the plant journal

SEB Society for Experimental Biology

The Plant Journal (2016) **87,** 423–441 doi: 10.1111/tpj.13208

RESOURCE

Gene coexpression network analysis of oil biosynthesis in an interspecific backcross of oil palm

Chloé Guerin[1], Thierry Joët[2], Julien Serret[2], Philippe Lashermes[2], Virginie Vaissayre[2], Mawussé D. T. Agbessi[3], Thierry Beulé[3], Dany Severac[4], Philippe Amblard[1], James Tregear[2], Tristan Durand-Gasselin[1], Fabienne Morcillo[3,†] and Stéphane Dussert[2,†,*]

[1]*PalmElit SAS, Montferrier-sur-Lez F-34980, France,*

[2]*IRD, UMR DIADE, 911 Av. Agropolis, Montpellier 34394, France,*

[3]*CIRAD, UMR DIADE, Montpellier 34398, France, and*

[4]*MGX-Montpellier GenomiX, c/o Institut de Génomique Fonctionnelle, 141 Rue de la Cardonille, Montpellier Cedex 5 34094, France*

Received 23 March 2016; revised 27 April 2016; accepted 27 April 2016; published online 19 July 2016.

*For correspondence (e-mail stephane.dussert@ird.fr).

†These authors contributed equally.

SUMMARY

Global demand for vegetable oils is increasing at a dramatic rate, while our understanding of the regulation of oil biosynthesis in plants remains limited. To gain insights into the mechanisms that govern oil synthesis and fatty acid (FA) composition in the oil palm fruit, we used a multilevel approach combining gene coexpression analysis, quantification of allele-specific expression and joint multivariate analysis of transcriptomic and lipid data, in an interspecific backcross population between the African oil palm, *Elaeis guineensis*, and the American oil palm, *Elaeis oleifera*, which display contrasting oil contents and FA compositions. The gene coexpression network produced revealed tight transcriptional coordination of fatty acid synthesis (FAS) in the plastid with sugar sensing, plastidial glycolysis, transient starch storage and carbon recapture pathways. It also revealed a concerted regulation, along with FAS, of both the transfer of nascent FA to the endoplasmic reticulum, where triacylglycerol assembly occurs, and of the production of glycerol-3-phosphate, which provides the backbone of triacylglycerols. Plastid biogenesis and auxin transport were the two other biological processes most tightly connected to FAS in the network. In addition to WRINKLED1, a transcription factor (TF) known to activate FAS genes, two novel TFs, termed NF-YB-1 and ZFP-1, were found at the core of the FAS module. The saturated FA content of palm oil appeared to vary above all in relation to the level of transcripts of the gene coding for β-ketoacyl-acyl carrier protein synthase II. Our findings should facilitate the development of breeding and engineering strategies in this and other oil crops.

Keywords: *Elaeis guineensis*, oil palm, fruit, fatty acid, glycolysis, starch, plastid, metabolism, palmitic acid, coexpression network.

INTRODUCTION

The growth of the world's population and the rising standards of living in emerging economies have greatly increased the global demand for vegetable oils, for both food and non-food products, including oleochemicals, cosmetics and biodiesel (Alexandratos and Bruinsma, 2012). Identifying the factors that regulate oil biosynthesis in plants is therefore a critical issue. Most of the oil we consume is extracted from seeds (e.g. soybean, rapeseed) and fruits (e.g. oil palm, olive) that store triacylglycerols (TAG), i.e. glycerol esterified with three fatty acids (FAs), in specialized tissues. The oil content of the mesocarp of the oil palm fruit is about 90% dry matter (DM), which is higher than in most oilseeds. This feature not only contributes to the exceptional yield of oil per hectare of this crop (Barcelos *et al.*, 2015), but also makes this tissue an attractive model for studying fundamental aspects of oil biosynthesis in plants.

The isolation of an Arabidopsis mutant specifically impaired in seed oil accumulation, termed *wrinkled1,* was a milestone in our understanding of the regulation of oil biosynthesis (Focks and Benning, 1998). The transcription factor (TF) WRINKLED1 (WRI1) directly activates genes involved in late glycolysis and fatty acid (FA) synthesis during seed maturation (Baud *et al.*, 2007; Maeo *et al.*, 2009). Earlier transcriptome analyses of the developing oil palm mesocarp revealed that a *WRI1* gene (*EgWRI1-1*) was massively transcribed at the onset of oil accumulation, suggesting that WRI1 also regulates FA synthesis in non-seed tissues (Bourgis *et al.*, 2011; Tranbarger *et al.*, 2011). The ability of *EgWRI1-1* to trigger FA synthesis was validated using transient expression in tobacco leaves (Dussert *et al.*, 2013). *EgWRI1-1* was also shown to complement Arabidopsis *wri1* mutants (Ma *et al.*, 2013). WRI1 is now considered to be a ubiquitous regulator of oil synthesis in higher plants (Ma *et al.*, 2013). However, despite this considerable step forward, our knowledge of the factors controlling oil accumulation in plant tissues remains partial. For instance, oil accumulation in Arabidopsis *wri1* mutant lines is not completely shut down (it is about two to four times lower) (Focks and Benning, 1998; Baud *et al.*, 2007), indicating that other central regulatory processes remain to be identified.

Identification of factors governing the FA composition of vegetable oils is also indispensable since it determines their uses (Dyer *et al.*, 2008). Palm oil is well known for its high level of saturated FA, mostly palmitic acid (16:0). Saturated oils have many advantages for the food industry thanks to their high oxidative stability and high melting point, making them a good alternative to trans fats (hydrogenated oils). Quantitative trait locus (QTL) analyses performed to date have shown limitations in their potential to explain the high 16:0 content of palm oil (Montoya *et al.*, 2013). Transcriptome comparisons of the three oil-accumulating tissues (mesocarp, endosperm and embryo) of oil palm did not resolve the molecular mechanisms involved either (Dussert *et al.*, 2013).

Genome-wide coexpression analysis is now widely recognized as an effective approach to elucidate the molecular determinants of various processes in plant biology (Aoki *et al.*, 2007; Usadel *et al.*, 2009). For instance, it enabled the identification of new enzymes involved in different metabolic pathways (Okazaki *et al.*, 2009), the discovery of key regulators of essential processes such as seed dormancy and germination (Bassel *et al.*, 2011) and, more generally, a complete overview of the transcriptional circuits underlying developmental (Palumbo *et al.*, 2014) and metabolic (Wei *et al.*, 2006) processes. Coexpression analysis is based on the 'guilt-by-association' paradigm (Saito *et al.*, 2008), which stipulates that two genes displaying transcript levels that are correlated across a wide range of conditions (e.g. developmental stages, genotypes, environments) are likely to be co-regulated or to be a TF–target gene pair. However, with the exception of a few studies, for example in the microalga *Chlamydomonas* (Gargouri *et al.*, 2015) and Arabidopsis (Mentzen *et al.*, 2008), up to now coexpression analysis has rarely been used to investigate oil biosynthesis in plants.

Together with the identification of gene-to-gene relationships, studies linking transcript and metabolic profiles have also proved to be useful for identifying new enzymes that could be key regulators of metabolic pathways (Bassel *et al.*, 2012). Quantitative trait transcripts (QTT), i.e. genes whose transcript amounts are significantly correlated with the quantitative trait of interest across genotypes (Vallabhaneni and Wurtzel, 2009) or environments (Joët *et al.*, 2010), are strong candidate regulatory genes for the rate-limiting steps of biosynthetic pathways. These genes can be identified by linear regression between transcript and metabolite levels or through multivariate analysis of transcriptomic-metabolomic datasets. The latter approach was recently applied to investigate changes in membrane glycerolipid composition in Arabidopsis (Szymanski *et al.*, 2014).

When designing a coexpression study, the challenge is to obtain large but continuous variation in gene expression levels. Among the sources of genetic variation that may alter the transcriptome conveniently for coexpression analysis, interspecific backcrossed hybrids (BCs) have been shown to be suitable for network inference and QTT detection (Kirst *et al.*, 2004; Filteau *et al.*, 2013). There are two species in the genus *Elaeis* which are interfertile. *Elaeis guineensis* (Eg) originates from Africa while *Elaeis oleifera* (Eo) occurs naturally in South and Central America. Like Eg, Eo also accumulates oil in the mesocarp, but to a much lesser extent (15–25% DM) (Barcelos *et al.*, 2015). Another use can be made of interspecific oil palm hybrids on account of the contrasting FA composition of mesocarp oil in the two species. In particular, the palmitic acid (16:0) level in Eo is about two times lower than in Eg (about 25 and 45%, respectively).

However, in some circumstances, using an interspecific BC population may have drawbacks. In an interspecific cross, divergence for *cis*- and *trans*-acting regulatory elements may indeed unbalance the expression of the parental alleles (Landry *et al.*, 2007; Combes *et al.*, 2015), thereby jeopardizing the detection of correlations at the gene level. In an extremely imbalanced context, coexpression analysis should theoretically be performed at the allele level. Because we know that WRI1 activates the transcription of fatty acid synthesis (FAS) genes via binding to AW-box elements in their promoter regions (Maeo *et al.*, 2009), the FA biosynthetic pathway is an appropriate model for testing the importance of *cis* and *trans* effects in interspecific hybrids.

Most published coexpression analyses in plant biology have taken advantage of the large microarray datasets that are publicly available. Unfortunately, no such datasets are available for oil palm. However, since the release of the Eg genome sequence (Singh *et al.*, 2013), it has become possible to use high-throughput RNA sequencing (RNA-seq) technologies for the measurement of gene expression in a large number of samples. This technology also has the major advantage of allowing both allele-specific expression (ASE) measurements (Gaur *et al.*, 2013) and coexpression analyses (Pfeifer *et al.*, 2014) using the same dataset. In the present work we used a multilevel approach to better understand oil biosynthesis in the oil palm mesocarp. We combined gene-to-gene coexpression analysis using a mixed guide-gene and non-targeted strategy (Aoki *et al.*, 2007), quantification of ASE and joint multivariate analysis of transcriptomic and lipid data. Using a new RNA-seq dataset (59 transcriptomes), specifically designed to exploit both genotypic and developmental variations in lipid gene expression, we built a gene coexpression network of the oil palm mesocarp, investigated the biological relevance of coexpression modules and searched for novel regulators of FA biosynthesis and the molecular mechanisms underlying the high 16:0 level of palm oil.

RESULTS

Assessment of *cis*- and *trans*-regulatory divergences for fatty acid synthesis in interspecific hybrid materials

The mesocarp oil content was very high, from 60 to 80% DM (Figure 1a), in all BCs studied, indicating no major alteration of oil biosynthesis in these hybrid materials. 16:0 and oleic acid (18:1) were the major FAs in all BCs, accounting for up to 80% of total FA (Table S1 in the Supporting Information). The percentage of 16:0 FAs in oil varied considerably within the BC population studied, from 32 to 47% (Figure 1b), and was negatively correlated with the percentage of 18:1 FAs (Figure 1c). Mesocarp oil content was not correlated with any of the individual FAs (Table S2). Interspecific single nucleotide polymorphisms (SNPs) enabled the determination of genotype (EgEg versus EgEo) at the *WRI1-1* and FAS gene loci in each BC. We then quantified ASE in both homozygous and heterozygous BCs at the *WRI1-1* locus and for heterozygous BCs at FAS gene loci. We did not detect any significant differences in FAS gene ASE between homozygous and heterozygous BCs at the *WRI1-1* locus, with the exception of *MAT*, which encodes a malonyl-CoA:ACP malonyltransferase and for which the Eg allele displayed slightly (but significantly, $P = 0.03$) higher transcript levels than the Eo allele in heterozygous genotypes (Figure 2a). Likewise, the *WRI1-1* genotype had no influence on correlations between FAS gene expression levels (Figure 2b). The slope and intercept of the regression lines were highly similar in homozygous and heterozygous genotypes at the *WRI1-1* locus. To better understand the independence of accumulation of FAS gene allele transcripts with respect to *WRI1-1* gene parentage, we compared the amino acid (AA) sequence of the two parental TF isoforms and analysed the proximal region of several FAS genes. The AA sequences of the Eg and Eo WRI1-1 proteins were highly conserved (Figure S1), an observation consistent with the absence of major differential *trans* effects on FAS gene ASE. The number, position and strand of AW motifs in the promoter region of FA genes were also remarkably conserved between Eg and Eo (Figure 2c). Even more noteworthy was the conservation of the motif sequence at a given position in the two species. Figure 2(c) depicts the sequences of the three AW-box elements present in the proximal upstream region of *ACC/CTα-1*, which codes for the carboxyltransferase α-subunit of heteromeric acetyl-CoA carboxylase (ACCase). Within each parental species, all three elements conformed to the consensus AW motif [CnTnG](n)7[CG] but a differences of a few nucleotides were observed among the three motif sequences. By contrast, at each of the three positions, the motif sequence in

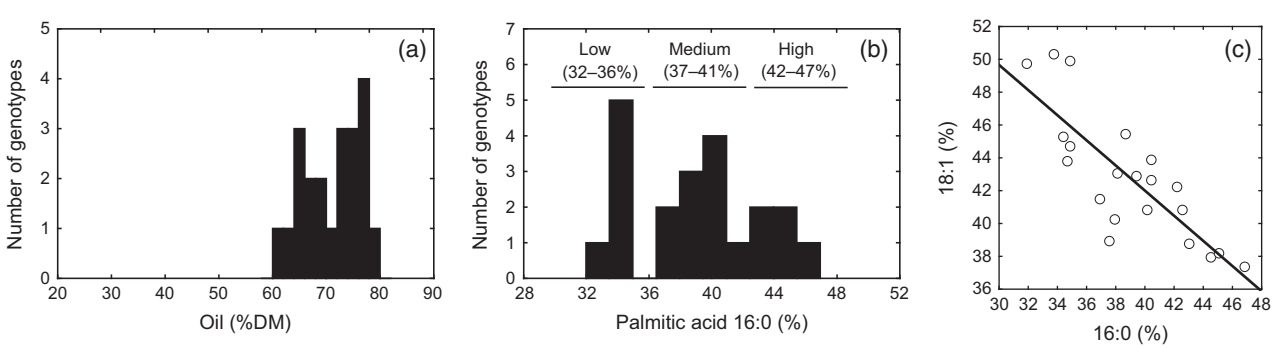


Figure 1. Distribution of mesocarp oil content and 16:0 percentage in the backcrossed hybrid (BC) population
Distribution of mesocarp oil content (a) and 16:0 percentage (b) in the BC population. BCs were classified in three groups according to the 16:0 level of the oil extracted from the mature mesocarp (low, medium and high 16:0).
(c) Negative correlation between palmitic and oleic (18:1) acids in the BC population.

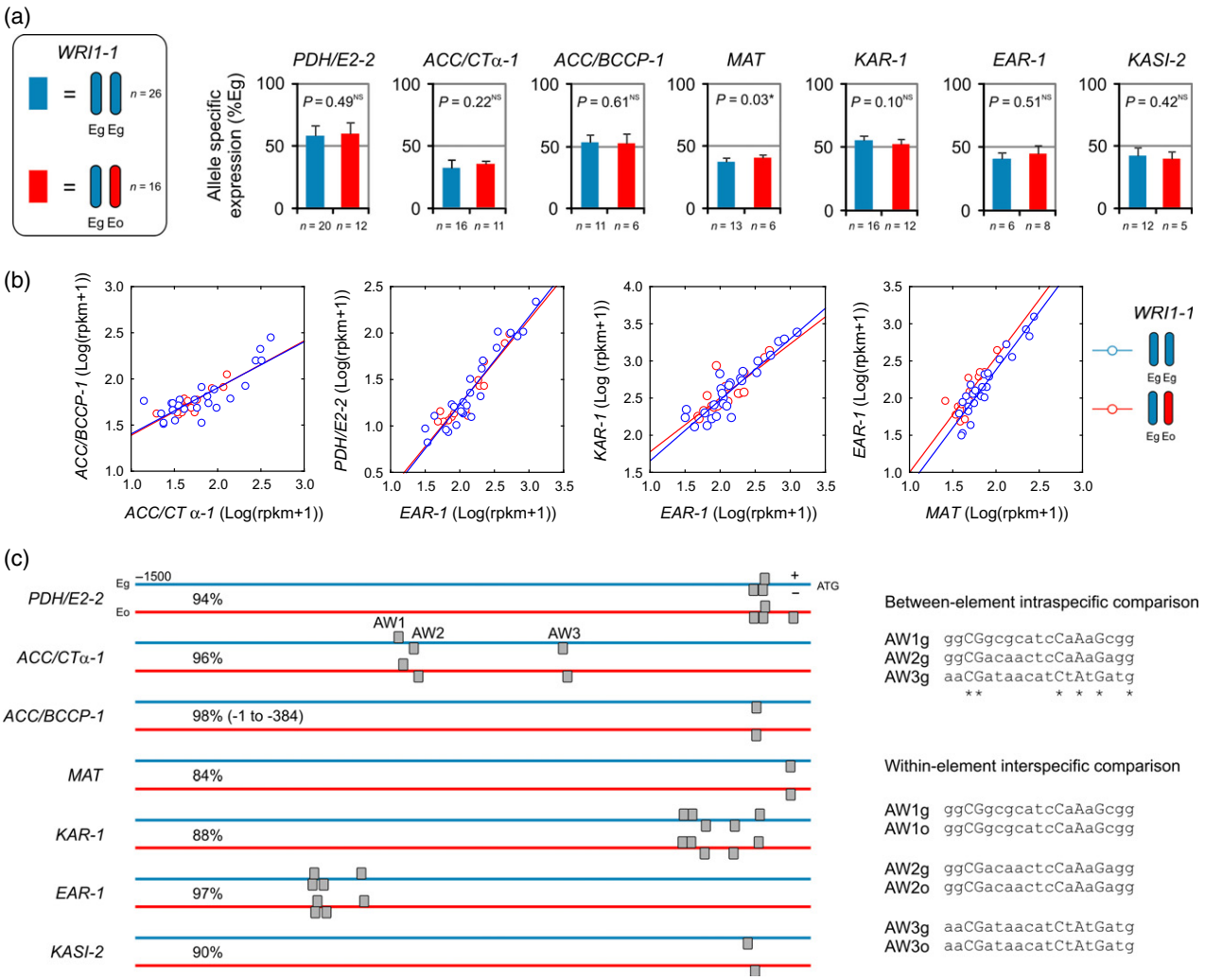


Figure 2. Effect of the genotype at the *WRI1-1* locus on transcription of fatty acid synthesis (FAS) genes.
(a) Allele specific expression (ASE) of seven FAS genes, as quantified by the proportion of *Elaeis guineensis* mRNA (%Eg), in homozygous (blue bars) and heterozygous (red bars) genotypes at the *WRI1-1* locus. The number of samples (*n*) used for ASE quantification is given below each bar.
(b) Correlation between transcript abundance of various FAS genes in homozygous (blue circles and line) and heterozygous (red circles and line) genotypes at the *WRI1-1* locus. rpkm, reads per kilobase and million reads.
(C) Number, position and strand of AW elements in the proximal upstream region of *E. guineensis* (blue sequence) and *Elaeis oleifera* (red sequence) alleles of seven FAS genes. The nucleotide identity of the proximal region of parental alleles is given. The nucleotide sequences of the three elements found in the promoter of *ACC/CTα-1* are also compared at the intraspecific and interspecific levels.

the two parental species was identical. This led us to hypothesize that *cis* effects also only play a minor role in determining species-specific differences in the WRI1–FAS gene regulatory system.

The oil palm mesocarp coexpression network

Two rounds of guided coexpression analysis using an $|R|$ threshold of 0.8 gathered 489 genes and the Markov clustering algorithm identified 10 major clusters, hereafter referred to as Modules 1–10 (Figures 3 and S2, Table S3). Module 1 contained 23 genes involved in the core FA biosynthetic machinery (Figure 3). This module was also characterized by a large number of genes involved in glycolysis and glycolysis-related processes, plastid biogenesis and starch metabolism. Module 2 had the second highest number of lipid-related genes. It was also rich in genes coding for enzymes of glycolysis and the pentose phosphate pathway (PPP). It was connected to Module 1 through a relatively small number of nodes. In contrast to Module 2, Module 3 closely surrounded Module 1. Correlations between the nodes of these two modules were all positive. Module 3 was particularly enriched in genes involved in several key aspects of plastid biogenesis and function: plastid division (*FtsZ2*), protein import (*TOC75III-1* and *-2*), protein folding (*Cpn60b-1* and *-2*), thylakoid membrane protein insertion (*ALB3*), plastid RNA editing

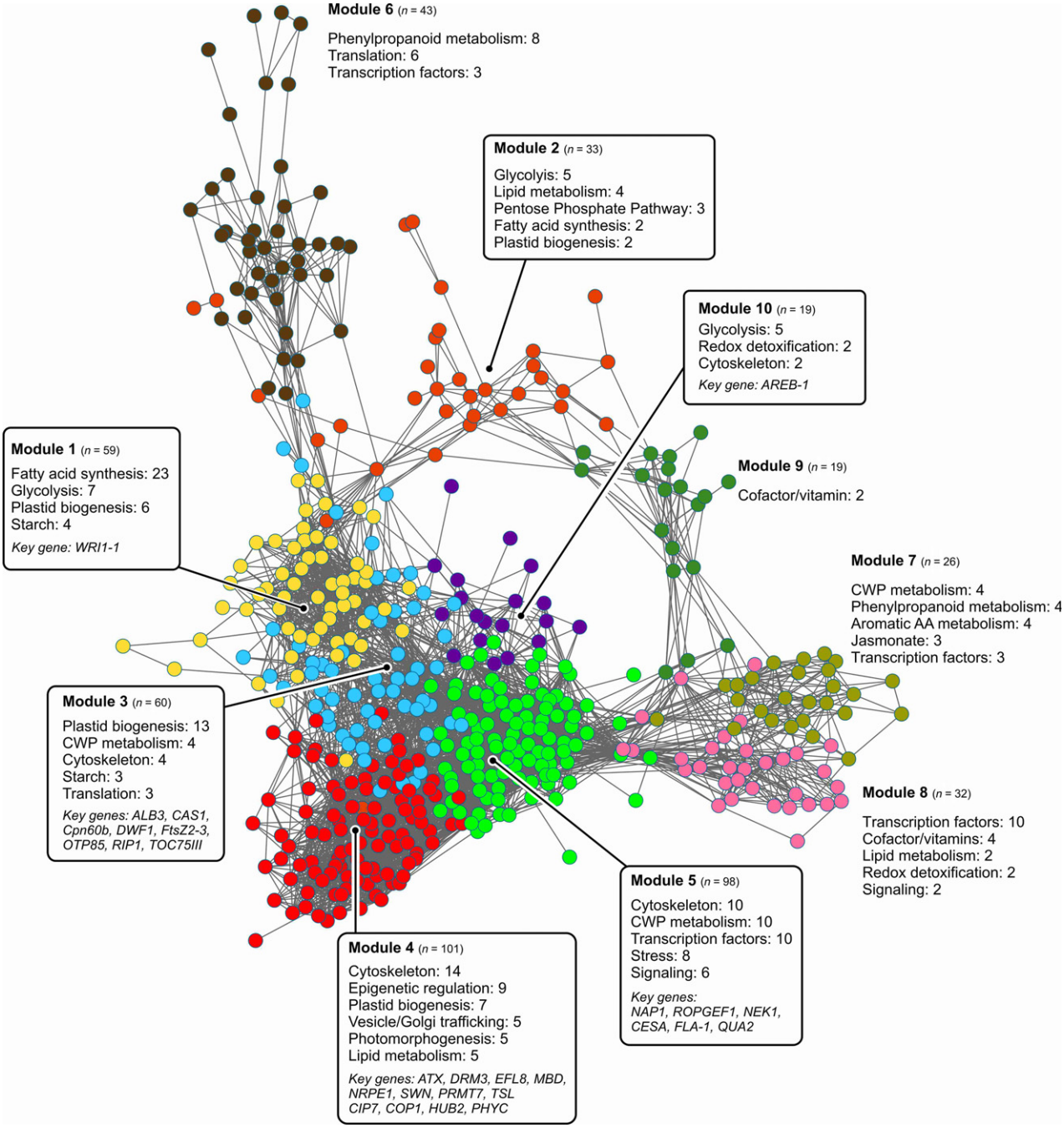


Figure 3. The oil palm mesocarp coexpression network.
The network was constructed using 35 lipid-related guide genes (Figure S7), two rounds of coexpression analysis and an $|R|$ threshold of 0.8. The Markov cluster algorithm was used for module detection. Modules are coloured to reveal the modular organization of the network. The number of nodes (genes) per module is given in brackets next to the name of the module. Each of the 489 genes of the network was manually annotated and assigned to a functional group. Within each module, the most represented functional groups and the number of nodes per functional group are given. Key genes discussed in the text are also shown. CWP, cell wall polysaccharide.

(*OTP85, RIP1-1* and *-2*). Two genes (*CAS1* and *DWF1*) coding for key enzymes of the sterol biosynthesis pathway, i.e. cycloartenol synthase and Δ24-sterol reductase, respectively, were also found in Module 3.

Two large modules (Modules 4 and 5) and one small module (Module 10) were located in the immediate vicinity of Module 3. Module 10 was enriched in glycolysis-related genes, most of which were positively linked to a TF of the

bZIP family (*AREB-1*). Module 4 was partly topologically intertwined with Module 3. In addition to the many proteins involved in actin and microtubule organization and dynamics, this module was enriched in genes involved in the epigenetic regulation of gene expression. All major mechanisms of epigenetic control were represented by one or several genes in Module 4: DNA methylation and demethylation (e.g. *MBD-1*), small RNA-directed DNA methylation (*NRPE1* and *DRM3*) and histone modifications. Among the diverse histone modifications that contribute to chromatin regulation, histone lysine (*ATX-1*, *EFL8*, *SWN*) and arginine (*PRMT7*) methylation, phosphorylation (*TSL*) and ubiquitination (*HUB2*) appeared to be involved in the processes governed by Module 4. Among the other functional categories that predominated in Module 4, photomorphogenesis was specifically overrepresented. It indeed contained a photoreceptor (*PHYC*), two copies of *COP1*, which codes for an E3 ubiquitin ligase that plays a key role in the degradation of light-induced TFs in darkness, and a potential direct downstream target of COP1 (*CIP7*).

Like Module 4, the adjacent Module 5 was particularly enriched in cytoskeleton-related genes and in particular contained three upstream regulators of actin and microtubule organization (*NAP1*, *ROPGEF1* and *NEK1*). Module 5 was also characterized by its richness in genes involved in cell wall polysaccharide metabolism, such as *CESA* and *QUA2*, which play a key role in cellulose and pectin biosynthesis, respectively, or *FLA-1*, which codes for a fasciclin-like arabinogalactan protein. A dense subnetwork of negative correlations topologically overlapped Module 5 (Figure S2). About 95% of these negative relationships were intramodular and 87% of them involved only 10 genes: four osmotin-like isoforms (*OSM34-1* to *-4*), two chitinase-like isoforms (*Chi-2* and *-3*), a sucrose synthase (*SUSY-1*), a hypersensitive response protein (*HIR1*), an aldehyde dehydrogenase (*ALDH2B-1*) and a plant natriuretic peptide (*EGC-1*). Finally, at the periphery of the network, Modules 6 and 7 could be distinguished by their high numbers of genes involved in the metabolism of phenylpropanoids, aromatic acids and cell-wall polysaccharides, suggesting they control important cell wall rearrangements in mesocarp cells.

Coexpression analysis assigned lipid genes to specialized functions and specific regulatory processes

With the exception of ketoacyl-ACP synthase III (KASIII) and the carboxyltransferase β-subunit of heteromeric ACCase (ACC/CTβ), all the biosynthetic steps necessary for the *de novo* formation of acyl chains in the plastid were represented by at least one gene in Module 1 (Figure 4). This module can thus be seen as the core module for biosynthesis of FA and oil. The absence of *ACC/CTβ-1* may be explained by the fact it is the only gene of the plastid genome among FAS genes and may thus be governed by different regulatory factors. For some of the FAS enzymes, all paralogues that had been previously identified by analysis of developmental time series (Dussert *et al.*, 2013) were retrieved in Module 1. For other enzymatic steps,

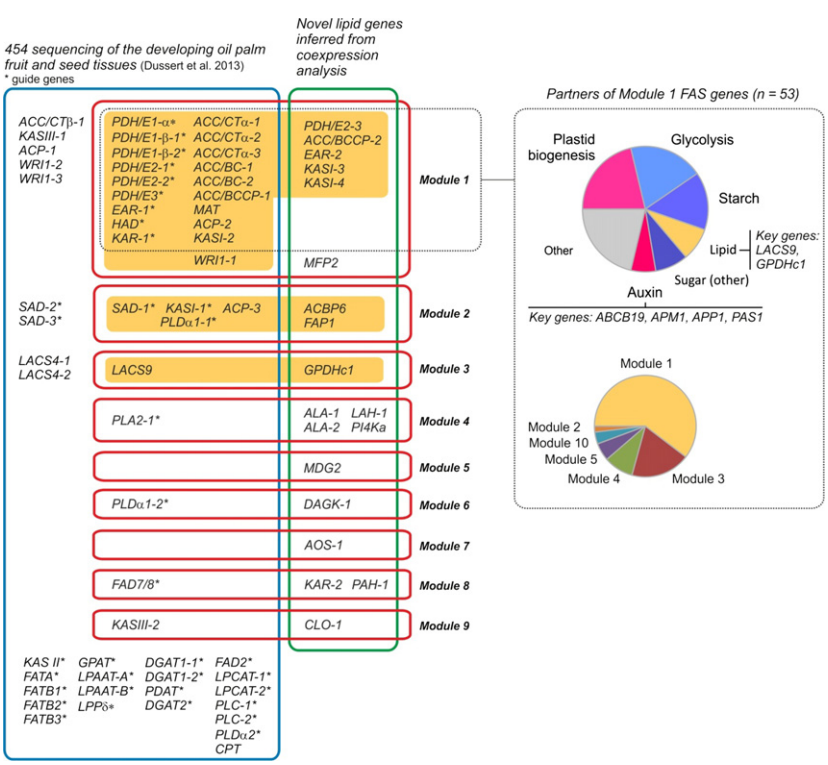


Figure 4. Lipid genes in the oil palm mesocarp coexpression network.
Lipid genes in Modules 1 to 9 are shown in red boxes (Module 10 contains no lipid genes). The green box separates the lipid genes identified in the present work by coexpression analysis of high-throughput RNA sequencing data from those previously revealed by 454 pyrosequencing-based transcriptome analysis of developing oil palm fruit and seed tissues, which are grouped in the blue box. Lipid genes thought to be primarily involved in oil biosynthesis pathways are in yellow boxes. The 23 fatty acid synthesis (FAS) genes in Module 1 are directly connected to 53 partners only, whose functions and module distribution are given in the pie charts to the right of the graph.

coexpression analysis revealed previously undescribed paralogues: *EAR-2*, which codes for an enoyl-ACP reductase, or *KASI-3* and *KASI-4,* which encode ketoacyl-ACP synthase I isoforms, for instance. The modular organization of the oil palm mesocarp coexpression network also provided insights into functional specialization between lipid gene paralogues in this species. For instance, of the three previously identified stearate desaturase (SAD) genes, only *SAD-1* was retained in the coexpression network and might therefore correspond to the main SAD isoform for 18:1 production (Figure 4). A similar observation was made for the acyl-carrier protein isoforms ACP-2 and ACP-3, which were found in Module 1 and at the interface between Module 1 and Module 2, respectively, whereas ACP-1 was not retrieved in the network. Several modules of the coexpression network contained lipid genes that do not code for biosynthetic enzymes but for other key aspects of acyl-lipid metabolism, such as lipid trafficking, signalling or catabolism. Among these, *FAP1* and *ACBP6*, which code for a plastidial FA-binding protein and a cytosolic acyl-CoA-binding protein (ACBP), respectively (both of which belong to Module 2), are of particular interest.

The partners of the 23 FAS genes identified in Module 1 were examined to uncover possible key *trans*-module regulatory processes involved in the activation of the core FA biosynthetic machinery (Figure 4). Almost half the direct partners of Module 1 FAS genes did not belong to Module 1, revealing the importance of between-module connections. The majority of the 53 partners of Module 1 FAS genes were genes coding for enzymes involved in glycolysis and starch metabolism and proteins needed for plastid biogenesis and function. Moreover, four genes involved in auxin transport, *ABCB19*, *APM1*, *APP1* and *PAS1*, were coexpressed with FAS genes. *ABCB19* was in fact one of the few connector hubs (high between-module degree) detected in the network, while *APM1* appeared to be a provincial hub (high intra-module degree) in Module 4 (Figure S3).

Among FAS partners, two key lipid genes, *LACS9* and *GPDHc1*, were found not in Module 1 but in Module 3. The long-chain acyl-CoA synthetase LACS9 contributes to the transfer of FA from the plastid to the endoplasmic reticulum (ER) for TAG assembly and the glycerol-3-P dehydrogenase GPDHc1 is involved in the synthesis of the glycerol backbone of TAG. These two genes therefore form a noteworthy bridge between FA synthesis in the plastid and TAG assembly in the ER.

Transcriptional orchestration of plastidial glycolysis, transient starch storage, methylglyoxal detoxification and sugar sensing during oil accumulation

Module 1 – the core FAS module – and modules in the immediate vicinity of Module 1 (Modules 2, 3 and 10) were remarkably enriched in genes involved in central carbon metabolism (Figures 3 and 6, Table S3), showing that the large amounts of pyruvate, energy and reducing power required for oil synthesis are supplied thanks to the concerted co-regulation of the route from sucrose to pyruvate and thereafter from pyruvate to FAs. However, in contrast to the FAS pathway, whose enzymatic steps are all quantitatively upregulated together, coexpression analysis found evidence for key transcriptional control points for glycolytic activity in the oil palm mesocarp. Our first noteworthy result was the fine transcriptional tuning of sucrose cleavage activity. Indeed, cell-wall invertase (*cwINV-1*) belonged to Module 3 and was a positive partner of FAS genes, while levels of transcription of *SUSY-1* (Module 5) and FAS genes were negatively correlated. The concomitant repression of *SUSY-1* and activation of *cwINV-1* are potentially the first control point that guides the carbon flux towards oil synthesis. Thereafter, 10 glycolytic steps occur in the pathway from glucose to pyruvate, each step occurring in both the cytosol and the plastid. Of the 20 steps, only five were represented by genes that were direct partners of FAS genes, and most were in Module 1, i.e. hexokinase (HXK1), fructokinase (PFKB-1), ATP-dependent phosphofructokinase (PFKp-1), glyceraldehyde 3-P dehydrogenase (GAPDHp-1) and pyruvate kinase (PKp) (Figure 5, Table S3). All these genes encode plastidial isoforms of the enzyme concerned. Coexpression analysis therefore revealed that the plastidial glycolytic route is only transcriptionally co-regulated with FA biosynthesis in the oil palm mesocarp. The plastidial glucose translocator (GLT1) also belonged to Module 1 and was a direct partner of FAS genes. Proteins involved in importing hexose into plastids and in plastidial glycolysis may thus be control points for carbon funnelling towards oil synthesis. Finally, network analysis revealed the coordinated activation of the non-oxidative phase of PPP and of the PDH bypass during oil synthesis in the oil palm mesocarp (Figure 5).

Coexpression analysis also enabled identification of several processes closely associated with sugar homeostasis and glycolysis. For instance, glycolysis may be a substantial source of methylglyoxal via the fragmentation of triose-phosphates. Methylglyoxal is a highly reactive dicarbonyl compound and a precursor of advanced glycation end-products, and is therefore considered highly toxic. Key enzymes of the cytosolic glyoxalase pathway (such as glyoxalase I and II, and lactate dehydrogenase), as well as the plastidial isoform of the recently described one-step glyoxalase enzyme (Kwon *et al.*, 2013), were all present in the network, suggesting that methylglyoxal detoxification is coordinately activated to handle glycolysis-induced methylglyoxal production in the oil palm mesocarp. As a second example, coexpression analysis revealed the tight transcriptional coordination of FAS genes with genes coding for proteins involved in sugar sensing and signalling, namely HXK, trehalose phosphate synthase (TPS), sucrose

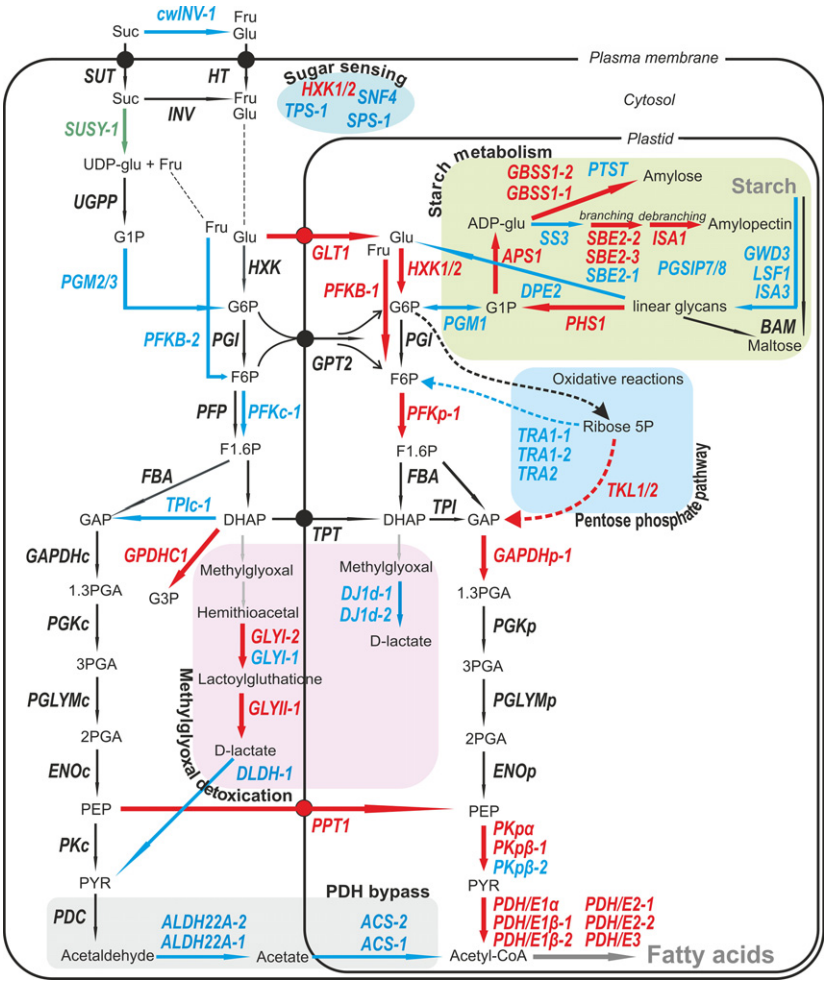


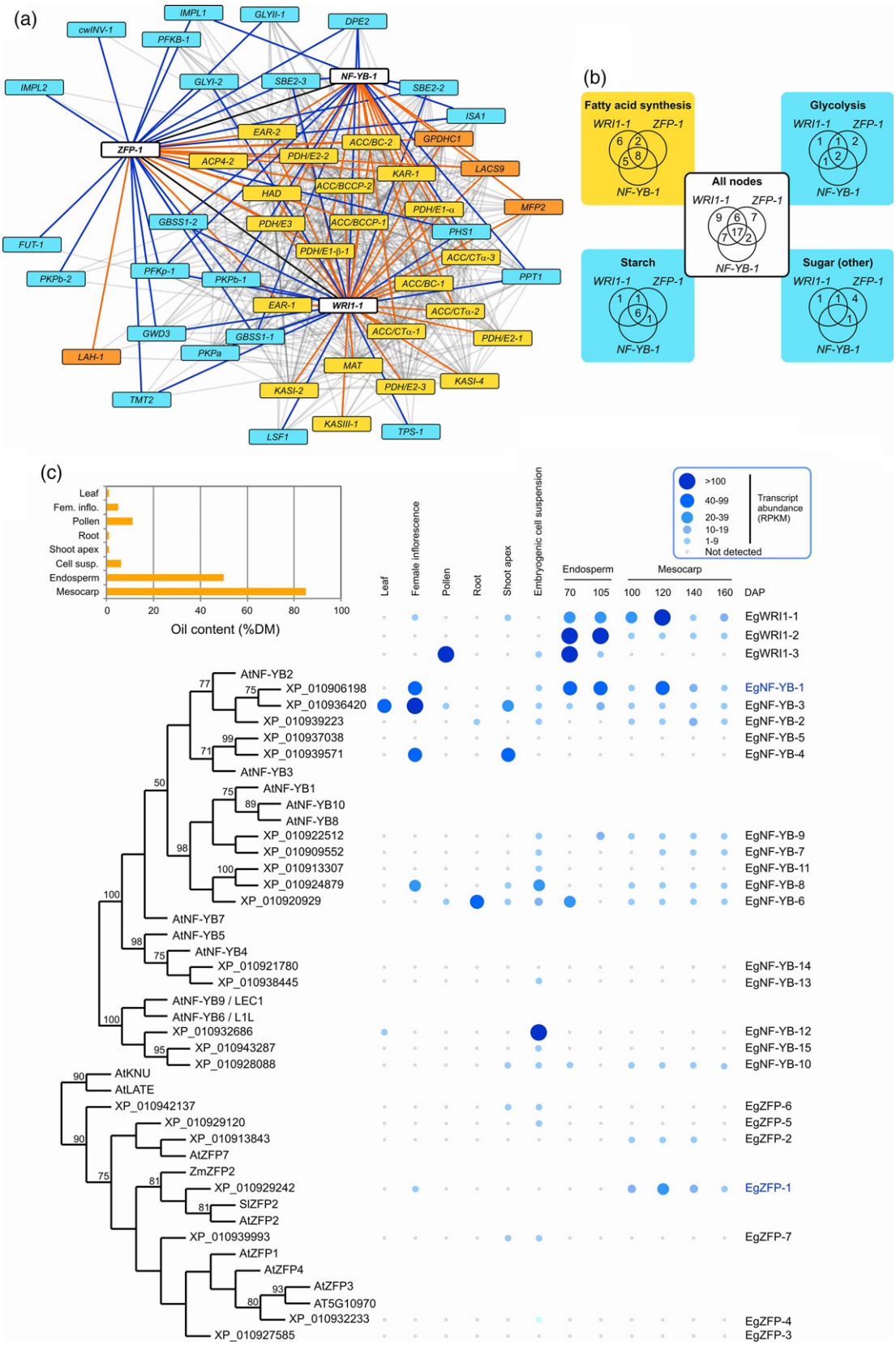
Figure 5. Transcriptional connections between the core fatty acid biosynthetic machinery and central carbon metabolism.
Genes in red were direct partners of fatty acid synthesis (FAS) genes as described in Figure 4. Genes in blue were identified in the mesocarp coexpression network. Genes in green were negatively correlated with FAS genes. Genes in black were not identified by coexpression analysis. For reasons of simplicity, unidirectional arrows are used even if most reactions are theoretically reversible, except for reactions catalyzed by hexokinase, phosphofructokinase, pyruvate decarboxylase, pyruvate dehydrogenase, pyruvate kinase, and most starch-related enzymes. Dashed lines indicate multistep pathways.

phosphate synthase (SPS) and the sucrose non-fermenting-related kinase (SNF4). Our second noteworthy result is the tight link between FA biosynthesis and starch metabolism, with virtually all enzymatic steps in starch synthesis and remobilization being represented in the network (Figure 5). Seven starch-related genes were seen to be direct partners of FAS genes and four of them played a major role as provincial hubs in Module 1 (*GBSS1-1*, *GBSS1-2* and *PHS1*) or as connector hubs (*SBE2-2 and -3*) in the architecture of the network (Figure S3).

The FAS coexpression module contains two other transcription factors in addition to WRI1

Using the 23 FAS genes found in Module 1 as guide genes, and an $|R|$ threshold of 0.7, we built a one-round coexpression subnetwork to uncover other TFs that may regulate FA biosynthesis. In this targeted subnetwork, we searched for all TFs connected to guide genes. The most connected TF was WRI1-1, which was coexpressed with 20 guide genes (Figure 6a). In addition to WRI1-1, we identified two other

Figure 6. The three transcription factors at the core of fatty acid synthesis (FAS) module: WRI1-1, NF-YB-1 and ZFP-1.
(a) A one-round guided coexpression subnetwork using an $|R|$ threshold of 0.7. The 23 FAS genes identified in Module 1 of the two-round coexpression network constructed with an $|R|$ threshold of 0.8 were used as guide genes. The FAS genes are in yellow and the other lipid genes are in orange. Genes involved in glycolysis, starch and other sugar metabolism are in blue. Edges coloured black, red, and blue correspond to connections between transcription factor (TF), TF and lipid genes, TF and sugar-related genes, respectively. Edges coloured grey correspond to connections between lipid- and sugar-related genes.
(b) Venn diagrams depicting the number of partners shared by *WRI1-1*, *NF-YB-1* and *ZFP-1* for all nodes of the subnetwork or for four distinct functional categories.
(c) Phylogenetic analysis of NF-YB-1 and ZFP-1, tissue-specific expression of all members of their respective subfamilies, and tissue oil content (% dry matter). For NF-YB-1, the maximum likelihood tree was built with the 15 oil palm proteins of the subfamily NF-YB of the NUCLEAR FACTOR Y (NF-Y) TF family, termed EgNF-YB-1 to -15, and proteins of this subfamily from Arabidopsis (At). For the phylogenetic analysis of ZFP-1, the eight Arabidopsis members of clade C1-1iAa of the one-domain ZINC FINGER PROTEIN subfamily (Englbrecht *et al.*, 2004), seven oil palm proteins of the same clade and other TFs from tomato (Sl; accession Solyc07 g006880) and maize (Zm; accession GRMZM2G357688) were used. Gene expression levels in different oil palm tissues (leaf, female inflorescence, pollen, root, shoot apex, embryogenic cell suspensions, endosperm and mesocarp) and at different developmental stages of the endosperm and the mesocarp (days after pollination, DAP), are represented by blue dots, whose size and depth of colour specify the number of reads per kilobase and million reads (RPKM).



TFs in this subnetwork: a member of the NF-YB subfamily of the NUCLEAR FACTOR Y (NF-Y) family (hereafter termed NF-YB-1; Figure 6c), which was connected to 13 guide genes (Figure 6a); and a member of the ZINC FINGER PROTEIN (ZFP) family (termed ZFP-1, Figure 6c), which was linked to 10 guide genes (Figure 6a). To determine common and specific partners of WRI1-1, NF-YB-1 and ZFP-1, we focused on genes involved in lipid and sugar metabolism (Figure 6a,b). It is worth noting that with the exception *KASIII-1*, all genes involved in lipid, glycolysis, starch and other sugar metabolic pathways found in the one-round subnetwork were already present in the two-round coexpression network, suggesting that the key players in these metabolic pathways were all uncovered with an $|R|$ threshold of 0.8. Most of these genes were linked to at least two of the three TFs (Figure 6b). WRI1-1 possessed six specific FAS partners and was connected to all FAS genes shared with NF-YB-1 and ZFP-1, suggesting that it plays a prominent role in the upregulation of FAS genes. However, this was not observed for glycolysis and starch metabolism, for which the three TFs had a similar number of partners.

To further examine the potential role played by NF-YB-1 and ZFP-1 in oil biosynthesis, we compared the tissue-specific expression of their transcripts with those of the three WRI1 paralogues (Dussert *et al.*, 2013) and other TFs of the same subfamilies (Figure 6c). For NF-YB-1, all members of the NF-YB subfamily were investigated. In contrast, for ZFP-1, the search was limited to members of the clade C1-1iAa (Englbrecht *et al.*, 2004), since ZFP-1 was assigned to this group by a preliminary phylogenetic analysis that included the 33 Arabidopsis members of the ZPF family with a single domain (subfamily C1-1i) (Figure S4). The closest Arabidopsis protein relatives of NF-YB-1 were AtNF-YB2 (At5 g47640) and AtNF-YB3 (At4 g14540). *NF-YB-1* was highly expressed in the mesocarp and the endosperm at the stages when FAS gene transcription peaked (Dussert *et al.*, 2013). Its transcription profile was similar to that of *WRI1-1* in the mesocarp and to that of *WRI1-2* in the endosperm. The amount of *NF-YB-1* transcript was also very high in the female inflorescence, whereas none of the three *WRI1* paralogues were expressed in this tissue. To confirm the possible role of NF-YB-1 in upregulation of the FAS gene and hence in lipid biosynthesis, the oil content of the different tissues analysed was measured using three independent biological samples (Figure 6c). A significant amount of oil was found in the female inflorescence (4.9%), the pollen (11.2%) and the embryogenic cell suspension (6.2%). Nile red staining coupled with fluorescence microscopy revealed the presence of highly fluorescent oil bodies in the female inflorescence tissue (Figure S5). Little or no *NF-YB-1* mRNA was detected in the pollen or in *in vitro* embryogenic cells. However, *WRI1-3* transcripts were strongly detected in the pollen and a gene coding for another TF of the NF-YB subfamily, which was within a branch that contained Arabidopsis LEC1 and LEC1-like (L1L), was significantly expressed in embryogenic cells. ZFP-1 was clearly grouped with AtZFP2 (At5g57520), SlZFP2 and ZmZFP2 (Figure 6c). As was also observed in the case of the *WRI1-1* and FAS genes, accumulation of *ZFP-1* transcript peaked at 120 days after pollination (DAP) in the oil palm mesocarp. *ZFP-1* transcripts were observed at very low levels in the female inflorescence and were not detected in other tissues. Transcripts coding for the other members of this clade did not accumulate in the different tissues analysed.

KASII plays a major role in determining the 16:0 level of palm oil

Principal component analysis (PCA) was used to investigate possible relationships between the FA composition of oil and the transcription levels of genes likely to control the FA composition of mesocarp oil, namely *FATA*, *FATB1*, *FATB2*, *KASII*, *SAD-1* and *SAD-2*. The first two PCs explained more than 50% of the total variance (Figure 7a). Factors with the highest loadings were *SAD-1*, *SAD-2*, *KASII*, *FATA,* 16:0 and 18:1 and, to a lesser extent, 18:0. As expected from correlation analyses (Figure 1c), 16:0 and 18:1 vectors were seen to be opposing. The only gene that clearly opposed 16:0 was *KASII*, which encodes the plastidial enzyme β-ketoacyl-ACP synthase II, suggesting that genes coding for acyl-ACP thioesterases (FATA, FATB1 and B2) and stearate desaturases (SAD-1 and -2) do not play a major role in the considerable variation found in the 16:0/18:1 ratio among BCs. When *FAP1* and *ACBP6,* both detected in Module 2*,* and *FAD2*, which encodes the enzyme that converts 18:1 into 18:2, were added as active variables, *KASII* remained the only gene opposing 16:0 (Figure S6) but *ACBP6* clearly opposed 18:0, suggesting that this cytosolic ACBP promotes unsaturated FAs among C18 FAs during TAG assembly. The PC1–PC2 score plot unambiguously separated low-, medium- and high-16:0 mesocarp samples (Figure 7b). The 21 BCs analysed were then grouped in three categories according to the level of 16:0 in their mesocarp oil: low (32–36%), medium (34–41%) and high (42–47%) 16:0 (Figure 1b). According to one-way ANOVA ($P = 0.0002$) and a *post hoc* Newman and Keuls test, amounts of *KASII* transcript were considerably higher in low-16:0 BCs than in medium- and high-16:0 BCs (Figure 7c). Heterozygous genotypes at the *KASII* locus were found in the three categories of BCs, suggesting that there is no association between the oil 16:0 level and the genotype at the *KASII* locus. However, the number of BCs studied was not high enough to test this result statistically. To provide further evidence that the KASII isoform did not influence the oil 16:0 level in our study, we measured ASE in heterozygous genotypes at the *KASII* locus and compared transcript levels for the two parental enzymes. Transcripts of the two alleles accumulated at the same level in

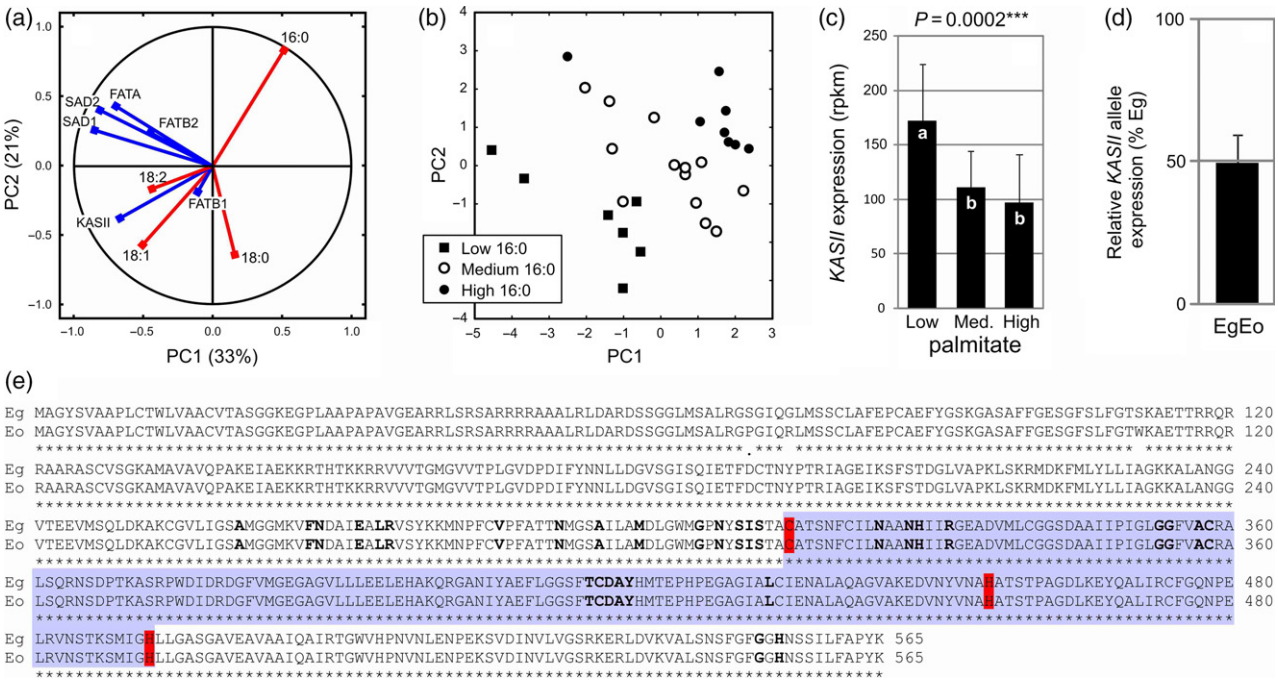


Figure 7. Influence of *KASII* transcript abundance during mesocarp development on oil 16:0 level.
(a), (b) Principal component analysis of fatty acid (FA) percentages at maturity and transcript abundance of six genes potentially involved in FA composition (*FATA*, *FATB1*, *FATB2*, *KASII*, *SAD-1* and *SAD-2*). (a) Correlations (factor loadings) of fatty acid percentages (red vectors) and transcript levels (blue vectors) with the first two principal components (PC). The proportion of variance explained by each PC is in brackets. (b) PC1–PC2 score plot of backcrossed hybrids (BCs) classified according to oil 16:0 percentage (low 16:0, 32–36%; medium 16:0, 37–41%; high 16:0, ≥42%).
(c) Mesocarp *KASII* transcript abundance in the three groups of BCs. Bars with the same letter at the top are not significantly different according to one-way ANOVA (*P*-value) and a *post hoc* Newman and Keuls test.
(d) Allele-specific *KASII* expression as measured by the proportion of *Elaeis guineensis* mRNA (%Eg) in heterozygous genotypes (EgEo) at the *KASII* locus.
(e) Alignment of *E. guineensis* (Eg) and *Elaeis oleifera* (Eo) KASII amino acid sequences. The blue box delimits the active site of the enzyme according to the NCBI's conserved domain database. Red boxes indicate the three residues of the catalytic triad. Residues in bold determine the dimer interface.

KASII heterozygous genotypes (Figure 7d). The AA sequences of Eg and Eo KASII were almost identical (Figure 7e). Only three AA differences were found (out of 565 AAs) at the N-terminal side, none of which were within the active site of the enzyme. The residues involved in dimerization and in the catalytic triad were also identical.

DISCUSSION

Investigating the metabolism of lipids through coexpression analysis

The regulatory system formed by WRI1 and its FAS target genes provided an excellent foundation upon which to develop our approach and evaluate its efficiency. By studying ASE, we were first able to ascertain that no 'genome clash' (Landry *et al.*, 2007) was affecting lipid metabolism in a way that would render *Elaeis* BCs unsuitable for coexpression analysis. Even though the two *Elaeis* species diverged from each other between 16 and 7 million years ago (Baker and Couvreur, 2013; Meerow *et al.*, 2015), they showed remarkable molecular conservation of the oil biosynthesis machinery. The regulatory system centred on WRI1 was also a very useful reference for network construction. Module 1 consistently contained *WRI1-1* and 23 FAS genes, thus reinforcing our confidence in the significance of the entire coexpression network. In the same way, these observations provide strong evidence that the two novel TFs identified in the FAS subnetwork play an important role in central carbon metabolism and FA biosynthesis.

Coexpression analysis enabled us to identify which lipid gene paralogues inferred from previous transcriptome analyses are likely to be the main contributors to oil biosynthesis in the oil palm mesocarp and to uncover multiple enzyme isoforms for several biosynthetic steps. We believe that the present study has allowed the delineation of an almost full set of plastidial genes/paralogues of key importance for oil biosynthesis in this crop plant, i.e. those that belong to Modules 1, 2 and 3. In contrast, key TAG assembly genes (*GPAT*, *LPAAT*, *DGAT* and *PDAT*, which code for glycerol-3-phosphate acyltransferase, 1-acylglycerol-3-phosphate acyltransferase, acyl-CoA:diacylglycerol acyltransferase and phospholipid:diacylglycerol acyltransferase enzymes, respectively) were all absent from the coexpression network. We first hypothesized that the developmental stages sampled were not appropriate for

coexpression analysis of TAG assembly, but the distribution characteristics (mean, minimum, maximum, standard deviation) of TAG assembly gene expression levels were actually similar to those of several Module 1 FAS genes (Table S4). No coexpression pairs formed among TAG assembly genes (Figure S7), even when the $|R|$ threshold was lowered to 0.5, suggesting that each of them has its own regulatory system. However, two Module 3 genes, *LACS9* and *GPDHc1*, which were coexpressed with Module 1 FAS genes, formed a significant bridge between the two compartments of oil biosynthesis. In plants, during *de novo* FA synthesis in the plastid the elongating acyl chain remains conjugated to ACP. The transfer of newly synthesized FA to the ER for TAG assembly first requires the hydrolysis of acyl-ACP by acyl-ACP thioesterases (FATA and FATB) to release free FAs, which are then activated by long-chain acyl-CoA synthetases (LACS). In the ER, enzymes of the Kennedy pathway use acyl-CoAs for the stepwise acylation of glycerol-3-phosphate (G3P). In Arabidopsis, the plastidial LACS9 and the ER-localized LACS1 and LACS4 generate the acyl-CoA pool used by TAG assembly enzymes (Zhao *et al.*, 2010; Jessen *et al.*, 2015). In the oil-accumulating avocado mesocarp, transcripts coding for the ortholog of LACS4 were shown to be the most abundant (Kilaru *et al.*, 2015). In a previous study, we observed that *LACS9* and *LACS4-1* were both highly transcribed in the oil palm mesocarp (Dussert *et al.*, 2013). Coexpression analysis now suggests that LACS9 plays the predominant role in this tissue. Genome-wide association studies allowed the identification of several candidate loci associated with kernel oil content in maize (Li *et al.*, 2013). *ZmWRI1a*, a gene coding for an ACP, and a gene with high homology to *AtLACS8/9*, were among the four genes that were further validated in a second maize germplasm group for their role in the determination of oil content. As mentioned above, G3P constitutes the backbone of TAGs. *GPDHc1* encodes a cytosolic NAD-dependent glycerol-3-phosphate dehydrogenase (Shen *et al.*, 2006), which is thought to supply G3P for the TAG synthesis process. By feeding glycerol to developing seeds *in planta* (Vigeolas and Geigenberger, 2004), or by overexpressing a yeast cytosolic GPDH using a seed-specific promoter (Vigeolas *et al.*, 2007), it is possible to obtain a significantly increased seed oil content in *Brassica*, suggesting that the supply of G3P is rate-limiting for TAG synthesis. In Arabidopsis, senescence-induced *LEC2* expression caused a three-fold increase in TAG levels in leaves (Kim *et al.*, 2015). LEAFY COTYLEDON2 (LEC2) is a master regulator of seed maturation and oil accumulation in Brassicaceae seeds (Baud *et al.*, 2007; Santos Mendoza *et al.*, 2008). Among genes that were upregulated in transgenic leaves (including *WRI1* and most FAS genes), *LACS9* and *GPDHc1* were among those that showed the highest fold change in expression. On the basis of these studies, it is tempting to hypothesize that TAG assembly in the oil palm mesocarp is driven by the flux towards the ER of FA and G3P generated by LACS9 and GPDHc1 respectively, without concerted transcriptional regulation of TAG assembly genes.

Systems analysis provides insights into central metabolism rewiring during oil biosynthesis

In contrast to FAS, which appears to be regulated as a whole (all enzymatic steps were represented in Module 1), only a few glycolytic steps were shown to be coregulated in the coexpression network. This suggests a different mode of control for linear metabolic pathways such as FAS, which is characterized by conserved stoichiometry of transcripts for the different enzymes, and central metabolism, in which pathways display branching, the flux towards a given branch depending on the regulation of specific steps. In addition to allosteric regulation and covalent modification of pre-existing enzymes, it is assumed that glycolytic fluxes are regulated by the concentration of proteins responsible for rate-limiting steps in sucrose cleavage, hexose transport and glycolytic activities (Plaxton, 1996; Angeles-Núñez and Tiessen, 2010; Geigenberger, 2011). The direct partners identified for FAS genes in our study included key players in these three categories, for example cwINV, GLT1 and PFKp. Interestingly, the few true glycolytic genes identified in the present network code for enzymes that catalyse irreversible reactions and are sites for allosteric regulation (HXK, PFKB, PFK and PK).

Network analysis also revealed transcriptional control exerted on carbon recapture pathways. The PPP may significantly contribute to the production of the NADPH required for FA synthesis (Baud and Lepiniec, 2010). By interconverting sugars that can re-enter glycolysis, the concerted activation of FAS and the non-oxidative phase of the PPP suggest the existence of an efficient mechanism for carbon recapture in the oil palm mesocarp. This mechanism is generally activated when cell demand for NADPH outstrips pentose requirements. Another example of carbon recapture is the transcriptional control exerted on the PDH bypass, which may contribute to carbon flux towards acetyl-CoA through the aerobic fermentation pathway (Mellema *et al.*, 2002). While the role of this pathway has been shown to be marginal in Arabidopsis seeds (Lin and Oliver, 2008), its quantitative contribution remains to be determined in fruit tissues, which may be subject to aerobic fermentation during maturation.

Starch synthesis and remobilization were substantially represented in the FAS module and adjacent modules, suggesting that starch metabolism plays a key role in oil synthesis in oil palm. Starch was found to accumulate concomitantly with oil synthesis in the oil palm mesocarp, but its content remained low and stable during development (Bourgis *et al.*, 2011). One may hypothesize that starch synthesis occurs when the supply of carbon exceeds the

capacity for lipid synthesis, while in the reverse situation starch is remobilized to fuel plastidial glycolysis (Borisjuk *et al.*, 2005; Geigenberger, 2011). Embryo-specific suppression of AGP-glucose pyrophosphorylase (AGPase), which catalyses the first committed step of starch biosynthesis, delayed oil accumulation during embryo development in rapeseed (Vigeolas *et al.*, 2004). Coordination between starch and lipid metabolism was recently evidenced in rapeseed, with bottom-up control of glycolysis and starch synthesis by lipogenic activity (Schwender *et al.*, 2015). High concentrations of the two glycolytic intermediates PEP and 3-PGA (resulting from low rates of FA synthesis) may shift carbon partitioning toward starch synthesis, since AGPase is allosterically affected by the 3-PGA/inorganic phosphate ratio (Crevillén *et al.*, 2003). However, this fine metabolic control depends on the prior establishment of the starch metabolic machinery. This prerequisite appears to be transcriptionally coordinated with FAS in the oil palm mesocarp.

Transcription of key players in sugar sensing was also quantitatively coordinated with that of FAS genes. Trehalose phosphate synthase is involved in the synthesis of trehalose 6-phosphate (T6P), a molecule thought to act as a signal of sugar availability, while *SNF4* codes for the activator subunit of SnRK1, the plant orthologue of the evolutionarily conserved SNF1/AMPK/SnRK1 protein kinase family that contributes to energy homeostasis in cells (Tsai and Gazzarrini, 2014). SnRK1 is activated in low-sugar conditions to inhibit growth and conserve energy, while T6P acts as a reporter of energy status and promotes growth and development in response to increasing sugar levels. Trehalose 6-phosphate inhibits SnRK1, thereby altering gene expression and promoting growth processes. Together with HXK, TPS and SnRK1 provide metabolite feedback regulation for sugar homeostasis (Nägele and Weckwerth, 2014). As seen in the case of starch metabolism, the production of sugar sensors for subsequent homeostasis during oil synthesis appears to be transcriptionally coordinated with the FA biosynthetic machinery.

Processes and transcription factors involved in FA biosynthesis

All the modules in the oil palm mesocarp coexpression network showed remarkable functional coherence. Among the processes that were abundantly represented in the network, plastid biogenesis appeared to be of primary importance for FA biosynthesis. Genes in this functional category were present not only in the FAS module (Module 1) but dominated in the module most tightly interconnected with the FAS module (Module 3). All the crucial processes of plastid division, development and functioning (Pogson and Albrecht, 2011; Jarvis and López-Juez, 2013) were represented by several key genes in these modules. The present study thus reveals the orchestrated transcription of genes involved in FA biosynthesis and plastid biogenesis in the oil palm mesocarp. Earlier experiments analysing top-down control showed that oil accumulation in oil palm is mostly driven by FA supply from plastids (Ramli *et al.*, 2009), i.e. a source effect (Ohlrogge and Jaworski, 1997), while in other species, including rapeseed, the demand exerted by TAG assembly enzymes controls the flux to oil (Tang *et al.*, 2012), i.e. a sink effect. In line with this observation, our findings suggest that biogenesis and maintenance of FA production sites is tightly controlled in the oil palm mesocarp during FA biosynthesis.

In connection with the possible role of plant hormones, coexpression analysis showed that auxin may be involved in FA biosynthesis and/or the main processes to which it is tightly linked in the network – glycolysis, starch metabolism and plastid biogenesis. Four genes involved in auxin transport (*ABCB19*, *APM1*, *APP1* and *PAS1*) were among the 53 direct positive partners of Module 1 FAS genes, and other auxin-related genes were detected in the network, including three IAA auxin-responsive proteins. ABCB19 is a member of the ATP-binding cassette (ABC) transporter family that is involved in auxin transport (Noh *et al.*, 2001). ABCB19 is required for normal auxin distribution in cotyledons, where oil accumulates during seed development in Arabidopsis (Lewis *et al.*, 2009). APM1 and APP1 are two plasma membrane aminopeptidases identified by their affinity with the auxin transport inhibitor NPA (Murphy *et al.*, 2002). The subcellular location of auxin transporters is altered in *apm1* loss-of-function mutants (Peer *et al.*, 2009). PAS1 is a chaperone involved in post-translational regulation of the ER fatty acid elongase complex required for the synthesis of very-long-chain FAs, which, in turn, are required for polar auxin transport (Roudier *et al.*, 2010). At this stage, establishing a model to unravel the links between auxin, FA biosynthesis and associated processes would be difficult. Nevertheless, numerous studies have suggested direct or indirect control of lipid metabolism by auxin. A link between auxin signalling and seed maturation has been substantiated by a significant number of studies in Arabidopsis (Santos Mendoza *et al.*, 2008). For instance, in Arabidopsis leaves, the ectopic expression of *LEC2* triggers TAG accumulation (Santos Mendoza *et al.*, 2005) and the activation of expression of not only *WRI1* (Baud *et al.*, 2007), through a regulatory cascade that remains to be determined (Marchive *et al.*, 2014), but also of auxin biosynthetic genes and auxin-responsive genes (Stone *et al.*, 2008). Interestingly, several auxin transporters have a transcript profile similar to those of the *WRI1* and FAS genes in the developing *Brassica* embryo (Deng *et al.*, 2015). A marked decrease in seed oil content and FAS gene transcription was observed in embryos of Transparent Testa16 RNA interference (RNAi) transgenic *Brassica* lines, accompanied by significant changes in the expression of several auxin-related genes (Deng *et al.*, 2012).

Other processes that showed gene enrichment in the network may also be associated with plastid biogenesis, auxin transport or both. For instance, Modules 3, 4 and 5 in the oil palm mesocarp coexpression network were characterized by their high numbers of cytoskeleton-related genes. Auxin transport and signalling depend on cytoskeleton organization (Li *et al.*, 2014). The cytoskeleton is also crucial for plastid biogenesis (Albrecht *et al.*, 2010). As a second example, Module 3, which was particularly enriched in genes involved in plastid biogenesis, contained two genes coding for key sterol biosynthetic enzymes (*CAS1* and *DWF1*). In Arabidopsis, *cas1* mutants showed severe defects in plastid biogenesis (Babiychuk *et al.*, 2008). Sterols are also indispensable for auxin transport, in particular for ABCB19 trafficking (Yang *et al.*, 2013). These examples underline not only the relevance of the processes revealed by coexpression analysis, but also the intricacy of the network they form.

WRI1 is known to activate late glycolysis genes in addition to FAS genes (Marchive *et al.*, 2014). Our data evidenced broader transcriptional orchestration, with early glycolysis and starch metabolism also coregulated with late glycolysis and FAS, suggesting that WRI1-1 does not work alone but in close cooperation with other transcriptional activators. In addition to WRI1-1, the FAS subnetwork contained two other TFs, termed NF-YB-1 and ZFP-1. Measurements of oil content and transcription levels of *NF-YB-1* and *ZFP-1* in different tissues and organs of oil palm corroborated a possible role in oil synthesis. The two closest Arabidopsis relatives of EgNF-YB-1 and EgZFP-1 were, respectively, AtNF-YB2 and AtZFP2. Although the precise roles of AtNF-YB2 and AtZFP2 remain to be determined (Cai and Lashbrook, 2008; Petroni *et al.*, 2012), publicly available microarray data suggest they are both highly transcribed in the early stages of seed development, at the onset of *AtWRI1* and FAS gene transcription (Winter *et al.*, 2007).

KASII is a major quantitative trait transcript for the saturated fatty acid content of palm oil

In plants, KASII is responsible for the elongation of 16:0-ACP to 18:0-ACP. Competition for 16:0 substrate thus exists between acyl-ACP thioesterases that are able to hydrolyse 16:0-ACP and KASII. The amount of 16:0 transferred to the ER for TAG assembly is the result of these competing activities. Acyl-ACP thioesterases are separated into two classes, termed FATA and FATB (Jones *et al.*, 1995). In most plants studied so far, FATA enzymes preferentially hydrolyse 18:1-ACP while 16:0-ACP is the preferential substrate of FATB enzymes (Salas and Ohlrogge, 2002). Accordingly, *FATB* transcription is usually high in tropical seeds that store 16:0-rich oil, such as those of coffee and cotton (Pirtle *et al.*, 1999; Joët *et al.*, 2009). In Arabidopsis seeds, overexpression of *FATB* enabled a three-fold increase in 16:0 content (Dörmann *et al.*, 2000). We identified three *FATB* paralogues and one *FATA* gene in the transcriptomes of the developing embryo, endosperm and mesocarp of oil palm (Dussert *et al.*, 2013). Based on comparisons of amounts of transcript between tissues, we previously hypothesized that high FATB2 activity controls the high level of 16:0 in the mesocarp. Our present findings rather suggest that low *KASII* transcription is the major contributory factor in accumulation of 16:0. Whatever the substrate preferences of oil palm FATA and FATB1 and B2 isoforms (Dussert *et al.*, 2013), levels of expression of genes encoding acyl-ACP thioesterases did not appear to influence the 16:0 level of mesocarp oil in BCs. This situation contrasts with the accumulation of medium-chain fatty acids in oil palm endosperm, which was observed to mainly rely on upregulation of the *FATB3* paralogue (Dussert *et al.*, 2013). In avocado, *SAD* upregulation is likely to be involved in determining the very high 18:1 level of mesocarp oil (Kilaru *et al.*, 2015). However, the present findings are consistent with the spectacular increase in 16:0 level in Arabidopsis seeds resulting from RNAi-mediated seed-specific *KASII* silencing (Pidkowich *et al.*, 2007). The identification of factors that control *KASII* transcription now appears to be of primary importance for a better understanding of this trait in oil palm.

The low levels of 18:0 in Eg, Eo and interspecific BCs indicated that almost the whole pool of 18:0-ACP generated by KASII was desaturated by Δ9 stearoyl-ACP desaturase (SAD) activity in the plastid. Coexpression analysis identified SAD-1 as the main isoform for 18:0-ACP desaturation in the oil palm mesocarp. *SAD-1* (Module 2) was coexpressed with two genes that received little attention in our previous studies: *FAP1* and *ACBP6*. In Arabidopsis, another FAP isoform, FAP3, was coexpressed with FAS genes (Mentzen *et al.*, 2008). FAP1 is a FA-binding protein which localizes in the plastid stroma and accumulates concomitantly with FAS synthesis in the developing Arabidopsis seed (Ngaki *et al.*, 2012). It preferentially interacts with saturated acyl-ACP but how it contributes to oil biosynthesis and FA composition is not yet known. The role of ACBPs in lipid biosynthesis and trafficking is better understood (Xiao and Chye, 2011; Block and Jouhet, 2015) . In particular, low-molecular-mass cytosolic ACBPs (lmACBPs), of which AtACPB6 is the only member in Arabidopsis, are thought to play a major role in oil biosynthesis in seeds (Engeseth *et al.*, 1996; Yurchenko and Weselake, 2011). Among possible mechanisms, lmACBPs may promote the exchange of 18:1-CoA between the acyl-CoA pool and the ER phosphatidylcholine (PC) pool through increased 1-acylglycerol-3-phosphocholine acyltransferase (LPCAT) activity (Yurchenko *et al.*, 2009). This process, also termed acyl editing, is crucial for enrichment of the acyl-CoA pool in polyunsaturated FAs (PUFAs; Bates *et al.*, 2009). In the present study, PUFA contents ranged

between 10 and 16% in BCs. Whether coregulation of Module 2 genes is indicative of a coordination of acyl-editing in the oil palm mesocarp deserves further research.

CONCLUSIONS

The release of the oil palm genome (Singh *et al.*, 2013) has made it possible to use the most advanced RNA-seq technologies and systems biology approaches, including gene coexpression network analysis and allele-specific expression analysis, to investigate lipid metabolism in the world's leading source of vegetable oil. Oil biosynthesis is investigated here for the first time in plants using gene coexpression analysis with a dataset specially designed for the purpose. The present strategy, which combined genotypic and developmental effects to maximize variation in gene expression levels and a mixed guide-gene and non-targeted strategy for network construction (Aoki *et al.*, 2007), proved to be very promising for future systems analyses of oil biosynthesis in plants. We took advantage of the 'guilt-by-association' paradigm (Saito *et al.*, 2008) and successfully identified enzymes, transcription factors and cellular processes involved in oil biosynthesis in oil palm. This should encourage plant scientists to use a similar approach in other oil crops.

EXPERIMENTAL PROCEDURES

Plant material

An interspecific F_1 hybrid palm LM10986D (female Eo H833D $\times$ male Eg LM3261D) was backcrossed to Eg LM2509D (used as the male parent) to generate the BC population used. H833D is a wild Eo from the Sona region (Panama) and LM3261D is an Eg type *dura* of Deli Socfin origin. LM2509D is an Eg type *dura* of Deli Dabou origin. Mesocarp samples were collected as previously described by Tranbarger *et al.* (2011) from BCs grown at Quinindé, Ecuador. Each BC was represented in the field by three to five individuals cloned by *in vitro* culture and planted in 2004. The legitimacy of the BC population used for this study was checked using 12 microsatellite loci as described in Montoya *et al.* (2013). Twenty-one BC genotypes were then selected for transcriptomic and biochemical analyses. Based on transcript accumulation profiles of lipid-related genes obtained previously (Tranbarger *et al.*, 2011), mesocarp samples were collected at three developmental stages, about 120, 140 and 160 DAP. After preliminary quantitative PCR (qPCR) analyses, 12 of the 63 samples collected were not retained for RNA-seq because they displayed no or very low lipid gene transcription. The method used for qPCR analysis is described in Morcillo *et al.* (2013). The set of primers allowing amplification of 18 lipid-related genes is given in Table S5. Eo mesocarp samples were collected at 124, 133, 143 and 152 DAP and the four Eg samples (100, 120, 140 and 160 DAP) were those used in Tranbarger *et al.* (2011). The Eo mesocarp samples were collected from palms having the same geographical origin as the BC female grandparent, cultivated at the INRAB Centre de Recherches Agricoles Plantes Pérennes (CRA-PP), Pobé, Benin. For the measurement of lipid content in other tissues, endosperm, pollen, female inflorescence, mesocarp and leaf samples were collected on Eg-type *dura* palms grown at CRA-PP. Roots and shoots were sampled from germinated seedlings. Cell suspensions were produced at IRD, Montpellier, France.

Lipid analysis and histology

Total lipids were extracted from 1-g samples of freeze-dried powder using a modified Folch method as described in Dussert *et al.* (2013). Fatty acid methyl esters (FAMEs) were then prepared according to the ISO-5509 standard, and analysed using gas chromatography to determine FA composition. Lipid analyses were performed in triplicate (from three biological repetitions). To visualize lipids, samples were fixed as previously described in Tranbarger *et al.* (2011), washed and stored in 1 $\times$ PBS, and then stained with Nile Red. Sections (100 μm) of female inflorescence tissues were cut with a vibratome before staining, and observed by epifluorescence microscopy. Pollen samples were analysed by confocal microscopy (Zeiss LSM 780, GaAsP detector), with laser excitation at 488 nm and emission at 490–560 nm (40$\times$).

RNA extraction and sequencing

Total RNA was extracted from 2 g of mesocarp tissue as previously described in Tranbarger *et al.* (2011). Fifty-nine libraries were constructed using the TruSeq Stranded mRNA Sample preparation kit (Illumina, http://www.illumina.com/), validated using the Agilent DNA 1000 chip kit, and quantified using SYBR-based quantitative PCR (Kapa, https://www.kapabiosystems.com/), then sequenced on a Illumina HiSeq 2000 (paired-end reads, 100 cycles) at the MGX platform (Montpellier Genomix, http://www.mgx.cnrs.fr/). A total of 1.41 billion reads were generated (with an average of 24 million reads per library). The entire dataset has been deposited at the European Nucleotide Archive (ENA) under the project number PRJEB11097.

RNA-seq data processing

Low-quality reads were removed using Cutadapt. Trimmed reads were mapped using the BWA-MEM package with default parameters (Li, 2013). Samtools (Li *et al.*, 2009) was used to count mapped reads and the number of reads per kilobase and million reads (rpkm) were then calculated. To assess whether reads from Eo alleles can be mapped without bias when the Eg genome is used as the reference, a set of 44 genes involved in the metabolism of lipids in oil palm (Dussert *et al.*, 2013) was chosen and their orthologues were searched for in Eo. Reads from mesocarp samples of Eg, Eo and BCs were then mapped on the two sets of parental species sequences and read counts were compared. Expression levels of the 44 lipid genes were identical, or almost identical, to the two sets of orthologous reference genes (Figure S8A–C). This may be explained by the very high orthologue nucleotide identity between the two species (Figure S8D). Owing to the low genetic divergence between parental species, reads of each of our 59 libraries were therefore mapped to a single Eg reference, which consisted of gene models generated by MyOPGP (EG5.Genes.V2; http://genomsawit.mpob.gov.my; 30 752 coding sequences; January 2014), in which manually curated lipid genes (Dussert *et al.*, 2013) were used to replace corresponding predicted coding sequences when the curated lipid genes were of higher quality. Genes represented by less than 250 rpkm across the 59 libraries were not retained for coexpression analysis. Network construction was therefore performed using the expression levels of 16, 836 genes. In addition, seven RNA-seq datasets generated by the Malaysian Oil Palm Genome Programme (MyOPGP) for various Eg tissues were downloaded from NCBI databases: leaf (SRX278048), 20-cm-long female flower (SRX278053), pollen (SRX278051), root (SRX278062), shoot apex (SRX278055) and

endosperm at 10 and 15 weeks after anthesis (SRX278021 and SRX278018, respectively). RNA-seq data (100-nucleotide Illumina reads) from embryonic cell suspensions were kindly provided by TB. For these eight additional RNA-seq datasets, the Eg coding sequence reference used for mapping was downloaded from the NCBI website (GCF_000442705.1_EG5_rna.fna; January 2015).

Allele-specific expression

For each BC library, aligned sequences were analysed for SNP discovery with the GATK toolkit (http://www.broadinstitute.org/gatk) using the UnifiedGenotyper module with default parameters to obtain a list of SNPs and allelic data, and the Depth-Of-Coverage module to obtain information on depth coverage. The SNPs detected using GATK were then compared with the corresponding nucleotides in both parental genomes using SNiPloid (Peralta *et al.*, 2013), a dedicated web-based tool (http://sniplay.cirad.fr/cgi-bin/sniploid.cgi). SNiPloid allows the user to set the minimum depth coverage for a sequence position to be taken into consideration. Minimum depth coverages of 16 and 4 were required for BCs and Eo, respectively.

Promoter analyses

The oil palm Eg and Eo genome sequences generated by MyOPGP (Singh *et al.*, 2013) were downloaded from the MPOB website (http://genomsawit.mpob.gov.my; EG5-linked assembly for Eg and O7-build assembly for Eo). MatInspector in the Genomatix software suite (http://www.genomatix.de) was used to search for AW boxes in promoter sequences (1500 bp).

Coexpression analysis, network construction, module and QTT detection

Our strategy to build a coexpression network for the oil palm mesocarp combined both the guide-gene approach and the non-targeted approach, as defined by Aoki *et al.* (2007), since the network was constructed in the vicinity of a set of guide genes but their partners may belong to the whole oil palm genome. A previous transcriptome analysis (Dussert *et al.*, 2013) provided a list of genes potentially involved in oil biosynthesis (Figure S7). We classified these genes into four major functional categories: *de novo* FA synthesis (FAS), plastidial FA elongation, desaturation and export (16:0/18:1 ratio), TAG assembly in the ER (TAG) and enrichment of TAG in polyunsaturated FA (PUFA). A set of 35 guide genes was chosen among them to represent the four lipid categories in a balanced manner. During the first round of network construction, all connections between guide genes, between guide genes and their partners (the genes to which they are linked, thereafter termed P1) and between P1 were computed (Figure S7). During the second round of network construction, partners of P1 were identified (P2) and then links between P2 were generated. Logarithmic transformation of read counts was applied in all coexpression analyses (Figure S9). Linear regressions were computed using in-house scripts. Preliminary trials showed that two rounds of coexpression analysis using an $|R|$ threshold of 0.8 provided a tractable number of nodes (489) and edges (4902) for network construction (Figure S7). The *P*-values of edges generated with these parameters were extremely low (Figure 3f). After Bonferonni–Holm or Benjamini–Hochberg false discovery rate correction, the *P*-values were still very low ($<9.2 \times 10^{-12}$ and $<2.9 \times 10^{-14}$, respectively). Preliminary trials also showed that an $|R|$ threshold of 0.7 was most suitable for the discovery of regulatory links (Figure S7). A one-round network with an $|R|$ threshold of 0.7 also provided a tractable network. These parameters were consequently used to identify novel potential regulators of the FAS subnetwork. Gene interactions were visualized using the open source software Cytoscape (Shannon *et al.*, 2003) and an organic layout. The Markov cluster (MCL) algorithm (inflation of 2; Enright *et al.*, 2002) was used for module detection. Each small (number of nodes $\leq$15) MCL cluster was merged with a larger cluster that was topologically adjacent or that surrounded the small cluster (Table S6). The result of this merging step was 10 modules, referred to as Modules 1 to 10. Genes of the oil palm coexpression network were manually annotated and 86% of them were assigned to functional groups and subgroups (Table S3). When possible (87% of nodes), a gene name was given based on the best Arabidopsis match. In addition, we used Blast2GO software to retrieve Gene Ontology terms. One-way ANOVA and PCA)were performed using Statistica software (Statsoft Inc., http://www.statsoft.com/). Relationships between oil FA composition and the transcription levels of genes likely to control FA composition were investigated in BC samples showing high FAS gene transcription. The criterion used for sample selection was based on *KAR-1* (which encodes a ketoacyl-ACP reductase) transcript accumulation, but very similar results were obtained with most FAS genes. For each BC, the developmental stage with the highest *KAR-1* rpkm value (termed KAR_{max}) was identified. Then, within each BC, all samples with *KAR-1* expression level $\geq$ 0.65 times KAR_{max} were used for PCA and ANOVA.

Phylogenetic analyses

Phylogenetic analyses were performed on the Phylogeny.fr platform (http://ww.phylogeny.fr). Briefly, amino acid sequences were aligned with ClustalW (v.2.0.3). Phylogenetic trees were constructed using the maximum likelihood method (PhyML, v.3.0). Node robustness was assessed by the approximate likelihood-ratio test with a default substitution model. Graphical representation of phylogenetic trees was performed using TreeDyn (v.198.3).

ACKNOWLEDGEMENTS

We are grateful to PalmElit SAS, the Association Nationale Recherche Technologie (CIFRE grant 2012/0742) and the Région Languedoc-Roussillon for their financial support. Our special thanks go to Claude Louise (PalmElit, Ecuador) and Roberto Poveda (DANEC, Ecuador) for their technical and logistical support during the collection of the material used in this study. We thank Michel Cazemajor (PalmElit, Pobé, Benin) and Leifi Nodichao (INRAB-CRAPP, Pobé, Benin) for collection of pollen, shoots, roots and female inflorescences. We also thank Christine Tranchant and Gauthier Sarah for advice on bioinformatics.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Amino acid alignment of *Elaeis guineensis* (Eg) and *Elaeis oleifera* (Eo) WRI1-1 proteins.

Figure S2. Three different views of the oil palm mesocarp gene coexpression network.

Figure S3. Module-specific distribution of the 489 nodes of the oil palm mesocarp coexpression network (two-round guide-gene approach, $|R|$ threshold of 0.8) as a function of their positive and negative intra-module and between-module degree.

Figure S4. Phylogenetic analysis of EgZPP-1.

Figure S5. Nile Red staining of oil bodies in the female inflorescence and pollen of oil palm.

Figure S6. Principal component analysis of fatty acid (FA) percentages at maturity and transcript abundance of nine genes potentially involved in FA composition (*ACBP6*, *FAD2*, *FAP1*, *FATA*, *FATB1*, *FATB2*, *KASII*, *SAD-1* and *SAD-2*).

Figure S7. A step-by-step approach to constructing the gene coexpression network of the oil palm mesocarp.

Figure S8. Impact of the reference genome on gene expression measurement.

Figure S9. Effect of logarithmic transformation of read counts for coexpression analysis illustrated with *EAR-1* and *KAR-1*.

Table S1. Fatty acid composition of the oil extracted from the mature mesocarp of the backcrossed hybrids studied: minimum, mean and maximum percentages.

Table S2. Correlations between lipid traits measured in the backcrossed hybrid population studied.

Table S3. The 489 genes of the two-round coexpression network.

Table S4. Distribution characteristics of lipid-related genes and the 489 genes of the coexpression network.

Table S5. List of primers used for quantitative PCR analyses.

Table S6. Cluster merging: each small ($n \leq 15$) cluster inferred from the Markov cluster algorithm was merged with a topologically adjacent cluster of high size.

REFERENCES

Albrecht, V., Simková, K., Carrie, C., Delannoy, E., Giraud, E., Whelan, J., Small, I.D., Apel, K., Badger, M.R. and Pogson, B.J. (2010) The cytoskeleton and the peroxisomal-targeted snowy cotyledon3 protein are required for chloroplast development in Arabidopsis. *Plant Cell*, **22**, 3423–3438.

Alexandratos, N. and Bruinsma, J. (2012) *World Agriculture Towards 2030/2050: The 2012 Revision*. ESA Working paper No. 12-03. Rome: FAO.

Angeles-Núñez, J.G. and Tiessen, A. (2010) Arabidopsis sucrose synthase 2 and 3 modulate metabolic homeostasis and direct carbon towards starch synthesis in developing seeds. *Planta*, **232**, 701–718.

Aoki, K., Ogata, Y. and Shibata, D. (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol.*, **48**, 381–390.

Babiychuk, E., Bouvier-Navé, P., Compagnon, V., Suzuki, M., Muranaka, T., Van Montagu, M., Kushnir, S. and Schaller, H. (2008) Allelic mutant series reveal distinct functions for Arabidopsis cycloartenol synthase 1 in cell viability and plastid biogenesis. *Proc. Natl Acad. Sci. USA*, **105**, 3163–3168.

Baker, W.J. and Couvreur, T.L.P. (2013) Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical Biogeography. *J. Biogeogr.*, **40**, 274–285.

Barcelos, E., Almeida Rios, S., Cunha, R.N.V., Lopes, R., Motoike, S.Y., Babiychuk, E., Skirycz, A. and Kushnir, S. (2015) Oil palm natural diversity and the potential for yield improvement. *Front. Plant Sci.*, **6**, 190.

Bassel, G.W., Lan, H., Glaab, E., Gibbs, D.J., Gerjets, T., Krasnogor, N., Bonner, A.J., Holdsworth, M.J. and Provart, N.J. (2011) Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. *Proc. Natl Acad. Sci. USA*, **108**, 9709–9714.

Bassel, G.W., Gaudinier, A., Brady, S.M., Hennig, L., Rhee, S.Y. and De Smet, I. (2012) Systems analysis of plant functional, transcriptional, physical interaction, and metabolic networks. *Plant Cell*, **24**, 3859–3875.

Bates, P.D., Durrett, T.P., Ohlrogge, J.B. and Pollard, M. (2009) Analysis of acyl fluxes through multiple pathways of triacylglycerol synthesis in developing soybean embryos. *Plant Physiol.*, **150**, 55–72.

Baud, S. and Lepiniec, L. (2010) Physiological and developmental regulation of seed oil production. *Prog. Lipid Res.*, **49**, 235–249.

Baud, S., Mendoza, M.S., To, A., Harscoët, E., Lepiniec, L. and Dubreucq, B. (2007) WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. *Plant J.*, **50**, 825–838.

Block, M.A. and Jouhet, J. (2015) Lipid trafficking at endoplasmic reticulum-chloroplast membrane contact sites. *Curr. Opin. Cell Biol.*, **35**, 21–29.

Borisjuk, L., Nguyen, T.H., Neuberger, T. *et al.* (2005) Gradients of lipid storage, photosynthesis and plastid differentiation in developing soybean seeds. *New Phytol.*, **167**, 761–776.

Bourgis, F., Kilaru, A., Cao, X., Ngando-Ebongue, G.F., Drira, N., Ohlrogge, J.B. and Arondel, V. (2011) Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proc. Natl Acad. Sci. USA*, **108**, 12527–12532.

Cai, S. and Lashbrook, C.C. (2008) Stamen abscission zone transcriptome profiling reveals new candidates for abscission control: enhanced retention of floral organs in transgenic plants overexpressing Arabidopsis ZINC FINGER PROTEIN2. *Plant Physiol.*, **146**, 1305–1321.

Combes, M.C., Hueber, Y., Dereeper, A., Rialle, S., Herrera, J.C. and Lashermes, P. (2015) Regulatory divergence between parental alleles determines gene expression patterns in hybrids. *Genome Biol Evol.*, **7**, 1110–1121.

Crevillén, P., Ballicora, M.A., Mérida, A., Preiss, J. and Romero, J.M. (2003) The different large subunit isoforms of Arabidopsis thaliana ADP-glucose pyrophosphorylase confer distinct kinetic and regulatory properties to the heterotetrameric enzyme. *J. Biol. Chem.*, **278**, 28508–28515.

Deng, W., Chen, G., Peng, F., Truksa, M., Snyder, C.L. and Weselake, R.J. (2012) Transparent testa16 plays multiple roles in plant development and is involved in lipid synthesis and embryo development in canola. *Plant Physiol.*, **160**, 978–989.

Deng, W., Yan, F., Zhang, X., Tang, Y. and Yuan, Y. (2015) Transcriptional profiling of canola developing embryo and identification of the important roles of BnDof5.6 in embryo development and fatty acids synthesis. *Plant Cell Physiol.*, **56**, 1624–1640.

Dörmann, P., Voelker, T.A. and Ohlrogge, J.B. (2000) Accumulation of palmitate in Arabidopsis mediated by the acyl-acyl carrier protein thioesterase FATB1. *Plant Physiol.*, **123**, 637–644.

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

Dyer, J.M., Stymne, S., Green, A.G. and Carlsson, A.S. (2008) High-value oils from plants. *Plant J.*, **54**, 640–655.

Engeseth, N.J., Pacovsky, R.S., Newman, T. and Ohlrogge, J.B. (1996) Characterization of an acyl-CoA binding protein from *Arabidopsis thaliana*. *Arch. Biochem. Biophys.*, **331**, 55–62.

Englbrecht, C.C., Schoof, H. and Böhm, S. (2004) Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome. *BMC Genom.*, **5**, 39.

Enright, A.J., Van Dongen, S. and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.*, **30**, 1575–1584.

Filteau, M., Pavey, S.A., St-Cyr, J. and Bernatchez, L. (2013) Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish. *Mol. Biol. Evol.*, **30**, 1384–1396.

Focks, N. and Benning, C. (1998) wrinkled1: a novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiol.*, **118**, 91–101.

Gargouri, M., Park, J.J., Holguin, F.O., Kim, M.J., Wang, H., Deshpande, R.R., Shachar-Hill, Y., Hicks, L.M. and Gang, D.R. (2015) Identification of regulatory network hubs that control lipid metabolism in *Chlamydomonas reinhardtii*. *J. Exp. Bot.*, **66**, 4551–4566.

Gaur, U., Li, K., Mei, S. and Liu, G. (2013) Research progress in allele-specific expression and its regulatory mechanisms. *J Appl Genet.*, **54**, 271–283.

Geigenberger, P. (2011) Regulation of starch biosynthesis in response to a fluctuating environment. *Plant Physiol.*, **155**, 1566–1577.

Jarvis, P. and López-Juez, E. (2013) Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Rev. Mol. Cell Biol.*, **14**, 787–802.

Jessen, D., Roth, C., Wiermer, M. and Fulda, M. (2015) Two activities of long-chain acyl-CoA synthetase are involved in lipid trafficking between the endoplasmic reticulum and the plastid in Arabidopsis. *Plant Physiol.*, **167**, 351–366.

Joët, T., Laffargue, A., Salmona, J., Doulbeau, S., Descroix, F., Bertrand, B., de Kochko, A. and Dussert, S. (2009) Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. *New Phytol.*, **182**, 146–162.

Joët, T., Salmona, J., Laffargue, A., Descroix, D. and Dussert, S. (2010) Use of the growing environment as a source of variation to identify the quantitative trait transcripts and modules of co-expressed genes that determine chlorogenic acid accumulation. *Plant, Cell Environ.*, **33**, 1220–1233.

Jones, A., Davies, H.M. and Voelker, T.A. (1995) Palmitoyl-Acyl Carrier Protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. *Plant Cell*, **7**, 359–371.

Kilaru, A., Cao, X., Dabbs, P.B. *et al.* (2015) Oil biosynthesis in a basal angiosperm: transcriptome analysis of Persea Americana mesocarp. *BMC Plant Biol.*, **15**, 203.

Kim, H.U., Lee, K.R., Jung, S.J., Shin, H.A., Go, Y.S., Suh, M.C. and Kim, J.B. (2015) Senescence-inducible LEC2 enhances triacylglycerol accumulation in leaves without negatively affecting plant growth. *Plant Biotechnol. J.*, **13**, 1346–1359.

Kirst, M., Myburg, A.A., De León, J.P., Kirst, M.E., Scott, J. and Sederoff, R. (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. *Plant Physiol.*, **135**, 2368–2378.

Kwon, K., Choi, D., Hyun, J.K., Jung, H.S., Baek, K. and Park, C. (2013) Novel glyoxalases from Arabidopsis thaliana. *FEBS J.*, **280**, 3328–3339.

Landry, C.R., Hartl, D.L. and Ranz, J.M. (2007) Genome clashes in hybrids: insights from gene expression. *Heredity*, **99**, 483–493.

Lewis, D.R., Wu, G., Ljung, K. and Spalding, E.P. (2009) Auxin transport into cotyledons and cotyledon growth depend similarly on the ABCB19 Multidrug Resistance-like transporter. *Plant J.*, **60**, 91–101.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, Preprint at arXiv:1303.3997v2 [q-bio.GN].

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) Subgroup 1000 genome project data processing: the sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li, H., Peng, Z., Yang, X. ***et al.*** (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.*, **45**, 43–50.

Li, G., Liang, W., Zhang, X., Ren, H., Hu, J., Bennett, M.J. and Zhang, D. (2014) Rice actin-binding protein RMD is a key link in the auxin-actin regulatory loop that controls cell growth. *Proc. Natl Acad. Sci. USA*, **111**, 10377–10382.

Lin, M. and Oliver, D.J. (2008) The role of acetyl-coenzyme a synthetase in Arabidopsis. *Plant Physiol.*, **147**, 1822–1829.

Ma, W., Kong, Q., Arondel, V., Kilaru, A., Bates, P.D., Thrower, N.A., Benning, C. and Ohlrogge, J.B. (2013) Wrinkled1, a ubiquitous regulator in oil accumulating tissues from Arabidopsis embryos to oil palm mesocarp. *PLoS ONE*, **26**, e68887.

Maeo, K., Tokuda, T., Ayame, A., Mitsui, N., Kawai, T., Tsukagoshi, H., Ishiguro, S. and Nakamura, K. (2009) An AP2-type transcription factor, WRINKLED1, of Arabidopsis thaliana binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. *Plant J.*, **60**, 476–487.

Marchive, C., Nikovics, K., To, A., Lepiniec, L. and Baud, S. (2014) Transcriptional regulation of fatty acid production in higher plants: molecular bases and biotechnological outcomes. *Eur. J. Lipid Sci. Technol.*, **116**, 1332–1343.

Meerow, A.W., Noblick, L., Salas-Leiva, D.E., Sanchez, V., Francisco-Ortega, J., Jestrow, B. and Nakamura, K. (2015) Phylogeny and historical biogeography of the cocosoid palms (Arecaceae, Arecoideae, Cocoseae) inferred from sequences of six WRKY gene family loci. *Cladistics*, **31**, 509–534.

Mellema, S., Eichenberger, W., Rawyler, A., Suter, M., Tadege, M. and Kuhlemeier, C. (2002) The ethanolic fermentation pathway supports respiration and lipid biosynthesis in tobacco pollen. *Plant J.*, **30**, 329–336.

Mentzen, W.I., Peng, J., Ransom, N., Nikolau, B.J. and Wurtele, E.S. (2008) Articulation of three core metabolic processes in Arabidopsis: fatty acid biosynthesis, leucine catabolism and starch metabolism. *BMC Plant Biol.*, **8**, 76.

Montoya, C., Lopes, R., Flori, A. ***et al.*** (2013) Quantitative trait loci (QTLs) analysis of palm oil fatty acid composition in an interspecific pseudo-backcross from *Elaeis oleifera* (HBK) Cortes and oil palm (*Elaeis guineensis* Jacq.). *Tree Genet. Genomes*, **9**, 1207–1225.

Morcillo, F., Cros, D., Billotte, N. ***et al.*** (2013) Improving world palm oil production: identification and mapping of the lipase gene causing oil deterioration. *Nat. Commun.*, **4**, 2160.

Murphy, A.S., Hoogner, K.R., Peer, W.A. and Taiz, L. (2002) Identification, purification, and molecular cloning of N-1-naphthylphthalmic acid-binding plasma membrane-associated aminopeptidases from Arabidopsis. *Plant Physiol.*, **128**, 935–950.

Nägele, T. and Weckwerth, W. (2014) Mathematical modeling reveals that metabolic feedback regulation of SnRK1 and hexokinase is sufficient to control sugar homeostasis from energy depletion to full recovery. *Front. Plant Sci.*, **5**, 365.

Ngaki, M.N., Louie, G.V., Philippe, R.N., Manning, G., Pojer, F., Bowman, M.E., Li, L., Larsen, E., Wurtele, E.S. and Noel, J.P. (2012) Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature*, **485**, 530–533.

Noh, B., Murphy, A.S. and Spalding, E.P. (2001) Multidrug resistance-like genes of Arabidopsis required for auxin transport and auxin-mediated development. *Plant Cell*, **13**, 2441–2454.

Ohlrogge, J.B. and Jaworski, J.G. (1997) Regulation of fatty acid synthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **48**, 109–136.

Okazaki, Y., Shimojima, M., Sawada, Y. ***et al.*** (2009) chloroplastic UDP-glucose pyrophosphorylase from Arabidopsis is the committed enzyme for the first step of sulfolipid biosynthesis. *Plant Cell*, **21**, 892–909.

Palumbo, M.C., Zenoni, S., Fasoli, M., Massonnet, M., Farina, L., Castiglione, F., Pezzotti, M. and Paci, P. (2014) Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major transcriptome reprogramming during grapevine development. *Plant Cell*, **26**, 4617–4635.

Peer, W.A., Hosein, F.N., Bandyopadhyay, A. ***et al.*** (2009) Mutation of the membrane-associated M1 protease APM1 results in distinct embryonic and seedling developmental defects in Arabidopsis. *Plant Cell*, **21**, 1693–1721.

Peralta, M., Combes, M.C., Cenci, A., Lashermes, P. and Dereeper, A. (2013) SNiPloid: a utility to exploit high-throughput SNP data derived from RNA-Seq in allopolyploid species. *Int. J. Plant Genomics*, doi: 10.1155/2013/890123.

Petroni, K., Kumimoto, R.W., Gnesutta, N., Calvenzani, V., Fornari, M., Tonelli, C., Holt, B.F. and Mantovani, R. (2012) The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. *Plant Cell*, **24**, 4777–4792.

Pfeifer, M., Kugler, K.G., Sandve, S.R., Zhan, B., Rudi, H. and Hvidsten, T.R. (2014) International Wheat Genome Sequencing Consortium, Mayer KF, Olsen OA. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science*, **345**, 1250091.

Pidkowich, M.S., Nguyen, H.T., Heilmann, I., Ischebeck, T. and Shanklin, J. (2007) Modulating seed beta-ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil. *Proc. Natl Acad. Sci. USA*, **104**, 4742–4747.

Pirtle, R.M., Yoder, D.W., Huynh, T.T., Nampaisansuk, M., Pirtle, I.L. and Chapman, K.D. (1999) Characterization of a palmitoyl-acyl carrier protein thioesterase (FATb1) in cotton. *Plant Cell Physiol.*, **40**, 155–163.

Plaxton, W.C. (1996) The organization and regulation of plant glycolysis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **47**, 185–214.

Pogson, B.J. and Albrecht, V. (2011) Genetic dissection of chloroplast biogenesis and development: an overview. *Plant Physiol.*, **155**, 1545–1551.

Ramli, U.S., Salas, J.J., Quant, P.A. and Harwood, J.L. (2009) Use of metabolic control analysis to give quantitative information on control of lipid biosynthesis in the important oil crop, *Elaeis guineensis* (oil palm). *New Phytol.*, **184**, 330–339.

Roudier, F., Gissot, L., Beaudoin, F. ***et al.*** (2010) Very-long-chain fatty acids are involved in polar auxin transport and developmental patterning in Arabidopsis. *Plant Cell*, **22**, 364–375.

Saito, K., Hirai, M.Y. and Yonekura-Sakakibara, K. (2008) Decoding genes with coexpression networks and metabolomics: 'majority report by precogs'. *Trends Plant Sci.*, **13**, 36–43.

Salas, J.J. and Ohlrogge, J.B. (2002) Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. *Arch. Biochem. Biophys.*, **403**, 25–34.

Santos Mendoza, M., Dubreucq, B., Miquel, M., Caboche, M. and Lepiniec, L. (2005) LEAFY COTYLEDON 2 activation is sufficient to trigger the

accumulation of oil and seed specific mRNAs in Arabidopsis leaves. *FEBS Lett.*, **579**, 4666–4670.

Santos Mendoza, M., Dubreucq, B., Baud, S., Parcy, F., Caboche, M. and Lepiniec, L. (2008) Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. *Plant J.*, **54**, 608–620.

Schwender, J., Hebbelmann, I., Heinzel, N. *et al.* (2015) Quantitative multilevel analysis of central metabolism in developing oilseeds of oilseed rape during in vitro culture. *Plant Physiol.*, **168**, 828–848.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schiwkowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

Shen, W., Wei, Y., Dauk, M., Tan, Y., Taylor, D.C., Selvaraj, G. and Zou, J. (2006) Involvement of a glycerol-3-phosphate dehydrogenase in modulating the NADH/NAD+ ratio provides evidence of a mitochondrial glycerol-3-phosphate shuttle in Arabidopsis. *Plant Cell*, **18**, 422–441.

Singh, R., Ong-Abdullah, M., Low, E.T.L. *et al.* (2013) Oil palm genome sequence reveals divergence of interfertile species in Old and New Worlds. *Nature*, **500**, 335–339.

Stone, S.L., Braybrook, S.A., Paula, S.L., Kwong, L.W., Meuser, J., Pelletier, J., Hsieh, T.F., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (2008) Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc. Natl Acad. Sci. USA*, **105**, 3151–3156.

Szymanski, J., Brotman, Y., Willmitzer, L. and Cuadros-Inostroza, Á. (2014) Linking gene expression and membrane lipid composition of Arabidopsis. *Plant Cell*, **26**, 915–928.

Tang, M., Guschina, I.A., O'Hara, P., Slabas, A.R., Quant, P.A., Fawcett, T. and Harwood, J.L. (2012) Metabolic control analysis of developing oilseed rape (*Brassica napus* cv Westar) embryos shows that lipid assembly exerts significant control over oil accumulation. *New Phytol.*, **196**, 414–426.

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

Tsai, A.Y. and Gazzarrini, S. (2014) Trehalose-6-phosphate and SnRK1 kinases in plant development and signaling: the emerging picture. *Front. Plant Sci.*, **5**, 119.

Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S. and Provart, N.J. (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant, Cell Environ.*, **32**, 1633–1651.

Vallabhaneni, R. and Wurtzel, E.T. (2009) Timing and biosynthetic potential for carotenoid accumulation in genetically diverse germplasm of maize. *Plant Physiol.*, **150**, 562–572.

Vigeolas, H. and Geigenberger, P. (2004) Increased levels of glycerol-3-phosphate lead to a stimulation of flux into triacylglycerol synthesis after supplying glycerol to developing seeds of *Brassica napus* L. *in planta*. *Planta*, **219**, 827–835.

Vigeolas, H., Möhlmann, T., Martini, N., Neuhaus, H.E. and Geigenberger, P. (2004) Embryo-specific reduction of ADP-Glc pyrophosphorylase leads to an inhibition of starch synthesis and a delay in oil accumulation in developing seeds of oilseed rape. *Plant Physiol.*, **136**, 2676–2686.

Vigeolas, H., Waldeck, P., Zank, T. and Geigenberger, P. (2007) Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. *Plant Biotechnol. J.*, **5**, 431–441.

Wei, H.R., Persson, S., Mehta, T., Srinivasasainagendra, V., Chen, L., Page, G.P., Somerville, C. and Loraine, A. (2006) Transcriptional coordination of the metabolic network in Arabidopsis. *Plant Physiol.*, **142**, 762–774.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. and Provart, N.J. (2007) An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS ONE*, **2**, e718.

Xiao, S. and Chye, M.L. (2011) New roles for acyl-CoA-binding proteins (ACBPs) in plant development, stress responses and lipid metabolism. *Prog. Lipid Res.*, **50**, 141–151.

Yang, H., Richter, G.L., Wang, X., Młodzińska, E., Carraro, N., Ma, G., Jenness, M., Chao, D.Y., Peer, W.A. and Murphy, A.S. (2013) Sterols and sphingolipids differentially function in trafficking of the Arabidopsis ABCB19 auxin transporter. *Plant J.*, **74**, 37–47.

Yurchenko, O.P. and Weselake, R.J. (2011) Involvement of low molecular mass soluble acyl-CoA-binding protein in seed oil biosynthesis. *N. Biotechnol.*, **28**, 97–109.

Yurchenko, O.P., Nykiforuk, C.L., Moloney, M.M., Ståhl, U., Banaś, A., Stymne, S. and Weselake, R.J. (2009) A 10-kDa acyl-CoA-binding protein (ACBP) from *Brassica napus* enhances acyl exchange between acyl-CoA and phosphatidylcholine. *Plant Biotechnol. J.*, **7**, 602–610.

Zhao, L., Katavic, V., Li, F., Haughn, G.W. and Kunst, L. (2010) Insertional mutant analysis reveals that long-chain acyl-CoA synthetase 1 (LACS1), but not LACS8, functionally overlaps with LACS9 in Arabidopsis seed oil biosynthesis. *Plant J.*, **64**, 1048–1058.