

# GENETIC DIVERSITY AND INBREEDING LEVEL IN DELI *Dura* AND AVROS ADVANCED BREEDING MATERIALS IN OIL PALM (*Elaeis guineensis* Jacq.) USING MICROSATELLITE MARKERS

TAY CHEE CHUN\*; CHUA KIA LING\*; LEE CHONG HEE\*; CHEAH SUAN CHOO\*\*; JAKIM BARIN†; AU WAI FONG‡; MOHD PUAD ABDULLAH†† and TAN SOON GUAN††

## ABSTRACT

Oil palm has a narrow genetic diversity due to intensive selection in breeding. Loss of diversity in breeding materials can lead to many consequences, despite uniformity is critical to breeders. The objectives of the study are to evaluate the genetic diversity in Deli and AVROS populations, and to determine the levels of inbreeding using CIRAD's Simple Sequence Repeat (SSR) markers. The parental populations evaluated were 186 palms from Deli and 188 palms from AVROS progenies sourced from two agencies. Genotyping using 32 SSR markers produced 230 alleles among the 17 oil palm progenies. The number of alleles scored ranged between 4 to 11. The average number of alleles per locus was 7.1875. The expected heterozygosity ( $H_e$ ) and observed heterozygosity ( $H_o$ ) were 0.7063 and 0.5270, respectively. The dendrogram revealed that the 17 populations were clustered into three main clusters namely Deli dura, AVROS and Ekona populations. The two different generations of AVROS resulted in Genting AgTech Sdn Bhd (GAT) AVROS had inbreeding coefficient of 0.0336 compared to Department of Agriculture (DOA) Sabah AVROS which was -0.0810. The diversity revealed via SSR showed that there are still present of variability in some of the progenies. The inbreeding level can serve as guidance during the inbred lines development.

**Keywords:** *Elaeis guineensis* Jacq., inbreeding coefficient, Simple Sequence Repeat (SSR).

**Date received:** 29 November 2016; **Sent for revision:** 22 March 2017; **Received in final form:** 3 April 2018; **Accepted:** 31 July 2018.

## INTRODUCTION

The African oil palm (*Elaeis guineensis* Jacq.) is an oil crop that is native to Africa. As the world's leading oil crop, oil palm produces 37% of the total vegetable oils and fats from just 18 million hectares, or 6% of total agricultural land cultivated with oil crops (Oil World, 2016). It yields five to 15 times more oil per hectare as compared to rapeseed or soyabean. In 2012, the estimated global oil palm planted area was 14.8 million hectares that produced 53.67 million tonnes of crude palm oil (CPO). In 2014, the oil palm planted area in Malaysia was 5.39 million hectares producing 19.67 million tonnes of palm oil. The great economic value generated by palm oil export

\* Genting AgTech Sdn Bhd, Lot L3-I-1, Enterprise 4, Technology Park Malaysia, Bukit Jalil, 57000 Kuala Lumpur, Malaysia.  
E-mail: cheechun224@gmail.com

\*\* ACGT Sdn Bhd, Lot L3-I-1, Enterprise 4, Technology Park Malaysia, Bukit Jalil, 57000 Kuala Lumpur, Malaysia.

† Department of Agriculture Sabah, Level 5, Wisma Pertanian Sabah, Jalan Tasik, Off Jalan Maktab Gaya, 88632 Kota Kinabalu, Sabah, Malaysia.

†† Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

has changed the oil palm from its ornamental status to that of a cash crop. In Malaysia, oil yield has increased four-fold in the last 50 years, half of this being attributable to genetic improvement of the planting material (Lee *et al.*, 1990). However, the study showed that selection for this improvement reduced the coefficient of variation for fruit bunch yield by nearly half. Intensive breeding from a narrow genetic base is leading to limited variation within the elite material, potentially reducing the rate of future breeding progress. Efforts to broaden the genetic variability of the current oil palm advanced breeding populations are crucial to enable the breeders to exploit the genetic potentials in order to produce planting materials that are high yielding with better oil quality and highly tolerance to diseases. Musa and Gurmit (2008) has indicated great potential for improvement through the introgression of selected palms from the MPOB germplasm into the advanced *dura* and *tenera* lines which reduced palm height (Isa *et al.*, 2008) and improved oil yield and fruit qualities (Junaidah *et al.*, 2008).

Conventional breeding takes 10 years for oil palm improvement due to the fact that it is perennial oil crop with a long generation cycle of breeding. These lead to slow progress and hamper the release of high yielding planting materials. However, there are various molecular techniques that are available today which provide a powerful tool to facilitate crop improvement and selection. Molecular markers have been exploited in oil palm research over the last two decades in areas such as DNA fingerprinting using restriction fragment length polymorphism (RFLP) (Cheah *et al.*, 1996). In addition, molecular markers are able to detect genetic variability (Billotte *et al.*, 2005), genome mapping and Quantitative Trait Loci (QTL) detection for Marker Assisted Selection (MAS) (Billotte *et al.*, 2010). Microsatellite marker or Simple Sequence Repeat (SSR) is one of the molecular markers that are in abundance which is well utilised for genetic diversity studies (Norziha *et al.*, 2008; Singh *et al.*, 2008). It has great advantages over other markers systems due to its high polymorphism and wide distribution of the loci within the genome (Vignal *et al.*, 2002). Besides, SSR is a co-dominant marker which is ideal for diversity measurements and is highly reproducible through the polymerase chain reaction (PCR). Hamblin *et al.* (2007) reported that SSR performed better to assign the inbred lines to sub-populations as compared to Single Nucleotide Polymorphism (SNP) in their research to evaluate the genetic diversity in a set of public maize inbreds. Comparisons of using SSR and SNP to reveal genetic diversity were discussed for several plant species such as maize (Van Inghelandt *et al.*, 2010), rice (Singh *et al.*, 2013), grape (Emanuelli *et al.*, 2013) and wheat (Ren *et al.*, 2013). These authors suggested that between SSR and SNP markers, a considerable

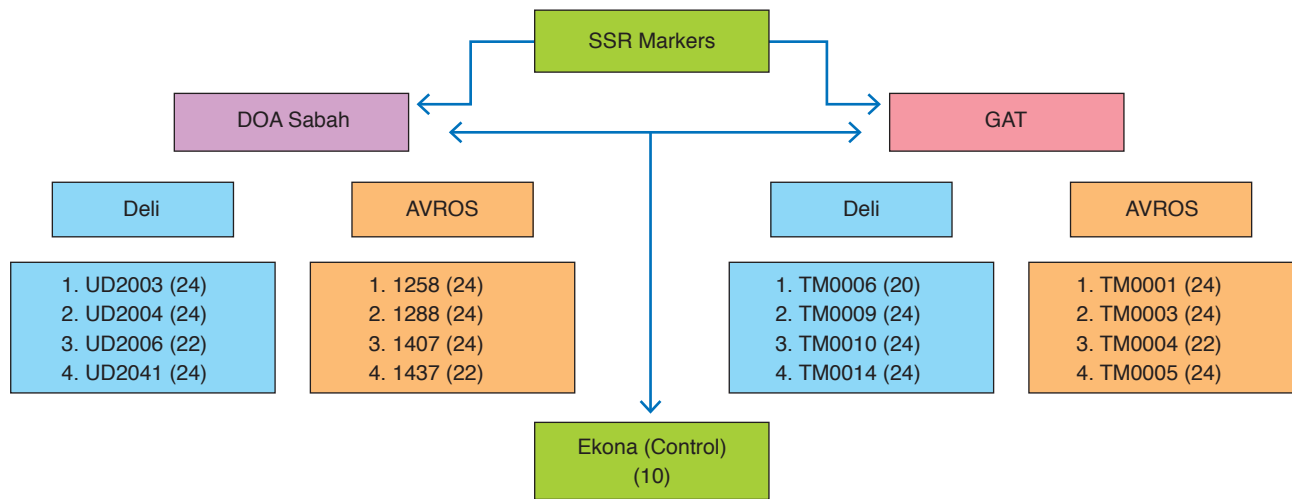
higher number of SNP marker was required in order to have an equivalent discriminating power as the SSR marker. They proposed between 7 and 11 times more SNP than SSR should be used for analysing population structure and genetic diversity. It was also found that SSR was able to provide more resolution in measuring genetic distance based on allele-sharing.

The Malaysian advanced breeding materials particularly the Deli population and AVROS population have been subjected to intensive selection since the 1930s (Maizura *et al.*, 2009). Breeding and selection for crop improvement using these populations by breeders had led to narrowed genetic diversity as shown by various reports (Lee *et al.*, 1990; Hayati *et al.*, 2004; Maizura *et al.*, 2006a). The majority of the studies were carried out using germplasms and not many focused on advanced breeding materials. Besides, no comparison of different generations of the same population was carried out previously. Thus, an understanding on the genetic diversity of the advanced breeding populations at the molecular level is of importance for breeders to exploit the remaining genetic potential for crop improvement. The findings from this study could serve as supplementary information for the breeders during their crossing design enabling more precise decisions on the selection of parental materials for seed production and inbred line development. Thus, the objectives of the current study were: 1) to determine the genetic diversity in the advanced breeding materials, Deli and AVROS populations, for future breeding and improvement, and 2) to determine the levels of inbreeding across different generations in AVROS populations.

## MATERIAL AND METHODS

### Plant Materials

Genting AgTech Sdn Bhd (GAT) germplasm materials were provided by the Malaysian Palm Oil Board (MPOB) to start GAT seed garden at Johor, Malaysia. As part of the collaboration between GAT and Department of Agriculture (DOA) Sabah on joint marker-assisted breeding programme, GAT received germplasm materials from Ulu Dusun Agriculture Research Station (UDARS). These DOA Sabah materials are planted in GAT Sandakan. The oil palm advanced breeding populations in this study were parental populations which consisted of Deli population, AVROS population and Ekona population was used as the control. These populations comprised of four AVROS progenies (94 palms from DOA Sabah), four Deli progenies (94 palms from DOA Sabah), and four AVROS progenies (94 palms from GAT) and four Deli progenies (92 palms from GAT) (Figure 1).



Note: DOA - Department of Agriculture.  
 GAT - Genting AgTech Sdn Bhd.  
 SSR - Simple Sequence Repeat.

Figure 1. Source of samples and number of palms per progeny from DOA Sabah and GAT for Deli and AVROS populations. A total of 186 samples representing Deli populations while AVROS populations consisted of 188 samples. Ten Ekona was included as control.

### DNA Extraction

DNA from the frozen green leaf samples were extracted using Qiagen Plant DNEasy Kit. The total genomic DNA was extracted following the manufacturer's instructions. Information of the materials analysed in this study is listed in Table 1. The quantity and concentration for the extracted DNA was measured using Infinite® 200 Pro NanoQuant microplate reader (Tecan, Switzerland). The final samples were diluted to a concentration of 10 ng µl<sup>-1</sup>. In order to confirm the suitability of DNA for enzymatic reaction, the extracted DNA was quality-checked by gel electrophoresis (1.0% agarose gel) and restriction enzyme digestion using *HindIII* (New England Biolabs).

### SSR Screening for Polymorphism

A total of 74 SSR primers from the CIRAD primer collection (<http://tropgenedb.cirad.fr/oilpalm/publications.html>) were screened on 16 sub-samples (consisted of one palm as representative from each progeny) for polymorphism. SSR were shortlisted according to their positions in each linkage group as reported by Billotte *et al.* (2005) and were spread evenly across the 16 linkage groups. The PCR amplification was performed in a 96-well plate (Axygen, USA) with 30 ng of DNA in a 10 µl final volume of 1 µl of 10X PCR buffer at the concentration of 1X, 0.5 µl of 2.5 mM MgCl<sub>2</sub>, 0.35 µl of 350 µM dNTP, 0.2 µl of 1U/reaction *Taq* DNA polymerase, 0.04 µl of 0.1 µM Forward primer and 0.2 µM Reverse primer mixture, 0.4 µl of 0.1 pmol µl M13 dye labelled forward primer, and final volume was topped up with 4.51 µl of autoclaved MilliQ

Ultrapure water. A touchdown PCR programme was performed as follows: initial denaturation at 94°C for 5 min; 5 cycles at 94°C for 30 s, primer melting temperature (T<sub>m</sub>) (62°C to 52°C, -2°C per cycle) for 1 min, and 72°C for 1 min 30 s; additional round of 32 cycles of 94°C for 30 s, 52°C for 1 min, and 72°C for 1 min 30 s and a final elongation step at 72°C for 10 min before being stored at 4°C after the PCR amplification was completed.

For post PCR multiplexing, 1.0 µl of PET (Applied Biosystems, USA) and 1.0 µl of VIC (Applied Biosystems, USA) PCR products were pooled together with 10 µl of a mixture of GeneScan™-600 LIZ® (Applied Biosystems, USA) and highly deionised Hi-Di™ Formamide (Applied Biosystems, USA). The mixture was prepared by adding 0.4 µl of size standard GeneScan™-600 LIZ® to 9.6 µl of Hi-Di™ Formamide. The mixture was denatured at 94°C for 3 min. The PCR products were subjected to capillary electrophoresis and detected by the laser system of the ABI 3730xL DNA Analyser. The output files from the sequencer were visualised using the GeneMapper® 4.1 software (Applied Biosystems, USA).

### SSR Genotyping and Scoring

A total of 32 CIRAD SSR primers were shortlisted for further use in genotyping of all the progenies involved in this study. The sequence information and expected amplicon sizes for each primer is listed in Table 2. The PCR amplification and post PCR multiplexing for capillary electrophoresis were carried out as described earlier. GeneMapper® 4.1 software is the latest fragment analysis software from Applied Biosystems, USA. The peak detection

TABLE 1. INFORMATION ON MATERIALS AND NUMBERS OF PALMS USED

Population Type	Population ID	Population information	No. of palm
(i) Maternal population	DOA Sabah UD 2003	UR x JL Deli <i>dura</i>	24
	DOA Sabah UD 2004	UR x UR Deli <i>dura</i>	24
	DOA Sabah UD 2006	B x B Deli <i>dura</i>	22
	DOA Sabah UD 2041	JL Deli <i>dura</i>	24
	GAT TM0006	MPOB Deli <i>dura</i>	20
	GAT TM0009	MPOB Deli <i>dura</i>	24
	GAT TM0010	MPOB Deli <i>dura</i>	24
	GAT TM0014	MPOB Deli <i>dura</i>	24
(ii) Paternal population	DOA Sabah 1258	3 <sup>rd</sup> Generation AVROS	24
	DOA Sabah 1288	3 <sup>rd</sup> Generation AVROS	24
	DOA Sabah 1407	3 <sup>rd</sup> Generation AVROS	24
	DOA Sabah 1437	3 <sup>rd</sup> Generation AVROS	22
	GAT TM0001	5 <sup>th</sup> Generation AVROS	24
	GAT TM0003	5 <sup>th</sup> Generation AVROS	24
	GAT TM0004	5 <sup>th</sup> Generation AVROS	22
	GAT TM0005	5 <sup>th</sup> Generation AVROS	24
(iii) Control	DOA Sabah Ekona	Ekona	10
Total			384

Note: DOA - Department of Agriculture.

GAT - Genting AgTech Sdn Bhd.

UD - Ulu Dusun. UR - Ulu Remis. JL - Johor Labis. B - Banting.

algorithms are optimised for detecting stutter bands in the SSR samples. The software enables the user to detect polymorphic DNA loci that contain repeated nucleotide sequences. The workflow of setting up the microsatellite analysis is as described by the GeneMapper® Software manual version 4.0.

### Genetic Variability Parameters

The genotype scores from the GeneMapper® 4.1 software was analysed using the PowerMarker software version 3.25 (Liu and Muse, 2005) to compute allele frequency parameters including allele frequency, number of alleles ( $N_a$ ), observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity, and polymorphic information content (PIC). The  $H_o$  is simply the proportion of heterozygous individuals in the population. The  $H_e$  is defined as the probability that two randomly chosen alleles from the population are different (Nei, 1972). The PIC estimation is crucial for shortlisting the informative markers for parental genotyping.

### Genetic Relatedness Analysis

Population structure analysis was performed on the 384 samples. The model-based software Structure v2.3.4 (Pritchard *et al.*, 2000) was used to infer the population structure using a burn-in of 100 000, run length of 100 000 and a model allowing for admixture and correlated allele frequencies. Five independent runs yielded consistent results. Analysis was performed from  $K=2$  to  $K=12$ . The mean log likelihood of the data for each given  $K$ ,  $L(K)$  was plotted against the  $K$  value. The rate of change of the likelihood distribution,  $L'(K)$  and the

absolute values of the second order rate change of the likelihood distribution,  $|L''(K)|$ , for each  $K$  value were calculated as  $L'(K)=L(K)-L(K-1)$  and  $|L''(K)|=|L'(K+1)-L'(K)|$  respectively (Evanno *et al.*, 2005). The  $L'(K)$  and  $|L''(K)|$  were plotted against the  $K$  value. These three plots were used to determine the number of subpopulations that were present among the 384 palms (Evanno *et al.*, 2005).

The Neighbor-Joining and principal component analysis (PCA) were performed using the NTSYSpc software version 2.20r (Rohlf, 2005). Individual-based genetic distance was calculated using the Nei distance (Nei, 1972). The similarity between all 384 samples was calculated using the SIMINT module. The coefficient used in the calculation was the Pearson product-moment correlation. The similarity matrix generated was used to compute eigenvalues and eigenvectors using the EIGEN module. The extracted three eigenvectors from the similarity matrix were used to plot a two dimensional PCA.

The advanced breeding populations were redefined based on the above individual-based genetic distance analysis. Population-based Neighbor-Joining was performed using the PowerMarker software version 3.25 (Liu and Muse, 2005). Population distance was calculated using the Nei distance (Nei, 1972). Levels of genetic variation within and among populations identified by the cluster analysis were estimated from allelic frequencies using the analysis of molecular variance, AMOVA (Weir and Cockreham, 1984; 1996). The software Arlequin version 3.5.1.2 (Excoffier *et al.*, 2005) performs the AMOVA procedure using SSR and standard multi-locus frequency data with 10 000 permutations. Wright's  $F$ -statistics, *i.e.*  $F_{ST}$  which

TABLE 2. MICROSATELLITE ID, PRIMER SEQUENCES, ANNUALING TEMPERATURE (TA) AND EXPECTED AMPLICON SIZE (ES)

No.	Microsatellite ID	Forward primer	Reverse primer	TA (°C)	ES (bp)
1	mEgCIR0037	CCA GTC TGC TAA CCA TCC TAT AC	TCT CAC TTC CTC CCC ACA TC	52	175
2	mEgCIR0067	TAC ACA ACC CAT GCA CAT	AAA AAC ATC CAG AAA TAA AA	52	167
3	mEgCIR0257	GCA GCT AGT CAC CTG AAC	GAC GAG ACT GGA AAG ATG	52	286
4	mEgCIR0580	GCA CCA GCA CTT TCA CAA	CAA ACC AAT ACC CAA GCT	52	212
5	mEgCIR0588	GCA AGA TGC AAT GGA GTT CA	CAA ACC GCA GCA AGT CAG A	52	215
6	mEgCIR0772	TAT AAT CCA CCC AGC ACA AC	CCA ATT ATA CAA TCC CAC AAA G	52	148
7	mEgCIR0774	TGG CCG AGG CAG AAG AAA AT	GCT TGG TGG GTA AGC TGG ATT ATT	52	161
8	mEgCIR0775	CAT ATG GCT AAG AAA GCC AAC AGA ATA	CGT GGT TGT GTA TAG GCT GAT TCT A	52	237
9	mEgCIR0779	AAT GCA GAC CAA GCT AAT CAT ATA C	GTT CAG GTG ATG GTG ACT CAG ATA G	52	238
10	mEgCIR0782	CGT TCA TCC CAC CAC CTT TC	GCT GCG AGG CCA CTG ATA C	56	187
11	mEgCIR0783	GAA TGT GGC TGT AAA TGC TGA GTG	AAG CCG CAT GGA CAA CTC TAG TAA	58	296
12	mEgCIR0788	ACA TTC CCT CTA TTA TTC TCA C	GTT TTG TTT GGT ATG CTT GT	52	129
13	mEgCIR0803	AAG CCA ACT TCA CAG ATA TGT TGA T	ATG AGC CTA ACA AAG CAC ATT CTA A	52	265
14	mEgCIR0804	GGA GTT AGT AAG TTA GTG AGA GAG A	GCG TTG TTT GGA TGA TG	52	205
15	mEgCIR0825	AGT GAG GTA TGG TTG ATT AGG A	TAT TGA TAG CAT TTG GGA TTA G	52	200
16	mEgCIR0874	TCC AGT TGT CGA GTT GTA GT	ATT ATG GGG TTA TGC TTT CA	52	235
17	mEgCIR1713	GCT GAA GAT GAA ATT GAT GTA	TTC AGG TCC ACT TTC ATT TA	52	252
18	mEgCIR2110	TGT TTT GTT TCG TGC ATG TG	GGC TGA CAT GCA ACA CTA AC	52	154
19	mEgCIR2212	CGG TTT TGT CGC ATC TAT G	GTC GTC AGG GAA CAA CAG T	52	125
20	mEgCIR2215	GAA CTT GGC GTG TAA CT	TGG TAG GTC TAT TTG AGA GT	52	121
21	mEgCIR2347	ATT TTG CAT GTG TTG AGA GC	CAA CCA ATT GCA CCC TAA AG	52	153
22	mEgCIR2387	TTG GTG AGC CAT TTG CTA CA	CCT CCT TCC ACC CCT CTA CT	52	243
23	mEgCIR2414	CAA TCA TTG GCG AGA GA	CGT CAC CTT TCA GGA TAT G	52	195
24	mEgCIR2569	TAG CCG CAC TCC CAC GAA GC	CCA GAA TCA TCA GAC TCG GAC AG	56	229
25	mEgCIR2595	TCA AAG AGC CGC ACA ACA AG	ACT TTG CTG CTT GGT GAC TTA	52	184
26	mEgCIR2600	GGG GAT GAG TTT GTT TGT TC	CCT GCT TGG CGA GAT GA	52	277
27	mEgCIR3260	AGG GCA AGT CAT GTT TC	TAT AAG GGC GAG GTA TT	52	221
28	mEgCIR3275	GAA GCC TGA GAC CGC ATA GA	TTC GGT GAT GAA GAT TGA AG	52	146
29	mEgCIR3310	ATC ATG GCC GAT CTG TAT TA	GGA ATG CTG GTC ATG GAA TAT A	52	104
30	mEgCIR3399	AGC CAA TGA AGG ATA AAG G	CAA GCT AAA ACC CCT AAT C	52	265
31	mEgCIR3439	ACT TGT AAA CCC TCT TCT CA	GTT TCA TTA CTT GGC TTC TG	52	247
32	mEgCIR3557	ATT CAT TGC CAT TCC CTT CA	TTG TCC CCT CTG TTC ACT CA	52	212

measures the genetic differentiation among all populations was generated from the analysis.

### Inbreeding Coefficient

The inbreeding coefficient ( $F_{IS}$ ) is the probability that two homologous alleles present in the same individual are identical by descent (Allard, 1960). It can be calculated for each population by comparing the  $H_e$  with  $H_o$  using the formula as follows:

$$F_{IS} = \frac{H_e - H_o}{H_e}$$

## RESULTS AND DISCUSSION

### Genetic Variability Parameters

The 32 CIRAD SSR primers produced a total of 230 alleles among the 17 oil palm progenies. The number of alleles scored per SSR ranged between 4

to 11. The average number of alleles per locus was 7.1875. The mean  $H_e$  and  $H_o$  were 0.7063 and 0.5270, respectively. The average of PIC value derived from the 32 SSR markers was 0.6593. The  $F_{IS}$  was 0.2550 based on all the progenies involved in the study. The lower average number of alleles (7.1875) detected as compared to Bakoume *et al.* (2015) probably was due to the variation in number of accessions samples and the origin of the samples from different germplasm collections. Decrease in the average number of alleles per locus is a clue to the loss of genetic variability in the breeding populations.

The genetic variability for each advanced breeding progenies based on each agency was further investigated. The allele frequency parameter analysis for the AVROS populations from DOA Sabah, the AVROS populations from GAT, the Deli populations from DOA Sabah and the Deli populations from GAT is summarised in Table 3. AVROS progenies from DOA Sabah produced a total of 133 alleles while those of GAT only had a

TABLE 3. SUMMARY OF ALLELE FREQUENCY PARAMETER ANALYSIS

Mean	DOA Sabah		GAT	
	AVROS	Deli	AVROS	Deli
Number of alleles, Na	3 - 6	2 - 5	2 - 5	2 - 7
Expected heterozygosity, $H_e$	0.4128	0.5544	0.5648	0.6399
Observed heterozygosity, $H_o$	0.4485	0.5003	0.5489	0.5660
Polymorphic Information content, PIC	0.3481	0.4816	0.4867	0.4984
Inbreeding coefficient, $F_{IS}$	-0.0810	0.1028	0.0336	0.1306

Note: DOA - Department of Agriculture.

GAT - Genting AgTech Sdn Bhd.

total of 106 alleles. An independent t-test conducted on number of alleles detected from DOA Sabah and GAT showed there was no significant difference. Nevertheless, the same set of markers when used to genotype the more advanced AVROS progenies (GAT) detected 25% less alleles. The number of alleles scored per SSR ranged between 3 to 6 and 2 to 5 for DOA Sabah and GAT, respectively. The mean number of alleles per locus was 4.1563 for DOA Sabah and 3.3125 alleles/locus for GAT. These observations of decreasing values of genetic variabilities indicated the loss of genetic diversity among the AVROS advanced breeding progenies in GAT. The selection process in different research stations led to different selection forces being applied to the respective AVROS populations. As such, due to these selection preferences, different types of alleles were favoured according to the objectives of each breeding programme for AVROS.

As for the Deli progenies, the total alleles detected for DOA Sabah was 109 while for GAT it was 122 alleles. The higher number of alleles was detected in the more advanced Deli progenies (GAT). GAT received the Deli populations from MPOB which has a wider Deli collection of different sub-populations which are known as Breeding Population Restricted Origins (BPRO). BPRO became one of the factors contributing to this observation while DOA Sabah only has a limited collection of Deli populations. The initial genetic pool for breeding the Deli populations from the two agencies was different and this caused the number of alleles detected in them to be different. Besides, this could also be explained by the breeding strategy carried out in MPOB which went through more generations of improvement either through selfing of Deli *dura* or intercrossing of Deli *dura* from different BPRO. The number of alleles detected for DOA Sabah ranged from 2 to 5 while for GAT it ranged from 2 to 7. The mean number of alleles per locus recorded for DOA Sabah and GAT were 3.4063 and 3.8125 respectively. Though there was a slight difference in the range of number of alleles detected among the Deli progenies from DOA Sabah and GAT, the mean number of alleles detected was close to each other. This was due to all of these progenies being the descendants of the historical four Deli palms established at the Bogor Botanical Garden in 1848.

### Genetic Relatedness Analysis

The distribution of the mean log likelihood of  $K$ ,  $L(K)$  from the STRUCTURE v2.3.4 software did not show a clear mode for the true number of populations in this study (Figure 2a). Commonly, once the real  $K$  is achieved,  $L(K)$  at larger  $K$ s plateaus or continues increasing slightly (Figure 2a). The difference between successive likelihood values of  $K$ ,  $L'(K)$  is shown in Figure 2b. Based on Evanno *et al.* (2005), it has been proven that the absolute values of the second order rate of change of the likelihood function,  $|L''(K)|$  has a mode at the true  $K$  for most of the situations investigated. Thus, in the current study,  $|L''(K)|$  was computed and the highest value of  $|L''(K)|$  was identified at  $K=2$  (Figure 2c). The model-based bar plot for the  $K=2$  is shown in Figure 3a indicating cluster 1 is AVROS population while cluster 2 is Deli population. When  $K$  was increased to 6, more detailed population structure was revealed (Figure 3b). The six clusters consisted of GAT D-self, DOA Sabah D-self, DOA Sabah and GAT DxD intercrosses, DOA Sabah AVROS, DOA Sabah Ekona and GAT AVROS. Even though Ekona palms were included in the current study as the control, the sample size of 10 palms was rather low and this might not give an indicative population structure as a whole. Thus, it is advisable to increase the number of samples to a similar sample size if the Ekona population were to be included as part of the diversity study.

The genetic analysis of all the 384 individual palms using the Nei distance (Nei, 1972) detected three major clusters which were AVROS, Deli and Ekona. The three major clusters further branched into several sub-clusters (Figure 4). The first cluster was branched into two sub-clusters which consisted of AVROS of DOA Sabah and GAT. The second cluster was further divided into three sub-clusters which consisted of one illegitimate palm, one sub-cluster of inter-crosses of Deli and one selfing Deli population from both agencies. The third cluster consisted of the Ekona population which was the control of the study. A summary of the populations contained in each cluster is listed in Table 4. This observation corresponded with the STRUCTURE result of  $K=6$  as discussed earlier.

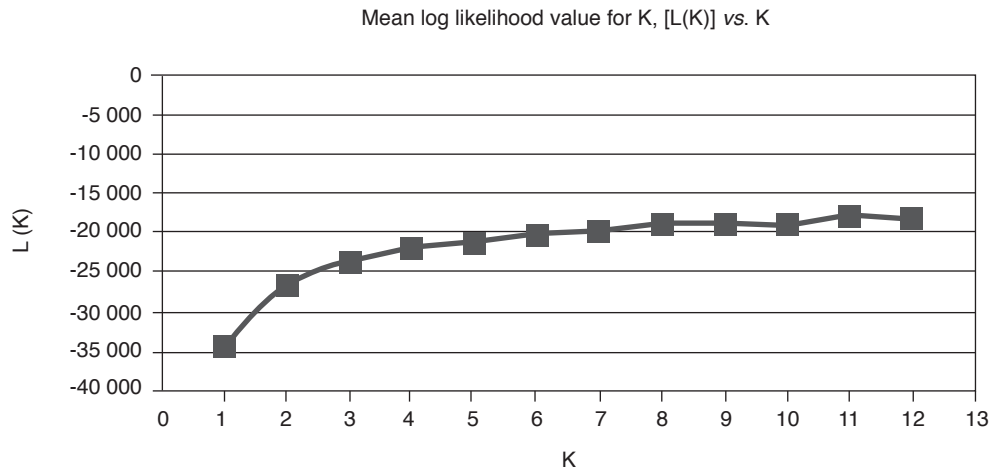


Figure 2a. Distribution of L (K) indicated once the real K is achieved, L (K) at larger K plateaus or continues increasing slightly.

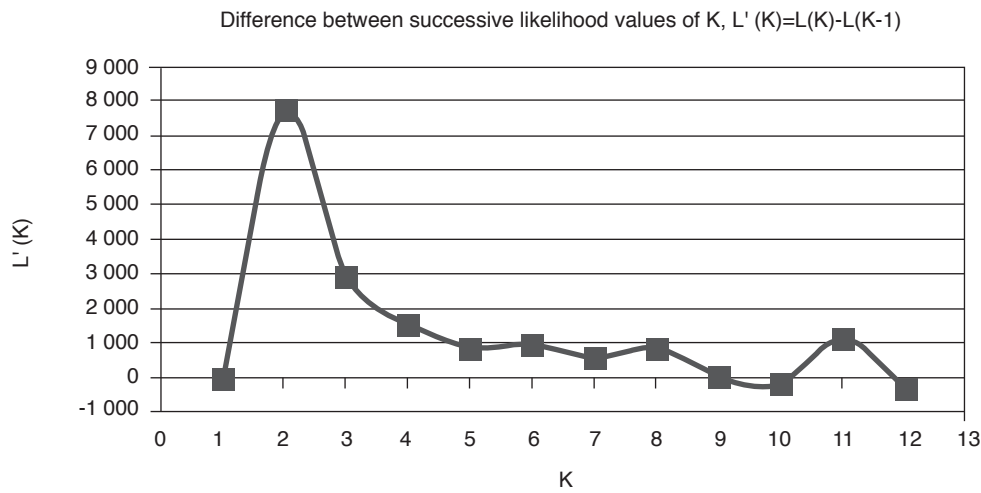


Figure 2b. Distribution of L' (K) indicated by the difference between successive likelihood values of K.

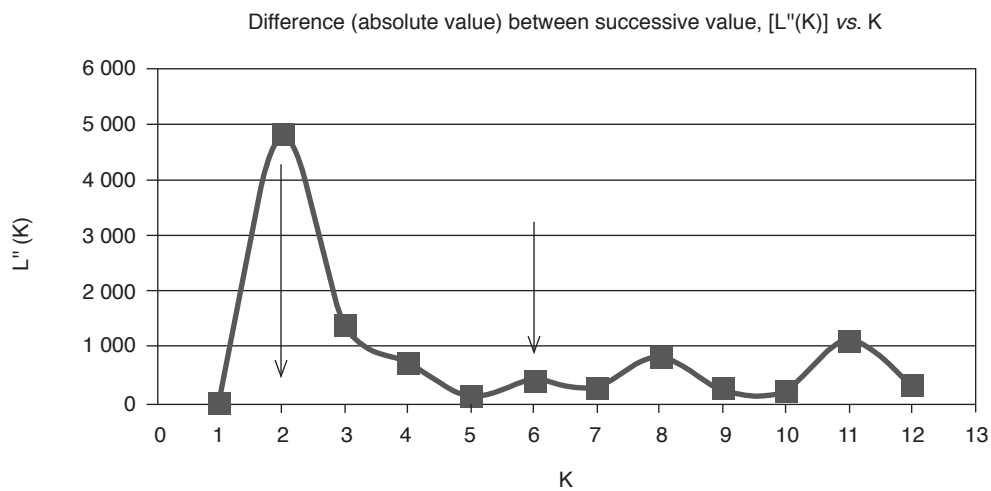
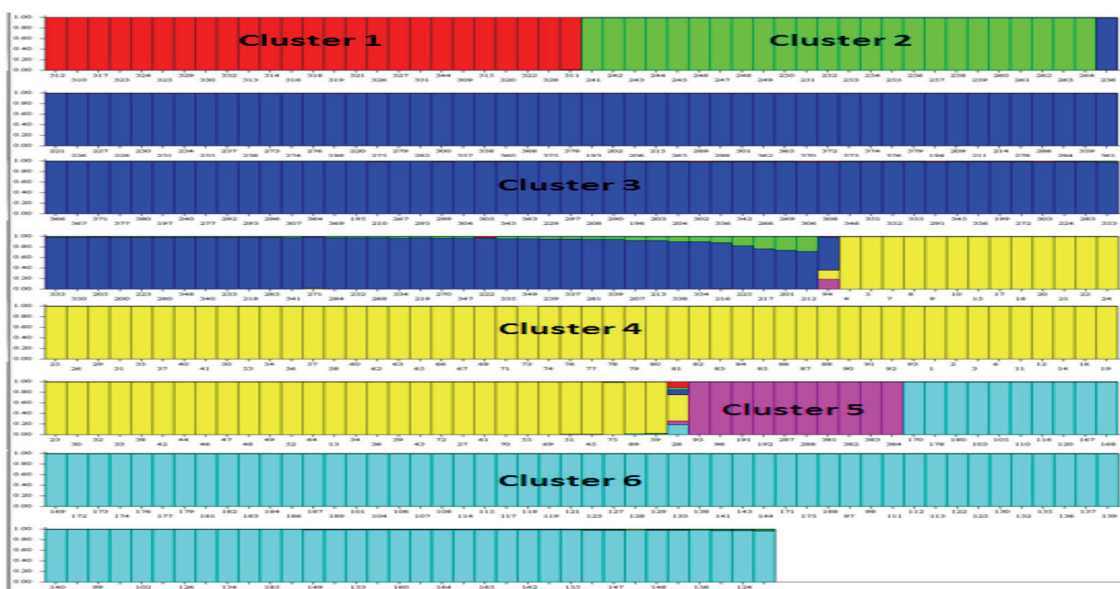


Figure 2c. Distribution of L'' (K) indicated by absolute value of second order rate of change of the likelihood function. Arrows indicated the K = 2 and K = 6.



Figure 3a. Model-based ancestry for oil palm advanced breeding populations indicated Cluster 1 as AVROS while cluster 2 as Deli population (K=2).



Note: DOA - Department of Agriculture. GAT - Genting AgTech Sdn Bhd.

Figure 3b. Model-based ancestry for oil palm advanced breeding populations indicated Cluster 1: GAT D-self, Cluster 2: DOA Sabah D-self, Cluster 3: DOA Sabah and GAT Dx D, Cluster 4: DOA Sabah AVROS, Cluster 5: DOA Sabah Ekona, Cluster 6: GAT AVROS (K=6).

TABLE 4. SUMMARY OF CLUSTERING OF INDIVIDUAL-BASED GENETIC DISTANCE OF NEI 1972

Cluster	Population	Sub-cluster	Population	Population size
Cluster 1	AVROS	1	DOA Sabah AVROS	93
		2	GAT AVROS	94
Cluster 2	Deli	3	Illegitimate palm	1
		4	DOA Sabah & GAT Deli: Inter-crossing	138
		5	DOA Sabah & GAT Deli: Selfing	48
Cluster 3	Ekona		DOA Sabah Ekona	10

Note: DOA - Department of Agriculture. GAT - Genting AgTech Sdn Bhd.

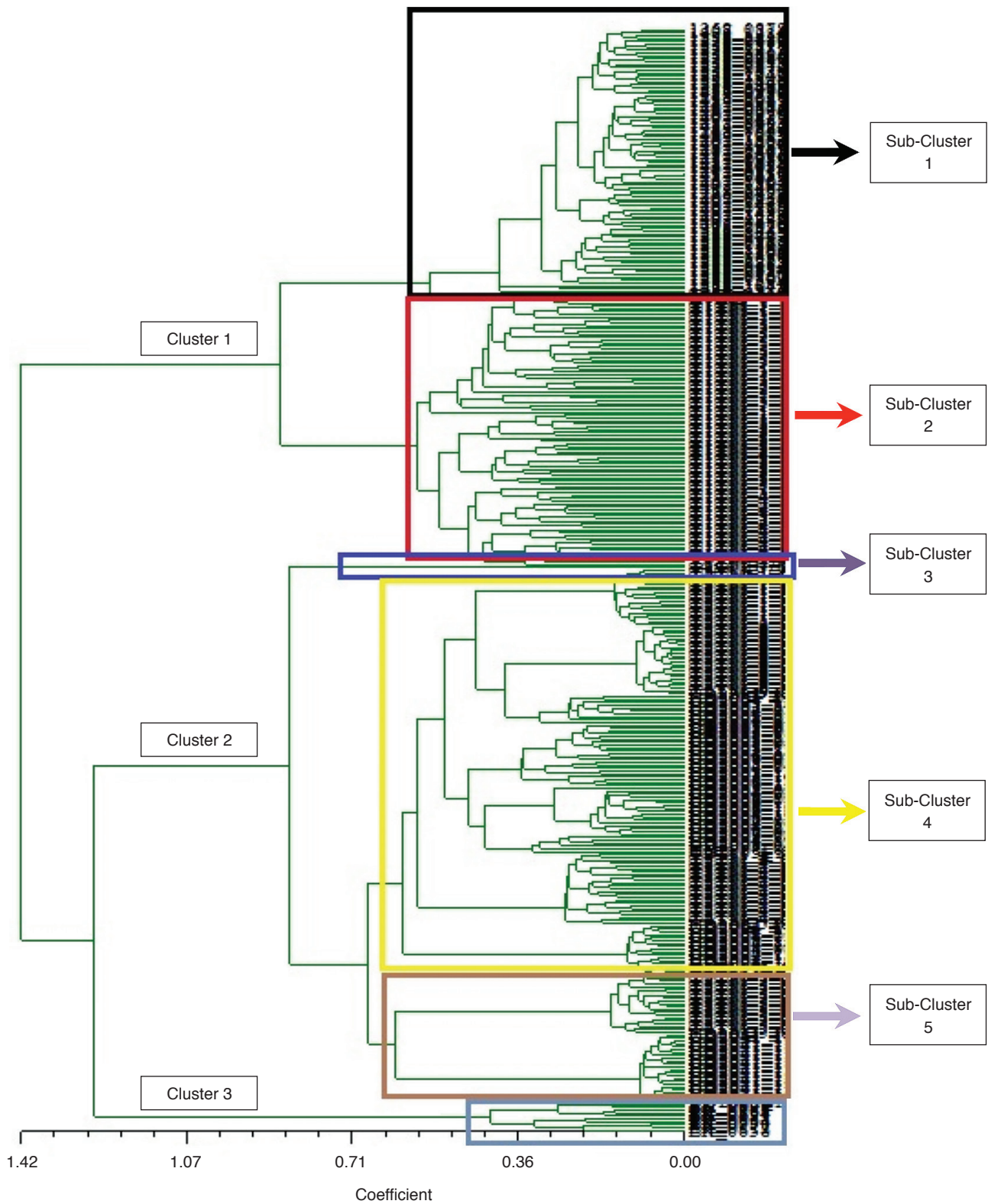


Figure 4. Dendrogram based on Neighbor-Joining clustering of individual palms using the genetic distance of Nei 1972.

Association among all the oil palm populations were revealed by PCA (Figure 5). The results recapitulated the findings from the population structure and Neighbor-Joining analyses. Three distinct clusters could be observed from the PCA, namely AVROS populations, Deli populations and a

cluster consists of the Ekona control. The advanced breeding progenies were redefined based on the above individual-based genetic distance analysis (Figure 4 and Table 4). The redefined population-based distance matrix is as shown in Table 5. Population-based Neighbor-Joining was performed using the

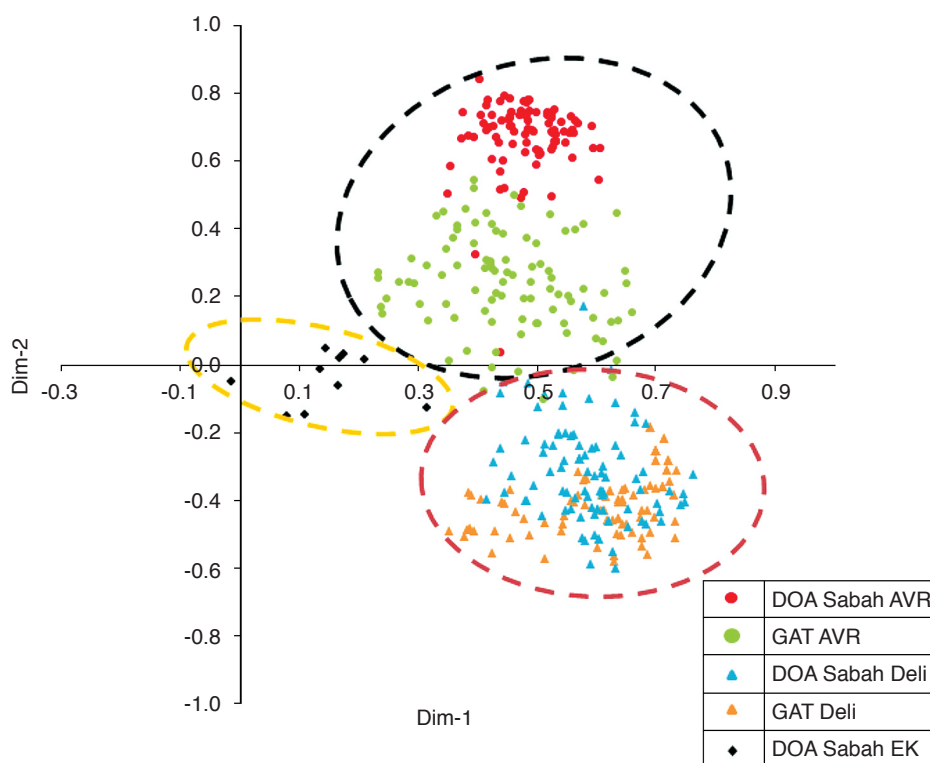


Figure 5. Principal component analysis (PCA) for all 384 palms using 32 Simple Sequence Repeat (SSR) showed the association of three clusters namely AVROS, Deli and Ekona.

PowerMarker software version 3.25 (Liu and Muse, 2005). Population distance was calculated using the Nei distance (Nei, 1972) and was illustrated using the TreeViewX software version 0.5.0 (Figure 6). The population-based Neighbor-Joining tree resembled the individual-based Neighbor-Joining tree.

The distribution of molecular genetic variation among populations listed in Table 4 identified using the Nei distance analysis was estimated by AMOVA. The suspected illegitimate was excluded from the AMOVA. Calculation of Wright's F statistics for all polymorphic SSR loci revealed that 65.41% of the genetic variation was due to the differentiation within populations (Table 6). The estimation of variation among populations was 0.3459 ( $F_{ST}$ ) and this was interpreted as very great genetic differentiation according to the classification proposed by Wright (1978).

#### Inbreeding Coefficient ( $F_{IS}$ ) for Different Generations of populations

The  $F_{IS}$  for different generations of populations are as shown in Table 7. Only the AVROS population of DOA Sabah had negative  $F_{IS}$  value ( $F_{IS} = -0.0810$ ) indicating the population had prominent deficiency of homozygote. The AVROS (GAT), Deli *dura* (DOA Sabah) and Deli *dura* (GAT) have positive  $F_{IS}$  values of 0.0336, 0.1028 and 0.1306, respectively indicating these populations are more inbred. The

$F_{IS}$  of AVROS Progeny 1407 ( $F_{IS}$ ) was -0.1871 while AVROS Progeny TM0005 in GAT (selfed progeny) was 0.1693 (Table 8).

The descendants from the SP540 *tenera* palms gave rise to AVROS populations, BM119, at the Oil Palm Research Station, Banting, Selangor (OPRS Banting), Malaysia in 1959. Family BM 119 consisted of 36 palms including nine *pisifera* which were progeny tested in PT38 planted in 1968 (Lee and Yeow, 1985). From OPRS Banting, 10 crosses of AVROS (T x T) were distributed to the Department of Agriculture, Serdang (DOA, Serdang) (Dr Lee Chong Hee, oil palm breeder/ex-General Manager of Golden Hope Agrotech Consultancy Sdn Bhd, pers. comm., 30 January 2014) and were planted as Trial 0.79 in 1965. The Palm Oil Research Institute of Malaysia (PORIM) was formed in 1979 and has taken over the oil palm genetic materials of DOA Serdang (Wahid *et al.*, 2005). Later, PORIM merged with PORLA in May 2000 and the new body, the Malaysian Palm Oil Board (MPOB) has taken over the management of the genetic materials ever since.

The genetic stock of AVROS population from PORIM was given as a set (10 T x T inter-crosses) to UDARS, DOA Sabah back in 1979 and were planted as Trial 21 in 1980 (Betty Kwan Keng Wah, ex-Principal Assistant Director of Research of Ulu Dusun, pers. comm., 26 February 2014). Today, the AVROS populations of DOA Sabah are at the third generation while the AVROS populations of

TABLE 5. POPULATION-BASED DISTANCE USING NEI 1972

OTU	A	B	C	D	E	F	G	H
A	0.0000							
B	1.7615	0.0000						
C	1.1559	0.4742	0.0000					
D	1.1964	1.1131	0.9072	0.0000				
E	1.0053	1.2390	0.6037	0.5193	0.0000			
F	0.9371	1.2979	0.6725	0.6658	0.1326	0.0000		
G	1.2153	1.4463	0.8280	1.0721	0.4511	0.4837	0.0000	
H	1.0918	1.8117	1.0274	0.7760	0.3801	0.3044	0.5082	0.0000

A: Cluster3\_DOA Sabah\_Ekona  
 B: Sub-Cluster1\_DOA Sabah\_AVROS  
 C: Sub-Cluster2\_GAT\_AVROS  
 D: Sub-Cluster3\_DOA Sabah\_AVROS  
 E: Sub-Cluster4\_DOA Sabah\_Deli  
 F: Sub-Cluster4\_GAT\_Deli  
 G: Sub-Cluster5\_DOA Sabah\_Deli  
 H: Sub-Cluster5\_GAT\_Deli

TABLE 6. ANALYSIS OF MOLECULAR VARIANCE (AMOVA) OF FIVE OIL PALM POPULATIONS RE-DEFINED USING NEI 1972 DISTANCE LISTED IN Table 4

Source of variation	d.f	Sum of squares	Variance components	Total variation (%)
Among populations	4	2 418.572	4.24119	34.59
Within populations	761	6 103.087	8.01983	65.41
Total	765	8 521.659	12.26101	-

Note:  $F_{ST} = 0.3459$ . d.f. - degree of freedom.

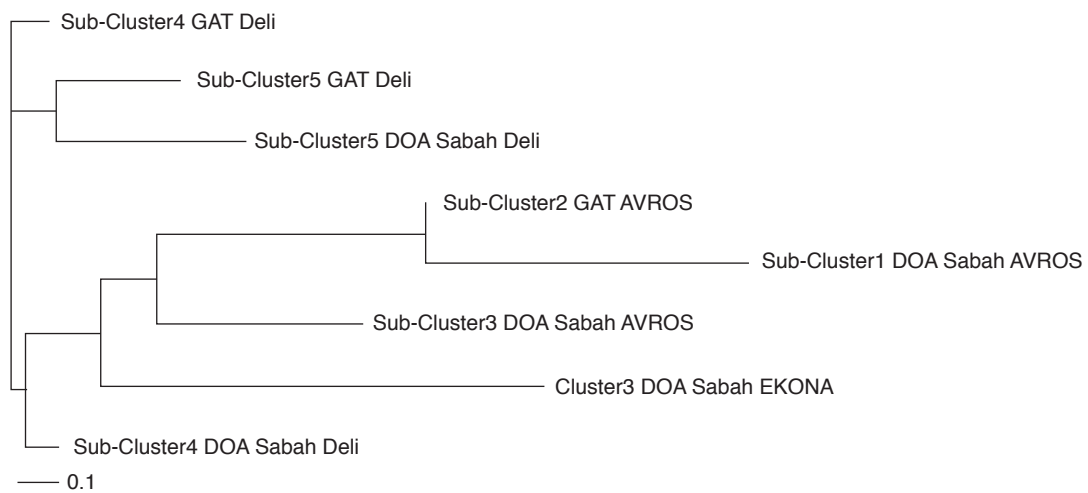


Figure 6. Dendrogram based on Neighbor-Joining clustering of re-defined populations using the genetic distance of Nei 1972. The population-based Neighbor-Joining tree resembled the individual based Neighbor-Joining tree.

TABLE 7. SUMMARY OF INBREEDING COEFFICIENT FOR EACH ADVANCE BREEDING POPULATIONS

Population, agency	Expected heterozygosity, $H_e$	Observed heterozygosity, $H_o$	Inbreeding coefficient, $F_{IS}$
AVROS, DOA Sabah	0.4128	0.4485	-0.0810
AVROS, GAT	0.5648	0.5489	0.0336
Deli <i>dura</i> , DOA Sabah	0.5544	0.5003	0.1028
Deli <i>dura</i> , GAT	0.6399	0.5660	0.1306

Note: DOA - Department of Agriculture. GAT - Genting AgTech Sdn Bhd.

TABLE 8. SUMMARY OF INBREEDING COEFFICIENT FOR 3<sup>RD</sup> (Progeny 1407) and 5<sup>TH</sup> (Progeny TM0005: Selfing) GENERATION OF AVROS

Population, agency	Expected heterozygosity, $H_e$	Observed heterozygosity, $H_o$	Inbreeding coefficient, $F_{IS}$
1407, DOA Sabah	0.3731	0.4505	-0.1871
TM0005, GAT	0.4603	0.3919	0.1693

Note: DOA - Department of Sabah. GAT - Genting AgTech Sdn Bhd.

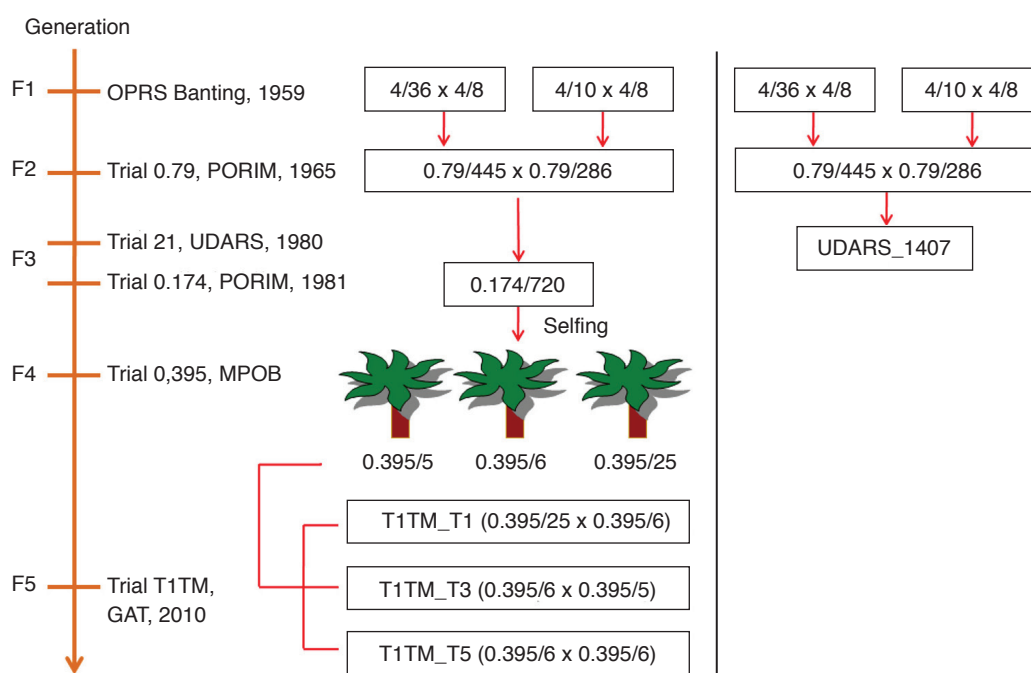
GAT are at fifth generation with the assumption that the AVROS progenies given by OPRS Banting were the first generation of AVROS (Figure 7). The comparison of AVROS at different generations of advanced breeding populations indicated that AVROS at DOA Sabah which is at third generation had a  $F_{IS}$  of -0.0810. The negative sign indicated deficiency of homozygotes while the AVROS at GAT had  $F_{IS}$  of 0.0336 indicating the deficiency in heterozygotes. The number of generations of inbreeding and selection pressure applied by the breeders had directly affected the inbreeding level present in the populations.

Based on personal communications with the MPOB's Senior Research Officer (Noh Ahmad, MPOB Kluang, 5 March 2014), the AVROS populations in GAT used in the current study shared the same ancestor of Progeny 1407 of DOA Sabah. Thus, the inbreeding coefficient of Progeny 1407 against the AVROS progeny in GAT (selfed progeny: AVROS Progeny TM0005) was further analysed in order to detect the inbreeding level using the specific progeny. The  $F_{IS}$  for AVROS Progeny TM0005 (GAT) was 0.1693 whereas AVROS progeny code 1407 (DOA Sabah) was -0.1871. The AVROS paternal lines provided by MPOB to GAT had been selected and improved through intensive breeding programmes. With the aim to develop high yielding planting materials, the breeders worked toward developing inbred lines for AVROS. After going through two cycles of selfing, the selection force on traits of interest had led to a reduction in  $H_o$ .

In the current study, the Deli populations from GAT have also been developed to a more advanced cycle as compared to those from DOA Sabah. Since there was no direct ancestors and descendants involved in any of the progeny involved in the current study, the comparison of inbreeding level of Deli populations was done at the populations by agency level and did not involve a specific progeny. Maizura *et al.* (2006b) described the screening of 16 Deli *dura* inter-crosses (16 DxD crosses) using 10 EST-SSR markers and found that there were still excess in heterozygotes among the populations studied. The excess in heterozygotes implicated the Deli populations used in her study were highly heterozygous and not inbred. This observation was contradicted by the finding in the current study where the Deli populations in GAT had higher inbreeding level ( $F_{IS} = 0.1306$ ) as compared to the Deli populations in DOA Sabah ( $F_{IS} = 0.1028$ ). On the other hand, the inbreeding level reported by Bakoume *et al.* (2015) which involved only one population of Deli MPOB indicated a high inbreeding level ( $F_{IS} = 0.5690$ ).

## CONCLUSION

Sustainable development of a breeding programme depends on having sufficient genetic variability to achieve high genetic gains in each selection cycle. The genetic variability revealed at the molecular level via markers on advanced breeding population,



Note: DOA - Department of Agriculture. GAT - Genting AgTech Sdn Bhd.

Figure 7. Ancestry of AVROS populations in GAT and DOA Sabah. Populations used in the current study were sourced from T1TM and UDARS\_1407.

Deli and AVROS, showed that even through the materials were improved for few generations, there are still presence of variability in some of the progenies while some of the progenies have less variability. Molecular information such as genetic variability between populations, inbreeding level among the progenies and genetic distance provides a better understanding to the breeder and can guide them in decision-making during the breeding programme planning. Importantly, this allows the breeder to focus on developing new populations with acceptable broad genetic base and adaptability. The utilisation of phenotypic assessment coupled with the use of molecular data will lead to great improvement in oil palm breeding while conserving genetic variability for future selection. Inbred line development for AVROS can be fast track through identifying the inbreeding level between the AVROS progenies that have the highest potential to be selected as the parental palms for further improvement. As for the less inbred progenies, they will maintain the variability within and between populations and served as the source for future selection. To a commercial oil palm research company, sources of diverse genetic populations and highly inbred populations are both equally important for crop improvement and seed production, respectively.

#### ACKNOWLEDGEMENT

The authors acknowledge the sponsorship given by Genting Plantations Berhad and permission to publish this article. We thank the Genting AgTech Sdn Bhd, ACGT Sdn Bhd, Department of Agriculture Sabah (DOA Sabah) and MPOB staff for all inputs and discussions on the project.

#### REFERENCES

- ALLARD, R W (1960). Systems of mating and their genetic consequences. *Principles of Plant Breeding*. p. 204-206.
- BAKOUMÉ, C; WICKNESWARI, R; SIJU, S; RAJANAIDU, N; KUSHAIRI, A and BILLOTTE, N (2015). Genetic diversity of the world's largest oil palm (*Elaeis guineensis* Jacq.) field genebank accessions using microsatellite markers. *Genetic Resources and Crop Evolution*, 62(3): 349-360.
- BETTY KWAN KENG WAH (2014). Personal communication. Genting Plantations Office Sabah (GPOS), Sandakan, Malaysia.
- BILLOTTE, N; JOURJON, M; MARSEILLAC, N; BERGER, A; FLORI, A; ASMADY, H; ADON, B; SINGH, R; NOUY, B and POTIER, F (2010). QTL detection by multi-parent linkage mapping in oil palm (*Elaeis guineensis* Jacq.). *Theoretical and Applied Genetics*, 120(8): 1673-1687.
- BILLOTTE, N; MARSEILLAC, N; RISTERUCCI, A; ADON, B; BROTTIER, P; BAURENS, F; SINGH, R; HERRAN, A; ASMADY, H and BILLOT, C (2005). Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). *Theoretical and Applied Genetics*, 110(4): 754-765.
- CHEAH, S C; LCL RAHIMAH, A; ABERLENC-BERTOSSI, F A; ADON, H; AMBLARD, B; BEULE, P; BILLOTE, T; DURAND-GASSELIN, N; DUVAL, T and FLORI, Y (1996). Quality control process for oil palm tissue culture using DNA probes. *PORIM Information Series No. 40*.
- EMANUELLI, F; LORENZI, S; GRZESKOWIAK, L; CATALANO, V; STEFANINI, M; TROGGIO, M; MYLES, S; MARTINEZ-ZAPATER, J M; ZYPRIAN, E; MOREIRA, F M and GRANDO, M S (2013). Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biology*, 13: 13-39.
- EVANNO, G; REGNAUT, S and GOUDET, J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8): 2611-2620.
- EXCOFFIER, L; ESTOUP, A and CORNUET, J M (2005). Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics*, 169(3): 1727-1738.
- HAMBLIN, M T; WARBURTON, M L and BUCKLER, E S (2007). Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PloS ONE*, 2(12): e1367.
- HAYATI, A; WICKNESWARI, R; MAIZURA, I and RAJANAIDU, N (2004). Genetic diversity of oil palm (*Elaeis guineensis* Jacq.) germplasm collections from Africa: Implications for improvement and conservation of genetic resources. *Theoretical and Applied Genetics*, 108(7): 1274-1284.
- ISA, Z; ONG, K; NORASYIKIN, I and SUBOH, O (2008). Performance of MPOB-Nigerian population 12 at Kulim - An update. *Proc. of the 3<sup>rd</sup> Seminar on Performance of PS1 and PS2 Materials and Elite Germplasm*. MPOB, Bangi. p. 139-149.
- JUNAIDAH, J; CHIN, C; RAFII, M Y and SYUHADA, W (2008). Performance and utilization of MPOB-

- Nigerian oil palm materials in FELDA. *Proc. of the 3<sup>rd</sup> Seminar on Performance of MPOB PS1 and PS2 Materials and Elite Germplasm* 2008. p. 73-90.
- LEE CHONG HEE (2014). Personal communication. Technology Park Malaysia, Kuala Lumpur, Malaysia.
- LEE, C and YEOW, K (1985). Progress in breeding and selection for seed production at HMPB Oil Palm Research Station. *The Planter Vol. 61 No. 706 (1985)*: 18-31.
- LEE, C; YONG, Y; DONOUGH, C; CHIU, S; SOH, A; RAJANAIDU, N; BASRI, M; PAMIN, K and MULUK, C (1990). Selection progress in the Deli *dura* population. *Proc. of the workshop on Progress of Oil Palm Breeding Populations*. Permatang Siantar, Sumatra, Indonesia, 25 November 1990. PORIM, Bangi. p. 81-89.
- LUI, K and MUSE, S (2005). PowerMarker: integrated analysis environment for genetic marker data. *Bioinformatics*, 21(9): 2128-2129.
- MAIZURA, I; NORZIHA, A; RAFIL, M Y and SALEH, G (2009). Relationship between genetic distance of oil palm parental palms and performance of their progenies based on microsatellite markers. Paper presented in the GSAS Seminar, 19 June 2009 at MPOB Head Office, Bangi, Selangor.
- MAIZURA, I; RAJANAIDU, N; ZAKRI, A and CHEAH, S (2006a). Assessment of genetic diversity in oil palm (*Elaeis guineensis* Jacq.) using restriction fragment length polymorphism (RFLP). *Genetic Resources and Crop Evolution*, 53(1): 187-195.
- MAIZURA, I; SINGH, R and RAJANAIDU, N (2006b). SSR analysis of genetic diversity among selected oil palm advanced breeding populations. *Proc. of the PIPOC 2011 International Palm Oil Congress and Exhibition*.
- MUSA, B and GURMIT, S (2008). Utilization of MPOB germplasm at United Plantations. *Proc. of the 3<sup>rd</sup> Seminar on Performance of PS1 and PS2 Materials and Elite Germplasm*. MPOB, Bangi. p. 43-60.
- NEL, M (1972). Genetic distance between populations. *American Naturalist*. p. 283-292.
- NOH AHMAD (2014). Personal communication. MPOB Kluang, Johor, Malaysia.
- NORZIHA, A; RAFIL, M; MAIZURA, I and GHIZAN, S (2008). Genetic variation among oil palm parent genotypes and their progenies based on microsatellite markers. *J. Oil Palm Res. Vol. 20(2)*: 533-541.
- OIL WORLD (2016). *Oil World Annual*. ISTA Mielke, Hamburg.
- PRITCHARD, J K; STEPHENS, M and DONNELLY, P (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2): 945-959.
- REN, J; CHEN, L; SUN, D; YOU, F M; WANG, J; PENG, Y; NEVO, E; BEILES, A; SUN, D and LUO, M (2013). SNP-revealed genetic diversity in wild emmer wheat correlates with ecological factors. *BMC Evolutionary Biology*, 13(1): 1.
- ROHLF, F (2005). NTSYS-pc: *Numerical Taxonomy and Multivariate Analysis System*. Version 2.2.
- SINGH, N; CHOUDHURY, D R; SINGH, A K; KUMAR, S; SRINIVASAN, K; TYAGI, R; SINGH, N and SINGH, R (2013). Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS ONE*, 8(12): e84136.
- SINGH, R; ZAKI, N M; TING, N; ROSLI, R; TAN, S; LOW, E L; ITHNIN, M and CHEAH, S (2008). Exploiting an oil palm EST database for the development of gene-derived SSR markers and their exploitation for assessment of genetic diversity. *Biologia*, 63(2): 227-235.
- VAN INGHELANDT, D; MELCHINGER, A E; LEBRETON, C and STICH, B (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theoretical and Applied Genetics*, 120(7): 1289-1299.
- VIGNAL, A; MILAN, D; SANCRISTOBAL, M and EGGEN, A (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34(3): 275-306.
- WAHID, M B; ABDULLAH, S N A and HENSON, I (2005). Oil palm - Achievements and potential. *Plant Production Science*, 8(3): 288-297.
- WEIR, B S and COCKERHAM, C C (1984). Estimating F-statistics for the analysis of population structure. *Evolution*. p. 1358-1370.
- WEIR, B S and COCKERHAM, C C (1996). *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Assoc. Inc., Sunderland, MA, USA.
- WRIGHT, S (1978). *Evolution and the Genetics of Populations: A Treatise in Four Volumes. Vol. 4: Variability within and among Natural Populations*. University of Chicago Press.