Analysing Population Structure of *Elaeis Oleifera* Germplasm using Model-based Approach Programme STRUCTURE

Wan Nurhayati Wan Hanafi[1], Farida Zuraina Mohd Yusof[1, 2], Rajinder Singh[3], Ahmad Kushairi Din[3], Rajanaidu Nookiah[3], Maizura Ithnin[3]

[1]Faculty of Applied Sciences, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia.
[2]Integrative Pharmacogenomics Institute (iPROMISE), Level 7, FF3 Building, UiTM Puncak Alam Campus, 42300 Bandar Puncak Alam, Selangor, Malaysia.
[3]Advance Biotechnology and Breeding Centre (ABBC), Malaysian Palm Oil Board (MPOB), No. 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia.
E-mail: [3]maizura@mpob.gov.my

Accepted: 31 May 2017
Received: 26 January 2017

ABSTRACT

Elaeis oleifera serves as a source of genetic foundation in oil palm improvement programme, as it possess several interesting agronomic traits such as slow growth, higher oil unsaturation and disease resistance. Malaysian Palm Oil Board (MPOB) has developed a collection of simple sequence repeats (SSRs) from Elaeis oleifera genome (E. oleifera-gSSRs). A total of 21 polymoprhic SSR markers were evaluated in the attempt to assess the population structure of E. oleifera populations. The appropriate common ancestry (K) value was determined to be seven from the likelihood scores. The profile from STRUCTURE analysis indicates considerable sharing of genetic components among E. oleifera population with an exception for Population 01 from Columbia and Population 02 from Costa Rica. The present study provides information on population structure of MPOB E. oleifera collection via model-based method for germplasm conservation and utilisation in breeding programmes.

Keywords: *E. oleifera, population structure, E. oleifera-gSSRs, STRUCTURE analysis*

INTRODUCTION

There are two oil palm species, namely *E. guineensis* and *E. oleifera*. Both species are diploid with chromosome number 2*n*=32 [1]. *E. guineensis*, which originates from Central and West Africa, is well known as the commercial oil palm due to its high oil yield of 4-5 tons/ha [2]. *E. oleifera*, the second species has its origins in South and Central America. *E. oleifera* possess lower oil yield, but exhibits certain favorable characteristics such as slower vertical growth, higher unsaturated fatty acids and tolerance to *Fusarium* wilt, lethal yellowing and bud rot when compared to *E. guineensis* [3]. These are the characteristics that can be used to improve commercial planting materials [4].

Generally the information on diversity and population structure of *E. oleifera* germplasm is still limited [5-7] compared to the *E. guineensis* germplasm which has been well characterised using variety of molecular markers such as isozymes [8]; RFLPs [9], SSRs derived from expressed sequence tags [10], genomic based SSRs [11] and SNPs [12]. The present study aims to unravel the population structure of MPOB *E. oleifera* germplasm collections using the model-based method.

MATERIALS AND METHOD

Plant material used in this study are part of MPOB *E. oleifera* germplasm collected at different sites in Central and South America: The samples used in the analysis were Population 01 (*n*=29) and Population 08 from Columbia (*n*=27). Population 03 (*n*=30) and Population 05 from Panama (*n*=26) were also utilised. The populations from Costa Rica were Population 02 (*n*=25) and Population 21 (*n*=25). Population 02 (*n*=30) and Population 03 from Honduras also formed part of the analysis. *E. oleifera* genomic deoxyribonucleic acid (DNA) was extracted and quantified for DNA concentration and DNA purity using Thermo Scientific µDrop Plate compatible with Thermo Scientific Multitask Go and ScanIt Software, following the manufacturer's protocol. The quality of DNA was further

confirmed by digestion with restriction enzyme. DNA was diluted to give a working stock of approximately 50 ng/µl.

The amplification was carried out by means of polymerase chain reaction (PCR) in 10 µl of reaction mixture containing DNA template, 10 x standard *Taq* reaction buffer (New England Biolabs, UK), 10 mM dNTPs (New England Biolabs), 5 U *Taq* DNA polymerase (New England Biolabs), 0.25 uM of forward primer with M13 tail, 0.25 uM reverse primer and 0.38 uM of selected fluorescent dye namely 6-FAM (blue), VIC (green), PET (red) and NED (yellow). The optimised amplification protocol was applied as follows: an initial denaturation at 94^0C for ten minute, followed by 35 cycles of denaturation at 94^0C for 30 second, annealing at 52°C for one minute, primer extension at 72^0C for one minute, followed by a final extension at 72^0C for ten minute. PCR was performed using GeneAmp System 9700 (Applied Biosystem). The presence of amplified PCR products were confirmed by electrophoreses in 4% SFR gel prepared in TAE buffer together with a 100 bp DNA marker (ThermoScientific, USA). Visualisation was carried out using gel documentation system (AlphaInnotech) after staining with ethidium bromide (0.5 µg/ml).

For the final analysis, a maximum of four PCR products, each labelled with 6-FAM, VIC, PET and NED were multiplexed at a ratio of 1:1:1:1. Two-ul of the multiplexed mix was denatured in 7.84 ul Hi-Di™ Formamide (Applied Biosystems, UK) and 0.16 ul GeneScan™ 400HD ROX™ Size Standard (Applied Biosystems, UK). The denatured sample was then fragmented and size-called on the ABI3100 genetic analyser [13]. All the raw data were imported from the database of ABI PRISM® 3100 data collection software to GeneMapper v4.1 software and finally generated in size (bp) for scoring purposes.

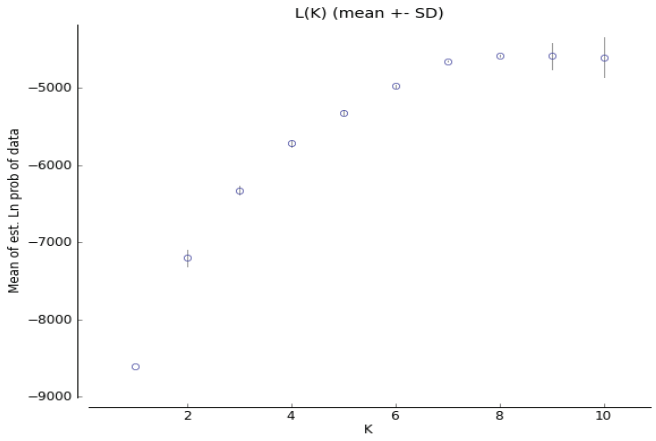
The membership of each genotype was run for range of genetic clusters from value of K=1 to 10 with the admixture model and correlated allele frequency using the STRUCTURE v2.3.4 [14]. For each K it was replicated 20 times. Each run was implemented with a burn-in-period of 50000 steps followed by 50000 Monte Carlo Markov Chain (MCMC) replicates. Maximum likelihood (LnPD) derived for each K and then plotted to find the plateau of the delta K (ΔK) values. Online programme known as 'Structure Harvester' was used to define number of subpopulations based

on the Evanno method [15], available at http://taylor0.biology.ucla.edu/structureHarvester/.

RESULT AND DISCUSSION

In this study, the genetic architecture of diverse germplasm was evaluated by model based clustering approach using the 21 *E. oleifera*-gSSRs. Estimating the suitable assumed population or the best K value in STRUCTURE is based on maximum likelihood (LnPD) and Delta K (ΔK) inferred from Structure Harvester (Figure 1). Plot of the mean likelihoods per K value shows that the K was approaching L(K) pleateaus at K=7. In ΔK plot, it shows that a clear peak was at K=7. The number of groups (K) obtained using the Evanno method (Table 1) indicates that the highest Delta K is best fit at K=7 with Delta K value of 43.739.

(a)



(b)

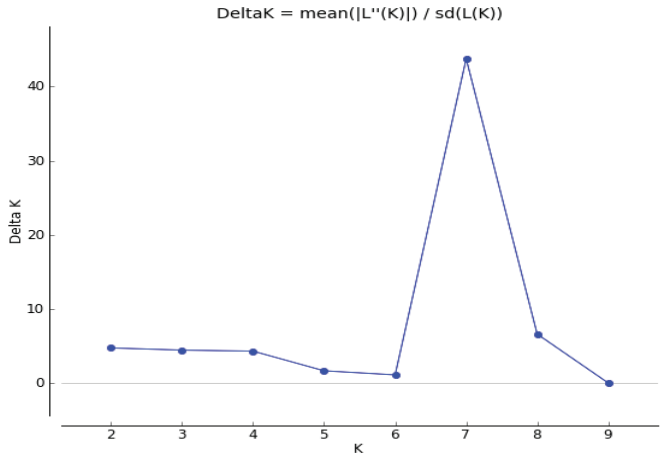


Figure 1: The output data from STRUCTURE generated using Structure Harvester (a) plot of the mean likelihoods per K value (b) Number of groups (K) indicated by the highest Delta K

Table 1: Table output of the Evanno method results

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	20	-8607.390	0.253	—	—	—
2	20	-7202.820	110.193	1404.570	527.050	4.783
3	20	-6325.300	60.528	877.520	271.675	4.488
4	20	-5719.455	46.960	605.845	204.210	4.349
5	20	-5317.820	35.653	401.635	60.715	1.703
6	20	-4976.900	15.794	340.920	18.210	1.153
7	20	-4654.190	5.723	322.710	250.310	43.739
8	20	-4581.790	11.648	72.400	77.415	6.646
9	20	-4586.805	172.143	-5.015	10.400	0.060
10	20	-4602.220	255.114	-15.415	—	—

The estimated membership coefficient of the analysed individuals in each cluster, K is represented by bar plot. Figure 2 shows a bar plot drawn in single line sorted by Q. Colour represent ancestry in each of the seven designated clusters. Colour represent ancestry in each of the seven designated clusters. Vertical bar represent individual *E. oleifera* ancestry: a single colour indicates pure ancestry in a given cluster and multiple colours indicate mixed ancestry. The Y-axis displays the estimated ancestry of each individual to a particular subpopulation. Overall proportion of membership of the sample in each of the seven clusters are: Inferred Cluster 1 = 0.175 (red) , Cluster 2= 0.166 (green), Cluster 3= 0.208 (blue), Cluster 4 = 0.114 (yellow), Cluster 5= 0.165, Cluster 6= 0.031 and Cluster 7= 0.142 (orange). In order to assign population, threshold level usually set at 80 % [16] but might vary between different research groups [17]. From this study, each of the seven optimal clusters has a considerable proportion of mixed memberships sharing among clusters.

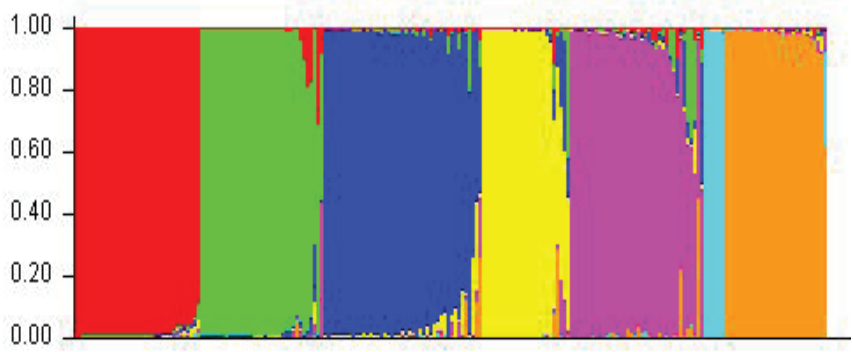


Figure 2: Clustering of ancestry values from STRUCTURE analysis for K=7

Further analysis on bar plot in the original order with K value ranging from K=2 to K=7 provides a better insight into the population structure of *E. oleifera* (Figure 3). K-values of 2-7 subpopulations are shown to right and naming of *E. oleifera* populations are as follow: C01-Population 01 from Columbia, C08-Population 08 from Columbia, P03-Population 03 from Panama, P05-Population 05 from Panama, K02-Population 02 from Costa Rica, K21-Population 21 from Costa Rica, H02-Population 02 from Honduras and H03-Population 03 from Honduras.

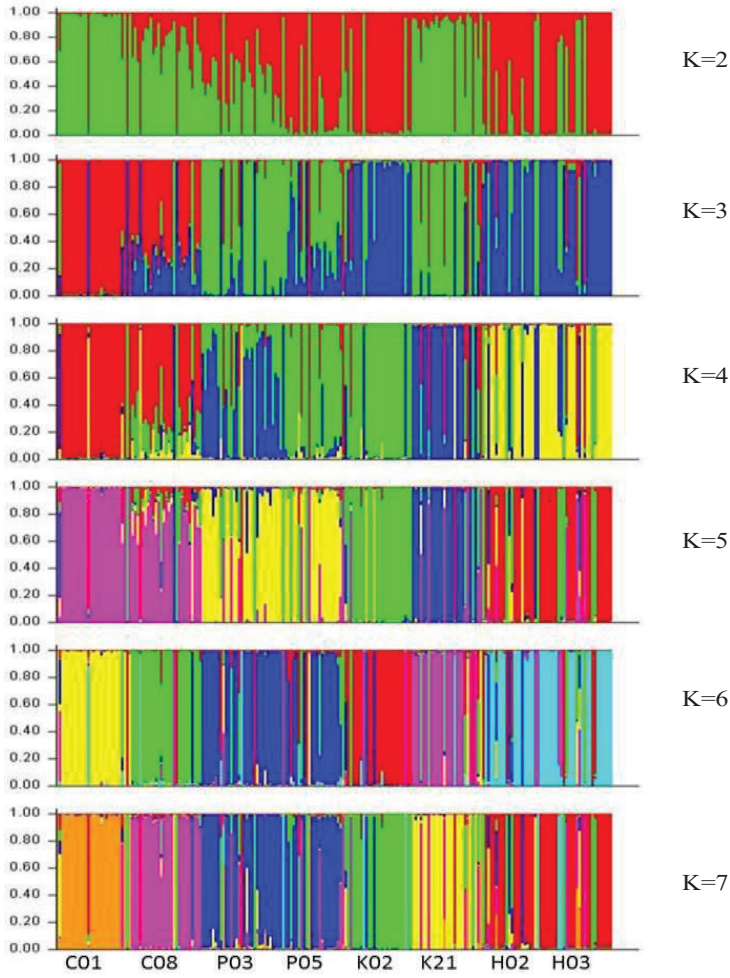


Figure 3: Bar plot in the original order K=2 to K=7

In present study, setting K=2 results in Population 01 from Columbia and Population 21 from Costa Rica fitting into one group which are predominantly in green while the other group in red consists of Population 05 from Panama, Population 02 from Costa Rica as well as Population 02 and Population 03 from Honduras. Population 08 from Columbia and Population 03 from Panama exhibit a genetic mixture of red and green groups.

For K=7, eight populations of *E. oleifera* were roughly divided into six groups. Group 1 consists of Population 01 from Columbia, denoted as the orange. Group 2 has Population 08 from Columbia and is denoted as magenta. Group 3 covers Populations 03 and 05 from Panama denoted in blue. Group 4 has Population 02 from Costa Rica which is green, while Group 5 is Population 21 from Costa Rica indicated as yellow. Finally, Group 6 includes Population 02 and 03 from Honduras (colour red).

The placement of Population 01 from Columbia and Population 02 from Costa Rica were more uniform with increasing K value while the remaining populations exhibit genetic mixture. Overall, no populations showed absolute uniformity. Least admixture was detected for populations C01 and and K02, based on the bar plots of all the K values. Relatively more admixture was observed in populations P03, P05, K21, H02 and H03.

CONCLUSION

In this study, 21 polymorphic *E. oleifera* genomic-SSR markers were used to reveal the population structure of selected *E. oleifera* populations. Model based clustering approach showed least admixture for populations from Colombia and Costa Rica. Populations from Panama and Honduras respectively indicated similarity within country but showed more admixture than those from Colombia and Costa Rica.

ACKNOWLEDGEMENT

The authors wish to thank the Director-General of MPOB for permission to publish this article.

REFERENCES

[1] M. Madon, M. M. Clyde and S. C. Cheah. 1998. Cytological analysis of *Elaeis guineensis* and *Elaeis oleifera* chromosomes, *Journal of Oil Palm Research, Vol. 10*, pp. 68–91.

[2] D. J. Murphy. 2014. The Future of Oil Palm as a Major Global Crop: Opportunities and Challenges, *Journal of Oil Palm Research, Vol. 26*, pp. 1–24.

[3] B. Cochard, P. Amblard and T. Durand-gasselin. 2005. Oil palm genetic improvement and sustainable development, *OCL, Vol. 12*, pp. 141–147. https://doi.org/10.1051/ocl.2005.0141

[4] I. Maizura, R. Singh and D. Kushairi. 2011. *Elaeis, in Wild Crop Relatives: Genomic and Breeding resources, Plantation and Ornamental Crops*, C. Kole, Ed. Verlag Berlin Heidelberg: Springer, pp. 113-124.

[5] E. Barcelos, P. Amblard, J. Berthaud and M. Seguin. 2002. Genetic diversity and relationship in American and African oil palm as revealed by RFLP and AFLP molecular markers, *Pesq. Agropec. Bras., Brasilia, Vol. 37*, pp. 1105–1114. http://dx.doi.org/10.1590/S0100-204X2002000800008

[6] C. K. Teh. 2010. Genetic diversity of Central and South American Wild Oil Palm (*E. oleifera*), Unpublished Master Thesis, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia.

[7] N. Mohd Zaki, R. Singh, R. Rosli and I. Ismail. 2012. *Elaeis oleifera* Genomic-SSR Markers : Exploitation in Oil Palm Germplasm Diversity and Cross-Amplification in Arecaceae, *Int. J. Mol. Sci., Vol. 13*, pp. 4069-4088. https://doi.org/10.3390/ijms13044069

[8] A. Hayati, R. Wickneswari, I. Maizura and N. Rajanaidu. 2004. Genetic diversity of oil palm (*Elaeis guineensis* Jacq.) germplasm collections from Africa: implications for improvement and conservation of genetic resources, *Theor Appl Genet., Vol. 108*, pp. 1274–1284. https://doi.org/10.1007/s00122-003-1545-0

[9] I, Maizura, N. Rajanaidu, A. Zakri and S. Cheah. 2006. Assessment of genetic diversity in oil palm (*Elaeis guineensis* Jacq.) using Restriction Fragment Length Polymorphism (RFLP), *Genet. Resour. Crop EVol., Vol. 53*, pp. 187–195. https://doi.org/10.1007/s10722-004-4004-0

[10] Y. Zulkifli, I. Maizura and R. Singh. 2012. Evaluation of MPOB Oil Palm Germplasm (*Elaeis guineensis*) Populations using EST-SSR, *Journal of Oil Palm Research, Vol. 24*, pp. 1368–1377.

[11] C. Bakoume, R. Wickneswari, R., S. Siju, N. Rajanaidu, A. Kushairi and N. Billotte. 2015. Genetic diversity of the world's largest oil palm (*Elaeis guineensis* Jacq.) field genebank accessions using microsatellite markers, *Genet. Resour. Crop EVol., Vol. 62*, pp. 349–360. https://doi.org/10.1007/s10722-014-0156-8

[12] P. W. Ong, I. Maizura, N. A. Abdullah, M. Y. Rafii, L. C. L. Ooi, E. T. L. Low and R. Singh. 2015. Development of SNP markers and their application for genetic diversity analysis in the oil palm (*Elaeis guineensis*), *Genetics and Molecular Research, Vol. 14*, pp. 12205–12216. https://doi.org/10.4238/2015.October.9.9

[13] N. C. Ting, Y. Zulkifli, K. Katialisa, S. Mayes, F. Massawe, R. Sambanthamurthi, J. Jansen, E. T. L. Leslie, I. Maizura, A. Kushairi, X. Arulando, R. Rosli, K.L. Chan, A. Nadzirah, , K. Sritharan, C. C. Lim, N. Rajanaidu, A. Mohd Din and R. Singh. 2016. Fine-mapping and cross valudation of QTLs linked to fatty acid composition in multiple independent interspecific crosses of oil palm, *BMC Genomics, Vol.17*, pp. 1-17. https://doi.org/10.1186/s12864-016-2607-4

[14] J. K. Pritchard, M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data, *Genetics Society of America, Vol. 155*, pp. 945 - 959.

[15] G. Evanno, S. Regnaut and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study, *Molecular Ecology, Vol. 14*, pp. 2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

[16] V. V. Nachimuthu, R. Muthurajan, S. Duraialaguraja and R. Sivakami. 2015. Analysis of Population Structure and Genetic Diversity in Rice Germplasm Using SSR Markers : An Initiative Towards Association Mapping of Agronomic Traits in Oryza sativa Analysis of Population Structure and Genetic Diversity in Rice Germplasm Using SSR, *Rice, Vol. 8*, pp. 1–24. https://doi.org/10.1186/s12284-015-0062-5

[17] A. M. Liakat, A. M. McClung, M. H. Jia, J. A. Kimball, S. R. McCouch and G.C. Eizenga. 2011. A Rice Diversity Panel Evaluated for Genetic and Agro-Morphological Diversity between Subpopulations and its Geographic Distribution, *Crop Science, Vol. 51*, pp. 2021–2036. https://doi.org/10.2135/cropsci2010.11.0641