



# Identification and Validation of Candidate Genes Involved in Fatty Acid Content in Oil Palm by Genome-Wide Association Analysis

Wei Xia<sup>1†</sup>, Tingting Luo<sup>3†</sup>, Yajing Dou<sup>1†</sup>, Wei Zhang<sup>3</sup>, Annaliese S. Mason<sup>4</sup>, Dongyi Huang<sup>1</sup>, Xiaolong Huang<sup>1</sup>, Wenqi Tang<sup>1</sup>, Jihua Wang<sup>5\*</sup>, Chunyu Zhang<sup>3</sup> and Yong Xiao<sup>2\*</sup>

<sup>1</sup>Institute of Tropical Agriculture and Forestry, Hainan University, Haikou, China, <sup>2</sup>Coconut Research Institute, Chinese Academy of Tropical Agricultural Sciences, Wenchang, China, <sup>3</sup>National Research Center of Rapeseed Engineering and Technology and College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, <sup>4</sup>Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, Giessen, Germany, <sup>5</sup>Guangdong Key Laboratory for Crops Genetic Improvement, Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangdong, China

## OPEN ACCESS

### Edited by:

Hanwei Mei,  
Shanghai Agrobiological Gene  
Center, China

### Reviewed by:

Francesca Taranto,  
Council for Agricultural and  
Economics Research, Italy  
Jesse Poland,  
Kansas State University,  
United States

### \*Correspondence:

Jihua Wang  
wangjihua@gdaas.cn  
Yong Xiao  
xiaoyong1980@catas.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 14 February 2019

**Accepted:** 11 September 2019

**Published:** 15 October 2019

### Citation:

Xia W, Luo T, Dou Y, Zhang W,  
Mason AS, Huang D, Huang X,  
Tang W, Wang J, Zhang C and Xiao Y  
(2019) Identification and Validation  
of Candidate Genes Involved in Fatty  
Acid Content in Oil Palm by Genome-  
Wide Association Analysis.  
*Front. Plant Sci.* 10:1263.  
doi: 10.3389/fpls.2019.01263

Oil palm (*Elaeis guineensis*) is the highest yielding oil crop per unit area worldwide, but its oil is considered unhealthy for human consumption due to its high palmitic acid content (C16:0). In order to facilitate breeding for fatty acid content in oil palm, genome-wide association analysis (GWAS) was used to identify and validate single-nucleotide polymorphism (SNP) markers and underlying candidate genes associated with fatty acid content in a diversity panel of 200 oil palm individuals. A total of 1,261,501 SNP markers previously developed using SLAF-seq (specific locus amplified fragment sequencing) were used for GWAS. Based on this analysis, 62 SNP markers were significantly associated with fatty acid composition, and 223 candidate genes were identified in the flanking regions of these SNPs. We found one gene (acyl-ACP thioesterase B genes) that was involved in fatty acid biosynthesis and that was associated with high palmitic acid content in the mesocarp. Over-expression of this gene caused a significant increase in palmitic acid content. Our study provides key loci that can be used for breeding oil palm cultivars with low palmitic acid content.

**Keywords:** *Elaeis guineensis*, FatB, fatty acid, genome-wide association analysis, palmitic acid

## INTRODUCTION

Oil palm (*Elaeis guineensis*, 2n = 32) is an important tropical oil crop. Among all oil crops, oil palm has the highest oil yield per unit area: its oil yield is more than 6 times that of peanut, 8 times that of soybean, and 10 times that of rapeseed (Huang, 2017). The world production of palm oil in 2017 was approximately 65 million tonnes (<http://faostat3.fao.org/home/E>). The two oil storage tissues of oil palm are the mesocarp and kernel, each of which produces oil with a different fatty acid composition. Palm oil usually refers to the oil extracted from the oil palm mesocarp, where palmitic acid (16:0) is the major fatty acid (50%). By contrast, in kernel oil, lauric acid (12:0) is the major fatty acid (50%). A principal objective in oil palm breeding is to decrease the palmitic acid content (16:0) and to increase the oleic acid (18:1) and linoleic acid content (18:2) in the oil product. Recently, by using conventional breeding methods, researchers in Malaysia have developed an oil palm variety

with high oleic acid (approximately 52%), although this oleic acid content is still far lower than that in peanut (>75% in high oleic acid varieties) or rapeseed (>80%) (Rajanaidu et al., 2017). Improving fatty acid composition via conventional breeding, however, is a slow process in oil palm because of the plant's long life cycle. The identification of genes controlling oil composition and the subsequent production of molecular markers linked to low palmitic acid content would be extremely helpful in accelerating the breeding and selection of oil palm.

In *Arabidopsis*, substantial research has been done to elucidate genes involved in the fatty acid biosynthesis pathway, which occurs mainly in the plastids and endoplasmic reticulum. Initially, acetyl CoA carboxylase (ACCase) catalyzes the conversion of the precursor converacetyl-CoA into malonyl-CoA (Sasaki et al., 1997). Catalyzed by ketoly-ACP synthase (KAS), malonyl-CoA is subsequently polymerized at a frequency of two carbons per cycle into the acyl carbon chain and is combined with acyl carrier protein (ACP) (Choi et al., 2000). Finally, the elongation of the carbon chain is terminated as catalyzed by acyl-CoA thioesterase (FAT) (Hunt et al., 2002). Previous research has shown that FAT exists as FATA and FATB (Salas and Ohlrogge, 2002). FATA was thought to be able to bring about the termination of C18:0-ACP and C18:1-ACP (Mekhedov et al., 2002). FATB, however, may also be involved in the termination of saturated fatty acyl-ACP (Bonaventure et al., 2004). In *Arabidopsis thaliana*, overexpression of the *AtFatB1* gene can significantly increase palmitic acid content. The disruption of *FatB* expression resulted in a 56% reduction in palmitic acid and a 50% reduction of stearate in *Arabidopsis* seeds (Wilson et al., 2001; Buhr et al., 2002). Moreover, enhanced expression of *FatB* from *Umbellularia californica* was observed to significantly increase the lauric acid content (C12:0) in *Brassica napus* (Voelker et al., 1996) (Figure 1).

Association mapping is an efficient method for identifying molecular markers associated with agronomic traits based on natural populations with or without pedigree relationships (Flint-Garcia et al., 2003). The resolution of an association study

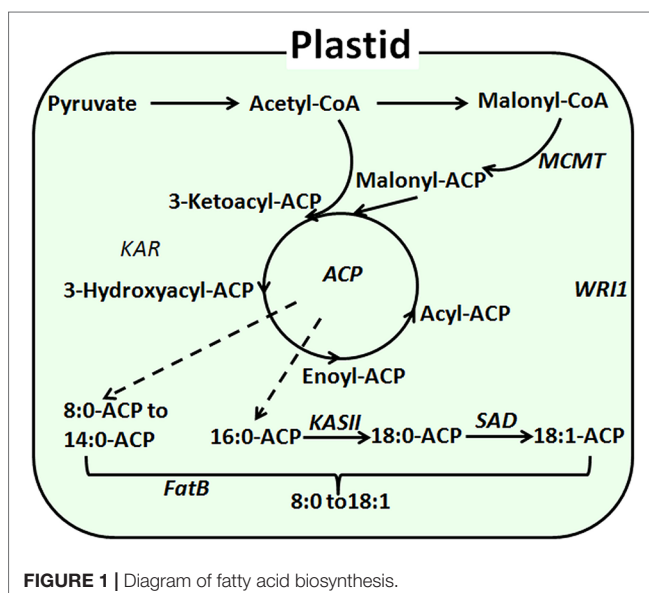
depends on the extent and structure of linkage disequilibrium (LD) in the selected population. With the development of sequencing technology and high-throughput single-nucleotide polymorphism (SNP) markers, association mapping has been widely applied in different crops to dissect complex agronomic traits (Atwell et al., 2010; Huang et al., 2011; Li et al., 2013). However, there is still the problem of how to reduce the detection of spurious associations between traits and markers resulting from population structure. One solution is to perform association analysis between allelic and phenotype variation in a less structured population (Pritchard et al., 2000). Another solution for structured populations is to apply a mixed-model approach, which can decrease spurious associations (Sorkheh et al., 2008). Compared with QTL (quantitative trait loci) mapping in biparental cross populations, association analysis using natural populations is more time- and cost-effective, especially for woody perennial crops that have long life cycles and/or that require large planting areas. In oil palm, QTL mapping has been used to resolve the genetic bases of several complex traits, including yield components (Rance et al., 2001; Billotte et al., 2010; Jeennor and Volkaert, 2014; Tisne et al., 2015), fatty acid composition (Singh et al., 2009; Montoya et al., 2013), sex ratio (Ukoskit et al., 2014), and embryogenesis (Ting et al., 2013). Unfortunately, the contribution of QTL mapping studies to marker-assisted breeding outcomes has been less than expected. Although association mapping has been validated to be a reliable method for identifying trait-associated markers for marker-assisted selection (MAS), this method has rarely been applied to oil palm.

In this study, we used genome-wide association analysis (GWAS) analysis to identify SNP markers associated with fatty acid contents and identified candidate genes adjacent to these SNP markers based on our previous characterization of the genes involved in lipid metabolism pathways in oil palm (Xiao et al., 2019). We then analyzed the expression patterns of these candidate genes in different tissues. To validate the role of top candidate genes in regulating palmitic oil content, we transferred these genes into *A. thaliana* and over-expressed them. Our results provide a comprehensive understanding of fatty acid content in oil palm, and the SNPs and candidate genes detected will facilitate breeding for fatty acid content in oil palm.

## MATERIALS AND METHODS

### Extraction and Measurement of Fatty Acid Contents in Oil Palm Mesocarp Tissues

Three fruits per oil palm individual (three biological replicates) were harvested, and fatty acid extraction and analyses for each mesocarp tissue were performed in triplicate (three different extractions as technical replicates). Approximately 60 mg of mesocarp (or 5 mg *Arabidopsis* seed) was used for extracting fatty acid according to methods described by Li-Beisson et al. (2013). Fatty acid composition was subsequently examined and measured using gas chromatography (Agilent DB-23, 30 m × 250). The heating procedure was as follows: the initial temperature was 180°C, followed by a temperature increase to 220°C at a rate



of 10°C increase per minute. The gas pressure was 17.392 psi. The nine values (three biological replicates × three technical replicates) obtained per oil palm individual were averaged for subsequent association mapping. The contents of decanoic acid (C10:0), lauric acid (12:0), myristic acid (C14:0), tripalmitelaidin acid (16:1), palmitic acid (16:0), stearic acid (18:0), oleic acid (18:1), linoleic acid (18:2), and oil were determined with reference to an internal standard. The calculation formula was as follows: 1) total fatty acid = (total area of all measure fatty acid peaks × quantity of heptadecanoic acid-methyl ester)/(peak area of heptadecanoic acid-methyl ester × quantity of quantity of the sample); 2) relative fatty acid percentage = peak area of a specific fatty acid/total area of all measured fatty acid peaks. The peak area was calculated using the Agilent software.

## SNP Genotyping

A total of 1,261,501 reliable SNPs markers (minor allele frequency > 0.05 and integrity > 0.8) were previously identified based on SLAF-seq (specific locus amplified fragment sequencing) by an Illumina HiSeq™ 2500 in a diversity panel of 200 oil palm individuals (Xia et al., 2019). The SLAF tag sequences were mapped to the African oil palm genome (Singh et al., 2013, the assembled oil palm genome of Version 5) using a mapping software: the Burrows-Wheeler alignment tool (Li and Drubin, 2009).

## Population Structure

Bayesian clustering was applied to analyze the population structure of 200 oil palm individuals using the software STRUCTURE (Pritchard et al., 2000). Based on the same set of SNPs, the number of subgroups (*K*) was predicted from 1 to 10, and the number of ancestors was determined according to the position of the minimum value, with an error rate obtained from 5-fold cross-validation. Maximum likelihood estimates for the ancestry proportion from each *K* subgroup for each accession were calculated. Five subgroups were indicated based on cross-validation errors (Xia et al., 2019).

## Association Mapping

To reduce spurious associations between alleles and phenotypic variation, mixed linear models (MLM) were used. Fixed effects were calculated using a *Q* (population) value matrix, and random effects were computed using a *K* (kinship) matrix. The *Q*+*K* value matrix was added to the MLM model. The *Q* matrix was obtained using STRUCTURE software (Pritchard et al., 2000), and the *K* matrix (genetic relationships among the 200 oil palm individuals) was obtained using SPAGeDi software (Hardy and Vekemans, 2002). Genome-wide association was performed using Tassel 5.0 (Bradbury et al., 2007). *P*-values for associations between SNP markers and fatty acid content were computed using the following formula:

$$y = X\beta + S\alpha + Qv + Zu + e$$

where  $X\beta$  represents the fixed effects other than the SNP under testing and the population structure;  $y$  is a vector of phenotypic

observation;  $\alpha$  is a vector of SNP effects;  $v$  is a vector of population effects;  $u$  is a vector of polygene background effects;  $e$  is a vector of residual effects;  $Q$  is a matrix from background effects relating  $y$  to  $v$ ; and  $X$ ,  $S$ , and  $Z$  are incidence matrices of 1s and 0s relating  $y$  to  $\beta$ ,  $\alpha$ , and  $u$ , respectively (Yu et al., 2006). Quantile–quantile plots (Q–Q plots) were drawn using the “ggplot2” software R package (Ginestet, 2011), and the Manhattan plots was drawn using the “qqman” software package (Turner et al., 2014). A total of 237,714 SNP markers, which had no missing data in the 200 oil palm individuals, were selected for association analysis. The threshold value ( $P < 4.2e^{-7}$ ,  $-\log_{10} P$  approximately equal to 6.3) was set for detecting reliable trait-associated markers at a cut-off FDR (false discovery rate) of 0.1 (Benjamini and Hochberg, 1995). Meanwhile, when FDR was set as 0.01, the threshold value of *P* value and  $-\log_{10} P$  are  $4.2e^{-8}$  and 7.3, respectively.

## Prediction of Candidate Genes Associated With Fatty Acid Content

All genes in LD blocks ( $r^2 > 0.6$ ) containing SNPs that were significantly associated with traits were identified for further candidate gene selection (Raman et al., 2015). The LD blocks present in the 200 oil palm individuals were estimated using the software Haploview v4.2. The number and size of the LD blocks on each chromosome were calculated according to previously established methods (Barrett et al., 2005). If SNP markers significantly associated with traits were located outside LD blocks, candidate genes were selected for further analysis following the criteria used by Zhou et al. (2017). Amino sequences of these selected genes were aligned using BLAST against the *Arabidopsis* protein database to predict the potential function of candidate genes. All candidate genes were selected based on gene ontology (GO) terms related to fatty acid biosynthesis and metabolism, and all transcription factors were selected based on Clusters of Orthologous Group of proteins (COG) within SNP-tagged genome regions.

## Transcriptome Data Downloaded and RPKM Calculation

We downloaded transcriptomic raw read data from the SRA (Short Read Archive) database, including SRR851069 (mesocarp 10 weeks after anthesis), SRR851067 (mesocarp 15 weeks after anthesis), SRR190699 (mesocarp 17 weeks after anthesis), SRR190700 (mesocarp 19 weeks after anthesis), SRR190701 (mesocarp 21 weeks after anthesis), SRR190702 (mesocarp 23 weeks after anthesis), SRR851070 (kernel 10 weeks after anthesis), SRR851068 (kernel 15 weeks after anthesis), SRR190703 (leaf), SRR851071 (root), SRR851103 (shoot), SRR851101 (female flower), and SRR851099 (pollen). RPKM (reads per kb per million reads) were used to calculate gene expression levels using the following formula (Mortazavi et al., 2008):

$$RPKM = \frac{10^6 C}{NL / 10^3}$$

where  $C$  is the number of reads that aligned exclusively with one expressed sequence,  $N$  is the total number of reads that aligned

with all expressed sequences, and L is the number of bases in the corresponding coding sequence.

## Vector Construction and Transformation for Over-Expression of *EgFatB*

Among the candidate genes, acyl-ACP thioesterase B genes *EgFatB1* and *EgFatB2* were significantly associated with high palmitic acid contents in the mesocarp. Primers used for *EgFatB* gene cloning were designed using the Snapgene Viewer software (Table 1). Polymerase chain reaction (PCR) amplifications were performed in 50  $\mu$ l reaction mixtures containing a 500 ng cDNA sample from the oil palm mesocarp, 1  $\times$  PCR buffer, 2 mM MgCl<sub>2</sub>, 5 U of TaqDNA polymerase (TaKaRa, China), 0.5  $\mu$ M of each primer, and 0.2 mM dNTP mix. The PCR program included denaturation at 98°C for 30 s, followed by 30 cycles at 98°C for 10 s, 55°C for 30 s, and 72°C for 60 s, and a final extension at 72°C for 5 min. The PCR products were electrophoretically visualized on a 1% agarose gel and recombined into the *pBinGlyRed3* vector (containing DsRed as a reporter gene and 35S as a promoter), which was then digested with EcoR I and XhoI. The constructed vectors were transformed into competent *Escherichia coli* cells (line Dh51), and inserts were validated by sequencing. A plasmid with the correct sequence insert was transformed into *Agrobacterium GV3101* and then transformed into *A. thaliana* using an *Agrobacterium*-mediated *in planta* transformation approach (Clough et al., 1998). The detailed protocols were as follows: the transformed *Agrobacterium* strain was streaked on solid LB medium containing Gent and Kana antibiotic

overnight (28°C) and then a single validated colony was propagated in the liquid LB medium (28°C, 250 r/min) for 16 h. The turbid *Agrobacterium* liquid (300 ml) was then centrifuged at a rotation speed of 5000 r/min for 20 min, and the precipitation was suspended by a solution with 5% sucrose and 0.001% SilvetL-77 after removing the liquid LB medium. *Arabidopsis* inflorescences at the 3–4 day post-bolting stage were immersed in the *Agrobacterium* suspension and covered with plastic wrap for 16 h in the dark. After 1 week, these steps were repeated to increase transformation efficiency. Positive transgenic *Arabidopsis* seeds were confirmed by detection of red autofluorescence and by PCR validation of targeted genes. Control and transgenic *Arabidopsis* plants were grown with a 16 h light/8 h dark photoperiod at 25°C. Fatty acid composition of transformed *Arabidopsis* plants was examined with the same protocol as was used for oil palm mesocarp tissue.

## RESULTS

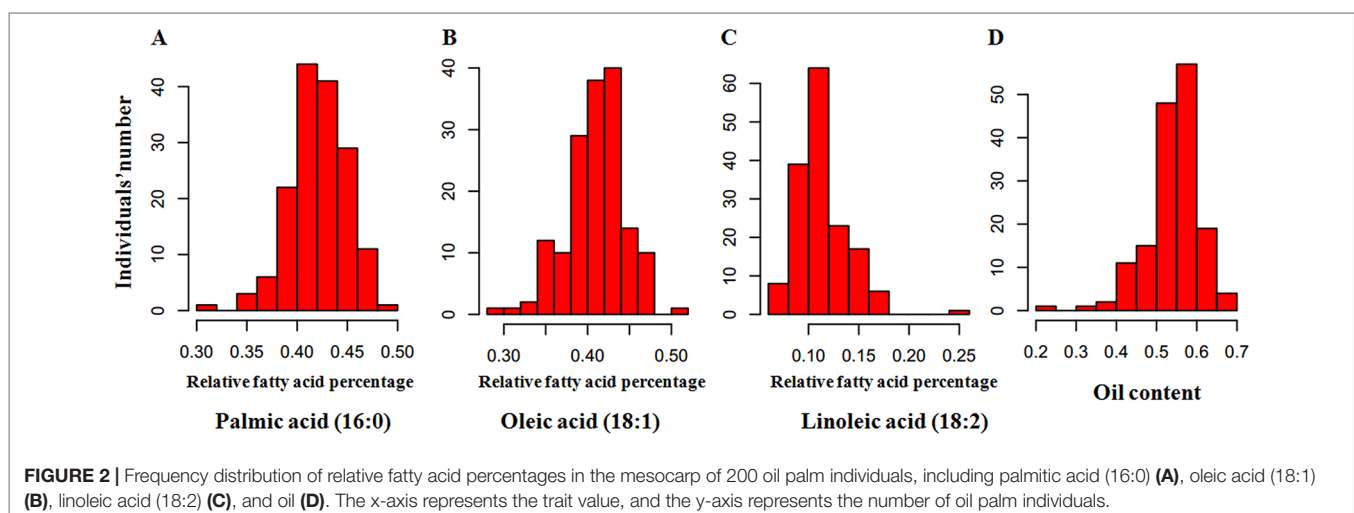
### Fatty Acid Composition in the Oil Palm Population

To perform an association analysis between SNP markers and relative fatty acid percentage, we first determined the relative fatty acid percentage of each of the 200 oil palm individuals from different geographical origins (Figure 2). Among the 200 oil palm individuals, relative palmitic acid percentage ranged from 31.3 to 48.8% with an average of 42.0%, oleic acid ranged from 31.3 to 50.1%, linoleic acid ranged from 7.1 to 18.5%,

**TABLE 1** | Primer sequences used to amplify the full CDS sequence of FatB1 and FatB2 in oil palm (*E. guineensis*).

	Forward primer	Reverse primer
FatB1	<b>CCGGAATT</b> CAGTGTCTCCATATCCCCATC	CCGCTCGAGTCAGTATTTCAAACGCAACA

Nucleotides in bold represent restriction enzyme sites for EcoR I and Xho I.



and total oil content ranged from 29.8 to 70.3%. Relative palmitic acid percentage (C16:0) was negatively associated with stearic acid ( $r = -0.627$ ;  $P < 0.001$ ) and relative oleic acid content ( $r = -0.657$ ;  $P < 0.001$ ). Relative oleic acid percentage was also negatively correlated with linoleic acid content ( $r = -0.689$ ;  $P < 0.001$ ).

## Genome-Wide Associations Between SNPs and Fatty Acid Content

To detect associations between allele variation and fatty acid content, we used MLM for GWAS; this controlled for the presence of population structure in the analyzed oil palm population. Meanwhile, five subgroups were previously indicated based on cross-validation errors (Xia et al., 2019) (see **Supplementary Table 1** for Q values). The MLM analysis revealed 62 SNP markers that were significantly associated ( $P < 4.2e^{-7}$ ) with fatty acid content, including palmitic acid content (32 SNPs), oleic acid content (4 SNPs), linoleic acid content (1 SNP), and total oil content (25 SNPs) (**Figures 3 and 4**).

## Identification of Candidate Genes Related to Fatty Acid Composition in *E. guineensis*

Of the 62 SNP markers significantly associated with fatty acid content, 6 were located in different LD blocks ( $r^2 > 0.6$ ), while the other SNP markers were not located in defined LD blocks. A total of 223 candidate genes were identified in the flanking regions of SNP markers that were significantly associated with different fatty acid contents (**Supplementary Table 2**). Based on GO annotation results, one candidate gene was involved in fatty acid biosynthesis: an acyl-ACP thioesterase B gene (*FatB*, involved in the termination of the fatty acid chain).

## Candidate Gene Expression in Different Oil Palm Tissues

The expression levels of candidate genes in fatty acid biosynthesis and metabolism pathways in different tissues and at different mesocarp developmental stages were estimated using RPKM values based on the method of Mortazavi et al. (2008). One candidate gene involved in fatty acid biosynthesis and metabolism was identified based on GO annotation results,

and this gene showed higher expression levels in the mesocarp tissues than in other tissue types (**Figure 5A**). The RPKM values for this gene were 41.8 and 37.9 at 10 and 15 weeks post-anthesis.

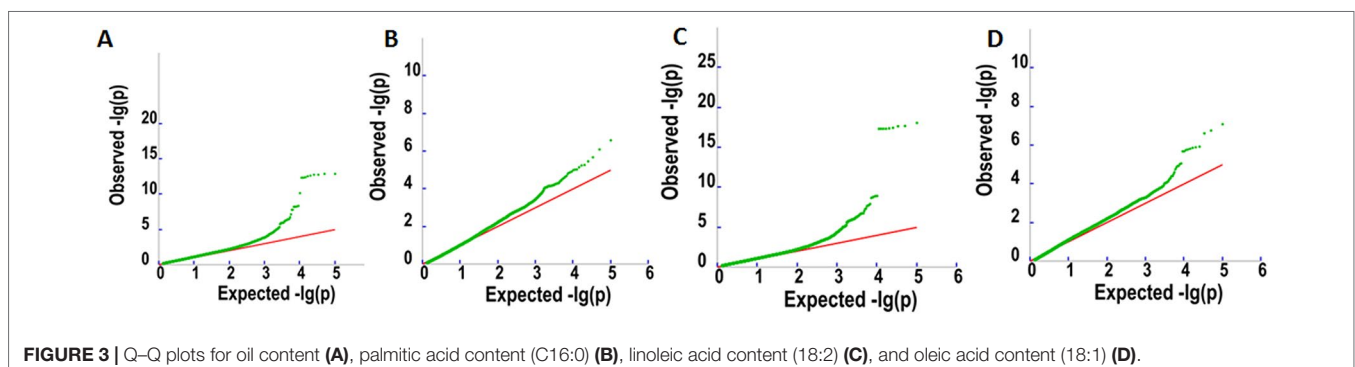
One *FatB* gene was detected as near to one significant SNP (EgChr7, 1731788). *FatB* genes have been reported to catalyze the termination of saturated fatty acyl-ACP. Interestingly, the expression level of *EgFatB1* was generally higher in the mesocarp than in the kernel, especially at 10 and 15 weeks post-anthesis.

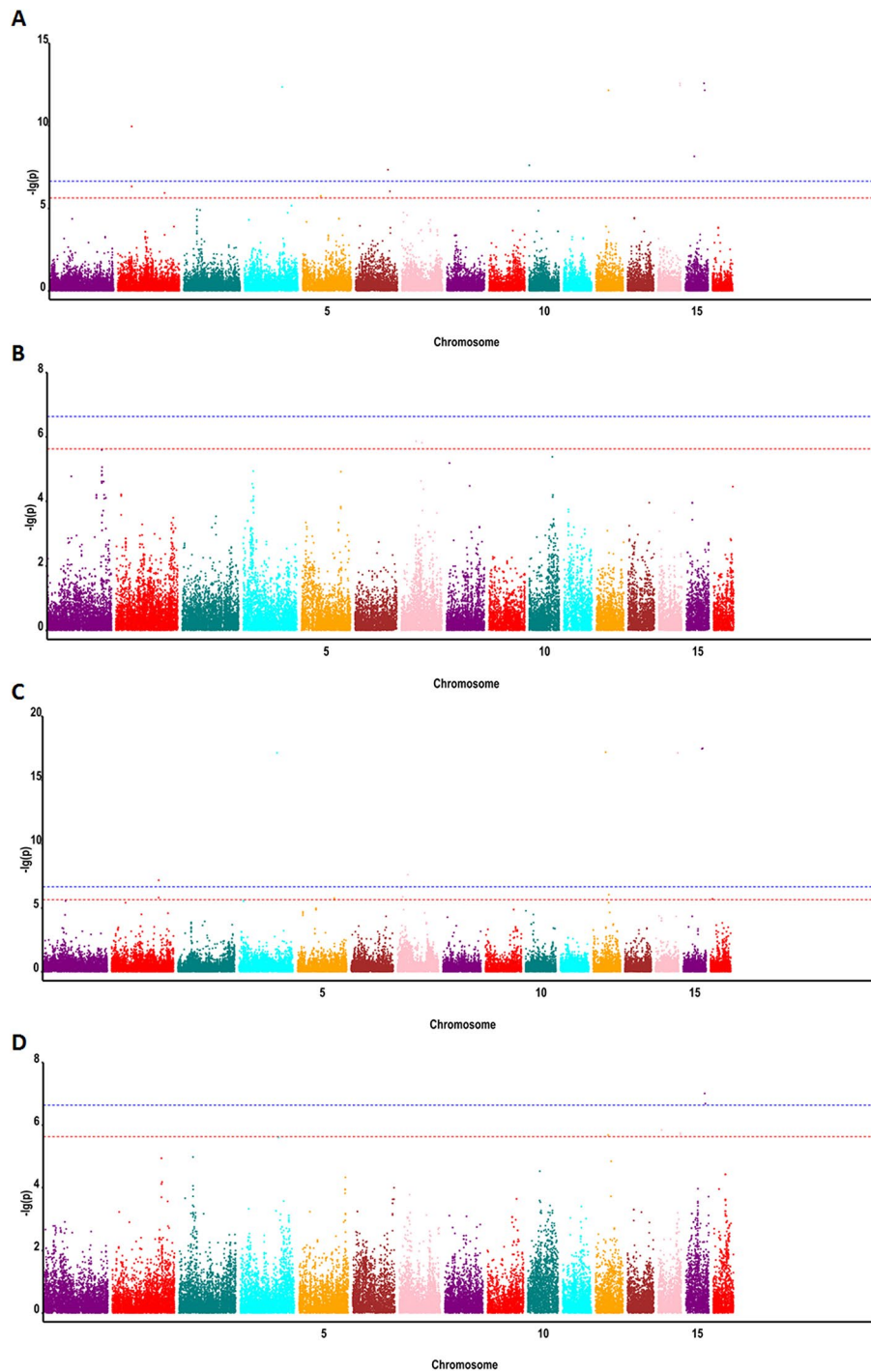
## Candidate Gene Expression Profiles in Different Oil Palm Individuals

In order to analyze the expression levels of candidate genes in different oil palm individuals with different fatty acid contents, we performed RNA-Seq on eight oil palm individuals. Four oil palm individuals were chosen on the basis of higher palmitic acid content (45.8%, 43.6%, 43.9%, and 41.3%) and showed *FatB* gene RPKM values of 52.8, 133.6, 3.45, and 49.9, respectively (**Figure 5B**). Four additional oil palm individuals were chosen on the basis of lower palmitic acid content (37.1%, 32.5%, 33.1%, and 30.3%) and showed *FatB* gene RPKM values of 37.1, 12.2, 29.9, and 11.6, respectively. The average expression level of *FatB* in oil palm individuals with higher palmitic acid content was two-fold higher than that of individuals with lower palmitic acid content.

## Functional Validation of *Egfatb1*

The full length coding sequence of *EgFatB1* was cloned and inserted into the pBinGlyRed3 vector. Genes inserted into this vector are promoted by a 35S promoter and in the same frame as the red autofluorescence reporter gene (*DsRed3*). The *EgFatB1* over-expression plasmid was transformed into *Arabidopsis*. Positive transformants were screened in the T1 generation by detection of red autofluorescence in the seeds and by PCR validation. Positive transgenic plants in the T2 generation were investigated for fatty acid contents. Transgenic plants with over-expressed *EgFatB1* increased the palmitic acid and stearic acid content but decreased the oleic acid and linoleic acid content in the seeds (**Figure 6**).



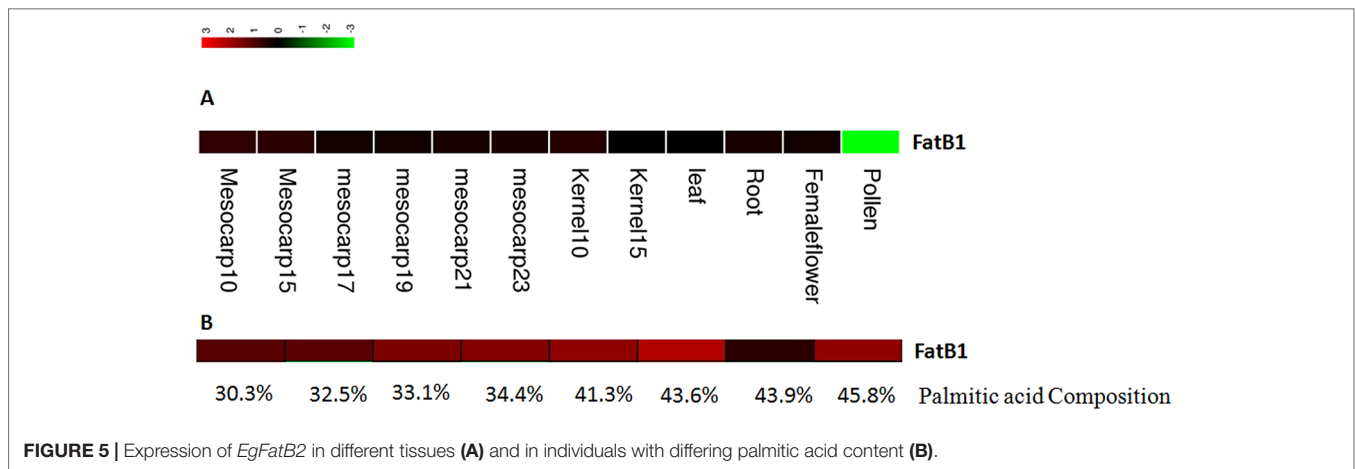


**FIGURE 4 |** Genome-wide associations between SNP markers and fatty acid contents in oil palm (*E. guineensis*) using MLM: **(A)** Manhattan plot for oil content, **(B)** Manhattan plot for palmitic acid content (C16:0), **(C)** Manhattan plot for linoleic acid content (18:2), and **(D)** Manhattan plot for oleic acid content (18:1). Red (6.3) and blue (7.3) lines indicate P-value cutoffs FDR of 0.1 and 0.01, respectively.

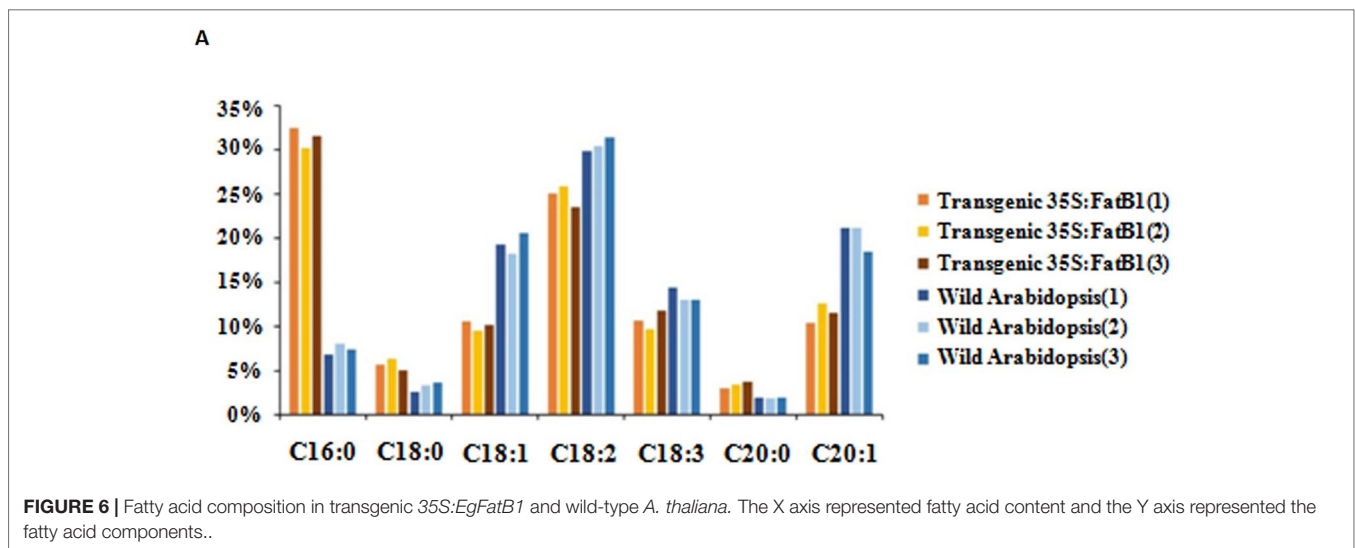
## DISCUSSION

Oil palm (*E. guineensis*) is an important crop for human nutrition, with tens of millions of tonnes of palm oil consumed every year. Although breeding could potentially improve the

nutritional value of palm oil for human consumption (e.g. by lowering the saturated fat content), breeding is hindered by the lack of genetic and genomic resources available for this crop. Using SLAF-seq technology, we genotyped a diversity panel of 200 oil palm individuals from four countries to obtain 1.2 million



**FIGURE 5 |** Expression of *EgFatB2* in different tissues (A) and in individuals with differing palmitic acid content (B).



**FIGURE 6 |** Fatty acid composition in transgenic *35S:EgFatB1* and wild-type *A. thaliana*. The X axis represented fatty acid content and the Y axis represented the fatty acid components..

genome-wide SNPs (Xia et al., 2019). We associated these SNPs with fatty acid content and identified dozens of candidate genes involved in fatty acid biosynthesis and metabolism pathways. One gene, *EgFatB1*, which was highly and specifically expressed in the mesocarp, was validated using *Arabidopsis* transformation: over-expression of *EgFatB1* resulted in increased saturated fat content in *Arabidopsis* seeds.

Oil palm (*E. guineensis*) originated from West and Southwest Africa, especially in the area between Angola and Gambia. Oil palm is well-known for its highest oil yield per unit area. Palm oil is an edible vegetable oil and derived from the mesocarp (reddish pulp) of the oil palm fruit. Along with coconut oil, palm oil is semi-solid at room temperature due to a high content of saturated fatty acid, especially palmitic acid (C16:0), which accounts for approximately 50% of palm oil fatty acid content. However, how fatty acid composition varies in different oil palm germplasm is unknown. In this study, we investigated the variation of fatty acid composition among 200 oil palm individuals. A large variation in fatty acid composition was detected between these oil palm individuals. Our study identified useful oil palm germplasm for improving fatty acid composition, especially for decreasing

palmitic acid content (lowest value: 31.3%) and increasing oleic acid content (highest value: 50.1%).

GWAS was firstly used as a powerful method to identify a broad range of complex diseases in human populations (Burton et al., 2007). Subsequently, with the completion and availability of additional genome sequences, this approach was used in flowering plants, including rice (Huang et al., 2011), foxtail millet (Jia et al., 2013), maize (Li et al., 2013), and *Arabidopsis* (Atwell et al., 2010). There are, however, some limiting factors in association analysis, such as insufficient marker density and population structure effects (Cappa et al., 2013). In a previous study, SLAF technology was used to develop a large number of SNP markers and to evaluate potential population structure, thus addressing these limiting factors and providing useful markers for MAS in oil palm. Palm oil accounts for 25% of vegetable oil traded worldwide annually and uses 5% of the global oil planting area. However, breeding progress in this tropical species is still slow, mostly due to the long life cycles of individual plants. Currently, only five to six generations of phenotype-based selection have been completed since the 1920s and 1930s (Hartley, 1967).

MAS has major potential to speed up the breeding of oil palm cultivars and reduce the length of conventional breeding cycles. Identifying trait-associated markers is a prerequisite for MAS. In the last decades, molecular markers have been used to identify QTL associated with important agronomic traits. Most marker development is based on conventional methods that result in modest density, including restriction fragment length polymorphism (RFLP) (Mayes et al., 1997), amplified fragments length polymorphism (AFLPs) (Singh et al., 2009), and simple sequence repeat (SSR) marker types (Billotte et al., 2010). To increase marker density, SNP markers were developed in oil palm. These high density SNP markers were successfully used in our study to identify molecular markers associated with important agronomic traits.

In oil palm, the first genetic lineage map was constructed using RFLP markers (Rance et al., 2001). QTL mapping was subsequently used to decipher the genetic bases of yield components (Rance et al., 2001; Billotte et al., 2010), fatty acid composition (Singh et al., 2009; Montoya et al., 2013), and sex ratio (Ukoskit et al., 2014), among other traits. These QTL mapping experiments involved populations produced from one to four parents. Montoya et al. (2013) reported 19 QTLs associated with fatty acid composition in palm oil using a backcross population of *Elaeis oleifera* with *E. guineensis*. Acyl-ACP thioesterase A and Stearoyl-ACP desaturase genes were found located in these QTL intervals. However, the confidence interval of these QTLs ranged from 6 to 30 cM, or approximately 6–30 Mbp of physical genome sequence in oil palm (Montoya et al., 2013). Teh et al. (2016) identified three key loci for high mesocarp oil content in oil palm, which were located on Chr5, Chr9, and Chr11. In our study, we identified 62 SNP markers ( $P < 4.2e^{-7}$ ) significantly associated with fatty acid content on Chr3, Chr5, and Chr11. Meanwhile, we also identified SNP markers associated with fatty acid on Chr1, Chr6, Chr10, Chr12, and Chr13. The *EgFatB1* gene was identified on Chr6 and validated to be associated with variation in palmitic acid content.

In our study, some SNP markers were associated with variation in fatty acid content, which is useful for selecting low palmitic acid genotypes in oil palm breeding and subsequently improving the nutritional value of palm oil. Moreover, there have been some previous examples where fatty acid composition in different crops has been improved using various biotechnology and mutation approaches. In rapeseed, the relative oleic acid content was increased to 71% by EMS and  $^{60}\text{Co}$  mutagenesis (Rucker and Robbelen, 1995). In soybean, oleic acid content

increased to 85% in a *fad2-1* mutant (Wagner et al., 2011). Meanwhile, in peanut, Norden et al. (1987) bred a high oleic acid cultivar (F435) from a natural mutant (Norden et al., 1987). Hence, artificial mutagenesis in oil palm could produce germplasm with significant divergence in fatty acid composition. In the present study, *FatB* was validated as a major effect gene for relative palmitic acid content, allowing us to screen for low palmitic acid content in oil palm by selecting for *FatB* mutations. The availability of molecular markers associated with agronomic traits will allow breeders to rapidly identify target traits in seedlings, thus accelerating selection and breeding schemes. Meanwhile, molecular markers associated with agronomic traits can also be used to detect the introgression of the trait of interest into elite varieties from more diverse germplasm.

## DATA AVAILABILITY STATEMENT

The raw SLAF-seq data for the 200 oil palm individuals has been deposited into the European Nucleotide Archive (<http://www.ebi.ac.uk/ena>). The bioproject number is PRJEB26466.

## AUTHOR CONTRIBUTIONS

YX, WX, and CZ participated in the design of the study. YX and WX performed the statistical analysis and drafted the manuscript. JW and YD were involved in analysis of the added transcriptome data. AM critically revised the manuscript. TL and WZ did the major experimental work including the extraction and measurement of oil content and relative fatty acid contents. WT, YD, DH, and XH contributed to and advised on the statistical analysis. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the Scientific and Technological Cooperation Projects of Hainan province (No. KJHZ2015-06). AM is funded by DFG Emmy Noether grant MA6473/1-1.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01263/full#supplementary-material>

## REFERENCES

- Atwell, S., Huang, Y. S., Vilhjalmsón, B. J., Vilhjalmsón, B. J., Willem, G., Horton, M., et al. (2010). Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631. doi: 10.1038/nature08800
- Barrett, J. C., Fry, B., Maller, J., Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21 (2), 263–265. doi: 10.1093/bioinformatics/bth457
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Billotte, N., Jourjon, M. F., Marseillac, N., Berger, A., Flori, A., Asmady, H., et al. (2010). QTL detection by multi-parent linkage mapping in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* 120, 1673–1687. doi: 10.1007/s00122-010-1284-y
- Bonaventure, G., Bao, X., Ohlrogge, J., and Pollard, M. (2004). Metabolic response to the reduction in palmitate caused by disruption of the *FatB* gene in *Arabidopsis*. *Plant Physiol.* 135, 1269. doi: 10.1104/pp.104.043372
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 223, 2633–2635. doi: 10.1093/bioinformatics/btm308



- Buhr, T., Sato, S., Ebrahim, F., Xing, A., Zhou, Y. M., Mathiesen, M., et al. (2002). Ribozyme termination of RNA transcript down-regulate seed fatty acid genes in transgenic soybean. *Plant J.* 30, 155–163. doi: 10.1046/j.1365-313X.2002.01283.x
- Burton, P. R., Clayton, D. G., Cardon, L., and Craddock, N. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 961–967. doi: 10.1038/nature05911
- Cappa, E. P., El-Kassaby, Y. A., Garcia, M. N., Acuna, C., Borralho, N. M. G., Grattapaglia, D., et al. (2013). Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. *PLoS One* 8, e81267. doi: 10.1371/journal.pone.0081267
- Choi, K. H., Heath, R. J., and Rock, C. O. (2000). Beta-ketoacyl-acyl carrier protein henthase III (FabH) is a determining factor in branched-chain fatty acid biosynthesis. *J. Bacteriol.* 182, 365–370. doi: 10.1128/JB.182.2.365-370.2000
- Clough, S. J., Bent, A. F., and Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313X.1998.00343.x
- Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 354–357. doi: 10.1146/annurev.arplant.54.031902.134907
- Ginestet, C. (2011). Ggplot 2: elegant graphics for data analysis. *J. R. Stat. Soc. A.* 174, 245–246. doi: 10.1111/j.1467-985X.2010.00676\_9.x
- Hardy, O. J., and Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2, 618–620. doi: 10.1046/j.1471-8286.2002.00305.x
- Hartley, C. W. S. (1967). *In the oil palm*. Ch.2 Vol. 2. London: Blackwell Professional, 1–36.
- Huang, H. (2017). The industrial survey of *Elaeis guineensis* in Malaysia. *World Trop. Agric. Inf.* 7, 7. doi: 10.3969/j.issn.1009-1726.2017.07.020
- Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., et al. (2011). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* 44, 32–39. doi: 10.1038/ng.1018
- Hunt, M. C., Solaas, K., Kase, B. F., and Alexson, S. E. (2002). Characterization of an acyl-CoA thioesterase that functions as a major regulator of peroxisomal lipid metabolism. *J. Biol. Chem.* 277, 1128–1138. doi: 10.1074/jbc.M106458200
- Jeennor, S., and 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. doi: 10.1007/s11295-013-0655-3
- Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., and Li, W. (2013). A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat. Genet.* 45, 957–961. doi: 10.1038/ng.2673
- Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., et al. (2013). Acyl-lipid metabolism. *Arabidopsis Book* 11 (e0161), 30–33. doi: 10.1199/tab.0161
- Li, H., and Drubin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Peng, Z., Yang, X., Wang, X., Fu, J., Wang, J., et al. (2013). Genome-wide association study dissects the genetic architecture of biosynthesis in maize kernel. *Nat. Genet.* 45, 43–50. doi: 10.1038/ng.2484
- Mayes, S., Jack, P. L., Corley, R. H. V., and Marshall, D. F. (1997). Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). *Genome* 40, 116–122. doi: 10.1139/g97-016
- Mekhedov, S., de Ilarduya, O. M., and Ohlrogge, J. (2002). Toward a functional catalog of the plant genome a survey of genes for lipid biosynthesis. *Plant Physiol.* 122, 389–402. doi: 10.1104/pp.122.2.389
- Montoya, C., Lopes, R., Flori, A., Cros, D., Cuellar, T., Summe, M., et al. (2013). Quantitative trait loci (QTLs) analysis of palm oil fatty composition in an interspecific pseudo-backcross from *Elaeis oleifera* (HBK) Cortes and oil palm (*Elaeis guineensis* Jacq.). *Tree Genet. Genomes* 9, 1207–1225. doi: 10.1007/s11295-013-0629-5
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Norden, A. J., Gorbet, D. W., and Knauff, D. A. (1987). Variability in oil quality among peanut genotypes in the Florida breeding program. *Peanut Sci.* 14, 7–11. doi: 10.3146/i0095-3679-14-1-3
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Rajanaidu, A., Kushairi, A., and Mohd-Pin, A. (2017). “Oil palm genetics resources,” in *Oil palm improvement through the use of genetic resources* (Malaysia: Malaysian Palm Oil Board), 221.
- Raman, H., Raman, R., Coombes, N., Song, J., Prangnell, R., Bandaranayake, C., et al. (2015). Genome-wide association analyses reveal complex genetic architecture underlying natural variation for flowering time in canola. *Plant Cell Environ.* 39, 1228–1239. doi: 10.1111/pce.12644
- Rance, K. A., Mayes, S., Price, Z., Jack, P. L., and Corley, R. H. V. (2001). Quantitative trait loci for yield components in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* 103, 1302–1310. doi: 10.1007/s122-001-8204-z
- Rucker, B., and Robbelen, G. (1995). Development of high oleic acid rapeseed. *Proceeding of 9th International Rapeseed Congress*. Cambridge, UK, 2, 389–391.
- Salas, J. J., and Ohlrogge, J. B. (2002). Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. *Arch. Biochem. Biophys.* 403, 25. doi: 10.1016/S0003-9861(02)00017-6
- Sasaki, Y., Kozaki, A., and Hatano, M. (1997). Link between light and fatty acid synthesis: thioredoxin-linked reductive activation of plastidic acetyl-CoA carboxylase. *Natl. Acad. Sci. U.S.A.* 94, 11096–11101. doi: 10.1073/pnas.94.20.11096
- Singh, R., Tan, S. G., Panandam, J. M., Rahman, R. A., Ooi, L. C., Low, E.-T. L., et al. (2009). Quantitative trait loci (QTLs) analysis of palm oil fatty acid composition in an interspecific pseudo-back cross from *Elaeis oleifera* (HBK) Cortes and oil palm (*Elaeis guineensis* Jacq.). *Tree Genet. Genomes* 9, 1207–1225. doi: 10.1007/s11295-013-0629-5
- Singh, R., Ong-Abdullah, M., Low, E. T., Manaf, M. A., Rosli, R., Nookiah, R., et al. (2013). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* 500, 335–339. doi: 10.1038/nature12309
- Sorkheh, K., Malyshevaotto, L. V., Wirthensohn, M. G., Tarkeshesfahani, S., and Martinezgomez, P. (2008). Linkage disequilibrium, genetic association mapping and gene localization in crop plants. *Genet. Mol. Biol.* 31, 5499–5503. doi: 10.1590/S1415-47572008000500001
- Teh, C. K., Ong, A. L., Kwong, Q. B., Apparou, S., Chew, F. T., Mayes, S., 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. doi: 10.1038/srep19075
- Ting, N. C., Jansen, J., Nagappan, J., Ishak, Z., Chin, C. W., Tan, S. G., 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. doi: 10.1371/journal.pone.0053076
- Tisne, S., Denis, M., Cros, D., Pomies, V., Riou, V., Syahputra, I., et al. (2015). Mixed model approach for IBD-based QTL Mapping in a complex oil palm pedigree. *BMC Genomics* 16, 798. doi: 10.1186/s12864-015-1985-3
- Turner, S. D. (2014). Qqman: an R package for visualizing GWAS results using QQ and Manhattan plots. *BioRxiv* 005165. doi: 10.1101/005165
- Ukoskit, K., Chanroj, V., Bhusudsawang, G., Pipatchartlearnwong, K., Tangphatsomruang, S., and Tragoonrun, S. (2014). Oil palm (*Elaeis guineensis* Jacq.) linkage map, and quantitative trait locus analysis for sex ratio and related traits. *Mol. Breed.* 33, 415–424. doi: 10.1007/s11032-013-9959-0
- Voelker, T. A., Hayes, T. R., Cranmer, A. M., Turner, J. C., and Davies, H. M. (1996). Genetic engineering of a quantitative trait: metabolic and genetic parameters influencing the accumulation of laurate in rapeseed. *Plant J.* 9, 229–241. doi: 10.1046/j.1365-313X.1996.09020229.x
- Wagner, N., Mroczka, A., and Roberts, P. D. (2011). RNAi trigger fragment truncation attenuates soybean FAD2-1 transcript suppression and yields intermediate oil phenotype. *Plant Biotechnol. J.* 9, 723–728. doi: 10.1111/j.1467-7652.2010.00573.x
- Wilson, R. F., Marquardt, T. C., Novitzky, W. P., Burton, J. W., Wilcox, J. R., Kinney, A. J., et al. (2001). Metabolic mechanisms associated with alleles governing the 16:0 concentration of soybean oil. *J. Am. Oil Chem. Soc.* 78, 335–340. doi: 10.1007/s11746-001-0265-4
- Xia, W., Luo, T., Zhang, W., Mason, A. S., Huang, D., Huang, X., et al. (2019). Development of high-density SNP markers and their application in evaluating genetic diversity and population structure in *Elaeis guineensis*. *Front. Plant Sci.* 10, 130. doi: 10.3389/fpls.2019.00130

- Xiao, Y., Xia, W., Mason, A. S., Cao, Z., Fan, H., Zhang, B., et al. (2019). Genetic control of fatty acid composition in coconut (*Cocos nucifera*), African oil palm (*Elaeis guineensis*), and date palm (*Phoenix dactylifera*). *Planta* 249, 333–350. doi: 10.1007/s00425-018-3003-x
- Yu, J., Pressoir, G., and Briggs, W. H. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zhou, Q., Han, D., Mason, A. S., Zhou, C., Zheng, W., Li, Y., et al. (2017). Earliness traits in rapeseed (*Brassica napus*): SNP loci and candidate genes identified by genome-wide association analysis. *DNA Res.* 0, 1–16. doi: 10.1093/dnares/dsx052

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Xia, Luo, Dou, Zhang, Mason, Huang, Huang, Tang, Wang, Zhang and Xiao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.