Running Head: Transcriptional regulation in a lipid rich fleshy fruit

**Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening and functional specialization in lipid and carotenoid metabolism**¹

**Corresponding Authors:** Fabienne Morcillo, Timothy J. Tranbarger
Institut de Recherche pour le Développement
IRD Centre de Montpellier
IRD/CIRAD Palm Development Group, DIADE
911 avenue agropolis BP 64501
34394 Montpellier cedex 5, France
Phone: (33) 0467416235
Fax: (33) 0467416181
E-mail: fabienne.morcillo@ird.fr, timothy.tranbarger@ird.fr

Timothy J. Tranbarger*, Stéphane Dussert², Thierry Joët, Xavier Argout, Marilyne Summo, Antony Champion, David Cros, Alphonse Omore, Bruno Nouy, and Fabienne Morcillo*

IRD (Institut de Recherche pour le Développement), UMR DIADE, 911 Avenue Agropolis
BP 64501, 34394 Montpellier cedex 5, France (T.J.T., S.D., T.J., A.C.)
CIRAD, UMR AGAP, F-34398 Montpellier, France (X.A., M.S., D.C.)
INRAB, CRA-PP, Pobé, Benin (A.O.)
PalmElit SAS, F-34980 Montferrier-sur-Lez, France (B.N.)
CIRAD, UMR DIADE, F-34398 Montpellier, France (F.M.)

¹Financial support provided by PalmElit SAS/IRD/CIRAD
²equal contribution
*Corresponding authors
ABSTRACT

Fruit provide essential nutrients and vitamins for human diets. Not only is the lipid rich fleshy mesocarp tissue of the oil palm (*Elaeis guineensis* Jacq.) fruit the main source of edible oil for the world, but it is also the richest dietary source of provitamin A. The present study examines the transcriptional basis of these two outstanding metabolic characters in the oil palm mesocarp. Morphological, cellular, biochemical and hormonal features defined key phases of mesocarp development. A 454-pyrosequencing derived transcriptome was then assembled for the developmental phases preceding, and during maturation and ripening when high rates of lipid and carotenoid biosynthesis occur. 2,629 contigs with differential representation revealed coordination of metabolic and regulatory components. Further analysis focused on the fatty acid (FA) and triacylglycerol (TA) assembly pathways, and during carotenogenesis. Notably, a contig similar to the Arabidopsis seed oil transcription factor WRINKLED1 was identified with a transcript profile coordinated with those of several FA biosynthetic genes and the high rates of lipid accumulation, suggesting some common regulatory features between seeds and fruits. We also focused on transcriptional regulatory networks of the fruit, in particular those related to ethylene transcriptional activation and MADs-box proteins. Our results suggest key roles for AGL2/SEPALLATA, AGAMOUS and GLOBOSA/PISTILLATA-like proteins in the mesocarp, and a central role for ethylene coordinated transcriptional regulation of type VII Ethylene Response Factors during ripening. Our results suggest divergence has occurred in the regulatory components in this monocot fruit compared to those identified in the dicot tomato fleshy fruit model.

INTRODUCTION

Fruit development, maturation and ripening are complex biological processes unique to plants. The monocotyledonous oil palm (*Elaeis guineensis* Jacq.) fruit is a drupe whose thick fleshy mesocarp is exceptionally rich in oil (80% Dry Mass, DM), making this species the highest oil-yielding crop in the world (Murphy, 2009). The mesocarp is also especially abundant in carotenoids and crude palm oil is the richest dietary source of provitamin A (Sambanthamurthi et al., 2000; Solomons and Orozco, 2003). Surprisingly, the molecular basis of oil palm fruit development, maturation and ripening has received very little attention. In contrast, the fleshy berries such as tomato and grape are considered models due to the wealth of genome resources, and genetic transformability. In these species hormones play key roles. For example, ripening appears to be controlled by both ethylene-dependent and independent pathways shown by the characterization of the tomato ripening related mutants
such as *ripening-inhibitor* (*rin*), *Colorless on-ripening* (*Cnr*), *Never-ripe* (*Nr*) and *Green-ripe* (*Gr*) (Wilkinson et al., 1995; Vrebalov et al., 2002; Barry and Giovannoni, 2006; Manning et al., 2006; Giovannoni, 2007). Abscisic acid (ABA) can also control tomato ripening through the activation of ethylene biosynthesis (Zhang et al., 2009a). In grape, whereas ripening was previously thought to be ethylene independent, ABA appears to control ripening along with auxin and ethylene (Coombe and Hale, 1973; Davies et al., 1997; Chervin et al., 2004; Chervin et al., 2008).

Another common regulatory feature of fruit has emerged from functional analyses with the MADs-box family of transcription factors, including RIN and TAGL1 from tomato, PLENA from peach, and FaMADS9 from strawberry, which have provided evidence for roles during ripening (Vrebalov et al., 2002; Tadiello et al., 2009; Vrebalov et al., 2009; Seymour et al., 2010). The recent identification of diverse MADS expressed during banana ripening raise the questions of whether the function of some MADS-box proteins are common to monocots, and whether divergence occurred after the separation of dicots and monocots (Liu et al., 2009; Elitzur et al., 2010). It is clear that tomato and grape with their different modes of ripening provide an excellent framework to compare the maturation and ripening regulatory mechanisms in diverse eudicot species. However, due to the lack of data, it is not clear whether similar or diverse regulatory mechanisms function during the maturation and ripening in monocot fruit species in general and oil accumulating fruits tissues such as the oil palm mesocarp in particular. Indeed, the low content or lack of triacylglycerols (TAG) and provitamin A in these eudicot model species make them inadequate to understand the regulatory mechanisms that function during the maturation and ripening in the oil-rich mesocarp.

Current knowledge about oil synthesis is derived mainly from research on seeds in which TAG accumulate during maturation (Bates et al., 2009; Baud and Lepiniec, 2010). In particular, the use of genomics and the model species *Arabidopsis thaliana* has allowed new insights into seed lipid synthesis (Lu et al., 2009; Zhang et al., 2009b; Li-Beisson et al., 2010) and its regulation (Baud and Lepiniec, 2009). Though genes and enzymes involved in some steps of fatty acid (FA) and TAG synthesis still have to be identified or functionally validated, the knowledge is advanced enough to enable oil synthesis pathway reconstruction from transcriptome data in non-model species (Joët et al., 2009). In contrast, far less is known about the molecular basis of lipid metabolism in the fleshy fruits that accumulate high amounts of TAG in their mesocarp, such as avocado, olive and oil palm. Furthermore, very few lipid-related genes have been cloned in oil palm (Othman et al., 2000; Nakkaew et al., 2007).
and most of the studies on oil metabolism in this species deal with metabolic control analysis experiments using callus cultures (Ramli et al., 2009). To our knowledge, lipid synthesis in the developing oil palm fruit has never been investigated through transcriptome analysis. In contrast, transcriptomic studies of developing seeds have revealed many key features of oil synthesis. The main processes leading to the accumulation of TAG in the seed begin in the plastid with the de novo formation of acyl chains (up to 18 carbons), followed by the desaturation at the Δ9 position and FA release, then continues in the endoplasmic reticulum (ER) with the sequential acylation on the three positions of the glycerol backbone, additional FA desaturation, esterification to phosphatidylcholine (PC) and acyl editing, and finally storage within specialized organelles called oil bodies. The identification of a very limited number of expression patterns for the numerous genes involved in FA and TAG synthesis first suggested that the transcriptional network that control seed lipid metabolism is remarkably coordinated (Ruuska et al., 2002; Baud and Lepiniec, 2009). In the Arabidopsis embryo, almost all genes involved in de novo FA synthesis show the same timing and pattern of expression, i.e. a bell-shape pattern whose peak coincides with the onset of lipid accumulation. In contrast, the transcription of genes required for TAG assembly in the ER rises later and remains high during the maturation process. These patterns have also been observed in the copious persistent endosperm of coffee, another seed tissue that accumulates oil (Joët et al., 2009). The analysis of Arabidopsis seed microarray data also revealed that the fold increase of most genes involved in TAG assembly and storage was significantly higher than that of genes of the core FA biosynthetic machinery (Baud and Lepiniec, 2009). Whether these important characteristics of seed oil synthesis are conserved and function in non-seed tissues such as the mesocarp of the oil palm fruit remains unknown.

The fact that most genes governing de novo FA synthesis in the plastid share the same temporal transcription pattern strongly suggests that they are co-regulated and probably share common cis- and trans-regulatory elements. The isolation of the Arabidopsis wrinkled1 (wri1) mutant constituted a key milestone in the validation of this hypothesis. This mutant is specifically impaired in seed TAG accumulation and shows lower transcript levels for key enzymes involved in lipid and carbohydrate metabolism (Focks and Benning, 1998; Ruuska et al., 2002). Furthermore, overexpression of WRII, that encodes an AP2/EREB family transcription factor, led to increased seed oil content and up-regulation of certain FA biosynthesis genes (Cernac and Benning, 2004; Maeo et al., 2009). Indeed, WRII interacts directly with an AW-box sequence [CnTnG(n)CG] in the upstream regions of several FA biosynthesis genes (Maeo et al., 2009). In addition, WRII expression appears to be under
control of LEAFY COTYLEDON1 (LEC1) and 2 (LEC2), two master regulators of seed maturation (Baud et al., 2007; Mu et al., 2008). While an analogous regulatory system appears to function in maize through the ZmWRIM and ZmLEC1 orthologs (Shen et al., 2010), a maize ortholog of LEC2 has not been identified, suggesting differences in this key regulatory network between monocots and dicots. Finally, whether similar regulatory systems function in the lipid rich mesocarp during fruit ripening compared to that during seed maturation is unknown.

The oil palm mesocarp not only accumulates the largest amount of TAG compared to all other crop species, but also large amounts of carotenoids (500 μg.g⁻¹ of mesocarp DM) predominantly α and β-carotene (Ikemefuna and Adamson, 1984). Carotenoids are precursors of vitamin A, an essential nutrient for human beings. Carotenoids and their oxygenated derivatives xanthophylls give fruits their pigmentation and play an important role as visible signals to attract animals for seed dispersal (Tanaka et al., 2008). The cleavage of carotenoids also generates apocarotenoids, including the plant hormone ABA with emerging roles during fruit ripening (Davies et al., 1997; Deluc et al., 2007; Gambetta et al., 2010). In addition, volatile apocarotenoid compounds such as β-ionone and geranylacetone have animal attracting characteristics with very low odor thresholds and play crucial roles in fruit dispersion (Auldridge et al., 2006a; Auldridge et al., 2006b).

Despite the attractive quantitative and qualitative carotenoid composition of the oil palm mesocarp, very few studies have focused on the transcriptional mechanisms underlying carotenoid gene metabolism or performed metabolite profiling during mesocarp development (Ikemefuna and Adamson, 1984; Khemvong and Suvachittanont, 2005; Rasid et al., 2008). Only two genes, PSY and DXS that encode phytoene synthase and 1-deoxy-D-xylulose-5-phosphate synthase respectively, have been investigated to date in oil palm. PSY and DXS are involved in the first committed step of carotenoid biosynthesis and the upstream 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway, respectively. By contrast, virtually all carotenogenic genes have been identified in model plants (Hirschberg, 2001; Howitt and Pogson, 2006; Chen et al., 2010) and this genomic knowledge should facilitate the identification of gene orthologues in the oil palm mesocarp. During the ripening of the fleshy tissue of tomato, chloroplasts redifferentiate to chromoplasts that accumulate large amounts of lycopene, a compound with strong antioxidant properties but no pro-vitamin A activity (Arango and Heise, 1998; Josse et al., 2000). While tomato is a reference for carotengenesis in fleshy fruit, it does not constitute an optimal model to study the transcriptional regulation of downstream pathways of carotene synthesis.
Our objective is to provide a basis to understand the molecular regulation and coordination of TAG and carotenoid biosynthesis during oil palm mesocarp maturation and ripening compared to those found in seeds and non-oily fruit. We use morphological, histological and biochemical analyses to define phases of oil palm fruit development, maturation and ripening, and in parallel, we generated and analyzed the mesocarp transcriptome during maturation and ripening using 454 pyrosequencing. By studying the accumulation of oil and carotenoids and hormonal profiling during oil palm fruit development and analyzing the differentially expressed transcripts, we have characterized key regulatory steps of the specific pathways and corresponding genes that are likely to function during fruit maturation and ripening. We describe the unique characters of lipid and carotenoid metabolism in this species, and provide insight into the transcriptional coordination and hormonal regulation in the mesocarp tissue compared to those in developing seeds and fleshy fruits. An overview of the metabolic networks involved in lipid and carotenoid accumulation and transcriptional networks in the oil palm mesocarp will allow comparison with model species and help identify candidate genes and molecular markers associated with these important chemical traits for breeding programs.
RESULTS

Phases of Oil palm Fruit Development, Maturation and Ripening Defined by Morphological, Cellular, Biochemical and Hormonal Parameters of the Mesocarp

Under the field conditions and with the genetic material used, the fruit of *E. guineensis* completed their development, maturation and ripening in approximately 160 days. Similar to what occurs in other drupes, a biphasic growth curve was observed, with an initial increase in fruit mass and size measured between 30 to 60 days after pollination (DAP; Fig. 1, A and B). After a 40-day lag period (60-100 DAP), a further increase in fruit mass was observed between 100 and 160 DAP, in particular between 140-160 DAP accompanied with an increase in fruit size (Fig. 1, A and B). A similar pattern is observed with the mesocarp fresh mass, which represents approximately 75% of that of the ripe fruit. Notably, a large increase in DM between 120 and 160 DAP reflects lipid accumulation in this tissue (Fig. 1, C and E). Further analysis of selected histological, biochemical and hormonal parameters allowed us to define five distinct phases of mesocarp development. Phase I between 30-60 DAP is defined by anticlinal cell divisions and expansion along with the initial increase in fruit mass and size (Fig. 1, A-C, and F). Phase II, between 60-100 DAP, is a transition period characterized by a lag in the accumulation of fresh mass, and also by peak amounts of IAA and IAA-conjugates (Fig. 1D; Supplementary Fig. S1). Phase III between 100-120 DAP is the end of the transition period during which decreases in auxin, GA and cytokinin metabolites are observed (Fig. 1D; Supplementary Fig. S1). Phase IV is concomitant with the beginning of maturation characterized by an increase in mesocarp fresh mass and beginning of lipid accumulation detected by 120 DAP (Fig 1, C, E and G). Phase IV between 120-140 DAP is characterized by lipid (more than 2 grams per fruit) and carotenoid accumulation in addition to the low amounts of all hormone metabolites examined including auxin, GA and cytokinin, ABA and ethylene (Fig. 1, D and E; Supplementary Fig. S1). Finally, during the ripening Phase V, there is a large increase in the hormones ABA and ethylene, and cell wall detachment related to ripening processes in the mesocarp are visualized (Fig. 1, D and H; Supplementary Fig. S1). During this phase, dry and fresh fruit mass increase massively, concurrently with lipid and carotenoid accumulation in the mesocarp (Fig. 1, A, C, E). As observed at 160 DAP, lipids accumulate within sub-cellular spherical organelles (10-15µm in diameter, 6-12 per mesocarp cell) that occupy the volume of the cells (Fig. 1H). In addition, as observed in ripe fruits, mesocarp cells contain distinct regions, presumably chromoplasts, with high carotenoid concentrations (Fig. 1H).
Statistical analysis and gene ontology annotations of contigs differentially represented during mesocarp development

A total of 29,034 contigs were obtained after clustering and assemblage of good quality sequence reads (Supplementary data S2 and S3). Read amounts associated with each of the 29,034 contigs were then analyzed using Audic-Claverie and FDR statistics leading to the identification of 2,629 differentially expressed contigs, which are henceforth referred to as contig group I, while the remaining 26,405 contigs with either non-differential representation or low read abundance are referred to as contig group II. Hierarchical Cluster Analysis of Group I contigs allowed the identification of 4 major clusters, termed A, B, C and D, which contain 695, 418, 1,167 and 349 contigs, each divided into 5, 4, 5 and 3 sub-clusters, respectively (Supplementary Fig. S4). In general, contigs that exhibited a transcription peak at 100, 120, 140 or 160 DAP were grouped in Cluster A, B, C and D, respectively. Gene Ontology (GO) accessions were assigned to the 2,629 group I contigs using the Blast2Go platform (Gotz et al., 2008). A total of 2,074 contigs (77%) were assigned GO accessions, 488 in Cluster A (70%), 377 in Cluster B (90%), 908 in Cluster C (78%) and 301 in Cluster D (86%). To identify differences in the GO annotations of the four clusters that reflect underlying molecular processes, an enrichment analysis of the annotations of each individual cluster compared with the annotations of the 2,074 contigs was performed using the GOSSIP platform (Bluthgen et al., 2005). Cluster B had the most over-represented GO annotations (131 terms), followed by Cluster D (22) and Cluster C (4), while Cluster A had only under-represented GO annotations (Supplementary Table S5). Cluster C had the most under-represented annotations (51 terms), followed by Cluster D (43), Cluster B (32) and Cluster A (17). The most significant under-represented GO terms in Cluster A were related to carbohydrate metabolic processes including glycolysis, while the most significantly over-represented terms in Cluster B were FA biosynthetic process, cellular carbohydrate metabolic process and lipid biosynthetic process. The results suggest in the 120 DAP mesocarp, at the beginning of phase IV, significant changes to the transcriptome coincide with the onset of lipid accumulation observed (Fig. 1E). By contrast, terms related to nucleic acid binding, nucleus, DNA binding and lipid catabolic process are over-represented in Cluster C, which suggest major transcriptional regulatory changes are initiated by the end of phase IV at the onset of ripening. Finally, the over-represented terms in Cluster D relate to carbohydrate metabolic processes including O-glycosyl compound hydrolysis, glycosyl bond hydrolysis, glucan endo-1,3-beta-D-glucosidase activity, chitinase activity and cellular aromatic
compound metabolic processes which suggest ripening processes such as cell wall modifications are taking place by 160 DAP. Overall, this analysis indicates that certain changes in the transcriptome are coordinated and reflect the biochemical and physiological processes observed during the maturation and ripening phases of the mesocarp (Fig. 1).

**Reconstructing FA and TAG biosynthetic pathways in the oil palm mesocarp using a pyrosequencing-based transcriptome**

Based on our current knowledge of oil synthesis in seeds, the pyrosequencing-based transcriptome approach enabled us to reconstruct the FA and TAG synthesis pathways in the oil palm mesocarp (Fig. 2). Indeed, for almost all lipogenic steps, contigs that exhibited high similarity to genes from Arabidopsis and other oleaginous species were found (Supplementary Table S6).

Three enzymes appear to participate in the acylation of DAG to TAG in the oil palm mesocarp. First, the presence of two contigs (CL875Contig1 and CL1Contig9221) similar to type 1 and type 2 diacylglycerol acyltransferases (DGAT), respectively, suggests both classes of DGAT play a role in TAG synthesis in this tissue. Moreover, the route involving PC as the acyl donor may make a similar contribution to the final acylation step of DAG given that a contig (CL1451Contig2) highly similar to an Arabidopsis phospholipid:diacylglycerol acyltransferase (AtPDAT1; At5g13640) had comparable total read amounts to that of DGAT1- and DGAT2-like contigs (about 40 reads per 200,000). Furthermore, no contig was found for *Arachis hypogea* cytosolic DGAT or the Arabidopsis bifunctional wax ester synthase/DGATs, nor did we identify contigs for an acyltransferase with MBOAT domains that displayed read levels higher than the DGAT- and PDAT-like contigs described above. Overall, these results suggest that the three DGAT1-, DGAT2- and PDAT-like contigs identified indeed encode the enzymes responsible for final TAG assembly in the oil palm mesocarp.

Because the reaction catalyzed by PDAT also produces lysophosphatidylcholine (LPC), a lysophospholipid acyltransferase (LPCAT) is necessary to cooperate with PDAT to regenerate PC from LPC. Interestingly, a contig (CL212Contig1) highly similar to AtLPLAT1 and AtLPLAT2, two lysophospholipid acyltransferases recently characterized in Arabidopsis (Stahl et al., 2008), had read amounts similar to the PDAT- and DGAT-like contigs described above. It is also possible that this LPCAT-like contig participates to acyl editing (Bates et al., 2009), though the requirement of this mechanism in a tissue producing an oil poor in polyunsaturated
unsaturated FAs (PUFAs) is unknown. In this regard, no contig with high similarity with Arabidopsis phospholipases A2 was found.

Finally, no contig was identified for the recently characterized phosphatidylcholine:diacetylgerol cholinephosphotransferase (PDCT), which is encoded by ROD1 and contributes to PC-DAG interconversion in developing Arabidopsis seeds together with CDP-choline:diacylglycerol cholinephosphotransferase (CPT) (Lu et al., 2006). Furthermore, the absence of any oleosin transcript in our libraries may provide an explanation for the considerable size of the oil droplets that accumulate in the oil palm mesocarp (10-15 µm; Fig. 1H). This size is similar to that of the large lipid bodies in the seeds of Arabidopsis lines in which oleosins were suppressed or severely attenuated (Siloto et al., 2006).

**FA synthesis in the plastid and TAG assembly in the ER are governed by two different transcriptional programs**

The reconstruction of the FA and TAG biosynthetic pathways revealed that de novo formation of acyl chains in the plastid and TAG assembly in the ER follow two distinct transcriptional programs. Indeed, almost all genes involved in glycolysis and FA synthesis in the plastid were from group I, which suggests strong developmental transcriptional regulation of this pathway in the mesocarp. In addition, almost all contigs of the core FA biosynthetic machinery showed maximal transcription at 120 DAP and grouped in Cluster B as determined by HCA (Supplementary Fig. S4), and displayed high read levels (average total of 424 per 200,000 reads, up to 1993 reads for EAR, enoyl-ACP reductase). Furthermore, the transcription peak of the core FA biosynthetic machinery coincided with the onset of oil accumulation in the mesocarp at the beginning of the maturation phase IV (Fig. 1E and 2). By contrast, for the TAG assembly pathway, only contigs for oleate and linoleate desaturases (FAD 2 and FAD3) were found within group I. For the enzymes involved in the sequential acylation on the three positions of the glycerol backbone, PC-DAG interconversion and acyl editing, contigs were identified in group II only, suggesting very little or no developmental control of transcription of genes involved in TAG assembly (Fig. 2). Moreover, contigs related to TAG assembly displayed no clear peak, but rather showed low (81 reads per 200,000) and constant read amounts. This unambiguous transcriptional dissimilarity between the pathways in the two compartments and the high homogeneity of patterns observed within each block suggest that FA synthesis in the plastid and TAG assembly in the ER are controlled by two different transcriptional programs.
Because transcripts encoding enzymes involved in de novo FA synthesis were grouped in cluster B which is characterized by profiles that peak at 120 DAP and in particular in sub-cluster B1 (Supplementary Fig. S4), a careful examination of the most highly represented contigs of this sub-cluster, which contains only 148 contigs, was carried out in order to identify potential regulatory elements of FA synthesis. This analysis revealed that the most abundant contig (CL1Contig558, renamed here EgLIP1) of this sub-cluster is highly similar to a TAG lipase class 3 family protein, which could account for the intensive TAG hydrolysis in the oil palm mesocarp found previously (Ngando Ebongue et al., 2006). The second most abundant contig within sub-cluster B1 (CL1Contig8373) has high similarity with the transcriptional regulator WRINKLED1 (Supplementary Table S6), demonstrating that the pyrosequencing-based approach chosen in the present study together with appropriate statistics constitutes a reliable and straightforward method to examine co-regulation of genes involved in a given pathway and identify candidates involved in their regulation.

A transcript encoding a transcription factor similar to WRINKLED1 is highly expressed in the mesocarp during FA Biosynthesis

CL1Contig8373 belongs to the WRI clade (Supplementary Fig. S7), which contains WRII (with 68% identity) and two other WRII-like sequences from Brassica napus (BnWRII, ABD72476) (Liu et al., 2010) and maize (ZmWRII, AY103852) (Shen et al., 2010) for which similar WRI function was shown in the seed. The contig CL1Contig8373, renamed EgAP2-2, peaks at 120 DAP, which corresponds to the onset of lipid accumulation and the transcript profile peak for many FA biosynthesis genes (Fig. 1E and Fig. 2). From the 19 transcripts expressed in the mesocarp encoding FA and TAG biosynthetic enzymes that contained a proximal upstream region long enough (>150bp), canonical AW-box elements were found in 6 transcripts (Supplementary Fig. S7). All of the transcripts encode enzymes similar to those that are active in the plastid and involved in glycolysis (ENO, PK) and FA synthesis (ACC CTα, KASIII, SAD and FaTB). The 15-bp element described by (Baud et al., 2009) was not found in the 5’UTR of oil palm sequences available.

WRII is known to be a target of LEC2 and LEC1, which are both master-regulators of the seed maturation process and also involved in FA biosynthesis regulation (Baud et al., 2007; Mu et al., 2008; Shen et al., 2010). The analysis revealed an absence of any sequences with significant similarity to transcripts encoding ABI3/VP1, FUS3 and LEC2, all of which belong to the B3 superfamily. Similarly, no contig with HPA3 subunits of the CCAAT-binding factor (IPR003956) was found among group I contigs.
The massive carotene accumulation precedes ABA biosynthesis

During mesocarp development, there is a spectacular increase in α- and β-carotene contents (expressed as total mesocarp carotenoid per fruit) during the maturation and ripening phases IV and V (Fig. 1E). While prior to 120 DAP only minimal amounts of the xanthophylls lutein and violaxanthin were detected, the total carotene content (α- and β-carotenes) reached very high amounts (718 mg g⁻¹ DM) by 160 DAP. Only trace to low quantities of ABA were observed from 80 to 140 DAP, whereas a six-fold increase in the ABA content was observed between 140 and 160 DAP (approximately 300 ng g⁻¹ DM). The amount of the conjugated storage form ABA glucose ester (ABA-GE) also increased 3-fold between 140 and 160 DAP (Fig. 3).

As with FA and TAG synthesis pathways, the 454 sequencing dataset enabled the identification of at least one contig for almost each step of the carotene, xanthophyll and ABA biosynthetic pathways (Fig. 3; Supplementary Table S6). However, only three contigs involved in carotenogenesis were found within group I. These three genes, namely PSY, PDS and PTOX, all encode for enzymes involved in the first two committed steps of carotenogenesis: condensation of two molecules of geranylgeranyl diphosphate (PSY), desaturation of the resulting phytoene (PDS), and recycling of the plastoquinol pool used as an electron acceptor for phytoene desaturation (PTOX). Moreover, these three contigs displayed maximal expression levels at 140-160 DAP, during which α- and β-carotene accumulated massively (Fig. 1E and Fig. 3). Therefore, these findings suggest that the two first committed steps of the carotenoid pathway, accomplished by PSY and PDS, are the major transcriptional control points for mesocarp carotenoid metabolism. Furthermore, despite the fact that the other carotenogenic genes were found only within group II, many displayed similar expression profiles. For example, two contigs similar to Z-ISO and ZDS, also involved in the desaturation steps, and fibrillins that are associated with carotenoid storage, all have higher read amounts at 160 DAP. Overall, these results suggest a coordinated transcriptional regulation of carotenogenesis in the oil palm mesocarp.

By contrast, genes involved in the upstream pathway, which is common with the biosynthesis of other isoprenoids, appeared to be expressed at lower levels, although there are putative paralogs for both DXS and FPS expressed during the carotenogenic phase. Similarly, genes acting downstream of the desaturation steps, in particular those encoding LCY-E and LCY-B which direct lycopene toward α- or β-carotenes, were expressed rather weakly. However, considering that lycopene was not detected and that the mesocarp accumulates
exceptional amounts of carotenes, LCY-E and LCY-B were obviously not rate-limiting steps. Therefore, as previously described above for the TAG assembly pathway, the present work provides evidence that a given metabolic pathway may be highly active without requiring that all biosynthetic steps display high and regulated transcript abundance.

**Coordinated transcriptional activation of ABA metabolism at ripening**

Genes encoding neoxanthin synthase (NSY), xanthosin dehydrogenase (XO) and abscisic acid glucosyl transferase (ABAGT), all show a significant (group I) increase in transcript abundance between 140-160 DAP (Fig. 3). In addition, genes encoding for carotenoid cleavage dioxygenases (CCDs) and ABA8OX contigs displayed a similar transcription pattern, even though they were found in group II. Furthermore, this marked upregulation of ABA biosynthetic genes during mesocarp ripening matched perfectly the dramatic increase in ABA content at this stage (Fig. 1D; Fig. 3). The metabolism of ABA therefore clearly displayed a coordinated transcriptional activation at ripening. The results also suggest local transcriptional control of ABA biosynthesis rather than transport from other tissues. Finally, the delay between carotenoid accumulation and subsequent ABA biosynthesis is consistent with the well-established control of ABA synthesis through the cleavage of 9-cis-xanthophylls rather than through the activation of *de novo* xanthophyll synthesis (Thompson et al., 2000).

**Chloroplasts and chromoplasts share a common repertoire of carotenogenesis regulators**

In contrast with our extensive knowledge of the enzymatic steps required for carotenoid synthesis in plants, relatively little is known about the regulatory mechanisms of this pathway (Lu and Li, 2008). Carotenoid content in the chromoplast has been related to the transcriptional control of genes involved in light signaling (*PHY-A, DDB1*) and plastid morphogenesis (*Or*) (Liu et al., 2004; Lu et al., 2006; Cazzonelli and Pogson, 2010), while for chloroplasts, genes involved in ethylene signaling (*RAP2.2*), chromatin modification (*SDG8*) and light signaling (*DET1, PIF1*) were reported (Mustilli et al., 1999; Welsch et al., 2007; Cazzonelli et al., 2009; Toledo-Ortiz et al., 2010). In the oil palm mesocarp, contigs similar to regulators from both types of plastids were found. For example, DDB1 and a histone methyltransferase (SET), and Or and DET1, for chloroplasts and chromoplasts, respectively, were found to have peak transcript amounts at 140 DAP, which presumably coincides with the chloroplast to chromoplast transition when carotenoids accumulate (Fig. 3). Similarly,
contigs encoding type VII ERF/RAP2.2 transcription factors, which may mediate PSY and PDS expression (Welsch et al., 2007), were found to be upregulated at 140 DAP (Fig. 4B). Finally, a contig similar to the gene encoding a phytochrome interacting factor (PIF), shown to down-regulate the accumulation of carotenoids by specifically repressing PSY (Toledo-Ortiz et al., 2010), was only expressed at 100 DAP, concomitantly with phytochromes (PHY-A, -B and –C), prior to the onset of carotene accumulation.

The oil palm mesocarp exhibits characteristics of a climacteric fruit with an autocatalytic burst of ethylene coordinated transcriptional activity

To determine the involvement of ethylene in the function, maturation and ripening of the oil palm mesocarp tissue, we first analyzed ethylene produced in the mesocarp during selected stages of development (Supplementary Fig. S1; Fig. 4A). The lowest amounts of ethylene were detected in the 120 DAP mesocarp (207 nl g⁻¹ FW⁻¹ h⁻¹) while the maximum amounts were detected at 160 DAP (7,263 nl g⁻¹ FW⁻¹ h⁻¹), which represents a 35 fold increase in ethylene production. Within group I, we found contigs with similarity to transcripts that encode the three key enzymes involved in ethylene biosynthesis (Fig. 4A). We also identified transcripts which encoded proteins similarity to ACC oxidase, the most prevalent of which (CL1Contig999) was similar to the Arabidopsis ACO4 and peaked at 140 DAP that coincides to the large increase in ethylene production in the mesocarp. Notably, some ACC oxidase transcripts were more abundant at 100 DAP, while other distinct transcripts strongly increased at 140 DAP. There were also three contigs within group I with similarity to aspartate aminotransferase-like (ATT), all (in particular CL1Contig505) with peak transcript abundance at 140 DAP. ACC synthase enzymes are part of the ATT family and based on their expression profiles, these contigs are good candidates to participate in ethylene biosynthesis in the mesocarp. We also identified two contigs with low read amounts similar to Arabidopsis ACC synthases. The data support the hypothesis that ethylene production in the mesocarp in controlled through a similar process as found in tomato, namely an autoinhibitory system-1 ethylene production during development, followed by a transition to a climacteric-like autocatalytic system-2 production during ripening through the transcriptional activation of specific transcripts for ACC synthase and ACC oxidase (Nakatsuka et al., 1998; Barry et al., 2000; Yokotani et al., 2009).

To further examine the activity of ethylene related processes in the mesocarp, we searched for contigs with similarity to transcripts that encode ethylene receptors, signal transduction and ethylene response factors (ERFs). Notably, we identified contigs with
similarity to key factors within group I, with the exception of ETHYLENE INSENSITIVE2 (EIN2) and CONSTITUTIVE TRIPLE RESPONSE (CTR1) that were within group II (Fig. 4B). As observed with the transcripts for ethylene biosynthesis enzymes, the transition from system-1 to system-2 ethylene production is marked by a coordinated increase in a large number of ethylene related transcripts, in particular those encoding ERFs. Furthermore, certain components of ethylene perception and signal transduction have transcript profiles that coincide with either system-1 or -2. For example, two contigs similar to the ethylene receptors ETR1 and EIN4 had peak read amounts at 100 DAP, while a contig similar to the key transcriptional activator EIN3 peaked at 140 DAP. The most striking observation is the number of ERF like transcripts with expression profile peaks at 140 DAP. Phylogenetic analysis with Arabidopsis AP2 domains distinguished the 10 major groups (Nakano et al., 2006), however no oil palm fruit sequences were assigned to groups VI and VIII (Supplementary Fig. S8). The results indicated the most predominant ERF found in the oil palm mesocarp was the type VII, the majority of which have peak transcript amounts at 140 or 160 DAP, during mesocarp maturation and ripening (Supplementary Fig. S8; Fig. 4B). In contrast, only a single ERF/AP2 type VII (CL1Contig256) and one type IX (CL1Contig6009) had peak transcript amounts at 100. At both 120 and 140 DAP type I and IV ERF/AP2s were most prevalent, while a single type X was identified at 140 DAP. Overall, a large number of ethylene related transcripts have expression profiles associated with either a system-1 basal amount of ethylene production, or system-2 ethylene production during ripening. In particular, type VII ERF/AP2s represent a major transcriptional response associated to the ethylene burst measured during the maturation and ripening phases IV and V of the mesocarp.

A Diversity of MADs-box Transcription Factors Expressed in the Oil Palm Mesocarp at Keys Stages

Our pyrosequencing-based approach allowed the identification of 14 sequences that contained at minimum the MADs-box domain and were retained for further phylogenetic analysis (Fig. 5; Supplementary Fig. S9). The 14 oil palm MADs-box sequences belong to 7 different subfamilies defined by the genes from Arabidopsis AP1 (APETALAI), PI (PISTILLATA), AP3 (APETAL3), AG (AGAMOUS), SEP (SEPALATTA), AGL6 (AGAMOUS-like 6) (Becker and Theissen, 2003), in addition to the tomato TM3 subfamily (Supplementary Table S6).

The class of MADs most represented in the mesocarp belong to the AGL2/SEPALATTA subfamily including 3 within group I and 2 within group II (Fig. 5A).
Notably, there were no oil palm sequences found in the SEP sub-clades 1 and 5 that contain either the strawberry FaMADS9 or the tomato LeRIN respectively, both of which perform key functions during fruit development and ripening (Vrebalov et al., 2002; Seymour et al., 2010). Overall, the five contigs within the AGL2/SEPALLATA subfamily are grouped within two distinct sub-clades supported by significant bootstrap values. Within the SEP3 sub-clade, EgAGL2-1, reported in oil palm male and female inflorescences (Adam et al., 2006), CL1Contig8010/EgAGL2-2 (100% amino acid identity) and CL1Contig6409 form a monophyletic group, while CL1Contig3848 forms a monophyletic group with the banana fruit MaMADS2 and MaMADS4 (Elitzur et al., 2010). Contig CL1Contig8010 is the most abundant MADS found in the mesocarp with an expression profile coordinated during maturation and ripening (Fig. 5A). CL1Contig2367 and CL1Contig1295 are found within another sub-clade that contains uniquely monocot sequences. Notably, CL1Contig2367 peaks at 100 DAP and decreases during maturation and ripening suggesting a function prior to phase III of mesocarp development. Furthermore, the oil palm sequences systematically form sub-clades with other monocot sequences, do not group with those from dicot fruit species, that suggests a diversification of function within this subfamily has occurred within monocots since the separation from eudicots.

There were three contigs identified within the MADs-box AG subfamily. CL1Contig1226 is the most abundant and peaks at 140 DAP (Fig. 5B). The three oil palm AG-like contigs analyzed comprise a monophyletic group with the banana fruit MaMADS5 (Elitzur et al., 2010) (Fig. 5B). The mesocarp fruit AG-like sequences are distinct from EgAG1 and EgAG2 reported in oil palm male and female inflorescences that group together in the eudicot AG-like (euAG-like) sub-clade (Adam et al., 2006). The cladogram indicates seven monophyletic groups supported by significant bootstrap values. The PLENA and euAG lineages include the TOMATO AGAMOUS-LIKE1 (TAGL1) and TOMATO AGAMOUS (TAG1) respectively that have regulatory functions during tomato flower and fruit development respectively (Seymour et al., 2008; Pan et al., 2010). Notably, these two groups contain sequences exclusively from eudicot fruit species such as peach, apple or grape.

From the GLOBOSA/PISTILLATA (GLO/PI) subfamily, a single contig from group I, CL290Contig1, shares 100% nucleic acid identity with EgGLO1 from oil palm previously identified (Supplementary Fig. S9) (Adam et al., 2006). The oil palm GLO-like protein is separated from the fleshy fruit eudicot sequences and positions closest to the banana MaMADS6, the only available sequence from a monocot fleshy fruit species in this subfamily.
The accumulation of CL290Contig1 is significantly increased during maturation and remains high during ripening.

Profiles of transcription factors and regulators in the mesocarp suggest coordinated transcriptional mechanisms occur in the mesocarp

A total of 1,930 group I contigs (73%) were assigned INTERPRO accession annotations, 150 contigs with functions related to transcriptional regulation, amongst which 127 contigs were transcription factors (TFs) and transcription regulators (TRs) not previously described in this study (Supplementary Tables S10-S13). Overall, there are 37 TF/TRs contigs with peak read amounts within Cluster A, 14 contigs at 120 DAP within Cluster B, 85 contigs at 140 DAP within Cluster C and 14 contigs at 160 DAP within Cluster D. As observed earlier from the GO annotation enrichment analysis (Supplementary Table S5), the largest proportion of contigs related to transcriptional regulation were found at 140 DAP within Cluster C.

Within Cluster A, with peak transcription amounts at 100 DAP, the most abundant TF/TR is a Cys2/His2-type zinc finger protein found within sub-cluster A3 (Supplementary Table S10). Notably, there are two NAC domain proteins that are also abundant within the same sub-cluster A3 (Supplementary Fig. S4). Indeed, NAC domain proteins are the most represented class of proteins at 100 DAP, three within sub-cluster A1 and two within sub-cluster A3 suggesting coordinated regulation of this class of TF/TR. AUXIN RESPONSE FACTORS (ARFs) are the second most represented class, with four ARFs all within sub-cluster A1 suggesting strong transcriptional coordination amongst this class of TF/TR. Another notable feature of Cluster A is the presence of contigs for genes encoding chromatin modifiers including a condensin complex component, a DNA methyltransferase, a polycomb group protein and two contigs for jumonji domain transcription factors implicated in chromatin regulation during development (Takeuchi et al., 2006). The results indicate that several factors involved in chromatin modifications that can affect changes in gene expression are present and may function at 100 DAP.

Remarkably only 14 TF/TRs contigs are found within Cluster B, which suggests restricted transcriptome regulation occurs at this stage in the mesocarp (Supplementary Table S11). In addition to EgAP2.2, the three ERF/AP2 TFs and the MADs-box described in previous sections (Fig. 4B; Fig. 5) there are two calmodulin related signaling proteins, one IAA/AUX protein, one zinc finger protein similar to the maize Indeterminate1 (ID1), one
golden2-like protein that regulates chloroplast development and the expression of the photosynthetic apparatus, one bZIP and a SEUSS-LIKE 3 transcription regulator.

In Cluster C in addition to the type VII ERF/AP2 TFs and the MADs-box TFs described above, two contigs for bZIP TF were found in sub-cluster C1 and C4 (Supplementary Table S12). Other classes that are well represented in Cluster C include five NAC domain proteins, four zinc finger proteins, three BELL1-LIKE HOMEODOMAIN proteins and three WRKY TF proteins. Finally, two contigs (CL1Contig6719 and CL2179Contig1) with similarity (E-value >1e-30) to the homeodomain-leucine zipper protein LeHB-1 that binds to the promoter of LeACO1 to activate transcription during ripening (Lin et al., 2008) were identified.

In Cluster D, eleven out of fourteen contigs are found within sub-group D3 (Supplementary Fig. S4) which suggests tight transcriptional coordination occurs at the ripening stage (Supplementary Table S13). In addition to the AGL2/SEPALLATA described above (Fig. 5A), other prevalent contigs found within sub-cluster D3 included a basic helix-loop-helix protein, a cop9 signalosome complex subunit 3, and a NAC domain protein. Indeed, NAC domain proteins made up the most represented class in Cluster D, notably two of the NAC domain proteins had the highest similarity (Evalue > 1e-50) to the tomato NAC-NOR reported as a component of the ethylene dependent ripening regulation in tomato (Giovannoni, 2004).

Finally, no contigs for CNR nor SlAP2a, regulatory factors that also directly or indirectly affect ethylene production during tomato ripening (Thompson et al., 1999; Chung et al., 2010) were found within group I.

DISCUSSION

The Oil Palm Mesocarp is an Original Fruit Model to Examine Regulatory Mechanisms that Function during Fruit Maturation and Ripening

The oil palm mesocarp presents an original model to examine the regulatory networks in a monocot fruit tissue subject to climacteric ripening and in which high amounts of lipids accumulate. The current study provides for the first time a detailed description of the original physiological and biochemical characteristics along with a thorough analysis of the underlying transcriptional activities, combined to develop a model that describes major events that occur during the defined phases of mesocarp development (Fig. 6).

Similar to during tomato development, a decrease in auxins, GAs and cytokinins (Gillaspy et al., 1993) is also observed prior to mesocarp ripening, however a second auxin
peak that overlaps with that of ethylene during tomato ripening is not observed in the oil palm mesocarp. ABA is also present early during tomato development and a recent detailed study in tomato revealed that an ABA peak occurs just prior to that of ethylene during ripening (Zhang et al., 2009a). By contrast, in the oil palm mesocarp there is a large simultaneous increase in both ethylene and ABA during fruit ripening that suggests regulatory functions and/or interactions for these two hormones in the mesocarp.

In the tomato fruit, the increase in ethylene production at the onset of ripening coincides with a transition from auto-inhibitory (system-1) to auto-stimulatory (system-2) ethylene production. The tomato model predicts system-1 functions prior to ripening, while system-2 operates during ripening and that the two systems are regulated by coordinated developmental and ethylene induced expression of individual ACO, ACS and tomato ethylene receptor genes (Nakatsu et al., 1998; Barry et al., 2000; Yokotani et al., 2009). In the present study, an increase in ethylene production characteristic of a climacteric burst is observed during phases IV and V. Furthermore, the coordinated transcriptional activation of subsets of transcripts for ACO and ACS-like (aspartate aminotransferase-like, AAT-like genes) genes suggests a transition from system-1 to system-2 occurs in the oil palm mesocarp during the maturation phase IV (Fig. 6). Indeed, by 140 DAP, a sharp increase in transcripts for ethylene biosynthetic enzymes and ERFs in the mesocarp are observed which resembles system-2 type ethylene production. Notably, six ERFs with peak read totals at 140 or 160 DAP are type VII ERFs similar to those implicated in the ripening of tomato, kiwi, and plum (Tournier et al., 2003; Wang et al., 2007; El-Sharkawy et al., 2009; Sharma et al., 2010; Yin et al., 2010). Phylogenetic analysis with the oil palm sequences and selected type VII ERF signature domains (Nakano et al., 2006) from Arabidopsis suggest conservation between monocots and dicots and that gene duplication events may have occurred for multiple type VII genes to be expressed in the mesocarp (Fig. 4; Supplementary Fig. S8).

A total of thirteen NAC domain-containing proteins within Clusters A, C and D indicates this class of transcription factors is one of the most highly represented in the mesocarp. Indeed, NAC domain proteins are key participants in transcription factor networks and play central roles in plant development (Olsen et al., 2005). While NAC domain TFs were the most prominent class at 100 DAP, four ARFs were also observed that coincides with the IAA metabolite peak suggesting auxin related functions at this stage (Fig. 6). Furthermore, two other NAC domain contigs were identified with similarity to the tomato NAC-NOR (AAU43922.1) gene, the mutation in which has been reported to be responsible for the nor ripening mutant (Giovannoni, 2004). Notably, both of these NAC Domain contigs have
expression profiles associated with mesocarp ripening that suggests functional similarity to the tomato NAC-NOR gene. While these results suggest conserved functions from some NAC domain proteins during ripening, very little is known about their roles during fruit development. In oil palm, the large number of transcripts observed for this class of TFs with different expression profiles that correspond to key phases of mesocarp development may reflect a high amount of gene duplication events followed by expression domain changes that occurred in this family since the separation between monocots and eudicots.

The Expression of MADs-box Genes from the SEP, AG and GLO Subfamilies Characterize Key Phases of Mesocarp Development and Function

In the current study we find a diversity of MADs-box transcripts with expression associated with the different phases of mesocarp development. MADs-box proteins are key regulators of fruit ripening (Vrebalov et al., 2002; Ito et al., 2008; Itkin et al., 2009; Vrebalov et al., 2009; Gimenez et al., 2010; Jaakola et al., 2010; Seymour et al., 2010). In particular, the protein RIN within the SEP4 sub-clade controls the induction of the tomato ACS transcripts involved in system-2 ethylene production (Barry et al., 2000). Recently, a SEP1/2 sub-clade MADs-box from the non-climacteric fruit strawberry (FaMADS9) was shown to play a key role in ripening (Seymour et al., 2010). To date, very little is known about the SEP subgroup involvement in monocot fruit ripening. Three SEP3 homologues MaMADS1, 2 and 4 expressed in ripening banana fruit tissues share low amino acid sequence similarity (55-62%) with RIN and none thus far complement the tomato rin mutation (Elitzur et al., 2010). However, MaMADS2 is expressed in the pulp prior to the burst of ethylene production and is not induced by ethylene treatments. In the oil palm mesocarp, there were at least five SEP homologues expressed, two of which (CL1Contig8010 and CL1Contig3848) within the SEP3 sub-clade increase during the ethylene burst. Interestingly, phylogenetic analysis indicated that CL1Contig3848 grouped closely to the banana MaMADS2, while neither of which group closely with the tomato RIN, nor with the strawberry FaMADSS9. These results support the hypothesis that the MADs-box SEP-like proteins are key factors of ripening not only in climacteric and non-climacteric, but also in monocot and dicot fruits. Interestingly, FaMADS9 and RIN belong to different SEP subgroups, SEP1/2 and SEP4 respectively, and the strawberry sequence most similar to RIN (FaSEP4) is expressed at very low levels in the fruit. These results demonstrate that while SEP-like proteins indeed operate in the ripening process of both climacteric and non-climacteric fruits, sequence divergence has occurred that may reflect differences between the two ripening models. While the downstream targets of
FaMADS9 are unknown, evidence indicates some common functions may exist during both climacteric and non-climacteric ripening for SEP subgroup MADs-box genes in these two dicots. Our data support the hypothesis that in monocot fruits, SEP3 homologues function upstream of ripening processes analogous to the tomato RIN and the strawberry FaMADSS9, and that the divergence observed within the SEP subgroup may reflect important differences between fruit ripening systems, including those of monocots.

Another tomato MADs-box protein, TAGL1 within the AG subgroup, also functions upstream of ethylene production, apparently through the control of \textit{LeACS2} expression yet independently from RIN (Itkin et al., 2009; Vrebalov et al., 2009; Gimenez et al., 2010). Very little is known about the role of AG clade MADs-box proteins during non-climacteric or monocot fruit ripening. In grape, there are at least three AG-like transcripts expressed in the fruit but unassociated with ripening (Boss et al., 2001; Boss et al., 2002; Diaz-Riquelme et al., 2009). In banana, there appears to be a least one AG-like transcript (\textit{MaMADS5/MuMADS1}) expressed in the pulp and is induced by ethylene (Liu et al., 2009; Elitzur et al., 2010). In the oil palm mesocarp, we identified at least three transcripts that encode AG clade MADs-box proteins. Notable, one transcript (CL1Contig5512) accumulates with no relation to the ethylene burst, while another (CL1Contig1226) accumulates during the initial burst of ethylene production and related transcriptional activity. The domains encoded by these two sequences along with that of a third oil palm sequence and the banana MaMADS5 form a monocot monophylegetic sub-clade. Furthermore, this monocot sub-clade is distinct from those containing the dicot tomato fruit TAGL1, TAG1, or the oil palm EgAG1 and 2 proposed to perform C-functions in the oil palm female flower (Adam et al., 2007a). The results suggest gene duplications in the oil palm genome have occurred that gave rise to at least three distinct yet closely related AG-like sequences with different expression profiles during mesocarp development, maturation and ripening, while \textit{EgAG1 and 2} expression and function may be limited to flower development. However, it is not known whether the three AG-like sequences expressed in the mesocarp are also expressed during flower development or in other oil palm tissues. Together, these results support the hypothesis that some AG clade MADs-box proteins have diverged to gain expanded expression with new functions in the fruit beyond those performed during flower formation. Furthermore, as with the SEP clade MADs-box proteins, a divergence of AG-like proteins between dicots and monocots is clearly observed that may reflect underlying functional differences between these two plant groups in relation to fruit development and ripening.
A transcript (EgGLO1) for a GLO/PI family MADs-box was also observed with expression related to the ethylene burst. The GLO/PI family perform B-functions and to our knowledge, EgGLO1 is the first within this class to be observed with a fruit ripening-related expression profile in any species. Indeed, while GLO members have been identified in eudicot fleshy fruit species such as grape, apple and peach, and in the monocot banana fruit, none thus far are associated with fruit ripening (Yao et al., 2001; Busi et al., 2003; Hileman et al., 2006; Poupin et al., 2007; Zhang et al., 2008; Diaz-Riquelme et al., 2009; Elitzur et al., 2010). In contrast, the oil palm mesocarp EgGLO1 was expressed in relation with the ethylene production and related transcriptional activity during maturation and ripening. Notably, the EgGLO1 nucleotide sequence is identical to the transcript observed previously during flower development (Adam et al., 2006). Indeed, there were two GLO/PI family MADs-box transcripts (EgGLO1 and EgGLO2) found to be expressed during flower development, but only EgGLO1 was observed in the mesocarp. Therefore, either the same gene which functions during flower development is also active and plays a role during fruit ripening, or a duplication has occurred and another similar gene has diverged to acquire a new fruit related function.

Overall, the data implicates the involvement of MADs-box genes, in particular SEP-like, AG-like and GLO-like, during maturation and ripening of the oil palm mesocarp. There appears to have been a substantial amount of functional diversification that has occurred in the SEP-like and AG-like subfamilies given the number of new transcripts found compared to those observed previously in studies on the oil palm flower structure (Adam et al., 2006; Adam et al., 2007a; Adam et al., 2007b). This diversification appears to have arisen through the expansion of the expression of genes that have either retained a function in the flower, or have specialized functions within the fruit. Indeed, several sequences identified in the present study, including those from the SEP-like and the GLO-like subfamilies, were identical at the nucleotide level to those observed in the oil palm flower and may be the same gene. Finally, the expression of certain MADs-box transcripts either positively or negatively coincides with the burst of ethylene production that occurs by 140 and 160 DAP. This raises the question of the relationship between ethylene production and related transcriptional activity and whether these MADs-box gene products play regulatory roles either upstream or downstream of the ethylene burst during mesocarp ripening.

The last acylating step of TAG assembly may involve several complementary routes
One of the most notable findings from the reconstruction of the oil synthesis pathways in the oil palm mesocarp is the possible contribution of three enzymes for the acylation of DAG to TAG, *i.e.* both type 1 and type 2 DGAT and PDAT. Although the current literature shows that the three enzymes, encoded by three separate gene families, are all capable of catalyzing the final acylation step, their individual contribution in plant tissues that accumulate oil remains poorly understood. The similar transcript amounts detected in the present study suggest that the contribution of these three enzymes could be of similar importance in the mesocarp of oil palm. Based on transcription patterns of DGAT1, DGAT2 and PDAT in developing seeds of various plants, it was suggested that DGAT2 and PDAT could be major contributors of TAG synthesis in seeds that store large amounts of epoxy and hydroxy FA (Li et al., 2010a). However, a recent study also indicates high transcription of both DGAT1 and DGAT2 in the mesocarp of olive during the stage of oil synthesis, while olive oil does not contain any unusual FA (Banilas et al., 2011). In that study, the authors hypothesize that DGAT2 could play a specific role during oil droplet enlargement that occurs at late stages of olive mesocarp maturation. Furthermore, it was recently demonstrated that both DGAT1 and PDAT1 contribute to TAG synthesis in the developing Arabidopsis embryo (Zhang et al., 2009b). A major role for PDAT in the developing oil palm mesocarp is substantiated by the identification of a contig very similar to its necessary partner – LPCAT – whose transcript levels were comparable to those of other transcripts involved in TAG assembly.

**Divergence of the FA synthesis and TAG assembly Transcriptional Regulation**

The second major result drawn from the analysis of the lipid-related transcriptome is that the core FA synthetic machinery of the oil palm mesocarp is remarkably coordinated at the transcriptional level (Fig. 2; Fig. 6). Large-scale transcriptomic studies focused on oil synthesis in plants remain scarce. However, since the same key result was identified in the developing embryo of Arabidopsis (Ruuska et al., 2002; Baud and Lepiniec, 2009) and the persistent living endosperm of albuminous coffee seeds (Joët et al., 2009), it is tempting to speculate that this feature is common to plant oil storing tissues, independent of their origin. In all three cases, a sharp increase in plastidial gene transcription occurs at the onset of lipid accumulation. Similarly to that observed in the Arabidopsis embryo and the coffee endosperm, ER genes involved in TAG assembly displayed an expression profile different from that of FA synthesis genes, suggesting distinct regulation programs. However, while the fold increase of most genes involved in TAG assembly was higher than those of the core FA
biosynthetic machinery in the Arabidopsis embryo (Baud and Lepiniec, 2009), transcript levels observed in the oil palm mesocarp for genes responsible of TAG assembly remained low and did not show a clear increase concomitant with oil deposition, such as that observed for FA synthesis genes. Since no other tissue accumulates more oil than the oil palm mesocarp, the present work clearly indicates that a marked up-regulation of TAG assembly gene transcription is not strictly required for oil accumulation and that the regulation of this pathway may differ considerably among plants and tissues. Based on the present transcriptomic data, one can therefore hypothesize that the metabolic control of oil synthesis in the oil palm mesocarp is mostly controlled by de novo FA synthesis. This hypothesis corroborates previous studies (Ramli et al., 2002; Ramli et al., 2009) that quantified flux control coefficients of the two blocks –FA synthesis and TAG assembly– in callus cultures of olive and oil palm using the technique of metabolic control analysis.

**Phytoene synthase and phytoene desaturase are the key players for carotenoid accumulation**

Our results clearly revealed a coordinated transcriptional activation of the carotenogenic genes involved in the synthesis of phytoene and subsequent desaturation steps leading to lycopene, while further steps involved in cyclization of lycopene towards carotenones and upstream isoprenoid synthesis appear to be poorly or not regulated at the transcriptional level. Despite that the oil palm mesocarp stores large amounts of carotenones but not lycopene, the transcriptional regulation of carotenogenesis nevertheless resembles that of the lycopene-rich tomato fruit, i.e. through the transcriptional activation of the first two committed steps of carotenoid synthesis (Ronen et al., 1999). This important finding suggests that the control of carotenoid metabolism in oil palm may first operate by directing the metabolic flux into the carotenoid pathway. This is consistent with previous reports which showed that PSY expression was the rate-limiting step for carotenogenesis in various plant models (Fraser et al., 1994; Shewmaker et al., 1999). In addition, the constitutive over-expression of PSY was found to be sufficient to trigger the accumulation and sequestration of carotenones into crystals in plastids from non-green tissues of Arabidopsis (Maass et al., 2009). Considering the importance of PSY and PDS upregulation for carotenogenesis in the oil palm mesocarp and that class VII ERF/RAP2.2 transcription factors are known to mediate PSY and PDS expression (Welsch et al., 2007), we can speculate that the contigs encoding ERF/RAP2.2 found to be upregulated at 140 DAP participate to the massive carotene accumulation in this fruit. Finally the concomitant upregulation of the Or, DDB1 and DET1 regulators suggest that
carotenoid accumulation may also be controlled by enhancing the sink strength, i.e. the formation of carotenoids sequestration structures during the chloroplast to chromoplast transition. Indeed, the Or gene was previously shown to cause high levels of β-carotene accumulation in an orange cauliflower mutant by enhancing sink strength (Lu et al., 2006). In addition, expression of the Or transgene in potato produces deep orange yellow-flesh tubers with an over sixfold increase of total carotenoid content. This increase in the Or transgenic tubers was found to be associated with the formation of chromoplasts and specific carotenoid sequestration structures, which serve as an effective metabolic sink to facilitate the sequestration and storage of carotenoids (Lopez et al., 2008).

The low abundance of lycopene cyclase transcripts (LCY) was at first quite puzzling since the oil palm mesocarp does not accumulate lycopene but carotenes. However, this discrepancy can be interpreted in two ways. First, it clearly evokes the low level of transcription of the whole TAG assembly pathway while the oil palm mesocarp is exceptionally efficient regarding TAG synthesis. One of the major lessons drawn from the combined analysis of both oil and carotenoid transcriptomes is that the coordinated upregulation of all steps of a given pathway is not a prerequisite for the activity of the corresponding metabolism. Besides, the amount of carotenoids synthesized may also depend on the properties of the encoded enzymes and therefore rely on the polymorphism observed at the sequence level. Indeed, a recent study carried out in cassava reported that a single nucleotide polymorphism in PSY2, leading to a non-conservative amino acid exchange, results in a dramatic increase of carotenoid accumulation in storage roots (Welsch et al., 2010). This polymorphism occurs in a region that is usually highly conserved among plants (ALDRWE amino acid sequence, Supplementary Fig. S14) and the A191D exchange leads to an increased PSY enzymatic activity in cassava. Notably, this conserved region is also affected in E. guineensis with an amino acid exchange in the vicinity of the Arginine (L198M exchange, Supplementary Fig. S14). This discovery opens up interesting prospects and a more in-depth characterization of the EgPSY enzyme should allow to test whether the exceptional carotene content of the oil palm mesocarp is also influenced by the peculiar amino acid sequence polymorphism in this region.

Towards the understanding of FA synthesis regulation in the oil palm mesocarp

The fact that most genes involved in the core FA biosynthetic machinery share the same temporal transcription pattern strongly suggests that they are co-regulated and probably share common regulatory elements. On the basis of recent literature on the control of oil
biosynthesis in seeds (Baud and Lepiniec, 2010) we hypothesize that an AP2 transcription factor similar to WRI1 could play this key role in the oil palm mesocarp (Fig. 6). Indeed, we not only found a highly expressed WRII-like contig, named EgAP2-2, which was perfectly co-regulated with FA biosynthetic genes, but also identified the same AW-box sequence in several genes belonging to the core FA biosynthetic machinery, which confirms that this cis element considered as the direct binding site for WRI1 (Maeo et al., 2009) is highly conserved among monocots and eudicots (Pouvreau et al. 2011). To our knowledge, the present work reveals for the first time that the transcriptional regulation of FA biosynthesis via WRII may not be restrained to seeds but may also operate in other tissues that accumulate lipids, such as the oil palm mesocarp.

These results raise the critical question whether EgAP2-2 expression in the mesocarp is regulated by master regulators analogous to those that function during seed maturation (Baud and Lepiniec, 2010). The absence of any LEC2-like transcript in the oil palm mesocarp transcriptome is not surprising since LEC2 presumably exists only in dicot genomes (Li et al., 2010b). In contrast, since LEC1 was recently proposed to regulate WRII in maize (Shen et al., 2010), it constituted a viable candidate for EgAP2-2 regulation. However, given that LEC1 is often considered specific to seed tissues, the regulation of EgAP2-2 in the mesocarp may be completely different. Nonetheless, since LEC1 was recently shown to be expressed in fern leaves in response to dehydration (Xie et al., 2008), its participation to oil synthesis regulation in the mesocarp of a monocot species could not be excluded. However, we did not identify any contig showing high homology with LEC1-type HAP3, suggesting a different regulatory cascade controls WRII expression in the oil palm mesocarp.

MATERIAL AND METHODS

Plant material production, histology and RNA preparation for 454 sequencing

Oil palm fruits were harvested at Pobè CRA-PP Station (INRAB) Benin, from a dura parent of Deli Dabou origin, within the same self-progeny of a single palm. In this progeny, after two successive self-pollinations, the percentage of heterozygozity is estimated at only 20-25 % (Cochard et al., 2009). For each stage of development studied, three independent bunches were collected on three distinct individuals of the same genotype. Fifteen spikelets were then collected in the centre of each bunch and 5 spikelets were randomly sampled from them. From all undamaged fruits withdrawn from each of the 5 spikelets, 3 fruits were then
randomly sampled. Then, for each stage × bunch × spikelet combination retained, mesocarp, shell and seed of each fruit were separated and weighed. Dry weight from all samples was measured after one night at 103°C. Variation in tissue weight across development was tested using 3-way ANOVA and post hoc Newman-Keuls test.

For histological analysis, samples were fixed in PFA 4% and PBS 1X in the presence of EDC 4% for 16h. Thereafter, samples were washed twice for 15 min in PBS 1X and glycine, then twice for 15 min in PBS 1X. Finally samples were dehydrated in ethanol 50%, 1 hour, and then 70° before embedding in Technovit 7100 resin (Heraeus Kulzer GmbH, Wehrhrim/Ts, Germany). Semi-thin sections of mesocarp (4 µm) were stained with Toluidine Blue (TB), which stains acid compounds such as pectin in blue in acid environment. To visualize lipids in the mesocarp, fixed samples were sectioned (100 µ) using vibratome, washed with PBS 1X, then stained with Nile red and analyzed by confocal microscopy (Zeiss LM510), with a laser excitation at 458 nm and a HFT filter 548/514 using the 40x/1.2 numeric aperture (NA). To visualize β-carotenes in mesocarp cells, fruit were sectioned (150 µ), after 16h in fixation buffer, using a vibratome, stored in PBS 1X, then observed by confocal microscope (Zeiss LM510 meta) with a laser excitation at 488 nm and an emission at 580-600 nm corresponding to β-carotenes emission spectra using the 63x/1.4 NA.

For RNA extractions, mesocarp tissue samples were collected and frozen immediately in liquid nitrogen. Total RNA from mesocarp was extracted as previously described (Morcillo et al., 2006). Using titanium kit (Roche), cDNA related to four stage of development (100, 120, 140 and 160 dap) were tagged independently then mixed together to one sample for 454 pyrosequencing carried out by GATC Biotech AG, Germany.

**Hormonal analysis**

For hormonal profiling analysis, mesocarp was collected from fruit aged between 80 to 160 DAP, lyophilized then grounded before to be sent (3 x 50 mg of dry powder) to NRC-CNRC (Canada). ABA and ABA metabolites, cytokinins, auxins, and gibberellins were quantified by UPLC ESI-MS/MS has been described in detail by (Chiwocha et al., 2003). For measure of ethylene release, spikelet consisted of 10 fruits are put in hermetic jars and ethylene production quantify after 24h by CPG (FID system, equipped with GSQ column). Values are means of duplicate.

**Lipid and carotenoid determination analysis**
Total lipids were extracted from 1g samples of freeze-dried powder using a modified Folch method as described previously (Laffargue et al., 2007). Carotenoid extraction was adapted from the method described previously (Achir et al., 2010), which includes a preliminary triacylglycerol removal step. Briefly, oil samples (100 mg) were carefully mixed with 2 mL acetone containing 15mg.L\(^{-1}\) echinenone as internal standard and left 30 min at – 20°C, leading to triacylglycerol crystallization. Triacylglycerols were separated by rapid sampling of the upper acetone phase that contains carotenoids and filtered through a 0.45-mm PTFE filter (Whatman, USA). The carotenoid extract was directly injected into the HPLC column. The column was a polymeric YMC-30 (4.6 mm id× 250 mm, 5 mm particle size) (YMC, Wilmington, NC, USA). Elution was performed using five successive steps at a flow rate of 1 ml/min: i) 100% A for 15 min, ii) linear gradient going from 0% B to 20% B for 10 min, iii) linear gradient going from 20% B to 90% B for 10 min, iv) linear gradient from 90% B to 100% B for 5 min, v) 100% B for 5 min (A= methanol:water (60:40, v/v) and B=MTBE:methanol:water (68:28:4, v/v/v). A UV-visible photodiode array detector (Dionex UVD 340U) was used to analyze the chromatograms at a detection wavelength of 450 nm, except for phytofluene (348 nm), phytoene (286 nm) and ξ-carotene (401 nm). Carotenoids were identified by the combined use of their relative retention times and their wavelength absorption maxima as determined using analytical standards (Carotenature, Lupsingen, Switzerland). Quantification was done combining the standard calibration curves and correction with internal standard peak areas. All analyzes were realized in triplicate (from 3 independent mesocarp samples).

454 sequences analysis, de novo assemblage and contig annotation

The 454 reads obtained from the 454 sequencing data were analysed using a specialized version of our tool ESTtik (Expressed Sequence Tag Treatment and investigation kit) (Argout et al., 2008) dedicated to the analysis of 454 data. To avoid problems of miss-assembly, we discarded sequences inferior to 120bp and low quality sequences. We then used the Megablast algorithm to detect and eliminate non coding sequences by comparing the whole reads against fRNAdb (Mituyama et al., 2009) with an Evalue cutoff of 1e-20. To avoid chimeric consensus sequences, which contain parts of different expressed genes, we modified the parameters of the TGICL (Pertea et al., 2003) assembler to obtain good quality contigs. We used a minimum overlap percent identity cutoff of 94 % and an overlap length cutoff of 60 bp for the clustering procedure of TGICL and a minimum overlap percent identity cutoff of 95% and an overlap length cutoff of 60 bp for the assembly. Finally, we predicted peptide sequences from the
transcriptomic data using a modified version of the prot4EST pipeline (Wasmuth and Blaxter, 2004). The peptide sequences were annotated using the InterProScan web service (Zdobnov and Apweiler, 2001) to find protein domains and functional sites by comparing the sequences against the InterPro signature database (Hunter et al., 2009). Contigs were then annotated using the standalone version 2.2.16 of Blast software (Altschul et al., 1990). Similar sequences were searched within the TIGR6 proteome of Oryza sativa japonica, the Uniprot Knowledgebase Swiss-Prot and TrEMBL (Boeckmann et al., 2003), the non-redundant protein sequences database (NR) from Genbank and the nucleotide sequences from the Genbank nucleotide collection NT. To obtain significant results for the whole Blast analyses, we used an Evalue cutoff of 1e-5 and the 10 best hits of each blast were retained for the annotation. In addition, we used the Blast2GO software (Gotz et al., 2008) to retrieve Gene Ontology terms based on our Blast Analyses.

**Statistical analysis of read abundance profiles**

A contig was considered differentially expressed during development when at least one of the 3 stage transitions (100-120, 120-140 and 140-160 DAP) exhibited a highly significant difference in read abundance at $P=0.01$. Difference in read abundance between two stages of development was tested using Audic-Claverie statistics (Audic and Claverie, 1997). $P$-values obtained by the AC test were then transformed using the Bonferroni correction to control false discovery rates. Hierarchical Clustering Analysis (HCA) was used to group contigs according to their transcription profile, using the tool developed by (Eisen et al., 1998) (http://rana.lbl.gov/eisen/).

**Phylogenetic Analysis**

Phylogenetic trees were constructed based on similarity searches performed with BLASTp programs with default parameters in protein sequence databases provided by the NCBI server (http://www.ncbi.nlm.nih.gov). Phylogenetic analyses were performed on the Phylogeny.fr plateform (http://www.phylogeny.fr) (Dereeper et al., 2008). Amino acid sequences were aligned with ClustalW (v2.0.3) (Thompson et al., 1994). After alignment, ambiguous regions (i.e. containing gaps and/or poorly aligned) were removed with Gblocks (v0.91b). The phylogenetic tree was reconstructed using the neighbor joining method implemented in Neighbor from the PHYLIP package (v3.66) (Saitou and Nei, 1987). Distances were calculated using ProtDist. The JTT substitution model was selected for the analysis. The robustness of the nodes was assessed by bootstrap proportion analysis.
(Felsenstein, 1985) computed from 100 replicates. Graphical representation and edition of the phylogenetic tree were performed with TreeDyn (v198.3).

**Supplementary Materials**

**Supplementary Figure S1**
- Complete data set of auxin, GA and cytokinin, ABA and ethylene metabolite analysis.

**Supplementary Figure S2**
- Overview of 454 pyrosequencing data.

**Supplementary Figure S3**
- FASTA list of assembled contigs.

**Supplementary Figure S4**
- Hierarchical Cluster Analysis (HCA) of group I contigs differentially represented.

**Supplementary Table S5**
- Gene Ontology enrichment analysis between developmental stages of group I contigs.

**Supplementary Table S6**
- Annotations of contigs discussed in text.

**Supplementary Figure S7**
- Phylogenetic analysis of EgAP2-2, a transcription factor similar to WRINKLED1

**Supplementary Figure S8**
- Phylogenetic analysis of the AP2 domains from the ERFs identified in the mesocarp.

**Supplementary Figure S9**
- Phylogenetic analysis of the MADs domains from mesocarp MADs-box proteins

**Supplementary Table S10**
- Transcription factors and regulators expressed in the mesocarp Cluster A

**Supplementary Table S11**
- Transcription factors and regulators expressed in the mesocarp Cluster B

**Supplementary Table S12**
- Transcription factors and regulators expressed in the mesocarp Cluster C

**Supplementary Table S13**
- Transcription factors and regulators expressed in the mesocarp Cluster D

**Supplementary Figure S14**
Amino acid alignments of PSYs conserved domains of oil palm contig (EgPSY) and others from *Manihot esculenta*, *Narcissus pseudonarcissus*, *Oryza sativa*, *Solanum lycopersicum*; *Zea mays*.

**ACKNOWLEDGEMENTS**

We thank Myriam Collin and Jean-Luc Verdeil for their excellent assistance with the histological analysis and James W. Tregear for his support and useful discussions during the preparation of this article. We also thank Nathalie Maréchal and Sylvie Avallone for assistance with the biochemical analyses. We would also like to especially thank all the personnel at INRAB, CRA-PP, Pobé, Benin for their technical and logistical support during the collection of the genetic material used in this study.

**LITERATURE CITED**


Baud S, Mendoza MS, To A, Harsoeet E, Lepiniec L, 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


Boss PK, Vivier M, Matsumoto S, Dry IB, Thomas MR (2001) A cDNA from grapevine (Vitis vinifera L.), which shows homology to AGAMOUS and SHATTERPROOF, is not only expressed in flowers but also throughout berry development. Plant Mol Biol 45: 541-553


A SQUAMOSA MADS box gene involved in the regulation of anthocyanin accumulation in bilberry fruits. Plant Physiol 153: 1619-1629


Li R, Yu K, Hildebrand DF (2010a) DGAT1, DGAT2 and PDAT expression in seeds and other tissues of epoxy and hydroxy fatty acid accumulating plants. Lipids 45: 145-157


Zdobnov EM, Apweiler R (2001) InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847-848


Zhang M, Fan J, Taylor DC, Ohlrogge JB (2009b) DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. Plant Cell 21: 3885-3901

Figure A: Fruit mass (g) over DAP.

Figure B: Fruit size (mm) over DAP.

Figure C: Mesocarp mass (g) over DAP.

Figure D: Hormones (ng g\(^{-1}\) DM) over DAP. ABA (ng g\(^{-1}\) DM) indicated.

Figure E: Total mesocarp lipids (g) over DAP.

Figure F: Image of mesocarp cells at 30 DAP.

Figure G: Image of mesocarp cells at 120 DAP.

Figure H: Image of mesocarp cells at 160 DAP.
Figure 1. Morphological, biochemical and histological analysis of *E. guineensis* fruit development from 30 to 160 days after pollination (DAP). A, Evolution of fresh (○) and dry (●) mass of whole fruits. B, Evolution of fruit length (○) and diameter (●). C, Evolution of fresh (○) and dry (●) mass of mesocarp. D, Quantification of auxin (IAA, □), cytokinin (trans Zeatin-O-glucoside, ○), gibberelin (GA3, ●) and abscisic acid (ABA, ▲). E, lipid (●), α-carotene (▲) and β-carotene (◇) content. Cellular characteristics of 30 DAP, F, 70 DAP, G, 160 DAP mesocarp. Lower left insets in G and H, confocal microscopy of mesocarp lipids detected by nile red staining. Lower right inset H, confocal microscopy of mesocarp β-carotene localization. Phases of development I, II, III, IV and V. Lipid and carotenoid content values expressed as total g mesocarp lipid or mg carotenoid on a per fruit basis. Arrow indicates cell separation and enlarged intercellular space. Scale bars, 50 µ. Scale bars inset 5µ. Differences in the lower case letters above data points indicate statistically significant differences.
Figure 2. Fatty acid synthesis and triacylglycerol assembly in the oil palm mesocarp as reconstructed from the 454 pyrosequencing transcriptome. Contig names followed by an asterisk (*) indicate contigs showing a significant variation during development according to Audic-Claverie and FDR statistics (contig Group I). Gene expression levels at 100, 120, 140 and 160 DAP are indicated in coloured bars. For the developmental stage displaying maximal expression level, the normalized transcript abundance, expressed as the number of transcripts per 200,000 transcripts, is given. For the other stages, expression levels are indicated as percentages of the maximal normalized transcript abundance of the gene, as described in the colour caption from 0% (white) to 100% (dark blue). ACC BC, Biotin Carboxylase; subunit of Heteromeric ACCase; ACC Cτα, Carboxyltransferase alpha Subunit of Heteromeric ACCase; ACC Cτβ, Carboxyltransferase beta Subunit of Heteromeric ACCase; CPT, Diacylglycerol Cholinephosphotransferase; DGAT, Acyl-CoA : Diacylglycerol Acyltransferase; EAR, Enoyl-ACP Reductase; ENOp, Enolase; FAD2, Oleate Desaturase; FAD3, Linoleate Desaturase; FATA, Acyl-ACP Thioesterase A; FATB, Acyl-ACP Thioesterase B; GPAT, Glycerol-3-Phosphate Acyltransferase; HAD, Hydroxyacyl-ACP Dehydrase; KAR, Ketoacyl-ACP Reductase; KAS I, Ketoacyl-ACP Synthase I; KAS II, Ketoacyl-ACP Synthase II; KASIII, Ketoacyl-ACP Synthase III; LACS, Long-Chain Acyl-CoA Synthetase; LPAAT, 1-Acylglycerol-3-Phosphate Acyltransferase; LPCAT, 1-Acylglycerol-3-Phosphocholine Acyltransferase; MAT, Malonyl-CoA : ACP Malonyltransferase; PAP, Phosphatidate Phosphatase; PDAT, Phospholipid : Diacylglycerol Acyltransferase; PDH, Dihydrolipoamide Acetyltransferase, E2 component of Pyruvate Dehydrogenase Complex; PK, Pyruvate kinase; SAD, Stearoyl-ACP Desaturase; WRI1, Wrinkled1.
**Figure 3.** Carotenoid and abscisic acid biosynthetic pathways with related metabolites and transcripts present during mesocarp development. Metabolite profiles are presented in graph insets and relative expression is as in Figure 2. ABA-GE, abscisic acid glucose ester; ABAGT, abscisic acid glucosyltransferase; ABAO, abscisic-aldehyde oxidase; ABA8ox, abscisic acid 8'-hydroxylase; CRTISO, carotenoid isomerase; CRTR-B, β-carotene hydroxylase; CRTR-E, ε-carotene hydroxylase; DPA, dihydrophaseic acid; FPS, farnesyl pyrophosphate synthase; GPPS, geranyl diphosphate synthase; GGPPS, geranyl geranyl pyrophosphate synthase; LCY-B, lycopene β-cyclase; LCY-E, lycopene ε-cyclase; MEP, methyl-erythritol phosphate; NCED, 9-cis-epoxycarotenoid dioxygenase; NSY, neoxanthin synthase; NXS, neoxanthin synthase; PA, phaseic acid; PDS, phytoene desaturase; PSY, phytoene synthase; PTOX, plastid terminal oxidase; VDE, Violaxanthin De-Epoxidase; XO, xanthosin dehydrogenase; ZDS, ζ-carotene desaturase; ZEP, zeaxanthin epoxidase; Z-ISO, 15-cis-ζ-carotene isomerase.
Figure 4. Ethylene production and transcript profiles of ethylene biosynthesis and core response pathway components in the oil palm mesocarp. A, Relative amounts of ethylene produced by the mesocarp from 120 to 160 DAP and transcript profiles for the methionine and ethylene biosynthetic enzymes S-adenosyl methionine (S-AdoMet, SAM) synthetase,, ACC (1-aminocyclopropane-1 carboxylic acid) synthase (ACS), ACC oxidase (ACO), aspartate aminotransferase-like (AAT-like) protein. B, transcripts for the ethylene signaling proteins ETR, ETHYLENE RESPONSE; EIN, ETHYLENE INSENSITIVE; ERF, ETHYLENE RESPONSE FACTOR; RAP, ERF/AP2 transcription factor; DREB DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN; CTR, CONSTITUTIVE TRIPLE RESPONSE. Oil palm type VII ERFs highlighted in grey. Relative expression is as in Figure 2.
Figure 5. Phylogenetic analysis of oil palm MADS with other MADS genes. A, Phylogenetic analysis of CL1Contig3848, CL1Contig8010, CL1Contig6409, CL1Contig2367 and CL1295Contig1 with other SEP clade protein sequences using NJ method based on the multiple alignment of the partial protein MIK sequences (139 residues). B, Phylogenetic analysis of CL1Contig5719, CL1Contig1226 and CL1Contig5512 with other AG clade protein sequences using NJ method based on the multiple alignment of the full-length MIKC protein sequences. ANR1 from Arabidopsis was used as root. Numbers on the branches are the bootstrap values for 100 replicates. The sequences included in this alignment are from Arabidopsis, maize (Zea mays), rice (Oryza sativa), oil palm (Elaeis guineensis), snapdragon (Antirrhinum majus) and from proteins in selected fruit species as apple (Malus domestica), banana (Musa acuminata), grape (Vitis vinifera), peach (Prunus persica), strawberry (Fragaria x ananassa) and tomato (Lycopersicon esculentum). The MADS box proteins from our study are in grey. Relative expression is as in Figure 2.
Figure 6. Scheme representing the major events that occur during the phases of mesocarp development. See text for details.