Analysis of genetic diversity and population structure of oil palm (*Elaeis guineensis*) from China and Malaysia based on species-specific simple sequence repeat markers

L.X. Zhou*, Y. Xiao*, W. Xia and Y.D. Yang

Hainan Key Laboratory of Tropical Oil Crops Biology/Coconut Research Institute, Chinese Academy of Tropical Agricultural Sciences, Wenchang, Hainan, China

*These authors contributed equally to this study.
Corresponding author: L.X. Zhou
E-mail: glzz_2009@163.com

Genet. Mol. Res. 14 (4): 16247-16254 (2015)
Received June 11, 2015
Accepted September 26, 2015
Published December 8, 2015
DOI http://dx.doi.org/10.4238/2015.December.8.15

ABSTRACT. Genetic diversity and patterns of population structure of the 94 oil palm lines were investigated using species-specific simple sequence repeat (SSR) markers. We designed primers for 63 SSR loci based on their flanking sequences and conducted amplification in 94 oil palm DNA samples. The amplification result showed that a relatively high level of genetic diversity was observed between oil palm individuals according a set of 21 polymorphic microsatellite loci. The observed heterozygosity (H_o) was 0.3683 and 0.4035, with an average of 0.3859. The H_o value was a reliable determinant of the discriminatory power of the SSR primer combinations. The principal component analysis and unweighted pair-group method with arithmetic averaging cluster analysis showed the 94 oil palm lines were grouped into one cluster. These results demonstrated that the oil palm in Hainan Province of China and the germplasm introduced from Malaysia may be from the same source. The SSR protocol was effective and reliable for assessing the genetic diversity of oil palm. Knowledge of the genetic

diversity and population structure will be crucial for establishing appropriate management stocks for this species.

Key words: Oil palm; SSR markers; Average genetic distance; Genetic structure

INTRODUCTION

Oil palm, *Elaeis guineensis* Jacq., is a perennial, monocotyledonous, monoecious, cross-pollinating species belonging to the Arecaceae family. The only other species in the genus *Elaeis* is the American oil palm, *Elaeis oleifera* (Montoya et al., 2014). Both the species have 16 chromosome pairs (2n = 32). Commercial *E. guineensis* (African oil palm) originated in intertropical Africa (Ting et al., 2014) and was imported into South Asia, where its industrial plantations started approximately 100 years ago. It has become one of the most important crops in Indonesia and Malaysia (Abram et al., 2014; Low et al., 2014). The history of oil palm introduction into China is more than 80-years-old. In 1926, the oil palm seeds were brought from Indonesia and Malaysia into China, and grown in Danzhou, Wanning, and other places of Hainan Province. After 1960, oil palm was widely planted in the Hainan Province (Corley and Tinker, 2003; Zhang et al., 2009; Xiao et al., 2014). Because of huge developmental potential, commercial cultivation of *E. guineensis* for palm oil production has come in focus. However, low temperature in these regions (generally lower than 20°C) results in slowing of flower bud differentiation, low fruit yield, and poor economy, subsequently severely affecting the development of oil palm industry in China (Ferwerda, 1977; Yunus et al., 2010; Lei et al., 2014). In-depth study of the species, focused on understating traits to improve breeding and cultivation, has significance and value in industry development. Being a commercial species, analyses of the genetic diversity and structure of oil palm are particularly important for the protection of germplasm resources, identification of oil palm populations, exploration of plant genetic resources and development of future breeding programs.

In plant genetic and structural studies, DNA-based assays, especially molecular markers, are recognized as efficient tools for genetic diversity assessment and population structure identification, molecular ecology studies, as well as for marker-assisted selection (Feng et al., 2009; Zaki et al., 2012). Among all the available molecular markers, simple sequence repeats (SSR) are considered one of the most efficient, providing abundant genetic information due to their co-dominant inheritance, a multi-allelic nature, chromosome specific location, high mutation rate, and ease of scoring. Especially, they are easily assayed using polymerase chain reaction (PCR; Powell et al., 1996; Qin et al., 2014). Currently, SSR markers are one of the most promising molecular marker systems for understanding the genetic diversity and structure of oil palm (Singh et al., 2008).

In this study, SSR was used to explore the genetic diversity and population structure of 2 oil palm populations and to define relationships among different provenances. The study provides a theoretical and experimental basis for future breeding, evaluation, management, and conservation of oil palm.

MATERIAL AND METHODS

Sample collection

Samples of the 2 populations (94 oil palm lines) of oil palm used in the study were collected

from two different areas of Hainan Province located in Southern China: one was composed of oil palm collected from the Coconut Institution in Wenchang city; the other including germplasm introduced from Malaysia. The distance between the individual trees in the same population was at least 50 m to avoid collecting ramets from the same genetic individual. Young, healthy leaves were collected and stored at -80°C in the Molecular Laboratory of the Coconut Research Institute.

DNA extraction

DNA samples were prepared from young leaves of the oil palm trees using the mini-CTAB method (Stewart and Via, 1993). The quality of DNA was tested by electrophoresis on a 1% agarose (w/v) gel. DNA concentration was measured using an ultraviolet spectrophotometer and was adjusted to 50 ng/μL. DNA was stored at -20°C and used subsequently for PCR amplification.

PCR amplification and electrophoresis

The software Msatfinder was used to identify gene-based SSRs across *Elaeis guineensis* transcriptome, and primers flanking SSRs were designed using Primer 3.0 software (Rozen and Skaletsky, 2000). Twenty four individual plant materials from different populations were used for the initial screening. Sixty-three polymorphic SSR loci from our library were used in this study. The 21 polymorphic SSR primer sequences are listed in Table 1.

PCR amplification was carried out in a total volume of 10 μL reaction mixture containing 2 μL 50 ng/μL genomic DNA, 1 μL 10X PCR buffer, 0.8 μL 25 mM $MgCl_2$, 0.2 μL 1 U Taq DNA polymerase (TaKaRa, China), 0.5 μL 0.5 μM of each primer, and 0.2 μL 0.2 mM dNTP mix. PCR amplification began with a 5 min denaturation at 94°C, followed by 35 cycles of 94°C for 30 s, 58.4°C for 30 s and 72°C for 30 s for elongation, and a final extension for 7 min at 72°C. PCR products were electrophoretically separated on 1% denaturing polyacrylamide gels and visualized by silver staining. Product sizes were determined by comparison with a 100-bp DNA ladder.

Data analysis

The results of the PCR amplified fragments were scored with 1 for presence and 0 for absence. When a binary data matrix was generated, the original document was converted with Popgene 1.31 for genetic diversity parameter calculation, including the calculation of observed heterozygosity (H_O) of polymorphic markers, Shannon's information index (I), and the observed number of alleles (N_A; Shen et al., 2010; Xiao et al., 2014; Yu and Cheng, 2014). The NTsys2.1 software was used to construct the principal component analysis (PCA), which presents the distribution of populations (Cameron et al., 2015; Wang et al., 2015). The unweighted pair-group method with arithmetic averaging (UPGMA) dendrogram was constructed from the distance matrix imported from PowerMarker V3.25 using MEGA 4 (Sneath and Sokal, 1973; Nei and Takezaki, 1983; Tamura et al., 2007).

RESULTS

SSR polymorphism and genetic variation

To evaluate the genetic variation between the two oil palm species (31 oil palm

lines were from the Coconut Institution in Hainan Province and 63 germplasms were the introductions from Malaysia), the analysis based on 21 SSR loci was performed (Table 2). The results showed that the 21 SSR loci were all polymorphic in the populations studied. The H_O of oil palm from the Coconut Institution was 0.3683, while the oil palm from Malaysia generated a higher H_O (0.4035). Furthermore, the observed N_A (2.4688) and I (0.611) obtained for oil palm from the Coconut Institution were both lower than that obtained for the species from Malaysia.

Table 1. Primer sequences used in SSR analysis of oil palm.

No.	Primer sequences	Tm (°C)	Size (bp)	Motif
1	F: TCCCTCTCACGCTCTCTGTT R: CTGGTGTGCCAACCTAAACC	58.4	202	$tct_{(6)}$
2	F: CCGGCTCAAGATCCAAAG R: ACTAGCGAGCCACTGAGAGC	58.4	213	$gcc_{(7)}$
3	F: GAAACGTTGGATCCATAGCAA R: GGACTAGCCTTTACTCATCAAAATG	58.4	215	$a_{(15)}$
4	F: ATTTGCAGTTGCAGGGTTCT R: GCAGCAGCAACAGATTCAAA	58.4	202	$tgt_{(5)}$
5	F: ACTCCAAAACCAAACCACCA R: CATAAAATCCGGATGCTGCT	58.4	199	$cac_{(5)}$
6	F: GGCTGGCTTCCCTAATTTTT R: CCTGCCCTGTCCATTCTTTA	58.4	236	$tctt_{(5)}$
7	F: GACGGCAGCTCCCTTCTT R: TGTGTCGATTGGTCCTCTTG	58.4	238	$cct_{(5)}$
8	F: CGCCCTCTGCTAAGTGCTAT R: TGAAAGAAAACCTTATGTGTCCA	58.4	180	$cag_{(6)}$
9	F: CTTCATCACCAGGCAGCTCT R: GCCCCCTTCTGCTCTTCTTA	58.4	200	$cag_{(6)}$
10	F: TAGAAGATGGCTTCCGACGA R: TTCCTCTCCTCCTCCTCCTC	58.4	233	$atg_{(5)}$
11	F: GATGGAGATGGAGGAAGTGG R: TCCCCTCCTTTTTCCTGTTT	58.4	206	$tgt_{(5)}$
12	F: TTCGGTTTGATTGCCGTTAT R: ATCTGTCCTCCCCGGTAACT	58.4	205	$ctc_{(5)}$
13	F: ACCTGTTTGCATGGAACCTT R: TTTCAACCGCCAAAGTCTTC	58.4	202	$a_{(14)}$
14	F: TGGCTGGTAATGCTAACTTGA R: CGGCAAGTATGGAAGGTGTT	58.4	189	$ag_{(10)}$
15	F: GGTTCCAAAGCACAGACCAT R: CTCTTAGTCTTTACCTCGACTACCA	58.4	222	$t_{(12)}$
16	F: TGCAGCTTCATCTGCTCGTA R: TATAAGACGGGCAACCCAAA	58.4	197	$tct_{(5)}$
17	F: GGGTCCAAAATCGAATATCCA R: TGAACAGATCCAGCATGTGA	58.4	169	$t_{(16)}$
18	F: CTGACTTACAGCCTCCGCTTA R: GGGCTTCACTGCTTTGAGAC	58.4	178	$a_{(15)}$
19	F: GGAGAGGTAACAAAGACCTCCA R: GTCACCATGGGTTCCATGTC	58.4	212	$cac_{(5)}$
20	F: CCGTAGCAAACCTCCGTTAT R: ACTTGATCGTCGCCATACG	58.4	164	$ct_{(9)}$
21	F: GATCCAACCACCGAATCAAC R: GAGGGAAATGGGGGAGAAT	58.4	183	$tct_{(5)}$

The results showed that oil palm from the Coconut Institution had lower diversity compared to oil palm germplasm from Malaysia. The oil palm germplasm from Malaysia was collected over widespread areas, resulting in more heterogeneous collections compared to oil palm from the Coconut Institution, which were mostly from scattered isolated populations across the Hainan Province, resulting in a relatively homozygous genome for the oil palm collections. Furthermore, these data indicated that SSR markers could be used to identify polymorphic loci for assessment of the genetic variation among oil palm.

Table 2. Details on observed heterozygosity (H_O); observed number of alleles (N_A), and Shannon's information index (I) of genetic variation within populations of oil palm.

No.	Samples	Population	H_O	N_A	I
1	31	Coconut Institution	0.3683	2.4688	0.611
2	63	Malaysia	0.4035	2.5	0.702
Mean			0.3859	2.4844	0.6565

Population structure

Using Popgene analysis, the average genetic distance of oil palm from different geographical origin was assessed (Yin et al., 2015). As shown in Table 3, the average genetic distance of oil palm from the Coconut Institution and Malaysia was 0.32 and 0.35, respectively. The distance between the Coconut Institution and Malaysia was 0.4. The results showed that the distance was almost the same between different geographical areas and single population.

Table 3. Average genetic distance of oil palm from different geographical origins.

Population	Coconut institution	Malaysia
Coconut Institution	0.32	
Malaysia	0.4	0.35

As shown in Figures 1 and 2, the dendrogram based on the genetic distance by the UPGMA cluster and the PCA analysis showed that the 2 oil palm populations were grouped into one cluster.

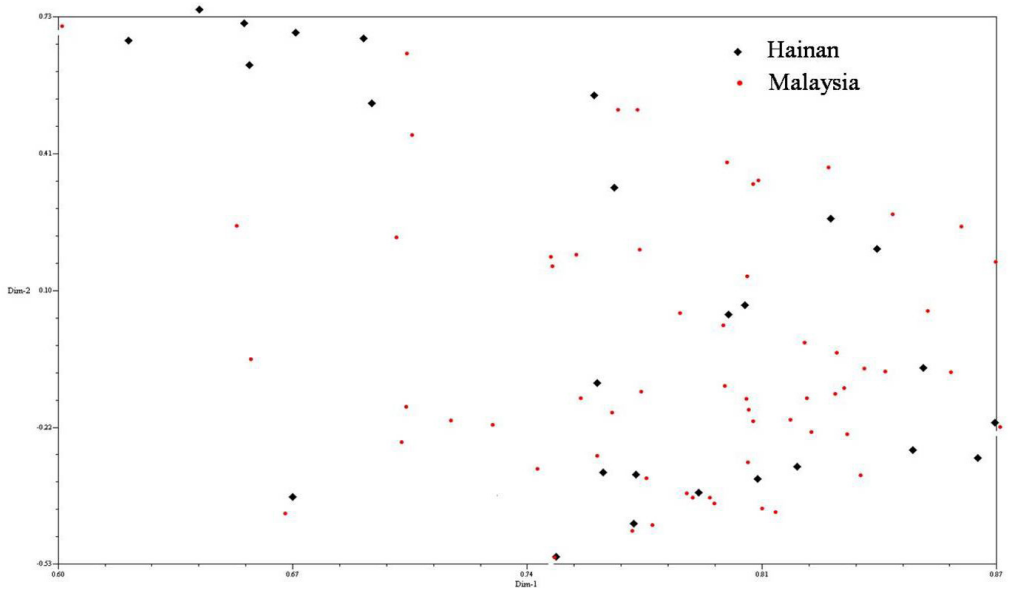

Figure 1. Principal component analysis of the 2 oil palm populations.

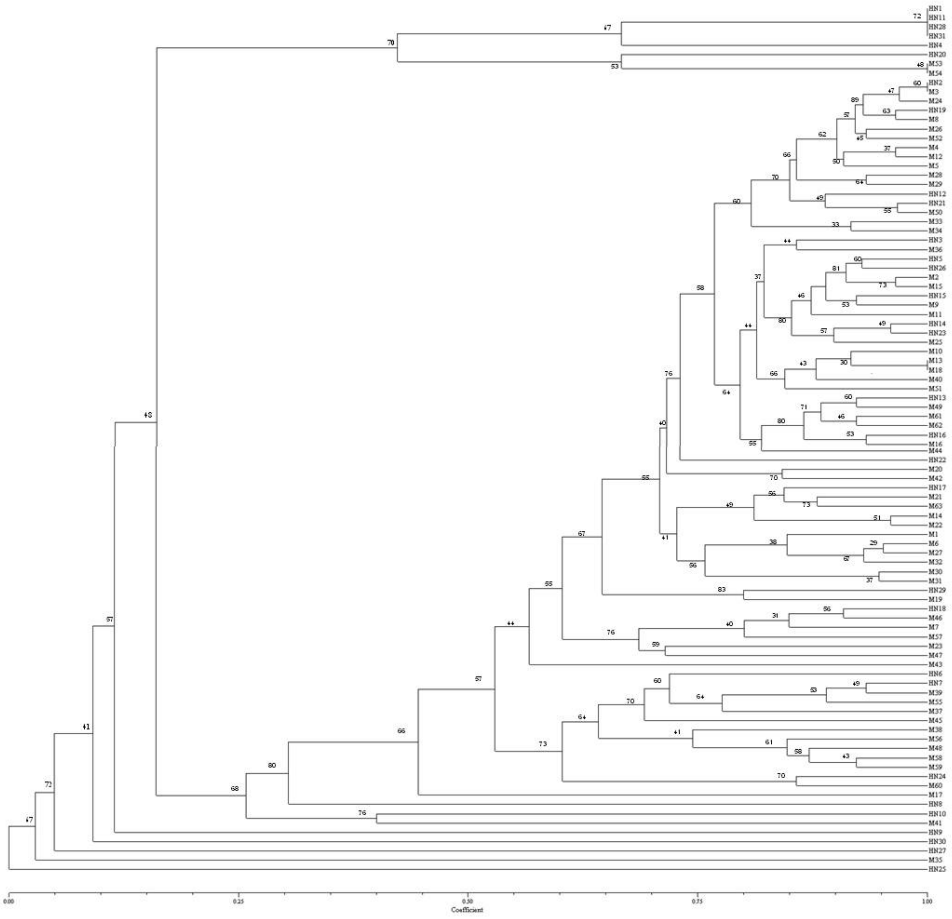

Figure 2. Phylogenetic dendrogram of the 2 oil palm populations based on the genetic distance by the UPGMA cluster.

In general, population clusters supported the origins and geographical distributions of the palms. Oil palm from the Coconut Institution and Malaysia showed a very close relationship. This was not surprising as Hainan Province of China and Malaysia are neighborhoods. Originally, there was no oil palm in China. Until 1926, oil palm seeds from Malaysia and Indonesia were brought in to China, and planted in several areas of Hainan Province. Subsequently, oil palm planting from Malaysia and Indonesia was introduced into Yunnan, Guangdong, and Guangxi Provinces in around 1941. After 1960, oil palm was widely introduced and planted in Hainan Province. Therefore, oil palm from the Coconut Institution could also be the germplasm from Malaysia, and the 2 populations had a similar genetic background.

DISCUSSION

Population genetic structure of a species can provide critical information for developing conservation and management strategies. In this study, 21 SSR primers were used to examine

the genetic diversity of 94 oil palm lines. The marker attribute, H_O, has been employed in many population variation studies at the genetic level to quantify the discriminatory power of primer combinations (Wang et al., 2014a). In this study, the average H_O was 0.3859 indicating that the genetic diversity in the species was relatively high. In addition, the average N_A (2.4844) also proved the same result. The populations with a relatively high average N_A suggest that it has good genetic variation and could be amenable to protection of germplasm resources. Genetic diversity is closely related to adaptive power, viability, and evolutionary potential. Finally, it also showed that this species has the genetic potential for breeding.

Shannon's information index can indicate the level of genetic diversity; a higher value indicates greater genetic diversity (Wang et al., 2014b). For the 94 oil palm lines, the average *I* was 0.6565, indicating various levels of genetic diversity.

On the basis of PCA and UPGMA cluster analysis, the 94 oil palm lines were grouped into one cluster. Originally, China had no oil palm, in the early 20th century, the oil palm seeds from Malaysia and Indonesia were planted in Hainan Province of China. In the 1960's, oil palm underwent a large-scale planting in Hainan, and the planting area was 18,700 hm^2 (Xia et al., 2014). At present, oil palm is distributed mainly in the tropical area of China. Therefore, Chinese oil palm development history, and the experiment findings supported the conclusion that the populations of Hainan and Malaysia may be from the same source, and they have a similar genetic background. Further research is needed in this area. Genetic diversity and population structure data based on SSR are a theoretical basis for deeper study and research into oil palm breeding. In the process of breeding selection, other reference indices, such as flower bud differentiation, fruit productivity, disease control, and cold stress should be used to determine breeding materials in the same group.

At present, the distribution of oil palm is scattered in Hainan Province. Genetic diversity and molecular systematic data can contribute to the development of effective conservation strategies. The genetic data obtained here for the oil palm based on microsatellite markers demonstrated indirectly the adaptive genetic diversity. These data provide genetic information for 94 oil palm lines, so that a great amount of genetic variation in the oil palm can be preserved. Further research should focus on characterization of genetic diversity, construction of genetic map, and the improvement of breeding for oil palm.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the Natural Science Foundation of Hainan Province (#20153070).

REFERENCES

Abram NK, Xofis P, Tzanopoulos J, MacMillan DC, et al. (2014). Synergies for improving oil palm production and forest conservation in floodplain landscapes. *PloS One* 9: e95388.

Cameron M, Ray R and Sabesan S (2015). Remote supervision of medical training via video conference in northern Australia: a qualitative study of the perspectives of supervisors and trainees. *BMJ Open* 5: e006444.

Corley RHV and Tinker PB (2003). The oil palm. 4th edn. Blackwell Science, Oxford.

Feng SP, Li WG, Huang HS, Wang JY, et al. (2009). Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). *Mol. Breed.* 23: 85-97.

Ferwerda JD (1977). Oil palm. In: Ecophysiology of tropical crop (Alvim PT and Kozlowski TT, eds.). Academic Press, London, 351-383.

Lei X, Xiao Y, Xia W, Mason AS, et al. (2014). RNA-seq analysis of oil palm under cold stress reveals a different C-repeat binding factor (CBF) mediated gene expression pattern in *Elaeis guineensis* compared to other species. *PloS One* 9: e114482.

Low ET, Rosli R, Jayanthi N, Mohd-Amin AH, et al. (2014). Analyses of hypomethylated oil palm gene space. *PloS One* 9: e86728.

Montoya C, Cochard B, Flori A, Cros D, et al. (2014). Genetic architecture of palm oil fatty acid composition in cultivated oil palm (*Elaeis guineensis* Jacq.) compared to its wild relative *E. oleifera* (H.B.K) Cortés. *PloS One* 9: 1-13.

Nei M and Takezaki N (1983). Estimation of genetic distances and phylogenetic trees from DNA analysis. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production. University of Guelph, Canada, 405-412.

Powell W, Morgante M, Andre C, Hanafey M, et al. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2: 225-238.

Qin Y, Sun DQ, Xu TJ, Liu XZ, et al. (2014). Genetic diversity and population genetic structure of the miiuy croaker, *Miichthys miiuy*, in the east China by microsatellite markers. *Genet. Mol. Res*. 13: 10600-10606.

Rozen S and Skaletsky H (2000). Primer 3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132: 365-386.

Shen J, Jia X, Ni H, Sun P, et al. (2010). AFLP analysis of genetic diversity of *Jatropha curcas* grown in Hainan, China. *Trees* 24: 455-462.

Singh R, Noorhariza MZ, Ting NC, Rozana R, et al. (2008). Exploiting an oil palm EST database for the development of gene-derived and their exploitation for assessment of genetic diversity. *Biologia* 63: 1-9.

Sneath PHA and Sokal RR (1973). Numerical Taxonomy: the principles and practice of numerical classification. W. H. Freeman, San Francisco.

Stewart CN and Via LE (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques* 14: 748-750.

Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol*. 24: 1596-1599.

Ting NC, Jansen J, Mayes S, Massawe F, et al. (2014). High density SNP and SSR-based genetic maps of two independent oil palm hybrids. *BMC Genomics* 15: 309-319.

Wang F, Zhang S, Liu MG, Lin XS, et al. (2014a). Genetic diversity analysis reveals that geographical environment plays a more important role than rice cultivar in Villosiclava virens population selection. *Appl. Environ. Microbiol*. 80: 2811-2820.

Wang F, Yang T, Burlyaeva M, Li L, et al. (2015). Genetic diversity of grasspea and its relative species revealed by SSR markers. *PLoS One* 10: e0118542.

Wang S, Liu Y, Ma L, Liu H, et al. (2014b). Isolation and characterization of microsatellite markers and analysis of genetic diversity in Chinese jujube (*Ziziphus jujuba* Mill.). *PLoS One* 9: e99842.

Xia W, Xiao Y, Yang Y D, Ma ZL, et al. (2014). Development and utilization of SSR markers of oil palm according to NCBI database. *Guangdong Agric. Sci*. 2: 144-148.

Xiao Y, Zhou L, Xia W, Mason AS, et al. (2014). Exploiting transcriptome data for the development and characterization of gene-based SSR markers related to cold tolerance in oil palm (*Elaeis guineensis*). *BMC Plant Biol.* 14: 384-396.

Yin M, Li H, McManus DP, Blair D, et al. (2015). Geographical genetic structure of *Schistosoma japonicum* revealed by analysis of mitochondrial DNA and microsatellite markers. *Parasit Vectors* 8: 150-158.

Yu J, Jing ZB and Cheng JM (2014). Genetic diversity and population structure of *Stipa bungeana*, an endemic species in Loess Plateau of China, revealed using combined ISSR and SRAP markers. *Genet. Mol. Res*. 13: 1097-1108.

Yunus R, Salleh SF, Abdullah N and Biak DR (2010). Effect of ultrasonic pre-treatment on low temperature acid hydrolysis of oil palm empty fruit bunch. *Bioresour. Technol*. 101: 9792-9796.

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

Zhang YS, Cao JH and Lin WF (2009). The industry development of oil palm industry in China. *China Trop. Agric*. 4: 15-18.