Euphytica

Comparison of quantitative trait loci (QTLs) associated with yield components in two commercial Dura x Pisifera breeding crosses

--Manuscript Draft--

Manuscript Number:	EUPH-D-20-00558R1
Full Title:	Comparison of quantitative trait loci (QTLs) associated with yield components in two commercial Dura x Pisifera breeding crosses
Article Type:	Original Article
Keywords:	oil palm; DxP; Quantitative trait loci; yield components; comparative QTL mapping
Corresponding Author:	Rajinder Singh, Ph.D Malaysian Palm Oil Board Kajang, Selangor MALAYSIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Malaysian Palm Oil Board
Corresponding Author's Secondary Institution:	
First Author:	Siti Hazirah Zolkafli
First Author Secondary Information:	
Order of Authors:	Siti Hazirah Zolkafli
	Ngoot-Chin Ting
	Nik Shazana Nik Mohd Sanusi
	Maizura Ithnin
	Sean Mayes
	Festo Massawe
	Ravigadevi Sambanthamurthi
	Ismanizan Ismail
	Mohd Isa Zainol Abidin
	Siti Habsah Roowi
	Yang Ping Lee
	Nurul Fatiha Farhana Hanafi
	Rajinder Singh
Order of Authors Secondary Information:	
Funding Information:	Malaysian Palm Oil Board (ABBC5-2013) — Dr Ngoot-Chin Ting
Abstract:	The high yielding tenera is the commercial oil palm planting material of choice in Southeast Asia. Notwithstanding this, there is continuous effort to further improve the yield and one way to do this is by addressing the yield components (YCs). Using 4,451 SNP and over 600 SSR markers , this study revealed quantitative trait loci (QTL) associated with YCs in two breeding populations, a Deli dura x Yangambi pisifera (P2) and a Deli dura x AVROS pisifera (KULIM DxP). Thirteen and 29 QTLs were identified in P2 and KULIM DxP, respectively . They were compared to other YC-linked QTLs reported previously for different genetic backgrounds by mapping the QTL-linked markers to the oil palm genome . The comparison revealed four common chromosomes containing QTLs influencing various YCs . The results reveal the possible presence of closely linked loci or pleiotropic genes influencing YCs

	in oil palm. Exploiting the genome data has also facilitated the discovery of candidate genes within or near the QTL regions including those related to glycosylation, fatty acid and oil biosynthesis, and development of flower, seed and fruit .
Response to Reviewers:	Authors responses to Reviewer's comments: Reviewer #1: The manuscript has lot of scientific lacuna. Following major points need to clarified 1. The parents used in the mapping population are have high variation for the traits under study? I don't think the parents have variation for all the traits. In linkage mapping the parents used to generate mapping should vary for the traits. How authors can do GWAS without following the simple logic in linkage mapping studies. It is a very important criteria for mapping QTLs Response: We thank the reviewer for this very important observation. The parent palms used to generate the two mapping families are the maternal Deli dura (in both cases) and AVROS and Yangambi pisifera (paternal parent). It is widely acknowledged by oil palm breeders that these parental palms, namely the Deli dura as well as the AVROS and Yangambi pisifera have significant variation in yield components (YCs). In fact, pisifera is female sterile and does not produce fruits that develop to maturity and hence, has no YCs associated with it. For crossing programmes, the pisifera palm is often selected based on the performance of its siblings (tenera that has fruit bunches), to indicate its yield potential. In a nutshell, the pisifera palms have no YCs directly associated with them, while the dura palms are selected for having favourable YCs. Thus, the pisifera and dura palms do vary in all aspects of YCs such as bunch weight, fruit-to-bunch ratio, kernel size and shell thickness. This has been well documented in literature e.g. Kushairi and Rajanaidu (2000). We have added a sentence in the Materials and methods section on this (under Mapping families, lines 187 – 192, page 5). As such, in this study, the phenotypic variance observed in the 16 YCs (presented in Supplementary Table S1) does reflect segregation of the parental palms, where the intraspecific hybrid populations are known to show hybrid vigor compared to both parents. 2. The English language should be improved Response: We would like to thank Reviewer 1 for pointing this out. We have edited the entire text to further improve the English. Changes made can be viewed in Main text with tracked changes.docx. 3. The brevity of the abstract can be improved Response: Thanks again for pointing this out. We have rewritten the abstract to avoid repetition. 4. Validation of the results must. So, the results must be validated. Response: We agree with Reviewer 1. One of the validation approaches we used in this study was to compare the QTLs identified in the populations utilized to those observed across different genetic backgrounds published previously. Our present results showed a handful of QTLs were common or located closely to those reported previously whereas, most of the QTLs identified were unique to P2 or KULIM DxP (Results and discussion: QTLs from different studies, lines 423 – 489, page 11 – 12 and Figure 4). However, the families used in this study form the important populations and the parental palms will be further improved to develop next generation of oil palm. As the breeding programme takes 10 – 12 years, the QTLs identified will be tested in the next generation as well as to determine stability of the QTL-linked markers in predicting the traits. 5. Many recent references are there. May be included. Response: We have updated the references with more recent publications in 2019 and 2020 throughout the text. 6. What about the replications. Since major phenotypic data involved, replication data is must.

	Response: This is a very important question and the authors agree that the quality of the phenotypic data will have a strong influence on the accuracy of marker-trait association. In oil palm breeding trials including the populations utilized in this study, the yield data (including the measurements for yield related components) is collected over a period of ~5 years or longer, starting at (or after) 6th year after planting in the open-field. The main reason for determining yield after the 6th year is to ensure consistency and reliability of the phenotypic data, as after the 6th year, oil palm fresh fruit production is more stable/consistent compared to the younger (< 6 years) plantings (Harun and Noor 2002, Corley and Tinker 2016). Within the data collection years, a minimum of three ripe bunches (replicates) per palm per year are normally sampled for bunch analysis according to the standard protocol practiced by oil palm breeders which has been well documented (Blaak et al. 1963, Rao et al. 1983, Isa et al. 2011). This is the standardized procedure used by the oil palm plantations/companies in Malaysia (and the rest of the world) for measuring yield and its related parameters. The standardized protocol for determination of yield parameters is also spelled out in the Malaysian national standards (MS157), which determine if parental palms are suitable for commercial seed production. We have described this in Materials and methods: Yield-related phenotypic data (lines 197 – 207, page 5 – 6). We have also highlighted in Results and discussion: Yield components (YCs) and correlations between them (lines 279 – 280, page 7) and Supplementary Table S1, that in this study, the average number (or replicates) of bunches analyzed per palm per year (MBN) was 13 bunches for both P2 and KULIM DxP families, respectively.

Comparison of quantitative trait loci (QTLs) associated with yield components in two commercial *Dura* x *Pisifera* breeding crosses

Siti Hazirah Zolkafli[1]*, Ting Ngoot-Chin[1,3]*, Nik Shazana[1]*, Maizura Ithnin[1], Sean Mayes[2], Festo Massawe[3], Ravigadevi Sambanthamurthi[1], Ismanizan Ismail[4], Mohd Isa Zainol Abidin[5], Siti Habsah Roowi[6], Lee Yang Ping[6], Nurul Fatiha Farhana Hanafi[6], Rajinder Singh[1]**

[1]Advanced Biotechnology and Breeding Centre, Malaysian Palm Oil Board (MPOB), P.O. Box 10620, 50720 Kuala Lumpur, Malaysia

[2] Plant and Crop Sciences, Sutton Bonington Campus, University of Nottingham, Sutton Bonington, Loughborough LE12 5RD, UK

[3] School of Biosciences, University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor, Malaysia

[4] School of Bioscience and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor

[5]Plant Breeding and Services Department, KULIM Plantations Berhad, 81900 Kota Tinggi, Johor, Malaysia

[6] FGV R&D Sdn Bhd,, FGV Innovation Centre (Biotechnology), PT23417, Lengkuk Teknologi, 71760 Bandar Enstek, Negeri Sembilan

*These authors contributed equally to this work.

** Corresponding author email: rajinder@mpob.gov.my

Abstract

The high yielding *tenera* is the commercial oil palm planting material of choice in Southeast Asia. Notwithstanding this, there is continuous effort to further improve the yield and one way to do this is by addressing the yield components (YCs). Using 4,451 SNP and over 600 SSR markers, this study revealed quantitative trait loci (QTL) associated with YCs in two breeding populations, a Deli *dura* x Yangambi *pisifera* (P2) and a Deli *dura* x AVROS *pisifera* (KULIM DxP). Thirteen and 29 QTLs were identified in P2 and KULIM DxP, respectively. They were compared to other YC-linked QTLs reported previously for different genetic backgrounds by mapping the QTL-linked markers to the oil palm genome. The comparison revealed four common chromosomes containing QTLs influencing various YCs. The results reveal the possible presence of closely linked loci or pleiotropic genes influencing YCs in oil palm. Exploiting the genome data has also facilitated the discovery of candidate genes within or near the QTL regions including those related to glycosylation, fatty acid and oil biosynthesis, and development of flower, seed and fruit.

Keywords Oil palm, DxP, Quantitative trait loci, Yield components, comparative QTL mapping

Abbreviations

ABW	: Average bunch weight
Acyl-ACP TE	: Acyl-acyl carrier protein thioesterase
AFLP	: Amplified fragment length polymorphism
AGL8	: Agamous-like MADS-box protein
AP	: Aspartic proteinase
Aux/IAA	: Auxin/indole-3-acetic acid
BN, BNO	: Bunch number
Bwt, BW	: Bunch weight
CHR	: Chromosome
CINV	: Alkaline/neutral invertase
cM	: Centimorgan
CTAB	: Cetyl trimethylammonium bromide
DMWM	: Dry mesocarp/wet mesocarp
EG5	: *E. guineensis* genome build
FELDA	: Federal Land Development Authority Malaysia
FFB	: Fresh fruit bunch(es) weight
FTB, FB	: Fruit/bunch
Fwt	: Fruit weight
GATA	: GATA-binding transcription factor
GA2OX	: Gibberellin 2-beta-dioxygenase
GGPP	: Geranylgeranyl diphosphate chloroplastic
GM	: G model
GPAT	: Glycerol-3-phosphate acyltransferase
GRF	: Growth-regulating factor
GRP	: Glycine-rich protein
GS	: Genomic selection
HXK1	: Hexokinase-1
IM	: Interval mapping
KASII, III	: beta-ketoacyl-ACP synthases II, III
KTB	: Kernel/bunch
KTF, KF	: Kernel/fruit
KW	: Kruskal-Wallis test
KY	: Kernel yield
LG	: Linkage group
LOD	: Logarithm of odds
MAS	: Marker-assisted selection
MBN	: Mean bunch number
MBOAT	: Membrane-bound O-acyltransferase
MFFB	: Mean fresh fruit bunch(es) weight
MFW	: Mean fruit weight

MKW	: Mean kernel weight
ML	: Maximum likelihood
MPW	: Mean mesocarp weight
MSW	: Mean shell weight
MTF	: Mesocarp/fruit
MQM	: Multiple-QTL model
NAC2	: NAC domain-containing protein 2
NDL1	: N-MYC downregulated 1
N.N. Stress	: Nearest neighbor stress
OTB, OB	: Oil/bunch
OTDP, O/DM	: Oil/dry mesocarp
OTF, OF	: Oil/fruit
OTWP	: Oil/wet mesocarp
OY	: Oil yield
PF	: Pulp/fruit
PME	: Pectinesterase
PG	: Polygalacturonase
PO	: Palm oil
QTL	: Quantitative trait loci
RFLP	: Restriction fragment length polymorphism
SAUR	: Small auxin-up RNA-like auxin-responsive protein
SNP	: Single nucleotide polymorphism
SRM1	: Salt-related MYB1
STF	: Shell/fruit
SSR	: Simple sequence repeat
TF	: Transcription factor
TOT	: Total oil
UGT	: UDP-glycosyltransferase
VQ	: Valine-glutamine motif-containing protein
WMF	: Wet mesocarp/fruit
WRI1	: WRINKLED1
YC	: Yield component

Introduction

Oil palm (*Elaeis guineensis* Jacq.) is the most productive oil crop in the world, and is currently grown on some 19 million hectares (ha) of land. This is only about 0.4 % of the total world agricultural land but accounts for almost 40.0 % of the global oils and fats (Kushairi et al. 2018). Comparatively, soybean (*Glycine max*) utilizes 40.1 % of the total agricultural land, followed by cottonseed (13.8 %), rapeseed (13.0 %) and sunflower (10.0 %) (Pirker et al. 2016).

In traditional oil palm breeding, the parental lines are continuously crossed to generate superior progenies, similar to producing hybrids in other crops. The progeny from crosses however, are not automatically acceptable just because they come from good parents. Thus, each cross is progeny tested, and only the confirmed combinations with superior yield are used to produce commercial seeds (Soh et al. 2003). It takes on average 10 – 12 years to develop a new variety, sometimes even up to 20 years for commercial application (Rajanaidu et al. 2000). The question begged is obviously whether the time can be shortened. The main challenge is collection of phenotypic data which is time consuming and labour-intensive, requring years for reliable data compilation. Yield is recorded for at least five years, from six to 10 years after planting in the field and vegetative measurements have to be done several times (Corley and Tinker 2016, Swaray et al. 2020).

In introgressing good trait(s) from Palm A into Palm B, the whole gamut of genes from A, both good and bad, are first incorporated with those from B, and then the undesirable genes weeded out by repeated subsequent self-pollination and selection. It would be faster if only the good gene alleles could be introgressed, but the question has always been how to do so. In recent years, enabling technologies have emerged, such as marker-assisted selection (MAS) and genomic selection (GS). In MAS, markers are used to predict the phenotype, saving time and money in gathering the phenotypic data, as selection can be made even on seedlings when the adult features are yet to show (Collard et al. 2005, Nadeem et al. 2017). More recently, GS, which uses genome-wide markers to estimate the effects of all loci, makes it possible to compute a genomic estimated breeding value for specific traits (Wang et al. 2018) and this approach, is gaining prominence for crop improvement. Both, MAS and GS increase the rate of genetic gain by reducing the necessary selection time for the desired traits. MAS- and GS-based programmes have been applied to improve yield in soybean (Concibido et al. 2003, Sebastian et al. 2010, Jarquín et al. 2014, Fallen et al. 2015, Stewart-Brown et al. 2019) and maize (Yousef and Juvik 2001, Massman et al. 2012, Liu et al. 2015, Pace et al. 2015, Beyene et al. 2016, Wang et al. 2020) and have enhanced disease resistance, yield, plant height and flowering time in wheat and rice (Gupta et al. 2010, Poland et al. 2012, Ragimekula et al. 2013, Spindel et al. 2015, Thavamanikumar et al. 2015, Borrenpohl et al. 2020). These molecular strategies are also applicable to oil palm.

In oil palm, the required tools and techniques for MAS and GS have been developed over the last two decades. For example, DNA-based markers and identification of genomic loci associated with monogenic as well as polygenic traits have been reported (Jack and Mayes 1993, Singh and Cheah 2005). The causal genes regulating the two most important monogenic traits - shell and fruit colour - have been identified and the discoveries translated into commercial diagnostic assays (Singh et al. 2013a, 2014, Ooi et al. 2016). For yield, the QTLs associated with oil yield (OY) and various other yield components (YCs) have been reported by Rance et al. (2001), Billotte et al. (2010), Jeennor and Volkaert (2014), Pootakham et al. (2015), Seng et al. (2016), Teh et al. (2016, 2020) and Bhagya et al. (2020). Many QTLs and markers have been associated with OY and various YCs across different genetic backgrounds, suggesting a complex genetic mechanism determining oil palm yield. The QTLs were uncovered using different marker systems, starting with restriction fragment length polymorphism (RFLP), which were largely replaced by amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR) and more recently, single nucleotide polymorphism (SNP) based markers. RFLP-based markers are codominant, but not popular at present as the technique for generating and identifying informative

RFLP markers is expensive and laborious. To overcome these shortfalls, AFLP markers can be used instead (Singh et al. 1999, Kularatne et al. 2001, Seng et al. 2007) although their dominant nature also posed some limitations in application. Subsequently, SSR markers (also codominant but requiring less DNA and with high reproducibility across laboratories) have become popular in oil palm research (Ting et al. 2010, Zaki et al. 2012, Ting et al. 2013). More recently, SNP markers have gained importance and are preferred due to their wide distribution in the genome, codominant nature and amenability to high throughput analysis (Mishra et al. 2014, Nadeem et al. 2017).

This study constructed a genetic linkage map for a Deli *dura* x AVROS *pisifera* family, a commercial planting material, and updated the Deli *dura* x Yangambi *pisifera* genetic map constructed previously by Ting et al. (2014). Both maps were constructed using the same oil palm customised array containing 4,451 SNP markers and over 600 SSR markers, making the comparison possible. The genetic maps were then used to identify QTLs associated with OY and YCs, and the results were compared to the QTLs published previously for oil palm. Linking and cataloguing the QTLs identified in different studies and by different marker systems is challenging, but has fortunately been made easier with the publication of the oil palm genome build (EG5) (Singh et al. 2013b). It is now possible to compare QTLs from different crosses and publications to determine if they fall within the same chromosomal regions. The ability to identify overlapping QTLs linked to a trait in a similar chromosomal region, adds confidence to the postulation that the genomic region strongly influences the trait concerned. Inclusion of QTL-linked markers consistently associated with a trait in a panel has increased the prediction accuracy of GS models in cattle improvement (Brøndum et al. 2015). More importantly, candidate genes within or near the QTL regions can now be identified for subsequent analysis to determine the actual causative genes for the yield trait(s).

Materials and methods

Mapping families

The first mapping family - P2 (05 Trial 1) - is an advanced breeding cross between an Ulu Remis Deli *dura* (ENL48) and a Yangambi *pisifera* (ML161). The P2 population consisted of 87 F_1 *tenera* palms currently grown at FGV R&D Sdn. Bhd., Kota Gelanggi, Pahang, Malaysia. The second family namely, KULIM DxP consisted of 135 F_1 *tenera* palms, planted at the Tereh Utara plantation of Kulim Plantation Bhd., Johor, Malaysia. The KULIM DxP palms were generated from a cross between an ex-Ulu Remis Deli *dura* (KT 910512/0804) and an AVROS *pisifera* (KT 911101/1203). The maternal *dura* and the paternal *pisifera* palms are known to have contrasting yield parameters, as *pisifera* is female sterile and rarely produces fruit bunches to maturity (Wonkyi-Appiah 1987, Kushairi et al. 1999, Kushairi and Rajanaidu 2000, Swaray et al. 2020). The maternal Deli *dura* palms are known to have higher bunch weight and lower bunch number compared to the paternal *pisifera* and the resulting intraspecific progenies of these two parental palms show hybrid vigour for yield (Gascon and de Berchoux 1964, Durand-Gasselin et al. 2000, Jin et al. 2017, Singh et al. 2020). Leaf materials from all the palms, including the parental ones, were sampled for DNA extraction and marker analysis.

Yield-related phenotypic data

Ripe bunches from both families were analysed for their YCs over a 5-year period according to the standard protocol used by oil palm breeders (Blaak et al. 1963, Rao et al. 1983, Isa et al. 2011). The standard protocol for determining YCs is also cited in the National standards (SIRIM standard MS157), as the recommended methodology to determine the suitable parental palms for commercial seed production. A minimum of three bunches per palm were analysed for 16 YC

parameters: mean bunch number (MBN, no/palm/year), mean fresh fruit bunch weight (MFFB, kg/palm/year), mean fruit weight (MFW, g/fruit), total mesocarp and kernel oils (TOT, ton/ha/year), mesocarp oil yield (OY, ton/ha/year), oil/bunch (OTB, %), oil/wet mesocarp (OTWP, %), oil/dry mesocarp (OTDP, %), mean mesocarp weight (MPW, g/fruit), mesocarp/fruit (MTF, %), kernel yield (KY, ton/ha/year), mean kernel weight (MKW, g/fruit), kernel/fruit (KTF, %), kernel/bunch (KTB, %), mean shell weight (MSW, g/fruit) and shell/fruit (STF, %). The distribution and correlations between the parameters were evaluated using the Kolmogorov-Smirnov normality and Pearson correlation tests in SPSS 16.0.

Genomic DNA extraction

Extraction of genomic DNA from frozen leaves stored at -80 ºC was done using the modified CTAB method (Doyle and Doyle 1990). DNA quality was checked by digestion with *Eco*RI and *Hae*III and electrophoresed on 0.8 % agarose gel (Rahimah et al. 2006). The acceptable purity values were 1.8 – 2.0, as measured by the NanoDrop spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE)

SNP and SSR analyses

SNP genotyping was performed by a service provider using the oil palm customized OPSNP3 Illumina Infinium II Bead-Chip array (Illumina Inc., San Diego, CA) containing 4,451 SNPs. For SSR genotyping, fragment analysis was carried out using the ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA). The SNP and SSR genotyping analyses were as described by Ting et al. (2013, 2014).

Construction of genetic linkage maps

An integrated genetic map of P2 was constructed previously (Ting et al. 2014). Additional SSR markers (sMo, sMh, sMg, _oSSR, sTE, sEg, sOleiSc, p5sc322 and sPSc) from the MPOB SSR database (http://opsri.mpob.gov.my/opsri/welcome.php, Rosli et al. 2019) and Billotte et al. (2010) (mEgCIR) were genotyped and added to the P2 map. The KULIM DxP genetic map was constructed using JoinMap® 4.1 (van Ooijen 2006) as described by Ting et al. (2014). In brief, the independent parental and integrated KULIM DxP genetic maps were constructed simultaneously using the maximum likelihood (ML) mapping algorithm, where each linkage group (LG) was formed from marker pairs with recombination frequency ≤ 0.2. The Haldane mapping function was used to determine the map distance in centimorgan (cM) and markers with nearest neighbour stress (N.N. Stress) value > 4 cM were excluded from the individual parental and integrated maps. Finally, a consistent marker-order was determined by four iterations of map calculation. The integrated genetic linkage maps for P2 and KULIM DxP were labeled as DP and DPK, respectively.

QTLs analysis

QTL analysis was carried out separately for DP and DPK as described by Ting et al. (2016). The default parameters in Interval Mapping (IM), the Multiple-QTL Model (MQM) and Kruskal-Wallis non-parametric ranking tests (KW) were used in MapQTL® 6 (van Ooijen 2009). The 95.0 % genome-wide (GW) and chromosome-wide (CW) LOD significance thresholds for each YC was determined by 1,000 permutations. In addition, G model (GM) (Bernardo 2013) was used to estimate the individual marker effect for the QTLs linked to each YC.

Mapping of QTLs to the oil palm genome build

Markers from the QTL regions were aligned to the oil palm reference genome (EG5) (Singh et al. 2013b) to identify their positions on the corresponding pseudo-chromosome using the program Exonerate (Slater et al. 2005) with its default parameters. Markers with low scores (< 90.0 % matched) and not uniquely mapped were removed. The genomic region corresponding to the QTLs were searched against the predicted oil palm gene model database (Chan et al. 2017) in PalmXplore (http://palmxplore.mpob.gov.my, Sanusi et al. 2018) to identify putative genes and their functions.

Results and discussion

Comparison of DP and DPK genetic maps

A DP (P2) genetic linkage map was constructed previously using AFLP, RFLP, SSR and SNP markers by Ting et al. (2014). A further 240 SSR markers, 151 from MPOB and 89 from Billotte et al. (2010) were added to the current DP map. The updated DP map now contains 1,595 markers across 16 LGs, spanning 1,714.3 cM. Interestingly, a small number of SNP markers (23 SNPM) that failed to map previously, are now in DP although the same mapping parameters were used. They helped bridge some gaps in the original map and further saturate some regions linked to QTLs e.g. OTB on LGDP2 and MSW and MKW on LGDP3. The DPK genetic map (KULIM DxP) had slightly fewer markers, only 57 SSRs and 1,449 SNPs in 16 LGs, covering a total map length of 1,902.3 cM. The average map distance per marker in DPK was 1.3 cM, which as expected was close to the 1.1 cM observed in DP. In DP, the LGs were 66.2 to 193.2 cM, and in DPK, the range observed was 60.7 to 192.4 cM. In both populations, LGDP/DPK5 was the shortest, and the longest was – LGDP/DPK4. There were in total 746 common markers across the 16 LGs, a comparison of which revealed relatively high collinearity of the markers in both maps (Supplementary Figure 1). This is likely due to both populations having female parents of the Deli *dura* pedigree. This suggests that major chromosomal rearrangements have not yet occurred in domestication of the closely related parental lines, as also observed for watermelon (Ren et al. 2014).

Yield components (YCs) and correlations between them

Of the 16 YCs evaluated, 11 were common in both P2 and KULIM DxP families - MBN, MFFB, TOT, OY, KY, OTB, KTF, KTB, MKW, MSW and OTWP. The data for MFW, MPW, STF, OTDP and MTF were only available for KULIM DxP. Almost all the YCs (except MSW) had a continuous and significant normal distribution ($p > 0.05$) in both populations. Normality of YC data was also observed in other oil palm mapping families analysed by Billotte et al. (2010), Seng et al. (2016) and Teh et al. (2020). For P2, YC data were available for 75 of its 87 palms, of which three outliers were removed for MBN based on a Boxplot analysis comparing the observed and expected mean values (5.0 % trimmed mean, SPSS 16.0). For KULIM DxP, the data was available for all of its 135 palms. However, for MSW, MPW and MKW, one, two and four outliers were removed, respectively, following Boxplot analysis.

MBN was determined for an average of 13 bunches/palm for both families, where the range of observations made for individual palms of P2 and KULIM DxP was 6 – 16 and 6 – 19, respectively. As MFFB is influenced by MBN, variation was also observed for it, 72.04 – 210.53 kg/palm/year in the two populations, while OY was 2.53 – 7.92 ton/ha/year. The variations for the different YCs are summarized in Supplementary Table S1. Wide distribution was also observed for fruit components, such as mesocarp measurements and their derivatives (MPW, OTWP, OTDP and MTF)

as well as the kernel- (KY, MKW, KTF and KTB) and shell-related traits (MSW and STF), suggesting that both populations are suitable for QTL analysis for all their YCs measured in this study.

The correlations between the various YCs were consistent in both P2 and KULIM DxP families, with three levels of positive relationships (Figure 1). Strong correlations were observed among MBN, MFFB, TOT and OY with r = 0.63 – 0.99. The second level of positive correlations was among the mesocarp and endocarp components. The mesocarp components (OTB, OTDP and MTF) and MPW had moderate correlation with r = 0.20 – 0.28 for KULIM DxP. Moderate to strong correlations (r = 0.30 – 0.77) were recorded among the endocarp components where KTF, STF, KY and KTB were correlated with MKW and MSW. Finally, the mesocarp and endocarp components contributing to MFW showed strong correlations with MPW (r = 0.87) and moderate correlations with MKW (r = 0.49). A graphical view of the correlations between the YCs is shown in Figure 1, while Supplementary Table S2 demonstrates the relationships of both the direct (those categorized in the same group) and contributory effects (those at different levels) of the YCs to the overall yield in oil palm.

Pearson correlation was negative between some YCs, mainly between the mesocarp (OTB, OTPM, OTDP, MPW and MTF) and endocarp (KTF, STF, KY, KTB, MKW and MSW) components. Among them, negative correlations with r = -0.29 to -0.95 occurred between MTF and the endocarp components in KULIM DxP. This clearly indicates that increasing mesocarp reduces kernel and shell, and *vice versa*, suggesting competition among the sinks for assimilates. Strong correlations among the YCs were also reported by Kushairi et al. (1999), Okwuagwu et al. (2008), Okoye et al. (2009), Seng et al. (2016), Osorio-Guarín et al. (2019) and Teh et al. (2020).

P2: QTLs linked to YCs

In the DP genetic map, 10 QTLs, significant at GW, were associated with various YCs. The traits for the QTLs and their LGs were MBN (LGDP13A), OTB (LGs DP2 and DP12), OTWP (LGDP12), KY (LGDP15), MKW (LGs DP3 and DP10), MSW (LGs DP2, DP3 and DP16) (Table 1). A QTL associated with MBN was identified at map interval 0.0 – 5.0 cM on LGDP13A. An AFLP marker, EAAG/MCTC-125, was closest to the QTL peak detected at LOD 3.9 for MBN. Both the IM and MQM methods revealed that the QTL explained ~20.5 % of the phenotypic variation for MBN, and a negative (paternal) effect (-0.59) was estimated using GM. When associating the MBN phenotype with the observed genotype profiles, without the AFLP locus from the paternal palm (denoted *aa* genotype) (Figure 2 A) MBN increased to 13.30 ± 1.53 bunches from 12.11 ± 1.53 bunches. The limitation of an AFLP marker here was its dominant nature, and it was not clear if the marker concerned, EAAG/MCTC-125, amplified a homozygous or heterozygous DNA segment. Therefore, other flanking markers (LOD 3.6) – namely, sMo00166, sMo00196, SNPM04999 and SNPM03169 - located ~2.6 cM (Figure S1) away were used as proxies, although the phenotypic variation explained was slightly reduced to 18.6

QTLs associated with OTB were found in the 48.0 – 52.0 cM (4.0 cM confidence interval) and 34.3 – 42.8 cM (8.5 cM confidence interval) regions of LGs DP2 and DP12, respectively. Markers from the two intervals showed negative effects from 0.9 – 1.2 % ($p = 0.007$). The closest markers flanking the QTLs were SNPM02314 (LGDP2) and SNPM04433 (LGDP12). Palms categorized in the genotypes *ab* and *aa* had significant differences in OTB ($p \leq 0.05$ T-test, SPSS 16.0). For the marker from the maternal palm - SNPM02314 - the homozygous genotype *aa* showed increased OTB (31.4 ± 2.6 %), ~1.9 % higher than the *ab* genotype (29.6 ± 2.9 %). The genotype of the paternal marker SNPM04433, meanwhile, had an opposite effect on OTB. The *aa* genotype (28.7 ± 2.8 %) had 2.6 % lower OTB than *ab* (31.3 ± 2.6 %) (Figure 2 B).

In addition to OTB, LGDP12 also hosted another GW significant QTL, OTWP, which interval overlapped that for OTB, with the same marker, SNPM04433, located closest to the QTL peaks for both traits. This explained why the two

YCs were strongly correlated (r = 0.81). However, SNPM04433 had a stronger effect of -2.14 (p = 0.000263) for OTWP than for OTB (only -1.20, p = 0.000160). This was likely due to the larger variation for OTWP (3.2 %) in the two genotypes *ab* (54.0 ± 3.5 %) and *aa* (50.9 ± 3.2 %) (Figure 2 C). QTLs associated with kernel and shell components, such as KY, MSW and MKW, were also identified on DP. The markers linked to them explained less of the phenotypic variation than those linked to the QTLs for fruit bunch, whole fruit and mesocarp components (Table 1). This is demonstrated for KY where marker SNPM01951 from the QTL interval 75.0 – 82.1 cM in LGDP15 showed an effect of only 0.07 (p = 0.013897). The average KY for the two genotypes *ab* and *aa* were 0.57 and 0.66 ton/ha/year, respectively, a difference of only 0.09 ton/ha/year (Figure 2 D). Similar observations were made for MSW and MKW where the genotypes *ab* and *aa* of SNPM02999 (LGDP2) and EAGC/MCAA-302 (LGDP10) showed only a small difference of not more than 0.18 g (Figure 2 E and F). Additional QTLs for MSW and MKW were observed in LGs DP3 and DP16 where markers showing clear codominant segregating profiles were detected close to their QTL peaks. The SSR marker mEgCIR3301 had three alleles <*ab*x*ac*>, which segregated into four genotype classes - *ab*, *aa*, *bc* and *ac*. Interestingly, *ab* and *aa* showed lower phenotypic values than *bc* and *ac* (Figure 2 E and F). Another interesting marker was SNPM02704 at the QTL interval associated with MSW on LGDP16. The two parental palms showed the same genotype <*ab*x*ab*> and therefore, their parental effects and contribution to the trait could not be determined *via* GM. However, among the three observed genotypes, *bb* had the lowest MSW (0.79 ± 0.3 g) compared to *aa* (0.96 ± 0.2 g) and *ab* (1.10 ± 0.2 g) (Figure 2 F).

In this study, QTL analysis also revealed a number of putative QTLs for YCs (Table 3). By permutating the entire 16 LGs, these QTLs had LOD scores lower than their GW significance thresholds but higher than their 95.0 % significant thresholds at the chromosome level. In this respect, three CW significant QTLs, termed putative, were identified for MBN, TOT and OY in LGDP2. Interestingly, these three production components are strongly related to each other (r = 0.79 – 0.99). In oil palm, a common QTL interval on the genetic map for related YCs, such as OTB, OTF, STF, KTF and DMWM, was also reported by Jeennor and Volkaert (2014). Similarly, in other crops, clustering of QTLs was reported for fiber quality and various yield traits in cotton (Keerio et al. 2018), weight, length, diameter and peduncle length in tomato (Portis et al. 2014), grain yield, harvesting index and grain weight in rice (Zhu et al. 2017) as well as maturity date, fruit development, fruit structure and the solid soluble content in sweet cherry (Calle and Wünsch 2020). The co-localization of multiple QTLs suggests the presence of closely linked loci or pleiotropic genes (Billotte et al. 2010, Lemmon and Doebley 2014).

KULIM DxP: QTLs linked to YCs

In this population, GW-significant QTLs were identified for nine YCs. The YCs with their associated QTLs and LGs were MBN and MFFB (LGDPK1), OTB (LGDPK8), OY and TOT (LGDPK1 and DPK8), KTB, KTF and MTF (LGDPK14) and STF (LGDPK4). A QTL was associated with MBN at interval 0 – 7.2 cM on LGDPK1, explaining ~15.9 % of the phenotypic variation for the trait. The QTL peak had LOD 5.1 and the closest marker was a SSR, mEgCIR3803, with four genotype classes among the progenies, namely *ac*, *ad*, *bc* and *bd*. Palms with the *ac* and *bc* genotypes had lower MBN of 12.61 ± 0.39 and 12.76 ± 0.38, respectively, than those with the *bd* (13.90 ± 0.33) and *ad* (14.85 ± 0.36) genotypes (Figure 3A). Within the same QTL interval, a smaller region (0.75 – 7.58 cM) was associated with MFFB, where the SNP marker, SNPM01086 was located closest to the QTL peak. In fact, MFFB is one of the most important traits that indicates the productivity of oil palm. This co-segregating <*ab*x*ab*> marker demonstrated that both the *aa* (157.92 ± 3.30 kg) and *ab* (156.56 ± 2.52 kg) genotypes contributed to significantly higher MFFB production than palms with the *bb* genotype (143.02 ± 4.28 kg). On LGDPK1, the slightly extended interval from 0.00 – 7.60 cM also hosted QTLs for OY and TOT,

where the co-segregating marker SNPM01086 was closest to the QTL peak. Higher OY (6.1 ± 0.2 ton/ha/year) and TOT (6.60 ± 0.1 ton/ha/year) were observed for the *aa* than in the *ab* (5.8 ± 0.1 ton/ha/year OY and TOT) and *bb* (5.24 ± 0.18 ton/ha/year OY and 5.76 ± 0.19 ton/ha/year TOT) genotypes.

The QTLs associated with OY and TOT were also identified on LGDPK8 (92.3 – 105.2 cM), with two SNP markers, SNPM02425 and SNPM02400, located closest to the QTL peaks, respectively. The OY-linked SNPM02425 showed a co-segregating profile <*ab*x*ab*>, *i.e.*, palms with the *bb* genotype had higher OY (6.18 ± 0.13 ton/ha/year) than those with *aa* (5.26 ± 0.2 ton/ha/year) and *ab* (5.76 ± 0.1 ton/ha/year). For the QTL associated with TOT, the maternally inherited marker SNPM02400 revealed significantly higher TOT (6.6 ± 0.1 ton/ha/year) for the homozygous genotype (*aa*) than *ab* (5.83 ± 0.1 ton/ha/year). Interestingly, SNPM02400 also pointed to another QTL associated with OTB located at the 101.1 – 103.4 cM interval. The *aa* genotype of this marker was also responsible for higher OTB (28.2 ± 0.2 %) than *ab* (26.8 ± 0.2 %) (Figure 3 C). The three YCs discussed above - OTB, OY and TOT - were significantly related with each another. Therefore, selection for higher OTB will also increase OY and TOT, although these three YC traits are highly influenced by the environment (Soh et al. 2017). The heritability for the three YCs are low, so their breeding improvement will be highly dependent on the environment and general operational management of the trials. If the environment is unfavourable and operational management is poor, the gains from MAS will be tentative.

On LGDPK4, the QTL interval associated with STF was 3.5 – 16.2 cM. It explained 18.6 % of the phenotypic variation in STF and the closest marker to the QTL peak was SNPM00151, which revealed a marker effect of -0.73 % (heterozygous in the paternal palm). The heterozygous (*ab*) group showed a significantly lower STF (10.60 ± 0.19 %) than *aa* (12.06 ± 0.19 %). On DPK14, the QTLs for three highly correlated traits – KTF, KTB and MTF were found within the same map interval (46.9 – 64.8 cM). For KTF and KTB, the markers closest to the QTL peak (54.0 cM) were SNPM04522 and SNPM04938 which mapped on the same locus, indicating they had similar segregation profiles in the mapping family. The phenotypic variation explained by the QTL for KTF (18.8%) was higher than that for KTB (21.1 %). Based on the genotypes of both markers, higher KTF and KTB were observed for the *ab* (7.69 ± 0.13 % KTF and 5.20 ± 0.09 % KTB) than the homozygous *aa* genotype (6.70 ± 0.13 % KTF and 4.46 ± 0.09 % KTB). Within the same map interval, SNPM01100, located closest to the QTL peak (57.4 cM), accounted for 15.6 % of the MTF phenotypic variation. In contrast with KTF and KTB, the *aa* genotype of SNPM01100 showed significantly higher MTF (82.45 ± 0.31 %) than *ab* (80.4 ± 0.28 %). Interestingly, marker SNPM01100 was also significantly associated with KTF and KTB, although it was not closest to their QTL peaks. This indicates that within the QTL interval, this marker influences multiple traits differently depending on its genotype, which is supported by the significant correlations of KTF and KTB with MTF. This suggests that the genes that contribute to increased kernel size (larger KTF and KTB) will reduce mesocarp (MTF). So, selection for MTF will reduce KTF, boosting the mesocarp oil yield (Kushairi et al. 1999).

This study also identified a number of putative QTLs for various YCs on LGs DPK2 (OTDP), DPK4 (MFW, MPW, MSW and KY), DPK5 (MPW, MFW, OTB, OTWP and OTDP), DPK7 (OTWP), DPK8 (MBN), DPK13 (KTF) and DPK14 (MKW, STF and KY). Information on the putative QTLs is summarized in Table 4.

Comparison of common QTLs between P2 and KULIM DxP

This study identified 42 QTLs (21 putative) in P2 and KULIM DxP, distributed across 12 LGs (except 06, 09 and 11). Within each family, a number of the QTLs were co-localized on the same regions, such as on LGs DP01 (MFFB, TOT and OY), DP02 (MBN, OY and TOT) and DP12 (OTB and OTWP) in P2. In KULIM DxP, common QTLs were found on LGs DPK05 (MFW, MPW, OTB, OTDP and OTWP), DPK08 (OTB and TOT) and DPK14 (MTF and STF and; KTB and KTF). However, comparing P2 and KULIM DxP, only a few QTLs were detected in the same LGs for both. The

QTLs on the same LGs were those associated with OTB, MBN, OY, TOT and MSW with OTDP in LG02, and MBN with KTF in LG13. However, the QTLs in the same LGs in P2 and KULIM DxP did not overlap, either in the genetic or physical map.

The lack of common QTLs in both families is likely due to differences in their genetic backgrounds, especially as their *pisifera* parents were different. The *pisifera* of P2 was Yangambi and that of KULIM DxP was AVROS, of quite separate origins. The *pisifera* of KULIM DxP contributed most of the alleles that revealed the GW QTLs for OTB (LGDPK08), KTB, KTF, MTF (LGDPK14) and TOT (LGDPK01). The maternal *dura*, as expected, contributed the alleles for the STF-related QTLs, as the shell trait is maternally inherited. However, in P2, the GW QTLs detected were contributed in equal numbers by both the paternal and maternal parents. Its paternally inherited QTLs were those associated with MBN (LGDP13A), OTB, OTWP (LGDP12) and KY (LGDP15).

QTLs from different studies

The QTLs identified in this study were compared with 144 previously reported for several oil palm crosses (Billotte et al. 2010, Jeenor and Volkaert 2014, Pootakham et al. 2015, Seng et al. 2016, Teh et al. 2016, Bai et al. 2017, Ithnin et al. 2017). Comparison was also made to the QTLs already detected for MFW, MPW, STF, MTF and OTDP in P2 (Ting et al. 2018). The sequences of all the published QTL-linked markers were first mapped to the EG5 genome build to locate them in their pseudo-chromosomes. The results showed that most of the QTLs identified in our study were unique to P2 or KULIM DxP, and have not been reported in other oil palm crosses. Nevertheless, genomic regions on CHR09 and 14 that hosted QTLs in LGs DP7 and DP3 was common to those reported in different genetic backgrounds (discussed below). And, another five QTLs detected in our study are located as close as 2,792 bp to the QTLs reported previously in CHR02, 06 and 15 (Figure 4).

In CHR02, marker SNPM00151, linked to the QTLs for STF and MSW, was located only ~236.4 kb away from the SSR marker sMg00022 that was reported to be associated with KB and KF by Seng et al. (2016). Interestingly, STF is positively related with both KB and KF, which explains why the same genomic region may influence both traits. In the window (2,092,554 – 2,328,938 bp) which encompasses both the QTL intervals, we identified two genes - *acyl-acyl carrier protein thioesterase* (*Acyl-ACP TE*) and *UDP-glycosyltransferase* (*UGT*) involved in the fatty acid (FA) biosynthesis and glycosylation modification, respectively, during fruit development and ripening (Pulsifer et al. 2014, Jing et al. 2011, Sun et al. 2017, Wu et al. 2017, Peng et al. 2020). In the oil palm fruit, the *Acyl-ACP TE* genes such as *FATA* and *FATB* encode protein that hydrolyse the FA acyl chains from ACPs. *FATA* is quite specific for unsaturated acyl ACPs e.g. C18:1-ACP for release of C18:1, and *FATB* for saturated acyl-ACPs, e.g. C16:0-ACP and C14:0-ACP for release of C16:0 and C14:0, respectively thus, playing essential roles in determining the FA composition of palm oil (Sambanthamurthi et al. 2000, Othman et al 2001). *UGT* is involved in anthocyanin glycosylation, the process of accumulating phenolic compounds which are responsible for the customary deep orange-to-red colour of oil palm exocarp. Based on their biological activities, the two genes have a direct impact on the composition of palm oil produced. However, their impact on the shell (and kernel) components, if any, require further investigation.

In CHR06, the marker EAGC/MCAA-302 closest to the QTL peak for MKW - was in the same QTL interval (37,012 – 38,280 kb) associated with PF and aBWT in a multi-parental DxP cross (Billotte et al. 2010). In the interval, a *valine-glutamine motif-containing protein* (*VQ*) was identified at chromosomal position 37,411,925 bp. In many plants, *VQ* has been reported to be responsive to biotic and abiotic stress, including pathogen infection, when interacting with the *WRKY* transcription factor (TF) (Chen et al 2012, Pecher et al. 2014, Liu et al. 2020). The specific interaction between the *VQ* motif FXhVQChTG (pfam05678) containing the gene *IKU1* and a *WRKY*, *MINI3,* reportedly controls endosperm growth

and seed size in *Arabidopsis* (Wang et al. 2010). Therefore, *VQ* is a good candidate gene to investigate for its regulatory effect on kernel and seed in oil palm. Additional analysis of the MKW-QTL region revealed that *VQ* was flanked by *gibberellin 2-beta-dioxygenase* (*GA2OX*) and a *GATA* TF (*GATA*), the putative functions of which are summarized in Table 5. Interestingly, these genes are significantly differentially expressed in low- and high-yielding oil palm (Wong et al. 2017). Furthermore, *GATA* is known to regulate biological functions in various plant organs, including the flower and seed.

In CHR09, the genomic region corresponding to 74.8 – 84.5 cM on LGDP7 of P2 was previously reported to be associated with MTF and STF (Ting et al. 2018). The same genomic region was also associated with QTLs for Bwt and Fwt which were identified in populations derived from Deli, La Me and Yangambi genetic backgrounds (Billotte et al. 2010). Although the correlations between MTF, Bwt and Fwt are not known, it is postulated that increased MTF (or decreased STF) will increase Fwt. A search for genes of interest was performed in the genomic region 8,208,977 to 9,198,501 bp, and two, *C3HC4-type zinc finger* TF (*RING finger*) and a *membrane-bound O-acyltransferase* (*MBOAT*), were shortlisted. In *Nicotiana benthamiana*, *RING finger* is in the chloroplasts and silencing it stops the growth of fruits (Wu et al. 2014). *MBOATs*, such as *diacylglycerol acyltransferase* (*DGAT*) and *lysophospholipid acyltransferase* (*LPLAT*), are involved in catalysing the synthesis and accumulation of lipids in developing seeds, including in the mesocarp of oil palm (Tranbarger et al. 2011, Li et al, 2013, Wang et al. 2012, Jin et al. 2017, Rosli et al. 2018).

The SSR marker mEgCIR3301 mapped to 6,491,270 bp in CHR14 was found associated to MKW in P2 and an DxP mapping family by Seng et al. (2016) as both families shared the same paternal parent (coded ML161). Interestingly, mEgCIR3301 was flanked by a lipid acylation-related gene, *glycerol-3-phosphate acyltransferase* (*GPAT*), at 6,480,850 bp and *WRI1*, at 6,510,932 bp. In many plants, including oil palm, *WRI1* has been reported to regulate genes encoding a number of key enzymes along the FA and triacylglycerol synthesis pathways (Maeo et al. 2009, Bourgis et al. 2011, Tranbarger et al. 2011, Chapman and Ohlrogge 2012, Qu et al. 2012, To et al. 2012, Vanhercke et al. 2013, Tajima et al. 2013, Grimberg et al. 2020, Kong et al. 2020). In fact, a wider group of genes, such as the sugar- and carbohydrate-responsive genes, are also reported to be regulated by *WRI1* (Masaki et al. 2005, Cernac et al. 2006). The storage compounds regulated by these genes eventually will affect development of the seed, embryo and even seedling, suggesting a possible role for *WRI1* in regulating MKW of oil palm.

Another common genomic region is the 19,804 – 20,124 kb interval on CHR15, which was associated with MTF and STF in KULIM DxP. The region was also reportedly linked to other important YCs, such as FFB, Fwt, Bwt and PO (Billotte et al. 2010). We identified a *pectinesterase* (*PME*) and a *small auxin-up RNA-like auxin-responsive protein* (*SAUR*) at 19,788,553 bp (to 19,805,976 bp) and 20,058,133 bp (to 20,059,096 bp), respectively. Both are related to cell metabolism, *PME* degrading pectin and modifying the cell wall in preparation for fruit ripening and softening, and *SAUR* involved in cell division, expansion and differentiation (Markakis et al. 2013, Abu-Sarra and Abu-Goukh 2015, Li et al. 2015, Wen et al. 2020). The presence of these genes in QTL regions influencing various bunch components suggests the importance of genes regulating cell wall development, cell division, expansion and differentiation for the appropriate development of all components in the fruit bunch. Extending the search beyond the common QTL regions (in CHR02, 06, 14 and 15), we also identified a number of genes and TFs involved in the regulation of sugar levels, FA/oil biosynthesis, growth and development of flower, seed and fruit (Table 5), all of which potentially impact development of the bunch components.

Conclusion

This study describes the QTLs associated with yield components in two advanced *dura* x *pisifera* populations. Several common QTLs were identified in both populations. The QTLs linked to MTF and OTWP in P2 and KULIM DxP that influence mesocarp formation, respectively, were located ~22,000 kb apart in CHR09 (LGDP/DPK07). In addition, another similar genomic region (~11,000 kb apart) in CHR08 (LGDP/DPK2) regulates OTB and OTDP in P2 and KULIM DxP, respectively, both directly contributing to oil yield. The QTLs associated with similar yield traits have been published previously in mapping populations of different genetic backgrounds. We collated all the information to identify the QTL regions influencing the related traits reported by the different studies in CHR02, 06, 09, 14 and 15. Search within and near the QTL regions in the different chromosomes revealed 29 candidate genes and transcription factors related to glycosylation, plant growth, development and architecture, glucose and hormone signalling, lipid metabolism, photosynthesis, flowering and fruit ripening. *UGT*, *PG*, *MYB*, *NAC2*, *AUX*/*IAA*, *RING finger* and *PME* are example of genes potentially regulating oil palm fruit formation, thus directly impacting yield. The current genome-based candidate gene approach is useful in identifying interesting genes that can assist in further understanding the genetic control of oil palm yield. In fact, *GATA* gene located within the QTL interval was shown previously to be differentially expressed in high- and low-yielding palms. Further validation of the association of the other candidate genes with the traits concerned can help develop useful tools for marker assisted selection in oil palm breeding. The markers linked to the QTLs could also be candidates for developing an appropriate marker panel for genomic selection in oil palm.

Acknowledgments The authors would like to thank the Director-General of the Malaysian Palm Oil Board (MPOB) for permission to publish this article. The authors thank Dr. Rex Bernardo from University of Minnesota for permission to use the G Model software. The authors also thank Mr. Andy Chang Kwong Choong for his critical comments and edits on the manuscript.

Declarations

Funding This study was funded by the Malaysian Palm Oil Board (MPOB).

Conflict of interest The authors declare that they have no conflict of interest.

Availability of data Summary data and genetic linkage maps used for this study are included in supplementary material.

References

Abu-Sarra AF, Abu-Goukh AA (1992) Changes in pectinesterase, polygalacturonase and cellulose activity during mango fruit ripening. J Hortic Sci Biotechnol 67:561–568. https://doi.org/10.1080/00221589.1992.11516284

Al'bert AV, Kavaĭ-ool UN, Ezhova TA (2014) The gene *NANA* regulates cell proliferation in *Arabidopsis thaliana* shoot apical meristem without interaction with *CLV1*, *CLV2*, *CLV3*. Russ J Dev Biol 45:326–32 https://doi.org/10.1134/S1062360414050026

Alamin M, Zeng D, Qin R et al (2017) Characterization and fine mapping of *SFL1*, a gene controlling screw flag leaf in rice. Plant Mol Biol Rep 35:491–503. https://doi.org/10.1007/s11105-017-1039-x

Behringer C, Bastakis E, Ranftl QL et al (2014) Functional diversification within the family of *B-GATA* transcription factors through the leucine-leucine-methionine domain. Plant Physiol 166:293–305. https://doi.org/10.1104/pp.114.246660

Bai B, Wang L, Lee M et al (2017) Genome-wide identification of markers for selecting higher oil content in oil palm. BMC Plant Biol 17:93. https://doi.org/10.1186/s12870-017-1045-z

Beyene Y, Semagn K, Mugo S et al (2015) Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci 55:154–163. https://doi.org/10.2135/cropsci2014.07/0460

Bhagya HP, Kalyana Babu B, Gangadharappa PM et al (2020) Identification of QTLs in oil palm (*Elaeis guineensis* Jacq.) using SSR markers through association mapping. J Genet 99:19. https://doi.org/10.1007/s12041-020-1180-4

Billotte N, Jourjon MF, Marseillac N. et al (2010) QTL detection by multi-parent linkage mapping in oil palm (*Elaeis guineensis Jacq.*). Theor Appl Genet 120:1673–1687. https://doi.org/10.1007/s00122-010-1284-y
Blaak G, Sparnaaij LD, Menendez T (1963) Methods of bunch analysis. Breeding and inheritance in the oil palm (*Elaeis guineensis Jacq.*) Part II. J W Afr Ins Oil Palm Res 4:146–155.
Borrenpohl D, Huang M, Olson E et al (2020) The value of early-stage phenotyping for wheat breeding in the age of genomic selection. Theor Appl Genet 133:2499–2520. https://doi.org/10.1007/s00122-020-03613-0
Bourgis F, Kilaru A, Cao X et al (2011) Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. Proc Natl Acad Sci USA 108:12527–12532. https://doi.org/10.1073/pnas.1106502108
Brøndum RF, Su G, Janss, L et al (2015) Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. J Dairy Sci 6:4107–4116. https://doi.org/10.3168/jds.2014-9005
Calle A, Wünsch A (2020) Multiple-population QTL mapping of maturity and fruit-quality traits reveals LG4 region as a breeding target in sweet cherry (*Prunus avium* L.). Hortic Res 7:127. https://doi.org/10.1038/s41438-020-00349-2
Cernac A, Andre C, Hoffmann-Benning S et al (2006) *WRI1* is required for seed germination and seedling establishment. Plant Physiol 141:745–757. https://doi.org/10.1104/pp.106.079574
Chan KL, Tatarinova TV, Rosli R. et al (2017) Evidence-based gene models for structural and functional annotations of the oil palm genome. Biol Direct 12:21. https://doi.org/10.1186/s13062-017-0191-4
Chapman KD and Ohlrogge JB (2012) Compartmentation of triacylglycerol accumulation in plants. J Biol Chem 287:2288–2294. https://doi.org/10.1074/jbc.R111.290072
Che R, Tong H, Shi B et al (2015) Control of grain size and rice yield by *GL2*-mediated brassinosteroid responses. Nat Plants 2:15195. https://doi.org/10.1038/nplants.2015.195
Chen L, Song Y, Li S et al (2012) The role of *WRKY* transcription factors in plant abiotic stresses. Biochim Biophys Acta 1819:120–128. https://doi.org/10.1016/j.bbagrm.2011.09.002
Cheng Y, Zhou Y, Yang Y et al (2012) Structural and functional analysis of *VQ* motif-containing proteins in Arabidopsis as interacting proteins of *WRKY* transcription factors. Plant Physiol 159:810–25. https://doi.org/10.1104/pp.112.196816
Chiang YH, Zubo YO, Tapken W et al (2012) Functional characterization of the *GATA* transcription factors *GNC* and *CGA1* reveals their key role in chloroplast development, growth, and division in Arabidopsis. Plant Physiol 160:332-48. https://doi.org/10.1104/pp.112.198705
Collard BCY, Jahufer MZZ, Brouwer JB et al (2015) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169–196. https//doi.org/10.1007/s10681-005-1681-5
Concibido VC, Denny R, Lange D et al (1997) Genome mapping of soybean cyst nematode resistance genes in 'Peking', PI 90763, and PI 88788 using DNA markers. Crop Sci 37:258–264. https://doi.org/10.2135/cropsci1997.0011183X003700010046x
Cooley MB, Yoder JI (1998) Insertional inactivation of the tomato polygalacturonase gene. Plant Mol Biol 38:521–30. https://doi.org/10.1023/A:1006086004262
Corley RHV, Tinker PB (2003) The Oil Palm. Blackwell Science, Oxford
Czolpinska M, Rurek M (2018) Plant glycine-rich proteins in stress response: an emerging, still prospective Story. Front Plant Sci 9:302. https://doi.org/10.3389/fpls.2018.00302
Dai N, Schaffer AA, Petreikov M et al (1995) *Arabidopsis thaliana* hexokinase cDNA isolated by complementation of yeast cells. Plant Physiol 108:879–80. https://doi.org/10.1104/pp.108.2.879
Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. FOCUS 12:13.
Duan P, Ni S, Wang J et al (2015) Regulation of *OsGRF4* by *OsmiR396* controls grain size and yield in rice. Nat Plants 2:15203. https://doi.org/10.1038/nplants.2015.203
Durand-Gasselin T, Kouame KR, Cochard B et al (2000) Diffusion varietale du palmier a huile (*Elaeis guineensis* Jacq.). OCL 7:207–214. https://doi.org/10.1051/ocl.2000.0207
Fallen BD, Allen FL, Kopsell DA et al (2015) Selective genotyping for marker assisted selection strategies for soybean yield improvement. Plant Genetics, Genomics and Biotechnology 2:95–119. https://doi.org/10.5147/pggb.v2i1.156
Ferrándiz C, Gu Q, Martienssen R, Yanofsky MF (2000) Redundant regulation of meristem identity and plant architecture by *FRUITFULL, APETALA1* and *CAULIFLOWER*. Development 127:725–34.
Gascon JP, de Berchoux C (1964) Caracteristiques de la production d'*Elaeis guineensis* (Jacq.) de diverses origines et leurs croisements. Application a la selection du palmier a huile. Oléagineux, 19:75–84.
Granot D, Kelly G, Stein O et al (2014) Substantial roles of hexokinase and fructokinase in the effects of sugars on plant physiology and development. J Exp Bot 65:809–19. https://doi.org/10.1093/jxb/ert400
Grimberg Å, Wilkinson M, Snell P et al (2020) Transitions in wheat endosperm metabolism upon transcriptional induction of oil accumulation by oat endosperm *WRINKLED1*. BMC Plant Biol 20:235. https://doi.org/10.1186/s12870-020-02438-9
Gu Q, Ferrándiz C, Yanofsky MF et al (1998) The *FRUITFULL MADS-box* gene mediates cell differentiation during Arabidopsis fruit development. Development 125:1509–17.
Gupta PK, Langridge P, Mir RR (2010) Marker-assisted wheat breeding: present status and future possibilities. Mol Breeding 26:145–161. https://doi.org/10.1007/s11032-009-9359-7

Hu Y, Chen L, Wang H et al (2013) Arabidopsis transcription factor *WRKY8* functions antagonistically with its interacting partner *VQ9* to modulate salinity stress tolerance. Plant J 74:730–745. https://doi.org/10.1111/tpj.12159
Hudson D, Guevara D, Yaish MW et al (2011) *GNC* and *CGA1* modulate chlorophyll biosynthesis and glutamate synthase (*GLU1/Fd-GOGAT*) expression in Arabidopsis. PLoS One 6:e26765. https://doi.org/10.1371/journal.pone.0026765
Isa Z, Kushairi A, Mohd Din A et al (2011) A critical re-examination of the method of bunch analysis in oil palm breeding – an update. Paper presented at the International Seminar on Breeding for Sustainability in Oil Palm. Kuala Lumpur, Malaysia. pp19–42.
Ithnin M, Xu Y, Marjuni M. et al (2017) Multiple locus genome-wide association studies for important economic traits of oil palm. Tree Genet Genomes 13:103. https://doi.org/10.1007/s11295-017-1185-1
Jack PL, Mayes S (1993) Use of molecular markers for oil palm breeding. II. Use of DNA markers (RFLPs). Oleagineux 48:1–8
Jarquín D, Crossa J, Xavier L et al (2014) A reaction norm model for genomic selection using high dimensional genomic and environmental data. Theor Appl Genet 127:595–607. https://doi.org/10.1007/s00122-013-2243-1
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 Genet Genomes 10:1–14. https://doi.org/10.1007/s11295-013-0655-3
Jin J, Sun Y, Qu J et al (2017) Transcriptome and functional analysis reveals hybrid vigor for oil biosynthesis in oil palm. Sci Rep 7:439. https://doi.org/10.1038/s41598-017-00438-8
Jing F, Cantu DC, Tvaruzkova J et al (2011) Phylogenetic and experimental characterization of an *acyl-ACP thioesterase* family reveals significant diversity in enzymatic specificity and activity. BMC Biochem 12:44. https://doi.org/10.1186/1471-2091-12-44
Jing R, Lin R (2015) The *VQ* motif-containing protein family of plant-specific transcriptional regulators. Plant Physiol 169:371-378. https://doi.org/10.1104/pp.15.00788
Keerio AA, Shen C, Nie Y et al (2018) QTL mapping for fiber quality and yield traits based on introgression lines Derived from *Gossypium hirsutum* × *G tomentosum.* Int J Mol 19:243. https://doi.org/10.3390/ijms19010243
Kong Q, Yang Y, Guo L et al (2020) Molecular basis of plant oil biosynthesis: Insights gained from studying the *WRINKLED1* transcription factor. Front Plant Sci 11:24. https://doi: 10.3389/fpls.2020.00024
Kushairi A, Loh SK, Azman I et al (2018) Oil palm economic performance in Malaysia and R&D progress in 2017. J Oil Palm Res 30:163–195. https://doi.org/10.21894/jopr.2018.0030
Kushairi A, Rajanaidu N (2000) Breeding populations, seed production and nursery management. In: Basiron Y, Jalani BS, Chan KW (ed) Advances in Oil Palm Research. Malaysian Palm Oil Board, Malaysia. pp 39–96
Kushairi A, Rajanaidu N, Jalani BS et al (1999) Agronomic performance and genetic variability of *dura* x *pisifera* progenies. J Oil Palm Res 11:1–24.
Lemmon ZH, Doebley JF (2014) Genetic dissection of a genomic region with pleiotropic effects on domestication traits in maize reveals multiple linked QTL. Genetics 198:345–353. https://doi:10.1534/genetics.114.165845
Li R, Hatanaka T, Yu K et al (2013) Soybean oil biosynthesis: role of *diacylglycerol acyltransferases*. Funct Integr Genomics 3:99-113. https://doi.org/10.1007/s10142-012-0306-z
Li S, Gao F, Xie K et al (2016) The *OsmiR396c-OsGRF4-OsGIF1* regulatory module determines grain size and yield in rice. Plant Biotechnol J 4:2134–2146. https://doi.org/10.1111/pbi.12569
Li ZG, Chen H-W, Li Q-T et al (2015) Three *SAUR* proteins *SAUR76, SAUR77* and *SAUR78* promote plant growth in Arabidopsis. Sci Rep 5:12477. https://doi.org/10.1038/srep12477
Lin SS, Martin R, Mongrand S et al (2008) RING1 E3 ligase localizes to plasma membrane lipid rafts to trigger *FB1*-induced programmed cell death in Arabidopsis 56:550–61. https://doi.org/10.1111/j.1365-313X.2008.03625.x
Liscum E, Reed JW (2002) Genetics of *Aux/IAA* and *ARF* action in plant growth and development. Plant Mol Biol 49:387–400. https://doi.org/10.1023/A:1015255030047
Liu C, Liu H, Zhou C et al (2020) Genome-wide identification of the VQ protein gene family of tobacco (*Nicotiana tabacum* L.) and analysis of its expression in response to phytohormones and abiotic and biotic stresses. Genes 11:284. https://doi.org/10.3390/genes11030284
Liu L, Du Y, Shen X et al (2015) *KRN4* Controls Quantitative Variation in Maize Kernel Row Number. PLoS Genet 11:e1005670. https://doi.org/10.1371/journal.pgen.1005670
Lo SF, Yang SY, Chen KT et al (2008) A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. Plant Cell 20:2603–2618. https://doi.org/10.1105/tpc.108.060913
Luo J, Zhou JJ, Zhang JZ (2018) *Aux/IAA* gene family in plants: molecular structure, regulation, and function. Int J Mol Sci 19:259. https://doi.org/10.3390/ijms19010259
Ma X, Zhang Y, Turečková V et al (2018) The *NAC* transcription factor *SlNAP2* regulates leaf senescence and fruit yield in Tomato. Plant Physiol 177:1286–1302. https://doi.org/10.1104/pp.18.00292
Maeo K, Tokuda T, Ayame A et al (2009) An *AP2-type* transcription factor, *WRINKLED1*, of *Arabidopsis thaliana* binds to the *AW*-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. Plant J 60:476–87. https://doi.org/10.1111/j.1365-313X.2009.03967.x
Mara CD, Irish VF (2008) Two *GATA* transcription factors are downstream effectors of floral homeotic gene action in Arabidopsis. Plant Physiol 147:707–18. https://doi.org /10.1104/pp.107.115634

Markakis MN, Boron AK, Van Loock B et al (2013) Characterization of a small auxin-up RNA (*SAUR*)-like gene involved in *Arabidopsis thaliana* development. PLoS One 8:e82596. https://doi.org/10.1371/journal.pone.0082596
Masaki T, Mitsui N, Tsukagoshi H et al (2005) ACTIVATOR of spomin: *LUC1/WRINKLED1* of *Arabidopsis thaliana* transactivates sugar-inducible promoters. Plant Cell Physiol 46:547–56. https://doi.org/10.1093/pcp/pci072
Mayes S, Jack PL, Corley RHV et al (1997) Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). Genome 40:116–22. https//doi.org/10.1139/g97-016
Mishra KK, Fougat RS, Ballani A et al (2014) Potential and application of molecular markers techniques for plant genome analysis. Int J Pure App Biosci. 2:169–188.
Mudgil Y, Ghawana S, Jones AM (2013) N-MYC down-regulated-like proteins regulate meristem initiation by modulating auxin transport and *MAX2* expression. PLoS One 8:e77863. https://doi.org/10.1371/journal.pone.0077863
Mudgil Y, Uhrig JF, Zhou J et al (2009) Arabidopsis *N-MYC DOWNREGULATED-LIKE1*, a positive regulator of auxin transport in a G protein-mediated pathway. Plant Cell 21:3591–609. https://doi.org/10.1105/tpc.109.065557
Muhammad Azhar Nadeem, Muhammad Amjad Nawaz, Muhammad Qasim Shahid et al (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol Biotechnol Equip 32:261–285. https://doi.org/10.1080/13102818.2017.1400401
Okada K, Saito T, Nakagawa T et al (2000) Five *geranylgeranyl diphosphate synthases* expressed in different organs are localized into three subcellular compartments in Arabidopsis. Plant Physiol 122:1045–1056. https://doi.org/10.1104/pp.122.4.1045
Okoye MN, Okwuagwu CO, Uguru MI (2009) Population improvement for fresh fruit bunch yield and yield components in oil palm (*Elaeis guineensis* Jacq.) AEJSR 4:59–63.
Okwuagwu CO, Okoye MN, Okolo, EC et al (2008) Genetic variability of fresh fruit bunch yield in Deli / *dura* x *tenera* breeding populations of oil palm (*Elaeis guineensis* Jacq.) in Nigeria. J Trop Agric 46:40–45.
Ooi Leslie C-L, Low Eng-Ti L, Abdullah Meilina O et al (2016) Non-*tenera* contamination and the economic impact of *SHELL* genetic testing in the Malaysian independent oil palm industry. Front Plant Sci 7:771. https://doi.org/10.3389/fpls.2016.00771
Osorio-Guarín JA, Garzón-Martínez GA, Delgadillo-Duran P et al (2019) Genome-wide association study (GWAS) for morphological and yield-related traits in an oil palm hybrid (Elaeis oleifera x Elaeis guineensis) population. BMC Plant Biol 19:533. https://doi.org/10.1186/s12870-019-2153-8
Osteryoung KW, Toenjes K, Hall B et al (1990) Analysis of tomato *polygalacturonase* expression in transgenic tobacco. Plant Cell 2:1239-48. https://doi.org/10.1105/tpc.2.12.1239
Othman A, Lazarus CM, Fraser T et al (2001) Cloning of a palmitoyl-acyl carrier protein thioesterase from oil palm. Biochem Soc Trans 28: 619–22. https://doi.org/10.1042/BST0280619
Pace J, Gardner C, Romay C et al (2015) Genome-wide association analysis of seedling root development in maize (*Zea mays L.*) BMC Genomics 16:47. https://doi.org/10.1186/s12864-015-1226-9
Paparelli E, Gonzali S, Parlanti S et al (2012) Misexpression of a *chloroplast aspartyl protease* leads to severe growth defects and alters carbohydrate metabolism in Arabidopsis. Plant Physiol 160:1237–50. https://doi.org/10.1104/pp.112.204016
Pecher P, Eschen-Lippold L, Herklotz S (2014) The *Arabidopsis thaliana* mitogen-activated protein kinases *MPK3* and *MPK6* target a subclass of '*VQ*-motif'-containing proteins to regulate immune responses. New Phytol 203:592–606. https://doi.org/10.1111/nph.12817.
Peng Z, Zhang H, Tian H et al (2020) The phylogeny and functional characterization of peanut acyl-acp thioesterases. J Plant Growth Regul 39:1381–1392. https://doi.org/10.1007/s00344-020-10073-9
Pirker J, Mosnier A, Kraxner F et al (2016) What are the limits to oil palm expansion? Global and Environ Chang. 40:73–81. https://doi.org/10.1016/j.gloenvcha.2016.06.007
Poland JA, Brown PJ, Sorrells ME et al (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253. https://doi.org/10.1371/journal.pone.0032253
Pootakham W, Jomchai N, Ruang-areerate P et al (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. https://doi.org/10.1016/j.ygeno.2015.02.002
Portis E, Barchi L, Toppino L et al (2014) QTL mapping in eggplant reveals clusters of yield-related loci and orthology with the tomato genome. PLoS One 9:e89499. https://doi.org/10.1371/journal.pone.0089499
Pulsifer IP, Lowe C, Narayaran SA et al (2014) Acyl-lipid thioesterase1-4 from *Arabidopsis thaliana* form a novel family of *fatty acyl-acyl carrier protein thioesterases* with divergent expression patterns and substrate specificities. Plant Mol Biol 84:549–63. https://doi.org/10.1007/s11103-013-0151-z
Qu J, Ye J, Geng YF et al (2012) Dissecting functions of *KATANIN* and *WRINKLED1* in cotton fiber development by virus-induced gene silencing. Plant Physiol 160:738–748. https://doi.org/10.1104/pp.112.198564
Ragimekula N, Varadarajula NN, Mallapuram SP et al (2013) Marker assisted selection in disease resistance breeding. J Plant Breed Genet 1:90–109.
Rahimah, AB, Cheah SC, Rajinder S (2006) Freeze-drying of oil palm (*Elaeis guineensis*) leaf and its effect on the quality of extractable DNA. J Oil Palm Res 18:296–304.

Rajanaidu N, Kushairi A, Rafii M et al (2000) Oil palm breeding and genetic resources. In: Basiron Y, Jalani BS, Chan KW (ed) Advances in Oil Palm Research. Malaysian Palm Oil Board, Malaysia. pp 171–227.

Rance K A, Mayes S, Price Z et al (2001) Quantitative trait loci for yield components in oil palm (*Elaeis guineensis* Jacq.). Theor App Genet 103:1302–1310. https://doi.org/10.1007/s122-001-8204-z

Ren Y, McGregor C, Zhang Y et al (2014) An integrated genetic map based on four mapping populations and quantitative trait loci associated with economically important traits in watermelon (*Citrullus lanatus*). BMC plant biology 14:33. https://doi.org/10.1186/1471-2229-14-33

Richter R, Behringer C, Müller IK et al (2010) The *GATA*-type transcription factors *GNC* and *GNL/CGA1* repress gibberellin signaling downstream from *DELLA* proteins and *PHYTOCHROME-INTERACTING FACTORS*. Genes Dev 24:2093–104. https://doi.org/10.1101/gad.594910

Roongsattham P, Morcillo F, Jantasuriyarat C (2012) Temporal and spatial expression of *polygalacturonase* gene family members reveals divergent regulation during fleshy fruit ripening and abscission in the monocot species oil palm. BMC Plant Biol 12:150. https://doi.org/10.1186/1471-2229-12-150

Rosli R, Amiruddin N, Ab Halim MA et al (2018) Comparative genomic and transcriptomic analysis of selected *fatty acid biosynthesis* genes and *CNL* disease resistance genes in oil palm. PLoS ONE 13:e0194792. https://doi.org/10.1371/journal.pone.0194792

Sambanthamurthi R, Sundram K, Tan Y (2000) Chemistry and biochemistry of palm oil. Progress in Lipid Research 39: 507–58. https://doi.org/10.1016/S0163-7827(00)00015-1

Sanusi NS, Rosli R, Ab Halim MA et al (2018) PalmXplore: oil palm gene database. Database https://doi.org/10.1093/database/bay095

Sebastian SA, Streit LG, Stephens PA et al (2010) Context-specific marker-assisted selection for improved grain yield in elite soybean populations. Crop Sci 50:1196–1206. https://doi.org/10.2135/cropsci2009.02.0078

Seng T-Y, Faridah QZ, Ho CL et al (2007) Flanking AFLP markers for the *virescens* trait in oil palm. J Oil Palm Res 19:381–392.

Seng T-Y, Ritter E, Mohamed Saad (2016) QTLs for oil yield components in an elite oil palm (*Elaeis guineensis*) cross. Euphytica 212:399–425. https://doi.org/10.1007/s10681-016-1771-6

Shan C, Mei Z, Duan J et al (2014) *OsGA2ox5*, a *gibberellin* metabolism enzyme, is involved in plant growth, the root gravity response and salt stress. PLoS One 9:e87110. https://doi.org/10.1371/journal.pone.0087110

Shockey J, Regmi A, Cotton K et al (2016) Identification of Arabidopsis *GPAT9* (At5g60620) as an essential gene involved in *triacylglycerol* biosynthesis. Plant Physiol 170:163–79. https://doi.org/10.1104/pp.15.01563

Shu K, Yang W (2017) *E3 Ubiquitin ligases*: ubiquitous actors in plant development and abiotic stress responses. Plant Cell Physiol 58:1461–1476. https://doi.org/10.1093/pcp/pcx071

Singer SD, Chen G, Mietkiewska E et al (2016) Arabidopsis *GPAT9* contributes to synthesis of intracellular glycerolipids but not surface lipids. J Exp Bot 67:4627-38. https://doi.org/10.1093/jxb/erw242

Singh R, Cheah SC (1999) Analysis of the inheritance of AFLP markers in an interspecific cross of oil palm using the pseudo-testcross strategy. J Oil Palm Res Special Issue:64–73

Singh R, Cheah SC (2005) Potential Application of Marker Assisted Selection (MAS) in Oil Palm. Oil Palm Bulletin 51: 1–9.

Singh R, Low ETL, Ooi LCL (2013a) The oil palm *SHELL* gene controls oil yield and encodes a homologue of *SEEDSTICK*. Nature 500:340-344. https://doi.org/10.1038/nature12356

Singh R, Ong-Abdullah M, Low ETL et al (2013b) Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. Nature 500:335–339. https://doi.org/10.1038/nature12309

Singh R, Low ETL, Ooi L C-L et al (2014) The oil palm *VIRESCENS* gene controls fruit colour and encodes a *R2R3-MYB*. Nature Commun 5:4106. https://doi.org/10.1038/ncomms5106

Singh R, Low ETL, Ooi L C-L et al (2020) Variation for heterodimerization and nuclear localization among known and novel oil palm SHELL alleles. New Phytologist 226:426–440. https://doi.org/10.1111/nph.16387

Soh AC, Mayes S, Roberts JA (2017) Oil palm breeding: genetics and genomics. CRC Press, US

Spindel J, Begum H, Akdemir D et al (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet 11:e1005350. https://doi.org/10.1371/journal.pgen.1004982

Stewart-Brown BB, Song Q, Vaughn JN et al (2019) Genomic selection for yield and seed composition traits within an applied soybean breeding program. G3(Bethesda) 9:2253–2265. https://doi.org/10.1534/g3.118.200917

Sun Y, Ji K, Liang B et al (2017) Suppressing *ABA uridine diphosphate glucosyltransferase* (SlUGT75C1) alters fruit ripening and the stress response in tomato. Plant J 91:574–589. https://doi.org/10.1111/tpj.13588

Swaray S, Amiruddin MD, Rafii MY et al (2020) Influence of parental *dura* and *pisifera* genetic origins on oil palm fruit set ratio and yield components in their D x P progenies. Agronomy 10:1793. https://doi.org/10.3390/agronomy10111793

Tajima D, Kaneko A, Sakamoto M et al (2013) Wrinkled 1 (*WRI1*) Homologs, AP2-type transcription factors involving master regulation of seed storage oil synthesis in castor bean (*Ricinus communis L.*). Am J Plant Sci 4:333–339. https://doi.org/10.4236/ajps.2013.42044

Teh C-K, Ong A-L, Kwong Q-B et al (2016) Genome-wide association study identifies three key loci for high mesocarp oil content in perennial crop oil palm. Sci Rep 6:19075. https://doi.org/10.1038/srep19075
Teh C-K, Ong A-L, Mayes S et al (2020) Major QTLs for trunk height and correlated agronomic traits provide insights into multiple trait integration in oil palm breeding. Genes 11:826. https://doi.org/10.3390/genes11070826
Ting N-C, Jansen J, Mayes S et al (2014) High density SNP and SSR-based genetic maps of two independent oil palm hybrids. BMC Genomics 15:309. https://doi.org/10.1186/1471-2164-15-309
Ting N-C, Jansen J, Nagappan J et al (2013) Identification of QTLs associated with callogenesis and embryogenesis in oil palm using genetic linkage maps improved with SSR markers. PLoS One 8:e53076. https://doi.org/10.1371/journal.pone.0053076
Ting N-C, Mayes S, Massawe F et al (2018) Putative regulatory candidate genes for QTL linked to fruit traits in oil palm (*Elaeis guineensis* Jacq.). Euphytica 214:214. https://doi.org/10.1007/s10681-018-2296-y
Ting N-C, Zulkifli Y, Kamaruddin K et al (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. https://doi.org/10.1186/s12864-016-2607-4
To A, Joubès J, Barthole G et al (2012) *WRINKLED* transcription factors orchestrate tissue specific regulation of fatty acid biosynthesis in Arabidopsis. Plant Cell 24:5007–5023. https://doi.org/10.1105/tpc.112.106120
Tranbarger TJ, Dussert S, Joët T et al (2011) Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. Plant Physiol 156:564–584. https//doi.org/10.1104/pp.111.175141
Van Ooijen, J.W. (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
Vanhercke T, El Tahchy A, Shrestha P et al (2013) Synergistic effect of *WRI1* and *DGAT1* coexpression on triacylglycerol biosynthesis in plants. FEBS Lett 587:364–9. https://doi.org/10.1016/j.febslet.2012.12.018
Wang A, Garcia D, Zhang H (2010) The *VQ* motif protein *IKU1* regulates endosperm growth and seed size in Arabidopsis. Plant J 63:670–9. https//doi.org/10.1111/j.1365-313X.2010.04271x
Wang B, Lin Z, Li X et al (2020) Genome-wide selection and genetic improvement during modern maize breeding. Nat Genet 52:565–571. https://doi.org/10.1038/s41588-020-0616-3
Wang L, Shen W, Kazachkov M et al (2012) Metabolic interactions between the Lands cycle and the Kennedy pathway of glycerolipid synthesis in Arabidopsis developing seeds. Plant Cell 24:4652–69. https://doi.org/10.1105/tpc.112.104604
Wang T, Tohge T, Ivakov A et al (2015) Salt-related *MYB1* coordinates abscisic acid biosynthesis and signaling during salt stress in Arabidopsis 169:1027–41. https://doi.org/10.1104/pp.15.00962
Wang X, Yang X, Zhongli H et al (2018) Genomic selection methods for crop improvement: current status and prospects. Crop J 6:330–340. https://doi.org/10.1016/j.cj.2018.03.001
Watson CF, Zheng L, DellaPenna D (1994) Reduction of tomato *polygalacturonase beta* subunit expression affects pectin solubilization and degradation during fruit ripening. Plant Cell 6:1623–34. https://doi.org/10.1105/tpc.6.11.1623
Wen B, Zhang F, Wu X et al (2020) Characterization of the tomato (*Solanum lycopersicum*) pectin methylesterases: evolution, activity of isoforms and expression during fruit ripening. Front Plant Sci 11:238. https://doi:10.3389/fpls.2020.00238
Wong YC, The HF, Mebus K et al (2017) Differential gene expression at different stages of mesocarp development in high- and low-yielding oil palm. BMC Genomics 18. https://doi.org/10.1186/s12864-017-3855-7
Wonkyi-Appiah JB (1987) Genetic control of fertility in the oil palm (*Elaeis guineensis* Jacq.). Euphytica, 36:505–511. https://doi.org/10.1007/BF00041494
Wu B, Gao L, Gao J et al (2017) Genome-wide identification, expression patterns, and functional analysis of *UDP glycosyltransferase* family in peach (*Prunus persica* L. Batsch) Front Plant Sci 8:389. https://doi.org/10.3389/fpls.2017.00389
Wu W, Cheng Z, Liu M et al (2014) *C3HC4-type RING* finger protein *NbZFP1* is involved in growth and fruit development in *Nicotiana benthamiana*. PLoS ONE 9:e99352. https://doi.org/10.1371/journal.pone.0099352
Xiang L, Le Roy K, Bolouri-Moghaddam MR et al (2011) Exploring the neutral invertase-oxidative stress defence connection in *Arabidopsis thaliana*. J Exp Bot 62:3849–62. https://doi.org/10.1093/jxb/err069
Yousef G, Juvik J (2001) Comparison of phenotypic and marker-assisted selection for quantitative traits in sweet corn. Crop Sci 41:645–655. https://doi.org/10.2135/cropsci2001.413645x
Zaki NM, Singh R, Rosli R et al (2012) *Elaeis oleifera* genomic-SSR markers: exploitation in oil palm germplasm diversity and cross-amplification in *arecaceae*. Int J Mol Sci 13:4069–88. https://doi.org10.3390/ijms13044069
Zhai Z, Liu H, Shanklin J (2017) Phosphorylation of *WRINKLED1* by *KIN10* results in its proteasomal degradation, providing a link between energy homeostasis and lipid biosynthesis. Plant Cell 29:871–889. https://doi.org/10.1105/tpc.17.00019
Zhu M, Liu D, Liu W et al (2017) QTL mapping using an ultra-high- density SNP map reveals a major locus for grain yield in an elite rice restorer R998. Sci Rep 7:10914. https://doi.org/10.1038/s41598-017-10666-7

848 Zulkifli Y, Rajinder S, Mohd Din A et al (2014) Inheritance of SSR of SSR and SNP loci in an oil palm interspecific
849 hybrid backcross (BC2) population. J Oil Palm Res 206:203–213.
850

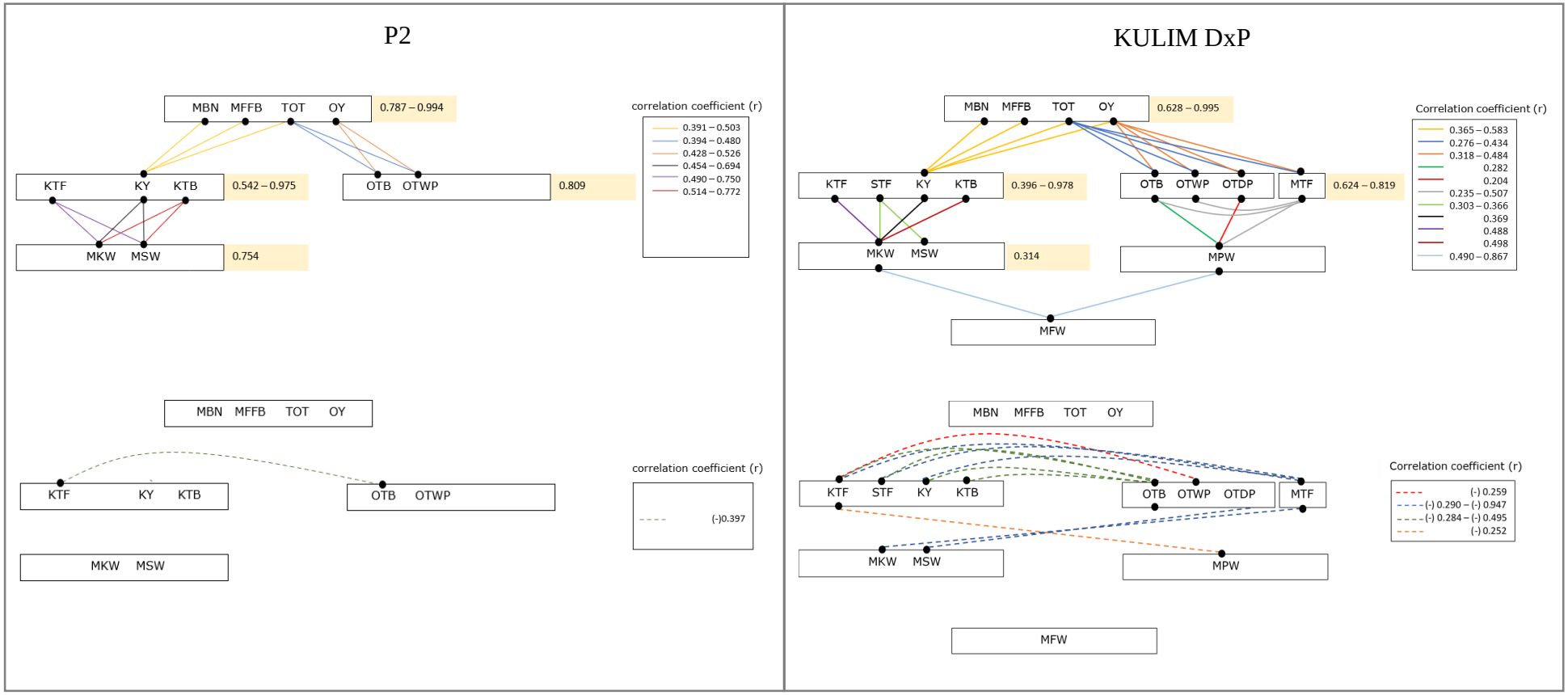


Fig. 1: Significant ($p \leq 0.01$, 2-tailed) positive (solid lines) and negative (dotted lines) correlations between YCs in P2 and KULIM DxP families.

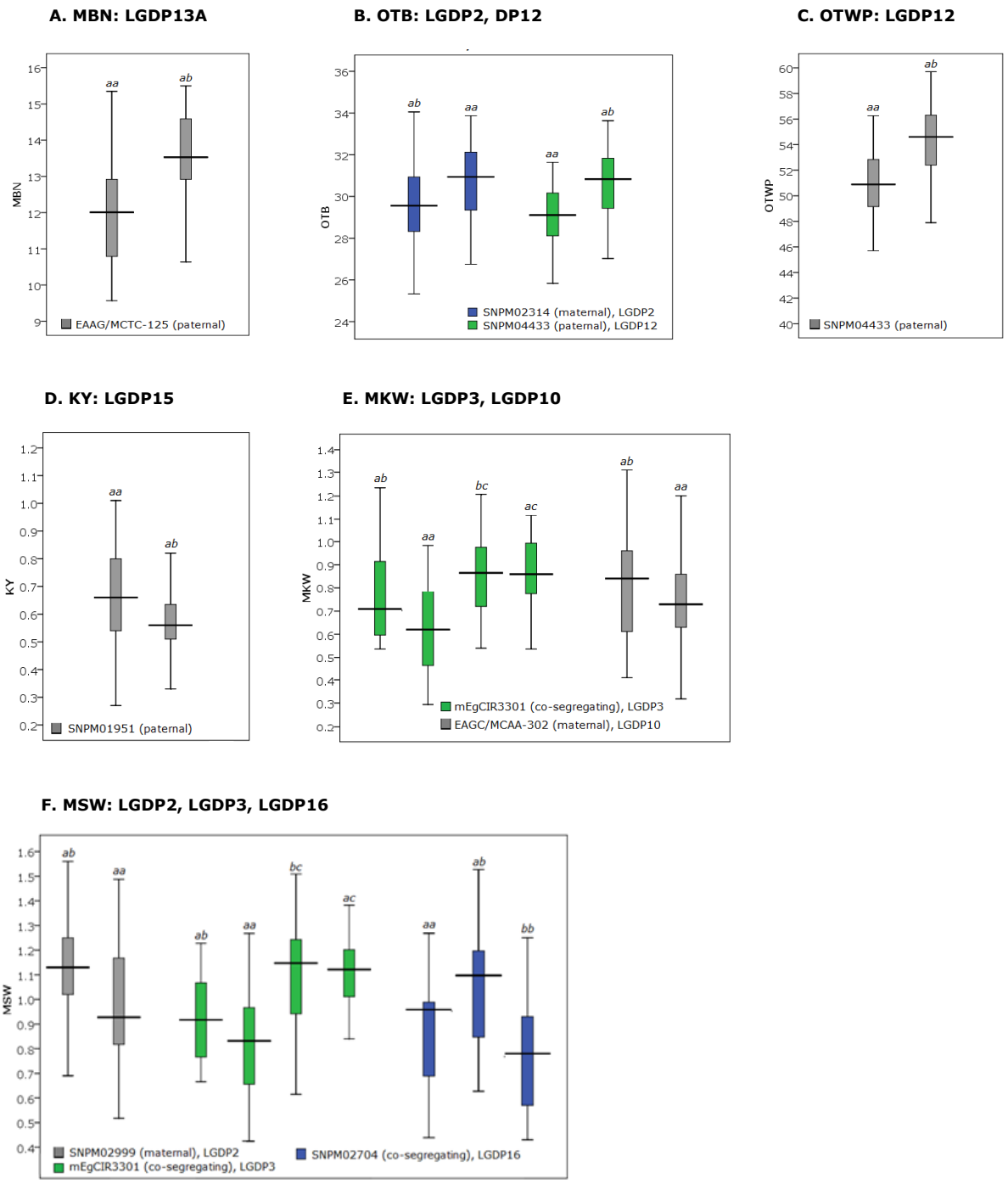


Fig. 2: Boxplot distribution of YCs by genotype of closest markers to QTL peaks in P2.

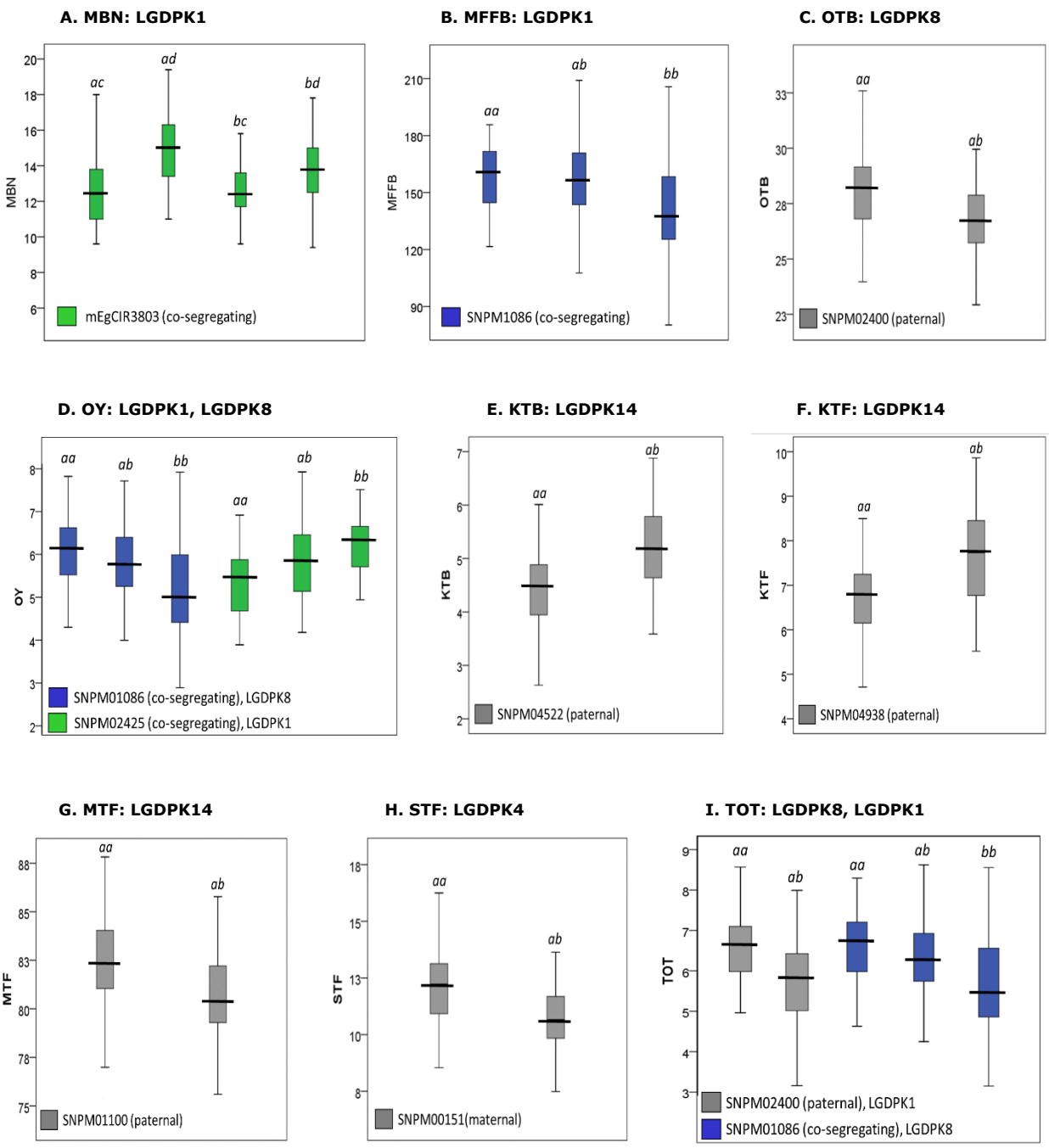
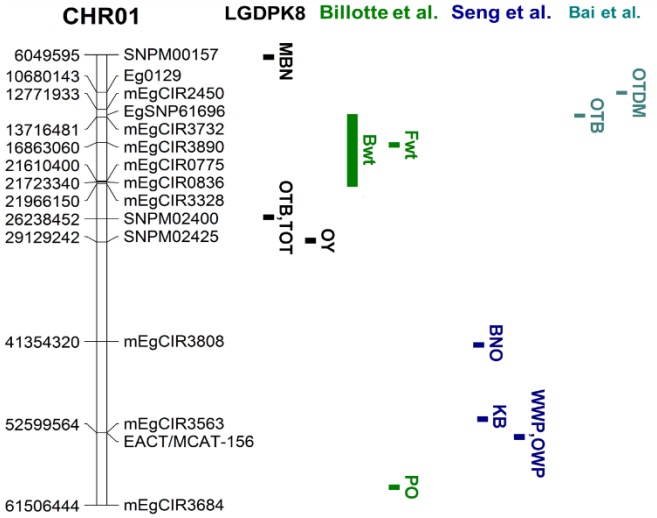
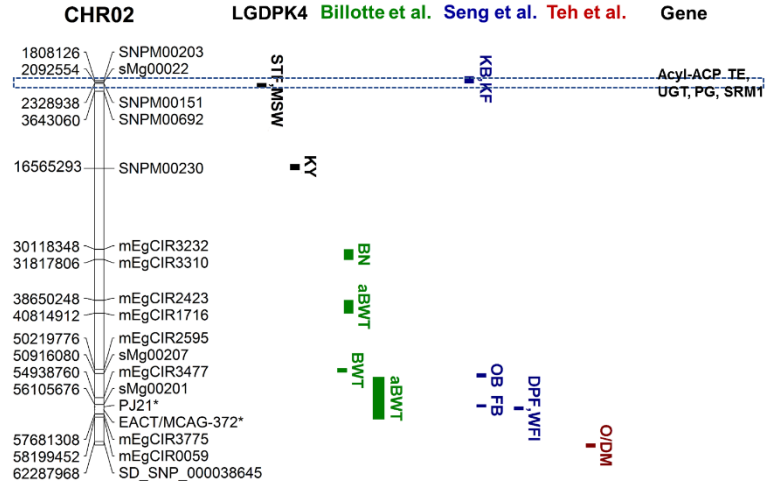
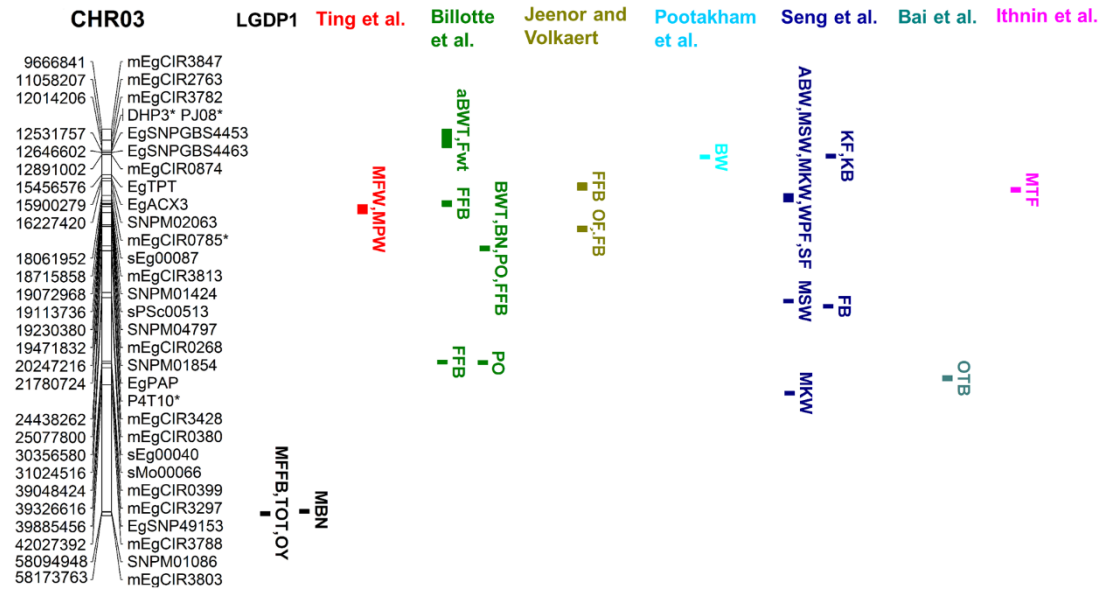
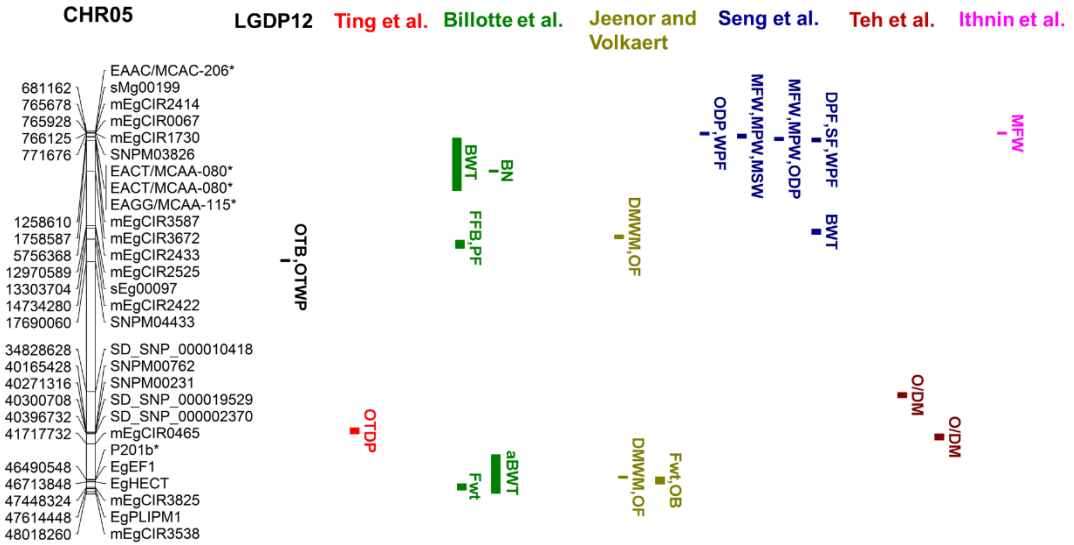


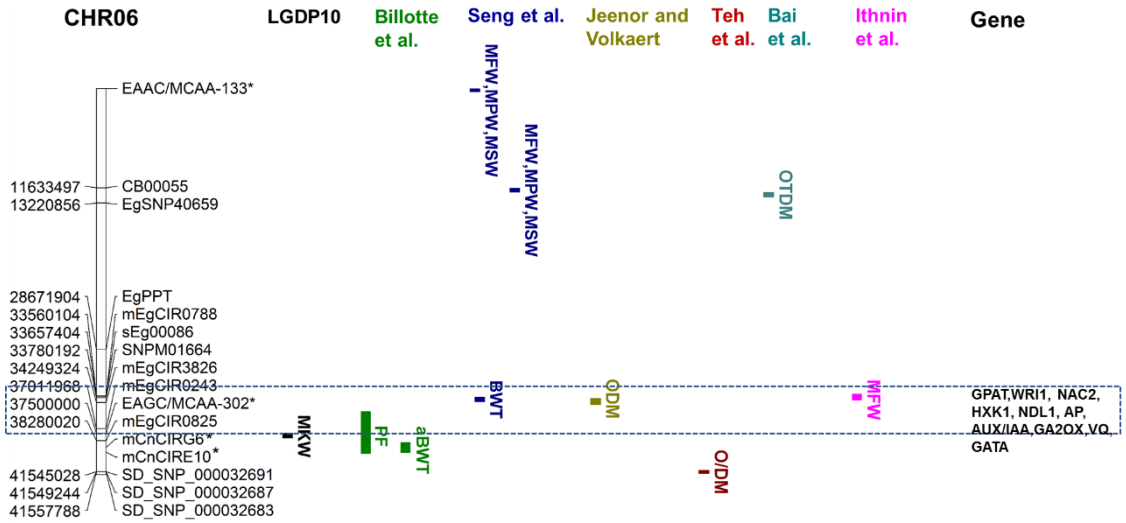
Fig. 3: Boxplot distribution of YCs by genotype of closest markers to QTL peaks in KULIM DxP.

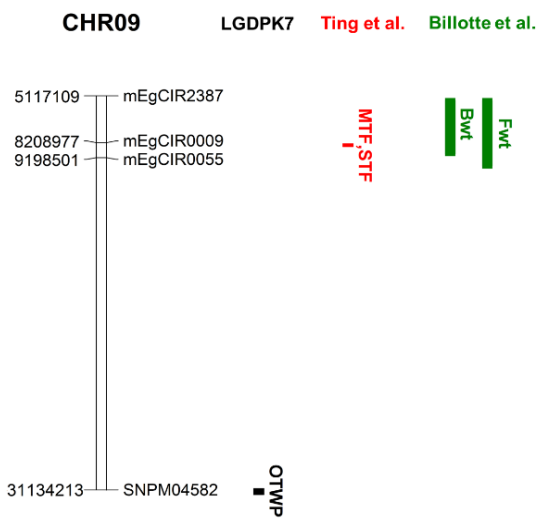

CHR01
LGDPK8
Billotte et al.
Seng et al.
Bai et al.
6049595 SNPM00157
10680143 Eg0129
12771933 mEgCIR2450
EgSNP61696
13716481 mEgCIR3732
16863060 mEgCIR3890
21610400 mEgCIR0775
21723340 mEgCIR0836
21966150 mEgCIR3328
26238452 SNPM02400
29129242 SNPM02425
41354320 mEgCIR3808
52599564 mEgCIR3563
EACT/MCAT-156
61506444 mEgCIR3684
MBN
Bwt
Fwt
OTB,TOT
OY
OTDM
OTB
BNO
KB
WWP,OWP
PO

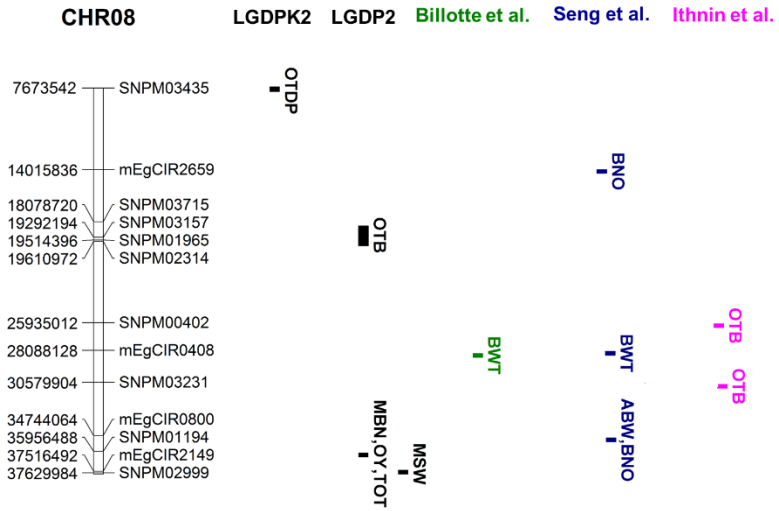

CHR02
LGDPK4
Billotte et al.
Seng et al.
Teh et al.
Gene
1808126 SNPM00203
2092554 sMg00022
2328938 SNPM00151
3643060 SNPM00692
16565293 SNPM00230
30118348 mEgCIR3232
31817806 mEgCIR3310
38650248 mEgCIR2423
40814912 mEgCIR1716
50219776 mEgCIR2595
50916080 sMg00207
54938760 mEgCIR3477
56105676 sMg00201
PJ21*
EACT/MCAG-372*
57681308 mEgCIR3775
58199452 mEgCIR0059
62287968 SD_SNP_000038645
STF,MSW
KY
KB,KF
Acyl-ACP_TE,
UGT_PG_SRM1
BN
aBWT
BWT
aBWT
OB
FB
DPF,WFI
O/DM

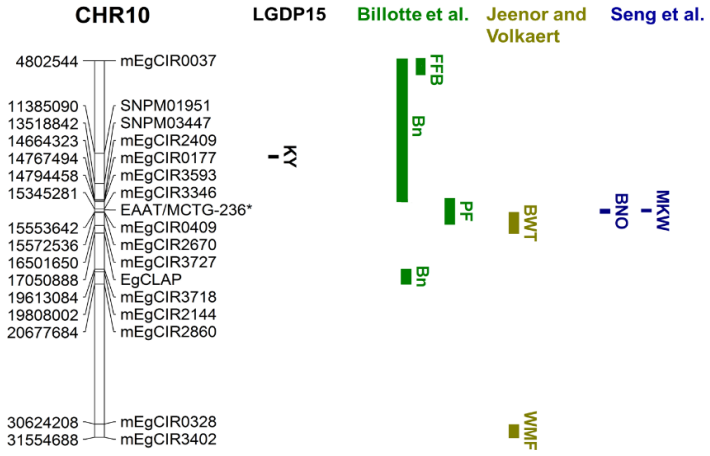

CHR03
LGDP1
Ting et al.
Billotte et al.
Jeenor and Volkaert
Pootakham et al.
Seng et al.
Bai et al.
Ithnin et al.
9666841 mEgCIR3847
11058207 mEgCIR2763
12014206 mEgCIR3782
DHP3* PJ08*
12531757 EgSNPGBS4453
12646602 EgSNPGBS4463
12891002 mEgCIR0874
15456576 EgTPT
15900279 EgACX3
16227420 SNPM02063
mEgCIR0785*
18061952 sEg00087
18715858 mEgCIR3813
19072968 SNPM01424
19113736 sPSc00513
19230380 SNPM04797
19471832 mEgCIR0268
20247216 SNPM01854
21780724 EgPAP
P4T10*
24438262 mEgCIR3428
25077800 mEgCIR0380
30356580 sEg00040
31024516 sMo00066
39048424 mEgCIR0399
39326616 mEgCIR3297
39885456 EgSNP49153
42027392 mEgCIR3788
58094948 SNPM01086
58173763 mEgCIR3803
MFW,MPW
MFFB,TOT,OY
MBN
aBWT,Fwt
FFB
FFB
BWT,BN,PO,FFB
PO
FFB OF,FB
BW
ABW,MSW,MKW,WPF,SF
MSW
MKW
KF,KB
FB
OTB
MTF

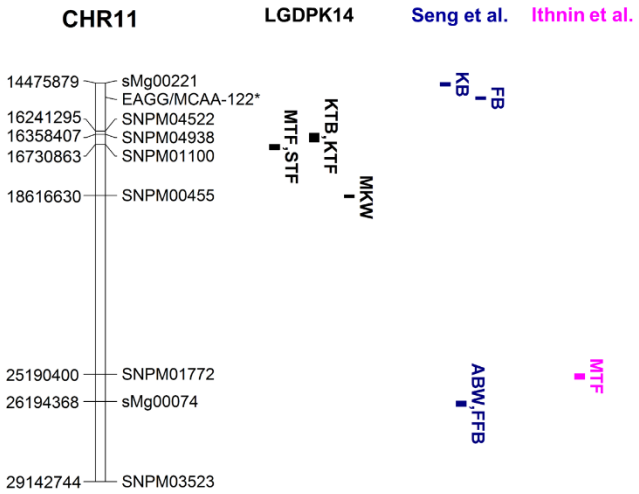

CHR05
LGDP12
Ting et al.
Billotte et al.
Jeenor and Volkaert
Seng et al.
Teh et al.
Ithnin et al.
EAAC/MCAC-206*
681162 sMg00199
765678 mEgCIR2414
765928 mEgCIR0067
766125 mEgCIR1730
771676 SNPM03826
EACT/MCAA-080*
EACT/MCAA-080*
EAGG/MCAA-115*
1258610 mEgCIR3587
1758587 mEgCIR3672
5756368 mEgCIR2433
12970589 mEgCIR2525
13303704 sEg00097
14734280 mEgCIR2422
17690060 SNPM04433
34828628 SD_SNP_000010418
40165428 SNPM00762
40271316 SNPM00231
40300708 SD_SNP_000019529
40396732 SD_SNP_000002370
41717732 mEgCIR0465
P201b*
46490548 EgEF1
46713848 EgHECT
47448324 mEgCIR3825
47614448 EgPLIPM1
48018260 mEgCIR3538
OTB,OTWP
OTDP
BN
BWT
FFB,PF
aBWT
Fwt
DMWM,OF
DMWM,OF
ODP,WPF
MFW,MPW,MSW
MFW,MPW,ODP
DPF,SF,WPF
BWT
Fwt,OB
O/DM
O/DM
MFW

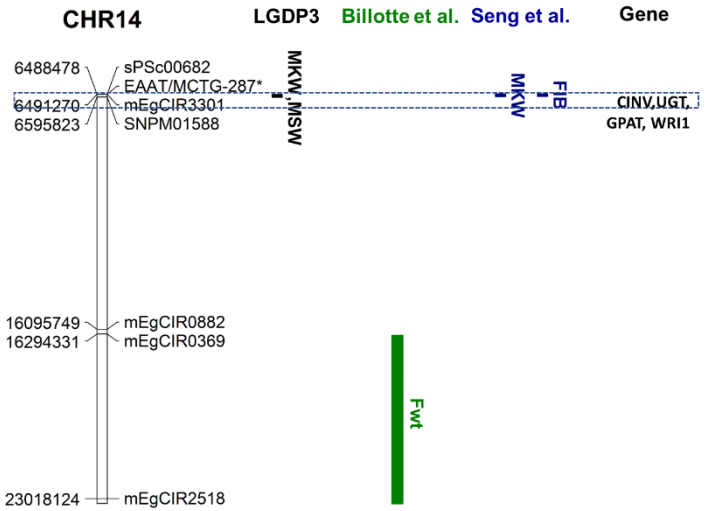

CHR06
LGDP10
Billotte et al.
Seng et al.
Jeenor and Volkaert
Teh et al.
Bai et al.
Ithnin et al.
Gene
EAAC/MCAA-133*
11633497
13220856
CB00055
EgSNP40659
28671904
33560104
33657404
33780192
34249324
37011968
37500000
38280020
EgPPT
mEgCIR0788
sEg00086
SNPM01664
mEgCIR3826
mEgCIR0243
EAGC/MCAA-302*
mEgCIR0825
mCnCIRG6*
mCnCIRE10*
41545028
41549244
41557788
SD_SNP_000032691
SD_SNP_000032687
SD_SNP_000032683
MFW,MPW,MSW
MFW,MPW,MSW
OTDM
MKW
PF
aBWT
BWT
O/DM
O/DM
MFW
GPAT,WRI1, NAC2, HXK1, NDL1, AP, AUX/IAA,GA2OX,VQ, GATA

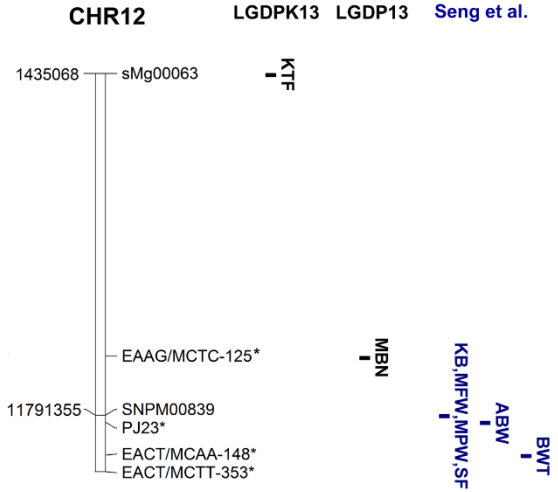

CHR09
LGDPK7
Ting et al.
Billotte et al.
5117109
8208977
9198501
31134213
mEgCIR2387
mEgCIR0009
mEgCIR0055
SNPM04582
MTF,STF
Bwt
Fwt
OTWP

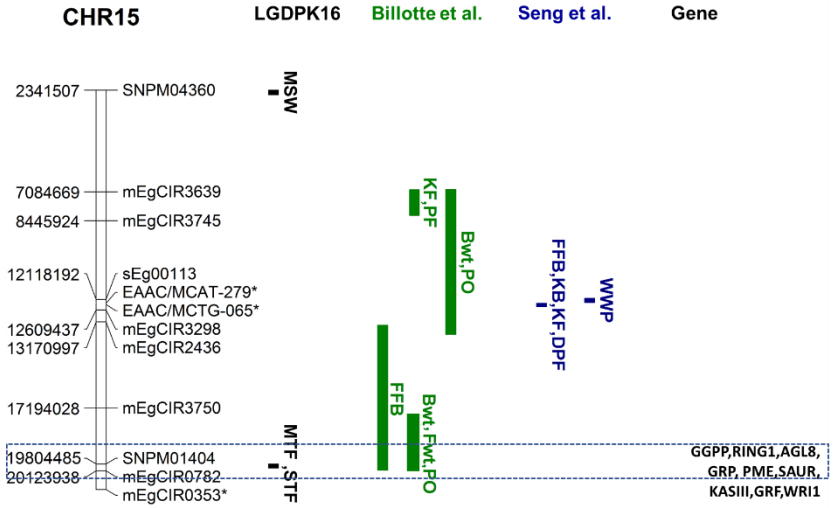

CHR08
LGDPK2
LGDP2
Billotte et al.
Seng et al.
Ithnin et al.
7673542
14015836
18078720
19292194
19514396
19610972
25935012
28088128
30579904
34744064
35956488
37516492
37629984
SNPM03435
mEgCIR2659
SNPM03715
SNPM03157
SNPM01965
SNPM02314
SNPM00402
mEgCIR0408
SNPM03231
mEgCIR0800
SNPM01194
mEgCIR2149
SNPM02999
OTDP
OTB
MBN,OY,TOT
MSW
BWT
BNO
BWT
ABW,BNO
OTB
OTB


CHR10
LGDP15
Billotte et al.
Jeenor and Volkaert
Seng et al.
4802544
11385090
13518842
14664323
14767494
14794458
15345281
15553642
15572536
16501650
17050888
19613084
19808002
20677684
30624208
31554688
mEgCIR0037
SNPM01951
SNPM03447
mEgCIR2409
mEgCIR0177
mEgCIR3593
mEgCIR3346
EAAT/MCTG-236*
mEgCIR0409
mEgCIR2670
mEgCIR3727
EgCLAP
mEgCIR3718
mEgCIR2144
mEgCIR2860
mEgCIR0328
mEgCIR3402
KY
FFB
Bn
PF
Bn
BWT
WMF
BNO
MKW


CHR11
LGDPK14
Seng et al.
Ithnin et al.
14475879
sMg00221
EAGG/MCAA-122*
16241295
SNPM04522
16358407
SNPM04938
16730863
SNPM01100
18616630
SNPM00455
25190400
SNPM01772
26194368
sMg00074
29142744
SNPM03523
MTF,STF
KTB,KTF
MKW
KB
FB
ABW,FFB
MTF


CHR14
LGDP3
Billotte et al.
Seng et al.
Gene
6488478
sPSc00682
EAAT/MCTG-287*
6491270
mEgCIR3301
6595823
SNPM01588
16095749
mEgCIR0882
16294331
mEgCIR0369
23018124
mEgCIR2518
MKW,MSW
MKW
FlB
CINV,UGT,
GPAT, WRI1
Fwt


CHR12
LGDPK13
LGDP13
Seng et al.
1435068
sMg00063
EAAG/MCTC-125*
11791355
SNPM00839
PJ23*
EACT/MCAA-148*
EACT/MCTT-353*
KTF
MBN
KB,MFW,MPW,SF
ABW
BWT


CHR15
LGDPK16
Billotte et al.
Seng et al.
Gene
2341507
SNPM04360
7084669
mEgCIR3639
8445924
mEgCIR3745
12118192
sEg00113
EAAC/MCAT-279*
EAAC/MCTG-065*
12609437
mEgCIR3298
13170997
mEgCIR2436
17194028
mEgCIR3750
19804485
SNPM01404
20123938
mEgCIR0782
mEgCIR0353*
MSW
KF,PF
Bwt,PO
FFB,KB,KF,DPF
WWP
FFB
Bwt,Fwt,PO
MTF,STF
GGPP,RING1,AGL8,
GRP, PME,SAUR,
KASIII,GRF,WRI1

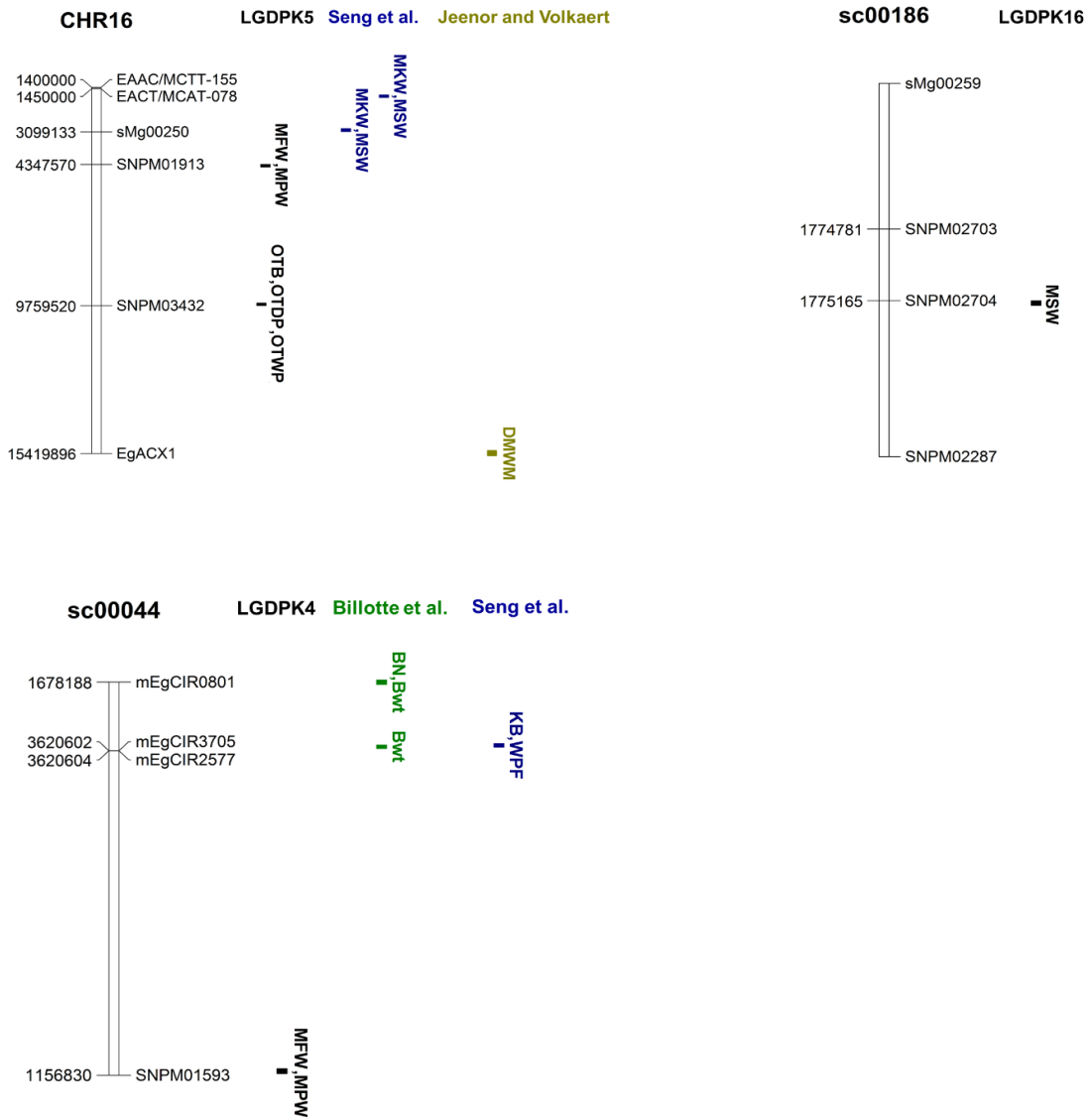


Fig. 4: Comparison of QTLs from different studies by mapping relevant information to oil palm EG5 genome build. Only closely linked markers defined the QTL regions for each trait on the chromosomes are shown.

Table 1: Genome-wide (GW) significant QTLs detected for YCs in P2. QTLs identified using Interval Mapping (IM), Multiple-QTL Model (MQM), Kruskal-Wallis non-parametric tests (KW) and G Model (GM).

Trait	IM					MQM		KW		GM	
LG	QTL interval (cM)	QTL peak (cM)	QTL peak (LOD)	Closest marker	Variation (%)	LOD	Variation (%)	K-value	p-value	Marker effect	p-value
Mean bunch number (MBN) (GW:3.4)											
DP13A	0 – 5.0	0	3.9	EAAG/MCTC-125	20.5	3.9	20.6	11.4	0.0010	0.59(-)	0.000969
Oil/bunch (OTB) (GW:3.2)											
DP2	48.0 – 52.0	48.7	3.8	SNPM02314	22.4	3.6	20.5	8.3	0.0050	0.92(-)	0.007228
				SNPM03157						0.92(-)	0.007228
				SNPM01965						0.92(-)	0.007228
				SNPM03715						0.92(-)	0.007228
DP12	34.3 – 42.8	39.8	3.6	SNPM04433	20.4	3.6	20.4	14.7	0.0005	1.20(-)	0.000160
Oil/wet mesocarp (OTWP) (GW:3.2)											
DP12	38.0 – 40.0	39.8	3.3	SNPM04433	18.8	3.3	18.8	15.3	0.0001	2.14(-)	0.000263
Kernel yield (KY) (GW:3.3)											
DP15	75.0 – 82.1	77.0	3.6	SNPM01951	34.1	3.6	34.1	5.4	0.0500	0.07(+)	0.013897
Mean kernel weight (MKW) (GW:3.4)											
DP3	54.2 – 58.5	54.2	3.8	mEgCIR3301	21.1	3.8	21.1	15.5	0.0050	0.18(-)	0.037698
				SNPM01588						0.11(+)	0.000661
				sPSc00682						0.09(-)	0.000187

DP10	11.7 – 22.9	20.7	3.7	EAGC/MCAA-302	22.7	3.4	22.5	4.3	0.0500	0.05(+)	0.043500
Mean shell weight (MSW) (GW:3.3)											
DP2	8.6	8.6	3.4	SNPM02999	19.1	3.4	19.1	9.0	0.0050	0.10(+)	0.002726
				mEgCIR2149						0.08(+)	0.002726
				SNPM03213						0.06(+)	0.029294
				sMg00194						0.07(+)	0.016289
DP3	53.0 – 57.5	54.2	4.6	mEgCIR3301	25.3	4.6	25.3	19.3	0.0005	0.25(-)	0.042563
				SNPM01588						0.16(+)	0.000064
				sPSc00682						0.12(-)	0.000046
DP16	0.5 – 2.5	1.2	3.8	SNPM02704	21.3	3.8	21.5	13.4	0.0050	NA	NA
				SNPM04360						NA	NA

NA: Not analysed using GM as both parents have same '*ab*' genotype <*ab*x*ab*>

Table 2: Genome-wide (GW) significant QTLs detected for YCs in KULIM DxP. QTLs identified using Interval Mapping (IM), Multiple-QTL Model (MQM), Kruskal-Wallis non-parametric tests (KW) and G Model (GM).

Trait	IM					MQM		KW		GM	
LG	QTL interval (cM)	QTL peak (cM)	QTL peak (LOD)	Closest marker	Variation (%)	LOD	Variation (%)	K-value	p-value	Marker effect	p-value
Mean bunch number (MBN) (GW = 4.5)											
DPK1	0.7 – 7.2	4.2	5.1	mEgCIR3803	15.9	5.1	15.9	22.7	0.0001	NA	NA
Mean fresh fruit bunch (MFFB) (GW = 4.5)											
DPK1	0.7– 7.6	2.3	5.4	SNPM01086	16.8	5.4	16.8	11.4	0.0050	NA	NA
Oil/bunch (OTB) (GW = 4.5)											
DPK8	101.1 – 103.4	102.8	4.6	SNPM02400	14.5	4.6	14.5	17.4	0.0001	0.72(+)	0.000011
Oil yield (OY) (GW = 4.3)											
DPK1	0.0 – 7.6	2.27	5.4	SNPM01086	16.5	5.3	16.5	23.6	0.0001	NA	NA
DPK8	92.3 – 105.2	98.8	5.1	SNPM02425	16.0	5.1	16.0	19.6	0.0001	NA	NA
Kernel/bunch (KTB) (GW = 4.6)											
DPK14	46.9 – 64.8	54.0	6.9	SNPM04522	21.1	6.9	21.1	28.5	0.0001	0.37(-)	0.000000
				SNPM04938	21.1	6.9	21.1	28.5	0.0001	0.37(-)	0.000000
Kernel/fruit (KTF) (GW = 4.5)											
DPK14	48.9 – 62.8	54.0	6.1	SNPM04938	18.8	6.1	18.8	24.6	0.0001	0.50(-)	0.000000
				SNPM04522	18.8	6.1	18.8	24.6	0.0001	0.50(-)	0.000000
Mesocarp/fruit (MTF) (GW = 4.4)											
DPK14	53.5 – 60.4	57.4	4.9	SNPM01100	15.6	5.0	15.8	20.2	0.0001	0.85(+)	0.000002
		54.0	4.7	SNPM04522	14.9	4.7	14.9	20.2	0.0001	0.94(+)	0.000005
				SNPM04938	14.9	4.7	14.9	20.2	0.0001	0.94(+)	0.000005

Shell/fruit (STF) (GW = 4.5)											
DPK4	3.5 – 16.2	6.7	6.0	SNPM00151	18.6	6.0	18.6	24.6	0.0001	0.73(-)	0.000000
Total oil (TOT) (GW = 4.5)											
DPK8	93.1 – 102.8	102.8	4.7	SNPM02400	14.8	4.7	14.9	19.3	0.0001	0.38(+)	0.000006
DPK1	0.0 – 7.6	2.3	5.2	SNPM01086	16.2	5.2	16.2	14.2	0.0010	NA	NA

NA: Not analysed using GM as both parents have same '*ab*' genotype <*ab*x*ab*>

Table 3: Chromosome-wide (CW) QTLs detected for YCs in P2. QTLs identified using Interval Mapping (IM), Multiple-QTL Model (MQM), Kruskal-Wallis non-parametric tests (KW) and G Model (GM).

Trait	**CW**	**IM**					**MQM**		**KW**		**GM**	
LG		**QTL interval (cM)**	**QTL peak (cM)**	**QTL peak (LOD)**	**Closest marker**	**Variation (%)**	**LOD**	**Variation (%)**	**K-value**	**p-value**	**Marker effect**	**p-value**
Mean bunch number (MBN)												
DP2	2.7	12.9	12.9	2.8	SNPM01194	14.6	2.8	14.6	12.0	0.0050	NA	NA
Total oil (TOT)												
DP2	3.1	12.7 – 12.9	12.9	3.2	SNPM01194	18.2	3.2	18.2	8.9	0.0500	NA	NA
Oil yield (OY)												
DP2	3.0	12.7 – 12.9	12.9	3.1	SNPM01194	17.6	3.1	17.6	7.8	0.0500	NA	NA

NA: Not analysed using GM as both parents have same '*ab*' genotype <*ab*x*ab*>.

Table 4: Chromosome-wide (CW) QTLs detected for YCs in KULIM DxP. QTLs identified using Interval Mapping (IM), Multiple-QTL Model (MQM), Kruskal-Wallis non-parametric tests (KW) and G Model (GM).

Trait	CW	IM					MQM		KW		GM	
LG		QTL interval (cM)	QTL peak (cM)	QTL peak (LOD)	Closest marker	Variation (%)	LOD	Variation (%)	K-value	p-value	Marker effect	p-value
Mean bunch number (MBN)												
DPK8	3.1	157.4 – 162.0	7.9	3.61	SNPM00157	11.6	3.6	11.6	12.9	0.0005	0.79(+)	0.000097
Mean fruit weight (MFW)												
DPK5	2.6	24.3 – 54.7	43.8	3.56	SNPM01913	11.6	3.6	11.5	14.0	0.0005	0.52(+)	0.000155
DPK4	3.5	147.1 – 171.9	159.5	4.11	SNPM01593	12.9	4.3	13.7	17.0	0.0001	0.57(+)	0.000017
Mean mesocarp weight (MPW)												
DPK4	3.2	150.5 – 174.8	159.5	4.2	SNPM01593	13.5	3.6	11.6	14.6	0.0005	0.46(+)	0.000077
DPK5	2.7	38.8 – 52.7	43.8	3.25	SNPM01913	10.6	3.3	10.6	16.6	0.0005	0.45(+)	0.009190
Mean shell weight (MSW)												
DPK4	3.3	3.8 – 9.7	6.7	3.56	SNPM00151	11.5	3.51	11.3	16.5	0.0001	0.10(-)	0.008394
Mean kernel weight (MKW)												
DPK14	3.0	61.4 – 65.8	62.8	3.3	SNPM00455	10.9	3.21	12.1	10.3	0.005	0.02(-)	0.000105

Kernel Yield (KY)												
DPK14	3	101.6 – 106.0	101.6	4.0	SNPM03523	12.9	3.0	10.8	12.1	0.0010	NA	NA
DPK4	3.4	54.0 – 62.8	55.7	3.1	SNPM00230	10.4	3.4	11.0	13.1	0.0050	0.06(-)	0.000529
Oil/wet mesocarp (OTWP)												
DPK7	3.1	29.2 – 31.9	31.9	3.4	SNPM04582	13.1	2.7	10.9	4.0	0.0500	0.64(-)	0.000709
DPK5	2.8	4.5 – 32.0	20.5	3.8	SNPM03432	12.8	3.7	12.0	12.7	0.0005	0.64(-)	0.000709
Oil/dry mesocarp (OTDP)												
DPK5	2.6	17.4 – 22.5	20.5	2.9	SNPM03432	10.3	2.8	9.2	4.0	0.0500	0.78(-)	0.000482
LGDP2	2.9	101.0	101.0	2.9	SNPM03435	9.4	3.0	9.7	12.5	0.0050	NA	NA
Oil/bunch												
DPK5	2.8	15.4 –20.5	20.5	2.9	SNPM03432	9.8	2.8	9.2	8.1	0.0050	0.45(-)	0.007484
Kernel/fruit (KTF)												
DPK13	2.9	52.3	52.3	2.9	SNPM00839	9.4	2.7	8.7	10.5	0.0100	NA	NA
Mesocarp/fruit (MTF)												
DPK16	2.9	43.7 – 52.8	48.4	3.8	SNPM01404	12.3	3.8	12.2	8.0	0.005	0.72(-)	0.001453
Shell/fruit (STF)												
DPK14	2.9	54.0 – 57.4	57.4	3.4	SNPM01100	10.9	3.5	11.4	15.0	0.0005	0.56(-)	0.000111
DPK16	3.0	44.1 – 52.7	48.4	4.0	SNPM01404	13.0	4.0	13.0	8.8	0.0050	0.49(+)	0.000790

Table 5: Putative biological functions for the candidate genes, proteins and transcription factors identified within the QTL region associated with yield components in the P2 and KULIM DxP mapping populations.

No.	Chromosome	Position (bp)	Within / flanking QTL region	Gene/transcription factor (TF)	NCBI accession number		Putative function for the encoded enzymes/protein/TF	Reference
					Gene	Protein		
1	CHR02	2,102,678 - 2,108,692	Within	*Acyl-acyl carrier protein thioesterase (Acyl-ACP TE)*	840418	Q9C7I5	*Acyl-ACP TE* plays an essential role in determining the fatty acid (FA) chain length by hydrolyzing the thioester bond which results in termination of acyl chain elongation during *de novo* biosynthesis of FAs in plants.	Uniprot (https://www.uniprot.org/uniprot/Q9C7I5); Pulsifer et al. 2014; Jing et al. 2011
2	CHR02	2,160,414 - 2,270,506	Within	*UDP-glycosyltransferase (UGT)*	N/A	K4CWS6	*UGT* mediates glycosylation modification such as anthocyanins, flavonols and flavor-related volatiles in development and ripening of fruits. It is also required for seed germination, abscisic acid (ABA)-mediated fruit ripening and negative responses to drought.	Uniprot (https://www.uniprot.org/uniprot/K4CWS6); Sun et al. 2017; Wu et al. 2017
3	CHR02	2,352,982 - 2,414,892	Flanking	*Polygalacturonase (PG)*	544051	P05117	*PG* is involved in pectin depolymerisation by hydrolyzing the O-glycosyl bonds in polygalacturonan, resulting in separation of cells in fruit abscission.	Uniprot (https://www.uniprot.org/uniprot/P05117); Osteryoung et al. 1990; Watson et al. 1994; Cooley et al. 1998; Roongsattham et al. 2012
4	CHR02	3,424,098 - 3,426,635	Flanking	*Salt-Related MYB1 (SRM1)*	830751	Q9FNN6	*SRM1* coordinates syntheses of ABA and signalling-related genes. In *Arabidopsis*, increasing ABA has negative effect on seed germination in saline conditions. It also promotes vegetative growth and leaf shape.	Uniprot (https://www.uniprot.org/uniprot/Q9FNN6); Wang et al. 2015
5	CHR06	33,837,283 - 33,840,111	Flanking	*Glycerol-3-phosphate acyltransferase (GPAT)*	836183	Q8GWG0	*GPAT* is involved in acylation of glycerol 3-phosphate in glycerolipid (e.g. triacylglycerol) biosynthesis in most plant seeds.	Uniprot (https://www.uniprot.org/uniprot/Q8GWG0); Singer et al. 2016; Shockey et al. 2016

6	CHR06	34,513,577 - 34,520,834	Flanking	*WRINKLED1 (WRI1)*	824599	Q6X5Y6	*WRI1* promotes sugar uptake and FA biosynthesis in developing seeds. The TF is also involved in embryo development, seed germination and seedling establishment.	Uniprot (https://www.uniprot.org/uniprot/Q6X5Y6); Zhai et al. 2017
7	CHR06	34,771,741 - 34,779,694	Flanking	*NAC domain-containing protein 2 (NAC2)*	101248665	K4BNG7	*NAC2* is a plant-specific TF involved in regulation of leaf senescence, fruit yield and sugar content in fruit ripening by establishing ABA homeostasis.	Uniprot (https://www.uniprot.org/uniprot/K4BNG7); Ma et al. 2018
8	CHR06	35,270,178 - 35,283,392	Flanking	*Hexokinase-1 (HXK1)*	829034	Q42525	In plants, *HXK1* encodes hexokinase, a sugar sensor in the glucose-signalling network.	Uniprot (https://www.uniprot.org/uniprot/Q42525); Dai et al. 1995; Granot et al. 2013
9	CHR06	35,818,699 - 35,823,166	Flanking	*N-MYC downregulated1 (NDL1)*	835777	Q9FJT7	*NDL1* interacts with the *G protein beta subunit (GB1)* is involved in regulation of lateral root formation and basipetal inflorescence auxin transport. Its overexpression will affect root architecture and reproductive organ development.	Uniprot (https://www.uniprot.org/uniprot/Q9FJT7); Mudgil et al. 2009; Mudgil et al. 2013
10	CHR06	36,319,863 - 36,322,283	Flanking	*Aspartic proteinase (AP)*	820452	Q9LTW4	*AP* plays an essential role in regulation of endogenous sugar levels, photosynthetic carbon metabolism in chloroplasts and general morphology and development of plant.	Uniprot (https://www.uniprot.org/uniprot/Q9LTW4); Paparelli et al. 2012; Al'bert et al. 2014
11	CHR06	36,637,313 - 36,642,286	Flanking	*Aux/IAA gene family (Aux/IAA)*	N/A	Q38825	*Aux/IAA* plays an important role in development and growth of roots, shoots, flowers and fruits. It is also a repressor of early auxin-inducible gene expression by interacting with *auxin response factors (ARFs)*.	Uniprot (https://www.uniprot.org/uniprot/Q38825); Liscum et al. 2002; Luo et al. 2018
12	CHR06	37,377,896 - 37,379,666	Within	*Gibberellin 2-beta-dioxygenase (GA2OX)*	4342182	Q8LGZ9	*GA2OX* regulates plant growth and architecture by inhibiting endogenous bioactive gibberellins.	Uniprot (https://www.uniprot.org/uniprot/Q8LGZ9); Lo et al. 2008; Shan et al. 2014
13	CHR06	37,411,925 - 37,413,912	Within	*Valine-glutamine motif-containing protein (VQ)*	6240987	Q1G3U8	*VQ* interacts with *WRKY* and is responsible for various developmental processes such as responses to biotic and abiotic stresses, seed development and size, and photomorphogenesis.	Uniprot (https://www.uniprot.org/uniprot/Q1G3U8); Hu et al. 2013; Jing et al. 2015; Wang et al. 2010; Cheng et al 2012, Pecher et al. 2014

14	CHR06	37,839,335 - 37,841,673	Within	*GATA* TF *(GATA)*	835788	Q5HZ36	*GATA* is involved in regulation of chlorophyll biosynthesis, chloroplast development, germination, senescence, elongation growth, flowering time and leaf starch content.	Uniprot (https://www.uniprot.org/uniprot/Q5HZ36); Mara et al. 2018; Richter et al. 2010; Hudson et al. 2011; Chiang et al. 2012; Richter et al. 2013; Behringer et al. 2014
15	CHR09	8,605,559 - 8,606,879	Within	*Zinc finger, C3HC4 type (RING finger)*	N/A	N/A	*RING finger* is involved in growth and fruit development.	Wu et al. 2014
16	CHR09	8,884,309 - 8,895,044	Within	*Membrane-bound O-acyltransferase (MBOAT)*	N/A	Q5GKZ7; Q9CAN8	Plant MBOATs, including *diacylglycerol acyltransferase* (*DGAT*) and *lysophospholipid acyltransferase* (*LPLAT*), play important role in lipid metabolism in developing seeds.	Uniprot (https://www.uniprot.org/uniprot/Q5GKZ7); Li et al. 2013; (https://www.uniprot.org/uniprot/Q9CAN8); Wang et al. 2012; Rosli et al. 2018
17	CHR14	6,284,291 - 6,291,463	Flanking	*Alkaline/neutral invertase (CINV)*	840454	Q9LQF2	*CINV* breaks sucrose down to fructose and glucose. It regulates root growth, leaf and silique development and floral transition.	Uniprot (https://www.uniprot.org/uniprot/Q9LQF2); Xiang et al. 2011
18	CHR14	6,369,304 - 6,371,721	Flanking	*UDP-glycosyltransferase (UGT)*	N/A	K4CWS6	In plants, *UGT* mediates glycosylation modification, such as in anthocyanins, flavanols and flavour-related volatiles, in development and ripening of fruits. It is also required for seed germination, ABA mediated fruit ripening and for negative response to drought.	Uniprot (https://www.uniprot.org/uniprot/K4CWS6); Sun et al. 2017; Wu et al. 2017
19	CHR14	6,480,850 - 6,486,840	Within	*Glycerol-3-phosphate acyltransferase (GPAT)*	836183	Q8GWG0	*GPAT* is involved in acylation of glycerol 3-phosphate in glycerolipid (e.g. triacylglycerol) biosynthesis in most plant seeds.	Uniprot (https://www.uniprot.org/uniprot/Q8GWG0); Singer et al. 2016; Shockey et al. 2016
20	CHR14	6,510,932 - 6,516,831	Within	*WRINKLED1 (WRI1)*	824599	Q6X5Y6	*WRI1* promotes sugar uptake and FA/oil biosynthesis in developing seeds which affects embryo development, seed germination and seedling establishment.	Uniprot (https://www.uniprot.org/uniprot/Q6X5Y6); Zhai et al. 2017
21	CHR15	19,353,857 - 19,357,974	Flanking	*Geranylgeranyl diphosphate chloroplastic (GGPP)*	N/A	N/A	*GGPP* is a precursor for various aspects of growth and development in plants, including biosynthesis of gibberellins, carotenoids, chlorophylls, isoprenoid quinones and geranylgeranylated proteins.	Okada et al. 2000
22	CHR15	19,399,974 - 19,400,633; 19,688,746 - 19,692,587	Flanking	*E3 ubiquitin-protein ligase RING1 (RING1)*	830902	Q9LX93	*RING1* is involved in development of plants, including dormancy and germination of seeds, root growth, flowering time and chloroplast development.	Uniprot (https://www.uniprot.org/uniprot/Q9LX93); Lin et al. 2008; Shu et al. 2017

23	CHR15	19,500,133 - 19,510,818	Flanking	*Agamous-like MADS-box protein (AGL8)*	836212	Q38876	*AGL8* regulates development of flowers and fruits by interacting with other MADS-box genes. For example, it promotes early floral meristem identity by interacting with *APETALA1*, *CAULIFLOWER* and *LEAFY* genes and together with *FRUITFULL* gene promotes carpel and fruit development. Therefore, mutations in these MADS-box genes could cause non-flowering phenotypes.	Uniprot (https://www.uniprot.org/uniprot/Q38876); Gu et al. 1998; Ferrandiz et al. 2000
24	CHR15	19,688,746 - 19,692,587	Flanking	*Glycine-rich protein (GRP)*	N/A	N/A	*GRP* is involved in cellular stress responses and signalling, floral development and auxin signalling.	Czolpinska et al. 2018
25	CHR15	19,788,553 - 19,805,976	Flanking	*Pectinesterase (PME)*	544090	P14280	*PME* is involved in modification of cell wall during fruit development in preparation for ripening and softening (induced by *PG*).	Uniprot https://www.uniprot.org/uniprot/P14280
26	CHR15	20,058,133 - 20,059,096	Flanking	*Small auxin-up RNA-like auxin-responsive protein (SAUR)*	832205	Q29PU2	*SAUR* induced by auxin, is involved in various biological processes, including cell division, expansion and differentiation.	Uniprot (https://www.uniprot.org/uniprot/Q29PU2); Markakis et al. 2013; Li et al. 2015
27	CHR15	20,227,498 - 20,233,306	Flanking	*3-oxoacyl-[acyl-carrier-protein (ACP)] synthase III (KASIII)*	4348632	Q7XEM4	*KASIII* enzyme catalyses condensation of malonyl-ACP to initialize carbon chain elongation during FA biosynthesis.	Uniprot (https://www.uniprot.org/uniprot/Q7XEM4); Alamin et al. 2017
28	CHR15	20,846,505 - 20,847,955	Flanking	*Growth-regulating factor (GRF)*	4330436	Q6ZIK5	*GRF* is involved in development and formation of root, leaf, stem, floral organ, seed and grain.	Uniprot (https://www.uniprot.org/uniprot/Q6ZIK5); Hu et al. 2015; Che et al. 2015; Duan et al. 2015; Li et al. 2016
29	CHR15	20,907,799 - 20,913,152	Flanking	*WRINKLED1 (WRI1)*	824599	Q6X5Y6	*WRI1* promotes sugar uptake and FA biosynthesis in developing seeds. The TF is also involved in embryo development, seed germination and seedling establishment.	Uniprot (https://www.uniprot.org/uniprot/Q6X5Y6); Zhai et al. 2017

Supplementary Table S1

Click here to view linked References

Supplementary Table S2

Click here to view linked References

Supplementary Figure S1

Click here to view linked References

Comparison of quantitative trait loci (QTLs) associated with yield components in two commercial *Dura* x *Pisifera* breeding crosses

Siti Hazirah Zolkafli[1]*, Ting Ngoot-Chin[1,3]*, Nik Shazana[1]*, Maizura Ithnin[1], Sean Mayes[2], Festo Massawe[3], Ravigadevi Sambanthamurthi[1], Ismanizan Ismail[4], Mohd Isa Zainol Abidin[5], Siti Habsah Roowi[6], Lee Yang Ping[6], Nurul Fatiha Farhana Hanafi[6], Rajinder Singh[1]**

[1]Advanced Biotechnology and Breeding Centre, Malaysian Palm Oil Board (MPOB), P.O. Box 10620, 50720 Kuala Lumpur, Malaysia

[2] Plant and Crop Sciences, Sutton Bonington Campus, University of Nottingham, Sutton Bonington, Loughborough LE12 5RD, UK

[3] School of Biosciences, University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor, Malaysia

[4] School of Bioscience and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor

[5]Plant Breeding and Services Department, KULIM Plantations Berhad, 81900 Kota Tinggi, Johor, Malaysia

[6] FGV R&D Sdn Bhd,, FGV Innovation Centre (Biotechnology), PT23417, Lengkuk Teknologi, 71760 Bandar Enstek, Negeri Sembilan

*These authors contributed equally to this work.

** Corresponding author

ABSTRACT

The high yielding *tenera* ~~from *dura* x *pisifera* intraspecific hybrids (DxP) of *Elaeis guineensis* Jacq~~ is the commercial oil palm planting material of choice in ~~most of~~ Southeast Asia. Notwithstanding this, there is continuous effort to further improve the yield and one way to do this is by addressing the yield components (YCs). Using 4,451 SNP ~~markers~~ and over 600 SSR markers, this study revealed quantitative trait loci (QTL) associated with YCs~~, including oil yield,~~ in two ~~DxP~~ breeding populations.~~:~~ ~~-~~a Deli *dura* x Yangambi *pisifera* (P2) and a Deli *dura* x AVROS *pisifera* (KULIM DxP). Thir~~Six~~teen and 29 QTLs were identified ~~for various YCs~~ in P2 and KULIM DxP, respectively. They were compared to other YC-linked QTLs reported previously for ~~oil palm of~~ different genetic backgrounds ~~. Comparison was also done~~ by mapping the QTL-linked markers~~-linked marker sequences~~ to the oil palm genome.~~, which facilitated the discovery of candidate genes within or near the QTL regions.~~ The Ccomparison ~~with other studies,~~ revealed four~~ive~~ common chromosomesal regions ~~(CHR02, 06, 09, 14 and 15)~~ containing QTLs influencing various YCs. The results reveal~~indicate~~ the possible presence of closely linked loci or pleiotropic genes influencing YCs~~yield~~ in ~~various DxP materials of~~ oil palm. Exploiting the genome data has also facilitated the discovery of candidate genes within or near the QTL regions including those~~Genes~~ related to glycosylation, fatty acid and oil biosynthesis, and development of flower, seed and fruit. ~~were identified within and flanking the chromosomal regions shared by multiple QTLs.~~

Keywords~~:~~

Oil palm, DxP, Quantitative trait loci, Yield components, comparative QTL mapping

Abbreviations

ABW	: Average bunch weight
Acyl-ACP TE	: Acyl-acyl carrier protein thioesterase
AFLP	: Amplified fragment length polymorphism
AGL8	: Agamous-like MADS-box protein
AP	: Aspartic proteinase
Aux/IAA	: Auxin/indole-3-acetic acid
BN, BNO	: Bunch number
Bwt, BW	: Bunch weight
CHR	: Chromosome
CINV	: Alkaline/neutral invertase
cM	: Centimorgan
CTAB	: Cetyl trimethylammonium bromide
DMWM	: Dry mesocarp/wet mesocarp
~~DPWT~~	~~: Dry mesocarp weight~~
EG5	: *E. guineensis* genome build
FELDA	: Federal Land Development Authority Malaysia
FFB	: Fresh fruit bunch(es) weight
FTB, FB	: Fruit/bunch
Fwt	: Fruit weight
GATA	: GATA-binding transcription factor
GA2OX	: Gibberellin 2-beta-dioxygenase
GGPP	: Geranylgeranyl diphosphate chloroplastic
GM	: G model
GPAT	: Glycerol-3-phosphate acyltransferase
GRF	: Growth-regulating factor
GRP	: Glycine-rich protein
GS	: Genomic selection
HXK1	: Hexokinase-1
IM	: Interval mapping
KASII, III	: beta-ketoacyl-ACP synthases II, III
KTB	: Kernel/bunch
KTF, KF	: Kernel/fruit
KW	: Kruskal-Wallis test
KY	: Kernel yield
LG	: Linkage group
LOD	: Logarithm of odds

MAS : Marker-assisted selection
MBN : Mean bunch number
MBOAT : Membrane-bound O-acyltransferase
MFFB : Mean fresh fruit bunch(es) weight
MFW : Mean fruit weight
MKW : Mean kernel weight
ML : Maximum likelihood
MPW : Mean mesocarp weight
MSW : Mean shell weight
MTF : Mesocarp/fruit
MQM : Multiple-QTL model
NAC2 : NAC domain-containing protein 2
NDL1 : N-MYC downregulated 1
N.N. Stress : Nearest neighbor stress
OTB, OB : Oil/bunch
OTDPM, O/DM : Oil/dry mesocarp
OTF, OF : Oil/fruit
OTWM, OTWP : Oil/wet mesocarp
OY : Oil yield
PF : Pulp/fruit
PME : Pectinesterase
PG : Polygalacturonase
PO : Palm oil yield/palm/year
QTL : Quantitative trait loci
RFLP : Restriction fragment length polymorphism
RING1 : Really interesting new gene 1
SAUR : Small auxin-up RNA-like auxin-responsive protein
SNP : Single nucleotide polymorphism
SRM1 : Salt-related MYB1
STF : Shell/fruit
SSR : Simple sequence repeat
TF : Transcription factor
TOT : Total oil
UGT : UDP-glycosyltransferase
VQ : Valine-glutamine motif-containing protein
WMF : Wet mesocarp/fruit
WPWT : Wet mesocarp weight
WRI1 : WRINKLED1
YC : Yield component

Introduction

Oil palm (*Elaeis guineensis* Jacq.) is the most productive oil crop in the world, and is currently grown on some 19 million hectares (ha) of land ~~(Kushairi et al. 2018)~~. This is only about 0.4 % of the total world agricultural land but accounts for almost 40.0 % of the global oils and fats (~~Pirker et al. 2016,~~ Kushairi et al. 2018~~9~~). Comparatively, soybean (*Glycine max*) utilizes 40.1 % of the total agricultural ~~world oilseeds crops~~ land, ~~,~~ followed by cottonseed (13.8 %), rapeseed (13.0 %) and sunflower (10.0 %) (Pirker et al. 2016~~Oil World 2013~~).

In traditional oil palm breeding, the parental lines are continuously crossed to generate superior progenies, similar to ~~that in~~ producing hybrids in other crops. The progeny from crosses however, are not automatically acceptable~~good~~ just because they come from good parents~~.~~, Thus~~o~~ each cross is progeny tested, and only the confirmed ~~good~~ combinations with superior yield are used to produce commercial seeds (Soh et al. 2003). It takes on average 10 ~~–~~ 12 years to develop a new variety, sometimes even up to 20 years for commercial application (Rajanaidu et al. 2000). The question begged is obviously whether the time can be shortened. The main challenge is ~~phenotypic data~~ collection of phenotypic data which is time consuming and labour-intensive, requring years for reliable data compilation. Yield is recorded for at least five years, from six to~~6 –~~ 10 years after planting in the field ~~from the nursery~~ and vegetative measurements have to be done several times (Corley and Tinker 2016, Swaray et al. 2020).

In introgressing good trait(s) from Palm A into Palm B, the whole gamut of genes from A, both good and bad, are first incorporated with those from B, and then the undesirable genes weeded out by repeated subsequent self-pollination and selection. It would be faster if only the good gene alleles could be introgressed, but the question has always been how to do so. In recent years, enabling technologies have emerged, such as marker-assisted selection (MAS) and genomic selection (GS). In MAS, markers are used to predict the phenotype, saving time and money in gathering the phenotypic data, as selection can be made even on seedlings when the adult features are yet to show (Collard et al. 2005, Nadeem et al. 2017). More recently, GS, which uses genome-wide markers to estimate the effects of all loci, makes it possible to compute a genomic estimated breeding value for specific traits (Wang et al. 2018) and this approach, is gaining prominence for crop improvement. Both, MAS and GS increase the rate of genetic gain by reducing the necessary selection time for the desired traits. MAS- and GS-based programmes have been applied to improve yield in soybean (Concibido et al. 2003, Sebastian et al. 2010, Jarquín et al. 2014, Fallen et al. 2015, Stewart-Brown et al. 2019) and maize (Yousef and Juvik 2001, Massman et al. 2012, Liu et al. 2015, Pace et al. 2015, Beyene et al. 2016, Wang et al. 2020) and have enhanced disease resistance, yield, plant height and flowering time in wheat and rice (Gupta et al. 2010, Poland et al. 2012, Ragimekula et al. 2013, Spindel et al. 2015, Thavamanikumar et al. 2015, Borrenpohl et al. 2020). These molecular strategies are also applicable to oil palm.

In oil palm, the required tools and techniques for MAS and GS have been developed over the last two decades. For example, DNA-based markers and identification of genomic loci associated with monogenic as well as polygenic traits have been reported (Jack and Mayes 1993, Singh and Cheah 2005). The causal genes regulating the two most important monogenic traits - shell and fruit colour - have been identified and the discoveries translated into commercial diagnostic assays (Singh et al. ~~2013b~~2013a, 2014, Ooi et al. 2016). For yield, the QTLs associated with oil yield (OY) and various other yield components (YCs) have been reported by Rance et al. (2001), Billotte et al. (2010), Jeennor and Volkaert (2014), Pootakham et al. (2015), Seng et al. (2016), ~~and~~ Teh et al. (2016,) 2020) and Bhagya et al. (2020). Many QTLs and markers have been associated with OY and various YCs across different genetic backgrounds, suggesting a complex

genetic mechanism determining oil palm yield. The QTLs were uncovered using different marker systems, starting with restriction fragment length polymorphism (RFLP), which were largely replaced by amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), and, more recently, single nucleotide polymorphism (SNP) based markers. RFLP-based markers are codominant, but not popular at present as the technique for generating and identifying informative RFLP markers is expensive and laborious. To overcome these shortfalls, AFLP markers can be used instead (Singh et al. 1999, Kularatne et al. 2001, Seng et al. 2007) although their dominant nature also posed some limitations in application. Subsequently, SSR markers (also codominant but requiring less DNA and with high reproducibility across laboratories) have become popular in oil palm research (Ting et al. 2010, Zaki et al. 2012, Ting et al. 2013). More recently still, SNP markers have gained importance and are preferred due to their wide distribution in the genome, codominant nature and amenability to high throughput analysis (Mishra et al. 2014, Nadeem et al. 2017).

This study constructed a genetic linkage map for a Deli *dura* x AVROS *pisifera* family, a commercial planting material, and updated the Deli *dura* x Yangambi *pisifera* genetic map constructed previously by Ting et al. (2014). Both maps were constructed using the same oil palm customised array containing 4,451 SNP markers and over 600 SSR markers, making the comparison possible. The genetic maps were then used to identify QTLs associated with OY and YCs, and the results were compared to the QTLs published previously for oil palm. Linking and cataloguing the QTLs identified in different studies and by different marker systems is challenging, but has fortunately been made easier with the publication of the oil palm genome build (EG5) (Singh et al. 2013a2013b). It is now possible to compare QTLs from different crosses and publications to determine if they fall within the same chromosomal regions. The ability to identify overlapping QTLs linked to a trait in a similar chromosomal region, adds confidence to the postulation that the genomic region strongly influences the trait concerned. Inclusion of QTL-linked markers consistently associated with a trait in a panel has increased the prediction accuracy of GS models in cattle improvement (Brøndum et al. 2015). More importantly, candidate genes within or near the QTL regions can now be identified for subsequent analysisidentification to determine of the actual causative genes for the yield trait(s).

Materials and methods

Mapping families

The first mapping family - P2 (05 Trial 1) - is an advanced breeding population (FGV R&D Sdn Bhd experimental plot, Kota Gelanggi, Pahang, Malaysia) a cross between an Ulu Remis Deli *dura* (ENL48) and a Yangambi *pisifera* (ML161)., The P2 population consisted ofwith 87 F_1 *tenera* palms currently grown at FGV R&D Sdn. Bhd., Kota Gelanggi, Pahang, Malaysia. The second family namely, (KULIM DxP) consisted of 135 F_1 *tenera* palms, planted at the (Tereh Utara Pplantation of, Kulim Plantation Bhd., Johor, Malaysia)., The KULIM DxP palms were generated from a cross between an ex-Ulu Remis Deli *dura* (KT 910512/0804) and an AVROS *pisifera* (KT 911101/1203). The maternal *dura* and the paternal *pisifera* palms are known to have contrasting yield parameters, as *pisifera* is female sterile and rarely produces fruit bunches to maturity (Wonkyi-Appiah 1987, Kushairi et al. 1999, Kushairi and Rajanaidu 2000, Swaray et al. 2020). The maternal Deli dura palms are known to have higher bunch weight and lower bunch number compared to the paternal pisifera and the resulting intraspecific progeny of these two parental palms show hybrid vigour for yield (Gascon and de Berchoux 1964, Durand-Gasselin et al. 2000, Jin et al. 2017, Singh et al. 2020). Leaf materials from all the palms, including the parental ones, were sampled for DNA extraction and marker analysis.

Yield-related phenotypic data

Formatted: Strikethrough

Ripe bunches from both families were analysed for their YCs over a ~~4.6 –~~ 5-year period according to the standard protocol used by oil palm breeders (Blaak et al. 1963, Rao et al. 1983, Isa et al. 2011). The standard protocol for determining YCs is also cited in the National standards (SIRIM standard MS157), as the recommended methodology to determine the suitable parental palms for commercial seed production. ~~the SIRIM standard oil palm fruit bunch analysis (Blaak et al. 1963, Rao et al. 1983, Isa et al. 2011).~~ A minimum of three~~two~~ bunches per palm were analysed for 16~~8~~ YC parameters: mean bunch number (MBN, no/palm/year), mean fresh fruit bunch weight (MFFB, kg/palm/year), mean fruit weight (MFW, g/fruit), total mesocarp and kernel oils (TOT, ton~~kg~~/ha~~palm~~/year), mesocarp oil yield (OY, ton~~kg~~/ha~~palm~~/year), oil/bunch (OTB, %), oil/wet mesocarp (OTWP, %), oil/dry mesocarp (OTDP, %), mean mesocarp weight (MPW, g/fruit), ~~wet mesocarp weight (WPWT, g), dry mesocarp weight (DPWT, g),~~ mesocarp/fruit (MTF, %), kernel yield (KY, ton~~kg~~/ha~~palm~~/year), mean kernel weight (MKW, g/fruit), kernel/fruit (KTF, %), kernel/bunch (KTB, %), mean shell weight (MSW, g/fruit) and shell/fruit (STF, %). The distribution and correlations between the parameters were evaluated using the Kolmogorov-Smirnov normality and Pearson correlation tests in SPSS 16.0.

Genomic DNA extraction

Extraction of genomic DNA from frozen leaves stored at -80 ºC was done using the modified CTAB method (Doyle and Doyle 1990). DNA quality was checked by digestion with *Eco*RI and *Hae*III and electrophoresed on 0.8 % agarose gel (Rahimah et al. 2006). The acceptable purity values were 1.8 – 2.0, as measured by the NanoDrop spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE)

SNP and SSR analyses

SNP genotyping was performed by a service provider using the oil palm customized OPSNP3 Illumina Infinium II Bead-Chip array (Illumina Inc., San Diego, CA) containing 4,451 SNPs. For SSR genotyping, fragment analysis was carried out using the ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA). The SNP and SSR genotyping ~~and~~ analyses were as described by Ting et al. (2013, ~~and~~ 2014).

Construction of genetic linkage maps

An integrated genetic map of P2 was constructed previously (Ting et al. 2014). Additional SSR markers (sMo, sMh, sMg, _oSSR, sTE, sEg, sOleiSc, p5sc322 and sPSc) from the MPOB SSR database (http://opsri.mpob.gov.my/opsri/welcome.php, Rosli et al. 2019) and Billotte et al. (2010) (mEgCIR) were genotyped and added to the P2 map. The KULIM DxP genetic map was constructed using JoinMap® 4.1 (van Ooijen 2006) as described by Ting et al. (2014). In brief, the independent parental and integrated KULIM DxP genetic maps were constructed simultaneously using the maximum likelihood (ML) mapping algorithm, where each linkage group (LG) was formed from marker pairs with recombination frequency ≤ 0.2. The Haldane mapping function was used to determine the map distance in centimorgan (cM) and markers with nearest ~~neighbor~~neighbour stress (N.N. Stress) value > 4 cM were excluded from the individual parental and integrated maps. Finally, a consistent marker-order was determined by four iterations of map calculation. The integrated genetic linkage maps for P2 and KULIM DxP were labeled as DP and DPK, respectively.

QTLs analysis

QTL analysis was carried out separately for DP and DPK as described by~~following~~ Ting et al. (2016). The default parameters in Interval Mapping (IM), the Multiple-QTL Model (MQM) and Kruskal-Wallis non-parametric ranking tests (KW) were used in MapQTL® 6 (van Ooijen 2009). The 95.0 % genome-wide (GW) and chromosome-wide (CW) LOD

significance thresholds for each YC was determined by 1,000 permutations. In addition, G model (GM) (Bernardo 2013) was used to estimate the individual marker effect for the QTLs linked to each YC.

Mapping of QTLs to the oil palm genome build

Markers from the QTL regions were aligned to the oil palm reference genome (EG5) (Singh et al. 2013ba) to identify their positions on the corresponding pseudo-chromosome using the program Exonerate (Slater et al. 2005) with its default parameters. Markers with low scores (< 90.0 % matched) and not uniquely mapped were removed. The genomic region corresponding to the QTLs were searched against the predicted oil palm gene model database (Chan et al. 2017) in PalmXplore (http://palmxplore.mpob.gov.my, Sanusi et al. 2018) to identify putative genes and their functions.

Results and discussion

Comparison of DP and DPK genetic maps

A DP (P2) genetic linkage map was constructed previously using AFLP, RFLP, SSR and SNP markers by Ting et al. (2014). A further 240 SSR markers, 151 from MPOB and 89 from Billotte et al. (2010) were added to the current DP map. The updated DP map now contains 1,595 markers across 16 LGs, spanning 1,714.3 cM. Interestingly, a small number of SNP markers (23 SNPM) that failed to map previously, are now in DP although the same mapping parameters were~~as~~ used. They helped bridge some gaps in the original map and further saturate some regions linked to QTLs e.g. OTB on LGDP2 and MSW and MKW on LGDP3. The DPK genetic map (KULIM DxP) had slightly fewer markers, only 57 SSRs and 1,449 SNPs in 16 LGs, covering a total map length of 1,902.3 cM. The average map distance per marker in DPK was 1.3 cM, which as expected~~,~~ was close to the 1.1 cM observed in DP. In DP, the LGs were 66.2 to 193.2 cM, and in DPK the range observed was 60.7 to 192.4 cM. In both populations, LGDP/DPK5 was the shortest, and the longest was – LGDP/DPK4. There were in total 746 common markers across the 16 LGs, a comparison of which revealed relatively high collinearity of the markers in both maps (Supplementary Figure 1). This is likely due to both populations having female parents of the Deli *dura* pedigree. This suggests that major chromosomal rearrangements have not yet occurred in domestication of the closely related parental lines, as also observed for watermelon (Ren et al. 2014).

Yield components (YCs) and correlations between them

Of the 16~~9~~ YCs evaluated, 11~~0~~ were common in both P2 and KULIM DxP families - MBN, MFFB, TOT, OY, KY, OTB, KTF, KTB, ~~OTB~~ MKW, MSW and OTWP. The data for ~~MKW, MSW, DPWT and WPWT were only determined in P2, while~~ MFW, MPW, STF, OTDP and MTF were only available for KULIM DxP. Almost all the YCs (except MSW) had a continuous and significant normal distribution ($p > 0.05$) in both populations. Normality of YC data was also observed in other oil palm mapping families analysed by Billotte et al. (2010), ~~and~~ Seng et al. (2016) and Teh et al. (2020). For P2, YC data were available for 75 of its 87 palms, of which three outliers were removed for MBN based on a Boxplot analysis comparing the observed and expected mean values (5.0 % trimmed mean, SPSS 16.0). For KULIM DxP, the data was available for all of its 135 palms. However, for MSW, MPW and MKW, one, two and four outliers were removed, respectively, following Boxplot analysis.

MBN was determined for an average of 13 bunches/palm for both families, where the range of observations made for individual palms of P2 and KULIM DxP was 6 – 16 and 6 – 19, respectively.~~MBN was 13 bunches harvested for both families, 6 – 16 and 6 – 19 for the individual palms in P2 and KULIM DxP, respectively.~~ As MFFB is influenced by MBN, variation was also observed for it, 72.04 – 210.53 kg/palm/year in the two populations, while OY was 2.53 – 7.92 ton~~kg~~/ha~~palm~~/year. The variations for the different YCs are summarized in Supplementary Table S1. Wide distribution

was also observed for fruit components, such as mesocarp measurements and their derivatives (MPW, ~~WPWT, DPWT,~~ OTWP, OTDP and MTF) as well as the kernel- and shell-related traits (KY, MKW, KTF and KTB, MSW and STF), suggesting that both populations are suitable for QTL analysis for all their YCs measured in this study.

The correlations between the various YCs were consistent in both P2 and KULIM DxP families, with three levels of positive relationships (Figure 1). Strong correlations were observed among MBN, MFFB, TOT and OY with r = 0.63 – 0.99. ~~Moderate correlations were observed for the YCs related to KY, OTB, OTWP, DPWT and MTF, with r = 0.39 – 0.50 and 0.28 – 0.58 for DP and DPK, respectively~~. The second level of positive correlations was among the mesocarp and endocarp components. The mesocarp components (OTB, OTDP, ~~DPWT~~ and MTF) and ~~(~~MPW ~~and WPWT)~~ had moderate correlation with r = ~~0.32 – 0.54 and~~ 0.20~~19~~ – 0.28~~51~~ for ~~P2 and~~ KULIM DxP~~, respectively~~. Moderate to strong correlations (r = 0.30 – 0.77) were recorded among the endocarp components where KTF, STF, KY and KTB were correlated with MKW and MSW. Finally, the mesocarp and endocarp components contributing to MFW showed strong correlations with MPW (r = 0.87~~1 – 0.98~~) and moderate correlations with MKW ~~and MSW~~ (r = 0.49~~ – 0.59~~). A graphical view of the correlations between the YCs is shown in Figure 1, while ~~and~~ Supplementary Table S2 demonstrates the relationships of both the direct (those categorized in the same group) and contributory effects (those at different levels) of the YCs to the overall yield in oil palm.

Pearson correlation was negative between some YCs, mainly between the mesocarp (OTB, OTPM, OTDP, ~~DPWT,~~ MPW~~,~~ ~~WPWT~~ and MTF) and endocarp (KTF, STF, KY, KTB, MKW and MSW) components. Among them, ~~strong~~ negative correlations with r = ~~-0.60 to -0.93 occurred between WPWT and the endocarp components in P2, while in KULIM DxP, they (r =~~ -0.29 to -0.95~~)~~ occurred~~were~~ between MTF and the endocarp components in KULIM DxP. This clearly indicates that increasing mesocarp reduces kernel and shell, and *vice versa*, suggesting competition among the sinks for assimilates. Strong correlations among the YCs were also reported by Kushairi et al. (1999), Okwuagwu et al. (2008), Okoye et al. (2009), ~~and~~ Seng et al. (2016), Osorio-Guarín et al. (2019) and Teh et al. (2020).

P2: QTLs linked to YCs.

In the DP genetic map, 10~~1~~ QTLs, significant at GW, were associated with various YCs. The traits for the QTLs and their LGs were MBN (LGDP13A), OTB (LGs DP2 and DP12), OTWP (LGDP12), ~~WPWT (LGDP7),~~ KY (LGDP15), MKW (LGs DP3 and DP10), MSW (LGs DP2, DP3 and DP16) (Table 1). A QTL associated with MBN was identified at map interval 0.0 – 5.0 cM on LGDP13A. An AFLP marker, EAAG/MCTC-125, was closest to the QTL peak detected at LOD 3.9 for MBN. Both the IM and MQM methods revealed that the QTL explained ~20.5 % of the phenotypic variation for MBN, and a negative (paternal) effect (-0.59) was estimated using GM. When associating the MBN phenotype with the observed genotype profiles, without the AFLP locus from the paternal palm (denoted *aa* genotype) (Figure 2 A) MBN increased to 13.30 ± 1.53 bunches from 12.11 ± 1.53 bunches. The limitation of an AFLP marker here was its dominant nature, and it was not clear if the marker concerned~~involved~~, EAAG/MCTC-125, amplified a homozygous or heterozygous DNA segment. Therefore, other flanking markers (LOD 3.6) – namely, sMo00166, sMo00196, SNPM04999 and SNPM03169 - located ~2.6 cM (Figure S1) away ~~from~~ were ~~be~~ used as proxies, although the phenotypic variation explained was slightly reduced to 18.6

QTLs associated with OTB were found in the 48.0 – 52.0 cM (4.0 cM confidence interval) and 34.3 – 42.8 cM (8.5 cM confidence interval) regions of LGs DP2 and DP12, respectively. Markers from the two intervals showed negative effects from 0.9 – 1.2 % (p = 0.007). The closest markers flanking the QTLs were SNPM02314 (LGDP2) and SNPM04433

(LGDP12). Palms categorized in the genotypes *ab* and *aa* had significant differences in OTB ($p \leq 0.05$ T-test, SPSS 16.0). For the marker from the maternal palm, - SNP02314 - the homozygous genotype *aa* showed increased OTB (31.4 ± 2.6 %), ~1.9 % higher than ~~for~~ the *ab* genotype (29.6 ± 2.9 %). The genotype of the paternal marker SNPM04433, meanwhile, had an opposite effect on OTB. The *aa* genotype (28.7 ± 2.8 %) had 2.6 % lower OTB than *ab* (31.3 ± 2.6 %) (Figure 2 B).

In addition to OTB, LGDP12 also hosted another GW significant QTL, OTWP, which interval overlapped that for OTB, with the same marker, SNPM04433, located closest to the QTL peaks for both traits. This explained why the two YCs were strongly correlated (r = 0.81). However, SNPM04433 had a stronger effect of -2.14 (p = 0.000263) for OTWP than for OTB (only -1.20, p = 0.000160). This was likely due to the larger variation for OTWP (3.2 %) in the two genotypes *ab* (54.0 ± 3.5 %) and *aa* (50.9 ± 3.2 %) (Figure 2 C). ~~For WPWT, a GW significant QTL was found on LGDP7 with the closest SSR marker pointing to it being mEgCIR0009 which revealed a difference of 2.36 g between the *ab* and *aa* genotypes (Figure 2 D).~~ QTLs associated with kernel and shell components, such as KY, MSW and MKW, were also identified on DP. The markers linked to them explained less of the phenotypic variation than those linked to the QTLs for fruit bunch, whole fruit and mesocarp components (Table 1). This is demonstrated for KY where marker SNPM01951 from the QTL interval 75.0 – 82.1 cM in LGDP15 showed an effect of only 0.07 (p = 0.013897). The average KY for the two genotypes *ab* and *aa* were 0.57 and 0.66 tonkg/hapalm/year, respectively, a difference of only 0.09 tonkg/hapalm/year (Figure 2 D~~E~~). Similar observations were made for MSW and MKW where the genotypes *ab* and *aa* of SNPM02999 (LGDP2) and EAGC/MCAA-302 (LGDP10) showed only a small difference of not more than 0.18 g (Figure 2 E and F~~, G~~). Additional QTLs for MSW and MKW were observed in LGs DP3 and DP16 where markers showing clear codominant segregating profiles were detected close to their QTL peaks. The SSR marker mEgCIR3301 had three alleles <*ab*x*ac*>, which segregated into four genotype classes - *ab*, *aa*, *bc* and *ac*. Interestingly, *ab* and *aa* showed lower phenotypic values than *bc* and *ac* (Figure 2 E and F~~, G~~). Another interesting marker was SNPM02704 at the QTL interval associated with MSW on LGDP16. The two parental palms showed the same genotype <*ab*x*ab*> and~~,~~ therefore, their parental effects and contribution to the trait could not be determined *via* GM. However, among the three observed genotypes, *bb* had the lowest MSW (0.79 ± 0.3 g) compared to *aa* (0.96 ± 0.2 g) and *ab* (1.10 ± 0.2 g) (Figure 2 F~~G~~).

In this study, QTL analysis also revealed a number of putative QTLs for YCs (Table 3). By permutating the entire 16 LGs, these QTLs had LOD scores lower than their GW significance thresholds but higher than their 95.0 % significant thresholds at the chromosome level. In this respect, three~~five~~ CW significant QTLs, termed putative, were identified for MBN, TOT and OY in LGDP2~~, and DPWT in LGs DP12 and DP15~~. Interestingly, these three production components are strongly related to each other~~ same genomic region was associated with multiple closely related YCs~~ (r = 0.798~~1~~ – 0.998~~7~~). ~~This was specifically observed in LGs DP2 and DP12. In LGDP2, the same genomic region hosted suggestive QTLs for MBN, TOT and OY, the three production components strongly related to each other (r = 0.79 – 0.99). Similarly, in LGDP12, marker SNPM04433, associated with the suggestive QTL for DPWT, was also located within the confidence intervals of the QTLs for OTB and OTWP. Similar genomic regions linked to different YCs were to be expected, especially for those strongly related (r = 0.63 – 0.99).~~ In oil palm, a common QTL interval on the genetic map for related YCs, such as OTB, OTF, STF, KTF and DMWM, was also reported by Jeennor and Volkaert (2014). Similarly, in other crops, clustering of QTLs was reported for fiber quality and various yield traits in cotton (Keerio et al. 2018), weight, length, diameter and peduncle length in tomato (Portis et al. 2014)~~, and~~ grain yield, harvesting index and grain weight in rice (Zhu et al. 2017) as well as maturity date, fruit development, fruit structure and the solid soluble content in sweet

cherry (Calle and Wünsch 2020). The co-localization of multiple QTLs suggests the presence of closely linked loci or pleiotropic genes (Billotte et al. 2010, Lemmon and Doebley 2014).

KULIM DxP: QTLs linked to YCs

In thise population, GW-significant QTLs were identified for nine YCs. The YCs with their associated QTLs and LGs were MBN and MFFB (LGDPK1), OTB (LGDPK8), OY and TOT (LGDPK1 and DPK8), KTB, KTF and MTF (LGDPK14) and STF (LGDPK4). A QTL was associated with MBN at interval 0 – 7.2 cM on LGDPK1, explaining ~15.9 % of the phenotypic variation for the trait. The QTL peak had LOD 5.1 and the closest marker was a SSR, mEgCIR3803, with four genotype classes among the progenies, namely *ac*, *ad*, *bc* and *bd*. Palms with the *ac* and *bc* genotypes had lower MBN of (12.61 ± 0.39) and (12.76 ± 0.38), respectively, than those with the *bd* (13.90 ± 0.33) and *ad* (14.85 ± 0.36) genotypes (Figure 3A). Within the same QTL interval, a smaller region (0.75 – 7.58 cM) was associated with MFFB, where the SNP marker, SNPM01086 was located closest to the QTL peak. In fact, MFFB is one of the most important traits that indicates the productivity of oil palm. This co-segregating <*ab*x*ab*> marker demonstrated that both the *aa* (157.92 ± 3.30 kg) and *ab* (156.56 ± 2.52 kg) genotypes contributed to significantly higher MFFB production than palms with the *bb* genotype (143.02 ± 4.28 kg). On LGDPK1, the slightly extended interval from 0.00 – 7.60 cM also hosted QTLs for OY and TOT, where the co-segregating marker SNPM01086 was closest to the QTL peak. Higher OY (6.1 ± 0.2 ton/ha/yearkg) and TOT (6.60 ± 0.1 ton/ha/yearkg) were observed for the *aa* than in the *ab* (5.8 ± 0.1 ton/ha/yearkg OY and TOT) and *bb* (5.24 ± 0.18 ton/ha/yearrkg OY and 5.76 ± 0.19 ton/ha/yearrkg TOT) genotypes.

The QTLs associated with OY and TOT were also identified on LGDPK8 (92.3 – 105.2 cM), with two SNP markers, SNPM02425 and SNPM02400, located closest to the QTL peaks, respectively. The OY-linked SNPM02425 showed a co-segregating profile <*ab*x*ab*>, *i.e.*, palms with the *bb* genotype had higher OY (6.18 ± 0.13 ton/ha/yearrkg) than those with *aa* (5.26 ± 0.2 ton/ha/yearkg) and *ab* (5.76 ± 0.1 ton/ha/yearkg). For the QTL associated with TOT, the maternally inherited marker SNPM02400 revealed significantly higher TOT (6.6 ± 0.1 ton/ha/yearkg) for the homozygous genotype (*aa*) than *ab* (5.83 ± 0.1 ton/ha/yearkg). Interestingly, SNPM02400 also pointed to another QTL associated with OTB located at the 101.1 – 103.4 cM interval. The *aa* genotype of this marker was also responsible for higher OTB (28.2 ± 0.2 %) than *ab* (26.8 ± 0.2 %) (Figure 3 C). The three YCs discussed above - OTB, OY and TOT - were significantly related with each another. Therefore, selection for higher OTB will also increase OY and TOT, although these three YC traits are highly influenced by the environment (Soh et al. 2017). The heritabilityies for the three YCs are low, so their breeding improvement will be highly dependent on the environment and general operational management of the trials. If the environment is unfavourable and operational management is poor, the gains from MAS will be tentative.

On LGDPK4, the QTL interval associated with STF was 3.5 – 16.2 cM. It explained 18.6 % of the phenotypic variation in STF and the closest marker to the QTL peak was SNPM00151, which revealed a marker effect of -0.73 % (heterozygous in the paternal palm). The heterozygous (*ab*) group showed a significantly lower STF (10.60 ± 0.19 %) than *aa* (12.06 ± 0.19 %). On DPK14, the QTLs for three highly correlated traits – KTF, KTB and MTF were found within the same map interval (46.9 – 64.8 cM). For KTF and KTB, the markers closest to the QTL peak (54.0 cM) were SNPM04522 and SNPM04938 which mapped on the same locus, indicating they had similar segregation profiles in the mapping family. The phenotypic variation explained by the QTL for KTF (18.8 %) was higher than that for KTB (21.1 %). Based on the

genotypes of both markers, higher KTF and KTB were observed for the *ab* (7.69 ± 0.13 % KTF and 5.20 ± 0.09 % KTB) than the homozygous *aa* genotype (6.70 ± 0.13 % KTF and 4.46 ± 0.09 % KTB). Within the same map interval, SNPM01100, located closest to the QTL peak (57.4 cM), accounted for 15.6 % of the MTF phenotypic variation. In contrast~~,~~ with KTF and KTB, the *aa* genotype of SNPM01100 showed significantly higher MTF (82.45 ± 0.31 %) than *ab* (80.4 ± 0.28 %). Interestingly, marker SNPM01100 was also significantly associated with KTF and KTB, although it was not closest to their QTL peaks. This indicates that within the QTL interval, this marker influences multiple traits differently depending on its genotype, which is supported by the significant correlations of KTF and KTB with MTF. This suggests that the genes that contribute to increased kernel size (larger KTF and KTB) will reduce mesocarp (MTF). So, selection for MTF will reduce KTF, boosting the mesocarp oil yield (Kushairi et al. 1999).

This study also identified a number of putative QTLs for various YCs on LGs DPK2 (OTDP), DPK4 (MFW, MPW, MSW and KY), DPK5 (MPW, MFW, OTB, OTWP and OTDP), DPK7 (OTWP~~M~~), DPK8 (MBN), DPK13 (KTF) and DPK14 (MKW, STF and KY). Information on the putative QTLs is summarized in Table 4.

Comparison of common QTLs between P2 and KULIM DxP

This study identified 4~~2~~5 QTLs (2~~1~~3 putative) in P2 and KULIM DxP, distributed across 1~~2~~3 LGs (~~01 to 16,~~ except 06, 09 and 11). Within each family, a number of the QTLs were co-localized on the same regions, such as on LGs DP01 (MFFB, TOT and OY), DP02 (MBN, OY and TOT) and DP12 (~~DPWT,~~ OTB and OTWP) in P2. In KULIM DxP, common QTLs were found on LGs DPK05 (MFW, MPW, OTB, OTDP and OTWP), DPK08 (OTB and TOT) and DPK14 (MTF and STF and; KTB and KTF). However, comparing P2 and KULIM DxP, only a few QTLs were detected in the same LGs for both. The QTLs on the same LGs were those associated with ~~WPWT and OTWP in LG07,~~ OTB, MBN, OY, TOT and MSW with OTDP in LG02, and MBN with KTF in LG13. However, the QTLs in the same LGs in P2 and KULIM DxP did not overlap, either in the genetic or physical map.

The lack of common QTLs in both families is likely due to differences in their genetic backgrounds, especially as their *pisifera* parents were different. The *pisifera* of P2 was Yangambi and ~~-~~that of KULIM DxP ~~-~~was AVROS, of quite separate origins. The KULIM DxP *pisifera* contributed most of the alleles that revealed the GW QTLs for OTB (LGDPK08), KTB, KTF, MTF (LGDPK14) and TOT (LGDPK01)~~.~~, ~~while t~~The maternal *dura*, as expected, contributed the alleles for the STF-related QTLs, as the shell trait is maternally inherited. However, in P2, the GW QTLs detected were contributed in equal numbers by both the ~~half and half from its~~ paternal and maternal parents. Its paternally inherited QTLs were those associated with MBN (LGDP13A), OTB, OTWP (LGDP12) and KY (LGDP15).

QTLs from different studies

The QTLs identified in this study were compared with 144 previously reported for several oil palm crosses (Billotte et al. 2010, ~~Singh et al. 2013,~~ Jeenor and Volkaert 2014, Pootakham et al. 2015, Seng et al. 2016, Teh et al. 2016, Bai et al. 2017, Ithnin et al. 2017). Comparison was also made to the QTLs already detected for MFW, MPW, STF, MTF and OTDP in P2 (Ting et al. 2018). The sequences of all the published QTL-linked markers were first mapped to the EG5 genome build to locate them in their pseudo-chromosomes. The results showed that most of the QTLs identified in our study were unique to P2 or KULIM DxP, and have not been reported in other oil palm crosses. Nevertheless, ~~two~~ genomic regions on CHR09 and CHR14 that hosted three QTLs in LGs DP7 and DP3, respectively, were common to those reported in different genetic backgrounds~~previously~~ (discussed below). And, another five QTLs detected in our study are located as close as 2,792 bp to the QTLs reported previously in CHR02, 06 and 15 (Figure 4).

In CHR02, marker SNPM00151, linked to the QTLs for STF and MSW, was located only ~236.4 kb away from the SSR marker sMg00022 that was reported to be associated with KB and KF by Seng et al. (2016). Interestingly, STF is positively related with both KB and KF, which explains why the same genomic region may influence both traits, In the window (2,092,554 – 2,328,938 bp) which encompasses both the QTL intervals, we identified two genes - *acyl-acyl carrier protein thioesterase* (*Acyl-ACP TE*) and *UDP-glycosyltransferase* (*UGT*) ~~–~~involved in the ~~–~~fatty acid (FA) biosynthesis and glycosylation modification, respectively, during fruit development and ripening (Pulsifer et al. 2014, Jing et al. 2011, Sun et al. 2017, Wu et al. 2017, Peng et al. 2020). In the oil palm fruit, the *Acyl-ACP TE* genes such as *FATA* and *FATB* encode protein that hydrolyse the FA acyl chains from ACPs. *FATA* is quite specific for~~to~~ unsaturated~~-~~ acyl ACPs e.g. C18:1-ACP for release of C18:1, and *FATB* for saturated acyl-ACPs, e.g. C16:0-ACP and C14:0-ACP for release~~production~~ of C16:0~~-~~ and C14:0~~-ACPs~~, respectively thus, playing essential roles in determining the FA composition of palm oil (Sambanthamurthi et al. 2000, Othman~~Abrizah~~ et al 2001~~, Sambanthamurthi et al. 2000~~). *UGT* is involved in anthocyanin glycosylation, the process of accumulating phenolic compounds which are responsible for the customary deep orange-to-red colour of oil palm exocarp. Based on their biological activities, the two genes have a direct impact on the composition of palm oil produced. However, their impact on the shell (and kernel) components, if any, require further investigation.

In CHR06, the marker EAGC/MCAA-302 closest to the QTL peak for MKW - was in the same QTL interval (37,012 – 38,280 kb) associated with PF and aBWT in a multi-parental DxP cross (Billotte et al. 2010). In the interval, a *valine-glutamine motif-containing protein* (*VQ*) was identified at chromosomal position 37,411,925 bp. In many plants, *VQ* has been reported to be responsive to biotic and abiotic stress, including pathogen infection, when interacting with the *WRKY* transcription factor (TF) (Chen et al 2012, Pecher et al. 2014, Liu et al. 2020). The specific interaction between the *VQ* motif FXhVQChTG (pfam05678) containing the gene *IKU1* and a *WRKY*, *MINI3*, reportedly controls endosperm growth and seed size in *Arabidopsis* (Wang et al. 2010). Therefore, *VQ* is a good candidate gene to investigate for its regulatory effect on kernel and seed in oil palm. Additional analysis of the MKW-QTL region revealed that *VQ* was flanked by *gibberellin 2-beta-dioxygenase* (*GA2OX*) and a *GATA* TF (*GATA*) the~~which~~ putative functions of which are summarized in Table 5. Interestingly, these genes are significantly differentially expressed in low- and high-yielding oil palm (Wong et al. 2017). Furthermore, *GATA* is known to regulate biological functions in various plant organs, including the flower and seed.

In CHR09, the genomic region corresponding to 74.8 – 84.5 cM on LGDP7 of P2 was previously reported to be associated with ~~WPWT was similar to those for~~ MTF and STF (Ting et al. 2018). The same genomic region was also associated with QTLs for ~~and~~ Bwt and Fwt which were identified in populations derived from Deli, La Me and Yangambi genetic backgrounds (Billotte et al. 2010). ~~This was supported by significant correlations between WPWT and MTF and STF in our YC data in P2 and KULIM DxP.~~ Although the correlations between MTF~~WPWT~~, Bwt and Fwt are not known, it is postulated that increased MTF (or decreased STF) will increase Fwt. A search for genes of interest was performed in the genomic region 8,208,977 to 9,198,501 bp, and two, *C3HC4-type zinc finger* TF (*RING finger*) and a *membrane-bound O-acyltransferase* (*MBOAT*), were shortlisted. In *Nicotiana benthamiana*, *RING finger* is in the chloroplasts and silencing it stops the growth of fruits (Wu et al. 2014). *MBOATs*, such as *diacylglycerol acyltransferase* (*DGAT*) and *lysophospholipid acyltransferase* (*LPLAT*), are involved in catalysing the synthesis and accumulation of lipids in developing seeds, including in the mesocarp of oil palm (Tranbarger et al. 2011, Li et al, 2013, Wang et al. 2012, Jin et al. 2017, Rosli et al. 2018).

The SSR marker mEgCIR3301 mapped to 6,491,270 bp in CHR14 was found associated to MKW in P2 and an DxP mapping family by Seng et al. (2016) as both families shared the same paternal parent (coded ML161). Interestingly, mEgCIR3301 was flanked by a lipid acylation-related gene, *glycerol-3-phosphate acyltransferase* (*GPAT*), at 6,480,850 bp and *WRI1*, at 6,510,932 bp. In many plants, including oil palm, *WRI1* has been reported to regulate genes encoding a number of key enzymes along the FA and triacylglycerol synthesis pathways (Maeo et al. 2009, Bourgis et al. 2011, Tranbarger et al. 2011, Chapman and Ohlrogge 2012, Qu et al. 2012, To et al. 2012, Vanhercke et al. 2013, Tajima et al. 2013, Grimberg et al. 2020, Kong et al. 2020). In fact, a wider group of genes, such as the sugar- and carbohydrate-responsive genes, are also reported to be regulated by *WRI1* (Masaki et al. 2005, Cernac et al. 2006). The storage compounds regulated by these genes eventually will affect development of the seed, embryo and even seedling, suggesting a possible role for *WRI1* in regulating MKW of oil palm.

Another common genomic region is the 19,804 – 20,124 kb interval on CHR15, which was associated with MTF and STF in KULIM DxP. The region was also reportedly linked to other important YCs, such as FFB, Fwt, Bwt and PO (Billotte et al. 2010). We identified a *pectinesterase* (*PME*) and a *small auxin-up RNA-like auxin-responsive protein* (*SAUR*) at 19,788,553 bp (to 19,805,976 bp) and 20,058,133 bp (to 20,059,096 bp), respectively. Both are related to cell metabolism, *PME* degrading pectin and modifying the cell wall in preparation for fruit ripening and softening, and *SAUR* involved in cell division, expansion and differentiation (Markakis et al. 2013, Abu-Sarra and Abu-Goukh 2015, Li et al. 2015, Wen et al. 2020). The presence of these genes in QTL regions influencing various bunch components suggests the importance of genes regulating cell wall development, cell division, expansion and differentiation for the appropriate development of all components in the fruit bunch. Extending the search beyond the common QTL regions (in CHR02, 06, 14 and 15), we also identified a number of –genes and TFs involved in the regulation of sugar levels, FA/oil biosynthesis, growth and development of flower, seed and fruit (Table 5), all of which potentially impact development of the bunch components.

Conclusion

This study describes the QTLs associated with yield components in two advanced *dura* x *pisifera* populations. Several common QTLs were identified in both populations. The QTLs linked to MTF~~WPWT~~ and OTWP in P2 and KULIM DxP that influence mesocarp formation, respectively, were located ~22,000 kb apart in CHR09 (LGDP/DPK07). In addition, another similar genomic region (~11,000 kb apart) in CHR08 (LGDP/DPK2) regulates OTB and OTDP in P2 and KULIM DxP, respectively, both directly contributing to oil yield. The QTLs associated with similar yield traits have been published previously in mapping populations of different genetic backgrounds. We collated all the information to identify the QTL regions influencing the related traits reported by the different studies in CHR02, 06, 09, 14 and 15. Search within and near the QTL regions in the different chromosomes revealed 29~~6~~ candidate genes and transcription factors related to glycosylation, plant growth, development and architecture, glucose and hormone signalling, lipid metabolism, photosynthesis, flowering and fruit ripening. *UGT*, *PG*, *MYB*, *NAC2*, *AUX/IAA*, *RING finger*~~*ZINGER*~~ and *PME* are example of genes potentially regulating oil palm fruit formation, thus directly impacting yield. The current genome-based candidate gene approach is useful in identifying interesting genes that can assist in further understanding the genetic control of oil palm yield. In fact, *GATA* gene ~~that~~ located within the QTL interval was shown previously to be differentially expressed in high- and low-yielding palms. Further validation of the association of the other candidate genes with the

traits concerned can help develop useful tools for marker assisted selection in oil palm breeding. The markers linked to the QTLs could also be candidates for developing an appropriate marker panel for genomic selection in oil palm.

Acknowledgments

The authors would like to thank the Director-General of the Malaysian Palm Oil Board (MPOB) for permission to publish this article. The authors also thank Mr. Andy Chang Kwong Choong for his critical comments and edits on the manuscript.

Funding

This study was funded by the Malaysian Palm Oil Board (MPOB).

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

References

Abu-Sarra AF, Abu-Goukh AA (1992) Changes in pectinesterase, polygalacturonase and cellulose activity during mango fruit ripening. J Hortic Sci Biotechnol 67:561–568. https://doi.org/10.1080/00221589.1992.11516284

Al'bert AV, Kavaĭ-ool UN, Ezhova TA (2014) The gene *NANA* regulates cell proliferation in *Arabidopsis thaliana* shoot apical meristem without interaction with *CLV1*, *CLV2*, *CLV3*. Russ J Dev Biol 45:326–32 https://doi.org/10.1134/S1062360414050026

Alamin M, Zeng D, Qin R et al (2017) Characterization and fine mapping of *SFL1*, a gene controlling screw flag leaf in rice. Plant Mol Biol Rep 35:491–503. https://doi.org/10.1007/s11105-017-1039-x

Bai B, Wang L, Lee M et al (2017) Genome-wide identification of markers for selecting higher oil content in oil palm. BMC Plant Biol 17:93. https://doi.org/10.1186/s12870-017-1045-z

Behringer C, Bastakis E, Ranftl QL et al (2014) Functional diversification within the family of *B-GATA* transcription factors through the leucine-leucine-methionine domain. Plant Physiol 166:293–305. https://doi.org/10.1104/pp.114.246660

Beyene Y, Semagn K, Mugo S et al (2015) Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci 55:154–163. https://doi.org/10.2135/cropsci2014.07/0460

Bhagya HP, Kalyana Babu B, Gangadharappa PM et al (2020) Identification of QTLs in oil palm (*Elaeis guineensis* Jacq.) using SSR markers through association mapping. J Genet 99:19. https://doi.org/10.1007/s12041-020-1180-4

Billotte N, Jourjon MF, Marseillac N. et al (2010) QTL detection by multi-parent linkage mapping in oil palm (*Elaeis guineensis Jacq.*). Theor Appl Genet 120:1673–1687. https://doi.org/10.1007/s00122-010-1284-y

Blaak G, Sparnaaij LD, Menendez T (1963) Methods of bunch analysis. Breeding and inheritance in the oil palm (*Elaeis guineensis Jacq.*) Part II. J W Afr Ins Oil Palm Res 4:146-155.

Borrenpohl D, Huang M, Olson E et al (2020) The value of early-stage phenotyping for wheat breeding in the age of genomic selection. Theor Appl Genet 133:2499–2520. https://doi.org/10.1007/s00122-020-03613-0

Bourgis F, Kilaru A, Cao X et al (2011) Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. Proc Natl Acad Sci USA 108:12527–12532. https://doi.org/10.1073/pnas.1106502108

Brøndum RF, Su G, Janss, L et al (2015) Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. J Dairy Sci 6:4107–4116. https://doi.org/10.3168/jds.2014-9005

Calle A, Wünsch A (2020) Multiple-population QTL mapping of maturity and fruit-quality traits reveals LG4 region as a breeding target in sweet cherry (*Prunus avium* L.). Hortic Res 7:127. https://doi.org/10.1038/s41438-020-00349-2

Cernac A, Andre C, Hoffmann-Benning S et al (2006) *WRI1* is required for seed germination and seedling establishment. Plant Physiol 141:745–757. https://doi.org/10.1104/pp.106.079574

Chan KL, Tatarinova TV, Rosli R. et al (2017) Evidence-based gene models for structural and functional annotations of the oil palm genome. Biol Direct 12:21. https://doi.org/10.1186/s13062-017-0191-4

Chapman KD and Ohlrogge JB (2012) Compartmentation of triacylglycerol accumulation in plants. J Biol Chem 287:2288–2294. https://doi.org/10.1074/jbc.R111.290072

Che R, Tong H, Shi B et al (2015) Control of grain size and rice yield by *GL2*-mediated brassinosteroid responses. Nat Plants 2:15195. https://doi.org/10.1038/nplants.2015.195

Chen L, Song Y, Li S et al (2012) The role of *WRKY* transcription factors in plant abiotic stresses. Biochim Biophys Acta 1819:120–128. https://doi.org/10.1016/j.bbagrm.2011.09.002

Cheng Y, Zhou Y, Yang Y et al (2012) Structural and functional analysis of *VQ* motif-containing proteins in Arabidopsis as interacting proteins of *WRKY* transcription factors. Plant Physiol 159:810–25. https://doi.org/10.1104/pp.112.196816

Chiang YH, Zubo YO, Tapken W et al (2012) Functional characterization of the *GATA* transcription factors *GNC* and *CGA1* reveals their key role in chloroplast development, growth, and division in Arabidopsis. Plant Physiol 160:332-48. https://doi.org/10.1104/pp.112.198705

Collard BCY, Jahufer MZZ, Brouwer JB et al (2015) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169–196. https//doi.org/10.1007/s10681-005-1681-5

Concibido VC, Denny R, Lange D et al (1997) Genome mapping of soybean cyst nematode resistance genes in 'Peking', PI 90763, and PI 88788 using DNA markers. Crop Sci 37:258–264. https://doi.org/10.2135/cropsci1997.0011183X003700010046x

Cooley MB, Yoder JI (1998) Insertional inactivation of the tomato polygalacturonase gene. Plant Mol Biol 38:521–30. https://doi.org/10.1023/A:1006086004262

Corley RHV, Tinker PB (2003) The Oil Palm. Blackwell Science, Oxford

Czolpinska M, Rurek M (2018) Plant glycine-rich proteins in stress response: an emerging, still prospective Story. Front Plant Sci 9:302. https://doi.org/10.3389/fpls.2018.00302

Dai N, Schaffer AA, Petreikov M et al (1995) *Arabidopsis thaliana* hexokinase cDNA isolated by complementation of yeast cells. Plant Physiol 108:879–80. https://doi.org/10.1104/pp.108.2.879

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. FOCUS 12:13

Duan P, Ni S, Wang J et al (2015) Regulation of *OsGRF4* by *OsmiR396* controls grain size and yield in rice. Nat Plants 2:15203. https://doi.org/10.1038/nplants.2015.203

Fallen BD, Allen FL, Kopsell DA et al (2015) Selective genotyping for marker assisted selection strategies for soybean yield improvement. Plant Genetics, Genomics and Biotechnology 2:95–119. https://doi.org/10.5147/pggb.v2i1.156

Ferrándiz C, Gu Q, Martienssen R, Yanofsky MF (2000) Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. Development 127:725–34

Granot D, Kelly G, Stein O et al (2014) Substantial roles of hexokinase and fructokinase in the effects of sugars on plant physiology and development. J Exp Bot 65:809–19. https://doi.org/10.1093/jxb/ert400

Grimberg Å, Wilkinson M, Snell P et al (2020) Transitions in wheat endosperm metabolism upon transcriptional induction of oil accumulation by oat endosperm *WRINKLED1*. BMC Plant Biol 20:235. https://doi.org/10.1186/s12870-020-02438-9

Gu Q, Ferrándiz C, Yanofsky MF et al (1998) The *FRUITFULL MADS-box* gene mediates cell differentiation during Arabidopsis fruit development. Development 125:1509–17

Gupta PK, Langridge P, Mir RR (2010) Marker-assisted wheat breeding: present status and future possibilities. Mol Breeding 26:145–161. https://doi.org/10.1007/s11032-009-9359-7

Hu Y, Chen L, Wang H et al (2013) Arabidopsis transcription factor *WRKY8* functions antagonistically with its interacting partner *VQ9* to modulate salinity stress tolerance. Plant J 74:730–745. https://doi.org/10.1111/tpj.12159

Hudson D, Guevara D, Yaish MW et al (2011) *GNC* and *CGA1* modulate chlorophyll biosynthesis and glutamate synthase (*GLU1/Fd-GOGAT*) expression in Arabidopsis. PLoS One 6:e26765. https://doi.org/10.1371/journal.pone.0026765

Isa Z, Kushairi A, Mohd Din A et al (2011) A critical re-examination of the method of bunch analysis in oil palm breeding – an update. Paper presented at the International Seminar on Breeding for Sustainability in Oil Palm. Kuala Lumpur, Malaysia. 19–42

Ithnin M, Xu Y, Marjuni M. et al (2017) Multiple locus genome-wide association studies for important economic traits of oil palm. Tree Genet Genomes 13:103. https://doi.org/10.1007/s11295-017-1185-1

Jack PL, Mayes S (1993) Use of molecular markers for oil palm breeding. II. Use of DNA markers (RFLPs). Oleagineux 48 1–8

Jarquín D, Crossa J, Xavier L et al (2014) A reaction norm model for genomic selection using high dimensional genomic and environmental data. Theor Appl Genet 127:595–607. https://doi.org/10.1007/s00122-013-2243-1

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 Genet Genomes 10:1–14. https://doi.org/10.1007/s11295-013-0655-3

Jin J, Sun Y, Qu J et al (2017) Transcriptome and functional analysis reveals hybrid vigor for oil biosynthesis in oil palm. Sci Rep 7:439. https://doi.org/10.1038/s41598-017-00438-8

Jing F, Cantu DC, Tvaruzkova J et al (2011) Phylogenetic and experimental characterization of an *acyl-ACP thioesterase* family reveals significant diversity in enzymatic specificity and activity. BMC Biochem 12:44. https://doi.org/10.1186/1471-2091-12-44

Jing R, Lin R (2015) The *VQ* motif-containing protein family of plant-specific transcriptional regulators. Plant Physiol 169:371-378. https://doi.org/10.1104/pp.15.00788

Keerio AA, Shen C, Nie Y et al (2018) QTL mapping for fiber quality and yield traits based on introgression lines Derived from *Gossypium hirsutum* × *G tomentosum.* Int J Mol 19:243. https://doi.org/10.3390/ijms19010243

Kong Q, Yang Y, Guo L et al (2020) Molecular basis of plant oil biosynthesis: Insights gained from studying the *WRINKLED1* transcription factor. Front Plant Sci 11:24. https://doi: 10.3389/fpls.2020.00024

Kushairi A, Loh SK, Azman I et al (2018) Oil palm economic performance in Malaysia and R&D progress in 2017. J Oil Palm Res 30:163–195. https://doi.org/10.21894/jopr.2018.0030

Kushairi A, Rajanaidu N (2000) Breeding populations, seed production and nursery management. In: Basiron Y, Jalani BS, Chan KW (ed) Advances in Oil Palm Research. Malaysian Palm Oil Board, Malaysia. pp 39–96

Kushairi A, Rajanaidu N, Jalani BS et al (1999) Agronomic performance and genetic variability of *dura* x *pisifera* progenies. J Oil Palm Res 11:1–24

~~Ithnin M, Xu Y, Marjuni M. et al (2017) Multiple locus genome-wide association studies for important economic traits of oil palm. Tree Genet Genomes 13:103. https://doi.org/10.1007/s11295-017-1185-1~~

~~Isa Z, Kushairi A, Mohd Din A et al (2011) A critical re-examination of the method of bunch analysis in oil palm breeding – an update. Paper presented at the International Seminar on Breeding for Sustainability in Oil Palm. Kuala Lumpur, Malaysia. 19–42~~

~~Jack PL, Mayes S (1993) Use of molecular markers for oil palm breeding. II. Use of DNA markers (RFLPs). Oleagineux 48 1–8~~

~~Jarquín D, Crossa J, Xavier L et al (2014) A reaction norm model for genomic selection using high dimensional genomic and environmental data. Theor Appl Genet 127:595–607. https://doi.org/10.1007/s00122-013-2243-1~~

~~Jin J, Sun Y, Qu J et al (2017) Transcriptome and functional analysis reveals hybrid vigor for oil biosynthesis in oil palm. Sci Rep 7:439. https://doi.org/10.1038/s41598-017-00438-8~~

~~Jing F, Cantu DC, Tvaruzkova J et al (2011) Phylogenetic and experimental characterization of an *acyl-ACP thioesterase* family reveals significant diversity in enzymatic specificity and activity. BMC Biochem 12:44. https://doi.org/10.1186/1471-2091-12-44~~

~~Kushairi A, Loh SK, Azman I et al (2018) Oil palm economic performance in Malaysia and R&D progress in 2017. J Oil Palm Res 30:163–195. https://doi.org/10.21894/jopr.2018.0030~~

~~Kushairi A, Ong-Abdullah M, Nambiappan B et al (2019) Oil palm economic performance in Malaysia and R&D progress in 2018. J Oil Palm Res 31:165–194. https://doi.org/10.21894/jopr.2019.0026~~

Lemmon ZH, Doebley JF (2014) Genetic dissection of a genomic region with pleiotropic effects on domestication traits in maize reveals multiple linked QTL. Genetics 198:345–353. https://doi:10.1534/genetics.114.165845

Li R, Hatanaka T, Yu K et al (2013) Soybean oil biosynthesis: role of *diacylglycerol acyltransferases*. Funct Integr Genomics 3:99-113. https://doi.org/10.1007/s10142-012-0306-z

Li S, Gao F, Xie K et al (2016) The *OsmiR396c-OsGRF4-OsGIF1* regulatory module determines grain size and yield in rice. Plant Biotechnol J 4:2134-2146. https://doi.org/10.1111/pbi.12569

Li ZG, Chen H-W, Li Q-T et al (2015) Three *SAUR* proteins *SAUR76*, *SAUR77* and *SAUR78* promote plant growth in Arabidopsis. Sci Rep 5:12477. https://doi.org/10.1038/srep12477

Lin SS, Martin R, Mongrand S et al (2008) RING1 E3 ligase localizes to plasma membrane lipid rafts to trigger *FB1*-induced programmed cell death in Arabidopsis 56:550-61. https://doi.org/10.1111/j.1365-313X.2008.03625.x

Liscum E, Reed JW (2002) Genetics of *Aux/IAA* and *ARF* action in plant growth and development. Plant Mol Biol 49:387–400. https://doi.org/10.1023/A:1015255030047

Liu C, Liu H, Zhou C et al (2020) Genome-wide identification of the VQ protein gene family of tobacco (*Nicotiana tabacum* L.) and analysis of its expression in response to phytohormones and abiotic and biotic stresses. Genes 11:284. https://doi.org/10.3390/genes11030284

Liu L, Du Y, Shen X et al (2015) *KRN4* ~~C~~controls ~~q~~Quantitative ~~v~~Variation in ~~m~~Maize ~~k~~Kernel ~~R~~row ~~n~~Number. PLoS Genet 11:e1005670. https://doi.org/10.1371/journal.pgen.1005670

~~Li S, Gao F, Xie K et al (2016) The *OsmiR396c-OsGRF4-OsGIF1* regulatory module determines grain size and yield in rice. Plant Biotechnol J 4:2134-2146. https://doi.org/10.1111/pbi.12569~~

~~Lin SS, Martin R, Mongrand S et al (2008) RING1 E3 ligase localizes to plasma membrane lipid rafts to trigger *FB1*-induced programmed cell death in Arabidopsis 56:550-61. https://doi.org/10.1111/j.1365-313X.2008.03625.x~~

Lo SF, Yang SY, Chen KT et al (2008) A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. Plant Cell 20:2603–2618. https://doi.org/10.1105/tpc.108.060913

Luo J, Zhou JJ, Zhang JZ (2018) *Aux/IAA* gene family in plants: molecular structure, regulation, and function. Int J Mol Sci 19:259. https://doi.org/10.3390/ijms19010259

Ma X, Zhang Y, Turečková V et al (2018) The *NAC* transcription factor *SlNAP2* regulates leaf senescence and fruit yield in Tomato. Plant Physiol 177:1286–1302. https://doi.org/10.1104/pp.18.00292

Maeo K, Tokuda T, Ayame A et al (2009) An *AP2-type* transcription factor, *WRINKLED1*, of *Arabidopsis thaliana* binds to the *AW*-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. Plant J 60:476–87. https://doi.org/10.1111/j.1365-313X.2009.03967.x

Mara CD, Irish VF (2008) Two *GATA* transcription factors are downstream effectors of floral homeotic gene action in Arabidopsis. Plant Physiol 147:707–18. https://doi.org /10.1104/pp.107.115634

Markakis MN, Boron AK, Van Loock B et al (2013) Characterization of a small auxin-up RNA (*SAUR*)-like gene involved in *Arabidopsis thaliana* development. PLoS ONE 8:e82596. https://doi.org/10.1371/journal.pone.0082596

~~Mara CD, Irish VF (2008) Two *GATA* transcription factors are downstream effectors of floral homeotic gene action in Arabidopsis. Plant Physiol 147:707–18. https://doi.org /10.1104/pp.107.115634~~

Masaki T, Mitsui N, Tsukagoshi H et al (2005) ACTIVATOR of spomin: *LUC1/WRINKLED1* of *Arabidopsis thaliana* transactivates sugar-inducible promoters. Plant Cell Physiol 46:547–56. https://doi.org/10.1093/pcp/pci072

~~Markakis MN, Boron AK, Van Loock B (2013) Characterization of a small auxin-up RNA (*SAUR*)-like gene involved in *Arabidopsis thaliana* development. PLoS ONE 8:e82596. https://doi.org/10.1371/journal.pone.0082596~~

Mayes S, Jack PL, Corley RHV et al (1997) Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). Genome 40:116–22. https//doi.org/10.1139/g97-016

Mishra KK, Fougat RS, Ballani A et al (2014) Potential and application of molecular markers techniques for plant genome analysis. Int J Pure App Biosci. 2:169–188

Mudgil Y, Ghawana S, Jones AM (2013) N-MYC down-regulated-like proteins regulate meristem initiation by modulating auxin transport and *MAX2* expression. PLoS One 8:e77863. https://doi.org/10.1371/journal.pone.0077863

Mudgil Y, Uhrig JF, Zhou J et al (2009) Arabidopsis *N-MYC DOWNREGULATED-LIKE1*, a positive regulator of auxin transport in a G protein-mediated pathway. Plant Cell 21:3591–609. https://doi.org/10.1105/tpc.109.065557

Muhammad Azhar Nadeem, Muhammad Amjad Nawaz, Muhammad Qasim Shahid et al (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol Biotechnol Equip 32:261–285. https://doi.org/10.1080/13102818.2017.1400401

Okada K, Saito T, Nakagawa T et al (2000) Five *geranylgeranyl diphosphate synthases* expressed in different organs are localized into three subcellular compartments in Arabidopsis. Plant Physiol 122:1045–1056. https://doi.org/10.1104/pp.122.4.1045

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

Okwuagwu CO, Okoye MN, Okolo, EC et al (2008) Genetic variability of fresh fruit bunch yield in Deli / *dura* x *tenera* breeding populations of oil palm (*Elaeis guineensis* Jacq.) in Nigeria. J Trop Agric 46:40–45

Ooi Leslie C-L, Low Eng-Ti L, Ong-Abdullah M~~Abdullah Meilina O~~ et al (2016) Non-*tenera* contamination and the economic impact of *SHELL* genetic testing in the Malaysian independent oil palm industry. Front Plant Sci 7:771. https://doi.org/10.3389/fpls.2016.00771

Osorio-Guarín JA, Garzón-Martínez GA, Delgadillo-Duran P et al (2019) Genome-wide association study (GWAS) for morphological and yield-related traits in an oil palm hybrid (*Elaeis oleifera x Elaeis guineensis*) population. BMC Plant Biol 19:533. https://doi.org/10.1186/s12870-019-2153-8

Osteryoung KW, Toenjes K, Hall B et al (1990) Analysis of tomato *polygalacturonase* expression in transgenic tobacco. Plant Cell 2:1239-48. https://doi.org/10.1105/tpc.2.12.1239

Othman A, Lazarus CM, Fraser T et al (2001) Cloning of a palmitoyl-acyl carrier protein thioesterase from oil palm. Biochem Soc Trans 28: 619–22. https://doi.org/10.1042/BST0280619

Pace J, Gardner C, Romay C et al (2015) Genome-wide association analysis of seedling root development in maize (*Zea mays L.*) BMC Genomics 16:47. https://doi.org/10.1186/s12864-015-1226-9

Paparelli E, Gonzali S, Parlanti S et al (2012) Misexpression of a *chloroplast aspartyl protease* leads to severe growth defects and alters carbohydrate metabolism in Arabidopsis. Plant Physiol 160:1237–50. https://doi.org/10.1104/pp.112.204016

Pecher P, Eschen-Lippold L, Herklotz S (2014) The *Arabidopsis thaliana* mitogen-activated protein kinases *MPK3* and *MPK6* target a subclass of '*VQ*-motif'-containing proteins to regulate immune responses. New Phytol 203:592–606~~0~~. https://doi.org/10.1111/nph.12817~~. 203(2) 592–606~~

Peng Z, Zhang H, Tian H et al (2020) The phylogeny and functional characterization of peanut acyl-acp thioesterases. J Plant Growth Regul 39:1381–1392. https://doi.org/10.1007/s00344-020-10073-9

Pirker J, Mosnier A, Kraxner F et al (2016) What are the limits to oil palm expansion? Global and Environ Chang. 40:73–81. https://doi.org/10.1016/j.gloenvcha.2016.06.007

Poland JA, Brown PJ, Sorrells ME et al (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253. https://doi.org/10.1371/journal.pone.0032253

Pootakham W, Jomchai N, Ruang-areerate P et al (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. https://doi.org/10.1016/j.ygeno.2015.02.002

Portis E, Barchi L, Toppino L et al (2014) QTL mapping in eggplant reveals clusters of yield-related loci and orthology with the tomato genome. PLoS ONE 9:e89499. https://doi.org/10.1371/journal.pone.0089499

Pulsifer IP, Lowe C, Narayaran SA et al (2014) Acyl-lipid thioesterase1-4 from *Arabidopsis thaliana* form a novel family of *fatty acyl-acyl carrier protein thioesterases* with divergent expression patterns and substrate specificities. Plant Mol Biol 84:549–63. https://doi.org/10.1007/s11103-013-0151-z

Qu J, Ye J, Geng YF et al (2012) Dissecting functions of *KATANIN* and *WRINKLED1* in cotton fiber development by virus-induced gene silencing. Plant Physiol 160:738–748. https://doi.org/10.1104/pp.112.198564

Ragimekula N, Varadarajula NN, Mallapuram SP et al (2013) Marker assisted selection in disease resistance breeding. J Plant Breed Genet 1:90–109

Rahimah, AB, Cheah SC, Rajinder S (2006) Freeze-drying of oil palm (*Elaeis guineensis*) leaf and its effect on the quality of extractable DNA. J Oil Palm Res 18:296–304

~~Rance K A, Mayes S, Price Z et al (2001) Quantitative trait loci for yield components in oil palm (*Elaeis guineensis* Jacq.). Theor App Genet 103:1302–1310. https://doi.org/10.1007/s122-001-8204-z~~

~~Ragimekula N, Varadarajula NN, Mallapuram SP et al (2013) Marker assisted selection in disease resistance breeding. J Plant Breed Genet 1:90–109~~

Rajanaidu N, Kushairi A, Rafii M et al (2000) Oil palm breeding and genetic resources. In: Basiron Y, Jalani BS, Chan KW (ed) Advances in Oil Palm Research. Malaysian Palm Oil Board, Malaysia. pp 171–227

Rance K A, Mayes S, Price Z et al (2001) Quantitative trait loci for yield components in oil palm (*Elaeis guineensis* Jacq.). Theor App Genet 103:1302–1310. https://doi.org/10.1007/s122-001-8204-z

Ren Y, McGregor C, Zhang Y et al (2014) An integrated genetic map based on four mapping populations and quantitative trait loci associated with economically important traits in watermelon (*Citrullus lanatus*). BMC plant biology 14:33. https://doi.org/10.1186/1471-2229-14-33

Richter R, Behringer C, Müller IK et al (2010) The *GATA*-type transcription factors *GNC* and *GNL/CGA1* repress gibberellin signaling downstream from *DELLA* proteins and *PHYTOCHROME-INTERACTING FACTORS*. Genes Dev 24:2093–104. https://doi.org/10.1101/gad.594910

Roongsattham P, Morcillo F, Jantasuriyarat C (2012) Temporal and spatial expression of *polygalacturonase* gene family members reveals divergent regulation during fleshy fruit ripening and abscission in the monocot species oil palm. BMC Plant Biol 12:150. https://doi.org/10.1186/1471-2229-12-150

Rosli R, Amiruddin N, Ab Halim MA et al (2018) Comparative genomic and transcriptomic analysis of selected *fatty acid biosynthesis* genes and *CNL* disease resistance genes in oil palm. PLoS ONE 13:e0194792. https://doi.org/10.1371/journal.pone.0194792

Sambanthamurthi R, Sundram K, Tan Y (2000) Chemistry and biochemistry of palm oil. Progress in Lipid Research 39: 507–58. https://doi.org/10.1016/S0163-7827(00)00015-1

Sanusi NS, Rosli R, Ab Halim MA et al (2018) PalmXplore: oil palm gene database. Database https://doi.org/10.1093/database/bay095

Sebastian SA, Streit LG, Stephens PA et al (2010) Context-specific marker-assisted selection for improved grain yield in elite soybean populations. Crop Sci 50:1196–1206. https://doi.org/10.2135/cropsci2009.02.0078

Seng T-Y, Faridah QZ, Ho CL et al (2007) Flanking AFLP markers for the *virescens* trait in oil palm. J Oil Palm Res 19:381–392

Seng T-Y, Ritter E, Mohamed Saad (2016) QTLs for oil yield components in an elite oil palm (*Elaeis guineensis*) cross. Euphytica 212:399–425. https://doi.org/10.1007/s10681-016-1771-6

Shan C, Mei Z, Duan J et al (2014) *OsGA2ox5*, a *gibberellin* metabolism enzyme, is involved in plant growth, the root gravity response and salt stress. PLoS One 9:e87110. https://doi.org/10.1371/journal.pone.0087110

Shockey J, Regmi A, Cotton K et al (2016) Identification of ~~a~~Arabidopsis *GPAT9* (At5g60620) as an essential gene involved in *triacylglycerol* biosynthesis. Plant Physiol 170:163–79. https://doi.org/10.1104/pp.15.01563

Shu K, Yang W (2017) *E3 Ubiquitin ligases*: ubiquitous actors in plant development and abiotic stress responses. Plant Cell Physiol 58:1461–1476. https://doi.org/10.1093/pcp/pcx071

Singer SD, Chen G, Mietkiewska E et al (2016) Arabidopsis *GPAT9* contributes to synthesis of intracellular glycerolipids but not surface lipids. J Exp Bot 67:4627-38. https://doi.org/10.1093/jxb/erw242

Singh R, Cheah SC (1999) Analysis of the inheritance of AFLP markers in an interspecific cross of oil palm using the pseudo-testcross strategy. J Oil Palm Res Special Issue:64–73

Singh R, Cheah SC (2005) Potential Application of Marker Assisted Selection (MAS) in Oil Palm. Oil Palm Bulletin 51: 1 – 9

~~Singh R, Ong-Abdullah M, Low ETL et al (2013a) Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. Nature 500:335–339. https://doi.org/10.1038/nature12309~~

Singh R, Low ETL, Ooi LCL (~~2013b~~2013a) The oil palm *SHELL* gene controls oil yield and encodes a homologue of *SEEDSTICK*. Nature 500:340-344. https://doi.org/10.1038/nature12356

Singh R, Ong-Abdullah M, Low ETL et al (2013a~~b~~) Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. Nature 500:335–339. https://doi.org/10.1038/nature12309

Singh R, Low ETL, Ooi L C-L et al (2014) The oil palm *VIRESCENS* gene controls fruit colour and encodes a *R2R3-MYB*. Nature Commun 5:4106. https://doi.org/10.1038/ncomms5106

Singh R, Low ETL, Ooi L C-L et al (2020) Variation for heterodimerization and nuclear localization among known and novel oil palm SHELL alleles. New Phytologist 226:426–440. https://doi.org/10.1111/nph.16387

Soh AC, Mayes S, Roberts JA (2017) Oil palm breeding: genetics and genomics. CRC Press, US

Spindel J, Begum H, Akdemir D et al (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genetics 11:e1005350. https://doi.org/10.1371/journal.pgen.1004982

Stewart-Brown BB, Song Q, Vaughn JN et al (2019) Genomic selection for yield and seed composition traits within an applied soybean breeding program. G3(Bethesda) 9:2253–2265. https://doi.org/10.1534/g3.118.200917

Sun Y, Ji K, Liang B et al (2017) Suppressing *ABA uridine diphosphate glucosyltransferase* (SlUGT75C1) alters fruit ripening and the stress response in tomato. Plant J 91:574–589. https://doi.org/10.1111/tpj.13588

Swaray S, Amiruddin MD, Rafii MY et al (2020) Influence of parental *dura* and *pisifera* genetic origins on oil palm fruit set ratio and yield components in their D x P progenies. Agronomy 10:1793. https://doi.org/10.3390/agronomy10111793

Tajima D, Kaneko A, Sakamoto M et al (2013) Wrinkled 1 (*WRI1*) Homologs, AP2-type transcription factors involving master regulation of seed storage oil synthesis in castor bean (*Ricinus communis L.*). Am J Plant Sci 4:333–339. https://doi.org/10.4236/ajps.2013.42044

Teh C-K, Ong A-L, Kwong Q-B et al (2016) Genome-wide association study identifies three key loci for high mesocarp oil content in perennial crop oil palm. Sci Rep 6:19075. https://doi.org/10.1038/srep19075

Teh C-K, Ong A-L, Mayes S et al (2020) Major QTLs for trunk height and correlated agronomic traits provide insights into multiple trait integration in oil palm breeding. Genes 11:826. https://doi.org/10.3390/genes11070826

Ting N-C, Jansen J, Mayes S et al (2014) High density SNP and SSR-based genetic maps of two independent oil palm hybrids. BMC Genomics 15:309. https://doi.org/10.1186/1471-2164-15-309

Ting N-C, Jansen J, Nagappan J et al (2013) Identification of QTLs associated with callogenesis and embryogenesis in oil palm using genetic linkage maps improved with SSR markers. PLoS ONE 8:e53076. https://doi.org/10.1371/journal.pone.0053076.

Ting N-C, Mayes S, Massawe F et al (2018) Putative regulatory candidate genes for QTL linked to fruit traits in oil palm (*Elaeis guineensis* Jacq.). Euphytica 214:214. https://doi.org/10.1007/s10681-018-2296-y

Ting N-C, Zulkifli Y, Kamaruddin K et al (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. https://doi.org/10.1186/s12864-016-2607-4

~~Ting N-C, Mayes S, Massawe F et al (2018) Putative regulatory candidate genes for QTL linked to fruit traits in oil palm (*Elaeis guineensis* Jacq.). Euphytica 214:214. https://doi.org/10.1007/s10681-018-2296-y~~(0123456789()., volV

To A, Joubès J, Barthole G et al (2012) *WRINKLED* transcription factors orchestrate tissue specific regulation of fatty acid biosynthesis in Arabidopsis. Plant Cell 24:5007–5023. https://doi.org/10.1105/tpc.112.106120

Tranbarger TJ, Dussert S, Joët T et al (2011) Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. Plant Physiol 156:564–584. https//doi.org/10.1104/pp.111.175141

~~Qu J, Ye J, Geng YF et al (2012) Dissecting functions of *KATANIN* and *WRINKLED1* in cotton fiber development by virus-induced gene silencing. Plant Physiol 160:738–748. https://doi.org/10.1104/pp.112.198564~~

Van Ooijen, J.W. (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

Vanhercke T, El Tahchy A, Shrestha P et al (2013) Synergistic effect of *WRI1* and *DGAT1* coexpression on triacylglycerol biosynthesis in plants FEBS Lett 587:364–9. https://doi.org/10.1016/j.febslet.2012.12.018

~~Van Ooijen, J.W. (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~~

Wang A, Garcia D, Zhang H (2010) The *VQ* motif protein *IKU1* regulates endosperm growth and seed size in Arabidopsis. Plant J 63:670–9. https//doi.org/10.1111/j.1365-313X.2010.04271.x

Wang B, Lin Z, Li X et al (2020) Genome-wide selection and genetic improvement during modern maize breeding. Nat Genet 52:565–571. https://doi.org/10.1038/s41588-020-0616-3

Wang L, Shen W, Kazachkov M et al (2012) Metabolic interactions between the Lands cycle and the Kennedy pathway of glycerolipid synthesis in Arabidopsis developing seeds. Plant Cell 24:4652–69. https://doi.org/10.1105/tpc.112.104604

Wang T, Tohge T, Ivakov A et al (2015) Salt-related *MYB1* coordinates abscisic acid biosynthesis and signaling during salt stress in Arabidopsis 169:1027–41. https://doi.org/10.1104/pp.15.00962

Wang X, Yang X, Zhongli H et al (2018) Genomic selection methods for crop improvement: current status and prospects. Crop J 6:330–340. https://doi.org/10.1016/j.cj.2018.03.001

~~Wang L, Shen W, Kazachkov M et al (2012) Metabolic interactions between the Lands cycle and the Kennedy pathway of glycerolipid synthesis in Arabidopsis developing seeds. Plant Cell 24:4652–69. https://doi.org/10.1105/tpc.112.104604~~

~~Wang A, Garcia D, Zhang H (2010) The *VQ* motif protein *IKU1* regulates endosperm growth and seed size in Arabidopsis. Plant J 63:670–9. https//doi.org/10.1111/j.1365-313X.2010.04271.x~~

Watson CF, Zheng L, DellaPenna D (1994) Reduction of tomato *polygalacturonase beta* subunit expression affects pectin solubilization and degradation during fruit ripening. Plant Cell 6:1623–34. https://doi.org/10.1105/tpc.6.11.1623

Wen B, Zhang F, Wu X et al (2020) Characterization of the tomato (*Solanum lycopersicum*) pectin methylesterases: evolution, activity of isoforms and expression during fruit ripening. Front Plant Sci 11:238. https://doi:10.3389/fpls.2020.00238

Wong YC, The HF, Mebus K et al (2017) Differential gene expression at different stages of mesocarp development in high- and low-yielding oil palm. BMC Genomics 18. https://doi.org/10.1186/s12864-017-3855-7

Wu B, Gao L, Gao J et al (2017) Genome-wide identification, expression patterns, and functional analysis of *UDP glycosyltransferase* family in peach (*Prunus persica* L. Batsch) Front Plant Sci 8:389 https://doi.org/10.3389/fpls.2017.00389

Wu W, Cheng Z, Liu M et al (2014) *C3HC4-type RING* finger protein *NbZFP1* is involved in growth and fruit development in *Nicotiana benthamiana*. PLoS ONE 9:e99352. https://doi.org/10.1371/journal.pone.0099352

Xiang L, Le Roy K, Bolouri-Moghaddam MR et al (2011) Exploring the neutral invertase-oxidative stress defence connection in *Arabidopsis thaliana*. J Exp Bot 62:3849–62. https://doi.org/10.1093/jxb/err069

Yousef G, Juvik J (2001) Comparison of phenotypic and marker-assisted selection for quantitative traits in sweet corn. Crop Sci 41:645–655. https://doi.org/10.2135/cropsci2001.413645x

Zaki NM, Singh R, Rosli R et al (2012) *Elaeis oleifera* genomic-SSR markers: exploitation in oil palm germplasm diversity and cross-amplification in *arecaceae*. Int J Mol Sci 13:4069–88. https://doi.org10.3390/ijms13044069

Zhai Z, Liu H, Shanklin J (2017) Phosphorylation of *WRINKLED1* by *KIN10* results in its proteasomal degradation, providing a link between energy homeostasis and lipid biosynthesis. Plant Cell 29:871–889. https://doi.org/10.1105/tpc.17.00019

Zhu M, Liu D, Liu W et al (2017) QTL mapping using an ultra-high- density SNP map reveals a major locus for grain yield in an elite rice restorer R998. Sci Rep 7:10914. https://doi.org/10.1038/s41598-017-10666-7

Zulkifli Y, Rajinder S, Mohd Din A et al (2014) Inheritance of SSR of SSR and SNP loci in an oil palm interspecific hybrid backcross (BC2) population. J Oil Palm Res 206:203–213

Authors responses to Reviewer's comments:

Reviewer #1: The manuscript has lot of scientific lacuna. Following major points need to clarified

1. The parents used in the mapping population are have high variation for the traits under study? I don't think the parents have variation for all the traits. In linkage mapping the parents used to generate mapping should vary for the traits. How authors can do GWAS without following the simple logic in linkage mapping studies. It is a very important criteria for mapping QTLs

Response: We thank the reviewer for this very important observation. The parent palms used to generate the two mapping families are the maternal Deli *dura* (in both cases) and AVROS and Yangambi *pisifera* (paternal parent). It is widely acknowledged by oil palm breeders that these parental palms, namely the Deli *dura* as well as the AVROS and Yangambi *pisifera* have significant variation in yield components (YCs). In fact, *pisifera* is female sterile and does not produce fruits that develop to maturity and hence, has no YCs associated with it. For crossing programmes, the *pisifera* palm is often selected based on the performance of its siblings (*tenera* that has fruit bunches), to indicate its yield potential. In a nutshell, the *pisifera* palms have no YCs directly associated with them, while the *dura* palms are selected for having favourable YCs. Thus, the *pisifera* and *dura* palms do vary in all aspects of YCs such as bunch weight, fruit-to-bunch ratio, kernel size and shell thickness. This has been well documented in literature e.g. Kushairi and Rajanaidu (2000). We have added a sentence in the **Materials and methods** section on this (under Mapping families, lines 187 – 192, page 5). As such, in this study, the phenotypic variance observed in the 16 YCs (presented in Supplementary Table S1) does reflect segregation of the parental palms, where the intraspecific hybrid populations are known to show hybrid vigor compared to both parents.

2. The English language should be improved

Response: We would like to thank Reviewer 1 for pointing this out. We have edited the entire text to further improve the English. Changes made can be viewed in Main text with tracked changes.docx.

3. The brevity of the abstract can be improved

Response: Thanks again for pointing this out. We have rewritten the abstract to avoid repetition.

4. Validation of the results must. So, the results must be validated.

Response: We agree with Reviewer 1. One of the validation approaches we used in this study was to compare the QTLs identified in the populations utilized to those observed across different genetic

backgrounds published previously. Our present results showed a handful of QTLs were common or located closely to those reported previously whereas, most of the QTLs identified were unique to P2 or KULIM DxP (**Results and discussion:** QTLs from different studies, lines 423 – 489, page 11 – 12 and Figure 4). However, the families used in this study form the important populations and the parental palms will be further improved to develop next generation of oil palm. As the breeding programme takes 10 – 12 years, the QTLs identified will be tested in the next generation as well as to determine stability of the QTL-linked markers in predicting the traits.

5. Many recent references are there. May be included.

Response: We have updated the references with more recent publications in 2019 and 2020 throughout the text.

6. What about the replications. Since major phenotypic data involved, replication data is must.

Response: This is a very important question and the authors agree that the quality of the phenotypic data will have a strong influence on the accuracy of marker-trait association. In oil palm breeding trials including the populations utilized in this study, the yield data (including the measurements for yield related components) is collected over a period of ~5 years or longer, starting at (or after) 6th year after planting in the open-field. The main reason for determining yield after the 6th year is to ensure consistency and reliability of the phenotypic data, as after the 6th year, oil palm fresh fruit production is more stable/consistent compared to the younger (< 6 years) plantings (Harun and Noor 2002, Corley and Tinker 2016). Within the data collection years, a minimum of three ripe bunches (replicates) per palm per year are normally sampled for bunch analysis according to the standard protocol practiced by oil palm breeders which has been well documented (Blaak et al. 1963, Rao et al. 1983, Isa et al. 2011). This is the standardized procedure used by the oil palm plantations/companies in Malaysia (and the rest of the world) for measuring yield and its related parameters. The standardized protocol for determination of yield parameters is also spelled out in the Malaysian national standards (MS157), which determine if parental palms are suitable for commercial seed production. We have described this in **Materials and methods:** Yield-related phenotypic data (lines 197 – 207, page 5 – 6). We have also highlighted in **Results and discussion:** Yield components (YCs) and correlations between them (lines 279 – 280, page 7) and Supplementary Table S1, that in this study, the average number (or replicates) of bunches analyzed per palm per year (MBN) was 13 bunches for both P2 and KULIM DxP families, respectively.