Genotyping by Sequencing for the Construction of Oil Palm (*Elaeis guineensis* Jacq.) Genetic Linkage Map and Mapping of Yield Related Quantitative Trait Loci

Fakhrur Razi Mohd Shaha[1,3], Pui Ling Liew[1], Qamaruz Zaman Faridah[3], Rosimah Nulit[3], Barin Jakim[2], Justina Rolland[2], Hui Yee Yong[1], Soo Heong Boon[1]

[1] ACGT Sdn Bhd, Lot L3-I-1, Enterprise 4, Technology Park Malaysia, 57000 Kuala Lumpur, Malaysia.
[2] Department of Agriculture Sabah, Wisma Pertanian Sabah, 88632 Kota Kinabalu, Sabah, Malaysia.
[3] Department of Biology, Faculty of Science, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Corresponding Author:
Fakhrur Razi Mohd-Shaha[1]
ACGT Sdn Bhd, Lot L3-I-1, Enterprise 4, Technology Park Malaysia, 57000 Kuala Lumpur, Malaysia.
Email address: rfakhrur@genting.com

Abstract

Background. Oil palm (*Elaeis guineensis* Jacq.) is one of the major oil-producing crops that provides great nutritional value and high economic returns. Genotyping-by-sequencing (GBS) provides a platform to accelerate the discovery of Single Nucleotide Polymorphisms (SNP) and insertion and deletions (InDel) markers for the construction of a genetic linkage map. A genetic linkage map facilitates the identification of significant DNA regions associated with the trait of interest via Quantitative Trait Loci (QTL) analysis. These genomic approaches will speed up the process in identifying significant DNA markers associated in the Marker Assisted Selection (MAS) for oil palm breeding programs.

Methods. A 112 full sibling mapping population from a cross of a female Deli *dura* and a male Serdang *pisifera* was used in this study. GBS libraries were constructed using the double digestion method with *Hind*III and *Taq*I enzymes. Reduced representation libraries (RRL) of 112 F_1 progeny and their parents were sequenced and the reads were mapped against the *E. guineensis* reference genome. The oil palm genetic linkage map was constructed using significant SNP and InDel markers which facilitated in QTL analysis to discover significant DNA regions associated with the trait of interests, namely fresh fruit bunch (FFB) yield, oil yield (OY), oil to bunch (O/B), oil to dry mesocarp (O/DM), oil to wet mesocarp (O/WM), mesocarp to fruit (M/F), kernel to fruit (K/F), shell to fruit (S/F) and fruit to bunch (F/B).

Results. A total of 2.5 million SNP and 153,547 InDel markers were identified. However, only a subset of 5,278 markers comprising 4,838 SNPs and 440 InDels were informative for the construction of a linkage map. The sixteen linkage groups spanned 2,737.6 cM for the maternal map and 4,571.6 cM for the paternal map with average marker densities of one marker per 2.9 cM and 2.0 cM, respectively. Significant QTLs of two for M/F, four for K/F and one for S/F were successfully detected. These QTLs explained 18.1 – 25.6% of the phenotypic variance. These findings, including a dense linkage map and identification of significant QTL regions associated with K/F, M/F and S/F could be applied in MAS for oil palm breeding programs. The present study highlighted the oil palm genetic linkage map and QTL analysis could fast-track the process in generating new oil palm varieties as compared to the conventional oil palm breeding approach.

Introduction

Oil palm, *Elaeis guineensis* Jacq., is one of the major oil-producing crops. There are two main globally commercialized species of oil palm which are *E. guineensis* and *E. oleifera*. The former originates from West Africa while, the latter is from South or Central America (Corley & Tinker, 2016). In Southeast Asia where the main global producing countries are situated, the planting of *E. guineensis* is favored. There are three different fruit types of oil palm available, *dura*, *pisifera* and *tenera*. The hybrid *Tenera* has higher oil yield compared to its *dura* and *pisifera* parents and resulted in 30% of the yield increment (Corley & Tinker, 2006).

Breeding progress for yield improvement in major crops, however, is still at a low rate which is about 1 – 2% per annum (Soh, 2004; OECD/FAO, 2019). It has been reported an almost stagnant yield of the oil palm performance in Malaysia from 2003 to 2018 which was different for soybean, rapeseed and sunflower which showed yield increment of 6 – 62% for the same period of time (Oil World, 2019). In fact, the best selection of plant for yield improvement should be done at later cycle when unstable genotypes, low heritability and differential plant competitive abilities may be observed at the early cycle (Soh, 2004). Such a breeding pipeline should be challenging to be applied in the perennial crops compared to the annual crop. Oil palm which is a perennial crop has a lifespan of up to 40 years (Kusairi et al., 2019). This could take a long-term investment to generate new oil palm generation with some yield improvement.

Recently, mining DNA markers using Next Generation Sequencing (NGS) approach is more feasible and affordable for many institutes as the sequencing cost reduces. This technology is able to produce genome wide markers. For oil palm, Pootakham et al. (2015) and Bai et al. (2017) reported that the construction of dense oil palm genetic linkage maps using the NGS platform would pave the way to QTL identification. Similar approaches were also found to be useful in various plant species; prominently sugarcane (Balsalobre et al., 2017), soybean (Liu et al., 2017), wheat (Gao et al., 2015), and alfalfa (Zhang et al., 2019). Leveraging on this marker technology in oil palm breeding program could not only reduce time taken to produce a new variety but also could possibly unlock the oil palm yield improvement in the near future.

This study was initiated to determine DNA markers associated with yield production traits, mainly yield and yield components such as fresh fruit bunch (FFB) and oil yield (OY), and bunch quality characters, including oil to bunch ratio (O/B), oil to dry mesocarp ratio (O/DM), oil to wet mesocarp ratio (O/WM), mesocarp to kernel ratio (M/F), kernel to fruit ratio (K/F), shell to fruit ratio (S/F), and fruit to bunch ratio (F/B), in an oil palm cross, Deli *dura* and Serdang *pisifera*. Many oil palm genetic linkage maps have been published (Pootakham et al., 2015; Bai et al., 2017). However, none of the publications had reported this study's cross followed by using NGS technique to discover polymorphic markers. Thus, this study could discover new DNA regions in the oil palm genome associated with oil yield related traits that might be applied in MAS to enhance oil palm yield performances.

Materials & Methods

Plant materials

A total of 112 *tenera* full sibs was generated from a cross of a Deli *dura* and a Serdang *pisifera*. The female parent is a descendent of a cross from Marihat Baris *dura* and Ulu Remis *dura*. Meanwhile, the male parent is a descendant of Dumpy Deli *dura* and Serdang fertile *pisifera*. The palms were managed and maintained by the Department of Agriculture Sabah (DOA, Sabah) at Quoin Hill, Tawau, Sabah, Malaysia following the standard management practices for oil palm plantation.

Commented [MARR1]: Suggested to add degree latitude and altitude

Commented [MARR2]: May add a reference for standards of management practices

DNA extraction

Leaf samples were collected from individual tree and immediately soaked in liquid nitrogen (LN) before being stored in a -80°C freezer until DNA extraction. Hundred milligrams of each individual leaf samples were used as a starting material for DNA extraction. The samples were frozen in the LN before being disrupted with bead using the TissueLyser II. Genomic DNA from the leaf tissue were extracted using DNeasy ® Plant Mini Kit (Qiagen, USA) following the manufacturer's protocol with minor modification on the incubation time.

Genomic DNA concentration and purity were quantified and qualified using Nanodrop-1000 spectrophotometer (Thermo Fisher Scientific, USA). The sample was normalized into 100 ng/µl using Buffer AE (Qiagen, USA). Then, the integrity of the extracted DNA was further examined on 1% TAE agarose gel (Choice Care, Malaysia). A total of 6 µl of sample, comprising 3 µl of 100 ng/µl of DNA samples and 3 µl of 6X DNA loading dye (New England, USA) were loaded on each well and run on the gel electrophoresis (Scie-Plas, England) for an hour at 100 V. The DNA ladder (Thermo Scientific, USA) was added on each row of gel to serve as an indicator for reference bands. After an hour on the electrical current, the agarose gel was then stained in ethidium bromide solution for half an hour before being viewed under an AlphaImager HP DE-500 (Alpha Innotech, Germany). DNA samples of all 112 F_1 and two parents that passed the quality control checks with acceptable ratio of absorbance reading at A260/A280 and A260/A230 1.8 and 2.0 – 2.2 (Thermo Scientific, 2010), respectively, were considered to be an

Commented [MARR3]: F1 is not a segregating population, hence could not be used for gene/QTL mapping, rather advanced to F2 segregating population, Why authors think it is good to use

acceptable DNA purity. These were stored in a 4°C chiller (Thermo Scientific, USA) prior to the sequencing works.

GBS library construction

A total of 114 DNA samples comprising of 112 DNA of progenies and their parents were sent for sequencing to construct the GBS library. The sequencing works and the GBS library construction were outsourced to Genting Laboratory Service Sdn. Bhd. (GLS). In this study, the good quality genomic DNA comprising of 112 F_1 and 2 parents were quantified using the Qubit® dsDNA HS Assay Kit (Life Technologies, USA). GBS libraries were constructed using the double digestion methods with *Hind*III and *Taq*I enzymes (New England Biolabs, Ipswich, MA) according to the standard GBS protocol (Elshire et al. 2011). *Hind*III is a type II endonuclease which recognizes a degenerate 6 bp sequence (5′…AAGCTT…3′) and cleavages between AA of the sequence resulting in 5′ overhang. Meanwhile, *Taq*I recognizes a degenerate 4 bp sequence (5′…TCGA…3′) and cuts between TC of the sequence to create 5′ overhang. A total of 200 ng of genomic DNA of each sample was digested with the enzymes for 2 hours at 75°C. The digested samples were examined by gel electrophoresis (Scie-Plas, England) on a 1% TAE agarose gel before the ligation adapters were added. The ligation products were pooled and those with sizes of 400–600 bp were selected using Pipin Prep (Sage Science, USA). Two different types of adapters, barcode and common adapters were used to enable paired-end and multiplex sequencing on the Illumina HiSeq2000 platform (Illumina Inc., San Diego, CA).

In this experiment, 12 unique adapters with specific barcodes were designed to carry out the multiplexing of 12 samples per Illumina flow cell lane. Library construction was performed using a manufacturer's protocol. Finally, libraries without adapter dimers were retained for DNA sequencing using Illumina HiSeq2000 sequencing platform following standard protocol.

Commented [MARR4]: Different studies have used this technique of sequencing, I suggest to add the reference for said "standards"

SNP and InDel calling

After the sequencing run, the sequencing reads were filtered using a FASTQ file with the quality confidence threshold of ≥ 25. The sequencing reads with the minimum read coverage of five for both parents and progenies were mapped and aligned to the EG5 genome build (Singh et al., 2013) which was obtained from GenomeSawit webpage (http://genomsawit.mpob.gov.my), using Burrows-Wheeler Aligner (Li & Durbin, 2009). Then, SNP and InDel calling were performed using the genome analysis toolkit, GATK ver 3.3.0, with default parameters (McKenna et al., 2010).

SNP and InDel filtering

Using sequence data with a minimum of ~~5x~~ 5X read depth and Q-value of 25 in each sample, markers across the progeny palms with more than 10% missing calls and major allele and genotype frequency more than 95% were excluded. The filtering task was performed using vcftools ver. 0.1.10 (Danecek et al., 2011) before the final versions of the SNP and InDel

Commented [MARR5]: Should it not the "less than"

Commented [MARR6]: What do you mean by "genotype frequency"?
How it measured?

genotyping calls were tabulated into PLINK format (Purcell et al., 2007). In the preparation of the input file for the construction of a genetic linkage map, the allele calls for the parental palms were also investigated. Markers with genotypes that were different from the expected Mendelian segregation patterns between the progenies and the parents for F_1 segregation were filtered out.

Genetic linkage map construction

The genetic linkage maps of the female parent, D200 and the male parent, QP447 were constructed using the significant SNP and InDel markers of 112 F_1. The analysis was performed using the JoinMap 4.1 software as described in the manual (Van Ooijen, 2006). Then, the genetic maps were drawn using the MapChart V2.2 software (Voorrips, 2002). Finally, the constructed linkage maps were used for QTL mapping to identify QTL regions associated with the traits of interest. Further data clean-up was performed prior to the construction of the genetic linkage map as described in the following details. In order to capture the population structure in this study, 2,291 informative SNP markers were used to plot principal component analysis (PCA) using PLINK ver. 1.90 (Weeks, 2010) and visualization was by using Excel to verify the progenies were full-sibs from F_1 mating design.

Genotype data preparation

Markers with major allele and genotype frequencies more than 95% were eliminated using Microsoft Excel 2010. Subsequently, the heterozygous markers were coded based on the JoinMap ver. 4.1 for a cross pollinator (CP) population. Markers coded with <lmxll> and <nnxnp> represents markers with one of the heterozygous parents, and <hkxhk>, <efxeg> and <abxcd> represents markers with both parents heterozygous with the present of two alleles, three alleles and four alleles, respectively.

Formatted: Highlight

The segregation ratios for each marker were examined using Chi-square goodness-of-fit test and the threshold p-value was set to 0.001. Segregation distortion markers with p-value less than 0.001 were excluded. Subsequently, the marker similarities were checked for reducing marker redundancy, as two identical markers will be located at the same position in the linkage map. If a pair of markers with the similarity value exactly equals to 1, one of the markers in pair was being excluded. After removing markers with the filtering parameters as described above, the data set was ready for the construction of genetic linkage maps.

Parental map construction

Maternal and paternal maps were constructed using the filtered markers set. To determine the significant LOD threshold that used to establish linkage groups, the start value was set at 2.0 and end value at 10.0 with a step size of 1.0 (Van Ooijen, 2006). The LOD score was calculated for the recombination frequency and performed by using the software based on G2 statistic, $G2 = 2 \sum O \log (O/E)$ where, O is the observed and E is the expected number of individuals in a cell, log is the natural algorithm and $\sum$ is the sum over all cells. Then, the values were multiplied by

0.217 to obtain the normal LOD score (Van Ooijen, 2006). The grouping parameter was selected at LOD-value more than 4.0 which indicates the likelihood for two loci to be linked are greater than 10000:1. Maximum likelihood (ML) mapping algorithm was selected to compute the mutual distance of loci in the respective groups.

The map order was further improved by eliminating markers exhibiting a Nearest Neighbor Stress value (N.N. stress) more than 3 cM. The N.N. stress is a measurement to indicate if markers are located outside the region they probably belong to (Van Ooijen, 2006). The maps were re-calculated after removing markers with N.N. Stress more than 3 cM in each of the linkage group until the optimal map order was achieved. Finally, the genetic map distances were calculated using the Haldane mapping function. Once the linkage group maps were successfully constructed, both parental maps were drawn and visualized by using MapChart 2.2 (Voorrips, 2002) and the genetic map information was used in the QTL analysis.

Phenotypic traits

In this study, five consecutive year for FFB yield of the individual palm were recorded from 1984 to 1988. The 5-year average FFB yield were calculated to derive as the FFB yield data. At least three bunches per palm were analyzed for bunch components, including O/B, O/DM, O/WM, M/F, K/F, S/F and F/B as per standard industry practice (Blaak et al., 1963). The OY was calculated from the multiplication of the 5-year average FFB yield and average O/B. A total of 9 traits (Supplemental data) were compiled for standard descriptive statistics. Out of a total, a subset of 111 and 97 palms were collected for FFB yield and other traits components, respectively. The phenotype data collections were performed by the Department of Agriculture (DOA), Sabah. During the field data recording, a few palms died but their leaves were already sampled earlier for the subsequent DNA sequencing and molecular analysis. Fourteen samples were excluded for the bunch analysis due to insufficient number of bunches recorded. A Pearson correlation analysis of the nine traits was also performed at the $p = 0.05$ threshold.

Commented [MARR7]: Is it not a too old data

QTL analysis

To declare the presence of a significant QTL, the threshold LOD values were estimated at the genome-wide (GW), and chromosome wide levels (CW). For both cases, the acceptable error level of the permutation test with 1000 iterations was 5%. The Interval Mapping analysis results were used to identify the nearest marker position to a detectable QTL region that surpasses the LOD threshold determined in the permutation test at GW level. Besides these two analyses, the Kruskal-Wallis (KW) test was performed to detect significant marker-trait associations at $P < 0.05$. The proportion of phenotypic variance explained (PVE) by a single QTL was calculated by the square of the partial correlation coefficient (r^2) by using the MapQTL ver. 6 (Van Ooijen, 2009).

Results

SNP and InDel Markers Identification

A total of 2,684,179 raw reads were obtained which showed variances to the reference genome with the minimum base quality confidence threshold of 25 and the minimum read depth coverage of 5 reads in each sample. The total raw reads comprised of 2,530,632 SNPs (94.3%) and 153,547 InDels (5.7%) with the file size of 1.6 Gb and 110 Mb sequence data, respectively.

For SNP marker discovery, 2,472,925 SNPs having missing genotype calls of more than 10% across the progeny samples were excluded. The remaining 57,707 SNPs which represented 2.33% of the total discovered SNP variances were retained. Meanwhile, a total of 3,854 InDels were retained after the elimination process. Of these, 2,154 InDels (55.9%) fitted into the genome build and 1,700 InDels (44.1%) resided on the scaffold.

Among the 57,707 SNPs with minimum genotype calls of 90%, 49,756 SNPs comprising 86.2% were filtered out and 7,951 informative SNPs were left. Meanwhile, a total of 665 InDels was retained for the subsequent analysis. In total, 3,113 SNPs and 225 InDels were filtered out after evaluating by ~~χ~~ χ^2 test for goodness-of-fit to the Mendelian segregation. As a result, 4,838 SNPs and 440 InDels were informative and retained for the construction of genetic linkage maps as shown in Table 1. A total of 2,291 SNP markers which was a subset of 4,838 SNPs was examined by using the PCA. The results presented in Figure 1 showed that all the progenies belonged to the same cross and no illegitimate was found.

The *dura* (maternal) and *pisifera* (paternal) linkage maps

Among 5,278 shortlisted markers, 3,224 SNPs (66.6%) and 300 InDels (68.2%) were heterozygous in the male parent. Meanwhile, 1,605 SNPs (33.2%) and 125 InDels (28.4%) were heterozygous in the female parent. The remaining markers showed the segregation patterns of being heterozygous in both parents with two alleles present, comprising 9 SNPs and 13 InDels. Meanwhile, two InDels are heterozygous with three alleles present as shown in Table 2. Based on the result, the heterozygosity markers were found to be more abundant in the paternal compared to the maternal oil palm.

A total of 5,278 markers, comprising 4,838 SNPs (91.7%) and 440 InDels (8.3%) were used to construct the genetic linkage map for the Deli *dura* x Serdang *pisifera*. Out of the 5,278 markers, a total of 1,754 markers, comprising markers with the segregation patterns of lmxll, hkxhk and efxeg were used for the maternal map construction. Meanwhile, a total of 3,548 markers, comprising markers with the segregation patterns of nnxnp, hkxhk and efxeg were used for the paternal map construction. Of these, a total of 280 and 521 redundant markers having identical genotype calls were excluded for the maternal and paternal maps, respectively. In addition, 236 markers for maternal map and 109 markers for the paternal map were excluded because of unsuccessful assignment to any grouping nodes.

Table 3 shows the markers which were successfully assigned to the groups at LOD score of 5.0 for the genetic linkage map. The major grouping nodes were confidently scored and assigned to 16 chromosomes corresponding to the physical map of the oil palm genome (Singh et al., 2013). Finally, 16 linkage groups for both parental maps were successfully constructed using 1,239 markers for the maternal map (D200) and 2,918 markers for the paternal map (QP447) as shown in Table 4. The linkage groups were named according to the physical map (Singh et al., 2013) for easy traceability. The LOD threshold was set at 5.0 to minimize erroneous grouping assigned for each of the markers. The mapping software we used provided us with LOD tree, starting from LOD1.0 to LOD10.0 scores. Having an advantage of the marker position on the physical map location of EG5 genome build, LOD value at 5.0 was the best synteny to form the grouping with the chromosome from the reference genome.

In addition, markers with N.N. Stress value more than 4.0 were excluded in order to retain only markers which were correctly assigned to the respective groups (Van Ooijen, 2006). The remaining 1,071 and 2,437 markers for the maternal and paternal maps, respectively, were fully informative in constructing the genetic linkage maps in this CP population. The marker distributions are presented in Figure 2 for the maternal map (D200) and Figure 3 for the paternal map (QP447). The maternal (D200, Deli *dura*) and the paternal (QP447, Serdang *pisifera*) maps are 2,737.6 cM and 4,571.6 cM long respectively. The average marker density for the maternal map is 2.9 cM; whereas for the paternal map is 2.0 cM.

The SNP and InDel markers used in the construction of the genetic linkage map were further investigated with the physical map's position. The results in both maternal and paternal maps showed that the genetic linkage map would be able to positioning the marker in the scaffold to the respective linkage map as shown in Figure 4.

Verification of map orders

The collinearity of the markers in the genetic linkage and physical map was investigated as shown in Figure 5. Only the markers that fall on the pseudo-chromosome were used in the investigation as these markers are successfully located in the oil palm reference genome with the physical marker location (Singh et al., 2013). Figure 5 shows that the genetic map constructed in this study was highly correlated to the physical map as indicated by R-squared (r^2) value, especially of linkage groups (LG) LG01, LG02, LG07, LG09, LG10, LG13 and LG14 for both the maternal and paternal maps with r^2 more than 0.9. For maternal maps, all groups showed high collinearity with r^2 value of more than 0.7 between the linkage map and the physical map, except for LG06. Meanwhile, all the groups from the paternal map showed between moderate to high collinearity, except for LG16 which showed very low collinearity with r^2 value of 0.18.

Distribution and correlation of the phenotypic data

Commented [MARR8]: May change to past tense

Commented [MARR9]: Figure 2 and 3 may merge to one as A and B

The coefficient of variations (CV) of the phenotypic data were lesser than 30% for all the analyzed traits with 17.7% for FFB, 20% for OY, 11.4% for O/B, 2.2% for O/DM, 7.2% for O/WM, 5.2% for M/F, 20.0% for K/F, 23.2% for S/F and 4.9% for F/B. The results showed that CV ranged from 2.2% to 23.2% (Table 5). The results of the pairwise Pearson correlation at the significant level at 0.05 was obtained (Table 6). From the result, no significant correlations were observed between FFB with other bunch component traits with the coefficient value of -0.13 for O/B, -0.06 for O/DM, -0.08 for O/WM, -0.03 for M/F, -0.06 for K/F, 0.09 for S/F and -0.17 for F/B. Meanwhile, FFB was strongly correlated with OY with coefficient value of 0.82. However, FFB and OY as well as O/DM and O/WM indicated significant positive correlations.

Identification of significant QTLs on the Deli *dura* (D200) and Serdang *pisifera* (QP447)

A total of eight significant markers associated with M/F, K/F and S/F were identified on LG13 derived from the female parent (D200). Of these, two markers for M/F, five markers for K/F and one marker for S/F were discovered. All associated markers located on chromosome 13 of the physical map (Singh, 2013).

Commented [MARR10]: These are associated markers, not the QTLs. How you define a QTL?

One of the QTLs was located near SNP_sc00239_385519 with phenotypic variation (PVE) of 20.6%. The SNP marker was located at 35.9 cM on the constructed genetic map and located on the scaffold of the physical map. Meanwhile, another QTL detected for M/F was detected on SNP_sc06366_4201 at 60.6 cM. This marker explained 23.9% of PVE and located on the scaffold of the physical map. Another QTL peak was detected on the scatter plot as shown in Figure 6 at 19.5 cM but this marker was excluded due to the insignificant value in the KW test.

Meanwhile, five other QTLs were successfully identified for K/F at GW significant threshold level of 4.1 as shown in Figure 7. These markers gave PVE value ranging from 19.4% to 25.6%. Out of the five markers, two of them were InDels namely as IND_Chr13_2331610 and IND_sc00391_495837. The former was located on the chromosome while, the latter was sited on the scaffold of the physical map. Other than InDels, three other SNP markers associated with K/F namely as SNP_Chr13_2672021, SNP_sc00239_385519 and SNP_sc00391_495539 were identified. Their PVE ranged between 19.5% and 25.6%. Of these, only the first SNP was located on the chromosome and the rest were on scaffold. The result showed that the QTLs for K/F were found on the maternal map.

Besides QTLs for M/F and K/F, a QTL was also detected on SNP_sc06366_4201 for S/F at the GW's significant threshold level of 4.2 and PVE was 18.1% (Figure 8). The SNP marker was located on the scaffold and at 60.6 cM in the genetic map. The QTL detected on S/F overlapped with K/F (Figure 9). Unfortunately, there was no QTL region with LOD value more than GW threshold identified for FFB, OY, O/B, O/DM, O/WM and F/B.

Discussion

In this study, GBS approach was successfully utilized to discover an abundance of SNP and InDel markers in oil palm which has a genome size of 1.8Gb. The high stringency of inclusion markers with 90% call rate was set up to minimize the false positive of the sequence data. Out of the retained SNPs, 44.6% (25,738 SNPs) managed to be mapped onto MPOB's reference genome (EG5 genome build). Meanwhile, the remaining 31,969 SNPs resided outside of the genome build which is known as scaffold. The scaffolds that could not be placed on any chromosomes of EG5 genome would probably result from the incomplete assembly of the reference genome. Furthermore, the mapping population in this study were *tenera* palms which differed from the reference genome that was generated was based on *pisifera* palms. Therefore, this could explain the obtaining result with high heterozygosity rate when using *pisifera*-based reference genome on *tenera* population. Genetically, only 50% of the parent genotype will be inherited in the next generation.

This study applied only a 5-read depth to discover variances across the population. In fact, selection at high read depth could minimize the sequencing error for generating more convincing results as discovered by Pootakham et al. (2015). A few strategies had been taken to reduce errors in the sequencing quality, including further filtering the obtained markers, selecting a good pair of endonucleases and utilizing the oil palm reference genome. In this study, a pair of endonucleases, *Hind*III and *Taq*I successfully produced high-quality variants in oil palm. *Hind*III which consists of only 6 bases long sequence is a frequent cutter RE, and leaving 4 bp overhang which serves efficiently for adapter ligation (Elshire et al., 2011). Meanwhile, the *Taq*I enzyme is a dam methylation-sensitive enzyme which does not cut the methylated base of the recognition sequence which represses gene transcription (Campbell & Kleckner, 1990). To verify the genotyping quality, the parental sources were also sequenced in this study together with their progenies to investigate any segregation distortion of the progenies against their parental genotypes. This strategy had been successfully performed and reported in bovine using a minimum read depth of only three with high genotyping quality stringency (Brouard et al., 2017). PCA result (Fig. 1) showed a tight cluster of all the progenies belonged to the same cross which shows the progenies' genetic origin of biparental progenies.

In oil palm, the homozygosity level is expected to be higher in *dura* than *pisifera* parents. This is due to the genetic background of *dura* being less diverse and variable due to its narrow origin (Corley and Tinker, 2016). In its breeding history, *dura* came from four different sources only that were originally planted in Bogor (Corley and Tinker, 2016). Therefore, such observations were expected and also reported by Ting et al. (2013) in the Deli *dura* which appeared to be more homozygous than the Yangambi *pisifera*. In this study, similar findings were also discovered when the male parents, *pisifera* having a double number of heterozygous SNPs and InDels than the female parent, *dura*. Furthermore, the constructed genetic maps showed the

Commented [MARR11]: Reference?

length of the maternal was half of the paternal map which explained high heterozygosity of the paternal line generating higher map resolution.

The linkage map constructed in this study was able to position markers from scaffolds onto the respective chromosomes in the physical map as shown in Figure 4. Linkage group 4 of the maternal and LG5 of the paternal map, for example, showed the most abundant markers from the scaffolds being able to be located on the chromosome. This might be due to the high recombination rate which occurred. The recombination rate and distribution along chromosomes are varied across different population which potentially influence a population's genetic background and the breeding strategy to be used (Ma et al., 2001; Ong et al., 2019). The linkage map reported in this study would be able to enhance the genome assembly of the oil palm by Singh et al. (2013). The enhancement of genome assembly through genetic linkage map was also reported across kingdoms such as *Daphnia magna* (Dukić et al., 2016), collared flycatcher (Kawakami et al., 2014) and soybean (Lee et al., 2020). Despite a total number of the retained markers of the maternal map in this study being about 50% less than the number of markers on the paternal map, the markers were successfully grouped into 16 linkage groups on the maternal map. Having a larger number of markers in a genetic linkage map could dissect more linkage information in QTL analysis. Thus, increasing marker size should be done in the future to obtain more genetic information of the oil palm. In order to capture more markers in the maternal line, increasing number of samples and using different RE combination should be performed in the future.

The total length of genetic linkage map obtained in this study are longer than those of the previous reports. Pootakham et al. (2015) constructed a genetic linkage map of 1,429.6 cM using 1,085 SNPs on Deli *dura* x Dumpy AVROS *pisifera*. Meanwhile, a genetic linkage map of 1,527.0 cM was constructed by Bai et al. (2017) on Deli *dura* x Ghana *pisifera* using 1,357 SNPs and 123 SSRs. Teh et al. (2020) reported a linkage map of Gunung Melayu *dura* x Gunung Melayu *dura* spanning 1,618.51 cM was constructed using 506 SNPs and 59 SSRs. The number of SNP markers used in this study was doubled those of the previous studies. This might be the reason for the longer map length we generated. The marker order obtained in this study were highly correlated to both the maternal and paternal physical maps, with moderate to high collinearity r^2 values with more than 0.70. In contrast the low collinearity results indicated by r^2 value could be due to insufficient number of the informative markers (Ong et al., 2019). Therefore, obtaining more significant markers could address the problem. In future, with the advance of mapping software to handle larger set of markers, the number of missing calls of 90% used in this study could be lowered down to 80% as used by Bai et al. (2017). This could potentially lead to the discovery of more significant markers in the mapping population. In this study, the stringency was set at the highest possible in order to reduce sequencing error due to the low read-depth.

The marker density reported in this study is higher than the previous findings of 1.03 cM (Bai et al., 2017) and 1.30 cM (Pootakham et al., 2015) obtained using the restriction-site associated DNA sequencing (RAD-seq) approach. Although GBS and RAD-seq are both restriction-site sequencing approaches, GBS targets for low coverage sequencing compared to RAD-seq (Beissinger et al., 2013). High target coverage sequencing may be able to reduce false-positive genotype calls. In addition, the sequencing error prone rate is higher when dealing with short reads. The highest average marker densities for maternal and paternal maps were 5.2 cM on LG15 and 2.8 cM on LG10 with a total number of markers of 15 and 55, respectively. This smaller number of markers anchored on these groups were possibly due to the high average marker density. Therefore, increasing marker size on these groups may increase marker densities leading to improve a map density.

In this study, the significant threshold for both GW and CW were obtained at the p-value at 0.05 as shown in Table 7. However, only GW threshold for each trait showed more stringency than CW was selected in declaring significance QTL in each parental line. The high stringency is necessary in order to avoid identifying the false QTL regions associated with any trait of interests. Obtaining the threshold based on the 1,000 iterations of the phenotype data would be able to increase the stringency of the data analysis. A total of eight significant markers associated with M/F, K/F and S/F were identified on LG13 of the maternal map and located on chromosome 13 of the physical map (Singh, 2013). This could explain that QTL regions found in this study would be specific to Deli *dura* germplasm because no other findings had been reported for these three traits we identified on chromosome 13 (Billotte et al., 2010; Jeenor & Valkaer, 2014; Teh et al., 2016; Ithnin et al., 2017; Teh et al., 2020).

The QTL for K/F found on the maternal map corresponded to that of the previous study by Okwuagwu and Okolo (1992) which stated K/F was generally inherited from the maternal side rather than paternal side. The overlapping QTLs for K/F, M/F and S/F were also discovered by Seng et al. (2016). In addition, the identified QTLs for these three traits in this study clustered and showed almost similar peak patterns in LG13 on the maternal map as illustrated in Figure 9. Based on the scatter plot, the sudden drop after the peak at 60.6 cM might be due to insufficient markers number around that region (Van Ooijen, 2009). Therefore, adding more markers to the region might improve the resolution of QTL mapping. This finding reconfirmed the previous claim that shell thickness is controlled by a single gene with two co-dominant alleles (Beirnaert & Vanderweyen, 1941). The unsuccessful QTL identification for FFB and oil yield-related components might be due to the fact that these traits are complex traits. These traits could be controlled by multiple genes. Using a powerful marker discovery approach such as gene-based marker and genome wide association study (GWAS) could overcome this challenge. Jeenor and Volkaert (2014) discovered QTLs for F/B and oil-to-fruit ratio of M/F and O/WM by using SNPs, SSRs and gene-based markers. Their findings gave showed gene-based marker is a powerful discovery approach for QTL detection.

FFB yield is a low heritability trait as it is controlled by environmental factors such as weather and harvesting practice (Okoye et al., 2009). Prolonged data of FFB recording of five consecutive years could increase the heritability of a trait (Corley & Tinker, 2016); thus, reducing the false positive results caused by the environmental effects. These results failed to discover any significant QTL regions for FFB. This could be due to their heritability levels which were reported as being only 5.02 and 5.12% for *pisifera* and *dura*, respectively (Rafii et al., 2002). These indicated genetic effect on the trait. Besides FFB, Rafii et al. (2002) also reported low values of narrow-sense heritabilities for both *pisifera* and *dura* with values of 15.9% and 1.45% for OY, 2.72% and 10.65% for O/B, 0.00% and 7.52% for O/DM, 0.00% and 18.26% for O/WM, and 9.45% and 4.36% for F/B. Meanwhile, moderate to high values of narrow-sense heritabilities were reported for M/F and S/F with 43.94% and 39.38% for *pisifera*, respectively (Rafii et al., 2002). For K/F, moderate to low values of narrow-sense heritabilities of 22.81% for *pisifera* and 37.26% for *dura* was observed by Rafii et al., (2002). These findings agreed with Menendez and Blaak (1964) which reported moderate to high narrow-sense heritabilities were for M/F (80%), K/F (83%) and S/F (61%), indicating major influence of genetics on these traits. Therefore, heritability could explain the result of this study when the identified significant markers were highly associated with only K/F, M/F and S/F. For example, for F/B which was a low heritable trait with only 28% heritability (Menendez & Blaak, 1964, as cited in Hardon et al., 1985), failing to discover significant marker associated with F/B might be due to the environmental influence on this trait. Fruit formation is highly dependent on the weevil population which may change the pollination rate due to the weather condition and population abundance in different climatic conditions (Prasetyo et al., 2014; The et al., 2020). Therefore, the failure to discover significant QTL for all nine traits, except for K/F, M/F and S/F in this study might be contributed to the traits' heritabilities. QTL regions showed high association with these three traits with the PVE ranging from 18.1 to 25.6% as summarized in Table 7.

In this study, the shortlisted markers influencing M/F, K/F and S/F were further investigated to compare the allele segregation pattern contributing to the traits. This type of analysis gave more confidence for the QTL analysis of these traits. The overall results showed that the shortlisted markers significantly contributed to the variation of the phenotypes. Palms with genotypes AG of SNP_sc06366_4201 had lower mean M/F than those with genotype GG (Figure 10). Meanwhile, the mean M/F shown in oil palms having genotypes AG was higher compared to those with genotype AA for SNP_sc00239_385519. As these markers were located on the scaffold position of the physical map, this finding could not be compared with the other findings.

Commented [MARR12]: Figure 10, 11 and 12 may also merge to one

For K/F traits, palms with homozygous genotypes AA of SNP_Chr13_2672021, SNP_sc00239_385519 and TT of SNP_sc00391_495539 exhibited higher means of K/F than those with heterozygous genotypes (Figure 11). For InDel markers, oil palms with genotypes

C/C showed a higher mean of K/F than those with genotypes CAT/C of IND_sc00391_495837. For S/F, higher mean was observed when oil palms exhibited genotype AG compared to GG for SNP_sc06366_4201 (Figure 12). These genotype effects in the mean phenotypes for genotype segregations are summarized in Table 8. In this study, markers located on chromosome 13 of the physical map were discovered. These identified QTL loci will be valuable for breeding program to select oil palms with trait of interest for Deli *dura or tenera* germplasm.

In this study, LG 13 appeared to be the QTL 'hot-spots' where all markers associated with K/F, M/F and S/F are located around this linkage group. Even though, the study did not manage to unlock the potential QTL for FFB and OY, kernel and mesocarp traits are still major components in determining the yield improvement because palm oil is extracted from these two fruit components. As previously discussed, K/F and M/F are inversely proportionate to shell thickness. Therefore, increasing the kernel and mesocarp ratio and decreasing shell thickness could generate oil palm with maximum oil yield. In most of these cases, the peaks for other than K/F, M/F and S/F traits were still detectable. However, they were at the lower than significant GW and CW thresholds causing less confidence to declare them as significant QTL. It might be possible to detect QTL associated with these traits by increasing sample size as well as provide phenotypic data with more variations. In addition, having more markers dispersed across the genome would also increase the detection power of these complex traits.

Conclusions

In this study, the efficiency of GBS using endonuclease pairs of *Hind*III and *Taq*I, provided a better degree of complexity reduction. A raw read with high level of marker detection was obtained with 2,530,632 SNPs (94.3%) and 153,547 InDels (5.7%) across 112 progenies from a cross of Deli *dura* and Serdang *pisifera*. Of these, a total of 5,278 polymorphic markers were informative and were shortlisted for the construction of a genetic linkage map for this population. The identified markers located along 16-chromosome gave genome wide marker detection across the oil palm genome. Leveraging on a set with a large number of markers, through this study also successfully constructed high-density genetic linkage maps of the maternal (Deli *dura*) and paternal (Serdang *pisifera*) maps of 2,737.6 cM and 4,571.6 cM, respectively. The average marker density for the former was 2.9 cM; whereas for the latter it was 2.0 cM. This study we positioned a large number of markers to reside on the scaffold which is a noticeable improvement of the oil genome assembly.

There was no significant marker discovery to identify QTLs of all nine investigated traits, nonetheless this study successfully discovered markers associated with K/F, S/F and M/F. The inability to identify markers for FFB, OY, O/B, O/DM, O/WM and F/B could be due to low heritabilities among these traits. Two SNP markers were discovered to be associated with M/F, namely SNP_sc00239_385519 and SNP_sc06366_4201. Meanwhile, a total of five significant markers associated with K/F were discovered that is three SNP markers, including

Commented [MARR13]: Conclusion part over lapped a major portion of results and discussion, may required to reduce for specific conclusion as per objective of study.

SNP_sc00239_385519, SNP_Chr13_2672021 and SNP_sc00391_495539, and two InDel markers which were IND_sc00391_495837 and IND_Chr13_2331610. Besides M/F and K/F, a single marker, SNP_sc06366_4201, was found to be associated with S/F. All the QTLs were reported to be located at chromosome 13, these findings could be new discoveries of DNA regions controlling K/F, S/F and M/F. Due to the fact that palm oil is extracted mainly from either oil palm mesocarp or kernel, these newly reported QTL regions associated with K/F, S/F and M/F by this study could be fully utilized in the oil palm breeding program.

The discovered markers in this study, however, require further validations to investigate their robustness. This could be achieved through fine-mapping. In future, generating F_2 population would produce more variation in phenotypic data and allelic segregation when the progenies comprise of all three different fruit types, *dura*, *pisifera* and *tenera*. As a result, the progenies would segregate into 1:2:1 ratio which could pave to better coverage of the construction of genetic linkage map and QTL mapping. In order to increase the power detection in QTL analysis, increasing the sample size should be one of the possible approaches. However, the QTLs detected in this study could be a population-specific to Deli *dura* D200 and Serdang *pisifera* QP447 cross; thus, further verification should be performed before applying them in MAS in other oil palm with different genetic backgrounds.

References

Bai B, Wang L, Lee M, Zhang, Y, Rahmadsyah, Alfiko Y, Ye BQ, Wan ZY, Lim, CH, Suwanto A, Chua N, Yue, GH. 2017. Genome-wide identification of markers for selecting higher oil content in oil palm. *BMC Plant Biology* 17(1):1–11 DOI 10.1186/s12870-017-1045-z.

Balsalobre TWA, da Silva Pereira G, Margarido GRA, Gazaffi R, Barreto FZ, Anoni CO, Cardoso-Silva CB, Costa EA, Mancini MC, Hoffmann HP, de Souza AP, Garcia AAF, Carneiro MS. 2017. GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genomics* 18(1):1–19 DOI: 10.1186/s12864-016-3383-x.

Beirnaert A, Vanderweyen R. 1941. Contribution a l'etudegenetique et biometrique des varieties d'*Elaeis- guineensis* Jacquin. *Publ Inst Nat Etude Agron Congo Belge Ser Sci* 27:1–101.

Beissinger TM, Hirsch CN, Sekhon RS, Foerster JM, Johnson JM, Muttoni G, Vaillancourt B, Buell CR, Kaeppler SM, de Leon N. 2013. Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics* 193(4):1073–1081 DOI: 10.1534/genetics.112.147710

Billotte N, Jourjon MF, Marseillac N, Berger A, Flori A, Asmady H, Adon B, Singh R, Nouy B, Potier F, Cheah SC, Rohde W, Ritter E, Curtois B, Charrier A, Mangin B. 2010. QTL detection by multi-parent linkage mapping in oil palm (*Elaeis guineensis* Jacq.). *Theoretical Applied Genetics* 120(8):1673–1687 DOI: 10.1007/s00122-010-1284-y.

Blaak G, Sparnaaij L, Menedez T. 1963. Breeding and inheritance in the oil palm (*Elaeis guineensis* Jacq.) II. Methods of bunch quality analysis. *JW African Institute for Oil Palm Research* 4:146–155.

Brouard JS, Boyle B, Ibeagha-Awemu EM, Bissonnette N. 2017. Low-depth genotyping-by-sequencing (GBS) in a bovine population: Strategies to maximize the selection of high

quality genotypes and the accuracy of imputation. *BMC Genetics* 18(1):1–14 DOI: 10.1186/s12863-017-0501-y.
Campbell JL, Kleckner N. 1990. *E. coli* oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell* 62(5):967–979 DOI: 10.1016/0092-8674(90)90271-F.
Corley RHV. 2006. Potential yield of oil palm – An update, symposium on yield potential in the oil palm. International Society of Oil Palm Breeders, Phuket.
Corley RHV, Tinker PB. 2016. The Oil Palm. John Wiley & Sons, Oxford.
Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics (Oxford, England)* 27(15):2156–2158. DOI: 10.1093/bioinformatics/btr330.
Dukić M, Berner D, Roesti M, Haag CR, Ebert D. 2016. A high-density genetic map reveals variation in recombination rate across the genome of *Daphnia magna*. *BMC Genetics* 17(1): 1–13 DOI: 10.1186/s12863-016-0445-7.
Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5):e19379 DOI: 10.1371/journal.pone.0019379.
Gao F, Wen W, Liu J, Rasheed A, Yin G, Xia X, Wu X, He Z. 2015. Genome-wide linkage mapping of QTL for yield components, plant height and yield-related physiological traits in the Chinese wheat cross Zhou 8425B/Chinese spring. *Frontiers in Plant Science* 6:1–17 DOI: 10.3389/fpls.2015.01099.
Hardon JJ, Rao V, Rajanaidu N. 1985. A review of oil palm breeding. In Russell GE, ed. *Progress in Plant Breeding*. New York: Butterworth and Co., 139–163.
Ithnin M, Xu Y, Marjuni M, Mohamed-Serdari N, Amiruddin MD, Low E-TL., Tan YC, Yap SJ, Ooi LCL., Nookiah R, Singh R, Xu S. 2017. Multiple locus genome-wide association studies for important economic traits of oil palm. *Tree Genetics and Genomes* 13(5):103 DOI: 10.1007/s11295-017-1185-1.
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 Genetics & Genomes* 10:1–14 DOI: 10.1007/s11295-013-0655-3.
Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, Olason P, Ellegren H. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular Ecology* 23(16):4035–4058. DOI: 10.1111/mec.12810.
Kushairi A, Ong-Abdullah M., Nambiappan B, Hishamuddin E, Bidin MNIZ, Ghazali R, Subramaniam V, Sundram S, Parveez GKA. 2018. Oil palm economic performance in Malaysia and R&D in 2018. *Journal of Oil Palm Research* 31(2):165–194 DOI: 10.21894/jopr.2019.0026.
Lee K, Kim MS, Lee JS, Bae DN, Jeong N, Yang K, Lee JD, Park JH, Moon JK, Jeong SC. 2020. Chromosomal features revealed by comparison of genetic maps of *Glycine max* and *Glycine soja. Genomics* 112(2):1481–1489 DOI: 10.1016/j.ygeno.2019.08.019.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760 DOI: 10.1093/bioinformatics/btp324.
Liu N, Li M, Hu X, Ma Q, Mu Y, Tan Z, Xia Q, Zhang G, Nian H. 2017. Construction of high-density genetic map and QTL mapping of yield-related and two quality traits in soybean

RILs population by RAD-sequencing. *BMC Genomics* 18(1):1–13 DOI: 10.1186/s12864-017-3854-8.

Ma XF, Wanous MK, Houchins K, Milla MAR, Goicoechea PG, Wang Z, Xie M, Gustafson JP. 2001. Molecular linkage mapping in rye (*Secale cereale* L.). *Theoretical and Applied Genetics* 102(4):517–523. DOI: 10.1007/s001220051676.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D., Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20(9):1297–1303 DOI: 10.1101/gr.107524.110.

OECD/FAO. 2019. OECD-FAO Agricultural Outlook 2019–2028, *OECD Publishing*, Paris. DOI: 10.1787/agr_outlook-2019-en.

Oil World. 2019. Oil world annual 2019. ISTA Mielke GmbH, Hamburg.

Okoye MN, Okwuagwu CO, Uguru MI. 2009. Population improvement for fresh fruit bunch yield and yield components in oil palm (*Elaeis guineensis* Jacq.). *American–Eurasian Journal of Scientific Research* 4(2):59–63.

Okwuagwu CO, Okolo EC. 1992. Maternal inheritance of kernel size in the oil palm, *Elaeis guneensis* Jacq. *Elaeis* 4(2).

Ong AL, Teh CK, Kwong QB, Tangaya P, Appleton DR, Massawe F, Mayes S. 2019. Linkage-based genome assembly improvement of oil palm (*Elaeis guineensis*). *Scientific reports* 9(1):6619. DOI: 10.1038/s41598-019-42989-y.

Pootakham W, Jomchai N, Ruang–areeta 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(5–6):288–295. DOI: 10.1016/j.ygeno.2015.02.002.

Prasetyo AE, Purba WO, Susanto A. 2014. *Elaeidobius kamerunicus*: Application of hatch and carry technique for increasing oil palm fruit set. *Journal of Oil Palm Research* 26(3):195–202.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81(3):559–575 DOI: 10.1086/519795.

Rafii MY, Rajanaidu N, Jalani BS, Kusairi A. 2002. Performance and Heritability Estimations on Oil Palm Progenies Tested in Different Environments. *Journal of Oil Palm Research* 14(1): 15–24.

Seng T, Ritter E, Saad SHM, Leao L-J, Singh RSH, Qamaruz–Zaman F, Tan S-G, Alwee, SSRS, Rao V. 2016. QTLs for oil yield components in an elite oil palm (*Elaeis guineensis*) cross. *Euphytica* 212:399–425 DOI: 10.1007/s10681-016-1771-6.

Singh R, Ong-Abdullah M, Low ETL, Manaf MAA, Rosli R, Nookiah R, Ooi LCL, Ooi SE, Chan KL, Halim MA, Azizi N, Nagappan J, Bacher B, Lakey N, Smith SW, He D, Hogan M, Budiman MA, Lee EK, ... , & Sambanthamurthi, R. 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New Worlds. *Nature* 500(7462):335–339 DOI: 10.1038/nature12309.

Soh AC. 2004. Selecting the ideal oil palm: what you see is not necessarily what you get. *Journal of Oil Palm Research* 16(2):121–128.

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. *Scientific Reports* 6:1–7 DOI: 10.1038/srep19075.

Teh CK, Ong, AL, Mayes S, Massawe F, Appleton DR. 2020. Major QTLs for Trunk Height and Correlated Agronomic Traits Provide Insights into Multiple Trait Integration in Oil Palm Breeding. *Genes* 11 DOI: 10.3390/genes11070826.

Ting N, Jansen J, Nagappan J, Ishak Z, Chin C, Tan G, Cheah SC, 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 DOI: 10.1371/journal.pone.0053076.

Van Ooijen JW. 2006. Joinmap® 4 Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands.

Van Ooijen JW. 2009. MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV, Wageningen, Netherlands.

Voorrips RE. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. *The Journal of Heredity* 93(1):77–78.

Weeks JP. 2010. plink: An R package for linking mixed-format tests using IRT-based methods. Journal of Statistical Software, 35(12), 1-33. DOI: 10.18637/jss.v035.i12.

Wu R, Ma CX, Casella G. 2007. Statistical genetics of quantitative traits linkage, maps and QTL. *Statistic for Biology and Health*, Springer, New York, pp. 45–51.

Zhang F, Kang J, Long R, Yu LX, Wang Z, Zhao Z, Zhang T, Yang Q. 2019. High-density linkage map construction and mapping QTL for yield and yield components in autotetraploid alfalfa using RAD-seq. *BMC Plant Biology* 19(1):1–12 DOI: 10.1186/s12870-019-1770-6.