# Genetic diversity analysis of Tenera × Tenera and Tenera × Pisifera Crosses and D self of oil palm (*Elaeis guineensis*) parental populations originating from Cameroon

**LALU FIRMAN BUDIMAN[1], ARDHA APRIYANTO[2], ADI PANCORO[3], SUDARSONO SUDARSONO[4,♥]**

[1]PT. Astra Agro Lestari Tbk. Jl. Pulo Ayang Raya, Blok OR-I, Kawasan. Industri Pulo Gadung, Jakarta, Indonesia.
[2]Biotechnology Laboratory, PT. Astra Agro Lestari Tbk. Jl. Pulo Ayang Raya, Blok OR-I, Kawasan. Industri Pulo Gadung, Jakarta, Indonesia.
[3]Genetics Laboratory, School of Life Science and Technology, Institut Teknologi Bandung. Jl. Ganesha 10, Bandung, West Java, Indonesia.
[4]PMB Lab., Departement of Agronomy and Horticulture, Faculty of Agriculture, Institut Pertanian Bogor. Jl. Raya Dramaga, IPB Campus, Bogor 16680, West Java, Indonesia. Tel.: +62-251-8629354, Fax.: +62-251-8629352, ♥email: s_sudarsono@ymail.com

**Abstract.** *Budiman LF, Apriyanto A, Pancoro A, Sudarsono S. 2019. Genetic diversity analysis of Tenera × Tenera and Tenera × Pisifera Crosses and D self of oil palm (*Elaeis guineensis*) parental populations originating from Cameroon. Biodiversitas 20: 937-949.* There are three types of oil palm (*Elaeis guineensis* Jacq.) based on the shell thickness, such as the Dura (D type, with a thick shell), the Pisifera (P type, with no or very thin shell) and the Tenera (T type, with medium shell thickness), respectively. The T type is a commercially grown oil palm, originated from hybridization between D × P types. The success of oil palm breeding depends on the availability of diverse parental populations, especially in the D and the P types. Unfortunately, the improved P type of oil palm may only be produced by crossing between Tenera (T × T) or between Tenera and Pisifera (T × P) while improved D type may easily be produced from selfing of a single Dura type palm (D self). Therefore, evaluation of the potential genetic diversity of Dura parental lines derived from D self and Pisifera lines derived from T × T or T × P is essential. The objectives of this research were to analyze the genetic diversity of T × T, T × P and D self oil palm progenies originated from Cameroon which would be used as parental population for breeding the commercial T types of oil palm in Indonesia, determine whether the progenies were from legitimate hybrids of the desired parents and evaluate their potential values for creating Tenera hybrid in the oil palm breeding programs. A total of 148 individuals from one combination of T × T and two T × P crosses and three D self-pollinations were evaluated. Genotyping was conducted using 16 SSR marker loci. The genotype data were analyzed using software for population genetic and genetic diversity analysis. Results of the analysis indicated the evaluated 16 SSR marker loci were either highly or moderately polymorphic based on their Polymorphic Information Content (PIC) values. Hence, they could be used for genetic diversity analysis of the evaluated oil palm progenies. Both the T × T and T × P progenies were more diverse than the D self-ones. Clustering and Principle Component Analysis (PCA) showed that all populations were grouped into three groups consisting of (1) B02 – T × P progenies, (2) B57 – T × T progenies, and (3) the rest of the populations (a mixture of the B01 – T × P progenies, and the three D self progenies). Moreover, the third group was further divided into five sub-groups, consisting of sub-group 3.1: the B01 progenies, and sub-group 3.2 to 3.5 comprising of a mix of individuals from members of at least two different D self progenies. All the studied T × T and T × P progenies could potentially be used as improved male parents for producing future Tenera oil palm hybrid varieties. The T x T and T × P progenies had a wider genetic distance than that of the D self progenies. Moreover, for practical breeding purposes, the members of D self oil palm progenies should not be grouped based only on the family but should be based on the results of the clustering analysis. The reported data should be beneficial for aiding future oil palm breeding in Indonesia.

**Keywords:** African oil palm, Dura, Pisifera, population structure, Simple Sequence Repeat

## INTRODUCTION

Palm oil is a major vegetable oil producing crop in the world. Palm oil supplies at least 32.9% of the total world vegetable oil demand while the rest is from other vegetable oils such as soybean (29.4%), rapeseed (16.0%), sunflower (9.1%), nut (3.2%), cotton (2.4%), and minor vegetable oils (9%) (Statista 2015). The demand outlook for palm oil is most probably still increasing in the years to come. It is predicted that in 2050, oil palm could fulfill all demands of world vegetable oil, reaching 240 million tons (Corley 2009; Barcelos et al. 2015).

The 2014 data from the Directorate General of Plantations, Republic of Indonesia (Ditjenbun 2014) indicated there was a rapid increase in oil palm plantation areas in Indonesia. The total area of oil palm plantations in 2004 was only 5,284,723 ha while in 2012 it was 10,956,231 ha. However, the Indonesian government through the Presidential Instruction Republic of Indonesia number 10/2011 has implemented a moratorium on forest to plantation land conversion since 2011 (Ditjenbun 2013). Moreover, the availability of arable land suitable for growing oil palm has also become a limiting factor for the opening of new oil palm plantations (Danielsen et al. 2009). Therefore, meeting the increasing future demand for palm oil has to come from more productive planting materials while those better yielding planting materials should come from effective oil palm breeding programs (Barcelos et al. 2015).

One of the critical success factors in generating more productive and superior quality oil palm seedlings as outputs of breeding programs is the availability of diverse oil palm parental materials. Diverse oil palm parental materials provide oil palm breeders with raw materials for developing new and improved cultivars with desirable characteristics (Govindaraj et al. 2015). Therefore, understanding genetic diversity and population structure of the oil palm parental materials is an important aspect of oil palm breeding (Thongthawee et al. 2010; Barcelos et al. 2015). Breeding for superior oil palm varieties is generally conducted through the reciprocal recurrent selection (RRS) method (Purba et al. 2000). In one generation of RRS, potential female (Dura [D] type) and male (Pisifera [P] or Tenera [T] type) parental lines are cross-hybridized to generate D × P or D × T (D × P/T) progenies. The D × P/T progenies are evaluated for their superior characters in the field. The cycle is continually repeated to eventually identify improved parental lines capable of producing new superior D × P/T cultivars.

In the RRS method, the presence of illegitimate (contaminant) progenies (progenies derived from unidentified parents) can occur in any stage of the controlled crosses in the oil palm breeding activities, from the initial stage of selecting and labeling parents to the final stage of field evaluation trials. Many factors affect the percentage of illegitimacy or contamination in oil palm breeding such as the biology of the plant flower (due to the existence of a few hermaphrodite flowers), human errors during pollen collection and damage of the pollination bag for covering the female flowers because of animals and the environment factors allowing insect pollinators to cross-pollinate the female flowers (Hama-Ali et al. 2015).

Improved parental lines in RSS oil palm breeding methods should not contain illegitimate progenies; therefore, identification of illegitimacy in the improved parental line populations is important. Moreover, new superior varieties should only be obtained from hybridization of genetically diverse parental populations. Therefore, generating basic information about the genetic diversity of parental populations is also necessary to identify the parental line combinations having the largest potential of hybrid vigor (Arias et al. 2012).

Phenotype variation and molecular markers have been used to evaluate genetic diversity of perennial crops such as cacao, coconuts, and oil palm (Moose and Mumm 2008; Tornincasa et al. 2010; Zulhermana et al. 2010; Ajambang et al. 2012; Solin et al. 2014; Ajijah et al. 2015; Maskromo et al. 2015). In oil palm, genetic diversity has been evaluated using RAPD, ISSR, AFLP, and SSR markers (Purba et al. 2000; Billotte et al. 2001, Lim and Rao 2005; Zulhermana et al. 2010; Ajambang et al. 2012). Although they are informative for studying the genetic diversity of oil palm, RAPD, ISSR, and AFLP are the dominant markers. Therefore, there are some limitations on the use of RAPD, ISSR and AFLP markers in genetic diversity because those markers are unable to differentiate the genotype of heterozygous and homozygous individuals (Okoye et al. 2016).

Simple-sequence repeats (SSRs), also known as microsatellites, are tandem repeats of two to four nucleotide motifs that are widely distributed in the eukaryotic genomes (Delseny et al. 1983; Tautz and Renz, 1984; Tautz, 1989; Billotte et al. 2001). The SSR markers derived from PCR amplification of SSR sequences have a high degree of allelic variability; therefore, it is a powerful tool for genetic analysis, germplasm identification and intra-or interspecific genetic mapping of many plants, including the genus Elaeis. The reasons SSR markers are preferable in plant genetic analysis include: (i) their existence is abundant and uniformly distributed in the genome, (ii) they are highly polymorphic and are co-dominant markers, (iii) they are relatively easy and fast to generate using the PCR technique and it is relatively simple and easy to interpret the allele configurations, and (iv) they are relatively easy to be accessed by other laboratories if the SSR sequences are available (Saghai-Maroof et al. 1994). The SSR marker is a co-dominant marker; therefore, it can differentiate genotypes of heterozygous and homozygous individuals (Akkaya et al. 1992; Jones et al. 2010). The SSR marker is also advantageous because of its highly discriminative character which makes it capable of detecting polymorphism even among closely related members of a population (Mohammadi and Prasanna 2003). It is also highly reproducible, relatively low-cost and requires minimum training for its implementation. More importantly, SSR markers can be used to analyze large numbers of samples since it is possible to automate genotyping using SSR markers (Hayden et al. 2008). The SSR marker has been used to evaluate oil palm genetic diversity and population structures of breeding materials and parental lines (Thongthawee et al. 2010; Okoye et al. 2016) and natural populations of African (*E. guineensis*) and American (*E. oleifera*)-origin oil palms (Ting et al. 2010).

To develop parental lines for generating new superior oil palm varieties, PT. Astra Agro Lestari Tbk. has set up collaborative activities with the Institute of Agricultural Research for Development (IRAD) of Cameroon by introducing selected Dura (D self) and Pisifera (T × P or T × T) parental lines, and the Tenera (D × P) progeny testing lines between the year of 2008 and 2011. All the parental lines (D self; T × P or T × T) and the progeny testing lines (D × P) were evaluated under Indonesian environmental conditions. Although they were generated through a comprehensive breeding program, the genetic background of these populations has never been evaluated using molecular tools. The objectives of this research were to analyze the genetic diversity of T × T, T × P and D self oil palm progenies originated from Cameroon which will be used as parental population for breeding of the commercial T types of oil palm in Indonesia, determine whether the progenies were from legitimate hybrids of the desired parents and evaluate their potential values for creating Tenera hybrid in the oil palm breeding programs in Indonesia.

## MATERIALS AND METHODS

### Plant materials

The data collection activities were conducted between April 2014 and April 2015 at the Biotechnology Laboratory, PT. Astra Agro Lestari (AAL) Tbk., Pangkalan Bun, Central Kalimantan. Samples were collected from the nursery and experimental station area of PT. AAL in Pangkalan Bun, Central of Kalimantan. A total of 148 individuals from six different crosses were evaluated. The populations consisted of one combination of a T × T and two T × P crosses and three D self progenies (Table 1). The number of evaluated progenies in each hybridization combination ranged from 20 to 73 individuals. Leaf tissues were sampled from the spear leaf of three-year-old plants in the trial field station.

### DNA extraction and SSR marker amplification

Total DNA was extracted from a total of 100 mg of leaf samples following standard protocols using a commercial plant DNA extraction kit (Geneaid). The RNA was removed from the whole nucleic acid by standard RNAse treatment (Sambrook et al. 1989). The quality and quantity of the extracted DNA were evaluated using a Nanodrop ND-1000 spectrophotometer (Nano-Drop Technologies Inc.) and by electrophoresis of 2 ml of the sample DNA in 1% agarose gel. Sixteen SSR marker loci, developed by CIRAD and available through TropGENE Database (http://tropgenedb.cirad.fr/) were used to genotype all of the evaluated individuals. The list of primers used to generate 16 SSR marker loci is presented in Table 2. These primers have previously been reported capable of generating polymorphic and informative SSR markers in oil palm (Billotte et al. 2001). To facilitate automatic fragment size detection, amplification of each SSR marker locus was done using FAM, ROX, HEX, and TAMRA fluorescent labeled (Schuelke 2000; Hayden et al. 2008) reversed primers. For each locus, the SSR marker PCR amplification was conducted using standard procedures as described in Kapa 2G Fast PolymeraseTM PCR Kit (Kapa Biosystems). The PCR amplification steps included one cycle of DNA denaturation at 95 °C for 3 mins, 35 cycles consisting of denaturation at 95 °C for 15 seconds, primer annealing at the appropriate temperature depending on each primer pair for 15 seconds, and primer extension at 72 °C for 5 seconds, and one cycle of final extension at 72 °C for 10 minutes. Fragment amplification was conducted using a 96-well Veriti Thermal cycler (Applied Biosystems). Analysis of the amplified fragments of all SSR loci was sub-contracted to a third-party genotyping provider (First Base Laboratories, Malaysia) and was conducted using the 96-capillary ABI Prism 3730 XL DNA Analyzer (Applied Biosystems Inc, USA) with the GENESCAN-500 (Liz) size standard.

### Data analysis

Allele scoring data provided by the third-party genotyping provider (First Base Laboratories, Malaysia) were organized for analysis using Microsoft Office Excel and the appropriate data files were prepared for subsequent analysis using various software for genetic diversity, population parameters, and population structure analysis. Population genetic parameters were analyzed using CERVUS version 3.0 (Kalinowski et al. 2007) and GENALEX version 6.3 (Peakall and Smouse 2012). Analysis of molecular variance (AMOVA) (Excoffier et al. 1992) using GENALEX version 6.3 (Peakall and Smouse 2012) was done to calculate the partition of the total genetic variation between and within the populations. Principal component and Neighbor-Joining analyses were done using DARWIN Software version 6.0.013 (Perier and Jacquemoud-Collet 2006), while the population structures were determined using the STRUCTURE software version 2.3.4 (Falush et al. 2003) and STRUCTURE HARVESTER version Web 0.6.94 (Earl et al. 2012). The approach by Evanno et al. (2005) was used to select the appropriate K clusters for the studied populations. To validate the presence of illegitimate individuals among the studied populations, the genotype data were subjected to analysis using the COLONY software (Jones and Wang 2010).

**Table 1.** List of the type of oil palm populations evaluated in this research consisting of individuals derived from Tenera/Pisifera crosses with La Me and La Me × SP540 genetic background and D self

| Family code | Genetic background | | | | Progenies sample no. | Crosses type[a] |
|---|---|---|---|---|---|---|
| | Female parent | Male parent | Female grand parent | Male grand parent | | |
| B01 | LM16844 | LM16844 | LM2T | LM2T | 20 | T × P |
| B02 | LM19029 | LM19029 | LM2T | LM2T | 36 | T × P |
| B57 | LM18228 | LM18228 | (LM5Tx LM10T) X RS3T AF | (LM5Tx LM10T) X RS3T AF | 29 | T × T |
| A140 | PO6690 | PO6690 | DA 115D | DA 115D | 23 | D self |
| A125 | PO6729 | PO6729 | DA10D x DA115D | DA10D x DA115D | 20 | D self |
| A127 | PO6844 | PO6844 | DA10D x DA3D | DA10D x DA3D | 20 | D self |

Note: [a] The T × P – cross-pollination between different palms of Tenera and Pisifera types. The T × T-cross pollination between different Tenera palms. The female (T) and the male (P or T) parents with the same number indicated they were siblings derived from the same grand-parents. The D self – self-pollination of a single palm of a Dura type, because the Dura type palm has both male and female flowers in a single palm

## RESULTS AND DISCUSSION

### Allele numbers in the evaluated oil palm population

In the combined six populations (all populations), the evaluated 16 loci of simple sequence repeat (SSR) markers generated using 16 primer pairs resulted in a total of 134 DNA bands (SSR alleles), with an average of 8 DNA bands (SSR alleles) per locus (Table 2). In the combined populations, both the mEgCIR0038 (LG 13) and mEgCIR3292 (LG 15) SSR loci generated the highest number of alleles/locus (11 alleles) (Table 2). On the other hand, both the mEgCIR2813 (LG 5) and mEgCIR0353 (LG 16) SSR loci generated the lowest alleles/locus (5 alleles) (Table 2). For the evaluated T × P/T progenies, the total numbers of alleles were 113, with an average of 7 alleles/locus. On the other hand, for the D self progenies, the total number of alleles was 90, with an average of 6 alleles/locus (Table 2). Moreover, the total number of alleles in each progeny ranged from 40 to 74 alleles while the average ranged from 3 to 5 alleles per locus (Table 2).

The observed number of alleles per locus in these oil palm populations is comparable to those reported previously for either improved or wild oil palm populations. Taeprayoon et al. (2015) reported the number of alleles per locus ranged from 3 to 8 with an overall average of 5.45 for the improved oil palm breeding population in Thailand while Okoye et al. (2016a, b) reported the average number of allele per locus was 2.96 for NIFOR Tenera materials from Nigeria. Okoye et al. (2016a, b) also reported the mean number of alleles in the entire population for each locus varied from 1.33 in Madagascar oil palm germplasm to 4.89 alleles per locus in NIFOR Tenera materials. In another study, Bakoume et al. (2007) indicated the presence of eight to 22 alleles per locus in the African wild oil palm populations.

The evaluated populations are derived from individual palm to palm crosses (Table 1); therefore, the expected number of alleles for each locus within each population depended on the genotype of the parents. There would be seven possible models of segregation based on Mendelian genetics, as presented in Table 3. The expected number of alleles per population and the expected allele frequencies in each model cross are also presented (Table 3).

The number of alleles per locus for some loci in the evaluated oil palm populations was higher than the expected values of the model crosses. Based on the hybridization model (Table 3), the expected number of alleles/locus ranged from 1 to 4 while in the evaluated populations, they ranged from 1 to 7 alleles/locus. Moreover, some loci showing the expected number of alleles per locus did not show the expected allele segregation ratios for each population (Table 4). This finding was similar to those of Okoye et al. (2016a, b) who evaluated the genetic diversity of NIFOR oil palm main breeding parent genotypes using SSR markers. Okoye et al. (2016a,b ) also found that alleles of the studied loci in the evaluated oil palm populations were also more than the expected and the allele frequencies did not fit the expected segregation ratios.

**Table 2.** The observed number of alleles for each SSR marker locus among loci in the evaluated oil palm populations.

| Locus | Linkage group | Number of alleles of each locus for the progeny populations[a] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | A125 | A127 | A140 | B01 | B02 | B57 | D Self[b] | TxT/P[c] | All[d] |
| mEgCIR0802 | 1 | 3 | 4 | 5 | 4 | 2 | 3 | 6 | 7 | 9 |
| mEgCIR3282 | 2 | 5 | 4 | 3 | 4 | 4 | 3 | 5 | 9 | 10 |
| mEgCIR0173 | 3 | 2 | 2 | 2 | 5 | 3 | 1 | 3 | 6 | 6 |
| mEgCIR3533 | 4 | 5 | 4 | 4 | 5 | 5 | 2 | 5 | 7 | 9 |
| mEgCIR2813 | 5 | 4 | 4 | 2 | 3 | 3 | 2 | 4 | 4 | 5 |
| mEgCIR3543 | 6 | 7 | 7 | 4 | 5 | 2 | 3 | 8 | 5 | 8 |
| mEgCIR0894 | 7 | 5 | 5 | 5 | 6 | 5 | 3 | 6 | 9 | 9 |
| mEgCIR0886 | 8 | 6 | 6 | 3 | 2 | 2 | 3 | 8 | 5 | 9 |
| mEgCIR3886 | 9 | 2 | 3 | 3 | 3 | 6 | 3 | 3 | 6 | 6 |
| mEgCIR3785 | 10 | 4 | 4 | 6 | 5 | 3 | 3 | 6 | 10 | 10 |
| mEgCIR3362 | 11 | 6 | 6 | 5 | 4 | 3 | 3 | 7 | 7 | 9 |
| mEgCIR2414 | 12 | 4 | 4 | 3 | 5 | 4 | 3 | 5 | 9 | 9 |
| mCnCIR0038 | 13 | 8 | 7 | 5 | 3 | 6 | 2 | 9 | 8 | 11 |
| mEgCIR3546 | 14 | 4 | 5 | 5 | 5 | 4 | 2 | 5 | 8 | 8 |
| mEgCIR3292 | 15 | 4 | 6 | 5 | 5 | 4 | 3 | 6 | 10 | 11 |
| mEgCIR0353 | 16 | 1 | 3 | 3 | 2 | 3 | 1 | 4 | 3 | 5 |
| Total alleles | | 70 | 74 | 63 | 66 | 59 | 40 | 90 | 113 | 134 |
| Average | | 4 | 5 | 4 | 4 | 4 | 3 | 6 | 7 | 8 |

Note: [a] A125, A127 and A140 were progenies of the D self; B01 and B02 – the progenies of T × P crosses; B57 – the progenies of the T × T cross. [b] D self – the total progenies of A125, A127 and A140. [c] The T × T/P – the total progenies of B01, B02 and B57; [d] All – the total progenies of A125, A127, A140, B01, B02 and B57.

**Table 3.** Expected allele frequencies based on Mendelian genetics model crosses of two parents

| Crosses between parents | | Expected alleles | |
|---|---|---|---|
| Model crosses | Examples of parent genotypes | No. | Frequency |
| Het. parents with different alleles | AB x CD | 4 | 1: 1: 1: 1 |
| Het. parents with one similar allele | AB x AC | 3 | 2: 1: 1 |
| Het. parents with two similar alleles | AB x AB | 2 | 1: 1 |
| Het, & Hom, parents with different alleles | AB x CC | 3 | 1: 1: 2 |
| Het. & Hom parents with one similar allele | AB x AA | 2 | 3: 1 |
| Hom, & Hom, parents with different alleles | AA x BB | 2 | 1: 1 |
| Hom. & Hom parents with similar alleles | AA x AA | 1 | 1 |

The number of alleles per locus for some loci in the evaluated oil palm populations is larger than the expected values of the model crosses. Based on the hybridization model (Table 3), the expected number of alleles/locus ranges from 1-4 while in the evaluated populations, they range from 1-7 alleles/locus. Moreover, some loci showing the expected number of alleles per locus do not show the expected allele segregation ratios for each population (Table 4). This finding was similar to those of Okoye et al. (2016a, b) who evaluated the genetic diversity of NIFOR oil palm main breeding parent genotypes using SSR markers. Okoye et al. (2016a, b) also found alleles of the studied loci in the evaluated oil palm populations were also more than the expected and the allele frequencies did not fit to the expected segregation ratios.

To validate the presence of illegitimate individuals among the studied populations, the genotype data were subjected to analysis using the COLONY software to infer the individual parentage and sibship from multilocus genotypes data. Results of the analysis (Figure 1) indicate that no illegitimate progeny were found among individuals belonging to the T × T/P crosses (B01, B02, or B57). The inference from the multilocus genotype data indicated that all individuals belonging to the T × T/P crosses were derived from a pair of parents, i.e. T1 and T2 for B01, T3, and T4 for B02, and T5 and T6 for B57 oil palm populations (Figure 1). On the other hand, not all of the D

self individuals were legitimate progenies of their respective parents. The individuals belonging to the D self populations were developed from the selfing of three Dura progenitors (Table 1). However, results of the COLONY analysis identified five progenitors for individuals belonging to the D self populations, namely A125, A127, and A140 populations. (Figure 1). Such results pointed out further possibilities of the presence of illegitimate progenies among the studied oil palm populations, especially for the D self populations.

**Table 4.** Loci having a number of alleles per locus and ratio of allele frequencies that were either fit or unfit the expectations based on the model crosses for each studied oil palm progeny populations

| Family | Total loci | Loci having number of allele/locus | | Loci having ratio of allele frequency[1] | |
|---|---|---|---|---|---|
| | | Fit to expectation | Unfit to expectation | Fit to expectation | Unfit to expectation |
| B01 | 16 | 8 | 8 | 2 | 6 |
| B02 | 16 | 12 | 4 | 6 | 6 |
| B57 | 16 | 16 | 0 | 8 | 8 |
| A140 | 16 | 9 | 7 | 5 | 4 |
| A125 | 16 | 9 | 7 | 4 | 5 |
| A127 | 16 | 9 | 7 | 3 | 6 |

Note: [1] Ratio of allele frequency were evaluated only for loci having a number of alleles fit to the expectations.
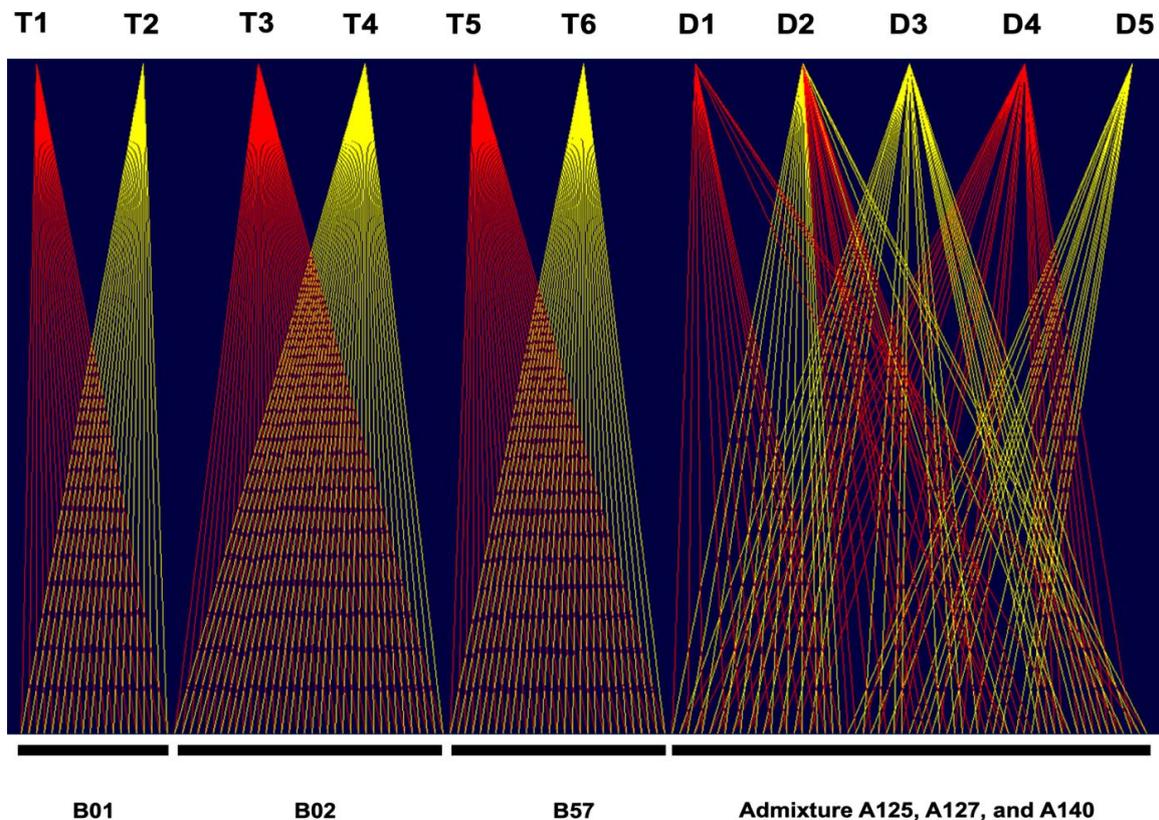


**Figure 1.** The most likely pedigree construction from COLONY. Identified parent IDs on the top bar and offspring IDs at the bottom. Red lines emanating downwards are from male parents and yellow lines are from females. T1, T2, T3, T4, T5, and T6 were the predicted parents of B01, B02, and B57 progeny populations. D1, D2, D3, D4, and D5 were the predicted parents of the admixture of A125, A127, and A140 progeny populations

## Mean allelic patterns and PIC Values

Out of the 16 simple sequence repeat (SSR) marker loci evaluated, 100% were polymorphic in the B01, B02, A140, or A127 populations. For the B57 and A125 populations, the percentages of polymorphic loci were 88% and 94%, respectively. The average percentage of polymorphic loci for all populations was 97%. In all populations, both the mEgCIR0038 (LG 13) and mEgCIR3292 (LG 15) SSR marker loci showed the highest average polymorphic information content (PIC=0.84). Meanwhile, the mEgCIR3785 (LG 10) showed the highest PIC in T x T/P (PIC=0.79) and the mEgCIR0038 (LG 13) was the highest in the D self population (PIC=0.79), (Table 5). The mEgCIR0353 (LG 16) SSR marker locus (Table 5) showed the lowest PIC in all the populations (PIC=0.41), in the T x T populations (PIC=0.39), and in the D self population (PIC=0.15). For all the populations, the average PIC over 16 loci of SSR markers was 0.74 while that for the evaluated T x T population was 0.70 and for D self was 0.56 (Table 6). Meanwhile, the average PIC in each locus for each of the evaluated oil palm populations ranged from 0.39-0.55 (Table 6).

According to Okoye et al. (2016a, b), the marker loci having a PIC value > 0.7 is informative while a PIC value between 0.4-0.7 is moderately informative and less than 0.4 is not informative. A high value of PIC for each locus indicates the locus is informative in distinguishing individuals within populations and among populations (Sajib et al. 2012). In the studied oil palm populations, most of the loci either belonged to informative or moderately informative categories. The 11 loci evaluated were identified as informative loci having a PIC > 0.7 and four loci were moderately informative having a PIC between 0.4 and 0.7 (Table 5).

The observed PICs in the studied D Self, T × T/P, and All populations were also higher than that reported by Solin et al. (2014) for D × P oil palm population. The results of Okoye et al. (2016a, b) revealed the existence of a high percentage of polymorphic loci (83.3%) in the entire oil palm materials with an average PIC=0.73. They also demonstrated that two SSR marker loci (sMg00016 and sMg00179) showed PIC=0.80 and PIC=0.85, respectively.

**Table 6.** The mean values of allele per locus, effective allele (Ne), private allele, locally common alleles present in < 50% of the populations, and polymorphic information content (PIC) in each locus for each oil palm population

| Popula -tions[a] | Total allele numbers | Mean values of | | | | |
|---|---|---|---|---|---|---|
| | | Allele/ locus | Ne | Private alleles | Common alleles (<50%) | PIC |
| A125 | 65 | 4.38 | 2.4 | 0.1 | 1.2 | 0.53 |
| A127 | 58 | 4.63 | 2.4 | 1.0 | 1.0 | 0.55 |
| A140 | 40 | 3.94 | 2.0 | 0.4 | 1.8 | 0.51 |
| B01 | 63 | 4.13 | 2.7 | 0.2 | 1.3 | 0.47 |
| B02 | 69 | 3.69 | 2.7 | 0.1 | 1.5 | 0.47 |
| B57 | 74 | 2.50 | 2.9 | 0.1 | 1.6 | 0.39 |
| D Self | 63 | 5.63 | - | - | - | 0.56 |
| TxT/P | 85 | 7.06 | - | - | - | 0.70 |
| All | 148 | 8.38 | - | - | - | 0.74 |

Note: [b] A125, A127 and A140 were progenies of the D self; B01 and B02 – the progenies of T × P crosses; B57 – the progenies of the T × T cross; D Self – the total progenies of A125, A127 and A140; The T × T/P – the total progenies of B01, B02 and B57; All – the total progenies of A125, A127, A140, B01, B02 and B57.

**Table 5.** The polymorphic information content (PIC) for each SSR marker locus among evaluated oil palm populations

| Locus | Linkage Group[a] | Polymorphic Information Content (PIC) for each locus among populations[b] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | A125 | A127 | A140 | B01 | B02 | B57 | D self | TxT/P | All |
| mEgCIR0802 | 1 | 0.45 | 0.5 | 0.69 | 0.43 | 0.33 | 0.48 | 0.58 | 0.69 | 0.77 |
| mEgCIR3282 | 2 | 0.53 | 0.53 | 0.43 | 0.47 | 0.56 | 0.47 | 0.5 | 0.76 | 0.8 |
| mEgCIR0173 | 3 | 0.33 | 0.22 | 0.11 | 0.53 | 0.08 | - | 0.26 | 0.52 | 0.49 |
| mEgCIR3533 | 4 | 0.64 | 0.49 | 0.46 | 0.45 | 0.61 | 0.25 | 0.54 | 0.7 | 0.76 |
| mEgCIR2813 | 5 | 0.5 | 0.45 | 0.3 | 0.44 | 0.4 | 0.37 | 0.42 | 0.65 | 0.66 |
| mEgCIR3543 | 6 | 0.58 | 0.58 | 0.61 | 0.71 | 0.37 | 0.57 | 0.61 | 0.63 | 0.69 |
| mEgCIR0894 | 7 | 0.68 | 0.58 | 0.67 | 0.54 | 0.55 | 0.46 | 0.66 | 0.76 | 0.79 |
| mEgCIR0886 | 8 | 0.59 | 0.69 | 0.48 | 0.05 | 0.38 | 0.48 | 0.71 | 0.7 | 0.8 |
| mEgCIR3886 | 9 | 0.34 | 0.57 | 0.47 | 0.21 | 0.59 | 0.35 | 0.54 | 0.74 | 0.79 |
| mEgCIR3785 | 10 | 0.62 | 0.55 | 0.65 | 0.62 | 0.54 | 0.48 | 0.62 | 0.79 | 0.82 |
| mEgCIR3362 | 11 | 0.65 | 0.7 | 0.64 | 0.54 | 0.55 | 0.52 | 0.7 | 0.76 | 0.83 |
| mEgCIR2414 | 12 | 0.51 | 0.59 | 0.54 | 0.67 | 0.58 | 0.54 | 0.55 | 0.77 | 0.78 |
| mCnCIR0038 | 13 | 0.73 | 0.76 | 0.6 | 0.5 | 0.68 | 0.37 | 0.79 | 0.78 | 0.84 |
| mEgCIR3546 | 14 | 0.59 | 0.61 | 0.66 | 0.65 | 0.61 | 0.37 | 0.63 | 0.75 | 0.77 |
| mEgCIR3292 | 15 | 0.69 | 0.76 | 0.71 | 0.59 | 0.56 | 0.47 | 0.73 | 0.76 | 0.84 |
| mEgCIR0353 | 16 | - | 0.24 | 0.19 | 0.05 | 0.15 | - | 0.15 | 0.39 | 0.41 |

Note: [a] The SSR marker loci and linkage group (LG) were based on Billotte et al. (2001). [b] A125, A127 and A140 were progenies of the D self; B01 and B02 – the progenies of T × P crosses; B57 – the progenies of the T × T cross; D self – the total progenies of A125, A127 and A140; The T × T/P – the total progenies of B01, B02 and B57; All – the total progenies of A125, A127, A140, B01, B02 and B57

**Table 7.** Allele classification according to allelic frequencies based on the approach of Buchert et al. (1997) (high, intermediate, low and rare) and Marshall and Brown (1975) (common and rare)

| Populations | Na | High P> 0.75 | Intermediate 0.75>P>0.25 | Low 0.25>P>0.01 | Rare P<0.01 | Common P>0.05 | Rare P<0.05 |
|---|---|---|---|---|---|---|---|
| B01 | 65 | 3 (4.6) b | 22 (33.8) | 40 (61.5) | 1 (1.5) | 50 (76.9) | 15 (23.1) |
| B02 | 58 | 2 (3.4) | 23 (39.7) | 3 (5.2) | 1 (1.7) | 47 (821.0) | 11 (19.5) |
| B57 | 40 | 3 (7.5) | 23 (57.5) | 14 (35.0) | 0 (0) | 39 (97.5) | 1 (2.5) |
| A140 | 63 | 3 (4.8) | 23 (36.5) | 37 (58.7) | 0 (0) | 52 (82.5) | 11 (17.5) |
| A125 | 69 | 1 (1.4) | 27 (39.1) | 41 (59.4) | 0 (0) | 56 (81.2) | 13 (18.8) |
| A127 | 74 | 2 (2.7) | 26 (35.1) | 45 (60.8) | 1 (1.4) | 60 (81.1) | 14 (18.9) |

[a] N, number of total alleles per population. [b] The x (y), x indicates the number of alleles and y indicates the percentage of the total alleles in each population

## Allelic frequency across populations

The calculated allele frequency for the studied oil palm population was grouped into either high (P≥ 0.75), intermediate (0.75>P≥0.25), low (0.25>P≥0.01) or rare (P<0.01) based on the system proposed by Buchert et al. (1997). It can also be grouped as either common (P≥0.05) or rare (P<0.05) according to Marshall and Brown (1975). The groupings of allele frequency in each locus for each of the studied oil palm population are presented in Table 7. In all of the studied oil palm populations (Table 7), only 1-3 (.4-7.5%) alleles were grouped as high and 0-1 (0-1.7%) as rare while the majority of alleles were either intermediate (22-27 or 33.8-57.5% alleles) or low frequency (3-45 or 5.2-61.5% alleles) according to the criteria proposed by Buchert et al. (1997). The B57 population showed the highest percentage of intermediate allele frequency (57.5%) while B01 showed the highest percentage of low allele frequency (61.5%). On the other hand, the B01 population showed the lowest percentage of intermediate (33.8%) and B02 population the lowest percentage (5.2%) of low allele frequency (Table 7). However, the majority of allele frequencies (Table 7) belonged to the common group and the rest belonged to the rare group according to the criteria developed by Marshall and Brown (1975).

In the oil palm population in the present study, rare alleles (P<0.05) were observed at a level ranging from 2.5% to 23.1% across loci in each population. Those numbers were lower than those found among wild oil palm populations, in which rare alleles (P<0.05) were found in as much as 38% and populations having similar agro-ecology were also reported to share rare alleles. It was suggested that natural selection had a larger impact on the percentages of rare alleles in the population than the genetic drift (Arias et al. 2013). Rare alleles have also been reported to be related to plant adaptation to abiotic and biotic stresses (Rajora et al. 2000). Therefore, it is important to maintain genetic variability in oil palm populations since it increases the possibility of having adaptive responses to biotic and abiotic stress (Maxted et al. 2006; Maxted et al. 2007). The presence of rare alleles may be because there is an enrichment process by having oil palm germplasm of different origins and/or countries (Arias et al. 2013). In the current studied oil palm population, the enrichment process may have been because of the presence of illegitimate progenies.

## Genetic diversity and differentiation among populations

The parameters of genetic diversity in each locus for each studied oil palm population are presented in Table 8. The estimation of genetic diversity within a population is important since it can be used to identify populations with a relatively high genetic diversity to design and maintain improvement strategies (Govindaraj et al. 2015). In the studied populations, the observed heterozygosity (H Obs.) values were all higher than the expected heterozygosity (H Exp.). The genetic diversity observed in each locus across the 6 populations studied varied within a range from the lowest H Exp.=0.47 (B57 population) to the highest H Exp.= 0.62 (A127 population) while the H Obs. in each locus for each population was from the lowest H Obs.= 0.627 (B57 population) to the highest H Obs=0.84 (B02 population). Meanwhile, the H Exp. for the T x T, the D Self, and All the populations were H Exp.=0.74, 0.62, and 0.77, while the H Obs. for the T x T, the D Self, and all the populations were H Exp.=0.74, 0.68, and 0.72, respectively (Table 6). Heterozygosity is an individual or population-level parameter and indicates the proportion of loci expected to be heterozygous in an individual (Gregorius, 1978). The H Obs. (Nei, 1978; Govindaraj et al. 2015) is the observed proportion of heterozygotes, averaged over the studied loci while H Exp. is also known as gene diversity (D), a calculated parameter (1.0 minus the sum of the squared gene frequencies). A greater value of H Obs. than H Exp. value indicates that the loci in the population have a high level of heterozygosity while a greater value of H Exp. than the H Obs. indicates a low level of heterozygosity (Govindaraj et al. 2015). In this study, the F values were negative for all loci for each of oil palm population (Table 8) and they ranged from F=-0.15 to F=-0.53. A negative F value further indicated an excess of heterozygotes among the six populations or the presence of undetected null alleles (Nei and Chesser, 1983). Okoye et al. (2016a, b) also reported the H Obs. varied from 0.17 for Madagascar material to 0.78 in Nigeria oil palm germplasm (mean = 0.575) and the H Exp. from 0.153 to 0.643.

Population structure is of interest to plant breeders and it can be estimated using (a) gene diversity in the average population, (b) levels of diversity in different populations, and (c) the degree of differentiation among populations (Brown, 1978). The $F_{ST}$ is the proportion of the total genetic variance contained in a subpopulation (the S

subscript) relative to the total genetic variance (the T subscript) and high $F_{ST}$ value indicates there is a considerable degree of differentiation among populations (Nei and Chesser, 1983). The $F_{IS}$ (inbreeding coefficient) is the proportion of the variance in the subpopulation contained in an individual and indicates the degree of inbreeding (Weir and Cockerham, 1984).

There was a high genetic differentiation ($F_{ST}$=0.38) among the six studied populations with a low number of migrants Nm=0.41 (Table 9). Out of the 100% total variance, 32% was distributed among the populations and 68% within individuals (Table 9). The $F_{IS}$ value over all loci and populations was also negative ($F_{IS}$=-0.31) and insignificant, whereas the $F_{IT}$ values were positive ($F_{IT}$=0.19) (Table 9). In this evaluation, the studied oil palm populations were improved breeding materials (T x T/P and D self populations) imported from IRAD, Cameroon. However, the population genetic parameters for these populations were similar to those previously reported for both improved oil palm populations and wild African ones.

The results of analysis of molecular variance (AMOVA) for improved oil palm breeding materials in Thailand indicated there was a 33% variation among populations and 67% among individuals within the populations. They also indicated the Nei's genetic distance among improved breeding materials in Thailand ranged from 0.53 to 0.62 (Taeprayoon et al. 2015). For the wild African oil palm germplasm, the high $F_{ST}$ was largely due to $F_{ST}$ among populations and the mean genetic distance across populations was 0.113 (Hayati et al. 2004). These authors also reported the oil palm population from Tanzania and the Democratic Republic of Congo showed the smallest genetic distance (D=0) while those from Madagascar and Sierra Leone were the furthest (D=0.568). In the meantime, Okoye et al. (2016a, b) demonstrated various wild oil palm provenances which showed significant genetic differentiation ($F_{ST}$ = 0.177) indicating the existence of genetic structures among the studied oil palm materials.

For the African wild population, the reported total gene flow across oil palm populations was low and the Nm value was 0.576. The mean expected heterozygosity was H Exp.=0.184, with values ranging from 0.109 to 0.261. The genetic differentiation among populations was high ($F_{ST}$=0.301), indicating a high genetic divergence (Hayati et al. 2004). Kularatne et al. (2001) also reported the presence of high overall genetic diversity among Nigerian oil palms and the genetic diversity gradually decreased in Senegal and Gambia, and also in Angola and Tanzania. Ghana oil palms were less diverse, while oil palms from Guinea showed a comparatively higher diversity (Kularatne et al. 2001).

**Population structure**

According to Pritchard et al. (2000) Structure software can perform groupings of individuals within the population more accurately than groupings based on similarity. Therefore, the output of the Structure software analysis would identify groupings of the population members better in the presence of individuals of a dubious subpopulation. The Structure software was also used to evaluate the studied oil palm populations. Results of the structure analysis using the Evanno method indicates that K = 2 was the best.

Based on the Structure analysis results (Figure 2), the six oil palm populations evaluated in this study were grouped into two populations, group I consisted of B01, A125, A127 and A140 populations and group II consisted of B02 and B57 populations. Although it was also a pisifera type of oil palm, similar to the B02 and B57 populations, the B01 populations were grouped into the same group as A125, A127, and A140 which were Dura type of oil palms. There were four individuals, one belonging to A127 and three to B01 populations, which were identified as probable admixtures, having a fraction of group 2 genetic contributions in a majority of group 1 genetic background (Figure 2).

**Table 8.** The mean values of genetic diversity parameters in each locus for each studied oil palm population

| Populations | H Exp[a] | H Obs[b] | F[c] |
|---|---|---|---|
| B01 | 0.60 | 0.68 | -0.37 |
| B02 | 0.62 | 0.70 | -0.53 |
| B57 | 0.58 | 0.67 | -0.35 |
| A140 | 0.53 | 0.74 | -0.16 |
| A125 | 0.54 | 0.84 | -0.15 |
| A127 | 0.47 | 0.62 | -0.16 |

Note: [a] H Exp., Nei's (1978) expected heterozygosity; [b] H Obs., observed heterozygosity; [c] F, fixation index Fis.

**Table 9.** Summary results of analysis of molecular variance (AMOVA) using input as allelic distance matrix for F-statistics analysis

| Source | df | SS | MS | Est. Var. | % |
|---|---|---|---|---|---|
| Among Populations (AP) | 5 | 552.299 | 110.460 | 2.686 | 32% |
| Among Individual (AI) | 142 | 430.384 | 3.031 | 0.000 | 0% |
| Within Individual (WI) | 148 | 846.232 | 5.718 | 5.718 | 68% |
| Total | 295 | 1828.916 | | 8.404 | 100% |
| **F-Statistics[a]** | **Value** | **P (rand >= data)** | | | |
| $F_{ST}$[b] | 0.380 | 0.001 | ** | | |
| $F_{IS}$[c] | -0.307 | 1.000 | ns | | |
| $F_{IT}$[d] | 0.190 | 0.001 | ** | | |
| $F_{ST}$ max | 0.502 | - | | | |
| Nm[e] | 0.407 | - | | | |

[a] Probability, P (rand >= data), Equations for calculating $F_{ST}$, $F_{IS}$ and $F_{IT}$ were based on standard permutation across the full data set. [b] $F_{ST}$ = AP / (WI + AI + AP) = AP /TOT; [c] $F_{IS}$ = AI / (WI + AI); [d] $F_{IT}$ = (AI + AP) / (WI + AI + AP) = (AI + AP) /TOT; [e] Nm = [ (1 /$F_{ST}$)-1] /4.
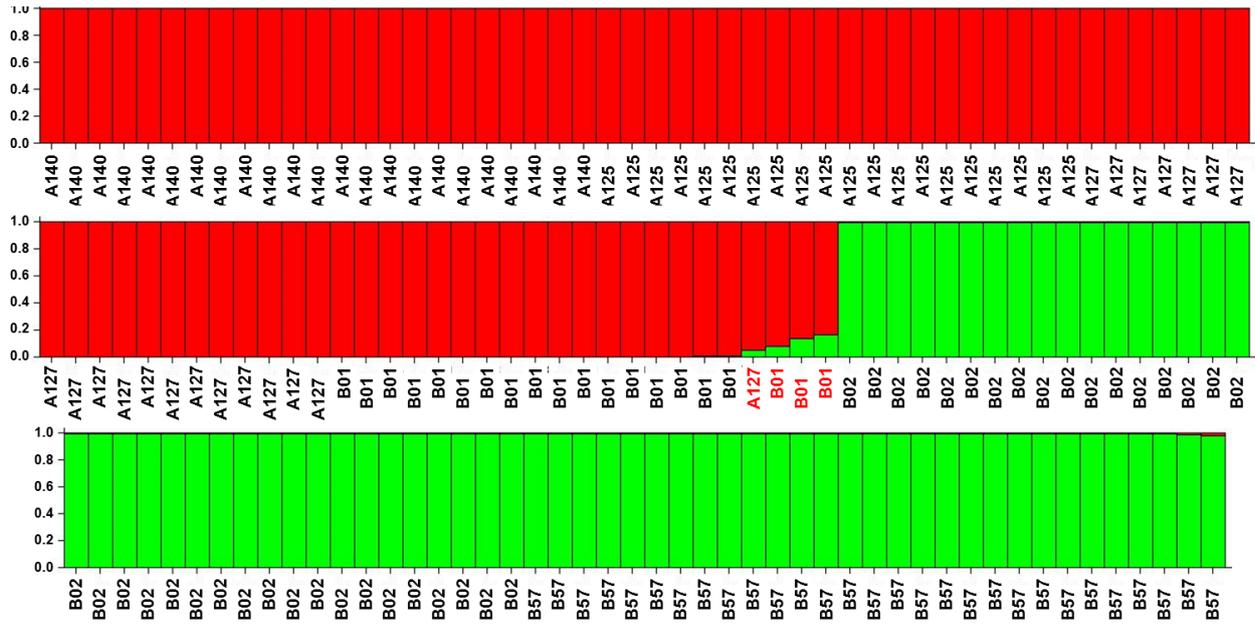
**Figure 2.** Six oil palm populations inferred from the STRUCTURE analysis. The vertical coordinate of each subgroup indicates the membership coefficients for each individual, and the digits on the horizontal coordinate represent the group IDs in Table 1. Different color in each individual bar indicated the genetic background. Individuals with a mixture of more than one color indicating they have more than one genetic background (admixtures)
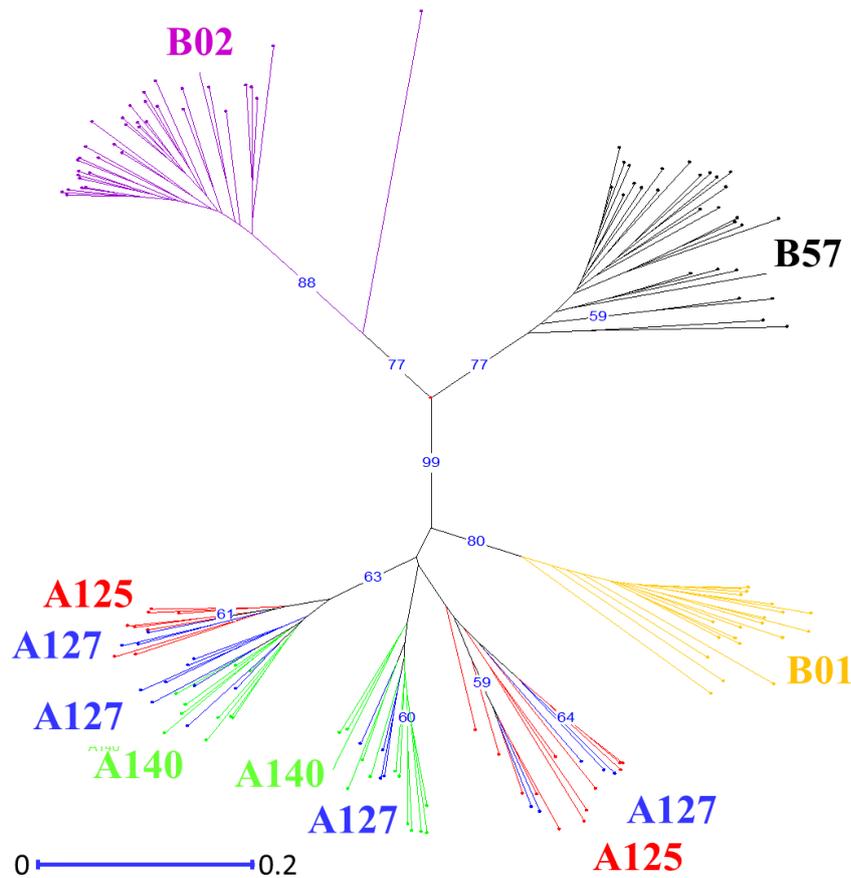


**Figure 3.** Unrooted neighbor-joining tree visualized from the dissimilarity matrix. The tree shows 148 individuals of six oil palms (*Elaeis guineensis*) populations originating from Cameroon. Accessions are identified by different lines color according to the populations (A125: red; A127: blue; A140: green; B01; yellow; B02: purple and B57: black)

## Cluster analysis of genetic relationship

Clustering analysis using the Neighbour-Joining approach and Simple Matching Dissimilarity Matrix of 148 oil palm individuals based on 16 SSR marker loci and the resulting clustering topology is presented in Figure 3. The result of analysis grouped the evaluated accessions of the studied oil palms into three major clusters. The first two major clusters consisted of the two T × T/P populations (B02, and B57) while the third major cluster consisted of the T × P population (B01) and a mixture of genotypes of the D self (A125, A127, and A140) oil palm populations (Figure 3). The results of cluster analysis clearly showed that one progeny of the T × T populations (B01 population) was closely related to the D self populations (A125, A127, and A140).

The individual members of each of the D self populations did not always fall into the same sub-group but they could belong to two to four different sub-groups. For example, members of the A127 D self population were grouped into four different sub-groups while those of the A125 and A140 populations were grouped into two different sub-groups. The mix in the three D self populations were probably associated with the progenitors of the populations. Individuals associated with A125 population were derived from (DA10D × DA115D) × (DA10D × DA115D), A127 from (DA10D × DA3D) × (DA10D × DA3D) and A140 from (DA115D × DA3D) × (DA115D × DA3D) parents (Table 1).

Results of the Principal Component Analysis (PCA) indicated that the representations of 16 SSR loci evaluated in this study were able to explain a total of 76.6% of the total variations in the populations. Moreover, the first five axes contributed to 63.73% of the total existing variations in the combined populations. Nearly identical groupings were also shown based on the accession divergence among the studied oil palm populations using PCA. Results of the PCA analysis (Figure 4) indicated members of the B01, B02, and B57 populations were placed in different groups while accessions belonging to the A125, A127, and A140 were intermixed in the same group (Figure 4). This may have happened because of the similar genetic background at the grandparent level. The A140 progenies have the genetic background of DA 115D × DA 115D grandparents. Meanwhile, the A125 progenies have the genetic background of (DA10D x DA115D) × (DA10D x DA115D). Therefore, between A140 and A125 progenies – they shared part of the same DA 115D which made some of the A140 and A125 progenies are intermixed. The A140 progenies also shared the same genetic background of DA 115D at the grandparent levels as the A127 progenies. Therefore, between A140 and A125 progenies – they also shared part of the same DA 115D which made some of the A140 and A125 progenies are intermixed. Moreover, The A125 and A127 progenies have also shared the same genetic background at the grandparent level since the shared part of the same DA 10D genetic background at the grandparent levels. The presence of the same genetic background at the grandparent level among A125, A127, and A140 might be the reason there were intermixed individuals in the grouping based on phylogenetics and PCA analysis. However, further studies are needed to confirm whether the intermixed individuals in the grouping based on phylogenetics and PCA analysis were because of the similar genetic background.

Characterization of wild African oil palm germplasm has been conducted by a number of research groups (Kularatne et al. 2001; Hayati et al. 2004; Bakoume et al. 2007; Maizura et al. 2006; Arias et al. 2013). Most of the evaluations were conducted to elucidate genetic information of wild oil palms from various countries or regions in Africa. These researchers used various molecular markers in their oil palm genetic evaluations such as AFLP (Kularatne et al. 2001), isozyme (Hayati et al. 2004), RFLP (Maizura et al. 2006), and SSR markers (Bakoume et al. 2007).

The reported results of the wild oil palm population evaluation using molecular markers indicated that many loci could be generated using multi-loci AFLP markers and the average number of generated markers could be as many as 47 loci per primer (Kularatne et al. 2001). However, AFLP is a dominant marker which is unable to differentiate heterozygous from homozygous individuals (Garcia et al. 2004). Using the isozyme system, Hayati et al. (2004) generated an average of 1.8 alleles per locus and out of the total markers, 54.5% of them were polymorphic while using RFLP markers. Maizura et al. (2006) generated an average of 1.9 alleles per locus and 62.7% of the total studied loci were polymorphic. Maizura et al. (2006) were also able to indicate the absence of as many as 36 alleles from the Deli related populations which still exist in the sampled wild oil palm populations. Although informative and co-dominant, the possible number of evaluated loci is limited for the isozyme system (Purba et al. 2000; Hayati et al. 2004), while it is no longer practical to apply RFLP for oil palm genetic analysis because of the limitation of this marker as reported by Garcia et al. 2004.

The SSR marker is more applicable for characterizing wild oil palm genetic materials as indicated by Bakoume et al. (2007). Using SSR markers, Bakoume et al. (2007) were able to generate a total of 209 alleles, with a mean number of 13.1 alleles per locus and the mean effective number of 3.3 alleles per locus (Ae). Characterization of improved oil palm populations and breeding materials have also been initiated such as in Indonesia (Solin et al. 2013; Ajambang et al. 2012; Zulhermana et al. 2010), Thailand (Taeprayoon et al. 2015), Nigeria and Malaysia (Okoye et al. 2016a, b).

Although it has progressed well, oil palm breeding in Southeast Asia (Malaysia and Indonesia) is still facing many problems such as the difficulty in conducting controlled crossing in oil palm which may result in contaminated/illegitimate progeny (Arias et al. 2012). Cross-pollination from unintended parents results in illegitimate progenies of the palm to palm crosses. The presence of contaminated/illegitimate progeny in oil palm breeding programs could cause a waste of time and labor (Hama-Ali et al. 2015). Hence, available methods for identifying illegitimate progenies should be beneficial. Hama-Ali et al. (2015) proposed to use a set of single-locus polymorphic SSR markers to identify illegitimate progenies.
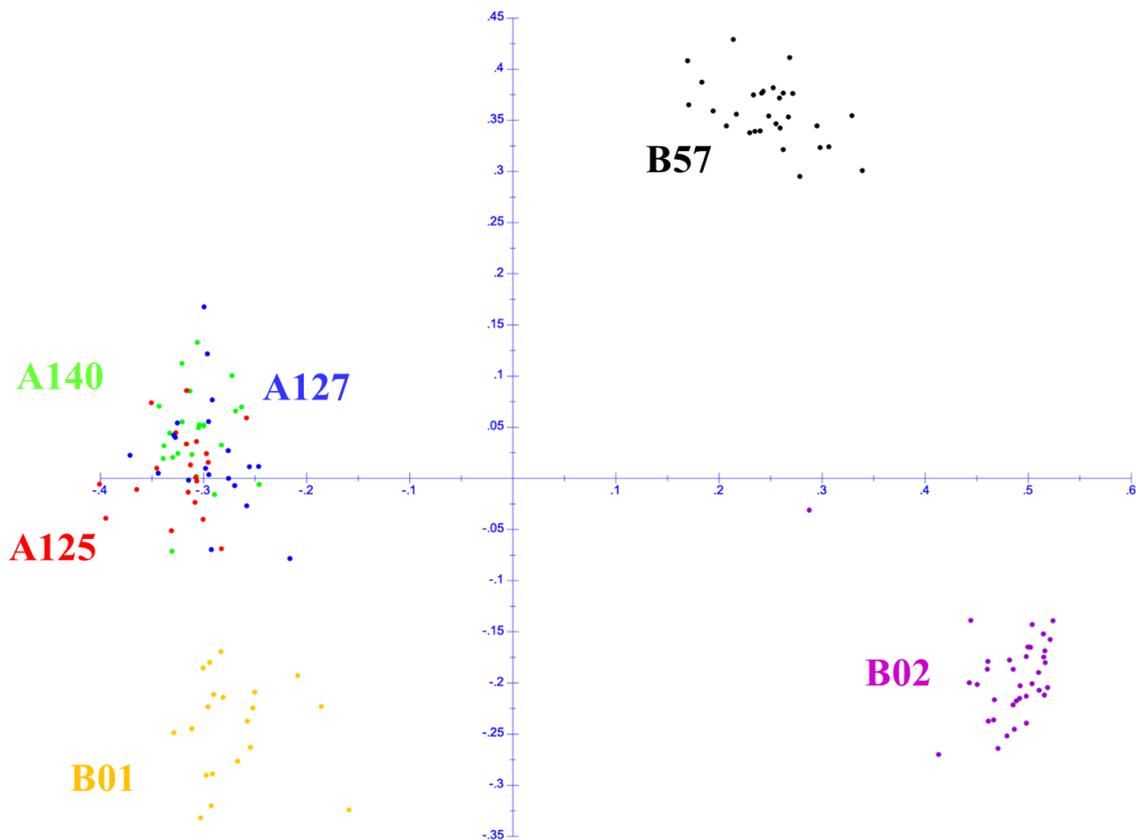
**Figure 4.** Principal component analysis of the 148 individuals in the six oil palm (*E. guineensis*) populations for the first two dimensions. It shows the conformation of the six populations into different groups (B01, B02, B57 and A125, A127, and A150). Accessions are identified by different line colors according to the populations (A125: red; A127: blue; A140: green; B01; yellow; B02: purple and B57: black)

The studied populations in this report represented both improved T x T/P and D self populations. They came from single palm to palm hybridization and consisted of three T x T/P crosses and three D self populations. According to Mendelian genetics of hybridization between two parents, the maximum expected allele numbers would be four alleles. Results of this investigation indicated the possible presence of illegitimate progenies among the studied T × T/P and D self populations. These hypotheses were supported by the observed numbers of the total alleles per locus and the distribution of allele frequency in each locus, which deviated from those expected for palm to palm hybridization model of Mendelian genetics.

Illegitimate (contaminant) progenies are quite common in the RRS method of oil palm breeding. The presence of illegitimate progenies (progenies derived from unidentified parents) occurred in any stage of the oil palm breeding activities. It may occur from the initial stage of selecting and labeling parents to the final stage of field evaluation trials. Factors affecting the percentage of illegitimacy in oil palm breeding include the biology of the oil palm flower, especially because of the existence of some hermaphrodite flowers, the human errors during pollen collection, the female flowers pollination bag damage because of animals,

or the many environment factors allowing insect pollinators to cross-pollinate the female flowers (Hama-Ali et al. 2015).

Similar findings had previously been reported by Okoye et al. (2016a, b) which also found a deviation of allele numbers and allele frequency from the expected model of Mendelian genetics. On the other hand, stuttering during PCR amplification of the SSR marker loci and other possible mechanisms may also result in variants having two to four base differences to the reference fragments. Occurrences of such variant alleles may also result in the identification of progenies as illegitimate, especially when the genotyping of SSR markers was done using sensitive capillary electrophoresis. However, the existence of suspected illegitimate progenies and their possible causes were not the subject of this investigation and the data were analyzed without excluding the possible illegitimate progenies.

Arias et al. (2013) hypothesized that the observed reduction of heterotic among African wild oil palm populations is the result of mixtures of different allele frequencies among different geographical regions. Such events have produced mixed populations with a heterozygous genotype deficiency when compared to the

expected according to the Hardy-Weinberg equilibrium. To be useful in breeding programs, the genetic base of the populations must be increased by oil palm breeders through the approach developed by Corley and Tinker (2003), which proposes to create crosses between plants from different origins, introducing new germplasm sources, followed by the selection of families and individuals to avoid inbreeding and placing emphasis on the general combining ability. These have been achieved through the RRS breeding strategy (Purba et al. 2000). The oil palm populations evaluated in this study were results of such a breeding strategy. As it has been shown previously, the studied populations maintain a high genetic diversity as indicated by H Obs, $F_{ST}$, $F_{IS}$, and other population parameters.

The long-term objectives of this study were to evaluate the improved male and female parents of oil palm without losing the potential for obtaining hybrid vigor. The generated molecular data indicated that the genetic diversity in the improved T × T/P and D self populations were mainly distributed within the populations, showing that all the sampled individuals within an improved population form a single, highly diverse population. Therefore, the future hybridization and new oil palm hybrid variety development could be based on the groups obtained in this investigation. Arias et al. (2013) suggested that the new variety and hybrid development of oil palms could be done using widely diverse parents. The two T × T/P populations (B02 and B57) could be selected as the potential male parents since they both demonstrated a wide genetic distance from the D self populations. On the other hand, one T × P population (B01) was genetically closer to the D self populations; therefore, it may potentially be suitable for evaluating correlations between genetic distance and hybrid vigor in oil palm as suggested by Cochard et al. 2009. Moreover, this experiment further proved the usefulness of SSR markers for characterizing and evaluating improved oil palm breeding populations and their use in future variety development. Such information derived from SSR marker analysis has the potential to support efficient oil palm genetic resource management and reduce the costs for new oil palm variety development.

In conclusion, results of this first evaluation of the genetic diversity of Astra Agro Lestari (AAL) six oil palm breeding populations demonstrated the presence of a high genetic variation within and between the AAL improved oil palm breeding populations. The T × T/P male parent populations were more diverse than the D self. The 16 evaluated microsatellite markers were either highly or moderately polymorphic based on their PIC values. Hence, the 16 SSR marker loci could be used for further analysis for larger numbers of AAL's T × T/P and D self population samples. The results of clustering and PCoA analysis indicated the T × T/P crosses of oil palm populations were clustered into three distinct groups. On the other hand, D self populations were clustered into four groups and each group comprised of a mixture of individual members of at least two different populations. Excluding the possibilities of some illegitimate individual existences, all of the three T × T/P could potentially be used as male parents for

producing future oil palm hybrid varieties since they have a wider genetic distance than that of the D self populations. Moreover, as the selected female parents, Dura types of oil palms should not be grouped based on their family names but should rather be based on the results of the clustering analysis using SSR marker data. Results of this study would be helpful for designing crosses among T × T/P and D self parents for future oil palm breeding and selection programs in Indonesia.

## REFERENCES

Ajambang W, Sudarsono, Asmono D, Toruan-Mathius N. 2012. Microsatellite markers reveal Cameroon's wild oil palm population as a possible solution to broaden the genetic basqe in the Indonesia-Malaysia oil palm breeding programs. Afr J Biotech 11 (69): 13244-13249.

Ajijah N. 2015. Development of Cacao Regeneration System through Somatic Embryogenesis and Identification of Variants Resistant to *Phytophthora palmivora* Butl. [Dissertation]. Bogor Agricultural University, Bogor, Indonesia. [Indonesian]

Akkaya MS, Bhagwat AA, Cregan PB. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132: 1131-1139.

Arias D, Montoya C, Rey L, Romero H. 2012. Genetic similarity among commercial oil palm materials based on microsatellite markers. Agric Colomb 30: 188-195.

Arias D, Montoya C, Romero H. 2013. Molecular characterization of oil palm *Elaeis guineensis* Jacq. materials from Cameroon. Plant Genet Resour 11 (2): 140-148.

Bakoume B, Wickneswari R, Rajanaidu N, Kushairi A, Amblard P, Billotte N. 2007. Allelic diversity of natural oil palm (*Elaeis guineensis* Jacq.) populations detected by microsatellite markers: implications for conservation. Plant Genet Resour: Characterization Utilization 5 (2): 104-107.

Barcelos E, de Almeida RS, Cunha RN, Lopes R, Motoike SYY, Babiychuk E, Skirycz A, Kushnir S. 2015. Oil palm natural diversity and the potential for yield improvement. Front Plant Sci 6: 1-16.

Billotte N, Risterucci AM, Barcelos E, Noyer JL, Amblard P, Baurens FC. 2001. Development, characterisation, and across-taxa utility of oil Palm (*Elaeis guineensis* Jacq.) microsatellite markers. genome 44 (3): 413-25.

Brown AH. 1978. Isozymes, plant population genetic structure and genetic conservation. Theor Appl Genet 52 (4): 145-57.

Buchert GP, Rajora OP, Hood JV, Dancik BP. 1997. Effects of harvesting on genetic diversity in old-growth eastern white pine (*Pinus strobus* L.) in Ontario, Canada. Conserv Biol 11: 747-758.

Cochard B, Adon B, Rekima S, Billotte N, Desmier R, Koutou A, Nouy B, Omore A, Purba AR, Glazsmann JC, Noyer JL. 2009. Geographic and genetic structure of african oil palm diversity suggests new approaches to breeding. Tree Genet Genomes 5: 493-504.

Corley RHV, Tinker PB. 2003. The Oil Palm, 4th ed. Oxford: Blackwell Publishers, UK

Danielsen F, Hendrien, Neil B, Faizal B, Carsten P, Paul AB, Daniel FD, Ben M, Lucas P, Matthew R, Emily BF. 2009. Biofuel plantations on forested lands: double jeopardy for biodiversity and climate. Conserv Biol 23 (2): 348-358.

Delseny M, Laroche M, Penon P. 1983. Detection of sequences with Z-DNA forming potential in higher plants. Biochem Biophys Res Comm 116: 113-120.

Ditjenbun. 2013. Diskusi Industri Sawit Pasca Moratorium, Mau Kemana? Direktorat Jendral Perkebunan, Republik Indonesia. Available from:

http: //ditjenbun.pertanian.go.id/tanhun/berita-182-diskusi-industri-sawit-pasca-moratorium-mau-kemana.html [22 July 2016]. [Indonesian]

Ditjenbun. 2014. Pertumbuhan Areal Kelapa Sawit Meningkat. Direktorat Jendral Perkebunan, Republik Indonesia. Available from: http: //ditjenbun.pertanian.go.id/setditjenbun/berita-238-pertumbuhan-areal-kelapa-sawit-meningkat.html. [22 July 2016]. [Indonesian]

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4 (2): 359-361.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14 (8): 2611-2620.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genet* 131: 479-491.

Falush D, Stephen M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genet 164 (4): 1567-87.

Garcia AAF, Benchimol LL, Barbosa AMM. Geraldi IO, Souza Jr CL, de Souza AP. 2004. Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. Genet Mol Biol 27 (4): 579-588.

Govindaraj M, Vetriventhan M, Srinivasan M. 2015. Review Article importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genet Res Int 14.

Gregorius HR. 1978. The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. Math Biosci 41: 253-432.

Hama-Ali EO, Alwee SSRS, Tan SG, Panandam JM, Ling HC, Namasivayam P, Peng HB. 2015. Illegitimacy and Sibship Assignments in Oil Palm (*Elaeis guineensis* Jacq.) Half-Sib Families Using Single Locus DNA Microsatellite Markers. Mol Biol Rep 42 (5): 917-25.

Hayati A, Wickneswari R, Maizura I, Rajanaidu N. 2004. Genetic diversity of oil palm (*Elaeis guineensis* Jacq.) germplasm collections from Africa: implications for improvement and conservation of genetic resources. Theor Appl Genet 108 (7): 1274-84.

Hayden MJ, Nguyen TM, Waterman A, Chalmers KJ. 2008. Multiplex-ready PCR: a new method for multiplexed SSR and SNP genotyping. BMC Genomics 9: 80.

Jones OR, Wang J. 2010. COLONY: a program for parentage and sibship inference from multilocus genotypes data. Mol Ecol Res 10: 551-555.

Jones AG, Small CM, Paczolt KA, Ratterman NL. 2010. A practical guide to methods of parentage analysis. Mol Ecol Res 10 (1): 6-30.

Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16: 1099-1106.

Kularatne RS, Shah FH, Rajanaidu N. 2001. The evaluation of genetic diversity of Deli Dura and African oil palm germplasm collection by AFLP technique. Tropical Agric. Res. 13: 1-12.

Lim CC, Rao V. 2005. DNA fingerprinting of oil palm. J Oil Palm Res 17: 136-44.

Maizura I, Rajanaidu N, Zakri AH, Cheah SC. 2006. Assessment of genetic diversity in oil palm (*Elaeis guineensis* Jacq.) using Restriction Fragment Length Polymorphism (RFLP). Genet Resour Plant Evol 53: 187-195.

Marshall DR, Brown AH. 1975. Optimum sampling strategies in genetic conservation. In: Franked OH and Hawkes JG (eds.). Crop Genetic Resources for Today and Tomorrow. Cambridge, London: Cambridge University Press, UK.

Maskromo I, Tenda ET, Tulalo MA, Novarianto H, Sukma D, Sukendah, Sudarsono. 2015. Phenotypic and genotypic variabilities among kopyor dwarf coconut varieties originated from Pati Central Java. J Litri 21 (1): 1-8.

Maxted N, Ford Lloyd BV, Jury SL, Kell SP, Scholten MA. 2006. Towards a definition of a crop wild relative. Biodiv and Conserv 15: 2673-2685.

Maxted N, Scholten MA, Codd R, Ford Lloyd BV. 2007. Creation and use of a national inventory of crop wild relatives. Biol Conserv 140: 142-159.

Mohammadi SA, Prasanna BM. 2003. Analysis of genetic diversity in crop plants-salient statistical tools and considerations. Crop Sci 43: 1235-1248.

Moose SP, Mumm RH. 2008. Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement. Plant Physiol 147: 969-977.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genet 89: 583-590.

Nei M, Chesser RK. 1983. Estimation of fixation indices and gene diversities. Ann Hum Genet 47 (3): 253-259.

Okoye M, Uguru MI, Bakoume C, Singh R, Okwuagwu CO. 2016a. Assessment of genetic diversity of NIFOR oil palm main breeding parent genotypes using microsatellite markers. Am Jour Plant Sci 7: 218-237.

Okoye M, Bakoume C, Uguru MI, Singh R, Okwuagwu CO. 2016b. Genetic relationships between elite oil palms from nigeria and selected breeding and germplasm materials from Malaysia via Simple Sequence Repeat (SSR) Markers. J Agric Sci 8 (2): 159-178.

Peakall R, Smouse PE. 2012. GenALEx 6.5: Genetic analysis in excel. population genetic software for teaching and research-an update. Bioinformation 28 (19): 2537-39.

Perrier X, Flori A, Bonnot F. 2003. Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (ed.). *Genetic Diversity of Cultivated Tropical Plants*. Enfield Science Publishers. Montpellier.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155 (2): 945-959.

Purba AR, Noyer JL, Baudouin L, Perrier X, Hamon S, Lagoda PJL. 2000. A new aspect of genetic diversity if Indonesian oil palm (*Elaeis guineensis* Jacq.) revealed by isoenzyme and AFLP markers and its consequences for breeding. Theor Appl Genet 101: 956-961.

Rajora OP, Rahman MH, Buchert GP, Dancik BP. 2000. Microsatellite DNA analysis of genetic efforts of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. Mol Ecol 9: 339-348.

Sambrook J, Fritch EF, Maniatis T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, USA.

Saghai-Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW. 1994. Extraordinarily polymorphic microsatellite DNA in barley species diversity, chromosomal location and population dynamics. Proc Natl Acad Sci USA 9: 5466-547.

Sajib AM, Hossain M, Mosnaz ATMJ, Hossain H, Islam M, Ali S, Prodhan SH. 2012. SSR marker-based molecular characterization and genetic diversity analysis of aromatic landraces of rice (*Oryza sativa* L.). J BioSci Biotech 1: 107-116.

Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18 (2): 233-234.

Solin NWNM, Sobir S, Toruan-Mathius N. 2014. Genetic diversity of DxP population yield component in oil palm's paternal half-sib family based on microsatellite markers. Energ Procedia 47: 196-203.

Tautz D, Renz M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12 (10): 4127-38.

Tautz D.1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 17 (16): 6463-6471.

Taeprayoon P, Tanya SH, Lee P, Srinives. 2015. Genetic background of three commercial oil palm breeding populations in Thailand revealed by SSR markers. Austr J Crop Sci 9 (4): 281-288.

Thongthawee S, Tittinutchanon P, Volkaert H. 2010. Microsatellites for parentage analysis in an oil palm breeding population. Thai J Genet 3 (2): 172-181.

Ting NC, Noorhariza MZ, Rozana R, Low ET, Ithnin M, Cheah SC. 2010. SSR mining in oil palm EST database: application in oil palm germplasm diversity studies. J Genet 89: 135-145.

Tornincasa P, Furlan M, Pallavicini A, Graziosi G. 2010. Coffee species and varietal identification. In: Nimis PL, Vignes Lebbe R (eds.). Tools for Identifying Biodiversity: Progress and Problems. Edizioni Università di Trieste, Italy.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Zulhermana, Sudarsono, Asmono D, Yulismawati. 2010. Intra-and inter-population genetic diversity of oil palm (*Elaeis guineensis* Jacq.) Pisifera clones originated from Nigeria based on SSR marker analysis. International Oil Palm Conference (IOPC), Yogyakarta, 1-3 Juni 2010 [Indonesian]