Journal of Oil Palm Research Vol. 30 (1) March 2018 p. 61 – 70
DOI: https://doi.org/10.21894/jopr.2017.0003

ASSOCIATION OF SNP MARKERS WITH HEIGHT INCREMENT IN MPOB-ANGOLAN NATURAL OIL PALM POPULATIONS

ONG, P W*; MAIZURA, I*; MARHALIL, M*; RAJANAIDU, N*; ABDULLAH, N A P; RAFII, M Y‡; OOI, L C L*; LOW, E T L* and SINGH, R***

ABSTRACT

*Low height increment is one of the desired traits in oil palm breeding and improvement programmes, as dwarf palms facilitate fruit harvesting and extend the economic life of the crop. In this study, 346 natural oil palms collected from Angola and maintained by the Malaysian Palm Oil Board (MPOB) were used. Analysis of variance (ANOVA) for height increment showed a significant difference (P ≤ 0.001) among families in populations, indicating substantial genetic variation for marker-trait association study. We applied nine carefully selected single nucleotide polymorphism (SNP) markers to genotype the oil palms via cleaved amplified polymorphic sequence (CAPS) method. Population structure analysis involving 18 SNP alleles divided the palms into two sub-populations, with no obvious relative kinship (values < 0.3). For association analysis between the SNP markers and height increment, three models were tested. The incorporation of population structure (Q) and relative kinship (K) as correction factors in the model had helped reduce false positive associations. Generally, the mixed linear model (MLM) with Q + K exhibited a more stringent model with less spurious associations detected. Based on this model, one significant marker SNPG00006_*Fat*I corresponding to indole-3-acetic acid (IAA)-amido synthetase gene was identified to be associated with height increment (P ≤ 0.05). The marker, although potentially specific to MPOB-Angolan germplasm, can assist in introgressing the dwarf phenotype into advanced breeding materials through marker-assisted selection (MAS).*

Keywords: height increment, marker-trait association, MPOB-Angolan natural oil palm, single nucleotide polymorphism.

Date received: 25 May 2017; **Sent for revision:** 26 May 2017; **Received in final form:** 31 July 2017; **Accepted:** 18 August 2017.

INTRODUCTION

Oil palm (*Elaeis guineensis* Jacq.) is the most important crop in Malaysia, being planted on over 80% (5.74 million hectares) of the total agricultural land use for commodity crops (6.91 million hectares) (MPOB, 2016; MPIC, 2016). Despite numerous efforts to improve the crop's productivity, challenges in harvesting ripe bunches from old palms still persist. Thus, one of the major thrusts in breeding is to reduce height increment rate of palms from 40 to 75 cm per year to below 30 cm per year (Kushairi *et al*., 2011; Rajanaidu *et al*., 2000). One of the ways to do this is to introgress the trait from germplasm samples into commercial materials.

Malaysian Palm Oil Board (MPOB) owns the largest oil palm germplasm collection in the world. The phenotypic evaluation revealed

* Malaysian Palm Oil Board,
6 Persiaran Institusi, Bandar Baru Bangi,
43000 Kajang, Selangor, Malaysia.
E-mail: maizura@mpob.gov.my

** Department of Crop Science, Faculty of Agriculture and Food Sciences, Universiti Putra Malaysia Kampus Bintulu, Jalan Nyabau, 97000 Bintulu, Sarawak, Malaysia.

‡ Laboratory of Climate-Smart Food Crop Production, Institute of Tropical Agriculture and Food Security, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

selected MPOB-Angolan palms with annual height increment of only 20 to 30 cm and showing high yields (Kushairi *et al*., 2003). However, introgressing this economically important trait into the present planting materials through conventional breeding will take an exceedingly long time. This is because oil palm is a perennial species. One breeding cycle of oil palm takes 10 years (Rajanaidu *et al*., 2000). Thus, the production of new and improved varieties in oil palm requires about 30 years, considering two to three generations of progeny testing. Recourse must, therefore, be made to biotechnology that offers tools for selecting palms carrying the traits of interest early (*e.g*.; in the nursery), without having to wait for the palms to mature. If an early selection is possible, the breeding cycle can be reduced by almost half of the time required in conventional programmes. This will subsequently speed up the development of new planting materials in oil palm.

Molecular markers such as microsatellites/ simple sequence repeats (SSR) and single nucleotide polymorphisms (SNP) (Cui *et al*., 2015; Banerjee *et al*., 2015; Shi *et al*., 2016) that show association with the traits of interest can be used to select individuals with the desirable traits. SNP are co-dominant molecular markers and can reveal single nucleotide changes in DNA sequences. They are the most abundant sequence variations in plant genomes, although studies have shown that their frequency varies among plant species, such as 1 SNP per 217 bp in rice (Lee *et al*., 2009), 1 SNP per 540 bp in wheat (Somers *et al*., 2003) and 1 SNP per 273 bp in soyabean (Zhu *et al*., 2003). The increasing number of expressed sequence tag (EST) and genomic sequence information available in public databases have opened up the opportunity to discover more SNP. Newly automated and high-throughput systems for SNP detection have been established which allow wide application of the marker in genetic studies, such as genetic diversity, population structure determination, linkage disequilibrium analysis and association mapping (Hamblin *et al*., 2010; Murray *et al*., 2009). Recently, a high-density oil palm SNP array was employed to identify markers linked to important agronomic traits (Kwong *et al*., 2016). However, the cost of genotyping with such arrays, makes it unattractive for routine application.

A candidate gene approach to study association between an allelic variant and phenotypic traits has been widely applied in plants. Candidate genes are identified based on the understanding of the biochemical pathway, mutation analysis and linkage analysis. The first candidate gene association between the allelic variants of the *Dwarf8* gene and the developmental traits (plant height and flowering time) was demonstrated in maize (Thornsberry *et al*., 2001). The results indicated significant association between the *Dwarf8* polymorphism with flowering time. Later, similar approaches were applied in wheat (Ravel *et al*., 2006) and rice (Lu *et al*., 2010). The candidate genes regulating plant height and stature have been described (Wang and Li, 2008). Among those, is the GRETCHEN HAGEN 3 (*GH3*), a gene family involved in hormonal regulation which has been implicated in dwarfism in both rice and *Arabidopsis* (Wang *et al*., 2008; Westfall *et al*., 2010).

In MPOB, efforts were initiated to mine for SNP in the oil palm genomic sequence database (Low *et al*., 2014). Mining SNP from publicly available EST sequence databases has also been reported for oil palm (Riju *et al*., 2007). One of the common and cost-effective techniques to assay SNP is the agarose gel-based method, known as cleaved amplified polymorphic sequence (CAPS) (Konieczny and Ausubel, 1993). The method amplifies a target DNA region containing a SNP, followed by digestion of the amplified product and fragment analysis on agarose gel. A simple and reliable SNP-based CAPS assay for oil palm has been developed. The ability of the SNP-based CAPS assay in showing polymorphism in natural oil palm germplasm as well as in differentiating the two oil palm species (*E. guineensis* and *E. oleifera*) has been reported (Singh *et al*., 2011).

A previous study had evaluated the genetic variability of selected MPOB-Angolan natural oil palm populations using random and candidate gene SNP markers (Ong *et al*., 2015). The objective of the present study is to use the informative SNP markers for marker-trait association. The study in particular was aimed at identifying SNP marker(s) significantly associated with height increment in the selected populations segregating for the trait. The marker(s) will be useful for use in breeding programmes involving MPOB-Angolan germplasm to reduce height increment, which is desirable to extend the economic life-span of the oil palm.

MATERIALS AND METHODS

Plant Materials and Phenotypic Data

The collection of the MPOB-Angolan natural genetic materials and their characteristics have been described previously (Rajanaidu *et al*., 1991; Kushairi *et al*., 2003). Four MPOB-Angolan populations namely AGO01, AGO02, AGO07 and AGO08 were used in this study. Of these, two of the populations (AGO01 and AGO08) were previously used in a genetic diversity analysis (Ong *et al*., 2015). These materials were planted in a randomised complete block design with two replications. A total of 346 palms were sampled from an experimental plot established at MPOB Kluang Research Station, Johor, Malaysia. The number of palms evaluated per population/family is shown in *Table 1*.

TABLE 1. NUMBER OF PALMS EVALUATED IN EACH POPULATION AND FAMILY

Population	Family (number of palms in parentheses)	Total
AGO01	AGO01-04 (31) and AGO01-05 (30)	61
AGO02	AGO02-03 (32)	32
AGO07	AGO07-01 (32), AGO07-02 (32) and AGO07-03 (31)	95
AGO08	AGO08-01 (32), AGO08-08 (30), AGO08-10 (32), AGO08-11 (32) and AGO08-12 (32)	158

Palm height was measured from the ground to the base of frond 41, at eight years after planting. Height increment was calculated based on Breure and Powell (1987). The variability of height increment such as mean, minimum, maximum and coefficient of variation (CV) was computed. Height increment was subjected to analysis of variance (ANOVA) using the following model, as described by Lawrence and Rajanaidu (1985): $Y_{ijk} = \mu + \alpha_i + \beta(\alpha)_{j(i)} + \varepsilon_{ijk}$ where Y_{ijk} is the observed value of height increment of the j^{th} family in the i^{th} population in the k^{th} replicate, μ is an overall mean, α_i is the fixed effect of the i^{th} populations, $\beta(\alpha)_{j(i)}$ is the random effect of the jth families nested within ith populations and ε_{ijk} is the random residual. All analyses were conducted using statistical analysis software (SAS) version 9.1 (SAS Institute Inc., Cary, North Caroline, USA).

DNA Extraction and SNP Genotyping

Spear leaves harvested from each palm were frozen in liquid nitrogen prior to storage in a -80°C freezer. Genomic DNA was extracted from frozen spear leaves using the modified cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1990). A set of nine informative SNP markers were used for genotyping using the CAPS assay, as described previously (Ong *et al*., 2015). These include three random SNP (653_*Aci*I, 3064_*Taq*I and 5963_*Alu*I) and six candidate gene SNPs (SNPG00002_*Hpy*188I, SNPG00004_*Aci*I, SNPG00005_*Bcg*I, SNPG00006_*Fat*I, SNPG00014_*Hpy*CH4III and SNPG00014_*Ssp*I).

Population Structure and Relative Kinship

STRUCTURE software version 2.3.3 (Pritchard *et al*., 2000) was used to infer population structure based on the model-based clustering method. Ten independent runs were performed by setting the number of populations (*k*) from 1 to 10 with 100 000 burn in and 100 000 Markov Chain Monte Carlo (MCMC) iterations. The admixture co-ancestry model was applied with correlated allele frequencies. The optimal *k* was chosen by an *ad hoc* statistic (Δk) based on the second order rate of change of the likelihood function with respect to *k* (Evanno *et al*., 2005). This analysis was performed using STRUCTURE HARVESTER (Earl and von Holdt, 2012). After the determination of optimum population structure (*k*), the inferred ancestry of individual matrix (*Q*) was estimated and then used for association analysis. The relative kinship matrix (*K*) was generated using Spatial Pattern Analysis of Genetic Diversity (SPAGeDi) software version 1.4 (Hardy and Vekemans, 2002). All negative kinship values between the individuals were assigned to zero.

Marker-trait Association and Model Testing

Marker-trait association analysis was performed according to the general linear model (GLM) and mixed linear model (MLM) using Trait Analysis by Association, Evolution and Linkage (TASSEL) software version 3.0.59 (Bradbury *et al*., 2007). Three different models were tested in which (1) GLM without *Q* model was the simplest model to explain the variation in height increment, as this model only considers the marker effect, (2) the GLM with *Q* model included *Q* matrix as a control for population structure and (3) the MLM with *Q* + *K* model took into account the both *Q* matrix and *K* matrix, which controls population structure as well as relative kinship. For all models, the marker effect and *Q* matrix were defined as fixed effects. The *K* matrix was considered a random effect. A quantile-quantile (QQ) plot was plotted to assess the efficiency for all the three different models in TASSEL. Markers were considered as significantly associated with height increment when $P \leq 0.05$. Only SNP markers with a minor allele frequency (MAF) higher than 0.05 were included in the association analysis. To determine physical location of the height increment associated SNP markers in oil palm genome, the flanking sequence of the SNP marker was run in a BLAST search against the oil palm reference genome sequence, EG5 (Singh *et al*., 2013). Subsequently, physical locations of previously identified quantitative trait locus (QTL) for stem/trunk height (Billotte *et al*., 2010; Lee *et al*., 2015; Pootakham *et al*., 2015) were also localised.

RESULTS

Phenotypic Data

The four selected MPOB-Angolan natural oil palm populations (AGO01, AGO02, AGO07 and AGO08) showed a mean height increment of 44.0 cm per year, with maximum and minimum height increment of 80.8 and 20.3 cm per year, respectively (*Table 2*). The CV for height increment was 26.9%, indicating that the MPOB-Angolan germplasm has

TABLE 2. DESCRIPTIVE PARAMETERS OF HEIGHT INCREMENT IN EACH POPULATION AND FAMILY

			Height increment (cm)		
	N	Mean	Minimum	Maximum	CV (%)
Overall	346	44.0	20.3	80.8	26.9
AGO01	61	49.4	23.7	69.3	19.7
AGO02	32	53.5	37.2	72.3	18.5
AGO07	95	44.2	25.0	80.8	24.2
AGO08	158	39.9	20.3	72.8	29.2
AGO01-04	31	46.8	34.7	66.5	18.0
AGO01-05	30	52.1	23.7	69.3	20.0
AGO02-03	32	53.5	37.2	72.3	18.5
AGO07-01	32	48.2	28.5	80.8	25.1
AGO07-02	32	39.4	25.0	63.7	22.5
AGO07-03	31	45.2	34.3	66.5	20.5
AGO08-01	32	30.9	21.5	50.7	21.6
AGO08-08	30	35.1	20.3	48.8	22.0
AGO08-10	32	39.2	20.7	56.7	22.5
AGO08-11	32	48.3	29.8	72.8	19.6
AGO08-12	32	45.8	21.0	71.2	31.5

Note: N - number of palms. CV - coefficient of variation.

reasonable genetic variation for height increment. The descriptive parameters for each population and family were also determined (*Table 2*). The population mean for height increment ranged from 39.9 to 53.5 cm per year while at the family level ranged from 30.9 to 53.5 cm per year. ANOVA revealed that there were significant differences among the family nested within populations for height increment ($P \leq 0.001$). However, no significant differences were detected among populations (*Table 3*).

Population Structure and Relative Kinship

A Bayesian model-based clustering method was used to assign 346 oil palms to two groups (k = 2) based on the Δk method of Evanno *et al*. (2005) (*Figure 1*). Of the 346 oil palms, 219 palms were grouped into group I and the remaining 127 palms were included in group II (*Figure 2*). All palms were categorised as admixed forms and no correlation was found between the clustering pattern and geographical distribution of the palms. Relative kinship estimates showed that about 65% of the pair-wise kinship estimates were between 0.00 and 0.05. Approximately 29% of estimates had a range of 0.00 to 0.20 and the remaining 6% showed various degrees of genetic relatedness (*Figure 3*). This result indicates that there was weak relationship among the palms analysed in the study.

TABLE 3. ANALYSIS OF VARIANCE OF HEIGHT INCREMENT

Source of variation	DF	MS	F value
Population	3	2 463.2	2.07
Family (population)	7	1 190	12.15*
Residual	335	97.9	-

Note: DF - degree of freedom. MS - mean square.
*Significant at $P \leq 0.001$.

Marker-trait Association and Model Testing

Based on the GLM without Q model, three SNP markers (SNPG0004_*Aci*I, SNPG00006_*Fat*I and SNPG00014_*Hpy*CH4III) were associated with height increment ($P \leq 0.05$). Among them, two markers namely SNPG0004_*Aci*I and SNPG00006_*Fat*I were identified to be associated with height increment by GLM with Q model at $P \leq 0.05$. Further association analysis using another model (MLM with $Q + K$) revealed that only one SNP marker (SNPG00006_*Fat*I) was significantly associated with height increment at $P \leq 0.05$. The number of significant markers identified under GLM without Q, GLM with Q and MLM with $Q + K$ models were 3, 2 and 1, respectively. The inclusion of population structure and kinship in the association models had reduced the number of significant markers associated with height increment. *Table 4* summarises the significant markers identified in association analysis using three models in TASSEL software.

The QQ plots for all the models showed a significant degree of deviation from the expected -Log$P$ values (*Figure 4*). MLM with $Q + K$ model had a slightly greater improvement in term of the degree of deviation from the expected -Log$P$ values compared to the GLM models (GLM without Q and GLM with Q). MLM model was more effective

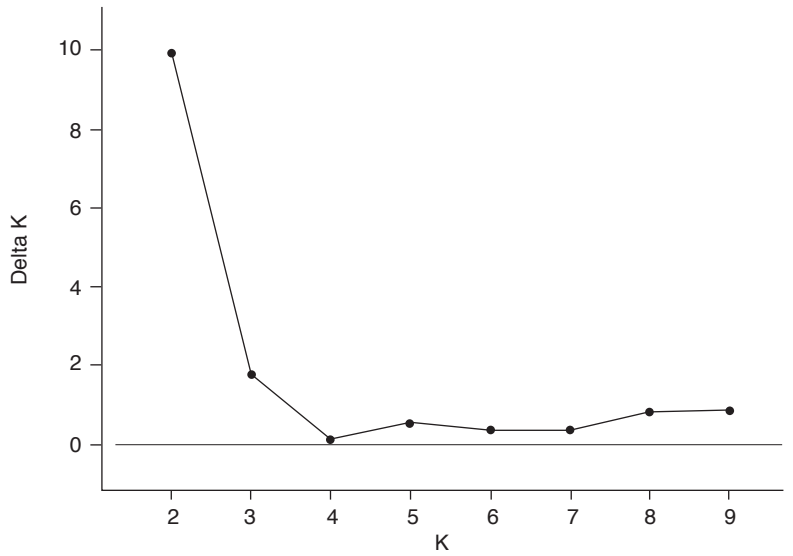

Figure 1. The Δk based on the second order rate of change of the likelihood function with respect to k.

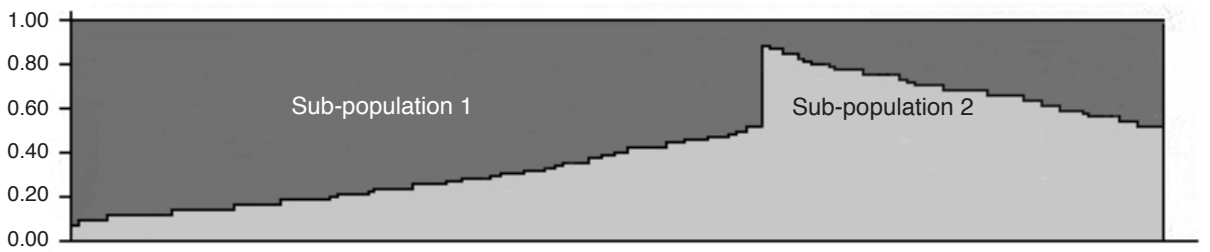

Figure 2. The STRUCTURE generated profiles of 364 MPOB-Angolan natural oil palms used to build the Q matrix.

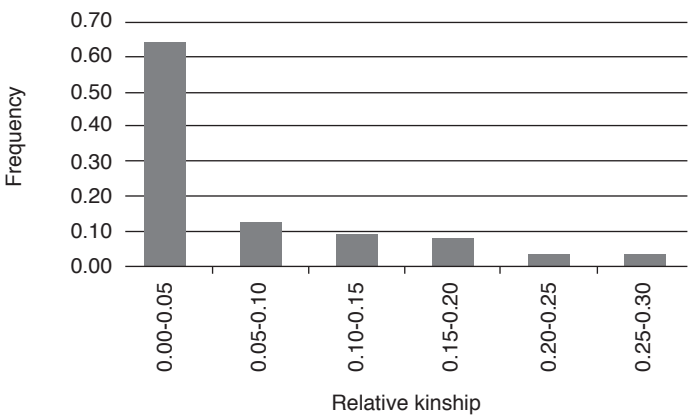

Figure 3. Distribution of the pairwise relative kinship estimates between 346 MPOB-Angolan natural oil palms.

TABLE 4. MARKER-TRAIT ASSOCIATIONS USING TASSEL SOFTWARE (GLM and MLM)

Marker	GLM without *Q* (*P* value)	GLM with *Q* (*P* value)	MLM with *Q* + *K* (*P* value)
653_*Aci*I	0.2825	0.6260	0.7122
3064_*Taq*I	0.3153	0.3343	0.1711
5962_*Alu*I	0.5792	0.7700	0.3989
SNPG00002_*Hpy*188I	0.1880	0.4717	0.4021
SNPG00004_*Aci*I	6.18E-03*	1.85E-05*	0.0510
SNPG00005_*Bcg*I	0.9096	0.7984	0.8354
SNPG00006_*Fat*I	6.32E-04*	3.05E-03*	7.60E-04*
SNPG000014_*Hpy*CH4III	1.22E-02*	0.0926	0.8050
SNPG00014_*Ssp*I	0.9623	0.5619	0.7802

Note: *Significant at $P \leq 0.05$. GLM – general linear model. MLM – mixed linear model.

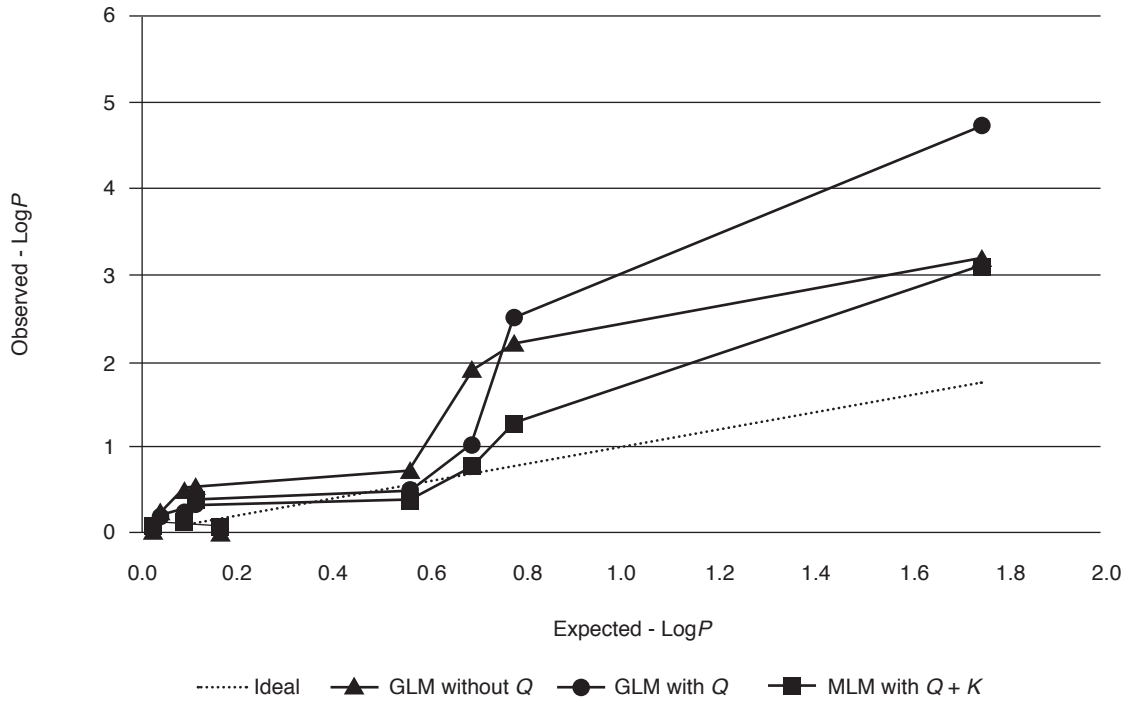

Figure 4. Quantile-quantile (QQ) plot of evaluated –LogP from association analysis of height increment using different models: (1) general linear model (GLM) without Q, (2) GLM with Q and (3) mixed linear model (MLM) with Q + K.

in screening out false positive markers compared to both GLM models. Therefore, MLM with *Q* + *K* models was the preferred choice for marker-trait association in this study. Applying this model, one SNP marker (SNPG00006_*Fat*I) was found to be significantly associated with height increment. This marker was within the indole-3-acetic acid (IAA)-amido synthetase gene. Interestingly, SNPG00006_*Fat*I was located on chromosome 1 about 32.81 Mb away from a QTL locus identified for height by Billotte *et al*. (2010) (*Tables 5* and *6*).

TABLE 6. PHYSICAL POSITION OF MARKERS ASSOCIATED WITH HEIGHT INCREMENT ON CHROMOSOME 1

Marker ID	Marker	Physical position (bp)
mEgCIR0246	Microsatellite	46 053 447
mEgCIR0886	Microsatellite	45 967 291
mEgCIR0439	Microsatellite	45 382 375
mEgCIR3711	Microsatellite	43 465 921
mEgCIR3808	Microsatellite	41 353 146
mEgCIR0778	Microsatellite	41 169 291
SNPG00006 *Fat*I	SNP	8 355 453

Note: SNP – single nucleotide polymorphism.

TABLE 5. MARKERS ASSOCIATED WITH STEM/TRUNK HEIGHT IDENTIFIED IN THE CURRENT STUDY AND THAT REPORTED PREVIOUSLY

Type of marker	Linkage group	Chromosome number	Reference
Microsatellite	4	2[a]	Billotte *et al*. (2010)
	8	1[a]	
	11	4[a]	
	15	10[a]	
SNP	10	10[b]	Pootakham *et al*. (2015)
	14	14[b]	
	15	15[b]	
Microsatellite	5	16[c]	Lee *et al*. (2015)
SNP	-	1[c]	-

Note: [a] Corresponding chromosome number assigned in Singh *et al*. (2013).
[b] Chromosome number assigned by Pootakham *et al*. (2015).
[c] Chromosome number assigned based on BLAST against EG5 genome build (Singh *et al*. 2013).
SNP – single nucleotide polymorphism.

DISCUSSION

The QTL associated with height in oil palm were reported previously (Billotte *et al*., 2010; Lee *et al*., 2015; Pootakham *et al*., 2015) using bi-parental populations segregating for the trait. In a long-lived outbreeding crop, such as oil palm, creating such populations takes a very long time and is laborious. Here, we have taken the initiative to utilise MPOB's germplasm collection and applied association mapping approach to identify molecular markers linked to height in oil palm. We have found wide variability in terms of height increment, among the selected MPOB-Angolan populations with CV of 26.9%. This suggested that there is sufficient variation for height in the populations to conduct marker-trait association analysis. The populations also possessed high CV (25.2%) in terms of stalk length (Noh *et al*., 2008) indicating another potential trait for conducting a similar analysis.

In the marker-trait association analysis, the results from GLM without *Q*, GLM with *Q* and MLM with $Q + K$ models were compared. Generally, the GLM without *Q* was the poorest, giving a skewed curve relatively far from ideal. GLM with *Q* showed slight improvement in terms of the number of significant markers. MLM with $Q + K$ models showed further improvement and the number of significant markers decreased relative to both GLM without *Q* and GLM with *Q* models. These results indicate that the marker-trait association attained in model GLM without *Q* was due to population structure. The inclusion of population structure and relative kinship reduce type I error, thus eliminating false positive associations (Yu *et al*., 2010). According to Li *et al*. (2011), correction factors concerning population structure and relative kinship are very crucial even though both had relatively small effect on marker-trait association. Our findings are consistent with results from other studies on pak-choi (Yu *et al*., 2010) and *Brassica rapa* (Pino Del Carpio *et al*., 2011).

More importantly in this study, marker SNPG00006_*Fat*I consistently showed significant association with height increment in all models tested. Li *et al*. (2011) reported that the level of association can be classified into four groups: strong ($P \leq 0.005$), moderate ($0.005 < P \leq 0.01$), weak ($0.01 < P \leq 0.05$) and none ($P > 0.05$). At a significance level of 0.005, SNP marker SNPG00006_*Fat*I constantly displayed strong association with height increment in all three models tested, thus increasing the confidence of its association with height increment. Hasjesh *et al*. (2008) reported candidate gene association mapping in diverse maize inbred lines and verified their findings through linkage mapping, gene expression and mutagenesis studies. We note that the SNPG00006_*Fat*I marker correspond to IAA-amido synthetase, which belongs to the family of *GH3* proteins that regulate auxin levels in plants growth and development. In rice, overexpression of *GH3-8* (Ding *et al*., 2008) and *GH3-13* (Zhang *et al*., 2009) resulted in dwarf phenotypes. The expression of *GH3* in rice is triggered by pathogenic or abiotic stress, which simultaneously causes dwarfness of the plant. This also explains the dwarfness phenotype among plants exposed to environmental stress. Similar studies can be initiated to evaluate the effect of both genes on oil palm. Thus, further work should be pursued perhaps through expression analysis of the associated candidate genes namely, IAA-amido synthetase and *GH3* across tall and short palms to confirm their role in determining height in the oil palm species.

In our analysis, we did not find a consistent association between height increment and random SNP markers (653_*Aci*I, 3064_*Taq*I and 5962_*Alu*I). Nevertheless, the candidate SNP marker developed here showed significant association with height increment. Candidate gene markers seemed to be more promising in marker-trait association in this study. A significant association between candidate gene markers and traits of interest were demonstrated in potato (Li *et al*., 2013), forest tree (Muller *et al*., 2015) and field pea (Jha *et al*., 2015). Thus, additional markers can be designed based on height genes published previously in other plants (Kujur *et al*., 2016). Applying the candidate gene approach would increase the chances of finding markers associated with height increment for development of MAS in oil palm.

To our knowledge, this is the first attempt of applying candidate gene approach to identify SNP markers associated with height increment on natural oil palm populations. Previously, there were several studies on conventional mapping of QTL associated with height in oil palm (Billotte *et al*., 2010; Lee *et al*., 2015; Pootakham *et al*., 2015). We found that marker SNPG00006_*Fat*I was on chromosome 1 but about 32.81 Mb away from the height QTL identified by Billotte *et al*. (2010). Additionally, Billotte *et al*. (2010) also reported another three QTL for the same trait on chromosome 2, 4 and 10. Similarly, Pootakham *et al*. (2015) identified SNP markers from three QTL regions for height located on chromosomes 10, 14 and 15, while Lee *et al*. (2015) reported SSR markers for height on chromosome 16. The QTL identified in this study could be unique to MPOB-Angolan germplasm, which is potentially useful for introgressing low height increment trait from the MPOB-Angolan materials to the advanced breeding populations.

CONCLUSION

This work was initiated to evaluate the usefulness of random and candidate gene SNP markers for

marker-trait association study in oil palm. The results revealed a significant association between candidate gene SNP marker (SNPG00006_*Fat*I) and height increment trait in oil palm populations from Angola, despite the small number of markers used. Admittedly, it is not convincing to rely only on one marker for MAS. Thus, the current study should be expanded to include a larger set of candidate gene SNP markers as this will help increase the chance of identifying additional significant markers for the development of MAS for height breeding in oil palm. Additionally, gene expression studies can be undertaken on extremely tall and short palms from MPOB-Angolan populations, as well as from other MPOB's germplasm populations such as Nigerian (*E. guineensis*) and Colombian (*E. oleifera*) which also shows great variability in height (Rajanaidu *et al*., 1999; Mohd Din *et al*., 2000). This approach may help capture the genes influencing height in oil palm.

ACKNOWLEDGEMENT

The authors would like to thank the Director-General of MPOB for permission to publish this article. Special thanks to Andy Chang Kwang Choong for his valuable comments on the manuscript. This research was partly supported by grant from the Ministry of Science, Technology and Innovation (MOSTI) of Malaysia (eScience Project #02-03-11SF0010).

REFERENCES

BANERJEE, N; SIRAREE, A; YADAV, S; KUMAR, S; SINGH, J; KUMAR, S; PANDEY, D K and SINGH, R K (2015). Marker-trait association study for sucrose and yield contributing traits in sugarcane (*Saccharum* spp. hybrid). *Euphytica, 205*: 185-201.

BILLOTTE, N; JOURJON, M F; MARSEILLAC, N; BERGER, A; FLORI, A; ASMADY, H; ADON, B; SINGH, R; NOUY, B; POTIER, F; CHEAH, S C; ROHDE, W; RITTER, E; COURTOIS, B; CHARRIER, A and MANGIN, B (2010). QTL detection by multi-parent linkage mapping in oil palm (*Elaeis guineensis* Jacq.). *Theoretical and Applied Genetics, 120*: 1673-1687.

BRADBURY, P J; ZHANG, Z; KROON, D E; CASSTEVENS, T M; RAMDOSS, Y and BUCKLER, E S (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*: 2633-2635.

BREURE, C J and POWELL, M S (1987). The one-shot method of establishing growth parameters in oil palm. *Proc. of the International Oil Palm/Palm Oil Conference*. PORIM, Bangi. p. 203-209.

CUI, D; XU, C; YANG, C; ZHANG, Q; ZHANG, J; MA, X; QIAO, Y; CAO, G; ZHANG, S and HAN, L (2015). Association mapping of salinity and alkalinity tolerance in improved japonica rice (*Oryza sativa* L. subsp. japonica Kato) germplasm. *Genetic Resources and Crop Evolution, 62*: 539-550.

DING, X; CAO, Y; HUANG, L; ZHAO, J; XU, C; LI, X and WANG, S (2008). Activation of the indole-3-acetic acid-amido synthetase *GH3-8* suppresses expansin expression and promotes salicylate- and jasmonate-independent basal immunity in rice. *The Plant Cell, 20*: 228-240.

DOYLE, J J and DOYLE, J L (1990). Isolation of plant DNA from fresh tissue. *Focus, 12*: 13-15.

EARL, D A and VONHOLDT, B M (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources, 4*: 359-361.

EVANNO, G; REGNAUT, S and GOUDE, J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology, 14*: 2611-2620.

HAMBLIN, M T; CLOSE, T J; BHAT, P R; CHAO, S; KLING, J G; ABRAHAM, K J; BLAKE, T; BROOKS, W S; COOPER, B; GRIFFEY, C A; HAYES, P M; HOLE, D J; HORSLEY, R D; OBERT, D E; SMITH, K P; ULLRICH, S E; MUEHLBAUER, G J and JANNINK, J L (2010). Population structure and linkage disequilibrium in US barley germplasm: Implication for association mapping. *Crop Science, 50*: 556-566.

HARDY, O J and VEKEMANS, X (2002). SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Resources, 2*: 618-620.

HARJES, C E; ROCHEFORD, T R; BAI, L; BRUTNELL, T P; KANDIANIS, C B; SOWINSKI, S G; STAPLETON, A E; VALLABHANENI, R; WILLIAMS, M; WURTZEL, E T; YAN, J and BUCKLER, E S (2008). Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. *Science, 319*: 330-333.

JHA, A B; TAR'AN, B; DIAPARI, M and WARKENTIN, T D (2015). SNP variation within genes associated with amylose, total starch and crude protein concentration in field pea. *Euphytica, 206*: 459-471.

KONIECZNY, A and AUSUBEL, F M (1993). A procedure for mapping *Arabidopsis* mutations using

co-dominant ecotype-specific PCR-based markers. *The Plant J., 4*: 403-410.

KUJUR, A; UPADHYAYA, H D; BAJAJ, D; GOWDA, C L L; SHARMA, S; TYAGI, A K and PARIDA, S K (2016). Identification of candidate genes and natural allelic variants for QTLs governing plant height in chickpea. *Scientific Reports, 6*: 27968.

KUSHAIRI, A; RAJANAIDU, N; MOHD DIN, A; ISA, Z A; NOH, A and JUNAIDAH, J (2003). Performance of Angola genetic materials. *Proc. of the Seminar on Progress of PS1 and PS2 Planting Materials and Release of Elite Germplasm to the Industry.* MPOB, Bangi. p. 75-90.

KUSHAIRI, A; MOHD DIN, A and RAJANAIDU, N (2011). Oil palm breeding and seed production. *Further Advances in Oil Palm Research* (Mohd Basri, W; Choo, Y M and Chan, K W eds.). MPOB, Bangi. p. 47-101.

KWONG, Q B; TEH, C K; ONG, A L; HENG, H Y; LEE, H L; MOHAMED, M; LOW, J Z-B; APPAROW, S; CHEW, F T; MAYES, S; KULAVEERASINGAM, H; TAMMI, M and APPLETON, D R (2016). Development and validation of a high-density SNP genotyping array for African oil palm. *Molecular Plant, 9*: 1132-1141.

LAWRENCE, M J and RAJANAIDU, N (1985). The genetical structure of natural populations and sampling strategy. *Proc. of the International Workshop on Oil Palm Germplasm and Utilisation*. PORIM, Bangi. p. 15-26.

LEE, G A; KOH, H J; CHUNG, H K; DIXIT, A; CHUNG, J W; MA, K H; LEE, Y S; LEE, J R; LEE, G S; GWAG, J G; KIM, T S and PARK, Y J (2009). Development of SNP-based CAPS and dCAPS markers in eight different genes involved in starch biosynthesis in rice. *Molecular Breeding, 24*: 93-101.

LEE, M; XIA, J H; ZOU, Z; YE, J; RAHMADSYAH; ALFIKO, Y; JIN, J; LIEANDO, J V; PURNAMASARI, M I; LIM, C H; SUWANTO, A; WONG, L; CHUA, N-H and YUE, C H (2015). A consensus linkage map of oil palm a major QTL for stem height. *Scientific Reports, 5*: 8232.

LI, L; TACKE, E; HOFFERBERT, H R; LUBECK, J; STRAHWALD, J; DRAFFEHN, A M; WALKEMEIER, B and GEBHARDT, C (2013). Validation of candidate gene markers for marker-assisted selection of potato cultivars with improved tuber quality. *Theoretical and Applied Genetics, 126*: 1039-1052.

LI, X; WEI, Y; MOORE, K J; MICHAUD, R; VIANDS, D R; HANSEN, J L; ACHARYA, A and BRUMMER, E C (2011). Association mapping of biomass yield and stem composition in a tetraploid alfalfa breeding population. *The Plant Genome, 4*: 24-35.

LOW, E T L; ROSLI, R; JAYANTHI, N; MOHD AMIN, A H; AZIZI, N; CHAN, K L; MAQBOOL, N J; MACLEAN, P; BRAUNING, R; MCCULLOCH, A; MORAGA, R; ONG-ABDULLAH, M and SINGH, R (2014). Analyses of hypomethylated oil palm gene space. *PLoS ONE, 9*: e86728.

LU, Y; XIAO, P; SHAO, Y; ZHANG, G; THANYASIRIWAT, T and BAO, J (2010). Development of new markers to genotype the functional SNPs of SSlla, a gene responsible for gelatinization temperature of rice starch. *J. Cereal Science, 52*: 438-443.

MPOB (2016). *Overview of Malaysian Oil Palm Industry 2015*. MPOB, Bangi.

MPIC (2016). *Annual Report 2015*. Ministry of Plantation Industries and Commodities, Putrajaya.

MOHD DIN, A; RAJANAIDU, N and JALANI, B S (2000). Performance of *Elaeis oleifera* from Panama, Costa Rica, Colombia and Honduras in Malaysia. *J. Oil Palm Res. Vol. 12*: 71-80.

MULLER, M; SEIFERT, S and FINKELDEY, R (2015). A candidate gene-based association study reveals SNPs significantly associated with bud burst in European beech (*Fagus sylvatica* L.). *Tree Genetics & Genomes, 11*: 116.

MURRAY, S C; ROONEY, W L; HAMBLIN, M T; MITCHELL, S E and KRESOVICH, S (2009). Sweet sorghum genetic diversity and association mapping for brix and height. *The Plant Genome, 2*: 48-62.

NOH, A; KUSHAIRI, A; MOHD DIN, A; MAIZURA, I; MARHALIL, M; OSMAN, A and RAJANAIDU, N (2008). Genetic variation for long stalk and high protein kernel in oil palm germplasm. *Proc. of the 3rd Seminar on Performance of MPOB PS1 and PS2 Materials and Elite Germplasm*. MPOB, Bangi. p. 150-167.

ONG, P W; MAIZURA, I; ABDULLAH, N A P; RAFII, M Y; OOI, L C L; LOW, E T L and SINGH, R (2015). Development of SNP markers and their application for genetic diversity analysis in the oil palm (*Elaeis guineensis*). *Genetics and Molecular Research, 14*: 12205-12216.

PINO DEL CARPIO, D; BASNET, R K; DE VOS, R C H; MALIEPAARD, C; PAULO, M J and BONNEMA, G (2011). Comparative methods for association studies: A case study on metabolite variation in a *Brassica rapa* core collection. *PLoS ONE, 6*: e19624.

POOTAKHAM, W; JOMCHAI, N; RUANG-AREERATE, P; SHEARMAN, J R; SONTHIROD, C; SANGSRAKRU, D; TRAGOONRUNG, S and TANGPHATSORNRUANG, S (2015). Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics, 105*: 288-295.

PRITCHARD, J K; STEPHENS, M and DONNELLY, P (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*: 945-959.

RAJANAIDU, N; JALANI, B S and DOMINGOS, M (1991). Collection of oil palm germplasm in Angola. *ISOPB Newsletter, 8*: 2-3.

RAJANAIDU, N; JALANI, B S; AHMAD, K D and RAO, V (1999). Oil palm genetic resources: collection, evaluation, utilization and conservation. *Proc. of the Symposium on the Science of Oil Palm Breeding*. PORIM, Bangi. p. 219-255.

RAJANAIDU, N; KUSHAIRI, A; RAFII, M; MOHD DIN, A; MAIZURA, I and JALANI, B S (2000). Oil palm breeding and genetic resources. *Advances in Oil Palm Research* (Yusof, B; Jalani, B S and Chan, K W eds.). MPOB, Bangi. p. 171-237.

RAVEL, C; PRAUD, S; MURIGNEUX, A; LINOSSIER, L; DARDEVET, M; BALFOURIER, F; DUFOUR, P; BRUNEL, D and CHARMET, G (2006). Identification of *Glu-B1-1* as a candidate gene for the quantity of high-molecular-weight glutenin in bread wheat (*Triticum aestivum* L.) by means of an association study. *Theoretical and Applied Genetics, 112*: 738-743.

RIJU, A; CHANDRASEKAR, A and ARUNACHALAM, V (2007). Mining for single nucleotide polymorphisms and insertions/deletions in expressed sequence tag libraries of oil palm. *Bioinformation, 2*: 128-131.

SHI, A; BUCKLEY, B; MOU, B; MOTES, D; MORRIS, J B; MA, J; XIONG, H; QIN, J; YANG, W; CHITWOOD, J; WENG, Y and LU, W (2016). Association analysis of cowpea bacterial blight resistance in USDA cowpea germplasm. *Euphytica, 208*: 143-155.

SINGH, R; MARIA, M; LOW, L E T; OOI, L C L; CHAN, P L; ROZANA, R; TING, N C and MAIZURA, I (2011). Oil palm genomics: a foundation for improved agricultural productivity. *Further Advances in Oil Palm Research* (Mohd Basri, W; Choo, Y M and Chan, K W eds.). MPOB, Bangi. p. 202-251.

SINGH, R; ONG-ABDULLAH, M; LOW, L E T; MANAF, M A; ROSLI, R; NOOKIAH, R; OOI, L C L; OOI, S E; CHAN, K L; HALIM, M A; AZIZI, N; NAGAPPAN, J; BACHER, B; LAKEY, N; SMITH, S W; HE, D; HOGAN, M; BUDIMAN, M A; LEE, E K; DESALLE, R; KUDRNA, D; GOICOECHEA, J L; WING, R A; WILSON, R K; FULTON, R S; ORDWAY, J M; MARTIENSSEN, R A and SAMBANTHAMURTHI, R (2013). Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. *Nature, 500*: 335-339.

SOMERS D J; KIRKPATRICK, R; MONIWA, M and WALSH, A (2003). Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. *Genome, 49*: 431-437.

THORNSBERRY, J M; GOODMAN, M M; DOEBLEY, J; KRESOVICH, S; NIELSEN, D and BUCKLER, E S (2001). *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics, 28*: 286-289.

WANG, H; TIAN, C E; DUAN, J and WU, K (2008). Research progress on *GH3*s, one family of primary auxin-responsive genes. *Plant Growth Regulation, 56*: 225-232.

WANG, Y and LI, J (2008). Molecular basis of plant architecture. *Annual Review of Plant Biology, 59*: 253-279.

WESTFALL, C S; HERRMANN, J; CHEN, Q; WANG, S and JEZ, J M (2010). Modulating plant hormones by enzyme action: The *GH3* family of acyl acid amido synthetases. *Plant Signaling & Behavior, 5*: 1607-1612.

YU, S; ZHANG, F; WANG, X; ZHAO, X; ZHANG, D; YU, Y and XU, J (2010). Genetic diversity and marker-trait associations in a collection pak-choi (*Brassica rapa* L. ssp. *chinensis* Makino) accessions. *Genes & Genomics, 32*: 419:428.

ZHANG, S W; LI, C H; CAO, J; ZHANG, Y C; ZHANG, S Q; XIA, Y F; SUN, D Y and SUN, Y (2009). Altered architecture and enhanced drought tolerance in rice via the down-regulation of indole-3-acetic acid by *TLD1/OsGH3.13* activation. *Plant Physiology, 151*: 1889-1901.

ZHU, Y L; SONG, Q J; HYTEN, D L; VAN TASSELL, C P; MATUKUMALLI, L K; GRIMM, D R; HYATT, S M; FICKUS, E W; YOUNG, N D and CREGAN, P B (2003). Single nucleotide polymorphism in soybean. *Genetics, 163*: 1123-1934.