers. The developments of molecular markers with low-cost analysis will accelerate the oil palm breeding for superior traits including palm oil quality improvement (Wong and Bernardo, 2008).

The SNP technology focuses on a single nucleotide difference within a gene. Mapping of intra gene SNPs candidate genes related to oleic acid biosynthesis was closed to acyl-ACP thioesterase type A (FATA) and Δ9 stearoyl-ACP desaturase (SAD) (Montoya *et al.,* 2013). SNP analysis in SAD was reported (Borlay *et al.,* 2017). Fatty acids synthesis takes place in plastid. Acyl carrier protein elongation occurs with several cycles that extend the acyl chain by 2 carbon atoms. The acyl-ACP that has been extended to 18 carbon atoms could be desaturated by SAD (stearoyl-ACP desaturase) to 18:1 acyl-ACP then processed by FATA into oleic acid (Bates *et al.,* 2013).

FATA gene in *Haematococcus pluvialis* has a positive correlation with fatty acid content with a carbon chain length of 18 and 20 (Lei *et al.,* 2012). Expression of FATA gene from sunflower seeds in *Escherichia coli* showed that FATA encodes thioesterase especially for monounsaturated acyl-ACPs (Serrano-Vega *et al.,* 2005). Jones *et al.* (2005) showed that FATA gene in *Cuphea hookeriana* encodes for 18:1-ACP thioesterase.

Since the role of FATA is very important in the synthesis of oleic acid, the SNP in this gene is very interesting to be investigated. The objective of this study is to obtain SNPs in FATA gene of the population composed of *E. guineensis*, *E. oleifera*, and *Eo*/*Eg* hybrid. These 3 types of major oil palm were chosen because they represent oil palms which contains low, high, and intermediate concentration of oleic acid, respectively, based on previous studies (Mozzon *et al*., 2013; Lieb *et al*., 2017). Then obtained SNPs were used as the basis for generating Single Nucleotide Amplified Polymorphisms (SNAP) markers. The SNAP markers could be applied to estimate the content of oleic acid in rape seed (Yang *et al*., 2012). We assume that the same principle can be applied hypothetically to oil palm The development of SNAP markers based on SNP from several samples based on bi-allelic criteria was reported (Pesik *et al.,* 2017). SNP is a marker with a relatively small occurrence per genome, a relatively low mutation rate, evenly distributed throughout the genome and detection is relatively easy. Insertion and deletion are the main bases as bi-allelic markers (Kurt *et al*., 2005). Currently, SNAP has been applied in the selection of high oleic acid in rapeseed (Yang *et al*., 2012), starch in maize (Kim *et al*., 2016), and for resistance to downy mildew disease in chestnuts (Xu *et al*., 2007).

MATERIALS AND METHODS

Plant materials

Genetic material for SNP identification were 13 accessions of oil palm consisted of 5 accessions of *E. guineensis*, 3 accessions of *E. oleifera*, and 5 accessions of *Eo/Eg* hybrids which were predicted to have difference in oleic acid contents. Ten oil palm accessions originated from Ghana, Evolution, Nigeria, Ekona, and Compact, and Eo/Eg hybrid were used for SNAP markers analysis. Genomic DNA was extracted from oil palm leaves using NucleoSpin Plant II kit (Macherey-Nagel, Germany) based on manufacturer instruction. DNA purity and quantification were measured by spectrophotometer NanoDrop 2000C (Thermo Scientific, United States of America).

Isolation of FATA gene

The primers were designed based on the database of FATA gene of *E. guineensis* (accession number JN003492.1) found in National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/) to amplify the exon region by using Primer3plus software (http://bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). The DNA was amplified by PCR using KOD FX Neo (Toyobo, Japan). The PCR mixture was prepared based on the Toyobo manual kit instruction in the final volume of 50 µL. Amplification was carried out by 35 cycles with the condition of pre-PCR at 95 °C for 5 min, denaturation at 98 °C for 20 s, annealing at 54-64 °C for 15 s, extension at 72 °C for 15 s, and post-PCR at 72 °C during 1 min. The PCR products were later electrophoresed in 1% agarose gel and purified using QIAquick PCR Purification Kit (Qiagen, Germany) based on manual instruction.

SNP analysis and SNAP primer design

The nucleotide sequences were aligned using Geneious software version 10.0.3 (https://www.geneious.com/). Gene reference for FATA was the accession of *E. oleifera*. The consistent SNPs among genotypes were selected and used for the template of SNAP primer design using the WebSNAPER Software (Drenkard *et al.*, 2000). Motif discovery on FATA gene sequence was performed using MEME software (Bailey *et al.,* 2006). Selected SNAP primers were used for the amplification of DNA of 13 oil palm accessions which were used in the sequencing process for consistency verification. Consistent and informative primers according to electropherogram are selected for further analysis.

Application of SNAP in the population and data analysis

Sixty oil palm accessions were amplified using selected SNAP primers. Amplification of DNA was performed by PCR in the final volume of 15 µL using DreamTaq DNA Polymerase (Thermo Scientific, United States of America). Amplification was carried by 35 cycles with the condition of pre-PCR at 95 °C for 3 min, denaturation at 95 °C for 20 s, annealing at 54-64 °C for 30 s, extension at a temperature of 72 °C for 1 min with post-PCR at a temperature of 72 °C for 1 min later analyzed with a 1% agarose gel. Data

were scored as binary data. Data were analyzed using GenAlex software version 6.5 (Smouse and Peakall, 2012), Darwin version 6 (Perrier and Jacquemoud-Collet, 2015), and Cervus version 3.07 (Marshall *et al.*, 1998).

RESULTS AND DISCUSSION

FATA genes were located on chromosomes 7 and 8 in the oil palm genome. The length of the FATA gene in chromosome 7 was 4827 bp and in chromosome 8 was 3482 bp (Figure 1). The FATA genes consisted of 7 and 8 exons in both chromosomes, respectively. Length of each exons in chromosome 7 were varied from 69 bp to 331 bp, while in chromosome 8 were ranged from 68 bp to 338 bp. A total of 11 primer pairs were designed to amplify all of the exon, consisting of 6 primer pairs for the chromosome 7 and 5 primer pairs for chromosome 8 (Table 1). The amplicon length was varied from 455 to 790 base pairs. The clear and single-band amplicon was purified and sequenced.

The nucleotide sequences of 13 accessions were aligned using Geneious software for single-base differences (SNP) identification. Total numbers of SNP found were 103, meaning that there was 1 SNP per 22 bp on average. In another study on oil palm, there were 1 SNP in every 21,471 bp with genotype by sequencing method (Pootakham *et al.,* 2015).

Based on allele consistency in each group, 10 SNPs were selected from 10 primer pairs. Nine SNPs were nonsynonymous mutation and 1 SNP from chromosome 7 was synonymous SNP (Table 2). There was no selected SNP obtained from the Primer FATAE4-57 which amplify exon 4 and 5 in chromosome 7. The SNPs from this primer was inconsistent among the group of oil palm accessions.

Sequences of FATA gene from *E. guineensis, E. oleifera*, *Eo/Eg* hybrids were aligned with FATA from other species with BLAST. Alignment analysis showed that FATA gene of *Eo*/*Eg* hybrid was close to *E. oleifera* than to *E. guineensis,* even though they are in 1 group. The FATA gene of 3 oil palm genotypes was the most identical with *Cocos nucifera* and *Phoenix dactylifera* (Figure 2) which belongs to palmae groups with similarity in sequence considerably high (91 to 99%). FATA gene from dicotyledonous was separated from other monocotyledonous plants. The FATA from two families, i.e. Poaceae and Iridaceae, is also close to oil palms. This result was different to the other FATA gene analysis where FATA gene of oil palm was very close to *Iris germanica* and *I. tectorum* belonging to Iridaceae, and far from Family of Poaceae like *Sorghum bicolor* and *Zea mays* (Rosli *et al*., 2018). Even though there is a duplication of data analyzed, the data can have different characteristics in the different benchmarks and in different organisms (Chen *et al*., 2017).

Nucleotide sequences were translated to protein using Geneious software. The protein sequences were analyzed using MEME software (Bailey *et al.,* 2006) for the discovery of new transcription factor binding sites and protein domains. The protein sequence analysis showed that both *E. oleifera* and *E. guineensis* have 3 motifs (Figure 3). The motifs found in

Table 1. Primers to amplify exon region of FATA gene.

No.	Primer Name	Position	Nucleotide Sequence
1	FATAE17	F	CGATATTTTTCATAACTTTTGTGG
		R	GAAAGAATAGAAATTTCCAAAGATCC
2	FATAE2-37	F	TGTACAAGATCAAACTAGTCATTGC
		R	GCATGCCATGTAAACTATCGTAAC
3	FATAE4-57	F	GTTATTTGTAACTCTGTGACAGCAA
		R	GTGAAAGCTTAAATTAGGCAAATAA
4	FATAE6-77	F	CCCAGGAGAGCCGATTTG
		R	TCATCGAACTAGCTTCCTCCACT
5	FATAE7-7	F	GGTTCGCCACGACCCCTACCA
		R	ATCCAATCACGCCTGGTGCCT
6	FATAE18	F	GTCGTATCGACCGAGGACCTTTA
		R	GAATGAATGGAACCAAAGATCC
7	FATAE2-38	F	TACTGGTGCCATGATGCAGGAAGTT
		R	AGAGCAGACACAATAGAAGTAATAACC
8	FATAE4-58	F	ATCAGTTCAATAGACCATGACAATTT
		R	TATAGAGAACCTCAAATGAAGATGGTGA
9	FATA6-78	F	ACTTCCTCTATGCTGGGGTAGTTT
		R	AGAGCAGACACAATAGAAGTAATAACC
10	FATAE88	F	TTGGGATGGAGACAATTCG
		R	TCATCGAACCAACTTCCTCCA
11	QFATA8-50	F	GTTTGATATGTGACCTTTGTCAT
		R	ACAACAGCGAAATGACAGCATA

Table 2. Selected SNP position in FATA gene.

Primer Number	SNP Position	Reference		Alternate		Mutation Type
		Allele	Amino acid	Allele	Amino acid	
FATAE17	233	T	Cysteine	G	Glycine	*Nonsynonymous*
FATAE2-37	207	A	Isoleucine	G	Glycine	*Nonsynonymous*
FATAE4-57	161	T	Leusine	A	Arginine	*Nonsynonymous*
FATAE6-77	248	G	Threonine	C	Threonine	*Synonymous*
FATAE18	88	G	Glycine	A	Glutamate	*Nonsynonymous*
FATAE2-38	199	C	Threonine	G	Arginine	*Nonsynonymous*
FATAE4-58	339	C	Leucine	T	Threonine	*Nonsynonymous*
FATA6-78	290	T	Cysteine	C	Arginine	*Nonsynonymous*
FATAE88	236	T	Alanine	G	Arginine	*Nonsynonymous*
QFATA8-50	292	T	Leucine	C	Serine	*Nonsynonymous*

the sequences are conserved regions. Motif discovery is an important step in protein analysis. The specific region that determines the protein structure and stability are usually located in a specific motive (Bailey *et al*., 2006). Sequence motifs are generally related to the regional functions of protein as catalytic sites, binding sites, and structural motifs between proteins (Falquet *et al.,* 2002). The motif sequences for most of the catalytic and binding site were known to have a conservative area that tends not to

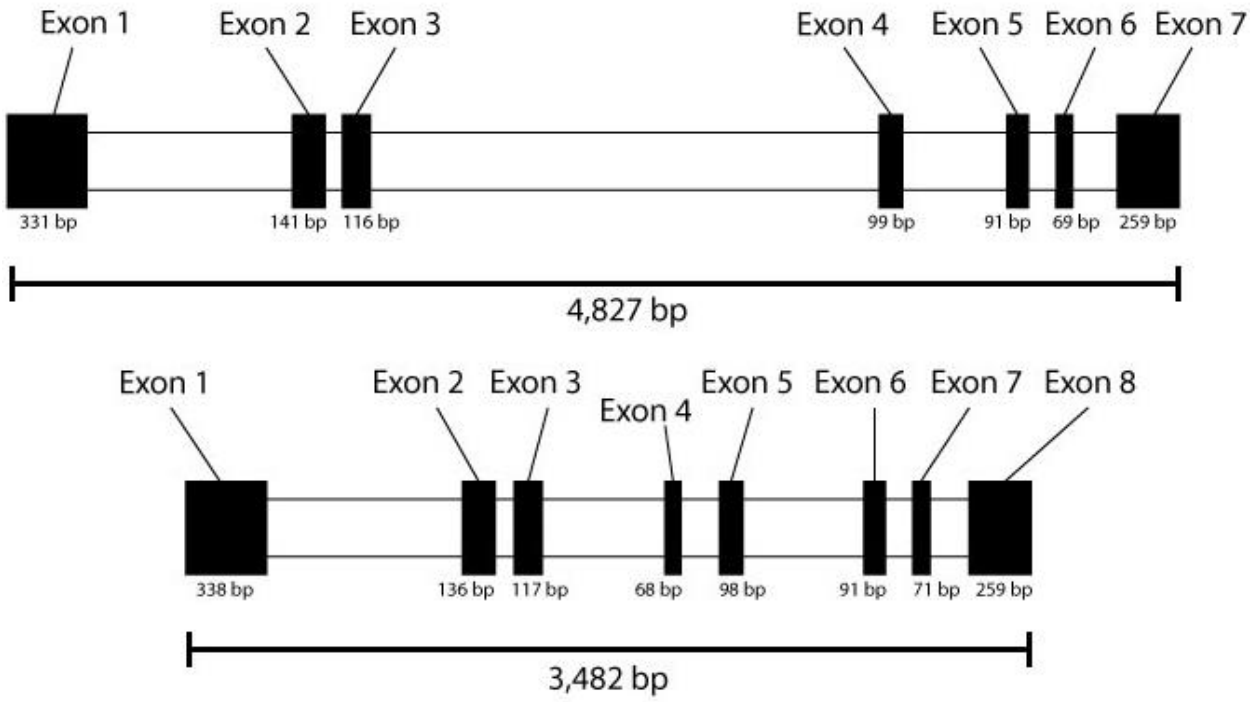

Figure 1. Visualization of FATA gene on chromosome 7 (top) and chromosome 8 (bottom).

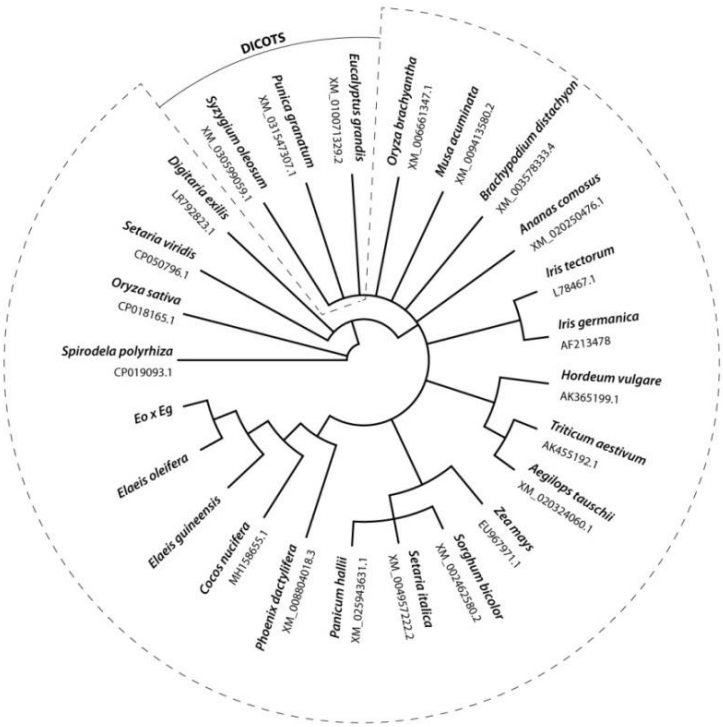

Figure 2. Phylogenetic tree based on FATA gene.

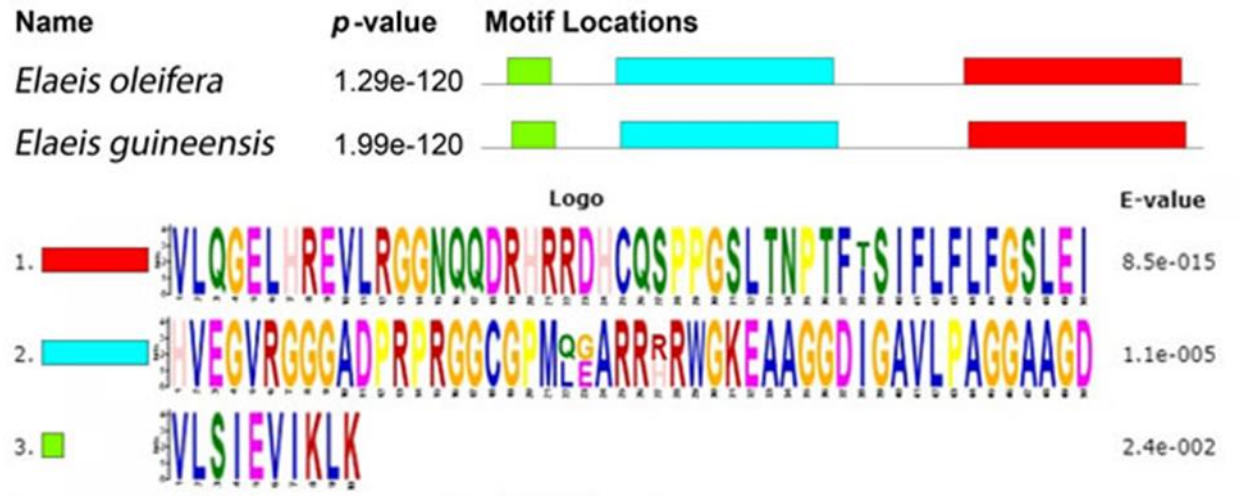

Figure 3. FATAE18 motif obtained from MEME Software.

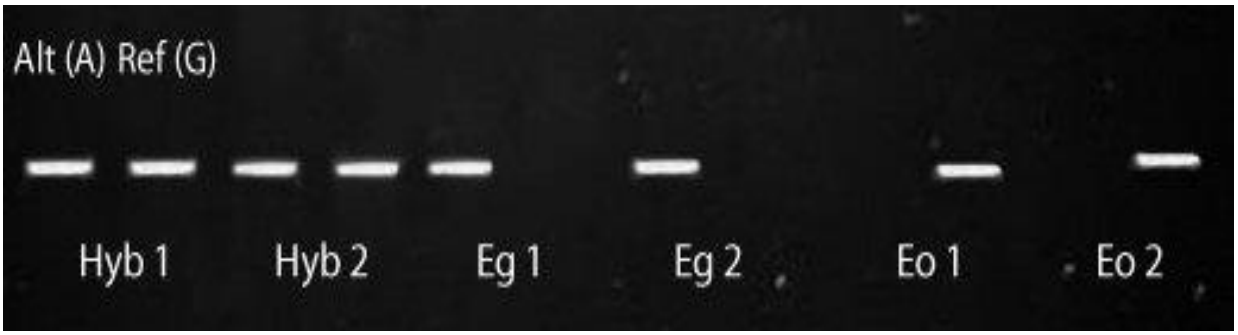

Figure 4. SNAP primer amplification (P1 88). Eo (*E. oleifera*), Eg (*E. guineensis*), Hyb (*Eo* × *Eg* hybrid).

change. The sequence of motif positions in conserved regions can be an identity of a gene or individual organism and could be used to differentiate between organism and evolution process.

Motif discovery is very useful in FATA gene because the regulatory mechanism of this gene is not yet known. Motif discovery helps to reveal a regulated region in the unknown areas (Bais *et al.,* 2011; Koschmann *et al.,* 2012; Machens *et al.,* 2013). The FATA gene works by splitting the ACP group on an 18:1-ACP substrate to produce oleic acid. Motif discovery might increase the sensitivity of the marker in detecting the desired trait in the MAS program.

FATA gene plays an important role in fatty acid synthesis, especially oleic acid. FATA gene product is Acyl-ACP specific thioesterase enzyme at 18:1 ACP. It is referred to as "long-chain" or "Oleoyl" ACP thioesterase. This enzyme acts in releasing free fatty acids from its Acyl-ACP. Decreased expression of the FATA gene showed a drastic reduction in the concentration of oleic acid in the

Arabidopsis seeds (Moreno-Pérez *et al*., 2012), thereby affecting the fatty acid concentration or triacylglycerol arrangement.

Development of SNAP marker and genetic diversity analysis

All selected SNP were analyzed by SNAPER Web program to obtain SNAP's primer candidates. Those candidates were sequences of combination that can be able to distinguish between a single base by changing the order that is closer to the 3' end of the SNP. There were 26 primer pair candidates for consistency testing by amplified with 13 accessions used for sequencing. We obtained 26 primer pair candidates from this analysis. These primer pair candidates were then tested to 13 oil palm accessions to confirm the previous results. The confirmed SNAP primers which consistently amplify the targ*et al*lele with sequencing electrophoregram were selected (Table 3). Heterozygote samples were amplified by both the reference (REF) and alternate (ALT) primer (Figure 4). Four of 9 SNAPs were consistent to distinguish *E. oleifera* and *E. guineensis*.

Heterozygosity is 1 way to measure the genetic diversity of the population. There are two types of heterozygosity, i.e. expected and observed. The expected heterozygosity of each primer was higher than the observed heterozygosity (Table 4). The expected heterozygosity showed the diversity of molecular marker. The value was calculated from each locus, assuming that the population has fulfilled Hardy-Weinberg equilibrium. Higher values indicated that the diversity of molecular marker is high. The observed heterozygosity, on the other hand, is directly calculated from observed data. Observed heterozygosity lower than the expected heterozygosity indicated that there was allele deviation in the population due to the inbreeding process. The inbreeding process followed by repeated selection of the population to increase the potential oil yield caused the Hardy-Weinberg law requirement to be unfulfilled. The heterozygosity values can be increased by the application of markers to larger populations with more diverse phenotypic characters (Chesnokov and Artemyeva, 2015).

The PIC value of each locus was 0.4, indicating that the locus was informative (Shete *et al.,* 2000). PIC value obtained were higher than in coconut (Larekeng *et al.,* 2018). PIC is a standard in molecular marker study to determine the genotypic range, including changes in single and more nitrogen bases. PIC value is 0 or close to 0 if there is no allele variation and can reach to 1 if genotyping has a new allele.

Codominant markers have the ability to detect polymorphism and are measured based on PIC value. A molecular marker with low PIC will be unable to detect polymorphism accurately (Lemos *et al.,* 2019). Factors that influence the PIC value are the size and number of the observed population. A good PIC value is determined by population selection and the consistency of molecular markers. In addition, the Shannon information index shows the number and distribution of allele in the population. A high Shannon index value indicates that the number and allele distribution are better.

Table 3. Selected SNAP primers.

S.No.	Primers	Base Sequence	Ta (°C)
1	P88 REF	CGAGAGGCGTCCGACAATGGG	64
	P88 ALT 9	CGAGAGGCGTCCGACAACGGA	64
	P88 REVERSE	ACGGTGGCGGTCTTGTTGATTCC	64
2	P207 REF	TTTCATCATTATTAGAATAAATTTAACTGTGAGCGA	54
	P207 ALT	TTTCATCATTATTAGAATAAATTTAACTGTGAG-TGG	56
	P207 REVERSE	GCCCTCTCCTACTGTTCTCATCTTCTTTCG	56
3	P248 REF	CAGGGTGATTGTCTGGAGTTCGTGG	58
	P248 REF	CCAGGGTGATTGTCTGGAGTTCTTGG	58
	P248 ALT	AATCCAGGGTGATTGTCTGGAGTTCATAC	58
4	P339 REF	TTCTAAAGAGCAGAGACAAAAGAGGGAGAAAAA G	62
	P339 REF	CATTCTAAAGAGCAGAGACAAAAGAGGGAGAATAAG	62
	P339 ALT	TCTAAAGAGCAGAGACAAAAGAGGGAGAACAAA	62

Table 4. Analysis of locus for SNAP primers.

S.No.	Locus	HObs	HExp	PIC	F(*Null*)	I	ChiSq
1	P88	0.3	0.503	0.445	0.2278	0.894	73.078*
2	P207	0.486	0.529	0.414	0.0271	0.799	73.018*
3	P248	0.389	0.55	0.455	0.1468	0.878	73.960*
4	P339	0.486	0.562	0.482	0.0363	0.920	75.623*
Average		0.415	0.536	0.449	0.1095	0.873	

HObs = Observed heterozygosity; HExp = Expected heterozygosity; PIC = Polymorphic information content; I = Shannon index of information; ChiSq = Chi Square test for Hardy-Weinberg equation with $P < 0.001$.

Further parameter is F null value. A negative value indicates poor allele distribution. If the F null value is 0, it indicates that there is no functional allele. If there is 0 allele, the allele will not be productive in providing information due to mutations.

A phylogenetic tree was constructed to determine the kinship among the accessions based on SNAP markers (Figure 5). A phylogenetic tree can be used in the parental selection of oil palm in the breeding program. Progenies with high genetic diversity can be obtained from crosses between genotypes having a distant genetic relationship. In contrast, a low genetic diversity is obtained from the crosses between parents that have a high genetic similarity. The SNP marker obtained from this research needs further analysis with a larger and distinct population. Together with SNAP markers from a SAD gene, SNAP markers of FATA gene can be used to predict oleic acid content of the oil palm progeny.

ACKNOWLEDGEMENTS

This research was supported by PT SMART Tbk. MDP was supported by LPDP (Indonesia Endowment Fund for Education) fellowship, Ministry of Finance, Republic of Indonesia.

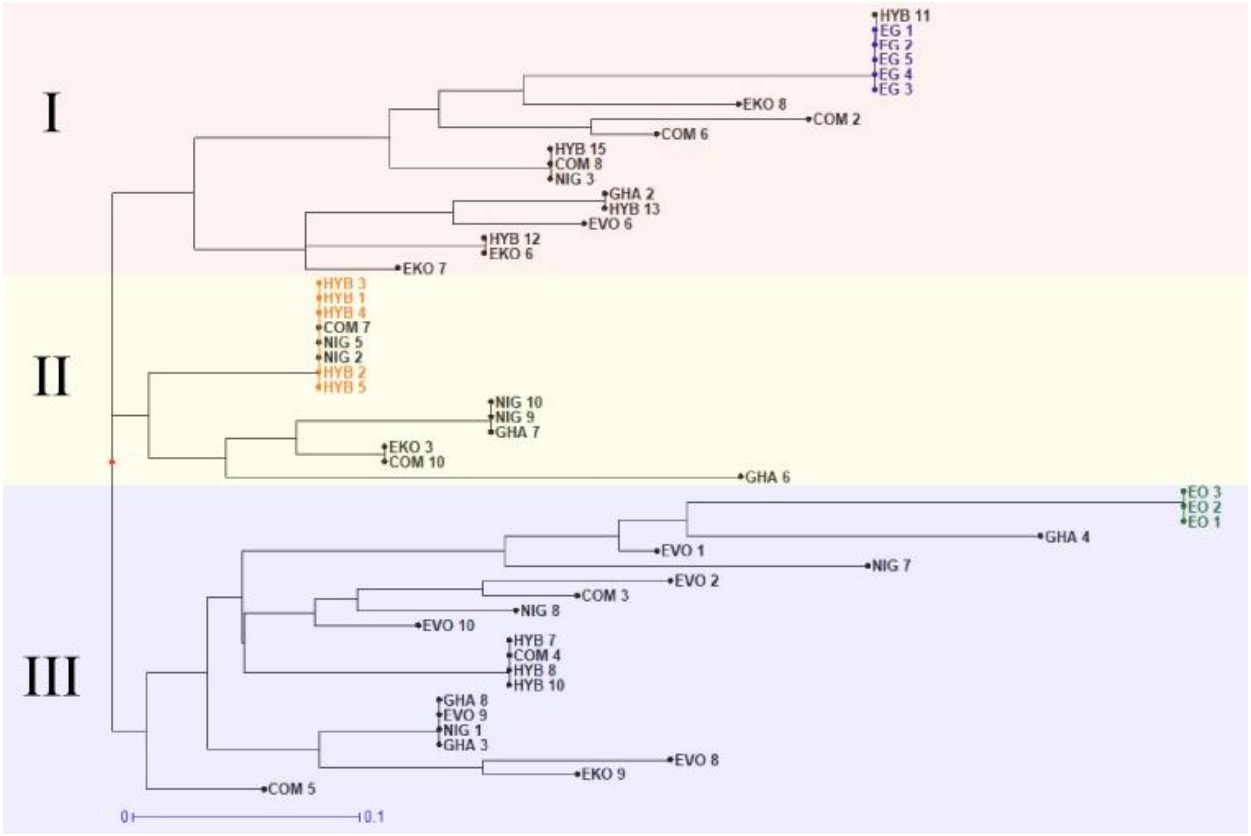

Figure 5. Phylogenetic tree of oil palm based on SNAP markers. Eo (*E. oleifera*), Eg (*E. guineensis*), HYB (*Eo* × *Eg* hybrid), COM (Compact), EKO (Ekona), NIG (Nigeria), GHA (Ghana), EVO (Evolution).

REFERENCES

Bailey TL, Williams N, Misleh C, Li WW (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucl. Acids Res*. 34: 369-373.

[USDA] United States Department of Agriculture (US) (2019). Oilseeds: World Markets and Trade. p. 19. http://apps.fas.usda.gov. Accessed on 11th December 2019.

Bais AS, Kaminski N, Benos PV (2011). Finding subtypes of transcription factor motif pairs with distinct regulatory roles. *Nucl. Acids Res.* 39 (11): e76-e76

Barcelos E, Rios SA, Cunha RNV, Lopes R, Motoike SY, Babiychuk E, Skirycz A, Kushnir S (2015). Oil palm natural diversity and the potential for yield improvement. *Front. Plant Sci.* 6: 190.

Bates PD, Stymne S, Ohlrogge J (2013). Biochemical pathways in seed oil synthesis. *Curr. Opin. Plant Biol*. 16: 358-364.

Borlay AJ, Suharsono, Roberdi, Liwang T (2017). Development of Single Nucleotide Polymorphism (SNP) marker for oleic acid content in oil palm (*Elaeis guineensis* Jacq.). *Pak. J. Biotechnol.* 14(1): 55-62.

Buckler ES, Thornsberry J (2002). Plant molecular diversity and applications to genomics. *Curr. Opin. Plant Biol*. 5: 107-111.

Chen Q, Zobel J, Verspoor K (2017). Benchmarks for measurement of duplicate detection methods in

nucleotide databases. *Database* 1-17.

Chesnokov YV, Artemyeva AM (2015). Evaluation of the measure of polymorphism information of genetic diversity. *Agric. Biol.* 50(5): 571-578.

Drenkard E, Richter BG, Rozen S, Stutius LM, Angell NA, Mindrinos M, Cho RJ, Oefner PJ, Davis RW, Ausubel FM (2000). A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. *Plant Physiol*. 124(4): 1483-1492.

Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJ, Hofmann K, Bairoch A (2002). The PROSITE database, its status in 2002. *Nucl. Acids Res*. 30(1): 235-238.

Jones A, Davies HM, Voelker TA (2005). Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. *Plant Cell* 7(3): 359-71.

Kim SG, Lee J-S, Shin S, Bae HH, Kim J-T, Son B-Y, Baek S-B (2016). Developing PCR-based SNP markers for distinguishing korean waxy corn F_1 hybrids. *Plant Breed. Biotechnol.* 4 (3): 315-323.

Koschmann J, Maschens F, Becker M, Niemeyer J, Schulze J, Bulow L, Stahl DJ, Hehl R (2012). Integration of bioinformatics and synthetic promoters leads to the discovery of novel elicitor-responsive cis-regulatory sequences in Arabidopsis. *Plant Physiol*. 160 (1): 178-191.

Kurt W, Nybom H, Pfenninger M, Wolff K, Kahl G. 2005. DNA Fingerprinting in Plants: Principles, Methods, and Applications. USA: CRC Press.

Kwong QB, Teh CK, Ong AL, Heng HY, Lee HL, Mohamed M, Low JZB, Apparow S, Chew FT, Mayes S (2016). Development and validation of a high-density SNP genotyping array for African oil palm. *Mol. Plant*. 9(8): 1132-1141.

Larekeng SH, Purwito A, Mattjik NA, Sudarsono S (2018). Microsatellite and SNAP markers used for evaluating pollen dispersal on Pati tall coconuts and Xenia effect on the production of 'Kopyor' fruits. In *IOP Conference Series: Earth and Environmental Science* (157(1):012042). IOP Publishing.

Lei A, Chen H, Shen G, Hu Z, Chen L, Wang J (2012). Expression of fatty acid synthesis genes and fatty acid accumulation in *Haematococcus pluvialis* under different stressors. *Biotechnol. Biofuels* 5(1): 18.

Lemos SC, Silveira RL, Buuron SK, dos Santos RS, Moro SC (2019). Determining the polymorphism information content of a molecular marker. *Gene* 726: 144175.

Lieb VM, Kerfers MR, Kronmüller A, Esquivel P, Alvarado A, Jiménez VM, Schmarr HG, Carle R, Schweiggert RM, Steingass CB (2017). Characterization of mesocarp and kernel lipids from *Elaeis guineensis* Jacq., *Elaeis oleifera* [Kunth] Cortés, and their interspecific hybrids. *J. Agric. Food Chem*. 65(18):3617-26.

Machens F, Becker M, Umrath F, Hehl R (2013). Identification of a novel type WRKY transcription factor binding site in elicitor-responsive cis-sequences from *Arabidopsis thaliana*. *Plant Molec. Biol.* 84 (4-5): 371-385.

Marshall TC, Slate JBKE, Kruuk LEB, Pemberton JM (1998). Statistical confidence for likelihood - based paternity inference in natural populations. *Mol. Ecol.* 7(5): 639-655.

Montoya C, Lopes R, Flori A, Cros D, Cuellar C, Summo M, Espeout S, Rivallan R, Risterucci AM, Bittencourt D, Zambrano JR, Alarcón WH, Villeneuve P, Pina M, Nouy B, Amblard P, Ritter E, Leroy T, Billotte N (2013). Quantitative trait loci (QTLs) analysis of palm oil fatty acid composition in an inter-

specific pseudo-backcross from *Elaeis oleifera* (H.B.K.) Cortés and oil palm (*Elaeis guineensis* Jacq.). *Tree Genet. Genome* 9: 1207-1225.

Moreno-Pérez AJ, Venegas-Calerón M, Vaistij FE, Salas JJ, Larson TR, Garcés R, Graham IA, Mart'inez-Force E (2012). Reduced expression of FatA thioesterases in Arabidopsis affects the oil content and fatty acid composition of the seeds. *Planta* 235(3): 629–639.

Mozzon M, Pacetti D, Lucci P, Balzano M, Frega NG (2013). Crude palm oil from interspecific hybrid *Elaeis oleifera* × *Elaeis guineensis*: Fatty acid regiodistribution and molecular species of glycerides. *Food Chem*. 141(1): 245-252.

Perrier X, Jacquemoud-Collet JP (2015). DARwin software. 2006.

Pesik A, Efendi D, Novarianto H, Dinarti D, Sudarsono S. 2017. Development of SNAP markers based on nucleotide variability of WRKY genes in coconut and their validation using multiplex PCR. *Biodiversitas J. Biol. Divers*. 18: 465–475.

Pootakham W, Jomchai N, Ruang-areerate P, Shearman JR, Sonthirod C, Sangsrakru D, Tragoonrung S, Tangphatsornruang S (2015). Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics* 105(5-6): 288-295.

Rafalski A (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol*. 5: 94–100.

Rajanaidu N, Kushairi A, Rafii M, Mohd DA, Maizura I, Jalani BS (2000). In: Advances in oil palm oil palm breeding and genetic resources. *Malaysian Palm Oil Board* 1: 171–237.

Rosli R, Amiruddin N, Ab Halim MA, Chan P-L, Chan K-L, Azizi N, Morris PE, Low E-TL, Ong-Abdullah M, Sambanthamurthi R, Singh R, Murphy DJ (2018). Comparative genomic and transcriptomic analysis of selected fatty acid biosynthesis genes and CNL disease resistance genes in oil palm. PLOS ONE 13(4): e0194792.

Serrano-Vega MJ, Garcés R, Martinez-Force E (2005). Cloning, characterization and structural model of a FatA-type thioesterase from sunflower seeds (*Helianthus annuus* L.). *Planta* 221(6): 868-80.

Shete S, Tiwari H, Elston RC (2000). On estimating the heterozygosity and polymorphism information content value. *Theor. Popul. Biol*. 57(3): 265–271.

Singh R, Ong-Abdullah M, Low ETL, Manaf MAA, Rosli R, Nookiah R, Ooi LCL, Ooi SE, Chan KL, Halim MA, *et al.,* (2013). Oil palm genome sequence reveals divergence of inter fertile species in Old and New worlds. *Nature* 500(7462): 335.

Smouse RPPE, Peakall R (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28(19): 2537–2539.

Wong CK, Bernardo R (2008). Genome wide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theor. Appl. Genet*. 116 (6): 815-824.

Xu Q, Wen X, Deng X. 2007. Cloning of two classes of PR genes and the development of SNAP markers for powdery mildew resistance loci in chestnut rose (*Rosa roxburghii* Tratt). *Mol. Breed*. 19(2): 179-191.

Yang Q, Fan C, Guo Z, Qin J, Wu J, Li Q, Fu T, Zhou Y. 2012. Identification of FAD2 and FAD3 genes in *Brassica napus* genome and development of allele-specific markers for high oleic and low linolenic acid contents. *Theor. Appl. Genet*. 125(4): 715-729.