Quantitative trait loci (QTLs) analysis of palm oil fatty acid composition in an interspecific pseudo-backcross from Elaeis oleifera (HBK) Cortès and oil palm (Elaeis guineensis Jacq.)


To cite this version:

Carmenza Montoya, Ricardo Lopes, Albert Flori, David Cros, Teresa Cuellar, et al.. Quantitative trait loci (QTLs) analysis of palm oil fatty acid composition in an interspecific pseudo-backcross from Elaeis oleifera (HBK) Cortès and oil palm (Elaeis guineensis Jacq.). Tree Genetics and Genomes, Springer Verlag, 2013, 9 (5), pp.1207 - 1225. 10.1007/s11295-013-0629-5. hal-01190742

HAL Id: hal-01190742
https://hal.archives-ouvertes.fr/hal-01190742
Submitted on 1 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Quantitative trait loci (QTLs) analysis of palm oil fatty acid composition in an interspecific pseudo-backcross from *Elaeis oleifera* (H.B.K.) Cortés and oil palm (*Elaeis guineensis* Jacq.)

Carmenza Montoya · Ricardo Lopes · Albert Flori · David Cros · Teresa Cuellas · Maryline Summo · Sandra Espeout · Ronan Rivallan · Ange-Marie Risterucci · Daniela Bittencourt · Jorge R. Zambrano · Wilmar H. Alarcón G · Pierre Villeneuve · Michel Pina · Bruno Nouy · Philippe Amblard · Enrique Ritter · Thierry Leroy · Norbert Billotte

Received: 23 August 2012 / Revised: 24 March 2013 / Accepted: 4 April 2013 / Published online: 21 June 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

**Abstract** We chose an *Elaeis* interspecific pseudo-backcross of first generation (*E. oleifera* × *E. guineensis*) × *E. guineensis* to identify quantitative trait loci (QTLs) for fatty acid composition of palm oil. A dense microsatellite linkage map of 362 loci spanned 1.485 cM, representing the 16 pairs of homologous chromosomes in the *Elaeis* genus from which we traced segregating alleles from both *E. oleifera* and *E. guineensis* grandparents. The relative linear orders of mapped loci suggested the probable absence of chromosome rearrangements between the *E. oleifera* and *E. guineensis* genomes. A total of 19 QTL associated to the palm oil fatty acid composition were evidenced. The QTL positions and the species origin as well as the estimated effects of the QTL marker alleles were in coherence with the knowledge of the oil biosynthesis pathway in plants and with the individual phenotypic correlations between the traits. The mapping of chosen *Elaeis* key genes related to oleic acid C18:1, using intra-gene SNPs, supported several QTLs underlying notably FATA and SAD enzymes. The high number of hyper-variable SSR loci of known relative linear orders and the QTL information make these resources valuable for such mapping study in other *Elaeis* breeding materials.

**Keywords** *Elaeis* · Oil palm · *E. oleifera* · QTL · Fatty acid

**Abbreviations**
ACCase  Acetyl-CoA carboxylase
ACP    Acyl carrier protein
AFLP   Amplified fragment length polymorphism

Communicated by D. Grattapaglia

C. Montoya
Oil Palm Biology and Breeding Program, Oil Palm Research Center—Cenipalma, Calle 20A # 43A-50, piso 4,
Bogotá, DC, Colombia

R. Lopes · D. Bittencourt
Laboratory of Molecular Biology, EMBRAPA Western Amazon,
Rodovia AM-10, Km 29, PO Box 319,
69010-970 Manaus, AM, Brazil

A. Flori · D. Cros · T. Cuellas · M. Summo · S. Espeout · R. Rivallan · A.-M. Risterucci · T. Leroy · N. Billotte
CIRAD Umr AGAP, Avenue Agropolis, TA A-108/03,
34398 Montpellier Cedex 5, France
e-mail: norbert.billotte@cirad.fr

J. R. Zambrano
Hacienda La Cabaña S.A., Carrera 8 No. 69-43,
Bogotá, DC, Colombia
Introduction

The oil palm (Elaeis guineensis Jacq.) is the most productive oil crop in the world, yielding an average of 4.1 t of vegetable oil per hectare per year. According to the United States Department of Agriculture, total world vegetable oil production in 2011 was 155.8 Mt, with the palm oil in first place (50.7 Mt), followed by soybean oil (42.4 Mt) and rapeseed oil (24.3 Mt). The consumption of vegetable oils falls into two major applications: food industry (with over 80 % of the market) and chemical industry for formulation of paints, inks, resins, varnishes, plasticizers, biodiesel production, etc. (Rosillo-Calle et al. 2009).

The oil palm is a monoic monocot, perennial, and allogamous plant (Hartley 1988). Its diploid genome consists of 16 chromosome pairs (Maria et al. 1995), and the content of its nuclear DNA has been estimated by flow cytometry as 3.9 pg/2C (Rival et al. 1997). There are only two species in the genus Elaeis: E. guineensis Jacq, originating from Africa and E. oleifera (H.B.K) from South America. The Elaeis fruit is a drupe, produced in high numbers within tight bunches. Its outer pulp (mesocarp) contains the (red) palm oil, and its kernel (endosperm) contains the kernel oil, minor oil similar to the coconut oil. The species E. oleifera has a very low yield in palm oil compared to commercial oil palm varieties (Tan et al. 1985), but it has several desirable agronomic characteristics: Elaeis oleifera presents a slow growth of the stem (Hardon 1969), shows resistance to the bud rot disease in South America (Meunier 1991), and a more edible palm oil in its fatty acid composition and concentrations of metabolites (carotenoids, tocopherols, and tocotrienols) also important for the phytopharmaceutical industry (Choo et al. 1997). Genetic advantageous features from both Elaeis species can be combined by interspecific hybridization for creating more performing oil palm varieties (Ong et al. 1981). The palm oil of E. guineensis contains approximately 50 % of saturated fatty acids, with 44 % of palmitic acid (C16:0), 5 % of stearic acid (C18:0), and trace amounts of myristic acid (C14:0). The unsaturated fatty acids are about 40 % of oleic acid (C18:1) and 10 % of polyunsaturated linoleic acid (C18:2) and linolenic acid (C18:3) (Cottrell 1991; Sambanthamurthi et al. 2000). The species E. oleifera is characterized by its high content of unsaturated fatty acids compared to E. guineensis. Among the E. oleifera palms from the Amazon region, the range for C18:1 is 47 to 52 %, while the range for C16:0 is 20 to 29 % (Meunier 1975; Mohd Din et al. 2000; Rey et al. 2003). The iodine value (IV) is a multi-parameter measure of the global unsaturation degree of the fatty acids in the vegetal oil. The IV for E. oleifera has been reported to be between 70 and 87 % (Meunier 1975; Mohd Din et al. 2000; Sambanthamurthi et al. 2000; Rey et al. 2004), whereas the value for E. guineensis tenera varieties is between 53 and 60 % (Tan et al. 1985; Ekpa et al. 1994; Noh et al. 2002; Rey et al. 2004).

In plants, de novo fatty acid synthesis takes place in the plastids and starting with the carboxylation of acetyl-CoA. Then, a series of condensation reactions catalyzed by the enzyme KAS (β-ketoacyl-ACP synthase) type I, II, and III allow the elongation of carbon chains of C4:0 (butyril acid) until C18:0 (stearic acid) (Ohlrogge and Jaworski 1997; Nishida 2004). Afterward, desaturase enzymes produce carbon chain unsaturation and originate fatty acids like C18:1 (oleic acid) and C18:3 (linolenic acid) (Browse et al. 1993; Okuley et al. 1994; Salas and Ohlrogge 2002). Simultaneously, there are subsequent elongations catalyzed by the fatty acid elongase (FAE) complex in the endoplasmic reticulum (ER), and finally, the fatty acids are incorporated into molecules of triacylglycerols (TAGs) which are assembled through the Kennedy pathway (Guschina and Harwood 2007 and references therein).

The traits controlled by multiple genes such as palm oil composition can be deciphered using modern DNA marker technologies to implement a marker-assisted selection (MAS) for more desirable fatty acid composition, as it is already done for other oil crops like oilseed rape (Brassica napus) (Hu et al. 1999, 2006; Smooker et al. 2011) soybean
(Glycine max (L.) Merr.) (Wilson et al. 2001; Monteros et al. 2008; Ha et al. 2010; Li et al. 2011; Chen et al. 2012) and olive (Olea europaea L.) (Poghosyan et al. 1999; Hernández et al. 2005; Banilas et al. 2011). With respect to genes involved in the fatty acid biosynthesis in oil palm, Nakkaew et al. (2008) identified and cloned the gene of biotin carboxylase (accD) from E. guineensis which is involved in the first step of the fatty acid synthesis. More recently, Al-Shanfari et al. (2011) constructed a suppression subtractive hybridization (SSH) library to identify genes involved in fruit ripening in E. guineensis. Bourgis et al. (2011) and Tranbarger et al. (2011) notably analyzed the gene expression and regulatory mechanisms underlying the fatty acid biosynthesis during the oil palm fruit mesocarp maturation and ripening. Singh et al. (2009) identified quantitative trait loci (QTLs) for fatty acid composition using 81 individuals of an Elaeis interspecific cross. However, these QTLs were only detected on the E. guineensis parent and not from the highly homozygous E. oleifera parent. In that way, no E. oleifera QTLs of palm oil composition are known.

A palm oil with a higher proportion of unsaturated fatty acids could mean new market opportunities for oil palm planters. This can be achieved by an introgressive hybridization between E. oleifera and E. guineensis (Tan et al. 1985; Amblard et al. 1995). The identification of QTLs or genes involved in the variation of fatty acid composition in both Elaeis species will allow a MAS. Therefore, the aim of our mapping study was to identify QTLs involved in the fatty acid biosynthesis of palm oil in both E. oleifera and E. guineensis genomes using an interspecific pseudo-backcross (E. oleifera × E. guineensis) × E. guineensis. As hypothesis, this population is suitable in identifying QTLs of palm oil composition as it is possible to distinguish the segregating alleles from E. oleifera and E. guineensis. We suppose that the proportions of fatty acids are traits involving QTLs with effects strong enough to be detected despite the limited size of our population, while minor QTLs might not be detected. Also, we focused on Elaeis key genes involved in the oleic acid (C18:1) proportion using intra-gene SNPs. The future purpose is to apply an introgression of E. oleifera genes into elite palm oil commercial varieties, by MAS in order to obtain a palm oil more rich in unsaturated fatty acids.

Materials and methods

Plant material

The mapping population consisted of 134 full-sibs derived from an interspecific pseudo-backcross obtained from the interspecific hybrid palm SA65T (E. oleifera SA49D × E. guineensis LM2466P) and an E. guineensis genitor PO3228D. This pseudo-backcross, named SA569, was planted in 1991 at two different locations in Colombia: Hacienda la Cabaña (LC) in the Department of Meta and Indupalma (IP) in the Department of Cesar. Palms were planted in progeny row trials in both sites, where the two agro-climatic environments are similar: alluvial soils, annual rainfall between 2,500 and 3,000 mm, medium temperature of 27 °C, a wet season between April and September, and a dry season between December and March. We assume that the environment effect is the same for all genotypes in both sites and that an additive correction of the phenotypic trait values from the planting site effect is sufficient.

This is a pseudo-backcross because the grandparents and parents are not totally homozygous and the E. guineensis parent is not the same genotype as the grandparent. Three fruit type varieties exist in the Elaeis genus depending on the shell thickness controlled by a major co-dominant gene called Sh (Beiraert and Vanderweyen 1941): dura (Sh+/Sh-), pisifera (Sh-/Sh+), and tenera (Sh+/Sh-) with thin shell. The female grandparent (SA49D) is a wild E. oleifera palm (dura) from the Coari region (Brazilian Amazon), and the male grandparent (LM2466P) is a pisifera oil palm obtained by selfing the LM2T oil palm genitor, an elite individual of the La Mé breeding population. The E. guineensis male parent PO3228D is a dura oil palm derived by selfing the DA115D oil palm genitor, an elite individual of the Deli breeding population. The fruit variety, either dura or tenera, was determined for each progeny as a morphological marker in order to map the Sh gene. About half of the progenies were tenera and the other half dura, without significant deviation from the 1:1 segregation ratio expected. Total genomic DNA was extracted for each grandparent, parent, and progeny from freeze-dried leaf samples using the NucleoSpin® 96 Plant II kit (Ref 740663.2) following the manufacturer’s (Macherey-Nagel GmbH, Germany) instructions.

Measurements of the fatty acid composition

A total of 115 progenies (90 from LC and 25 from IP), as well as the E. oleifera grandparent SA49D and the interspecific hybrid parent SA65T, were analyzed for fatty acid composition and the iodine value of the palm oil in mature fruits. The number of SA569 individuals planted was initially high (200), but some palms died before sampling. Also, SA569 had abort/sterility problems which affected the bunch development, resulting in no or low bunch production, which is a common problem in interspecific oil palm materials (Obasola et al. 1977; Pamin et al. 1995). The mapping population is quite small and borderline acceptable for QTL mapping, a somewhat inevitable situation considering the limitations that the crop presents for accessing larger families and the phenotyping work involved. An alternative solution could have been to use the
multi-parent QTL mapping strategy applied in oil palm by Billotte et al. (2010), using in our case pseudo-backcrosses sharing common parents, but no such mating design exists yet.

For each palm, two ripe bunches were collected, and 30 random fruits were sampled per bunch (two fruit samples/palm), to provide a good estimate of the palm oil quality (Wuidart and Gascon 1975), based on previous experiments of the former “Institut de Recherche des Huiles et des Oléagineux” (I.R.H.O) in France. Palm oil composition analysis was carried out on these two biological replicates per palm, and results are reported as mean values. No phenotypic data were available for the E. guineensis grandparent LM2466P and parent PO3228D. Fruits were boiled (100 °C) in water and then hot-pressed using a screw press to extract the palm oil. The extracted oil and hot water were recovered in a 25-ml round bottom flask. Reaction medium was refluxed for 10 min. A volume of 3 ml of methanolic HCl was added until phenolphthalein discoloration occurred. The mixture was refluxed again for 10 min and then cooled down to room temperature. A volume of 8 ml of hexane and 10 ml of water were added, and the organic phase was recovered, dried over anhydrous sodium sulfate, and filtered for subsequent GC analysis with an Agilent 6890 series (Bios Analytique, France) using a Supelcowax 10 capillary column (SGE, Courtabœuf, France) with the following characteristics: length, 30 m; internal diameter, 0.32 mm; film thickness, 0.25 μm; FAMEs (1 μl) directly injected into the GC; carrier gas, helium flow 1.0 ml/min; splitting ratio, 1/20; injector temperature, 250 °C; and FID detector temperature, 270 °C. The temperature settings were as follows: initial temperature 185 °C for 2 min and 185 to 225 °C at 5 °C/min. The final results are expressed as the relative percentage of the peak area. Quantitative phenotypic traits in this study considered the nine main fatty acids: myristic acid (C14:0), palmitic acid (C16:0), palmitoleic acid (C16:1), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), linolenic acid (C18:3), arachidic acid (C20:0), and gadoleic acid (C20:1). In addition, the iodine value (IV) was determined by the Wijs method described in the ISO 3961:2009 standard.

Statistical analysis of phenotypic data

An analysis of variance (ANOVA) test (post hoc test of Tukey, α=0.05) was realized for all traits to evidence eventual differences between the planting sites LC and IP. If significant differences were detected for a given trait, the individual phenotypic data for the IP site were standardized according to the LC site, by a mean correction for further statistical analyses as follows: IP Standardized data=IP raw data+(LC mean−IP mean). The Gauss distribution of the quantitative data was checked by a normality test of Shapiro–Wilks at an α threshold of 5%. The relationships between phenotypic traits, at the individual palm level, were estimated by calculating the Pearson’s correlation coefficients. The trait heritability h² (broad sense) was estimated by ANOVA with one fixed factor (palm) using the two bunches measured per palm assuming that for these oil traits, variability between bunches on the same tree is not far from variability between bunches on separate trees of a clone in the same trial, as follows: model: Yij=m+Ai+Eij, where i (1 to 115) is the palm, j (1 to 2) is the bunch measured, Yij is the phenotypic value of the palm i for its bunch j, Ai is the trait genetic value of the palm, Eij is the error (environment effect on the trait phenotypic value), and h²=σ²/E²+σ²+σ², where σ² is the genetic variance, E² is the error variance, MS_p is the mean square for palm, and MS_E is the error mean square.

Microsatellite analyses

SSR genotyping of SA569 along with its grandparents and parents was performed by the commercial laboratory ADNid (Montpellier, France, http://www.adnid.fr). The legitimacy of the progenies was checked using 12 SSR loci according to Durand-Gasselin et al. (2009). An initial set of 700 SSR loci was pre-screened for polymorphism on grandparents and parents. A total of 384 SSR loci were genotyped and consisted of 250 oil palm SSR loci (Billotte et al. 2001, 2005 and unpublished by CIRAD), 21 coconut SSR loci (Billotte et al. 2010), and 113 oil palm EST-SSR loci (Tranbarger et al. 2012). Our choice favored the SSRs with best amplification patterns, a priori segregating from the SA65T and PO3228D parents, and with high polymorphism. The plant material was analyzed by the SSR technique in fluorescent conditions according to Roy et al. (1996) using the capillary electrophoresis and detection systems of a 3500xl Genetic Analyzer (Applied Biosystems, USA). The PCR amplification was performed in a 384-well Eppendorf Mastercycler with 25 ng of DNA in a 10-μl final volume of buffer (10 mM Tris–HCl (pH 8.3), 50 mM KCl, 0.001 %w/v glycerol,) containing 1.50 mM MgCl₂, 0.08 μM of the forward M13-tailed primer, 0.1 μM of the reverse primer, 200 μM of dNTP, 1 U of Taq DNA polymerase (Life Technologies, USA), and 0.1 μM of M13 primer-fluorescent dye FAM, NED, PET, or VIC (Applied Biosystems, USA). A touchdown PCR program was

Statistical analysis of phenotypic data

An analysis of variance (ANOVA) test (post hoc test of Tukey, α=0.05) was realized for all traits to evidence eventual differences between the planting sites LC and IP. If significant differences were detected for a given trait, the individual phenotypic data for the IP site were standardized according to the ISO 3961:2009 standard.
performed as follow: initial denaturation at 94 °C for 4 min; 10 cycles at 94 °C for 45 s, primer melting temperature (Tm) (55 to 50 °C, -0.5 °C/cycle) for 1 min, and 72 °C for 1 min 15 s; additional round of 25 cycles of 94 °C for 45 s, 50 °C for 1 min, 72 °C for 1 min, and a final elongation step at 72 °C for 30 min. For post-PCR multiplexing 2.0 μl of FAM, 2.5 μl of NED, 3.5 μl of PET, and 2.0 μl of VIC PCR products were pooled together, and the final volume was adjusted to 20 μl with sterile water. PCR pooled products (2 μl) were mixed with 15 μl of a mixture of GeneScan™-600 LIZ® (Applied Biosystems, USA) and highly deionized Hi-Di™ Formamide (Applied Biosystems, USA). The mixture was prepared by adding 12 μl of size standard GeneScan™-600 LIZ® to 1.5 ml of Hi-Di™ formamide. PCR products were subject to the capillary electrophoresis, detected by the laser system of a 3500xL Genetic Analyzer (Applied Biosystems, USA), and visualized using the GeneMapper® software 4.1 (Applied Biosystems, USA).

The presence or absence of each amplified SSR allele in each individual was identified, cross-checked by two different operators, and the genotype configuration of each locus was coded according to the nomenclature of Ritter et al. (1990), with one to four alleles segregating in our cross between two partially heterozygous parents SA65T and PO3228D.

Intra-gene SNP detection and mapping of chosen key genes

Five *Elaeis* genes of the palm oil biosynthesis were chosen for their mapping and a QTL/gene co-localization study focusing on the oleic acid C18:1 in palm oil: for the enzymes β-ketoacyl-ACP synthase of type I and II (KAS I, KAS II), the Acyl-ACP thioesterases (FATA, FATB) and the Δ9 stearoyl-ACP desaturase (SAD). cDNA sequences of these genes were identified from eight *Elaeis* full-length cDNA libraries of pulp (unpublished) previously made from developing fruits between 60–75 and 150–180 days after pollination. *E. guineensis* sequences of these genes are available from Bourgis et al. (2011) at http://www.biomemb.cnrs.fr/contigs.html in file oil-contigs.doc (Table 4). Each *Elaeis* species was represented by four libraries from different genetic backgrounds: La Mé, Deli, Yangambi, and Nigeria for *E. guineensis* and Colombia, Brazil, Costa Rica, and Surinam for *E. oleifera*. These libraries of the International Oil Palm Genome Projects (OPGP) Consortium were sequenced by the Genoscope-CEA (Evry, France) using the 454 Roche/GS-FLX Pyrosequencing technology (1/2 run per library). Each read library was analyzed using the ESTtik tool (Expressed Sequence Tag treatment and investigation kit) (Argout et al. 2008) especially modified for the analysis of 454 sequence data. The 454 reads obtained from the eight pulp libraries were assembled using the TGICL (Pertea et al. 2003) assembler. Contigs were automatically annotated using the standalone BLAST software (Altschul et al. 1990). We searched for similar sequences on the *Oryza sativa japonica* proteome TIGR6, the Uniprotn Knowledgebase Swiss-Prot and TrEMBL (Boeckmann et al. 2003), the GenBank non-redundant protein database NR, and the GenBank nucleotide database NT. We used an e-value cutoff of 1e-5 and retrieve the 10 best hits to annotate each contig. We used the Blast2GO pipeline (Götz et al. 2008) to annotate the contigs with the Gene Ontology terms based on our BLAST results. Also, a BLAST analysis was performed with the *E. guineensis* cDNA libraries of Bourgis et al. (2011) and Tranbarger et al. (2011). We used the Prot4EST pipeline (Wasmuth and Blaxter 2004) to obtain a protein prediction from our contig sequences. We used the program IDEG6 (Romualdi et al. 2003) for data normalization and statistical analysis of the read libraries (general chi-square test applying the Bonferroni correction, p < 0.0001), to test the differential expression of our five genes between the two *Elaeis* species. Finally, the ACE assembly output from TGICL was treated to obtain a collection of FASTA alignment using the AceToFastaAlignment script from the SNiPlay utilities (Dereeper et al. 2011). We used the SNiPlay pipeline on our FASTA alignment collection for single nucleotide polymorphism (SNP) analysis. SNP loci were identified from cDNA reads for our five targeted genes, with eventually several SNPs for a given gene. Designed SNPs were genotyped by the commercial laboratory ADNid (Montpellier, France, http://www.adnid.fr), combining the GoldenGate Genotyping Assay with the VeraCode technology (Illumina, San Diego, USA) and using the Illumina’s BeadXpress Reader System for detection.

Linkage mapping

A linkage map of SA569 was constructed based on 115 progenies, without missing data for the polymorphic SSRs, SNPs, and the Sh gene. SSR loci showing a significant distortion (p<0.01) according to the expected Mendelian segregating ratio of their alleles were removed from the analysis. The linkage phases between SSR marker alleles were determined using the JoinMap v.4.0 software (Van Ooijen and Voorrips 2001) in our pseudo-backcross considered to be a double pseudo-test cross (Grattapaglia and Sederoff 1994). Marker loci with estimated allele phases were then analyzed using the CarthaGene software (Schliex and Gaspin 1997) at LOD 7.5 with a maximum recombination threshold of 0.3. The most probable relative linear orders of markers within each linkage group (LG) were estimated using the several multi-point maximum likelihood algorithms of CarthaGene. The Haldane mapping function was used to convert recombination frequencies into map distances (Haldane 1919). The chart of the linkage map was generated using BioMERCATOR v.3.1 (Arcade et al. 2004).
QTL analysis

QTL analyses using the SSR-SNP linkage map and all available genotypic and phenotypic data were performed with the MapQTL® software (Van Ooijen 2004). As a first step, the non-parametric Kruskal–Wallis (K-W) test was performed to identify significant marker-trait associations at \( p < 0.005 \). Subsequently, the interval mapping (IM) method was performed using the LOD statistic test with a mapping step size of 1 cM and a maximum number of 5 neighboring markers. For each trait, threshold LOD values were estimated at the genome-wide (GW) global risk \( \alpha \) of 5 and 1 % for declaring the presence of a QTL by the re-sampling method and permutation of the trait data (1,000 iterations) of the software. Finally, a multiple-QTL model (MQM) method was undertaken along with the automatic iterations) of the software. The confidence interval of each significant QTL was estimated under MapQTL, considering that the variance explained by an identified QTL is, as estimated under MapQTL,

\[
\% \text{variance explained} = 100 \left( \frac{\sigma_a^2}{\sigma_p^2} \right)
\]

with \( \sigma_a^2 \) corresponding to the genetic variance due to additive effect and \( \sigma_p^2 \) corresponding to the phenotypic variance.

The corrected variance explained by this QTL was re-estimated as follows:

\[
\% \text{Corrected variance explained} = \frac{\sigma_a^2}{\sigma_p^2} \left( 1 - \frac{1}{2 \ln(10) \times \text{LOD}} \right)
\]

\[
= \frac{\sigma_a^2}{\sigma_p^2} \left( 1 - \frac{1}{4.605 \times \text{LOD}} \right)
\]

where LOD corresponds to the LOD value of the identified QTL.

In addition, the QTL marker loci were used to perform an ANOVA test (type III, post hoc test of Tukey at \( \alpha = 0.05 \)) to estimate the mean effects of the parent QTL marker alleles on the mean of each phenotypic trait. For the hybrid parent SA65T, the species origin of the QTL marker alleles were identified, and the allelic effects at the QTL were therefore estimated by contrast of \( E. oleifera \) (grandparent SA49D) against \( E. guineensis \) (grandparent LM2466P). The QTLs (position, confidence interval, variation explained) were visualized on the linkage map using BioMERCATOR v.3.1 (Arcade et al. 2004).

Results

Palm oil fatty acid composition

The principal palm oil fatty acids found on the SA569 progenies were (Table 1) the following: oleic acid C18:1 (mean 45.0 %), palmitic acid C16:0 (mean 38.8 %), and linoleic acid C18:2 (mean 11.0 %), followed by stearic acid C18:0 (mean 3.5 %). Their coefficients of variation were high to medium, with 8.2 % (C18:1), 7.5 % (C16:0), 14.5 % (C18:2), and 38.7 % (C18:0). The other fatty acids (C14:0, C16:1, C18:3, C20:1, and C20:0) were found in trace amounts (<1 %). The interspecific hybrid parent SA65T showed rather similar proportions of C16:0 and C18:1 compared to its \( E. oleifera \) parent SA49D. The individual phenotypic values of the SA569 progenies were normally distributed for the four traits C16:0, C18:1, C18:2, and iodine value (IV) while not for C18:0 and fatty acids in traces (C14:0, C16:1, C18:3, C20:0, and C20:1). The data histograms (data not shown) presented a continuous variation in all traits with the exception of C20:1, which was excluded from the QTL analysis. The ANOVA test (data not shown) showed significant mean differences between the two plantation sites LC and IP for the traits C18:2, C20:0, and IV. The individual phenotypic data were therefore corrected as mentioned above for further statistical analyses.

Pearson’s correlations between the individual fatty acid proportions and IV are given in Table 2. The myristic acid C14:0 was positively correlated with the palmitic acid C16:0 and palmitoleic acid C16:1 while negatively correlated with the stearic acid C18:0 and oleic acid C18:1. The palmitic acid C16:0 was negatively correlated with the oleic acid C18:1 and with the iodine value IV, indicative for the total proportion of unsaturated fatty acids. Two other correlations for C16:0 were positive with C14:0 and negative with C18:3. Additional correlations were negative between C16:1 and C18:0, C18:0 and C18:1, and C18:1 and C18:2 and positive between C18:2 and C18:3. The broad sense heritabilities \( h^2 \) were high in the pseudo-backcross (0.8 and above) for the principal fatty acid proportions C16:0, C18:0, C18:1, and C18:2 as well as for the iodine value.
Intra-gene SNPs and microsatellite polymorphism

In the eight *Elaeis* full-length cDNA libraries of pulp, the number of cDNA consensus contigs, i.e., distinct copies in *Elaeis* for the studied key genes expressed in the pulp, and their related genotyped SNPs are: 3 copies for SAD (15 SNPs), 3 copies for KAS I (10 SNPs), 1 copy for KAS II (3 SNPs), 1 copy for FATA (2 SNPs), and 2 copies for FATB (4 SNPs). In total, 14 out of the 34 intra-gene SNPs were polymorphic on the SA569 parents. Every gene copy had at least one polymorphic SNP, except one copy of KAS I.

Up to 84 % of the 700 tested SSRs were correctly amplified in the grandparents and parents of SA569. For these amplified SSRs, the number of heterozygous loci was 161 (28 %) in the *E. oleifera* grandparent SA49D, 147 (25 %) in the *E. guineensis* grandparent LM2466P, and 128 (22 %) for the *E. guineensis* parent PO3228D (data not shown).

About 19 % of the SSRs were homozygous in both parents, and 81 % (471) were mappable on SA569, including 19 % in one parent only and 69 % in both parents. About 12 % of the SSRs showed at least one null allele in the *E. oleifera* palm SA49D, compared to 1 and 3 %, respectively, in the *E. guineensis* palms LM2466P and PO3228D. Out of the 384 SSRs genotyped on SA569, 372 (97 %) showed effectively a traceable segregation of their alleles, allowing the mapping of their loci (Fig. 1). The SSR allele patterns of the latter showed six of the nine genotype configurations defined by

### Table 1

Means, ranges, variances, variation coefficients, and broad sense heritability for iodine value and fatty acid composition in SA569 and the parent SA65T and grandparent SA49D

<table>
<thead>
<tr>
<th>Trait</th>
<th>On the SA569</th>
<th>Parents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acronym</td>
<td>Mean (n=115)</td>
</tr>
<tr>
<td>Myristic acid</td>
<td>C14:0</td>
<td>0.6</td>
</tr>
<tr>
<td>Palmitic acid</td>
<td>C16:0</td>
<td>38.8</td>
</tr>
<tr>
<td>Palmitoleic acid</td>
<td>C16:1</td>
<td>0.2</td>
</tr>
<tr>
<td>Stearic acid</td>
<td>C18:0</td>
<td>3.5</td>
</tr>
<tr>
<td>Oleic acid</td>
<td>C18:1</td>
<td>45.0</td>
</tr>
<tr>
<td>Linoleic acid</td>
<td>C18:2</td>
<td>11.0</td>
</tr>
<tr>
<td>Linolenic acid</td>
<td>C18:3</td>
<td>0.3</td>
</tr>
<tr>
<td>Arachidic acid</td>
<td>C20:0</td>
<td>0.3</td>
</tr>
<tr>
<td>Gadoleic acid</td>
<td>C20:1</td>
<td>0.1</td>
</tr>
<tr>
<td>Saturated</td>
<td></td>
<td>43.5</td>
</tr>
<tr>
<td>Monounsaturated</td>
<td></td>
<td>45.3</td>
</tr>
<tr>
<td>Polyunsaturated</td>
<td></td>
<td>11.4</td>
</tr>
<tr>
<td>Iodine value</td>
<td>IV</td>
<td>58.7</td>
</tr>
</tbody>
</table>

*a Coefficient of variation  
b Broad sense heritability estimated from offsprings

### Table 2

Individual Pearson’s correlation coefficients for fatty acid proportions and iodine value in the pseudo-backcross population SA569

<table>
<thead>
<tr>
<th>Trait</th>
<th>Acronym</th>
<th>C16:0</th>
<th>C16:1</th>
<th>C18:0</th>
<th>C18:1</th>
<th>C18:2</th>
<th>C18:3</th>
<th>C20:0</th>
<th>C20:1</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myristic acid</td>
<td>C14:0</td>
<td>0.30**</td>
<td>0.31**</td>
<td>-0.30**</td>
<td>-0.20*</td>
<td>0.05</td>
<td>0.06</td>
<td>-0.22*</td>
<td>-0.07</td>
<td>-0.11</td>
</tr>
<tr>
<td>Palmitic acid</td>
<td>C16:0</td>
<td>0.09</td>
<td>-0.09</td>
<td>-0.79**</td>
<td>-0.01</td>
<td>-0.20*</td>
<td>0.05</td>
<td>-0.11</td>
<td>-0.81**</td>
<td>0.18</td>
</tr>
<tr>
<td>Palmitoleic acid</td>
<td>C16:1</td>
<td>-0.78**</td>
<td>0.18*</td>
<td>-0.05</td>
<td>0.13</td>
<td>-0.57**</td>
<td>-0.19*</td>
<td>0.07</td>
<td>0.61**</td>
<td>0.18</td>
</tr>
<tr>
<td>Stearic acid</td>
<td>C18:0</td>
<td>-0.34**</td>
<td>0.09</td>
<td>-0.13</td>
<td>0.67**</td>
<td>0.10</td>
<td>-0.28**</td>
<td>0.10</td>
<td>-0.28**</td>
<td>0.18</td>
</tr>
<tr>
<td>Oleic acid</td>
<td>C18:1</td>
<td>-0.47**</td>
<td>0.08</td>
<td>0.04</td>
<td>-0.06</td>
<td>0.40**</td>
<td>0.02</td>
<td>0.37**</td>
<td>-0.31**</td>
<td>0.02</td>
</tr>
<tr>
<td>Linoleic acid</td>
<td>C18:2</td>
<td>0.26**</td>
<td>0.04</td>
<td>-0.06</td>
<td>0.40**</td>
<td>0.02</td>
<td>0.37**</td>
<td>-0.31**</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Linolenic acid</td>
<td>C18:3</td>
<td>-0.16</td>
<td>0.05</td>
<td>0.05</td>
<td>0.37**</td>
<td>-0.31**</td>
<td>0.02</td>
<td>0.02</td>
<td>0.37**</td>
<td>-0.31**</td>
</tr>
<tr>
<td>Arachidic acid</td>
<td>C20:0</td>
<td>0.38**</td>
<td>-0.31**</td>
<td>0.02</td>
<td>0.37**</td>
<td>-0.31**</td>
<td>0.02</td>
<td>0.02</td>
<td>0.37**</td>
<td>-0.31**</td>
</tr>
<tr>
<td>Gadoleic acid</td>
<td>C20:1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iodine value</td>
<td>IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Asterisks indicate significant values at \( p \leq 0.05 \) (*) and \( p \leq 0.01 \) (**)
Ritter et al. (1990) for a cross between two heterozygous parents. These SSR configurations were cases No. 1, 3, 4, 7, 8, and 9. On the pseudo-backcross, 67% (258) SSR loci belonged to genotype configuration No. 3, where only one parent is heterozygous, which was the hybrid SA65T in most (251) cases. In fact, the large majority of the SSRs was segregating only from the hybrid SA65T (genotype configurations No. 1 and 3) and most others from both parents (genotype configurations No. 7, 8, and 9 with 3 to 4 segregating alleles). In parallel, 30% of the SSRs were segregating from the recurrent parent PO3228D. Considering all genotype configurations, 365 (95.0%) genotyped SSRs showed alleles segregating from the interspecific hybrid parent SA65T and 117 (30.5%) from the *E. guineensis* parent PO3228D.

Microsatellite and intra-gene SNP linkage map

Few SSRs (11) were removed from the genotypic data set because of significant segregation deviation from the Mendelian ratio or due to missing data (12) or locus redundancy (2). A total of 347 segregating SSRs, 14 SNPs of genes, and the *Sh* locus were finally used to establish the linkage map of SA569. This linkage map consisted of 16 linkage groups corresponding to the 16 pairs of homologous chromosomes in *Elaeis* and spanned 1.485 cM with an average marker density of 4 cM (Fig. 2). The standard error in estimating the map position of a given co-dominant marker was ±0.5 cM in our backcross, according to Allard (1965). The linkage group length ranged from 49.1 to 175.9 cM, with an average of 92.8 cM. There were 10 to 38 marker loci per linkage group (average of 22), generally uniformly distributed along the chromosomes, except for a few regions lacking markers (max. ± 30 cM) on linkage groups No. 7, 9, 11, and 12. Regarding each map parent, 73% (253) of the mapped SSRs segregated only from the hybrid parent SA65T, 2% (7) from PO3228D only, and 27% (93) were common SSRs segregating from both parents. Out of the 14 mapped SNPs of genes, 11 segregated only from the hybrid parent SA65T and 4 from both parents.

A total of 156 SSRs (45%) and the *Sh* locus were previously mapped on the *E. guineensis* consensus multi-parent map published in oil palm by Billotte et al. (2010). In general, common loci were co-linear between the interspecific pseudo-backcross SA569 and the *E. guineensis* multi-parent map, except few cases of closely linked markers. Two new SSR loci (mEgCIR3676 and mEgEST0248) were proved on LG 4 to flank the *Sh* gene controlling the fruit type, at 12.5 cM for each of them (Fig. 2).

Identified QTLs

A total of 19 QTLs associated with palm oil fatty acid composition were evidenced by the K-W analysis at *p*<
0.005, with one to three QTLs per fatty acid or iodine value IV (Fig. 2). Subsequently, 13 QTLs were also evidenced by the IM and 11 by MQM analyses at the significant genome-wide threshold \( \alpha \) of 1 or 5 %. These latter QTLs were located on four linkage groups (LGs) 4, 5, 6, and 10. The other QTLs, only identified by the K-W method and therefore considered as “putative” QTLs, were mapped to LGs 2, 3, 5, and 15. For these putative QTLs, a maximum LOD value was also observed (data not shown) with the IM

\[ \text{Legend:} \]

**Marker:**
- mEgCIR: Oil palm genomic SSR
- mEgEST: Oil palm gene SSR
- sEgOPGP: Oil palm gene SNP
- mCnCIR: Coconut genomic SSR

**Position:**
- R²
- Confidence interval

**QTL:**
- Position

**Fig. 2** Nineteen QTLs of fatty acid proportion and iodine value of the palm oil in the interspecific *Elaeis* pseudobackcross SA569. Note: one star (\*) or two stars (\**\): QTL detected by the MQM method at the genome-wide \( \alpha \) threshold value of 5 or 1 %, respectively. No star: putative QTL as only detected by the Kruskall–Wallis test at \( p<0.005 \). The linkage map encompasses 347 SSRs, 14 SNPs, and the Sh locus distributed on 16 linkage groups corresponding to the 16 homologous chromosomes of the plant. The names and the positions (cM) of the markers are given on the right side of the linkage groups.

and/or MQM methods at the same or nearby location, despite being not significant. Some of these putative QTLs were closely linked, such as for C16:0 and IV on LG 15, or co-localized with another significant QTL determined by the IM and/or MQM methods, as for instance a putative QTL of C16:0 with a QTL of C18:1 on LG 5. The confidence interval for the positions of the putative or significant QTLs was 16.5 cM on average (min 6.0 and max 30 cM). Some traits showed common significant QTLs, as for C18:2 and

\[ \text{mEgCIRxxx and mEgESTxxx: *E. guineensis* SSR loci. *E. guineensis* sEgOPGpxxx: SNP loci. mCnCIRxxx: *Cocos nucifera* SSR loci. Reference marker loci in the cross LM2T × DA10D (Billotte et al. 2010) are indicated by an extension “_R”. The names, positions, and confidence regions of the QTLs are given on the left side of the linkage groups. In red are figured the QTLs of saturated fatty acid proportion; in blue: the QTLs of unsaturated fatty acid proportion and of iodine value. The names of mapped genes are given on the right side of corresponding SNP names.} \]
C18:3 sharing the same location at 11.1 cM on LG 10 or for traits C16:1 and C18:0 at 27.4 cM on LG 6. Other significant QTLs were closely linked, such as for C18:1 and C20:0 on LG 6.

The 14 intra-gene SNPs were distributed on five LGs (Fig. 2 and Table 4). They proved by mapping the different copies in *Elaeis* of the genes KAS I (2 on LGs 1 and 15), SAD (3 on LGs 2, 6, and 7), and FATB (2 on LGs 1 and 6), while the single copy for FATA was mapped on LG 6. Obviously, the third copy of KAS I, monomorphic in SA569, was not mapped. Two SNPs, of SAD (sEgOPGP00038) and FATA (sEgOPGP00050), co-localized on LG 6 with the common QTL for C16:1 and C18:0 and were located within the confidence intervals of the QTLs for C18:1 and C20:0. Two other SNPs of KAS I (sEgOPGP00082) and KAS II (sEgOPGP00037) co-localized on LG 14 with the putative QTL of C14:0. Other SNPs did not co-localize with detected QTLs. Regarding the expression of mapped genes at the species level (Table 5), the statistical analysis of the read libraries showed that five mapped gene copies were differentially expressed between *E. oleifera* and *E. guineensis*: FATB (sEgOPGP00017) on LG1, SAD (sEgOPGP00063) on LG 2, FATA (sEgOPGP00050), and SAD (sEgOPGP00038) co-localizing with QTL(s) on LG 6.

The percentage of the phenotypic variation explained by a significant QTL, corrected from the sampling error, was medium to high and ranged between 12 and 50 % (Table 3). The corrected phenotypic variations explained by the QTLs for the principal fatty acids were 16 % (C16:0), 50 % (C18:0), 15 % (C18:1), and 21 % (C18:2). For IV, the explained cumulative variation was 31 % (two QTLs). The ANOVA (data not shown), of the individual phenotypic data depending on the QTL marker alleles for each trait, proved that only the hybrid parent SA65T had a significant parental effect on the trait, due to its inherited *E. oleifera* and *E. guineensis* alleles, while no significant allelic effect was found for the *E. guineensis* parent PO3228D for the few loci (5) heterozygous in that latter (data not shown). The effects associated to the *E. oleifera* QTL marker alleles were...
negative for the proportions of saturated fatty acids C14:0, C16:0, C18:0, and C20:0 (Table 3). In parallel, they were positive for the percentage of the unsaturated fatty acids C16:1, C18:1, C18:2, and C18:3 and for the iodine value IV. Only for C18:3, the *E. oleifera* allele of the QTL locus mEgCIR0801 presented a negative effect.

**Discussion**

**Mapping population and linkage map**

No line exists in the allogamous *Elaeis* species, and subsequently, no true backcross is available for QTL analysis. We chose an *Elaeis* interspecific pseudo-backcross of first generation (*E. oleifera* × *E. guineensis*) × *E. guineensis* to develop molecular markers for fatty acid composition of *Elaeis* palm oil. The pseudo-backcross was highly heterozygous and allowed to map a large number of loci based on different segregating *E. oleifera* and *E. guineensis* grandparent alleles which were traceable in the progenies. The LM2466P and PO3228D palms were the respective selfings of two unrelated LM2T and DA115D genitors. The homozygosity degrees of LM2T and DA115D were 37 and 59 % based on SSR data (Billotte 2004) and increased beyond 75 % in LM2466P and in PO3228D. Due to this reason, the proportion of heterozygous loci useful for QTL analysis on the SA6ST interspecific genome went up to 98 % with traceable *E. oleifera* alleles. On the other side, this proportion was rather low (28 %) for the highly homozygous PO3228D, leading to reduced power for QTL detection in this *E. guineensis* parent. Most loci, segregating only from the hybrid SA6ST (Fig. 1), improved the estimates for linkage phase and genetic distances. The total map length...
Table 3  List of QTLs identified by the Kruskal–Wallis method (at p<0.005) and by the interval mapping (IM) and/or multiple QTL model (MQM) methods for the palm oil

<table>
<thead>
<tr>
<th>Trait</th>
<th>Acronym</th>
<th>Interval mapping analysis</th>
<th>MQM analysis</th>
<th>QTL marker alleles and associated effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LG^a QTL peak (cM)^b</td>
<td>Marker^c</td>
<td>LG^a QTL peak (cM)^b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Maximum LOD^d</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MapQTL estimated Expl. Var.^f</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myristic</td>
<td>C14:0</td>
<td>2</td>
<td>mEgEST0191</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>52.2</td>
<td>30.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>mEgCIR3286^g</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>mEgCIR201^s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>mEgCIR0521</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16.1</td>
<td>11.3</td>
</tr>
<tr>
<td>Palmitic</td>
<td>C16:0</td>
<td>4</td>
<td>mEgCIR3413^s</td>
<td>4.1*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51.7</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>mEgCIR2307^s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20.5</td>
<td>13.9</td>
</tr>
<tr>
<td>Palmitoleic</td>
<td>C16:1</td>
<td>6</td>
<td>sEgOPGP00038</td>
<td>14.9**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>27.4</td>
<td>27.4</td>
</tr>
<tr>
<td>Stearic</td>
<td>C18:0</td>
<td>6</td>
<td>sEgOPGP00038</td>
<td>17.7**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>27.4</td>
<td>27.4</td>
</tr>
<tr>
<td>Oleic</td>
<td>C18:1</td>
<td>6</td>
<td>mEgCIR3676^s</td>
<td>4.2*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>16.1</td>
</tr>
<tr>
<td>Linoleic</td>
<td>C18:2</td>
<td>3</td>
<td>mEgCIR1498</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>58.4</td>
<td>30.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>21.1</td>
</tr>
<tr>
<td>Linolenic</td>
<td>C18:3</td>
<td>4</td>
<td>mEgCIR0801</td>
<td>5.6**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>153.9</td>
<td>14.9</td>
</tr>
<tr>
<td>Arachidic</td>
<td>C20:0</td>
<td>6</td>
<td>mCnCIR236^s</td>
<td>8.9**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>24.1</td>
<td>8.7**</td>
</tr>
<tr>
<td>Gadoleic</td>
<td>C20:1</td>
<td>Not analysed</td>
<td>Not analysed</td>
<td>Not analysed</td>
</tr>
<tr>
<td>Iodine value</td>
<td>IV</td>
<td>6</td>
<td>mEgEST166^s</td>
<td>6.5**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>35.9</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>3.2</td>
</tr>
</tbody>
</table>

a LG = linkage group  
b Cumulative distance from the top marker of the linkage group  
c #: neighborhood locus if not at the QTL position; δ: cofactor marker for MQM analysis  
d α significance threshold: * at 5 %, ** at 1 %  
e Percentage of the phenotypic variance explained at the QTL corrected according to Luo et al. (2003)  
f genome-wide significance threshold at 5 or 1 % of level of probability  
g Eo = allele inherited from the E. oleifera grandparent SA49D; Eg1 = allele inherited from the E. guineensis grandparent LM246P

1218 Tree Genetics & Genomes (2013) 9:1207–1225
is about the same with the SSR linkage map of the oil palm LM2T × DA10D cross (1479 cM) published by Billotte et al. (2010) and shorter than the 1815 or 2247 cM of other published maps for *Elaeis* (Singh et al. 2009; Seng et al. 2011). In addition, only a few (3 %) skewed markers were identified. All loci in Mendelian segregation could be assigned to 1 of the 16 linkage groups. Our linkage map is the first saturated one to be published for an *Elaeis* pseudo-backcross. The co-linearity of loci common to the *E. guineensis* map in oil palm (Billotte et al. 2010) indicated the probable absence of chromosome rearrangements between the *E. oleifera* and *E. guineensis* genomes. This is in coherence with the karyological examination of interspecific meiotic pairing with a balanced distribution of chromosomes in gametes and formation of regular tetrads (Schwendiman et al. 1982).

The size of our mapping population is limited due to the lack of a larger pseudo-backcross in current *Elaeis* breeding materials. In that way, our experiment is a preliminary study where a certain number of QTLs with small or medium effects might not be detected. Studies showed that by increasing the population size, it is possible to increase the total number of QTLs detected and to reveal those with small effects (Vales et al. 2005; Raghavan and Collard 2012). This is well known that the power to detect QTLs is rather low using a small number of individuals, depending on the trait heritability and QTL effects, with also the non-negligible risk to evidence false QTLs especially for those with weak effects (Beavis 1998). Considering that the best detection of QTLs is observed in highly heritable traits, highheritabilities may provide clues about the accuracy of our QTLs found. The broad sense heritabilities in our interspecific backcross were very high for principal fatty acids and relevant compared to those reported in oil palm by Noiret and Wuidart (1976) or Ollagnier and Olivin (1984), with values of 0.8 for oleic acid and 0.9 for iodine value. In soybean RIL populations, heritabilities between 0.6 and 0.7 for fatty acid composition were reported (Panthee et al. 2006). In peanut, values of 0.9 for oleic acid and linoleic acid were observed in RIL population (Sarvamangala et al. 2011), and for rapeseed, values between 0.8 and 0.9 were founded for fatty acids, with one exception for stearic acid (0.4) in doubled haploid lines (Zhao et al. 2008). The heritability values reported by previous authors for oil palm and other oil crops indicate the reliability of our experimentation.

### Table 4

<table>
<thead>
<tr>
<th>Reference Contig no. from Bourgis et al. (2011)</th>
<th>Gene</th>
<th>Gene description</th>
<th>Blast TAIR9</th>
<th>cDNA length (bp)</th>
<th>Intra-gene <em>Elaeis</em> SNP locus</th>
<th>SNP map position and co-localized QTL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>Contig no.</td>
<td>Gene</td>
<td>Description</td>
<td>Accession</td>
<td>Id. (%)</td>
<td>Name</td>
</tr>
<tr>
<td>M01000059296</td>
<td>KAS I</td>
<td>β-Ketoacyl-ACP synthase I</td>
<td>AT5G46290</td>
<td>88</td>
<td>1,304</td>
<td>sEgOPGP00081</td>
</tr>
<tr>
<td>M01000051962</td>
<td>KAS I</td>
<td>β-Ketoacyl-ACP synthase I</td>
<td>AT5G46290</td>
<td>86</td>
<td>1,695</td>
<td>sEgOPGP00082</td>
</tr>
<tr>
<td>M01000052495</td>
<td>KAS II</td>
<td>β-Ketoacyl-ACP synthase II</td>
<td>AT1G74960</td>
<td>85</td>
<td>1,768</td>
<td>sEgOPGP00037</td>
</tr>
<tr>
<td>M01000052807</td>
<td>SAD</td>
<td>Δ9 stearoyl-ACP desaturase</td>
<td>AT2G43710</td>
<td>78</td>
<td>1,816</td>
<td>sEgOPGP00063</td>
</tr>
<tr>
<td>M01000052807</td>
<td>SAD</td>
<td>Δ9 stearoyl-ACP desaturase</td>
<td>AT2G43710</td>
<td>78</td>
<td>1,840</td>
<td>sEgOPGP00038</td>
</tr>
<tr>
<td>M0100005019</td>
<td>SAD</td>
<td>Δ9 stearoyl-ACP desaturase</td>
<td>AT2G43710</td>
<td>79</td>
<td>1,640</td>
<td>sEgOPGP00022</td>
</tr>
<tr>
<td>M0100003668</td>
<td>FATA</td>
<td>Acyl-ACP thioesterase type A</td>
<td>AT4G13050</td>
<td>75</td>
<td>1,192</td>
<td>sEgOPGP00050</td>
</tr>
<tr>
<td>M0100002790</td>
<td>FATB</td>
<td>Acyl-ACP thioesterase type B</td>
<td>AT1G08510</td>
<td>74</td>
<td>1,380</td>
<td>sEgOPGP00017</td>
</tr>
<tr>
<td>M0100000939</td>
<td>FATB</td>
<td>Acyl-ACP thioesterase type B</td>
<td>AT1G08510</td>
<td>69</td>
<td>822</td>
<td>sEgOPGP00057</td>
</tr>
</tbody>
</table>
experiment was capable of identifying QTLs of medium to high effects, while smaller and undetected QTLs explaining a non-negligible part of the variation. The QTL power detection of our experiment can be precised according to the simulations of Muranty (1996). In our pseudo-backcross of about 100 individuals, for a biallelic QTL of an expected heterozygosity He (Nei 1973) of 0.5, the chance to detect a QTL is about 75 % when the explained phenotypic variance (EPV) by that QTL is 20 % or above. Increasing the number of individuals for a single cross allows such a QTL detection power, even for QTLs with an EPV as low as 5 %, while the power is not increased for QTLs with an EPV of 20 % and over, even with 500 individuals. In the future, a multi-parent QTL mapping design (Muranty 1996) as used in oil palm by Billotte et al. (2010) will be used to detect QTLs of palm oil composition in interspecific materials.

Using a population of a small size leads also to overestimation of the additive variance associated with correctly detected QTL. The bias is the sum of two biases: due to the sampling error (contribution of the environmental variance to the estimate of the additive variance of the QTL) and of the Beavis effect (Beavis 1998). Similar with those proposed by Luo et al. (2003) and Xu (2003), we corrected the variances explained by the QTLs from the bias due to the sampling error, knowing that the major part of the overestimation, which is the Beavis effect itself, cannot be corrected. Notwithstanding, the identified QTLs and mapped homologous genes involved in the palmitic and oleic acids biosynthesis are new information in the Elaeis genus and on a saturated genetic map. The next step will be to expand our population to reach at least 200 individuals for a better detection and effect estimation of the QTLs. Also, in a near future, other single pseudo-backcrosses will be used or again a multi-parent QTL mapping design (Muranty 1997) like that tested in oil palm by Billotte et al. (2010). An association mapping study with unrelated Elaeis accessions in collections could also precise the accuracy of the identified QTLs (associated marker effects) and evidenced those undetected one.

### Fatty acid composition

We have analyzed the two classes of saturated (C14:0, C16:0, C18:0, and C20:0) and unsaturated (C16:1, C18:1, C18:2, C18:3, and C20:1) fatty acids in the pulp of mature

### Table 5  Elaeis cDNAs of five key gene functions related to the oleic acid (C18:1) biosynthesis. Identification of differentially expressed genes between E. guineensis and E. oleifera species by analysis of read matrices of full-length cDNA libraries of pulp from developing fruits

<table>
<thead>
<tr>
<th>Gene</th>
<th>Associated Elaeis SNP locus</th>
<th>Total reads</th>
<th>E. guineensis reads</th>
<th>E. oleifera reads</th>
<th>Normalized E. guineensis counts (1)</th>
<th>Normalized E. oleifera counts (2)</th>
<th>1-2 statistical tests(\alpha)</th>
<th>AC test(b) R test(i)</th>
<th>(\chi^2) test</th>
</tr>
</thead>
<tbody>
<tr>
<td>KAS I</td>
<td>sEgOPGP000081</td>
<td>58</td>
<td>34</td>
<td>24</td>
<td>0.40</td>
<td>0.40</td>
<td>0.0465</td>
<td>0.5111</td>
<td>0.5144</td>
</tr>
<tr>
<td>KAS I</td>
<td>sEgOPGP000082</td>
<td>182</td>
<td>120</td>
<td>62</td>
<td>1.50</td>
<td>0.90</td>
<td>0****</td>
<td>0****</td>
<td>0****</td>
</tr>
<tr>
<td>KAS I</td>
<td>SNPs monomorphic</td>
<td>465</td>
<td>411</td>
<td>54</td>
<td>5.20</td>
<td>0.90</td>
<td>0****</td>
<td>0****</td>
<td>0****</td>
</tr>
<tr>
<td>KAS II</td>
<td>sEgOPGP00037</td>
<td>173</td>
<td>108</td>
<td>65</td>
<td>1.4</td>
<td>0.1</td>
<td>0.0002</td>
<td>0.0015</td>
<td>0.0017</td>
</tr>
<tr>
<td>SAD</td>
<td>sEgOPGP00063(c)</td>
<td>6,336</td>
<td>4,524</td>
<td>1,812</td>
<td>57.50</td>
<td>27.40</td>
<td>0****</td>
<td>0****</td>
<td>0****</td>
</tr>
<tr>
<td>SAD</td>
<td>sEgOPGP00038</td>
<td>5,444</td>
<td>2,769</td>
<td>2,675</td>
<td>35.20</td>
<td>40.50</td>
<td>0****</td>
<td>0****</td>
<td>0****</td>
</tr>
<tr>
<td>SAD</td>
<td>sEgOPGP00022(d)</td>
<td>47</td>
<td>31</td>
<td>16</td>
<td>0.40</td>
<td>0.20</td>
<td>0.0180</td>
<td>0.1059</td>
<td>0.1103</td>
</tr>
<tr>
<td>FATA</td>
<td>sEgOPGP00050</td>
<td>303</td>
<td>101</td>
<td>202</td>
<td>1.30</td>
<td>3.10</td>
<td>0****</td>
<td>0****</td>
<td>0****</td>
</tr>
<tr>
<td>FATB</td>
<td>sEgOPGP00017</td>
<td>108</td>
<td>99</td>
<td>9</td>
<td>1.30</td>
<td>0.10</td>
<td>0****</td>
<td>0****</td>
<td>0****</td>
</tr>
<tr>
<td>FATB</td>
<td>sEgOPGP00057</td>
<td>36</td>
<td>12</td>
<td>24</td>
<td>0.20</td>
<td>0.40</td>
<td>0.0030</td>
<td>0.0111</td>
<td>0.0113</td>
</tr>
</tbody>
</table>

Each Elaeis species was represented by four cDNA libraries made from distinct genetic backgrounds. The cDNA libraries were sequenced by the 454 Roche/FLX Pyrosequencing technology

- Accession No. under TAIR9 of the similar Arabidopsis thaliana cDNA
- \(\%\) of sequence identity
- SNP position from the \(5'\) extremity of the Elaeis cDNA
- LG = linkage group
- Cumulative Haldane distance from the top marker of the linkage group
- QTL co-localized at the SNP position
- As well as sEgOPGP00064, 65, 66, and 67 at the respective cDNA positions 1,236, 316, 683, and 897 bp
- As well as sEgOPGP00023 at the cDNA position 876 bp
- \(\alpha\) significance threshold: **** at 1/10,000
- According to Audic and Claverie (1997)
- According to Stekel et al. (2000)
fruits. An interspecific cross *E. oleifera × E. guineensis* shows mean values for fatty acid proportions which are intermediate between the mean values of its *E. oleifera* and *E. guineensis* parents (Harden 1969; Macfarlane et al. 1975; Meunier and Boutin 1975; Opute and Obasola 1979; Tan et al. 1985). This is in accordance with the co-dominance theory for genes involved in palm oil biosynthesis and their alleles inherited from each *Elaeis* parent (Harden 1969; Ong et al. 1981). Surprisingly, the proportion of oleic acid of the interspecific hybrid parent SA65T was quite similar to the one of its *E. oleifera* parent of the Brazilian Coari origin (Table 1). This is probably due to equivalent proportions of oleic acid in both *E. oleifera* and the *E. guineensis* parent of the La Mé origin, which is known to have a high oleic acid proportion (Gascon and Wuidart 1975; Monde et al. 2009).

Pearson correlations do not determine the cause-and-effect relationships between the phenotypic traits but estimate the strength of association between them, at the individual palm level, which is useful for breeding purposes. The oil palm breeders are using for a long time the iodine value as an indicator and a selection parameter for the degree of unsaturation of palm oil (Wuidart and Gascon 1975; Corley and Tinker 2003). The iodine value showed the well-known positive correlations with the proportions of unsaturated fatty acids and negative one with the saturated fatty acids. These correlations are congruent with those of the study by Singh et al. (2009) in an *Elaeis* interspecific cross and also with other correlations in *E. guineensis* (Noh et al. 2002). The positive correlation between C14:0 and C16:0 or C16:1 is in accordance with reports in rice by Ying et al. (2012) and in oil palm by Singh et al. (2009). The negative correlation between C16:0 and C18:1 has been reported in oil palm by several authors (Meunier 1975; Wuidart and Gascon 1975; Noh et al. 2002; Singh et al. 2009). In other crops such as *Sesamum indicum* L (Were et al. 2006), *Brassica napus* L (Zhao et al. 2008), or *Zea mays* (Yang et al. 2010), this negative correlation is frequently justified as pleiotropic effects of single genes or closely linked genes controlling different traits (Yang et al. 2010). The negative correlation between C18:2 and C18:1 has been reported in *Sesamum indicum* L (Uzun et al. 2008), maize (Yang et al. 2010), and *Arachis hypogaea* L. (Sarvamangala et al. 2011). However, estimates of genetic (additive) and pure environmental correlations cannot be precisely estimated on our single cross where all offsprings have the same kinship degree. Such estimates are also not available in the literature. Therefore, our conclusions on correlations among fatty acids should be considered with caution as they are just phenotypic correlations.

**Mapped QTLs**

The hybrid SA65T holds several QTLs with alleles inherited from both *Elaeis* genomes, while no statistical effect of any *E. guineensis* allele was evidenced in the *E. guineensis* parent PO3228D. This is mainly due to the high homozygosity rate of PO3228D and probably to the narrow genetic diversity of its Deli origin descending from an early introduction of only four “historic” palms to Indonesia (Hartley 1988). The confidence interval is rather large for many of the detected QTLs, and such regions represent several Mb of genomic DNA considering the large genome size of *Elaeis* species, of about 1.9 Gb in oil palm (Zieler et al. 2010). Many functional genes and regulatory elements may underlie a single QTL.

In our full-length *Elaeis* cDNA libraries, the number of distinct expressed copies for our five genes KAS I, KAS II, SAD, FATA, and FATB were in accordance with the cDNA data published in *E. guineensis* by Bourgis et al. (2011) and Tranbarger et al. (2011). They were proved as distinct by mapping (different loci and not allelic variants). They might be the true number of copies in both *E. oleifera* and *E. guineensis* genomes for these genes involved in the proportion of C18:1. The intra-gene SNP markers of the *Elaeis* genes SAD and FATA (respectively, sEgOPGP00038 and sEgOPGP00050) co-localized and were mapped on LG 6 within the confidence intervals overlap of the five QTLs for C16:1, C18:0, C18:1, C20:0, and IV (Fig. 2 and Table 3). The SNP marker associated to SAD could indicate that this mapped gene is involved in the final proportions of C18:0 and C18:1. The *E. oleifera* versus *E. guineensis* allele effects of SAD are coherent on the two traits as, respectively, negative and positive. Also, based on its SNP marker, the FATA enzyme, which is known to export C18:1 from plastid to ER, is a good candidate to explain the QTL of C18:1 in that region. Also, the gene KAS I seems to be a good candidate underlined by the putative QTL for C14:0 on LG 15. KAS II (chain lengthening of C16:0 to C18:0) mapped at that same locus could also support the accuracy of this QTL, as it has been demonstrated that KAS II enzyme can use C10:0 to C14:0 as substrates, although substrate C16:0 is preferred (Shimakata and Stumpf 1982). Other gene SNP markers were mapped outside QTLs. This fact might be due to insufficient explained phenotypic variation (true absence of QTL at their positions) and/or again to the limited size of our mapping population for detecting co-localized QTLs (undetected QTL at these positions). Nevertheless, these SNP markers are useful information upon FATB and other gene copies of KAS I and SAD in the *Elaeis* genomes. We will not discuss neither cannot interpret without complementary data the differential expression of the mapped genes at the species level, only note the absence of clear relation with QTLs detected, except the co-localized SAD and FATA on LG 6.

**Conclusion**

The dense microsatellite linkage map of our interspecific pseudo-backcross represents the 16 pairs of homologous
In the Elaeis genus from which we could trace segregating alleles from both E. oleifera and E. guineensis grandparents. The number of mapped SSR loci, with accurate relative linear orders, and their molecular hyper-variability make this resource valuable for other research centers wishing to undertake rapidly any genetic map in Elaeis breeding materials. The probable absence of chromosome rearrangement between the E. oleifera and E. guineensis genomes would ease the identification, validation, and use of QTL markers in both species.

The QTL positions, the E. guineensis or E. oleifera species origin of the QTL marker alleles, and the estimated allelic effects at the QTLs were in good coherence with both the knowledge of the oil biosynthesis pathway in plants and with the individual correlations estimated between the fatty acid proportions in the palm oil. Moreover, one chromosome region on LG 6 presented good candidates for enzyme genes FATA and SAD involved in the variation of the C18:1 proportion in palm oil.

This preliminary study represents a first stone in view to assist a backcross strategy to exploit genetic differences at level of fatty acid composition inside the Elaeis genus. Our results are specific to our mapping population, and QTL information should be confirmed in other independent and larger Elaeis segregating populations. Since species of the Elaeis genus are outcrossing species, QTLs mapped on our genetic background may not hold on other crosses due to loss of polymorphism and to the fact that family linkage disequilibrium structure may not resemble linkage disequilibrium at population levels. Therefore, the markers identified as linked to the QTLs for fatty acid composition may not be useful while screening other crosses in a MAS strategy. For assigning known genes to QTLs and for practical MAS, the next step of our study will be to search all genes known to be involved in fatty acid synthesis, in our Elaeis cDNA libraries of pulp, and to identify and map their polymorphic intra-gene SNPs. On the other hand, it will be possible to anchor the genetic map(s) to a whole-genome sequence of Elaeis guineensis, as soon as this latter will be available, in order to search SSRs and SNPs for a QTL fine-mapping strategy. Also, such anchoring allows us to identify candidate genes or genes known to be involved in fatty acid synthesis among the genes underlying the fined-mapped QTL regions, for which intra-gene markers may be developed for practical use of MAS (once candidate genes are validated). At the phenotypic level, the individual palms of the mapping population(s) would be characterized for the kinetic of the palm oil biosynthesis, by fatty acid analysis all along the development of ripening fruits, and for the transcript profiles of the related genes, by parallel RT-QPR analysis.

The current reference method of MAS for quantitative traits is the genomic selection (GS) developed by Meuwissen et al. (2001). In GS, all marker effects are estimated simultaneously in the training population and used to estimate the value of selection candidates, without testing the significance of individual marker effects. This minimizes the bias of marker effects and is allows capturing small effects. Odegard et al. (2009) showed that GS could be efficient in introgressing a polygenic trait from an inferior to a superior population by backcrossing. As our pseudo-backcross used E. guineensis individuals from elite families, this is a relevant starting point to introgress the fatty acid profile of E. oleifera into breeding populations of E. guineensis. The progeny SA569 could be used as training population, while the test population would be produced by backcrossing its best individuals (in terms of fatty acid profile) to elite E. guineensis genitors. The availability of intra-gene SNPs should also prove useful, as Meuwissen and Goddard (2010) showed that including causative mutations increased the accuracy of GS. However, the size of the training population is a key parameter of accuracy. A study is therefore required to estimate the accuracy that GS could achieve with our dataset. This could be done by cross-validation.

Acknowledgments The experiment was part of a Ph.D. fellowship granted by the Centro de Investigaciones en Palma de Aceite (CENIPALMA), Colombia. This study was funded by the International Consortium Oil Palm Genome Projects (OPGP) and its members: Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), France; Instituto Vasco de Investigación y Desarrollo Agrario (NEIKER), Spain; Advanced Agricolecolical Research Sdn. Bhd (AAR), Malaysia; PT Astra Agro Lestari Tbk, Indonesia; PT Inti Indosawit Subur and its affiliates (Asian Agri Group), Indonesia; PT Bakrie Sumatera Plantations Tbk, Indonesia; Centro de Investigaciones en Palma de Aceite (CENIPALMA), Colombia; Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Brazil; Felda Agricultural Services Sdn Bhd, Malaysia; Genting Plantations Berhad, Malaysia; IOI Corporation Berhad, Malaysia; Indonesien Oil Palm Research Institute (IOPRI), Indonesia; PT Matahari Kahuripan Indonesia, Indonesia; PT SMART Tbk, Indonesia; PT Sampoerna Agro Tbk, Indonesia; and SOCFIN Group, Luxemburg. Authors are thankful to Plantation Hacienda La Cabaña S.A. (Colombia), Indupalma LTAD (Colombia), the Centre de Recherche Agronomique des Plantes Pérennes (CRA-PP) of the Institut National de Recherche Agronomique du Bénin (INRAB), the Indonesian Oil Palm Research Institute (IOPRI), and the Centre National de Recherche Agronomique (CNRA) in Côte d’Ivoire for providing vegetal samples and/or phenotypic data. We would like to thank the Genoscope-CEA (Evry, France) for its support and the 454 Roche/454 FLX Pyrosequencing of the Elaeis cDNA libraries. We wish to thank Dr. Timothy Tranbarger for his early provision of the PCR primer sequence information of the EST-SSR loci used in this study. We are also grateful to Dr. Brigitte Courtois, Dr. Jean-Marc Gion, Dr. Brigitte Margin, and Dr. Dominique This for their scientific advice. Finally, we would like to thank the reviewers of this journal for their corrections and kind help to improve this article.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References


Choo YM, Ma AN, Yap SC (1997) Carotenoids, vitamin E and sterols in oils from Elaeis guineensis, Elaeis oleifera and their hybrids. Palm Oil Dev 27:1–9


Ha BK, Monteros MJ, Boerma HR (2010) Development of SNP assays for high-density microsatellite markers (SSR) during the oil palm variety selection and production processes. “International seminar on Oil Palm Genomics and its application to Oil Palm Breeding.” ISOPB (The International Society for Oil Palm Breeders), conferences, Kuala Lumpur Convention Centre Malaysia

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 7:309–3435


Van Ooijen JW (2004) MapQTL5, software for mapping of quantitative loci in experimental populations. Kyazma BV. Wageningen, the Netherlands