Correction

PLANT BIOLOGY

The authors note the following: “A relevant reference by Tranbarger et al. that describes the regulation of oil palm fruit ripening should be added to the list of references in our article. In addition to providing data that complement our study, Tranbarger et al. reached similar conclusions with regard to the importance of palm orthologs of the Arabidopsis WRINKLED1 transcription factor.”

Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning

Fabienne Bourgis 1,2,3, Aruna Kilaru 1,2, Xia Cao 4, Georges-Frank Ngando-Ebongue 5, Noureddine Drira 1, John B. Ohlrogge 6, and Vincent Arondeau 1,2,3

1 Université de Bordeaux Ségalen, Laboratoire de Biogenèse Membranaire, Unité Mixte de Recherche 5200, F-33000 Bordeaux, France; 2 Centre National de la Recherche Scientifique, Laboratoire de Biogenèse Membranaire, Unité Mixte de Recherche 5200, F-33000 Bordeaux, France; 3 Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI 48824; 4 Department of Plant Biology, Michigan State University, East Lansing, MI 48824; 5 Centre de Recherches sur le Palmier à Huile de la Dibamba, Institut de Recherche Agricole pour le Développement, BP243 Douala, Cameroon; 6 Laboratoire de Biotechnologie Végétale, Faculté des Sciences de Sfax, Sfax 3038, Tunisia

Oil palm can accumulate up to 90% oil in its mesocarp, the highest level observed in the plant kingdom. In contrast, the closely related date palm accumulates almost exclusively sugars. To gain insight into the mechanisms that lead to such an extreme difference in carbon partitioning, the transcriptome and metabolite content of oil palm and date palm were compared during mesocarp development.

Compared with date palm, the high oil content in oil palm was associated with much higher transcript levels for all fatty acid synthesis enzymes, specific plastid transporters, and key enzymes of plastidial carbon metabolism, including phosphofructokinase, pyruvate kinase, and pyruvate dehydrogenase. Transcripts representing an ortholog of the WRI1 transcription factor were 57-fold higher in oil palm relative to date palm and displayed a temporal pattern similar to its target genes. Unexpectedly, despite more than a 100-fold difference in flux to lipids, most enzymes of triacylglycerol assembly were expressed at similar levels in oil palm and date palm. Similarly, transcript levels for all but one cytosolic enzyme of glycolysis were comparable in both species. Together, these data point to synthesis of fatty acids and supply of pyruvate in the plastid, rather than acyl assembly into triacylglycerol, as a major control over the storage of oil in the mesocarp of oil palm. In addition to greatly increasing molecular resources devoted to oil palm and date palm, the combination of temporal and comparative studies illustrates how deep sequencing can provide insights into gene expression patterns of two species that lack genome sequence information.

triacylglycerol biosynthesis | *Elaeis guineensis* | fruit ripening | *Phoenix dactylifera*

Oil palm (*Elaeis guineensis Jacq.*) originates from intertropical Africa. It was imported into South Asia where industrial plantations started about 100 years ago. Oil palm is now the most productive world oil crop (3.5 tons/ha/yr), with 36% of world production (1). Because present-day genetic material can produce up to 10 tons of oil/ha/yr, it is likely that palm oil will keep increasing its share of the market. Most palm oil is derived from the fruit mesocarp where it can comprise up to 90% of the dry weight. This is, by far, the highest oil content reported for any plant tissue. Despite its obvious scientific and economic interest, literature and molecular resources available for oil palm remain scarce (2).

Presently, knowledge of triacylglycerol (TAG) accumulation in plants is based almost entirely on studies of oil seeds, which contain no more than 60% oil (3). The main source of carbon for storage oil synthesis in higher plants is sucrose, which in non-green tissues is converted to pyruvate via glycolysis and the pentose phosphate pathway (PPP). Pyruvate is the main precursor for the acetyl-CoA molecules destined to fatty acid synthesis. Plastid pyruvate kinase (PK), pyruvate dehydrogenase (PDH), and acetyl-CoA carboxylase are considered as key enzymes for fatty acid synthesis (3) and in oil seeds are regulated by transcription factors, including WRINKLED1 (WR1) (4, 5).

Fatty acids synthesized from acetyl-CoA in the plastid are exported as acyl-CoA esters to the endoplasmic reticulum where they enter glycerolipid metabolism. Reactions generating fatty acids, phosphatidic acid (PA), and diacylglycerol (DAG) are common to both membrane phospholipid and TAG biosynthetic pathways. Only the last acylation step, conversion of DAG into TAG, is specific to oil synthesis. Oil stored in seeds is used to fuel postgerminative growth of seedlings; however, mesocarp oil provides an attractant for disseminating animals and does not undergo further plant metabolism.

To identify features specific to high oil production in nonseed tissues such as oil palm mesocarp, we generated several million ESTs for five developing stages of oil palm mesocarp. To complement and strengthen our study, we carried out similar experiments in date palm (*Phoenix dactylifera*), a closely related species (6) that stores almost exclusively sugars rather than oil (Fig. 1A). The 100-fold difference in total fatty acid content in their mesocarp is most striking (Fig. 1B). The focus of this temporal and comparative analysis was to gain insight into factors responsible for this dramatic difference in carbon partitioning. In addition, a better understanding of oil accumulation in fruits may present strategies for engineering oil accumulation in other vegetative tissues.

Results and Discussion

As an initial step toward understanding how photosynthate is directed into very different pathways and end products, pyrosequencing ESTs were generated from mesocarp of oil palm and date palm fruits and oil palm leaves. To aid interpretation of the palm transcriptomes, we also conducted metabolite analyses, supplemented with transmission electron microscopy. Aside from lipid and terpenoid compositions, mesocarp metabolites and subcellular organization of oil storage during oil palm ripening are largely unknown.
Metabolite Contents Differed Markedly Between Palm Mesocarps.

The total fatty acid content in oil palm mesocarp rose from 2% to 88% (dry weight) during ripening whereas it remained below 1% in date palm mesocarp (Fig. 1C), indicating a 100-fold difference in total fatty acid content. The rate of oil accumulation was 1.6 μmol fatty acid/h/g fresh weight (FW), which is comparable to rapeseed (1.7 μmol fatty acid/h/g FW). Although TAGs accounted for more than 95% of total fatty acids in ripe oil palm, they were undetectable (<0.002% dry weight) in date palm, indicating that storage of fatty acids as TAG differed by more than 1,000-fold. Analysis of fatty acid composition in polar and neutral lipids of oil palm mesocarp indicated less palmitate and stearate in oil palm whereas sugar content was 14% and decreased to 1.3% at 23 WAP (Dataset S1B). This indicated that uptake of sugars, presumably destined to oil synthesis, started before major oil accumulation.

NMR metabolite analysis revealed that organic acids (mostly citrate) of oil palm declined from 3.3% to 0.2%, a trend similar to that of sugars (Dataset S1B). In contrast, during date palm mesocarp development, levels of citrate were 20-fold below that of oil palm, and organic acid content remained nearly constant (1.1%). Amino acids were 7- to 10-fold higher in oil palm than in date palm at early stages, and their composition was markedly different in the two species. Starch content increased in the last ripening stages of oil palm to represent more than 1% of dry weight, whereas it was almost undetectable in date palm. The analyses of these major metabolites indicate that sugars represent the major source of carbon available for oil synthesis.

Electron microscopy of 20 WAP oil palm mesocarp showed most notably, cells with numerous oil droplets, which, in contrast to oil seeds varied greatly in size (0.2–4 μm diameter; Fig. 1D). A few cells were almost completely filled by a single oil droplet (Fig. S1A). Most plastids showed a chloroplast-like structure (Fig. S1B), and a few amyloplasts were visible in ripe fruits (Fig. S1C).

Oil Palm and Date Palm Transcript Sequences Were Highly Similar and Closely Matched with Arabidopsis Orthologs. For this study, about 4 and 2 million ESTs, with an average read length of 389 and 362 nt, were obtained for oil palm and date palm mesocarp, respectively (Dataset S1C). After oil palm EST assembly, 95.9% of contigs with ≥50 ESTs were highly similar (BLASTX E-value < 10^-15) to Arabidopsis proteins. Comparable results were obtained for date palm (Dataset S1C). Contigs were annotated on the basis of Arabidopsis proteins because it is by far the best-annotated proteome among the plant kingdom. ESTs were assigned to the most closely related Arabidopsis protein with the understanding that palms may have gene families with different complexity that may not be reflected in these assignments. Genes most relevant to this study are listed in Dataset S2A and B, with expression data for each stage of development. Oil palm and date palm nucleotide sequences were highly conserved (92% identity). This high identity is further discussed in relation to palm phylogeny in SI Materials and Methods.

Analysis of Lipid-Related Genes: Transcripts for All Fatty Acid Synthesis Enzymes Showed Up-Regulation Patterns. Annotations of over 600 genes of acyl-lipid metabolism in Arabidopsis were recently updated (7). Orthologs of about 450 of these genes were expressed in oil palm mesocarp during fruit ripening, and these were categorized on the basis of their biochemical pathway and subcellular localization (Dataset S2A). Unexpectedly, for almost every pathway of lipid metabolism, EST levels were similar in oil palm and date palm (Fig. 2A) and did not show temporal changes during palm ripening (Fig. 2B). Only the fatty acid synthesis subcategory displayed major differences between oil palm and date palm together with distinct temporal increases during oil palm ripening.

EST levels for 18 plastidial proteins that are involved in conversion of pyruvate to fatty acids were, on average, 13-fold higher in oil palm than in date palm (Fig. 2A and Fig. 3). The largest individual differences were noted for ketoacyl-acyl carrier protein (ACP) reductase, ketoacyl-ACP dehydratase, and the E3 and E1α subunits of PDH, for which the ESTs were more than 50-fold higher in oil palm relative to date palm (Dataset S2A).

In addition to the high oil-palm-to-date-palm ratio, ESTs for fatty acid synthesis from pyruvate increased on average 5.7 ± 1.3-fold between 15 and 21 WAP (Fig. 2B). These increases were within a narrow range and displayed a highly correlated temporal expression pattern (R² = 0.99; Fig. S2). These data indicate closely coordinated expression during oil synthesis in a nonseed tissue, a pattern also noted in Arabidopsis and other seeds (3).

It is clear from both the species comparisons and the temporal data that the very high oil content of oil palm mesocarp correlates with high and increasing levels of transcripts coding for all fatty acid synthesis proteins. Contrary to oil seeds (3), the levels...

Fig. 1. Oil palm and date palm fruits show a completely different carbon-partitioning pattern. (A) Phylogenic tree constructed with RBCL gene sequences. Bar indicates the percentage of divergence. (B) Open fruits and mesocarp composition (% dry weight) (SI Materials and Methods). (C) Fatty acid and soluble sugar content of mesocarp during fruit ripening. (D) Transmission electron micrograph of cell containing oil droplets from oil palm mesocarp, harvested 20 WAP. See also Fig. S1.

Fig. 2. Lipid subcategories except for fatty acid synthesis show similar expression pattern in oil palm and date palm. (A) The ratio of ESTs in oil palm versus date palm for each lipid category (calculated per enzyme to account for different number of enzymes per pathway). (B) Temporal profile of ESTs for lipid categories in developing mesocarp of oil palm. See Dataset S2A for the list of genes included in each category. Pt, plastid; TAG, triacylglycerol; ExPL, extraplastidial polar lipids; PL, polar lipids; Mt, mitochondria; SphL, sphingolipids; β-ox, β-oxidation.
of fatty acid synthesis transcripts remained high up to the end of oil accumulation in oil palm mesocarp.

**TAG Assembly Transcript Levels Were Similar in Oil and Date Palm and Remained Mostly Constant.** In sharp contrast to the EST patterns for plastidial fatty acid enzymes, ESTs for most enzymes involved in TAG assembly showed low or no up-regulation during ripening and no substantial difference between oil palm and date palm (Fig. 3 and Fig. S3). For example, the acylating enzymes G3P acyltransferase (GPAT9); LysoPA acyltransferases (LPAATs); and phospholipid:DAG acyltransferase (PDAT), an enzyme that catalyzes the transfer of an acyl group from PC to DAG, showed no up-regulation during oil palm ripening and a less than twofold difference between palm mesocarps (Fig. 3 and Dataset S24). Although patterns differed for the two isoforms of DAG acyltransferase (DGAT1 and -2) that catalyze the last (acyl-CoA-dependent) acylation step to TAG, the sum of their ESTs increased only 2-fold during ripening and were 2.5-fold higher in oil palm than in date palm (Fig. 3 and Fig. S3).

Taken together, our data implicate transcriptional regulation of fatty acid synthesis rather than of TAG assembly enzymes as a major factor associated with very high oil synthesis in oil palm mesocarp.

**PA Phosphatase Isoforms Were Distinct.** Although the overall EST levels of TAG assembly enzymes did not show substantial differences temporally or between oil and date palm, in some cases a specific isoform varied more (Fig. S3). Presently, it is unclear which isoforms of PA phosphatase (PAP) are involved in TAG biosynthesis (8). Orthologs of *Arabidopsis* PAH1 and PAH2, which are involved in phospholipid synthesis but apparently not in TAG assembly, showed much lower EST levels in oil palm than in date palm mesocarp. Therefore, more probable candidates for generating DAG destined to TAG are two other putative PAPs, LPP6 and LPP6 (9), which remain largely uncharacterized. LPP6 was expressed at similar levels in both palms and was not up-regulated during oil palm ripening (Dataset S24). However, oil palm ESTs for LPP6 were 13 times higher than in date palm, were not detected in leaves, and increased during oil palm ripening (Fig. S3). The EST levels of LPP6 were comparable to that of GPAT and LPAAT and showed a pattern that correlated with oil synthesis (Fig. S3). Thus, it is tempting to speculate that LPP6 might generate a DAG pool destined to oil synthesis in oil palm mesocarp.

**PC-Related Enzymes Showed Distinct Changes.** Although detailed understanding of the molecular mechanisms involved is still lacking, flux analyses and other studies have demonstrated the importance of PC metabolism and acyl editing in the process of TAG assembly (10–12). Indeed, ESTs coding for some enzymes possibly involved in these processes showed important variations (Fig. S4). For example, PC synthesis can proceed via either the methylation or nucleotide pathway (13). The methylation pathway was down-regulated in oil palm compared with date palm, whereas the nucleotide pathway was up-regulated. Interestingly, down-regulation of the methylation pathway was linked to higher oil content in yeast (14), liver (15), and *Brassica napus* transgenic lines overexpressing DGAT (16). Surprisingly, PC:DGAT phosphocholine transferase (PDCT), an enzyme that plays an important role in TAG composition of *Arabidopsis* (12), showed 17-fold less ESTs in oil palm than in date palm. The low EST levels of PDCT provide leads for future research on the interconnection of pathways involved in phospholipid and TAG synthesis and their possible importance for high oil synthesis.

**A Transcription Factor with High Similarity to WR11 Was Strongly Up-Regulated.** Of 784 transcription factors (with >40 ESTs), only 6 had EST levels at least 15 times higher in oil palm than in date palm and were up-regulated at least fivefold during oil palm ripening. One of these six transcription factors is an AP2-domain-containing transcription factor with a very high similarity to maize (*E*-value = 10^-59) (17, 18) and *Arabidopsis* WR11 (*E*-value = 10^-30; Fig. S5). The EST levels for WR11-like were 57-fold higher in oil palm mesocarp, relative to date palm, and also increased by 7.5-fold during ripening (Fig. 4). This suggests that a WR11 ortholog in oil palm might play a role similar to that in *Arabidopsis*. This hypothesis is also strongly supported by the higher oil palm EST levels for each of the 10 genes that are reported to be WR11-regulated (4, 5, 17). Furthermore, as with *Arabidopsis* seeds, the temporal profile of ESTs coding for WR11-like in oil palm was similar to that of its putative targets (Fig. 4). Our data provide a strong indication that a WR11 ortholog plays a major role in oil accumulation not only in seeds but also in nonseed vegetative tissues such as mesocarp. In *Arabidopsis*, WR11 is under the control of seed maturation master regulators such as LEAFY COTYLEDON1 and -2, FUSCA3 and ABSICIC ACID INSENSITIVE3 (19, 20). However, no obvious homologs to these genes were identified in oil palm mesocarp, suggesting that *Eq*WR11-like is likely to control oil synthesis independently of the upstream factors that participate in seed development and may involve a different regulatory network, possibly fruit-specific.

**Increased Transcripts for Specific Enzymes and Transporters That Provide Pyruvate for Fatty Acid Synthesis.** A major flux through glycolysis is expected to provide the large amounts of pyruvate required for high oil synthesis in oil palm. Each of the eight steps
of glycolysis occurs in both cytosol and plastid. Notably, only 6 glycolytic enzymes of 16 increased, either temporally during oil palm ripening or when comparing oil palm and date palm (Fig. 5). Only two of these six differed by both criteria. These were plastid isoforms of ATP-dependent phosphofructokinase (PFK) and PK, which were 6.3- and 4.1-fold higher than in date palm, respectively (Fig. 5). These two enzymes also increased 5- and 3.3-fold during ripening of oil palm (Fig. 5). Both PK and PFK are considered to catalyze key regulatory steps of glycolysis (21), whereas most other enzymes of glycolysis are estimated to be in large excess with regards to flux in B. napus embryos (22). Therefore, our results imply that an increase in expression of plastid glycolysis plays a major role in providing pyruvate for high rates of fatty acid synthesis.

Unlike the plastid isoforms, none of the ESTs for cytosolic glycolysis showed increases during ripening of oil palm mesocarp (Fig. 5). In addition, EST levels for cytosolic glycolysis were similar in both palms, except for PFK (Fig. 5). The ESTs coding for cytoplasmic PFK, which is an ATP-dependent enzyme, were 3.4 times higher in oil palm than in date palm mesocarp whereas pyrophosphate-dependent phosphofructokinase (PFP) EST levels remained unchanged during ripening and were at similar levels compared with date mesocarp (Fig. S6). This strongly contrasts with Arabidopsis and B. napus seeds, where PFP is
much more highly expressed than PFK and has been hypothesized to carry most of the flux through glycolysis (23).

The most striking changes with regards to carbohydrate pathways concerned plastid transporters (Fig. 6B). Oil palm ESTs coding for a GPT2 ortholog, which in Arabidopsis transports hexose, pentose, and triose phosphate (24), increased 16-fold during ripening, were on average 9 times higher than in date (Fig. 6B and C), and represented one of the highest EST counts of all carbohydrate-related genes of oil palm mesocarp. Interestingly, antisense inhibition of a GPT transporter in Vicia seeds altered carbon partitioning and promoted protein storage at the expense of carbohydrates (25). ESTs for the phosphoenolpyruvate (PEP) transporter PPT1 were 3.7 times higher than in date palm and increased fourfold during oil palm ripening (Figs. 5 and 6B). PFP and malate have been shown to provide some contribution to oil synthesis, both as carbon units and as a reductant source (3). Although plastid-localized enzymes of these pathways showed EST increases in oil palm, these were moderate compared with plastid PK, PFK, and GPT2 and fatty acid synthesis enzymes (Figs. 5 and Dataset S2B). Only a strong decrease in cytosolic malic enzyme was noted in oil palm compared with date palm.

Taken together, our results suggest that (i) plastid glycolysis is up-regulated in oil palm compared with date palm and temporally during ripening, and (ii) GPT2 and PPT provide glycolytic substrates (hexose P) and intermediates (triose phosphate, PEP) to the plastid by transport from the cytosol, a trend that is accentuated during ripening. This implies a strong funneling of carbon toward pyruvate in the plastids of oil palm but not of date palm. Thus, strongly increased fatty acid synthesis (Figs. 2 and 3), together with plastid carbon supply, is likely crucial for the eventual accumulation of 90% oil in the mesocarp.

**Ratio of Cell-Wall Invertase to Sucrose Synthase ESTs Is Much Higher in Oil Palm than in Oil Seeds.** Carbon is supplied to the fruit mostly as apoplastic sucrose. Higher cell-wall invertase activity is known to lead to increased fruit sugar levels and increased seed weight in tomato fruit (26). Because the sum of ESTs encoding cell-wall invertases increased about 3-fold during oil palm ripening and were 3.7-fold higher than in date palm (Fig. 5 and Fig. S6B) and 50-fold higher than in oil palm leaf (Dataset S2B), we speculate that a significant part of sucrose is hydrolyzed at the cell wall. The level of ESTs coding for sucrose synthase, which initiates a pyrophosphate-dependent pathway to hexose-P via UDPG pyrophosphorylase, was similar in oil palm mesocarp and in leaf, was 2.2-fold higher than in date, and increased by 2-fold during oil palm ripening. The ratio of cell-wall invertase to sucrose synthase ESTs was 35-fold higher than this ratio in Arabidopsis seeds (4). This suggests, as noted above for the PFP versus PFK ratios, that oil palm mesocarp and oil seeds differently use ATP- and pyrophosphate-dependent enzymes.

Aside from glycolysis, hexoses provide carbon for starch biosynthesis. ESTs coding for ADPG pyrophosphorylase (AGPase) and starch synthase showed 13- and 9.4-fold increases during oil palm ripening, respectively (Fig. 5, Fig. S6C, and Dataset S2B). Interestingly, plastid phosphoenolpyruvate orthophosphate dikinase (PPDK) showed a similar expression pattern (Fig. S6C). In maize, PPDK belongs, together with AGPase, to a starch synthesizing multienzyme complex (27) that likely facilitates energy and proton exchange between AGPase and PPDK. Taken together, these data explain the late increase of starch that was revealed by our metabolite analysis of oil palm mesocarp (Dataset S1B).

**Conclusions**

The temporal and cross-species comparisons made in this study provide a concurrent exploration of transcripts and metabolites associated with major species-specific differences in central carbon partitioning. In almost all cases, if transcripts increased more than threefold during oil palm ripening, there was also a greater or similar difference between oil and date palm. This comparison between oil and date palm reinforced and strengthened the interpretations derived from transcript temporal patterns. The high oil content of oil palm was not associated with increased gene expression for many of the enzymes required for hexose-to-TAG metabolism. Instead, high oil production correlated most strongly with temporal increases and high oil-palm-to-date palm ratios for transcripts coding for all fatty acid synthesis enzymes (including PDH), specific plastid transporters, select plastid glycolytic enzymes (namely PFK and PK), and a WR11-like transcription factor (WR11-like). The expression levels of these ESTs are 13- and 9.4-fold higher, respectively, in oil palm mesocarp and date palm, and that all its known targets were up-regulated, implies a remarkable similarity between regulation of fatty acid synthesis destined to oil in seeds and in nonseed tissues. Intriguingly, although oil palm and date palm appeared to express similar levels of transcripts for the complete pathway of TAG synthesis from plastid fatty acids, date palm accumulates no detectable TAG.

Although many features of the transcription patterns described here showed similarities to that in oil seeds, important differences could be noted. WR11-like does not appear to be under control of the same upstream factors (e.g., LEC2) as in oil seeds. This may explain in part the very different temporal profiles of expression between oil palm and oil seeds. For most oil seeds, transcripts related to lipid synthesis peak and decline well before the later stages of TAG accumulation (3). In contrast, in oil palm, almost all of the up-regulated transcripts did not decline but continued to increase until the end of oil accumulation. Also, in contrast to oil seeds, oil palm showed a preferential up-regulation of ATP-dependent enzymes rather than one for pyrophosphate-dependent enzymes of sucrose breakdown and early glycolysis, suggesting that ATP availability might be an important factor related to higher levels of oil. Finally, such a large change in expression of the GPT plastid transporter gene was not reported in oil seeds. All these differences may provide insights into how oil palm mesocarp accumulates more oil than oil seeds do.

The oil palm and date palm EST datasets also provide information on several lipid-related enzymes and their specific isoforms, such as PAPs, which may give possible leads for future research. More generally, the data illustrate that such a dual developmental and comparative approach, applied to diverse biological systems of interest, can be as powerful as developments in transcriptomic studies carried out in model organisms. Finally, our results increase tremendously the molecular resources available for palms and provide a large EST collection for fruits. In addition to knowledge generated on the regulation of carbon partitioning in oil palm mesocarp, these resources should also...
allow breeders to identify candidate genes involved in high oil synthesis. For example, polymorphism of WRI1 in rapeseed relates highly to oil content (28). This will pave the way to marker-assisted selection of oil palm, a key technique for more rapid improvement of crop trees.

Materials and Methods
For complete details, see SI Materials and Methods.

Biological Material. Oil and date palm fruits were harvested at 65.1206.50:2242.484.514.16:3304 and 15447.26:150:55 2811.29021.80 nm thickness fl 68:2232 rbcl 23999.902. Plant Cell < 53.783. Oil palm mesocarp (20 WAP) was subjected to cryo-pretreatment and cryosubstitution before sectioning. Slices of 50 μm thickness were observed by transmission electron microscope mostly as described (31). cDNA Library Construction and Sequencing. Total RNA was extracted from five stages of oil palm (32) and date palm mesocarp (33) and oil palm leaves. Oil palm and date palm cDNA libraries were synthesized according to GS20 DNA Library Preparation (Roche) and Roche cDNA Library Preparation Method, respectively. cDNA libraries were subjected to emulsion PCR, and libraries with amplified library were loaded onto a picotiter plate and sequenced. Pyrosequencing was performed at the Department of Energy Joint Genome Institute following protocols for the Genome Sequencer GS FLX Titanium System (Roche Diagnostic).

Bioinformatics and Data Analyses. EST reads obtained from 454 Sequencing were trimmed and filtered to remove low-quality sequences. De novo assembly was by CAP3 software. Blast similarity search using BLASTX (E-value < 10-10) against the The Arabidopsis Information Resource 8 and RefSeq (National Center for Biotechnology Information) databases was performed to assign an annotation to the transcripts. EST sequences generated from this study are available from National Center for Biotechnology Information Sequence Read Archive (accession nos. SRX059258-62, SRX059116-20, and SRX059798-802). Information on data files containing annotation and temporal expression along with unprocessed sequences is provided in SI Materials and Methods. Contigs with more than 10 ESTs are listed in Dataset S2 and sequences can be downloaded at http://www.biomemb.cnrs.fr/contigs.html.

The number of ESTs/100,000 ESTs was used as an estimate of gene expression to enable fold change comparison between stages and species. Values greater or equal to twofold that are presented in figures or text are significant at P < 0.05 on the basis of the DESeq method (34).

Acknowledgments. We thank J. Malek and R. Sambanthamurthi for sharing unpublished results, A. Troncose-Ponce for help with data analysis and P. Koon for support. Metabolite analyses were carried out at the Metabolome Facility of Bordeaux Functional Genomics Center (University of Bordeaux-Ségalen/Institut National de la Recherche Agronomique/Centre National de la Recherche Scientifique) by M. Maucourt and C. Debore (NMR) and by Y. Gibon and V.A. (starch). Electron microscopy was carried out at the Bordeaux Imaging Center (Plant Imaging Facility) by M. Peyypel, with help from B. Batailler, C. Cheniclet, and V. Rouyère. DNA sequencing was supervised by E. Lindingqul and C. Pennacchio. EST assembly and data bases were provided by Curt Wilkerson and Nick Thrower (Michigan State University). F.B. was the recipient of a fellowship from the Conseil Régional d’Aquitaine. G.F.-N.E. was the recipient of a short-term European Molecular Biology Organization fellowship. This work was supported in part by the Office of Science of the Department of Energy under Contract DE-AC02-05CH11231.
Supporting Information

Bourgis et al. 10.1073/pnas.1106502108

SI Materials and Methods

Biological Material for Metabolite Analyses and RNA Extraction. 
Harvest dates were chosen to represent comparable stages of ripening for oil palm and date palm. Oil palm fruits (cv. Dura) were harvested from five closely related trees (resulting from self-pollination of DA115D tree) at the Center for Oil Palm Research (Centre de Recherches sur le Palmier à Huile de la Dibamba) in Cameroon. Five stages of fruits—15, 17, 19, 21, and 23 wk after pollination (WAP)—were harvested from external spikelets of these trees on February 10, 2009. Date palm fruits (cv. Deglet Noor) were from a single late-flowering (mid-June) tree. Five stages of fruits—16, 19, 22, 25, and 27 WAP—were harvested from mid-October to end of December 2008 in the Sfax area in Tunisia. All fruits were flash-frozen in liquid nitrogen after harvest and were transported on dry ice to the Laboratoire de Biogenèse Membranaire and stored at −80 °C until analysis. Mesocarp from each developing stage was quickly dissected from frozen fruits and ground in liquid nitrogen to a fine powder, which was used for metabolite analyses and RNA extraction.

Electron Microscopy. A spikelet carrying 20 WAP fruits was harvested and conserved at ambient temperature (20–30 °C) for less than 48 h after harvest. Mesocarp fragments from distal and proximal (close to the attachment) parts of the fruit were selected for fixation. Mesocarp area with minimal fiber content was sectioned by hand or vibratome (Microm) set at 150 μm. Sectioned samples were immersed in a drop of hexadecane and their size rapidly adjusted to fit fixation container dimensions (1,500 × 200 μm in diameter and thickness). Cryo-fixation was carried out immediately, after removal of excess hexadecane, using Leica EMPLAC. Cryo-substitution was performed (using Leica AFS) by 2% osmium-containing acetone, followed with or without a 1% tannic acid treatment (in acetone). Samples were then embedded in Epon resin. Slices of 1-μm thickness were obtained using a Reichert 2040 microtome and stained with toluidine blue to check quality and to select appropriate areas for further examination. Ultrathin slices were obtained using an ultramicrotome Reichert Ultracut Leica with a diamond blade. Slices of 50- to 70-nm thickness deposited on a 200-mesh grid wrapped with parlodion were treated with 7% uranyl acetate (in 50% ethanol), rinsed several times with ultrapure distilled water, and further treated with lead citrate. Tannic acid treatment did not improve the quality of slices. Distal material with lower oil content allowed for better resolution. Observations were carried out on a transmission electron microscope Phillips CM10 80 kV.

Metabolite Profiling by Proton NMR. Primary metabolites were quantified using 1H-NMR for polar extracts (1). The preparation of extracts and NMR acquisition parameters were optimized to allow absolute quantification of individual metabolites. Briefly, polar metabolites were extracted (40–50 mg dry weight per technical replicate) with an ethanol−water series at 80 °C. The lyophilized extracts were titrated with potassium deuteroxide to pH 6 in 200 mM potassium phosphate buffer in D2O, added with EDTA and diosidum salt at a final concentration of 2 mM, and lyophilized again. Each dried titrated extract was solubilized in 0.5 mL D2O with (trimethylsilyl)propionic-2,2,3,3-d4 acid sodium salt (0.01% final concentration) for chemical shift calibration. 1H-NMR spectra were recorded at 500.162 MHz on a Bruker Avance spectrometer using a 5-mm broadband inverse probe and an electronic reference for quantification. Sixty-four scans of 32,000 data points each were acquired with a 90° pulse angle, a 6,000-Hz spectral width, a 2.73-s acquisition time, and an optimized recycle delay (20 or 25 s). Preliminary data processing was conducted with TOPSPIN 2.1 software (Bruker Biospin). The assignments of metabolites in the NMR spectra were made by comparing the proton chemical shifts with literature or database values (MeRy-B 2010 at http://www.cbib.u-bordeaux2.fr/MERYB), by comparing with spectra of authentic compounds, and by spiking the samples.

Two-dimensional experiments (1H–1H COSY, 45° read pulse; Heteronuclear Single Quantum Coherence) were conducted to verify the identity of compounds. The metabolite concentrations were calculated using AMIX (version 3.9.1; Bruker) and Excel software. The metabolites were quantified using the glucose calibration curve and the proton amount corresponding to each resonance for all compounds except fructose, glutamate, and glutamine that were quantified using specific calibration curves. The metabolite concentrations were calculated from concentrations in the NMR tube and sample dry weight. Data for three technical replicates were calculated as mean ± SD.

The 15 1H-NMR spectra of each species were converted into JCAMP-DX format and deposited with associated metadata into the Metabolomics Repository of Bordeaux MeRy-B http://www.cbib.u-bordeaux2.fr/MERYB, accession no. D10001 and D10002.

Lipid Analyses. Mesocarp ground in liquid nitrogen was homogenized in boiling isopropyl alcohol for 10 min. The solid residue was extracted according to modified Folch procedure (2) and both extracts were pooled. Neutral lipids were separated by TLC as described (2). Triacylglycerol (TAG) and diacylglycerol (DAG) spots were scraped off the plate, and the lipids were extracted according to Bligh and Dyer method (2) and transmethylated using methanol containing 2% sulfuric acid. Fatty acid methyl esters were analyzed by gas chromatography as described (3). Lipids were quantified using either C17 fatty acid (oil palm) or glyceryl triheptadecanoate (date palm) as the internal standard. Polar lipids were extracted from the total lipid by phase partition (hexane/ethanol/water, 75/25/2.5; chloroform/methanol/acetic acid/water, 80/9/12/2), and fatty acid composition was determined.

RNA Extraction. Tris-HCl (pH 8) was prepared using diethylpyrocarbonate (DEPC)-treated water, and all other solutions were treated with 0.1% DEPC.

Oil palm. Frozen powder was homogenized (5 mL of buffer per gram of powder) in 5 M guanidine thiocyanate, 100 mM Tris-HCl (pH 8), 0.5% lauroyl sarcosine, 1% insoluble polyvinylpolyrylidone (PVPP), and 1% β-mercaptoethanol. Debris were removed by centrifugation at 10,000 × g for 10 min, and the supernatant (avoiding fat layer on top) was carefully overlaid onto a 9-mL cushion made of 5.7 M CsCl and 10 mM EDTA (pH 7.5). Ultracentrifugation was carried out in 36-mL tubes using a Sorval AH629 rotor at 20 °C and centrifuged at 25,000 × g (112,000gmax) for 26 h. Supernatant was carefully removed and discarded. The RNA pellet was resuspended in 10 mM Tris-HCl (pH 8), 1 mM EDTA, 0.1% SDS buffer and stored at −80 °C.

Date palm. We were unable to recover sufficient date palm RNA using protocols based on aqueous phenol extraction alone or in combination with guanidine thiocyanate or hydrochloride, with LiCl precipitation of RNA, or with the protocol that proved successful for oil palm. This was likely due to very high carbo-
hydrate content. RNA was successfully obtained when frozen powder was homogenized at 65 °C using an Ultraturax set at high speed into prewarmed Tris-HCl 100 mM (pH 8), 25 mM EDTA, 2.5 M NaCl, 2% Hexadecyltrimethylammonium bromide, 2% PVPP, 0.05% spermidine, and 2% β-mercaptoethanol, using at least 10 mL of buffer per gram of frozen mesocarp powder. The mixture was incubated for 15 min at 65 °C, and the protocol described by Meisel et al. (5) was followed exactly.

Library Construction and Sequencing. mRNA was isolated from total RNA using the Dynabeads mRNA Direct Kit (Invitrogen). Purity and quantity of RNA was evaluated on RNA 6000 Pico chip using the Agilent 2100 Bioanalyzer. The oil palm cDNA library was synthesized using SuperScriptIII (Invitrogen) reverse transcriptase and dT\(_{15}\)VN\(_{15}\) primer according to GS20 DNA Library Preparation (Roche). The date palm cDNA library was constructed using the Roche cDNA Rapid Library Preparation Method, using avian myeloblastic virus reverse transcriptase and Roche Random primer. Subsequent to second-strand synthesis, adaptor ligation was done, and the quality of cDNA was assessed after removal of small fragments. After fragmentation, cDNA fragments were amplified by emulsion PCR and processed for sequencing. DNA sequencing was performed at the Department of Energy Joint Genome Institute following protocols for the Genome Sequencer GS FLX Titanium System (Roche Diagnostic).

Bioinformatics and Data Analysis. Reads obtained from 454 Sequencing were trimmed to remove both low-quality and primer sequences using Seq-Clean (6). To obtain larger contigs, the 454 reads obtained for all stages of mesocarp of oil palm or date palm were assembled to generate contigs (and subsequently deconvoluted to obtain ESTs at each stage). Initially, 5% of the data were assembled with CAP3 (7) to identify high abundance ESTs, which were then removed from the full dataset using BLAT (8). The reduced dataset then underwent two rounds of assembly with CAP3. First-round CAP3 parameter settings for percentage match, overlap length, maximum overhang percentage, gap penalty, and base quality cutoff for clipping were p90, o50, h15, g2, and c17, respectively. For the second round, “−o” was changed to 100.

The contigs were matched to Arabidopsis Genome Initiative (AGI) locus ID by BLASTX to the Arabidopsis proteome (The Arabidopsis Information Resource 8) with a cutoff E-value < 10\(^{-10}\). There may be more than one contig assigned to each AT locus ID. The EST numbers for multiple contigs that matched to the same At locus ID are summed in Dataset S2. To enable comparison of EST levels across stages and between species, data at each stage are expressed as ESTs/100,000 ESTs. The nucleotide sequences for individual ESTs will be available upon publication in the National Center for Biotechnology Information Sequence Read Archive (accession nos. SRX059258-62, SRX059116-20, and SRX059798-802).

For the enzymes and proteins shown in Figs. 3 and 5, assignments of palm contigs to a protein function (i.e., an enzyme carrying out a given reaction) were manually verified by examination of BLASTX reports. In all cases, the difference in BLASTX score between the highest AGI locus hit and the second highest hit was large enough to clearly attribute an enzymatic reaction to a contig’s deduced protein. In addition, plastid or cytosol assignments (shown in Figs. 3 and 5) were individually verified by examination of deduced protein sequence for N-terminal extensions (plastid) and for strict conservation of the initiating methionine (cytosol). This analysis confirmed localization for all full-length contigs except 6GPDH. The Arabidopsis plastid 6GPDH isoforms lack N-terminal extensions. However, BLASTX results together with data from Arabidopsis plastid proteome sequences allowed unambiguous assignment of plastid isoforms. The third isoform (ortholog of A2SG75670) might have a dual peroxisome/cytosol location.

As noted in Results, palm may have gene/protein families with more or fewer members than Arabidopsis. However, as noted above, for data presented from this study, contig assignments to enzyme/protein function were not ambiguous. Therefore, for multi-isof orm families represented in Dataset S2 A and B and that belonged to the same subcellular compartment, ESTs for isoforms were summed for presentation in Figs. 3 and 5. Both date and oil palm contigs that were assigned to a given Arabidopsis AGI were indeed closer (nucleotide sequence identity) to each other than to contigs assigned to another Arabidopsis AGI. However, these sequence comparisons suggested that in some cases isoform genesis occurred independently in palms and Arabidopsis after these had diverged from a common ancestor, so that strict orthology between Arabidopsis and palms for a given isoform might not always be true.

Comparative analyses of transcriptome data were conducted to determine the fold difference in EST levels between (i) oil palm and date palm (sum of ESTs for five developing stages for oil palm divided by date palm sum), (ii) oil palm mesocarp at 23 and 15 WAP, and (iii) oil palm mesocarp and leaf. Data for select genes related to lipid and carbohydrate metabolism are presented in Dataset S2 A and B. Fold differences presented in figures or text (in main text) that were greater or equal to twofold were significant at P < 0.05 on the basis of the DEGseq method (random sampling model) (9). A plot of the log2 transformation of the oil-palm-to-date-palm EST ratios displayed a normal distribution with average ratio of 1.03 (log2 = 0.04).

Comments on Nucleotide Sequence Identity and Phylogeny of Palms. Major lineages of monocots are considered to have diverged 100–146 million years ago (10). The minimal age of palm trees is considered to be about 120 million years, and the Cocoseae, which comprise Elaeis guineensis but not Phoenix dactylifera, started to radiate more than 60 million years ago (11, 12). Therefore, oil palm and date palm diverged probably about 100 million years ago but kept a high nucleotide sequence similarity (about 92%) across our data. By contrast, the average nucleotide sequence identity of Arabidopsis thaliana and Brassica napus is about 85% (13), and they are considered to have diverged about 20 million years ago (14). Our data are in agreement with other palm studies based on RBCL and ADH genes (15), which have shown that synonymous nucleotide substitution rates were 5- and 2.5-fold lower in palms than in grasses, respectively. This lower nucleotide substitution might be due a longer generation time of palms compared with annual plants (15).


**Fig. S1.** Transmission electron micrographs (TEM) of oil palm mesocarp. (A) TEM of a cell from proximal mesocarp at 20 WAP. Oil appears to be stored predominantly as a large, single droplet. (B) TEM of a cell depicting a chromoplast (C) and mitochondria (M). Chromoplasts are the only type of plastids that were observed in 20-WAP mesocarp cells. (C) Starch granules (S) appeared only in amyloplasts from mesocarp of a ripe fruit of *Elaeis oleifera*. CW, cell wall; OD, oil droplet; OPM, outer plastid membrane. A cryo-fixation procedure was used to preserve the ultrastructure. However, because *E. guineensis* ripe fruit contains about 90% oil (soft material) and 5% fibers (hard material), it has not been possible to obtain intact ultrathin slices. Therefore, we used fruits at 20 WAP, which contained about 30% oil. To observe amyloplasts in ripe fruit, we used *E. oleifera* because this closely related species contains less oil (∼50%), which allowed us to obtain good quality ultrathin slices.
EST profiles for select lipid-related genes. ESTs levels and temporal patterns for most genes involved in glycerolipid assembly were similar in oil palm and date palm mesocarp (e.g., GPAT9, LPAAT2, DGAT1). Exceptions were noted for DGAT2, LPPβ, and FAD2 isoforms in oil palm, which showed patterns similar to that of fatty acid synthesis genes (Fig. 2B and Fig. S2).

Fig. 52. Expression profiles of 18 plastid fatty acid synthesis proteins indicate highly coordinated expression in oil palm mesocarp. To compare the profiles, EST data for each protein were calculated as the percentage of ESTs at each stage to total ESTs for all stages. The mean 21:15 WAP ratio was 5.7 ± 1.3 (n = 18), and the linear regression of 15–21 WAP showed a $R^2$ value >0.99.

Fig. 53. EST profiles for select lipid-related genes. ESTs levels and temporal patterns for most genes involved in glycerolipid assembly were similar in oil palm and date palm mesocarp (e.g., GPAT9, LPAAT2, DGAT1). Exceptions were noted for DGAT2, LPPβ, and FAD2 isoforms in oil palm, which showed patterns similar to that of fatty acid synthesis genes (Fig. 2B and Fig. S2).
Fig. S4. Oil-palm-to-date palm EST ratio for select lipid-related genes. Most enzymes and proteins related to acyl-CoA reactions or binding had higher levels of ESTs in oil palm than in date palm, suggesting a role in TAG synthesis. Enzymes involved in phosphatidylycholine synthesis also showed changes. ACBP, acyl-CoA binding protein; ACC, acetyl-CoA carboxylase; CCT, CTP: choline-phosphate cytidylyltransferase; CK, choline kinase; CPT, CDP-choline:DAG choline phosphotransferase; DGAT, diacylglycerol acyltransferase; LACS, long-chain acyl-CoA synthase; LPCAT, lysophosphatidylcholine acyltransferase; LPEAT, lysophosphatidylethanolamine acyltransferase; PAH and LPP, isoforms of phosphatic acid phosphatase; PDCT, phosphatidylycholine:diacylglycerol cytidylyltransferase; PEAMT, phosphoethanolamine methyl transferase; PLAT, phospholipid acyl transferase-like (uncharacterized); PLMT, phosphatidylethanolamine methyl transferase.

Fig. S5. Phylogenetic tree indicates close relationship of oil palm (Eg-WRI1-like) sequence in oil palm to that of WRI1-like in maize (Zm) and Arabidopsis (At). Other sequences included in this comparison contain a double AP2 domain (1). The tree was built from multiple alignments of protein sequences, using ClustalW and TreeView software. Bar indicates the percentage divergence between the sequences.

Temporal profile of select genes involved in carbohydrate pathway. (A) In cytosol, EST levels for ATP-dependent phosphofructokinase (PFK) in oil palm mesocarp increased with ripening but not in date palm. Interestingly, pyrophosphate-dependent phosphofructokinase (PPF) showed no difference in profile during development or in comparison with date palm. (B) EST pattern for sucrose-hydrolyzing enzymes, cell-wall invertase (CW-INV), and sucrose synthase (SuSy) indicate higher expression levels in oil palm relative to date palm and a temporal increase that is consistent with oil accumulation. (C) In oil palm mesocarp, ADPG pyrophosphorylase (AGPase), an enzyme involved in starch metabolism, showed an expression pattern that is temporally associated with starch accumulation. Interestingly, orthophosphate pyruvate dikinase (PPDK) also followed a similar pattern. In maize, plastid PPDK has been shown to belong to a multienzymatic complex that associates with starch metabolism (1). It has been suggested that this complex might help to channel pyrophosphate generated by PPDK to AGPase. This hypothesis is consistent with the similar patterns of expression of PPDK, starch biosynthetic enzymes, and starch accumulation in mesocarp.


Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLS)