Putative regulatory candidate genes for QTL linked to fruit traits in oil palm (*Elaeis guineensis* Jacq.)

Ting Ngoot-Chin ● Sean Mayes ● Festo Massawe ● Ravigadevi Sambanthamurthi ● Johannes Jansen ● Sharifah Shahrul Rabiah Syed Alwee ● Seng Tzer-Ying ● Maizura Ithnin ● Rajinder Singh*

N-C. Ting ● R. Sambanthamurthi ● M. Ithnin ● R. Singh* (corresponding author)
Advanced Biotechnology and Breeding Centre, Malaysian Palm Oil Board (MPOB), P.O. Box 10620, 50720 Kuala Lumpur, Malaysia
e-mail: rajinder@mpob.gov.my

S. Mayes
Plant and Crop Sciences, Sutton Bonington Campus, University of Nottingham, Sutton Bonington, Loughborough LE12 5RD, UK

N-C. Ting ● F. Massawe
Biotechnology Research Centre, School of Biosciences, University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor, Malaysia

J. Jansen
Biometris, Wageningen University and Research Centre, P.O. Box 100, 6700 AC Wageningen, the Netherlands

S. S. R. Syed Alwee ● T-Y. Seng
FELDA Global Ventures Research and Development Sdn. Bhd. (FGV R&D), Kuala Lumpur, Wilayah Persekutuan Malaysia

T-Y. Seng
IOI Palm Biotech Sdn. Bhd., IOI Resort, 62502 Putrajaya, Malaysia (Current address)

ACKNOWLEDGEMENTS

The authors would like to thank the Director-General of Malaysian Palm Oil Board for permission to publish this paper. Part of the work was carried out at Biometris, Wageningen University and Research Centre, the Netherlands. The authors thank Dr. Rex Bernardo from University of Minnesota for permission to use the G Model software. We would also like to extend our appreciation to Ms. Maizura Azwanie Mohd Zarawi for her technical assistance in PCR. We thank Bioinformatics team at Malaysian Palm Oil Board for providing the automatic computational pipeline used for sequence homology search and sequence extraction.

ABSTRACT

Palm oil is among the most important vegetable oils, contributing to a quarter of the world's oils and fats market. The oil palm (*Elaeis guineensis* Jacq.) fruitlets, which are the source of palm oil, vary from 8 to 20 g in weight. Palm oil content in the fruitlets is approximately 45 – 50 % by weight and an increase in the percentage of mesocarp-to-fruit is likely to have a positive effect on oil yield. In this study, we report a quantitative trait loci (QTL) associated with two yield related components, namely fruit and mesocarp content in a commercial breeding population (Deli dura x Yangambi pisifera). The QTL confidence interval of about 12 cM (~6.7 Mbp) was fine-mapped with 31 markers (17 SNPs and 14 SSRs) consisting of 20 nuclear markers derived from the maternal parent, six paternal and five co-segregating markers. Interestingly, inheritance of the paternal alleles leads to a larger difference in both fruit and mesocarp weight, when comparing genotypes in the progeny palms. Candidate genes and transcription factors were mined from the QTL region by positioning markers on the oil palm EG5 genome build. Putative genes and transcription factors involved in various biological processes including flower organ development, flowering, photosynthesis, microtubule formation, nitrogen and lipid metabolism were identified within this QTL interval on pseudo-chromosome 3. This genome-based approach allowed us to identify a number of potential candidate gene markers associated with oil palm fruit and mesocarp weight which can be further evaluated for potential use in marker-assisted breeding.

KEYWORDS

Deli dura; Yangambi pisifera; Yield components; Pseudo-chromosome; Marker-assisted selection

INTRODUCTION

The oil palm (*Elaeis guineensis* Jacq.) is a diploid ($2n = 2x = 32$) outbreeding perennial monocot with a genome of approximately 1.8 billion base-pairs (Singh et al. 2013). The oil palm is a major oil producing crop and is widely grown in Southeast Asia, mostly in Indonesia and Malaysia. In Malaysia, an over 5.7 million hectares of agricultural land has been planted with oil palm (Kushairi et al. 2017). The palm is a naturally highly productive crop where the annual fresh fruit bunch (FFB) yield averages about 19.0 metric tons per hectare per year (mt/ha/yr) during the economic lifespan of the palm (Teoh 2002, Wahid and Simeh 2009). Depending on the genotype and the agroecological conditions, the FFB yield could increase to 25 – 30 mt/ha between the 7th to 18th year after planting (Woittiez et al. 2017). After the 18th year, FFB yield starts to decrease until the end of its commercial lifespan, which is normally 25 – 30 years after field planting (Zulkifli et al. 2010, Woittiez et al. 2017). Since the oil palm is a perennial crop with a long lifespan, conventional breeding to develop new and improved varieties is a long process. Approximately 20 years are needed for a modified recurrent selection (Rosenquist 1990) or reciprocal recurrent selection (Gascon and de Berchoux 1963) breeding scheme, from the selection of elite palms for crossing programmes to the commercialization of quality seeds (Rajanaidu et al. 2000). In comparison, the annual oil crop *e.g.* soybean (*Glycine max*) breeding programme requires a much shorter time (four-and-a-half to seven years) to make the relevant crosses and evaluate yield performance, while release of improved varieties to the market takes an additional three to three-and-a-half years (Pathan and Sleper 2008).

To assist oil palm breeding efforts, molecular markers are being associated with yield traits. Numerous marker systems have been exploited for this purpose such as restriction fragment length polymorphism (RFLP), simple sequence repeat (SSR) and single nucleotide polymorphism (SNP). Using these marker systems, at least 268 QTLs have been linked

to various yield components (YC) mostly involving Deli x Yangambi and Deli x AVROS genetic backgrounds (Rance et al. 2001, Billotte et al. 2010, Jeennor and Volkaert 2014, Pootakham et al. 2015, Seng et al. 2016). These published QTL studies were based on genetic linkage maps developed from a specific bi- or a multi-parental mapping family which limited the resolution of the QTL confidence interval to wide genetic regions and only accounted for genetic variation (polymorphic allelic markers) present within a specific family (Korte and Farlow 2013). This limits the application of such QTL-linked markers across diverse genetic backgrounds. To help overcome this limitation, association mapping studies involving germplasm collections have been initiated in oil palm (Teh et al. 2016, Ithnin et al. 2017). However, a point to consider is that in Southeast Asia, the commercial planting materials – tenera palms – are mainly derived from very limited genetic backgrounds. Commercial lines usually involve Deli dura maternal and pisifera paternal palms that are either Yangambi or AVROS based. Therefore, markers tightly linked to important QTLs such as YC that are informative in these genetic backgrounds may be desirable to assist in the development of improved parental lines *via* marker-assisted selection (MAS).

Ultimately, it is important to understand the molecular mechanisms regulating yield traits in oil palm. Model and annual plants have revealed interesting information in this direction. A number of genes and transcription factors (TFs) involved in the biosynthesis of oil, fatty acids and development of embryos and seeds were found to affect oil yield in *Arabidopsis thaliana*, *Brassica napus*, *Zea mays* and *Sesamum indicum* (Roesler et al. 1997, Baud et al. 2007, Liu et al. 2009, Wei et al. 2015, Elahi et al. 2016). For oil palm, the availability of the genome and transcriptome sequence data (Tranbarger et al. 2011, Bourgis et al. 2011, Dussert et al. 2013, Singh et al. 2013, Guerin et al. 2016, Jin et al. 2016) have now made it possible to identify candidate genes from the QTL intervals. To date, a few of the QTLs linked to oil yield have been found co-localized with genes involved in biosynthesis of fatty acids and lipids (Jeennor and Volkaert 2014, Teh et al. 2016), although a formal proof that these candidates are the genes responsible for the trait variation is difficult in oil palm. This is due to the long tissue culture (*e.g.* somatic embryogenesis) and maturity time for genetic modification and trait confirmation. A number of genes involved in metabolic activities *e.g.* carbohydrate, amino acid, energy, stress and structure have also been correlated to differences in oil yield using proteomics-based studies (Ooi et al. 2015). These studies suggest that yield is a complex trait in oil palm, as for most crops.

In this study, a well-saturated genetic map from a commercial breeding cross (Deli dura x Yangambi pisifera) (Ting et al. 2014) was used to identify QTLs associated with mean fruit weight (MFW) and mean mesocarp weight (MPW). A specific QTL interval on linkage group (LG) DP1 was linked with high confidence to pseudo-chromosome 3 (CHR03) of the published oil palm EG5 genome build (Singh et al. 2013). The corresponding QTL region on CHR03 revealed a number of interesting candidate genes and TFs involved in various biological activities related to development and formation of fruits which could possibly affect oil palm fruit yield. This approach allowed us to identify potential candidate genes and TFs associated with MFW and MPW.

MATERIALS AND METHODS

Mapping population

The mapping population named as P2, was derived from a cross involving Ulu Remis Deli dura (ENL48) with a Yangambi pisifera (ML161) and consisted of 87 F_1 tenera palms (05 Trial 1). This high-yielding breeding cross, which was created by the FELDA Agricultural Services Sdn. Bhd. (FASSB) (now owned by FELDA Global Ventures, FGV) has been actively exploited for production of commercial planting materials (Chin and Suhaimi 1996).

Fruit yield phenotypic data

Phenotypic data measurement was carried out by FASSB over four years and six months period (June 1995 – December 1999) using the standard oil palm fruit bunch analysis procedure (Blaak et al. 1963, Rao et al. 1983, Isa et al. 2011). The two fruit traits were measured as follows: MFW(g) for a sub-sample = fruit weight (g) / number of fresh nuts in samples and; MPW(g) for a sub-sample = fruit weight (g) - fresh nut weight (g) / number of fresh nuts. The Kolmogorov-Smirnov normality and Pearson correlation tests were performed using SPSS 16.0.

Genetic linkage map and QTL analysis

The published P2 genetic linkage map (Ting et al. 2014) was used for analysis of marker-trait association. The 16-linkage group integrated map consists of 1,331 SNP and SSR markers that spanned 1,867 cM. The QTL analysis was performed using a combined method of Interval Mapping (IM), Multiple-QTL Model (MQM) and Kruskal-Wallis test (KW) implemented *via* MapQTL® 6 (van Ooijen 2009). The 95.0 % genome-wide LOD significant threshold was determined by 1,000 permutations. G model (GM) analysis was carried out to estimate the effect of marker for each QTL identified (Bernardo 2013).

Mapping of QTL to EG5 and mining of candidate genes

SNP and SSR markers within the QTL interval were mapped to CHR03 of the EG5 genome build using the MPOB in-house homology search pipeline. The chromosomal region associated with QTL interval was extracted and searched for significant homology (BLASTN and BLASTX) in the NCBI databases (nt, nr and refseq_protein). SSR markers (sPSc) were developed to flank the candidate genes and TFs of interest.

Mapping of candidate SSR markers

Oil palm genomic DNA was extracted from spear leaves and used to genotype the SSR markers linked to the candidate genes. Preparation of DNA stock and PCR for SSR fragment analysis using the ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA) was done as described by Ting et al. (2013, 2014). Candidate markers were mapped by pre-fixing the existing P2 genetic map order using the '*Start Order*' function implemented *via* JoinMap® 4.1 (van Ooijen 2006).

RESULTS

Phenotypic data associated with yield components (YC)

The phenotypic data for mean fruit weight (MFW), mean mesocarp weight (MPW), mesocarp/fruit ratio (MTF), shell/fruit ratio (STF) and oil/dry mesocarp (OTDP) was available for about 75 of the 87 P2 progeny palms. Phenotypic data was not available for the two parental palms therefore, a comparison between progenies and the parents could not be carried out. The MFW on average was 12.34 ± 2.20 g of which, 10.57 ± 1.97 g consisted of fleshy pulp (mesocarp) measured as MPW (Table 1). For STF, MTF and OTDP, the average percentage is 8.20 (± 1.82), 85.58 (± 2.95) and 79.81 (± 1.72), respectively. All the YC demonstrated a continuous and normal distribution with p values greater than 0.05 (Kolmogorov-Smirnov test, SPSS 16.0). The population also showed good segregation for the traits concerned, which facilitated QTL analysis.

The Pearson correlation test showed that MPW is strongly correlated with MFW (r = 0.98) but moderately correlated with MTF (r = 0.337), OTDP (r = 0.252**)** and STF (r = -0.345). The shell component (STF) also as expected showed negative correlation with other mesocarp related measurements. Among these, the strongest correlation was with MTF (r = -0.930) (Table 2).

QTL associated with fruit traits

Using the published P2 genetic map (Ting et al. 2014), the map interval ranging from 75.1 – 92.8 cM on LGDP1 was found significantly associated with MFW and MPW as determined using the combined methods of IM, MQM, KW and GM (Figure 1, Table 3). The QTL interval contained 17 SNPs and three SSRs and the QTL peak was flanked by three SNP markers SNPM01424 (co-localized with SNPM04797) and SNPM01854 at 67.9 and 70.2 cM, respectively. The LOD scores of these flanking markers were almost similar for MFW (~LOD 4.6) and MPW (LODs 4.1 – 4.2). The percentage of phenotypic variation explained by the flanking markers was 25.7 % for MFW and 23.4 % for MPW as measured by IM and MQM. For SNPM01424 (and SNPM04797), marker effect was -0.80 g (p = 0.000305) for MFW and -1.27 g (p = 0.000553) for MPW, as determined using GM. The genetic effect for SNPM01854 was estimated to be 1.43 g (p = 0.004037) and 1.17 g (p = 0.011879) for MFW and MPW, respectively.

Similarly, QTLs were identified for STF, MTF and OTDP on other linkage groups (Table 2). On LGDP7, the QTL interval was from 74.8 – 84.5 cM, with mEgCIR0009 being the closest marker to the QTL peak for STF with an effect of -0.61 (*p* = 0.005230). Based on the observed '*ab*' and '*aa*' genotypes, palms were categorized into two distinct groups with mean phenotypic values for STF of 7.7 % (SD 1.4) and 9.0 % (SD 2.1), respectively. The difference between the two groups was 1.4 %. In fact, the same marker was also associated with the QTL for MTF where, palms grouped according to the two genotypes revealed a difference of 2.35 % for the trait. In addition to LGDP7, LGDP12 hosted another genome-wide significant QTL, namely OTDP at the map position ranging from 58.0 – 70.4 cM. Two co-localized SNP markers namely, SNPM00762 and SNPM00231 were the closest to the QTL peak showing a moderate effect of -0.74 for the trait, as estimated by GM. A difference of 1.1 % was observed for OTDP between the '*ab*' ('*aa*') (80.2 % ± 1.6) and the '*aa*' ('*ab*') (79.1 % ± 1.8) genotypes of SNPM00762 (and SNPM00231). The sequence information for the SNP markers is available at http://genomsawit.mpob.gov.my.

MFW and MPW linked QTL corresponding region in CHR03 revealed interesting candidate genes

The 20 markers within the QTL confidence interval were mapped to pseudo-chromosome CHR03 of the oil palm EG5 genome build, anchoring at 16,227,200 – 22,962,000 bp. The QTL-linked genomic sequence corresponding to approximately 6.7 Mbp was searched for homology to known genes in the NCBI databases. The BLASTN results revealed almost 100.0 % sequence similarity (E-value 0; identity 99.0 – 100.0 %) to at least 14 predicted genes and TFs in *E. guineensis*. These were *MYB21*, *oxysterol-binding protein-related protein 1C* (*ORP1C*), *DELLA SLR1* (*SLR1*), *zinc finger NUTCRACKER* (*NUC*), *α,α-trehalose-phosphate synthase* (*TPS6*), *chlorophyll a-b binding protein of LHCII type 1* (*Lhcb1*), *microtubule-associated protein 70-1* (*MAP70.1*), *T-complex protein 1 subunit delta* (*CCT4*), *tubulin beta chain* (*TUBB*), *glutamine synthetase nodule isozyme* (*GS*), *lipid phosphate phosphatase 2* (*LPP2*), *ORP1C*, *3-ketoacyl-CoA synthase 6* (*KCS6*), *protein VAC14 homolog* and *chalcone synthase* (*CHS*). These protein-encoding genes and TFs have been reported to be involved in various biological processes, particularly development of stamen and pollen as well as flowering. Putative functions of these genes and TFs are summarized in Online Resource 1.

Mapping candidate SSR markers to the QTL interval

Using the candidate gene strategy described by Ting et al. (2016), primer-pairs were designed to amplify SSRs flanking the candidate genes, based on their physical position (Figure 1, Online Resource 2). Ten SSR markers developed for *Lhcb1*, *KCS6*, *VAC14*, *CCT4*, *SLR1* and *ORP1C* were informative and successfully mapped back to the QTL confidence interval. The SSR markers developed for other candidate genes were either not polymorphic {nine SSRs for *NUC* (one), *GS1* (one), *MAP70.1* (one), *TPS6* (one), *TUBB* (one), *CHS* (two) and *LPP2* (two)} or could not be mapped confidently (one SSR for *MYB21*) onto the genetic map. Mapping of candidate SSR markers increased the total number of markers to 31 (17 SNPs and 14 SSRs) within the 11.9-cM QTL interval. The 31 SNP and SSR markers consisted of 20 nuclear markers polymorphic in the maternal parent, six in the paternal palm, with five co-segregating. The distance between these markers ranged from 0 – 2.3 cM with the majority (81.0 % markers) co-localizing with at least one other marker, thus resulting in seven clusters (with zero map distance). This suggests that the QTL interval had a high enough density of markers and additional markers are unlikely to further improve the map resolution, as the population size is the limiting factor.

By mapping these candidate SSR markers, it was observed that the LOD scores for markers flanking the QTL peak increased slightly for MFW (LODs 4.6 – 4.8) and MPW (LODs 4.2 – 4.3) (Online Resource 3). The candidate SSR marker for a tubulin-related gene (*CCT4*) sPSc00513, consisting of AC and GT compound motifs [$(AC)_{13}$GTACATAGACACACAC$(GT)_9$], was mapped close to the QTL peak. The candidate SSR marker co-localized with the existing SNP markers (SNPM01424 and SNPM04797) on the P2 genetic map, on the right of the QTL peak (Figure 2). *CCT4* is known to play an important role in assisting the folding of tubulin protein, the main structure of microtubules in cell wall and fibres that are abundantly found in the oil palm columnar stem and fruits (Online Resource 1). In CHR03, *CCT4* is located at 19,110,773 bp and sPSc00513, located at a distance of ~2.8 kb from *CCT4*. The distance is much closer than the two existing SNP markers which were 38.0 kb (SNPM01424) and 119.4 kb (SNPM04797) away from *CCT4*. For SNPM01424 and SNPM04797, the single nucleotide variation observed in the two parents (maternal dura ENL48 x paternal pisifera ML161) were TT x TG and CC x TC, respectively. Similar to the two SNP markers, sPSc00513 was also heterozygous in the paternal ML161. On the left-side of the QTL peak (towards the end of CHR03), SNPM01854 revealed heterozygous SNP alleles (AG) in the maternal ENL48 parent and was homozygous GG in ML161.

The heterozygous SNP alleles *e.g.* AG, TC and TG were scored as '*ab*' whereas, the homozygous allelic genotypes *e.g.* TT, CC and GG were coded as '*aa*' for further analyses. Evaluation by comparing the phenotypic mean values between the '*ab*' and '*aa*' genotypes (T-test, SPSS 16.0) showed that the paternal alleles (sPSc00513, SNPM01424 and SNPM04797) revealed a larger difference in MFW (difference of 1.64 g) and MPW (difference of 1.41 g), compared to the alleles inherited from the maternal parent for marker SNPM01854. The maternal alleles for marker SNPM01854 revealed a slightly lower difference of 1.33 g in MFW and 1.14 g in MPW among the '*ab*' and '*aa*' genotypes. Palms that inherited the paternal '*ab*' genotype of sPSc00513, SNPM01424 and SNPM04797 had a higher MFW (13.30 g) compared to the '*aa*' genotype group (11.66 g). If a fresh fruit bunch for oil palm is estimated to contain 2,000 fruitlets, the 13.30 g MFW in the '*ab*' genotype can produce an average of 26.6 kg FFB compared to 23.3 kg in the '*aa*' genotype (not including the stalk and spikelet).

For MPW, the '*ab*' genotype demonstrated a higher weight (11.40 g) compared to that observed for the '*aa*' group (9.99 g). The same trend was also noticed for the paternally inherited alleles for other markers (sPSc00520, mEgCIR0268A and SNPM00813) within the QTL confidence interval, showing a relatively larger difference in MFW

and MPW when compared to the maternal heterozygous alleles. For the paternal alleles, their differential power decreases towards left of the QTL interval whereas for maternal alleles, the differential power decreases in an opposing direction (Figure 2).

DISCUSSION

This study revealed genome-wide QTLs associated with YC namely, MFW, MPW, MTF, STF and OTDP. Moving forward, this study further examined the QTL confidence interval on LGDP1 that was associated with MFW and MPW by saturating the region with SSR markers linked to specific candidate genes. The informative markers linked to candidate genes were successfully positioned on the QTL interval. The improved resolution obtained was obvious when the entire QTL interval was mapped to CHR03 of the EG5 genome build. These SSR markers such as sPSc00513, sPSc00516 and sPSc00520 are physically positioned closer to the candidate genes, and thus can act as proxies for the genes for further validation.

The QTL interval revealed 14 candidate genes and TFs related to plant architecture, photosynthesis, nitrogen metabolism, lipid transportation and metabolism, stress response, flowering and formation of stamen and microtubule (Online Resource 1). From the QTL interval, clustering of genes was obvious particularly in the region 18,043,640 – 19,230,181 bp (flanked by sPSc00373 and SNPM04797) positioned next to the QTL peak, on the right (Figure 2). Six candidate genes were identified in this cluster and the shortest adjacent distance was only 11,509 bp, observed between *GS* and *VAC14*. *GS* encodes for glutamine synthetase nodule isozyme which is involved in nitrogen metabolism (Seabra et al. 2013) whereas, *VAC14* regulates lipid kinase activity (Zhang et al. 2012, Stecker et al. 2014).

In close proximity to *GS* and *VAC14*, four other genes *KCS6*, *MAP70.1*, *TPS6* and *CCT4* were identified. *KCS6* was detected at a distance of about 489,600 bp (towards the beginning of the chromosome) and it is the only fatty acid related gene detected in the QTL interval. The gene codes for the KCS6 enzyme, known to be involved in the synthesis of very-long-chain fatty acids for production of cuticular wax, required for the formation of coat layer on fruits, leaves, flowers and pollen (Lee and Suh 2013, Lokesh et al. 2013). Its function is to prevent uncontrolled water loss and protect plants from biotic stresses (Domínguez et al. 2011). However, *KCS6* has also been reported to affect formation of flower organs. In tomato (*Solanum lycopersicum*), wax deficiency resulting from the disruption of *KCS6*, impairs fertility, due to fusion between epidermal cells in flower (Smirnova et al. 2013).

At a distance of ~122,300 bp from *GS* and *VAC14* towards the end of CHR03, *MAP70.1* encoding the microtubule-associated protein was identified. It is required for the assembly of microtubules that are essential for cell division and for the formation of cell wall, xylem, fibres and wood in plants (Mandelkow and Mandelkow 1995, Pesquet et al. 2010). Structurally, the microtubules are made up of many (α- and β-) tubulin proteins (Snustad et al. 1992, Blume et al. 2010, Rao et al. 2016). Two tubulin-related genes *CCT4* and *TUBB* were identified in the QTL region. This points to a region that may contain a set of genes that are candidates for further investigation, especially on their role in oil palm fruit formation. The other two genes within the confidence interval namely, *TPS6* and *NUC* are both involved in modulating sucrose metabolism although, they have also been linked to other functions such as stress response (*TPS6*) and regulation of photoperiodic flowering (*NUC*) (Seo et al. 2011, Goddijn and van Dun 1999, Chary et al. 2008, Ponnu et al. 2011).

A tissue-specific TF, *MYB21* was found to be located (at 20,928,385 bp) close to the QTL peak on the left. The TF is reported to be expressed only in flowers in *Antirrhinum* (Jackson et al. 1991, Moyano et al. 1996) and *A. thaliana* (Shin et al. 2002). *MYB21* (together with other members *e.g. MYB24* and *MYB57*) is known to interact with jasmonic acid (JA) and gibberellic acid (GA) to promote stamen development. This is because deficiency in either JA or GA will cause male sterility (Mandaokar et al. 2006, Cheng et al. 2009, Song et al. 2011). For oil palm, the involvement of JA and GA during inflorescence induction was reported by Ajambang et al. (2015). The association of *MYB21* with oil palm yield was demonstrated recently, where *MYB21* expression was found to be significantly higher in the high-yielding palms compared to low-yielding palms (Wong et al. 2017). The findings by Wong et al. (2017) have drawn attention to the role of the male inflorescence, which has a direct impact on production of fruits in oil palm.

In addition to *MYB* TFs, accumulation (overexpression) of the DELLA protein encoded by *SLR1* also acts as a negative regulator of GA, causing male sterility and severe dwarf phenotypes (Dai and Xue 2010, Hirano et al. 2012). Interestingly, this gene was detected at a distance of 1,411,537 bp from *MYB21*. In the case of male sterility, limited or no pollen is produced, thus the female flowers are insufficiently pollinated which result in poor or no fruit formation. In order to get a sufficient supply of pollen, it is important to have palms producing appropriate ratio of male and female inflorescences. Ajambang et al. (2016) suggested planting a number of pollen-supply palms among the commercial oil palm trees in the plantation. Information on the regulatory genes and markers that are linked to the fruit traits and their pattern will help to implement these strategies efficiently.

Nevertheless, the expression of *MYB21* is induced by a photoreceptor (*COP1*) which responds to light in the nucleus (Shin et al. 2002). Disruption or malfunction in *COP1* is found to affect proper functioning of *MYB21*. This has been shown by overexpressing *MYB21* in the *COP1* mutant which resulted in malformation of flower tissues in *A. thaliana* (Shin et al. 2002). Another light responsive gene *Lhcb1* encoding a light-harvesting antenna complex II (LHCII) protein, was also identified in the QTL confidence interval. The main function of LHCII is to harvest and transfer photons into biochemical energy during the photosynthesis process (Xia et al. 2012, Pietrzykowska et al. 2014). The capacity and efficiency of light capture and conversion and, the photosynthesis process as a whole, is reported to have a strong positive relationship with the biomass and yield production in grain crops (Zhu et al. 2010, Parry et al. 2011). A study identified nucleotide variation in the *Lhcb1* gene was associated with grain traits in barley (Xia et al. 2012).

The finding of various genes and TFs in close proximity to each other within the QTL interval associated with MFW and MPW potentially supports the complex genetic mechanism that is likely to be involved in determining oil palm fruit yield (Ooi et al. 2015). Literature reveals that many of these genes (including TFs) are multi-functional with some having similar biological function (pleiotropic genes) or are involved in a complex regulatory network impacting oil palm fruit development. The co-localization of QTLs as indicated by overlapping confidence intervals for MFW and MPW in LGDP1 are consistent with these observations, as also reported previously by Billotte et al. (2010), Seng et al. (2016), and Ithnin et al. (2017).

The QTL regions in this study are also similar to those associated with other YC with moderate to low correlation in previous studies. The QTL region for MFW and MPW in LGDP1, corresponds to QTLs for mesocarp/fruit ratio (MTF) (Ithnin et al. 2017), mean shell weight (MSW), mean kernel weight (MKW), wet mesocarp/fruit ratio (WPF), shell/fruit ratio (SF), dry mesocarp/fruit ratio (DPF) and fruit/bunch ratio (FB) reported by Seng et al. (2016) (Figure 3A). Interestingly, a QTL reported for FFB by Billotte et al. (2010) is the same region hosting the QTLs for MFW and MPW

in this study. Co-localization of multiple QTLs associated with various YC that only have moderate or low significant correlation may be due to multiple tightly-linked QTLs or tightly linked trait-specific genes (Hall et al. 2006, Shi et al. 2009). Furthermore, the percentage of variation accounted for by the linked loci could be relatively small and may not be in the same direction, particularly if there have been historical recombination events to give dissociative haplotypes.

Interestingly, the QTL for OTDP at interval 58.0 – 70.4 cM in LGDP12 corresponds to the physical interval 34,828,628 – 40,396,733 bp in CHR05. The same region was reported to be associated with a similar trait (named as O/DM by Teh et al. 2016) identified *via* a genome-wide association study (Figure 3B). The three significant SNP markers (SD_SNP_000010418, SD_SNP_000019529 and SD_SNP_000002370) reported by Teh et al. (2016) are located within the confidence interval in the present study when anchored to CHR05. The SD_SNP_000019529 (reported as the most significant locus for O/DM) is located at a distance of 30,995 bp and 136,882 bp respectively from the two SNP markers, SNPM00231 and SNPM000762, detected in the present study as the flanking markers to the QTL peak for OTDP. In P2, phenotypic difference for OTDP (1.1 %) observed between the heterozygous and homozygous allelic profiles is almost similar to the phenotypic variation observed by Teh et al. (2016) (1.3 – 1.4 %) when validating the most significant marker across 2,091 palms from the Deli x AVROS and Nigerian x AVROS backgrounds. Thus, this adds confidence to the reliability of the QTL for OTDP identified in the present study.

CONCLUSION

This study revealed a number of QTLs associated with yield components namely, MFW, MPW, MTF, STF and OTDP in an advanced dura x pisifera population. Some of the QTLs were consisted with those reported previously, signifying the potential application in breeding programmes. The QTL interval associated with MFW and MPW, when mapped to the oil palm EG5 genome build revealed a number of interesting genes and transcription factors. A total of 14 candidate genes and transcription factors related to plant architecture, photosynthesis, nitrogen metabolism, lipid transportation and metabolism, stress response, flowering and formation of stamen and microtubule resided in the QTL region. Of these, *MYB21*, *Lhcb1*, *KCS6*, *MAP70.1*, *TUBB*, *CCT4*, *TPS6* and *NUC* have biological functions that could possibly regulate fruit formation in oil palm. SSR markers designed to target the candidate genes were successfully mapped to the QTL interval, where both the genome and map locations of these genes were obtained. The candidate genes will be further evaluated across different genetic backgrounds in future studies. The genome-based candidate gene approach as presented in this study has proven efficient for identifying potential candidate genes influencing yield traits in oil palm. Cross-comparison can also be made to published genetic maps and genomes of other plant species to facilitate identifying causal genes linked to oil yield.

LIST OF TABLES

Table 1. Summary of mean fruit weight (MFW) and mean mesocarp weight (MPW) measured for the P2 mapping population.

Table 2. Pearson correlation coefficient observed among the five yield components.

Table 3. Genome-wide (GW) significant QTLs detected for five fruit traits using Interval Mapping (IM), the Multiple-QTL Model (MQM), the Kruskal-Wallis non-parametric tests (KW) and G Model (GM).

LIST OF FIGURES

Figure 1. Identification of QTLs associated with mean fruit weight (MFW) and mean mesocarp weight (MPW) in linkage group (LG) DP1 (A). Subsequently, the QTL linked markers (in bold) were mapped onto the EG5 pseudo-chromosome 3 (CHR03) (B). The various genes and transcription factors identified in the corresponding QTL interval in CHR03 are indicated in red and candidate SSR markers are represented in blue fonts (C).

Figure 2. Association of marker genotypes with phenotypic values *via* mean fruit weight (MFW) and mean mesocarp weight (MPW). The QTL region on LGDP1 (inverted) is aligned to the EG5 pseudo-chromosome 3 (CHR03) and markers flanking the QTL peak are indicated by *.

Figure 3. Comparison of QTLs associated with MFW and MPW (A.) and OTDP (B.) identified in the present study (LGDP) to those published previously. Only the corresponding portions of the linkage groups containing QTLs and common markers are shown. Coloured bars next to linkage groups indicate the QTL regions and the flanking markers are coloured in purple.

ELECTRONIC SUPPLEMENTARY MATERIAL

Online Resource 1. Putative biological function for the candidate genes and transcription factors identified within the QTL region (in LGDP1) associated with mean fruit weight (MFW, g) and mean mesocarp weight (MPW, g) in the P2 mapping population.

Online Resource 2. Candidate SSR markers information.

Online Resource 3. Increased LOD scores after mapping of candidate SSR markers. LGDP1 with candidate SSR markers is indicated by orange dots and the blue dots indicate the markers mapped in published LGDP1 (Ting et al. 2014). Map positions are based on the updated LGDP1.

LIST OF ABBREVIATIONS

BLASTN : Similarity search of the NCBI nucleotide database using a nucleotide query
BLASTX : Similarity search of the NCBI protein database using a translated nucleotide query
CCT4 : *T-complex protein 1 subunit delta*
CHR03 : pseudo-chromosome 3
CHS : *Chalcone synthase*
COP1 : *Constitutively photomorphogenic 1*
cM : Centimorgan map distance
DPF : Dry mesocarp/fruit ratio
EG5 : *E. guineensis* 5th genome build as published as Singh et al. (2013)
FB : Fruit/bunch ratio
FELDA : Federal Land Development Authority Malaysia
FFB : Fresh fruit bunch
GM : G model

GS : *Glutamine synthetase nodule isozyme*

IM : Interval Mapping

KCS6 : *3-ketoacyl-CoA synthase 6*

KW : Kruskal-Wallis test

Lhcb1 : *Chlorophyll a-b binding protein of LHCII type 1*

LG : Linkage group

LOD : Logarithm of odds

LPP2 : *Lipid phosphate phosphatase 2*

MAP70.1 : *Microtubule-associated protein 70-1*

MAS : Marker-assisted selection

Mbp : Million base-pairs

MFW : Mean fruit weight

MKW : Mean kernel weight

MPW : Mean mesocarp weight

MQM : Multiple-QTL Model

MSW : Mean shell weight

MTF : Mesocarp/fruit ratio

MYB21 : *TF MYB21*

nr : NCBI non-redundant protein sequences database

nt : NCBI nucleotide collection

NUC : *Zinc finger protein NUTCARACKER*

ORP1C : *Oxysterol-binding protein-related protein 1C*

OTDP : Oil/dry mesocarp ratio

QTL : Quantitative trait loci

refseq_protein : NCBI reference proteins database

RFLP : Restriction fragment length polymorphism

SF and STF : Shell/fruit ratio

SLR1 : *DELLA protein SLR1*

SNP : Single nucleotide polymorphism

SSR : Simple sequence repeat

TF : Transcription factor

TPS6 : *α,α-trehalose-phosphate synthase [UDP-forming] 6*

TUBB : *Tubulin beta chain*

VAC14 : *Protein VAC14 homolog*

WPF : Wet mesocarp/fruit ratio

YC : Yield components

FUNDING

This study was funded by the Malaysian Palm Oil Board.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Ajambang W, Ardie SW, Volkaert H, Ngando-Ebongue GF, Sudarsono S (2015) Comparative expression profiling of three early inflorescence stages of oil palm indicates that vegetative to reproductive phase transition of meristem is regulated by sugar balance. Funct Plant Biol. 42(6):589-598.

Ajambang W, Mondjeli C, Ntsefong GN, Sudarsono S (2016) RNA-Seq analysis reveals influence of sugar level and photoperiod on seasonality in oil palm (*Elaeis guineensis* Jacq.) sex-specific inflorescence emergence. J App Biol Biotech. 4(02):016-022.

Baud S, Wuillème S, Dubreucq B, de Almeida A, Vuagnat C, Lepiniec L, Miquel M, Rochat C (2007) Function of plastidial pyruvate kinases in seeds of *Arabidopsis thaliana*. Plant J. 52(3):405-19.

Billotte N, Jourjon MF, 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, Mangin B (2010) QTL detection by multi-parent linkage mapping in oil palm (*Elaeis guineensis* Jacq.). Theor Appl Genet. 120(8):1673-1687.

Blaak G, Spaarnaaij LD, Menendez T (1963) Breeding for inheritance in the oil palm (*Elaeis guineensis* Jacq.) Part II Methods of bunch quality analysis. J W Afr Inst Oil Palm Res. 4:146.

Blume Y, Yemets A, Sheremet Y, Nyporko A, Sulimenko V, Sulimenko T, Dráber P (2010) Exposure of beta-tubulin regions defined by antibodies on an *Arabidopsis thaliana* microtubule protofilament model and in the cells. BMC Plant Biology 10:29.

Bourgis F, Kilaru A, Cao X, Ngando-Ebongue GF, Drira N, Ohlrogge JB, Arondel V (2011) Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. Proc Natl Acad Sci USA 108(30):12527-32.

Chary SN, Hicks GR, Choi YG, Carter D, Raikhel NV (2008) Trehalose-6-phosphate synthase/phosphatase regulates cell shape and plant architecture in *Arabidopsis*. Plant Physiol. 146(1):97-107.

Cheng H, Song S, Xiao L, Soo HM, Cheng Z, Xie D, Peng J (2009) Gibberellin acts through jasmonate to control the expression of *MYB21*, *MYB24*, and *MYB57* to promote stamen filament growth in *Arabidopsis*. PLoS Genet. 5(3):e1000440.

Chin CW, Suhaimi S (1996) FELDA oil palm planting materials. In proceedings of seminar on sourcing of oil palm planting materials for local and overseas joint venture: 22-23 July 1996; Selangor. Edited by Rajanaidu N, Jalani BS. Bangi: Palm Oil Research Institute of Malaysia.

Dai C, Xue H-W (2010) Rice *early flowering*1, a CKI, phosphorylates DELLA protein SLR1 to negatively regulate gibberellin signalling. EMBO J. 29(11):1916-27.

Domínguez E, Heredia-Guerrero JA, Heredia A (2011) The biophysical design of plant cuticles: an overview. New Phytologist 189:938–949.

Dao TTH, Linthorst HJM, Verpoorte R (2011) Chalcone synthase and its functions in plant resistance. Phytochem Rev. 10:397–412.

Dussert S, Guerin C, Andersson M, Joët T, Tranbarger TJ, Pizot M, Sarah G, Omore A, Durand-Gasselin T, Morcillo F (2013) Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. Plant Physiology 162:1337–1358.

Elahi N, Duncan RW, Stasolla C (2016) Molecular regulation of seed oil accumulation. J Adv Nutr Hum Metab. 2:e1296.

Gascon JP, de Berchoux C (1963) Caractéristiques qualitatives du régime d'*Elaeis guineensis* Jacq. Quelques relations entre les Dura et Tenera d'une même descendance et leur application a l'amélioration des semences. Oleagineux 18:411-5.

Goddijn OJ, van Dun K (1999) Trehalose metabolism in plants. Trends Plant Sci. 4(8):315-319.

Guerin C, Joët T, Serret J, Lashermes P, Vaissayre V, Agbessi MD, Beulé T, Severac D, Amblard P, Tregear J, Durand-Gasselin T, Morcillo F, Dussert S (2016) Gene coexpression network analysis of oil biosynthesis in an interspecific backcross of oil palm. Plant J. 87(5):423-41.

Hall MC, Basten CJ, Willis JH (2006) Pleiotropic quantitative trait loci contribute to population divergence in traits associated with life-history variation in *Mimulus guttatus.* Genetics 172(3):1829-44.

Hirano K, Kouketu E, Katoh H, Aya K, Ueguchi-Tanaka M, Matsuoka M (2012) The suppressive function of the rice DELLA protein SLR1 is dependent on its transcriptional activation activity. Plant J. 71(3):443-53.

Isa ZA, Kushairi A, Mohd Din A, Suboh O, Junaidah J, Noh A, Choo KW, Musa B (2011) A critical re-examination of the method of bunch analysis in oil palm breeding – an update. Paper presented at ISOPB conference, Kuala Lumpur Convention Centre, Kuala Lumpur.

Ithnin M, Xu Y, Marjuni M, Mohamed Serdari N, Mohd Din A, Low ETL, Tan Y-C, Yap S-J, Ooi LC-L, Rajanaidu N, Singh R, Xu S (2017) Multiple locus genome-wide association studies for important economic traits of oil palm. Tree Genet Genomes 13:103.

Itoh H, Ueguchi-Tanaka M, Sato Y, Ashikari M, Matsuoka M (2002) The gibberellin signaling pathway is regulated by the appearance and disappearance of *SLENDER RICE1* in nuclei. Plant Cell 14(1):57-70.

Jackson D, Culianez-Macia F, Prescott AG, Roberts K, Martin C (1991) Expression patterns of myb genes from *Antirrhinum* flowers. *Plant Cell* 3(2):115-25.

Jeennor S, Volkaert H (2014) Mapping of quantitative trait loci (QTLs) for oil yield using SSRs and gene-based markers in African oil palm (*Elaeis guineensis* Jacq.). Tree Genet Genomes 10:1–14.

Jin JJ, Lee M, Bai B, Sun Y, Qu J, Rahmad syah, Alfiko Y, Lim CH, Suwanto A, Sugiharti M, Wong L, Ye J, Chua N-H, Yue GH (2016) Draft genome sequence of an elite Dura palm and whole-genome patterns of DNA variation in oil palm. DNA Research 23(6):527–533.

Katagiri T, Ishiyama K, Kato T, Tabata S, Kobayashi M, Shinozaki K (2005) An important role of phosphatidic acid in ABA signaling during germination in *Arabidopsis thaliana*. Plant J. 43(1):107-17.

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29.

Kushairi A, Singh R, Ong-Abdullah M (2017) The oil palm industry in Malaysia: thriving with transformative technologies. J. Oil Palm Res. 29:431-439.

Lee SB, Suh MC (2013) Recent advances in cuticular wax biosynthesis and its regulation in *Arabidopsis*. Mol Plant. 6(2):246–249.

Liu Z, Yang X, Fu Y, Zhang Y, Yan J, Song T, Rocheford T, Li J (2009) Proteomic analysis of early germs with high-oil and normal inbred lines in maize. Mol Biol Rep. 36:813–21.

Lokesh U, Kiranmai K, Pandurangaiah M, Sudhakarbabu O, Nareshkumar A, Sudhakar C (2013) Role of plant fatty acid elongase (*3 keto acyl-CoA synthase*) gene in cuticular wax biosynthesis. RRJAAS 2(4):35-42.

Mandaokar A, Thines B, Shin B, Lange BM, Choi G, Koo YJ, Yoo YJ, Choi YD, Choi G, Browse J (2006) Transcriptional regulators of stamen development in *Arabidopsis* identified by transcriptional profiling. Plant J. 46(6):984-1008.

Mandelkow E, Mandelkow E-M (1995) Microtubules and microtubule-associated proteins. Curr Opin Cell Biol. 7:72-81.

Moyano E, Martínez-Garcia JF, Martin C (1996) Apparent redundancy in myb gene function provides gearing for the control of flavonoid biosynthesis in *antirrhinum* flowers. Plant Cell 8(9):1519–1532.

Ngo MH, Colbourne TR, Ridgway ND (2010) Functional implications of sterol transport by the oxysterol-binding protein gene family. Biochem J. 429(1):13-24.

Ooi TEK, Yeap WC, Leona DJD, Ng BZ, Lee FC, Ainul MO, Appleton DR, Chew FT, Kulaveerasingam H (2015) Differential abundance analysis of mesocarp protein from high- and low yielding oil palms associates non-oil biosynthetic enzymes to lipid biosynthesis. Proteome Sci.13:28.

Paradis S, Villasuso AL, Aguayo SS, Maldiney R, Habricot Y, Zalejski C, Machado E, Sotta B, Miginiac E, Jeannette E (2011) *Arabidopsis thaliana* lipid phosphate phosphatase 2 is involved in abscisic acid signalling in leaves. Plant Physiol Biochem. 49(3):357-62.

Parry MAJ, Reynolds M, Salvucci ME, Raines C, P. Andralojc J, Zhu X-G, Price GD, Condon AG, Furbank RT (2011) Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. J Exp Bot. 62(2):453–467.

Pathan MS, Sleper DA (2008) Advances in soybean breeding. In Genetics and Genomics of Soybean. Edited by Stacey G. Springer-Verlag, New York.

Paul S, Gable K, Beaudoin F, Cahoon E, Jaworski J, Napier JA, Dunn TM (2006) Members of the Arabidopsis FAE1-like *3-ketoacyl-CoA synthase* gene family substitute for the Elop proteins of *Saccharomyces cerevisiae*. J Biol Chem. 281(14):9018-29.

Pesquet E, Korolev AV, Calder G, Lloyd CW (2010) The microtubule-associated protein AtMAP70-5 regulates secondary wall patterning in *Arabidopsis* wood cells. Curr Biol. 20(8):744-9.

Pierrugues O, Brutesco C, Oshiro J, Gouy M, Deveaux Y, Carman GM, Thuriaux P, Kazmaier M (2001) Lipid phosphate phosphatases in *Arabidopsis*. Regulation of the AtLPP1 gene in response to stress. J Biol Chem. 276(23):20300-8.

Pietrzykowska M, Suorsa M, Semchonok DA, Tikkanen M, Boekema EJ, Aro EM, Jansson S (2014) The light-harvesting chlorophyll a/b binding proteins Lhcb1 and Lhcb2 play complementary roles during state transitions in *Arabidopsis*. Plant Cell 26(9):3646-60.

Ponnu J, Wahl V, Schmid M (2011) Trehalose-6-phosphate: connecting plant metabolism and development. Front Plant Sci. 2:70.

Pootakham W, Jomchai N, Ruang-areerate P, Shearman JR, Sonthirod C, Sangsrakru D, Tragoonrung S, 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.

Rajanaidu N, Kushari A, Rafii M, Mohd Din A, Ithnin M, Jalani BS (2000) Oil palm breeding and genetic resources. In Advances in oil palm research. Edited by Basiron Y, Jalani BS, Chan KW. Bangi:Malaysian Palm Oil Board.

Rance KA, Mayes S, Price Z, Jack PL, Corley RHV (2001) Quantitative trait loci for yield components in oil palm (*Elaeis guineensis* Jacq.). Theor Appl Genet. 103:1302-1310.

Rao G, Zeng Y, He C, Zhang J (2016) Characterization and putative posttranslational regulation of α- and β-tubulin gene families in *Salix arbutifolia.* Sci Rep. 6:19258.

Rao V, Soh AC, Corley RHV, Lee CH, Rajanaidu N, Tan YP, Chin CW, Lim KC, Tan ST, Lee TP, Ngui M (1983) A critical reexamination of the method of bunch quality analysis in oil palm breeding. PORIM Occasional Paper 9:1-28.

Roesler K, Shintani D, Savage L, Boddupalli S, Ohlrogge J (1997) Targeting of the Arabidopsis homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. Plant Physiol. 113(1):75-81.

Rosenquist EA (1990) An overview of breeding technology and selection in *Elaeis guineensis.* In proceedings 1989 PORIM International Palm Oil Development Conference, Module Agriculture: 5 – 9 September 1989; Kuala Lumpur. Bangi:Palm Oil Research Institute of Malaysia.

Seabra AR, Silva LS, Carvalho HG (2013) Novel aspects of glutamine synthetase (GS) regulation revealed by a detailed expression analysis of the entire GS gene family of *Medicago truncatula* under different physiological conditions. BMC Plant Biol. 13:137.

Seo PJ, Ryu J, Kang SK, Park CM (2011) Modulation of sugar metabolism by an INDETERMINATE DOMAIN transcription factor contributes to photoperiodic flowering in *Arabidopsis*. Plant J. 65(3):418-29.

Seng T-Y, Ritter E, Mohamed Saad SH, Leao L-J, Singh R, Qamaruz Zaman F, Tan S-G, Syed Alwee SSR, Rao V (2016) QTLs for oil yield components in an elite oil palm (*Elaeis guineensis*) cross. Euphytica 212:399–425.

Shi J, Li R, Qiu D, Jiang C, Long Y, Morgan C, Bancroft I, Zhao J, Meng J (2009) Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus.* Genetics 182(3):851-61.

Shin B, Choi G, Yi H, Yang S, Cho I, Kim J, Lee S, Paek N-C, Kim J-H, Song P-S, Choi G (2002) *AtMYB21*, a gene encoding a flower-specific transcription factor, is regulated by *COP1*. Plant J. 30(1):23-32.

Singh R, Ong-Abdullah M, Low ETL, Abdul Manaf MA, Rosli R, Nookiah R, Ooi LC-L, Ooi S-E, Chan K-L, Halim MA, Azizi N, Nagappan J, Bacher B, Lakey N, Smith SW, He D, Hogan M, Budiman MA, Lee EK, DeSalle R, Kudrna D, Goicoechea JL, Wing RA, Wilson RK, Fulton RS, Ordway JM, Martienssen RA, Sambanthamurthi R (2013) Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. Nature 499(7459):doi:10.1038/nature12309.

Skirpan AL, Dowd PE, Sijacic P, Jaworski CJ, Gilroy S, Kao TH (2006) Identification and characterization of PiORP1, a *Petunia* oxysterol-binding-protein related protein involved in receptor-kinase mediated signaling in pollen, and analysis of the *ORP* gene family in *Arabidopsis*. Plant Mol Biol. 61(4-5):553-65.

Snustad DP, Haas NA, Kopczak SD, Silflowa CD (1992) The small genome of *Arabidopsis* contains at least nine expressed β-tubulin genes. Plant Cell 4:549-556.

Smirnova A, Leide J, Riederer M (2013) Deficiency in a very-long-chain fatty acid β-ketoacyl-coenzyme a synthase of tomato impairs microgametogenesis and causes floral organ fusion. Plant Physiol. 161(1):196-209.

Song S, Qi T, Huang H, Ren Q, Wu D, Chang C, Peng W, Liu Y, Peng J, Xie D (2011) The Jasmonate-ZIM domain proteins interact with the R2R3-MYB transcription factors MYB21 and MYB24 to affect Jasmonate-regulated stamen development in *Arabidopsis*. Plant Cell. 23(3):1000-13.

Stecker KE, Minkoff BB, Sussman MR (2014) Phosphoproteomic analyses reveal early signalling events in the osmotic stress response. Plant Physiol. 165(3):1171–1187.

Sun W, Meng X, Liang L, Jiang W, Huang Y, He J, Hu H, Almqvist J, Gao X, Wang Li (2015) Molecular and biochemical analysis of chalcone synthase from *Freesia hybrid* in flavonoid biosynthetic pathway. PLoS One 10(3):e0119054.

Teh CK, Ong AL, Kwong QB, Apparow S, Chew FT, Mayes S, Mohamed M, Appleton DR, Kulaveerasingam H (2016) Genome-wide association study identifies three key loci for high mesocarp oil content in perennial crop oil palm. Sci Rep. 6:19075.

Teoh CH (2002) The palm oil industry in Malaysia: from seed to frying pan. WWF Malaysia. http://assets.panda.org/downloads/oilpalmchainpartaandb_esri.pdf.

Ting N-C, Jansen J, Nagappan J, Ishak Z, Chin CW, Tan S-G, Cheah S-C, Singh R (2013) Identification of QTLs associated with callogenesis and embryogenesis in oil palm using genetic linkage maps improved with SSR markers. PLoS One 8(1):e53076.

Ting N-C, Jansen J, Mayes S, Massawe F, Sambanthamurthi R, Ooi LC-L, Chin CW, Arulandoo X, Seng T-Y, Syed Alwee SSR, Ithnin M, Singh R (2014) High density SNP and SSR-based genetic maps of two independent oil palm hybrids. BMC Genomics 15:309.

Ting N-C, Yaakub Z, Kamaruddin K, Mayes S, Massawe F, Sambanthamurthi R, Jansen J, Low ETL, Ithnin M, Kushairi A, Arulandoo X, Rosli R, Chan K-L, Amiruddin N, Sritharan K, Lim CC, Nookiah R, Mohd Din A, Singh R (2016) Fine-mapping and cross-validation of QTLs linked to fatty acid composition in multiple independent interspecific crosses of oil palm. BMC Genomics 17:289.

Tranbarger TJ, Dussert S, Joët T, Argout X, Summo M, Champion A, Cros D, Omore A, Nouy B, Morcillo F (2011) Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening and functional specialization in lipid and carotenoid metabolism. Plant Physiol. 156:564–584.

van Ooijen JW (2006) JoinMap®4.1, software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, the Netherlands.

van Ooijen JW (2009) MapQTL® 6, software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma B.V., Wageningen, the Netherlands.

Xia Y, Ning Z, Bai G, Li R, Yan G, Siddique KHM, Baum M, Guo P (2012) Allelic variations of a light harvesting chlorophyll a/b-binding protein gene (*Lhcb1*) associated with agronomic traits in Barley. PLoS One 7(5):e37573.

Xie DW, Wang XN, Fu LS, Sun J, Zheng W, Li ZF (2015) Identification of the trehalose-6-phosphate synthase gene family in winter wheat and expression analysis under conditions of freezing stress. J Genet. 94(1):55-65.

Wahid MB, Simeh MA (2009) Issues related to production cost of palm oil in Malaysia. Oil Palm Industry Economic Journal 9(2):2009.

Wei X, Liu K, Zhang Y, Feng Q, Wang L, Zhao Y, Li D, Zhao Q, Zhu X, Zhu X, Li W, Fan D, Gao Y, Lu Y, Zhang X, Tang X, Zhou C, Zhu C, Liu L, Zhong R, Tian Q, Wen Z, Weng Q, Han B, Huang X, Zhang X (2015) Genetic discovery for oil production and quality in sesame. Nat Commun. 6:8609.

Woittiez LS, van Wijk MT, Slingerland M, van Noordwijk M, Giller KE (2017) Yield gaps in oil palm: a quantitative review of contributing factors. Eur J Agron. 83:57–77.

Wong YC, Teh HF, Mebus K, Ooi TEK, Kwong QB, Koo KL, Ong CK, Mayes S, Chew FT, Appleton DR, Kulaveerasingam H (2017) Differential gene expression at different stages of mesocarp development in high- and low-yielding oil palm. BMC Genomics 18(1):470.

Zhang Y, McCartney AJ, Zolov SN, Ferguson CJ, Meisler MH, Sutton MA, Weisman LS (2012) Modulation of synaptic function by VAC14, a protein that regulates the phosphoinositides PI(3,5)P_2 and PI(5)P. EMBO J. 31(16):3442–3456.

Zhu XG, Long SP, Ort DR (2010) Improving photosynthetic efficiency for greater yield. Annu Rev Plant Biol. 61:235-61.

Zulkifli H, Halimah M, Chan KW, Choo YM, Mohd Basri W (2010) Life cycle assessment for oil palm fresh fruit bunch production from continued land use for oil palm planted on mineral soil (Part 2). J. Oil Palm Res. 22:887-894.