Development and Characterization of Simple Sequence Repeats Derived from Mitochondrial Genome of Oil Palm Using Next Generation Sequencing

Pichahpuk Uthaipaisanwong

Interdisciplinary Graduate Program in Genetic Engineering, The Graduate School, Kasetsart University, Bangkhen Campus, Ladyao, Chatuchak, Bangkok 10900

Suthasinee Somyong, Sithichoke Tangphatsornruang and Thippawan Yoocha

National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Phaholyothin Road, Khlong Nueng, Khlong Luang, Pathum Thani 12120

Chatchawan Jantasuriyarat*

Department of Genetics, Faculty of Science, Kasetsart University, Bangkhen Campus, Ladyao, Chatuchak, Bangkok, 10900

Abstract

The advancement of the high throughput sequencing and computational technologies allow us to perform genome sequencing and analyzing the genome sequence. In this study, the oil palm mitochondrial (mt) sequences (258 contigs) were assembled from reads that aligned to the homologous regions of date palm mt genome (covered ~66.7 %). The 51 from 258 oil palm mt contigs contained 78 SSRs with a variety of repeat motifs. Mononucleotide repeats were the largest group (80.7 %) consisting of A/T (98.4 %) and G/C (1.6 %) motifs. Trinucleotide repeats formed the second most repeats (8.9 %) consisting of TTA/TAA (33.2 %), TTC/GAA (16.7 %), TAG/CTA (16.7 %), TAT/ATA (16.7 %) and CTA/TAG (16.7 %). This was followed by the dinucleotide repeats (7.7 %) consisting of AT/TA (100 %) and tetranucleotide repeats (2.7 %), which consist of TATT/AATA (50 %) and ATTT/AAAT (50 %). Ten mt simple sequence repeat (mtSSR) markers were developed from the candidate mt sequences of *E. guineensis* to study the genetic diversity among fifteen plant species distributed in different subfamilies in Arecaceae. Nine of ten markers, developed from oil palm mt genome sequence, were transferable and polymorphic in 15 Arecaceae species. The average number of alleles detected by these ten mtSSR markers was 7 ranged from 4 to 14. The average polymorphic information content (PIC) score for the mtSSRs was 0.766. This is the first report revealing the use of mtSSR, developed from oil palm mt sequences, as a resource for the genetic diversity study in Arecaceae species.

*Corresponding author: fscicwj@ku.ac.th

DOI: 10.14456/tjst.2017.2

Keywords: oil palm; mitochondria; high throughput sequencing; simple sequence repeat

1. Introduction

Oil palm (*Elaeis guineensis* Jacq.) is an important oil crop in the world. Oil palm is a monocot plant. It is a member of the Arecaceae or Palmae family, which contains 183 genera and 2,364 species (Dransfield *et al*., 2005; Gomez-Navarro *et al*., 2009). Palm oil is an important component in food industry, e.g. as liquid cooking ingredients and margarines and is used in the production of cosmetics and pharmaceutical industry (Soh *et al*., 2009). Particularly economical importance, in the last decade, demand for palm oil as a source of biodiesel, which is an alternative energy source, has been grown very fast with the high rise in petroleum price. Oil palm produces the highest oil yields compared with other oil producing plants (3.5 ton/ha/year), and accounts for 36 % of world production (Bourgis *et al*., 2011). Since the 1990s, the global oil palm harvested areas are increase to 17 Million hectare in 2012 (Pirker *et al*., 2016). Of these, Malaysia and Indonesia are the top two countries increasing oil palm plantation area follows by Nigeria and Thailand. However, the limitation of the oil palm planting areas is driving the destruction of the rainforest in Southeast Asia, especially, Thailand (Vijay *et al*., 2016). Therefore, the improvement of the yield and quality of oil palm is required for sufficient need of the oil. Initially, the conventional plant breeding has been used to improve the productivity of oil palm. However, oil palm conventional breeding is time consuming with approximately 10-12 years for a cycle with low efficiency process due to the long reproductive cycle and seed maturation period. This method also requires large planting area and expensive process (Mayes *et al*., 2000). Molecular biotechnology approaches, such as marker assisted selection linked to useful traits such as yield, shell thickness and embryogenesis rate, can be carried out to reduce time for conventional breeding and may prove to be a better way to further improve productivity of oil palm (de Touchet *et al*., 1991; Mayes *et al*., 2000). The molecular markers were used not only to study marker-assisted selection but also as the efficient tools to study population structure, molecular ecology, and genetic diversity. Even though the research mainly focus on using the molecular markers designed from oil palm nuclear DNA (ncDNA), the other two organellar DNAs, mitochondrial DNA and chloroplast DNA, are also valuable for molecular biotechnology applications.

Mitochondria are widely known that they evolved from endosymbiosis of cyanobacteria cell and proteobacterium-like cell into the common ancestor of eukaryotes more than a billion year ago. This endosymbiosis theory was published by Lynn Margulis in 1970 (Archibald, 2011). The flowering plant mitochondrial DNAs (mtDNA) have high copy number and are usually distributed among multiple organelles per cell. Technically, high mtDNA copy number is relatively easy to amplify because it appears in

multiple copies in cell whereas the ncDNA is two copies for most somatic cells in diploid plants. In addition, the variable DNA regions of mitochondria are suitable molecular marker to study genetic diversity in the species-specific level. The mtDNA in most angiosperms, including oil palm, are uniparentally-inherited from maternal parent. Their haploid genome is only one allele per cell and per organism, which is not perturbed by the occurrence of recombination between two alleles (Galtier *et al*., 2009). These characteristics can reduce the difference in the genetic variability among population and can represent the within-species history of mtDNA, which traces back the original resource and geographic movement of maternal lineages (Avise *et al*., 1987). The importance of the study of oil palm genetic diversity is the identification of oil palm populations, exploration of plant genetic resources, protection of germplasm resources, and development of future breeding programs (Zhou *et al*., 2015).

The SSR marker is one of the most efficient molecular markers in detecting the polymorphism among crosses and closely related lines due to their co-dominant inheritance, a multi-allelic nature, and high mutation rate (Tangphatsornruang *et al*., 2009; Zhou *et al*., 2015). Especially, it is easily assayed using PCR technique, which requires small amount of DNA and provides the advantages of reliability, reproducibility, discrimination, and ease of scoring (Smith *et al*., 1997). Moreover, the SSR marker, which is transferable across taxa, is advantageous as they save time and cost in developing SSR markers for members of taxa that have not been extensively studied (Lin *et al*., 2014).

An advance of massively parallel DNA sequencing approaches, applied reversible terminator technology for rapid and accurate large-scale sequencing, is feasible to sequence the whole genomic DNA including nuclear and organellar DNA (Wheeler *et al*., 2008). Therefore, in this study, we developed the mitochondrial SSR (mtSSR) markers in oil palm using the Illumina sequencing platform and study the transferability of these mtSSR markers among fifteen plant species in Palmae family.

2. Materials and Methods

2.1 Plant materials

Fifteen species in Arecaceae used in this work included *Areca catecuhu* L. (dwarf betel nut palm), *Veitchia merrillii* (manila palm), *Cyrtostachys renda* Blume (sealing wax palm), *Chrysalidocarpus lutescens* (yallow palm), *Cocos nucifera* L. (coconut palm), *Phoenix dactylifera* L. (date palm), *Livistona rotundifolia* (Lam.) Mart (anahaw palm), *Copernicia macroglossa* (cuban petticoat palm), *Calamus* sp. (rattan), *Borassus flabellifer* L. (palmyra palm), *Nypa fruticans* Wurmb. (atap palm), *Corypha umbraculifera* L. (fan palm), *Caryota mitis* Lour. (burmese fishtail palm), *Arenga pinnata* (Wurmb) Merr. (sugar palm), *Elaeis guineensis* (oil palm).

2.2 DNA extraction

Genomic DNA of fifteen plant species in this study was extracted from fresh leaf

tissues by CTAB method (Doyle and Doyle, 1987; Doyle and Doyle, 1990). The method was first performed from cutting and grinding oil palm fresh leaves in liquid nitrogen with pestle. The powder sample was transferred to 1.5 mL tube (1/3 of tube). 600 µL of extraction buffer, 10 µL of beta-mercaptone, 20 µL of 20 % SDS were added and mixed in the tubes. The sample was incubated at 65 °C for 20-30 min and mixed once after 10 min. The sample was left at room temperature for 1 min and added with 1 volume of chloroform : isoamyl alcohol (24:1). The tubes were gently mixed and centrifuged at 12,000 rpm at room temperature for 10 min. The aqueous phase was transferred into new 1.5 mL tube with avoiding transferring any chloroform. One volume of cold isopropanol was added and mixed by inverting tubes 20-30 times. To collect DNA pellet, the solution was centrifuged at 12,000 rpm at room temperature for 5 min and then supernatant was discarded. The DNA pellet was cleaned by adding 1 mL of 70 % EtOH by inverting tubes 5-10 times and then was centrifuged at 12,000 rpm at room temperature for 1 min. The supernatant was discarded and the tube was inverted on a clean kimwipe to allow the DNA pellet to dry for 30-45 min. Finally, the DNA pellet was dissolved with 50 µL TE.

2.3 SSR markers discovery and primer design

Total genomic DNA of oil palm dura fruit type (R14/15D) was extracted from fresh leaf tissue by CTAB method. The genomic sample was sequenced using the Illumina Hiseq platform by Macrogen, Inc. Korea. The paired-end reads were generated from a sequencing library with 8 kb inserts. The adaptor sequences and low quality bases (quality < 15) were trimmed by Trimmomatic v27 (Bolger *et al*., 2014). The remaining paired-end reads were mapped against the reference sequence of date palm (*P. dactylifera*) complete mitochondrial genome (accession number NC_016740.1) (Yang *et al*., 2010) with GS Reference Mapper software v2.8 (Margulies *et al*., 2005). Oil palm candidate mt sequences, receiving from mapping, were then run through MISA software to identify SSR (microsatellite) sequences under the following conditions: repeat units must be at least 10 nucleotides for mononucleotide repeats, at least 14 nucleotides for di-nucleotide repeats, and at least 15 nucleotides for tri, tetra, penta and hexa nucleotide repeats. Primers were designed following the parameters as primer size between 18-25 bp with a GC content of 50-60 %, and primer melting temperature (Tm) of 55-60 °C, and PCR product size of 180-250 bp. The mtSSR primers were blasted through the oil palm nuclear DNA sequence from The Malaysia Palm Oil Board (Singh *et al*., 2013) and the oil palm chloroplast genome (accession number NC_017602.1) (Uthaipaisanwong *et al*., 2012) for checking the primer specificity.

2.4 Marker validation

The SSR specific loci that succeeded to design primer were used to amplify with genomic DNA of *E. guineensis* for primers test and other fourteen species among Arecaceae family for mtSSR markers validation. PCRs were performed under the conditions of a final volume

of 10 µL consisting of 1 ng of DNA template, 1 Unit of iTaq™ DNA polymerase, 1 x iTaq™ reaction buffer that included Mg^{+2}, 0.25 mM dNTPs, 10 µM of each forward and reverse primer and distilled water to the PCR reaction to 10 µL. The condition of amplification the mtSSR markers was in the thermocycler as pre-incubation at 94 °C for 2 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, and elongation at 72 °C for 1 min and a final extension step at 72 °C for 10 min. PCR products were resolved on 1 % agarose gel in 0.5X TBE buffer (100V for 40 min) for presence of fragments corresponding to the expected sizes. The polymorphism were identified by separating the PCR fragments and detected the exact DNA fragments length by DNA Analyzer based on capillary gel electrophoresis technology (BIOptics Qsep100 dna-CE, Taiwan, China). Here, we have determined the number of allele and the polymorphism information content (PIC) for each primer to evaluate the genetic diversity. The PIC value was calculated according to the formula: $PIC_i = 1-\sum P_{ij}^2$, where, P_{ij}, where, P ij is the frequency of the jth allele for the ith marker (Abedian and Talebi, 2012).

3. Results and Discussion

3.1 SSR marker discovery

We sequenced the oil palm dura to obtain about 81 million (100 bp) paired-end reads, which were generated from an 8-kb sequencing library. The adaptor sequences and low-quality bases were removed from the raw reads. Trimmed and cleaned reads (78 million paired-end reads) were subsequently mapped to the date palm mitochondrial genome with the GS Reference Mapper software. One million and two hundred thousand reads (1,268,317 reads) mapped to the reference genome covered approximately 66.7 % of the reference date palm mt genome. Two hundred and fifty-eight contigs were assembled from reads that aligned to the homologous regions of the reference sequences. A total of 78 SSRs were identified with a frequency of 0.164 SSR per kb in 258 mt genome contigs of oil palm (in total size 476,688 bp). The frequency distribution of mtSSR in oil palm mitochondrial sequences was shown in (Table 1). Mononucleotides were the most abundant repeat type (80.78 %), and showed a strong bias toward the A/T repeat-motifs. The second most prevalent repeats were tri-nucleotides (8.97 %), followed by di-nucleotide (7.69 %) and tetra-nucleotide (2.56 %). The nucleotide repeats were compared between oil palm mt sequences in this study and date palm mt genome (Yang *et al*., 2010) under the same criteria. The most abundant nucleotide repeat in oil palm mt genome is mononucleotides. The A/T (98.4 %) motifs were the most abundant followed by C/G (1.6 %) as reported in the date palm mt genome. Tri-nucleotides were the second most abundant repeat in oil palm mt genome, which comprise TTA/TAA (33.2 %), TTC/GAA (16.7 %), TAG/CTA (16.7 %), TAT/ATA (16.7 %) and CTA/TAG (16.7 %). Whereas, the second most abundant repeat in date palm mt genome was di-nucleotides. The most abundant in di-

nucleotides were AT/TA (100 %) motifs, which were the most abundant repeats as reported in date palm mt genome. Moreover, the most abundant in tetra-nucleotide motifs were AAAT/ATTT as date palm mt genome. Total 499,254,157 bp (218,183 assembled contigs) was generated from genomic DNA of oil palm dura using illumine HiSeq platform. The SSR motifs were searched in these dataset and 130,840 potential SSRs were identified. The A/T, AT/TA, and TTA/TAA motifs were the most abundant repeats in mono-, di-, and tri-nucleotides respectively. While the ACAT/ATGT motifs were the most abundant in genome of oil palm. Total nucleotide repeats containing motifs between one and six nucleotides in size were occurred at a frequency of 0.262 SSR per kb of sequences (Taeprayoon *et al*., 2015). Moreover, an available 41,374 oil palm genome sequences from NCBI, 3,024 genomic-SSRs (gSSRs) were detected in oil palm sequences. Total nucleotide repeats were occurred at a frequency of 0.22 gSSR per kb of sequences (Palliyarakkal *et al*., 2011). The frequency of SSR repeat of oil palm mt sequence (0.164 SSR per kb) is different to the frequency of SSR genome sequences.

Table 1 Frequency and distribution of SSRs in *E. guineensis* mitochondrial genome sequence

SSR motif	Number of repeat units												Total
	5	6	7	8	9	10	11	12	13	14	15	>15	
Mononucleotide													
A/T	-	-	-	-	-	43	11	5	2	1	-	1	62
C/G	-	-	-	-	-	1	-	-	-	-	-	-	1
Di-nucleotide													
AT/AT	-	-	1	2	1	-	-	-	-	-	-	-	4
TA/TA	-	-	1	-	-	1	-	-	-	-	-	-	2
Tri-nucleotide													
GAG/CTC	1	-	-	-	-	-	-	-	-	-	-	-	1
TTA/TAA	1	-	-	-	-	-	1	-	-	-	-	-	2
TTC/GAA	1	-	-	-	-	-	-	-	-	-	-	-	1
TAG/CTA	1	-	-	-	-	-	-	-	-	-	-	-	1
TAT/ATA	-	1	-	-	-	-	-	-	-	-	-	-	1
CTA/TAG	1	-	-	-	-	-	-	-	-	-	-	-	1
Tetra-nucleotide													
TATT/AATA	1	-	-	-	-	-	-	-	-	-	-	-	1
ATTT/AAAT	-	-	1	-	-	-	-	-	-	-	-	-	1
Total													78

3.2 Primer design

Fifty-four mononucleotides were shorter than 12 bp and were excluded from the analysis. Primer pairs were successfully designed for 14 mtSSRs by Primer3 software, and the amplified product sizes ranged between 180 to 250 bp with a GC content of 50-60 %, and primer melting temperature (Tm) of 55-60 °C. None of these primer pairs were able to match the oil palm genome sequence and the oil palm complete cp genome, suggesting that the fourteen primer pairs were specific to oil palm mt sequence. Of these 14 markers, ten markers were randomly selected for empirical validation with genomic DNA of oil palm dura to test the PCR condition (Table 2). Primers were synthesized by Integrated DNA Technologies, Inc. (IDT).

Table 2 Mitochondrial SSR markers and PCR product sizes (base pairs, bp) for determining the polymorphism between/within oil palm species.

Marker ID	SSR type	SSR size (bp)	SSR location	Date palm Position	Product size (bp)	Primer length (bp)	Seq 5' to 3'
mtSSR01	(GAG)5	15	IGS	176922	180	20	AGCGACAATACCGTACACCA
				176936		20	AGATCAAAGTGTCCGTCCGT
mtSSR02	(A)13	13	IGS	234497	200	20	GGTCCGGTCCATCTTTTCCT
				234509		20	CTACATTCAACTGTGCCCCG
mtSSR03	(TTC)5	15	IGS	366967	203	20	AGACCGACCTACCTACCTGT
				366981		20	AGTCATTCCATTCCTTCCGG
mtSSR04	(CTA)5	15	IGS	400614	238	20	ATGAAAGCTAAGCGGGAAGG
				400628		20	AAAGGGCCTTGTAAGACCGA
mtSSR05	(TA)7	14	IGS	443795	224	20	TTGGGCCCTATTCGTTCGAT
				443808		20	GGCAAAATGTTCAGTAGCCC
mtSSR06	(AT)6	12	IGS	453075	225	20	GATAATTCTGGCCGGGTTCG
				453087		20	GCGGCCTACAATGTTGAGAC
mtSSR07	(TAG)5	15	IGS	655695	182	20	GGTCAGAATGGATCCAGGGA
				665722		20	ACAGCAACGTGTGGAATCTG
mtSSR08	(A)16	16	IGS	665707	230	20	AGTTTCATTAGCTCCCGAGG
				665722		20	TCGGCAATAGTGTCCCTACC
mtSSR09	(A)14	14	IGS	688873	237	20	ACACCAATTGAGCTCCAGGA
				688886		20	TGAGCCCACGGATACAGATC
mtSSR10	(AT)7-(N)13-(AT)8	43	IGS	711959	211	18	TGCAGTGAACCGCAATCA
				712001		20	ATAAGTCCAGCCCCTGTGTC

3.3 Marker validation

To test the effectiveness of ten developed mtSSR markers, the markers were amplified with the genomic DNA sample of oil palm dura and visualized their products with 1 % agarose gel electrophoresis. The optimal PCR

condition for these mtSSR markers had been varied between annealing temperature and DNA template concentration. To examine the optimal PCR parameters, the amplification products for each mtSSR marker should appear as a single locus-specific band in genomic DNA of E. guineensis under the condition at 55 °C annealing temperature and 1 ng of DNA template at final concentration.

The successful mtSSR markers in primer amplification testing were used to amplify with 15 species among Arecaceae family including dwarf betel nut palm, manila palm, sealing wax palm, yellow palm, coconut palm, date palm, anahaw palm, petticoat palm, rattan, palmyra palm, nypa palm, talipot palm, fishtail palm, sugar palm, and oil palm. In a cross-taxa transferability testing, nine mtSSR markers were successfully amplified in fifteen species from the Arecaceae. All mtSSR markers showed polymorphism among the fifteen species in Arecaceae with PIC score in range of 0.4933 to 0.893 as shown in Table 3. The conclusion of ten markers amplified cross 15 species transferability in Arecaceae is shown in Table 4. In a cross-taxa amplification testing, six mtSSR markers (mtSSR01, mtSSR02, mtSSR03, mtSSR08, mtSSR09, and mtSSR10) amplified a single band through 15 Arecaceae species because of single specific target region. While other four markers (mtSSR04, mtSSR05, mtSSR06, and mtSSR07) amplified multiple bands in some Arecaceae species, due to non-specific amplification, duplicated, or repetitive nucleotide sequences within their mt genome. Besides, mtDNA can be transferred and integrated into chloroplast or nuclear genome, which can cause multiple bands. Even though, nine of ten markers were completely cross-transferable (100 %) and showed the polymorphism among 15 Arecaceae species, the mtSSR09 was able to amplify only 12 of 15 species, which is 80 % cross-transferable. Three species that mtSSR09 cannot amplify included yellow palm, coconut palm, and fishtail palm which are in subfamily Arecoideae, Arecoideae, and Coryphoideae respectively. Of these, coconut is in the same subfamily as oil palm. No amplification may be due to sequence variation diverging from the oil palm mt genome. Comparison with the previous study of eleven of the American oil palm gSSR markers which produced the banding profiles of various sizes in both American and African oil palm were used to evaluate cross transferable in Arecaceae. These gSSR marker set was successful amplify (100 % transferability) in the African oil palm and coconut palm, while the frequencies of transferability of other palms were in range of 45.5-72.7 % (Zaki *et al*., 2012). Even though, these gSSR marker set provides the various size of PCR products, the mean PIC score for the gSSR was 0.402 which lower than the mtSSR from this study. These results demonstrated that the mtSSR marker development method by searching the different microsatellite variation between oil palm and date palm (different genus) increases the potential and efficiency of transferability of SSR marker across taxa. Moreover, the SSR markers, which were

developed from organellar genomes, provide higher genetic diversity more than the nuclear genome sequence among Arecaceae species. Although the Arecaceae contains over 2,300 members only the few species are important economic plants. Of these, only major economic plants (coconut, oil palm and date palm) are available of genome sequence for using molecular marker development. The cross-taxa transferability of mtSSR markers can enrich the mtSSR markers, which provide the tool for study the genetic diversity within and between related species in Arecaceae. Primers designed based on conserved sequences can generate the complete mt sequence. Therefore, the mtSSR markers, which provided a single locus-specific band (mtSSR01, mtSSR02, mtSSR03, mtSSR08, mtSSR10) and can amplify through 15 Arecaceae species, might be available for screening and assemble mt genome in other Arecaceae species.

Table 3 Polymorphic information content (PIC), allele number and size range of 10 successful mitochondrial markers.

Marker ID	Number of alleles	PIC value	Size range (bp)
mtSSR01	4	0.6489	168-180
mtSSR02	8	0.8444	189-197
mtSSR03	4	0.4933(min)	189-198
mtSSR04	4	0.6666	232-244
mtSSR05	11	0.8540	208-246
mtSSR06	7	0.7111	218-338
mtSSR07	6	0.8265	176-206
mtSSR08	13	0.8933 (max)	207-275
mtSSR09	7	0.8333	222-234
mtSSR10	8	0.8533	187-223

5. Conclusion

This study used an efficient method, high throughput sequencing technology, for developing mitochondrial SSR markers in oil palm. Seventy-eight putative SSR loci were identified. Mononucleotide of A/T repeats was the most abundant, following with the dinucleotide of AT/TA repeats. Ten of seventy-eight mtSSR markers were tested in oil palm DNA samples, and all of them were successfully amplified. In a cross-taxa transferability testing, nine mtSSR markers were successfully amplified in fifteen species from the Arecaceae. All mtSSR markers showed polymorphism among the fifteen species in Arecaceae with PIC score in range of 0.4933 to 0.8933. The transferability of

developed mtSSR exhibits a good resource for genetic diversity analysis in 15 Arecaceae species. This is the first time to report the mtSSR markers that could be a new resource for the genetic diversity study in Arecaceae species.

Table 4 The amplified products of ten markers (mtSSR01-10) across 15 species transferability in Arecaceae were separated by capillary gel electrophoresis using Qsep100 DNA analyzer. Numbers represent of PCR product base pair size of fragments in base pairs: NA = not amplifiable; * = highest concentration. More than one product size means multiple PCR products.

Sample ID	Betel palm	Manila palm	Sealing wax palm	Yellow palm	Coconut palm	Date palm	Anahaw palm	Petticoat palm	Rattan	Palmyra palm	Nypa palm	Talipot palm	Fishtail palm	Sugar palm	Oil palm
mtSSR01	180	174	168	177	177	177	180	174	177	174	177	174	177	177	174
mtSSR02	195	193	193	194	197	189	192	192	190	190	192	191	190	189	193
mtSSR03	195	192	195	195	198	195	195	189	198	192	192	192	195	195	192
mtSSR04	46, 107, 161*, 329	60, 102, 241*, 335, 680	32, 68, 87, 161*, 252, 296, 436, 541, 838	37, 238*, 281, 865	238	232	238	238	238	238*, 241*, 284	241	238	107*, 244*, 480, 690	238	238
mtSSR05	222	216	218	214	220	210, 226, 246*	218	224	212*, 285*, 487	216	216	218	208	212	214
mtSSR06	224	222	222	224	179, 298, 338*	218	220	220	224	62*, 336, 486	73, 217, 228*	228	41, 71, 180, 227, 307, 710*	226	222
mtSSR07	188	188	188	188	200	176	176	176	176	182	206	182	100, 180, 413, 736*	169, 198*	191
mtSSR08	224	214	222	221	225	215	213	207	239	237	228	224	231	275	215
mtSSR09	226	230	226	228	229	228	NA	NA	NA	234	234	228	222	232	232
mtSSR10	199	213	213	199	203	203	187	205	203	193	193	187	223	193	211

6. Acknowledgements

We acknowledge the funding support by the National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Thailand (NSTDA) and Agricultural Research Development Agency (ARDA). We acknowledge the scholarship support by Thailand Graduate Institute of Science and Technology (TGIST), NSTDA. We acknowledge the oil palm germplasm support by the Golden Tenera Company Limited located in Karbi, Thailand.

7. References

Abedian, M. and Talebi, M., 2012, Chloroplast Microsatellite Diversity among and within *Prunus mahaleb* L. and *P. avium* L. Species, J. Agric. Sci. 4: 191-202.

Archibald, J.M., 2011, Origin of eukaryotic cells: 40 years on, Symbiosis 54: 69-86.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. and Saunders, N.C., 1987, Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics, Ann. Rev. Ecol. Syst. 18: 489-522.

Bolger, A.M., Lohse, M. and Usadel, B., 2014, Trimmomatic: A flexible trimmer for illumina sequence data, Bioinformatics 30: 2114-2120.

Bourgis, F., Kilaru, A., Cao, X., Ngando-Ebongue, G.F., Drira, N., Ohlrogge, J.B. and Arondel, V., 2011, Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning, Proc. Natl. Acad. Sci. USA. 108: 12527-12532.

de Touchet, B., Duval, Y. and Pannetier, C., 1991, Plant regeneration from embryogenic suspension cultures of oil palm (*Elaeis guineensis* Jacq.), Plant Cell Rep. 10: 529-532.

Doyle, J.J. and Doyle, J.L., 1987, A rapid DNA isolation procedure for small quantities of fresh leaf tissue, Phytochem. Bull. 19: 11-15.

Doyle, J.J. and Doyle, J.L., 1990, Isolation of Plant DNA from fresh tissue, Focus 12: 13-15.

Dransfield, J., Uhl, N.W., Asmussen, C.B., Baker, W.J., Harley, M.M. and Lewis, C.E., 2005, A new phylogenetic classification of the palm family, Arecaceae, Kew Bull. 60: 559-569.

Galtier, N., Nabholz, B., Glemin S. and Hurst, G.D., 2009, Mitochondrial DNA as a marker of molecular diversity: A reappraisal, Mol. Ecol. 18: 4541-4550.

Gomez-Navarro, C., Jaramillo, C., Herrera, F., Wing, S.L. and Callejas, R., 2009, Palms (Arecaceae) from a Paleocene Rainforest of Northern Colombia, Am. J. Bot. 96: 1300-1312.

Lin, Y., Lu, J.J., Wu, M.D., Zhou, M.B., Fang, W., Ide, Y. and Tang, D.Q., 2014, Identification, cross-taxon transferability and application of full-length cDNA SSR markers in *Phyllostachys pubescens*, Springerplus 3: 486.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis,

G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley R.F. and Rothberg, J.M., 2005, Genome sequencing in microfabricated high-density picolitre reactors, Nature 437: 376-380.

Mayes, S., Jack, P.L. and Corley, R.H., 2000, The use of molecular markers to investigate the genetic structure of an oil palm breeding program, Heredity (Edinb) 85 Pt 3: 288-293.

Palliyarakkal, M.K., Ramaswamy, M. and Vadivel, A., 2011, Microsatellites in palm (Arecaceae) sequences, Bioinformation 7: 347-351.

Pirker, J., Mosnier, A., Kraxner, F., Havlik, P. and Obersteiner, M., 2016, What are the limits to oil palm expansion?, Glob. Environ. Change Hum. Pol. Dimens. 40: 73-81.

Singh, R., Ong-Abdullah, M., Low, E.T., Manaf, M.A., Rosli, R., Nookiah, R., Ooi, L.C., Ooi, S.E., Chan, K.L., Halim, M.A., Azizi, N., Nagappan, J., Bacher, B., Lakey, N., Smith, S.W., He, D., Hogan, M., Budiman, M.A., Lee, E.K., DeSalle, R., Kudrna, D., Goicoechea, J.L., Wing, R.A., Wilson, R.K., Fulton, R.S., Ordway, J.M., Martienssen, R.A. and Sambanthamurthi, R., 2013, Oil palm genome sequence reveals divergence of interfertile species in old and new worlds, Nature 500: 335-339.

Smith, J.S.C., Chin, E.C.L., Shu, H., Smith, O.S., Wall, S.J., Senior, M.L., Mitchell, S.E., Kresovich, S. and Ziegle, J., 1997, An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): Comparisons with data from RFLPs and pedigree, Theor. Appl. Genet. 95: 163-173.

Soh, A.C., Wong, C.K., Ho, Y.W. and Choong, C.W., 2009, Oil Crops (Handbook of Plant Breeding), Springer, New York.

Tangphatsornruang, S., Somta, P., Uthaipaisanwong, P., Chanprasert, J., Sangsrakru, D., Seehalak, W., Sommanas, W., Tragoonrung, S. and Srinives, P., 2009, Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean [*Vigna radiata* (L.) Wilczek], BMC Plant Biol. 9: 137.

Taeprayoon, P., Tanya, P., Kang, Y.J., Limsrivilai, A., Lee, S.H. and Srinives, P., 2015, Genome-wide SSR marker development in oil palm by Illumina HiSeq for parental selection, Plant Genet. Resour. 14: 157-160.

Uthaipaisanwong, P., Chanprasert, J., Shearman, J.R., Sangsrakru, D., Yoocha, T., Jomchai, N., Jantasuriyarat, C., Tragoonrung, S. and Tangphatsornruang, S., 2012, Characterization of the chloroplast genome sequence of oil palm (*Elaeis guineensis* Jacq.), Gene 500: 172-180.

Vijay, V., Pimm, S.L., Jenkins, C.N. and Smith, S.J., 2016, The impacts of oil palm on recent deforestation and biodiversity loss, PLoS One 11(7): e0159668.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W.,

Chen, Y.J., Makhijani, V., Roth, G.T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C.L., Irzyk, G.P., Lupski, J.R., Chinault, C., Song, X.Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D.M., Margulies, M., Weinstock, G.M., Gibbs, R.A. and Rothberg, J.M., 2008, The complete genome of an individual by massively parallel DNA sequencing, Nature 452: 872-876.

Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., Zhao, D., Al-Mssallem, I.S. and Yu, J., 2010, The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.), PLoS One 5(9): e12762.

Zaki, N.M., Singh, R., Rosli, R. and Ismail, I., 2012, *Elaeis oleifera* genomic-SSR markers: Exploitation in oil palm germplasm diversity and cross-amplification in Arecaceae, Int. J. Mol. Sci. 13: 4069-4088.

Zhou, L.X., Xiao, Y., Xia, W. and Yang, Y.D., 2015, Analysis of genetic diversity and population structure of oil palm (*Elaeis guineensis*) from China and Malaysia based on species-specific simple sequence repeat markers, Genet. Mol. Res. 14: 16247-16254.