PCP
PLANT & CELL PHYSIOLOGY

Identification of Genes Involved in Lipid Biosynthesis through de novo Transcriptome Assembly from *Cocos nucifera* Developing Endosperm

Kyle B. Reynolds [1,2,3,6,*], Darren P. Cullerne[4,5,6], Anna El Tahchy[1], Vivien Rolland[1], Christopher L. Blanchard[3], Craig C. Wood[1], Surinder P. Singh[1] and James R. Petrie[1]

[1]Commonwealth Scientific and Industrial Research Organisation, Agriculture and Food, Canberra, ACT, 2601, Australia
[2]Department of Primary Industries, Graham Centre for Agricultural Innovation, Charles Sturt University, Wagga Wagga, NSW, Australia
[3]ARC Industrial Transformation Training Centre for Functional Grains, Charles Sturt University, Wagga Wagga, NSW, Australia
[4]School of Environmental and Life Sciences, University of Newcastle, Newcastle, NSW, Australia
[5]Research School of Biology, Australian National University, Canberra, ACT, Australia
[6]These authors contributed equally to this work.
*Corresponding author: E-mail, Kyle.Reynolds@csiro.au; Fax, +61 2 62464950.
(Received April 9, 2018; Accepted December 19, 2018)

Regular Paper

***Cocos nucifera* (coconut), a member of the Arecaceae family, is an economically important woody palm that is widely grown in tropical and subtropical regions. The coconut palm is well known for its ability to accumulate large amounts of oil, approximately 63% of the seed weight. Coconut oil varies significantly from other vegetable oils as it contains a high proportion of medium-chain fatty acids (MCFA; 85%). The unique composition of coconut oil raises interest in understanding how the coconut palm produces oil of a high saturated MCFA content, and if such an oil profile could be replicated via biotechnology interventions. Although some gene discovery work has been performed there is still a significant gap in the knowledge associated with coconut's oil production pathways. In this study, a *de novo* transcriptome was assembled for developing coconut endosperm to identify genes involved in the synthesis of lipids, particularly triacylglycerol. Of particular interest were thioesterases, acyltransferases and oleosins because of their involvement in the processes of releasing fatty acids for assembly, esterification of fatty acids into glycerolipids and protecting oils from degradation, respectively. It is hypothesized that some of these genes may exhibit a strong substrate preference for MCFA and hence may assist the future development of vegetable oils with an enriched MCFA composition. In this study, we identified and confirmed functionality of five candidate genes from the gene families of interest. This study will benefit future work in areas of increasing vegetable oil production and the tailoring of oil fatty acid compositions.**

Keywords: Coconut • *De novo* assembly • DGAT • Endosperm • FATA • Fatty acid • GPAT9 • Medium chain • *Nicotiana* • Oleosin • PDAT • Thioesterase • Transcriptome • Triacylglycerol.

Abbreviations: BLAST, Basic Local Alignment Search Tool; BUSCO, benchmarking universal single-copy orthologues; CEGMA, core eukaryotic genes mapping approach; CTAB, cetyl trimethylammonium bromide; DAG, diacylglycerol; DGAT, diacylglycerol acyltransferase; dpm, disintegrations per minute; FATA, class A acyl-ACP thioesterase; FATB, acyl-ACP thioesterase class B; G3P, glycerol-3-phosphate; GPAT9, glycerol-3-phosphate acyltransferase; LPAAT, lysophosphatidic acid acyltransferase; MCFA, medium-chain fatty acid; PC, phosphatidylcholine; PDAT, phospholipid:diacylglycerol acyltransferase; RNA, ribonucleic acid; TAG, triacylglycerol; TLC, thin-layer chromatography; WRI1, WRINKLED1.

Introduction

Cocos nucifera (coconut palm) is a member of the Arecaceae family, being widely grown in tropical and subtropical regions as an economically important plant, accounting for 1.8% of the global vegetable oil market with production of 3.2 million metric tonnes per year (OECD 2017). The coconut palm is well known for its ability to accumulate and store large amounts of oil, approximately 65% of the seed weight (Laureles et al. 2002, Gunstone et al. 2007, Kumar 2011). In physiochemical terms, coconut oil varies significantly from other vegetable oils as it is rich in saturated fatty acids (93%) with 85% of these being medium-chain fatty acids (MCFA; C8–C16), predominantly lauric (C12:0) and myristic (C14:0) acids. Oils that are rich in lauric acid exhibit characteristics that are important for application within both the food and chemical industries including increased oxidative stability, low melting points and the ability to produce stable emulsions (Arkcoll 1988, Kumar 2011). Currently, lauric oils are used in the manufacture of soaps, detergents, textiles, paints, varnishes, cosmetics, pharmaceutical products and are also used within the food industry (Arkcoll 1988, Laureles et al. 2002, Kumar 2011). Lauric oils, are produced by only two industrially important species being coconut palm and *Elaeis guineensis* (African oil

Plant Cell Physiol. 60(5): 945–960 (2019) doi:10.1093/pcp/pcy247, Advance Access publication on 4 January 2019, available online at https://academic.oup.com/pcp

palm), both being limited to growth in tropical and subtropical climates. The large investment required and the long wait before the palms yield a return of profit is discouraging the establishment of new plantations (Arkcoll 1988, Basiron 2007, Carter et al. 2007, Murphy 2007), and is also a major driver for the development of new MCFA-producing crops that can be grown in more temperate climates.

The unique composition of coconut oil raises interest in understanding how the coconut palm produces oil of a high saturated MCFA content, and if such an oil profile could be replicated in another species via biotechnology interventions. To date, there has been limited success in a variety of species. In transgenic *Brassica napus* (canola) seeds, high lauric acid content has been achieved to a peak of 67% (Dehesh et al. 1996) with the co-expression of a class B fatty acid thioesterase (*FATB*) from *Umbellularia californica* (California bay laurel; *UcFATB*) and lysophosphatidic acid acyltransferase from coconut palm (*CnLPAAT*). More recently, transcriptomic analyses have enabled the identification of *FATB* and *LPAAT* genes from many *Cuphea* species (cigar plant), which have been used to both modify the fatty acid profiles and improve the incorporation of MCFA into the triacylglycerol (TAG), respectively, of transgenic *Camelina sativa* seeds (Kim et al. 2015a, Kim et al. 2015b). However, during seed development the laurate content in phosphatidylcholine (PC) increases, suggesting that there is a bottleneck in the efficient utilization of lauric acid during oil synthesis. This concept has also been demonstrated in attempts to accumulate unusual fatty acids in the seed oils of *Arabidopsis thaliana* (Bates and Browse 2011). In the high-laurate canola lines, laurate levels in PC remained high (up to 26%) until the late stages of TAG deposition (Wiberg et al. 1997). The subsequent work from Knutzon et al. (1999) showed that the addition of the *LPAAT* from coconut (*CnLPAAT*), although specific for the acylation of laurate to the *sn*-2 position of glycerol-3-phosphate (G3P), did not reduce the flux of laurate into PC. We hypothesize that the MCFA content of transgenic oils may be influenced through the specificity and activity of biosynthetic enzymes. Therefore, better manipulation of fatty acid and oil biosynthesis pathways will be essential for improving the performance of MCFA accumulation in transgenic oils. The discovery of genes involved in the assembly of MCFA-enriched oils is an important step forward in achieving this. However, such efforts have been hindered by the limited genomic resources available.

In prospect of improving the MCFA content of oils the identification of genes involved in TAG assembly were of particular interest, specifically MCFA-specific acyltransferases. Although *LPAAT* from coconut palm (Knutzon et al. 1995, Knutzon et al. 1999) and many *Cuphea* species (Kim et al. 2015a, Kim et al. 2015b) have previously been identified and characterized, other MCFA-specific acyltransferases remain unknown. The *AtGPAT9* (glycerol-3-phosphate acyltransferase) was recently identified and was demonstrated to have functions in both initiating TAG assembly through the mediation of *sn*-1 esterification and increasing TAG content (Shockey et al. 2016, Singer et al. 2016). The final step of TAG assembly through the Kennedy pathway involves the esterification of a third fatty acid to the *sn*-3 position of G3P, which is performed by diacylglycerol acyltransferase (*DGAT*). *DGAT* has been shown to exhibit variations in substrate specificity (Frentzen 1998, Durrett et al. 2010, Dussert et al. 2013, Aymé et al. 2015, Aznar-Moreno et al. 2015), and hence has a significant contribution towards the final fatty acid composition of plant oils. Recent transcriptomic analyses have enabled the identification of a new *DGAT* gene from *Cuphea avigera* var *pulcherrima*, *CpuDGAT1*, which was combined with the expression of *CvFatB1* and *CvLPAT2* to achieve high levels of C10:0 incorporation in the oils of transgenic *C.sativa* seeds (Iskandarov et al. 2017). Another acyltransferase that contributes to TAG synthesis is phospholipid:diacylglycerol acyltransferase (*PDAT*), which uses DAG as the substrate and PC as the fatty acid donor for TAG production (Dahlqvist et al. 2000). Thus the involvement of acyltransferases in both the efficient production of oil and the tailoring of fatty acid composition makes these important targets for the improvement of MCFA oils.

Oleosins have also become an attractive target of metabolic engineering in the production of stable oils in transgenic plants, particularly for non-seed platforms for the production of oil (Vanhercke et al. 2014a). They stabilize oil by covering the surface of lipid droplets and protect against the degradation by lipases (Hsieh and Huang 2004, Maurer et al. 2013). Oleosin proteins contain a hydrophilic N-terminus, a central hydrophobic stretch of 72 residues and a hydrophilic C-terminus. The central stretch of 72 hydrophobic residues is the hallmark of oleosins as no other protein contains such a long hydrophobic stretch. Another key feature of an oleosin protein is the formation of the proline-knot in the center of the long hydrophobic region that forms via the interaction of one serine and three proline amino acid residues (Hsieh and Huang 2004). Based on the high oil content of the seeds from the coconut palm (67%) (Padolina et al. 1987, Arkcoll 1988), it was hypothesized that *oleosins* must be playing an integral role in the accumulation and maintenance of the oil.

Recent gene discovery studies of coconut palm have focused on identifying genes involved in yield, with transcriptomes built using material from leaves and mature fruits (H. Fan et al. 2013, Huang et al. 2013, Huang et al. 2014, Liang et al. 2014, Nejat et al. 2015). Consequently, there is still a significant gap in knowledge associated with the oil biosynthesis pathways in coconut palm. In this study, an RNA-Seq approach was used to build a *de novo* transcriptome from developing coconut endosperm. This transcriptome was utilized to identify and isolate genes involved in the synthesis of lipids, particularly TAG. It was hypothesized that some of these genes may exhibit a strong substrate preference for MCFA, and so could be used for the future development of vegetable oils with an enriched MCFA content. Of particular interest are thioesterases, acyltransferases and oleosins (**Fig. 1**), which are involved in the processes of releasing fatty acids for assembly, the esterification of fatty acids into glycerolipids and the protection of oils from degradation, respectively. These candidate genes were then confirmed to be functional using the transient *Nicotiana benthamiana* leaf assay system.

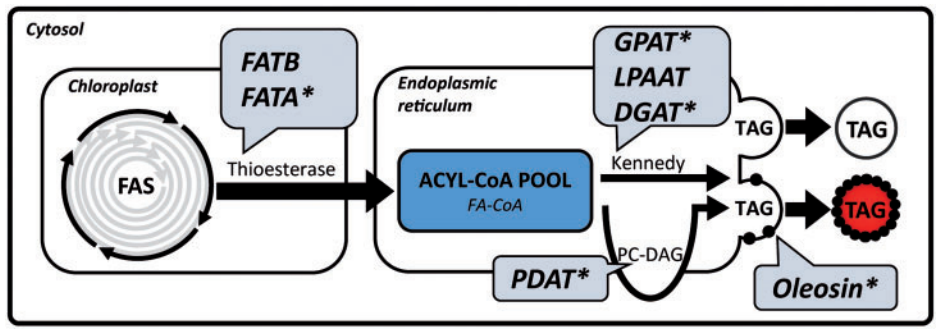

Fig. 1 Graphical representation of the primary pathways involved in triacylglycerol (TAG) production, highlighting the lipid synthesis genes of interest (grey bubbles). The asterisk represents those genes that have been investigated in this study following the construction of the *Cocos nucifera* (coconut palm) endosperm transcriptome. FAS, fatty acid synthase; *FAT*, fatty acid thioesterase, class A (*FATA*) and class B (*FATB*); *GPAT*, glycerol phosphate acyltransferase; *LPAAT*, lysophosphatidic acid acyltransferase; *DGAT*, diacylglycerol acyltransferase; *PDAT*, phospholipid:diacylglycerol acyltransferase; PC,phosphatidylcholine; DAG, diacylglycerol.

Results

Coconut endosperm transcriptome

Before proceeding with RNA extractions, total lipids were extracted from a sample of coconut endosperm to determine the fatty acid composition. The total lipids were analyzed via thin-layer chromatography (TLC) (**Fig. 2A**) using a two phase system, which revealed that the white endosperm layer was composed almost entirely of TAG, hence confirming that the endosperm is a layer of oil deposition (Oo and Stumpf 1979). The fatty acid composition of the endosperm (**Fig. 2B**) was similar to the expected profile of 85–90% MCFA, determined in previous studies (Arkcoll 1988, Laureles et al. 2002, Kumar 2011). The oil content of coconut endosperm was quantified to be 31.41 ± 0.13% ($n = 3$; data not shown), hence verifying that this stage is representative of both active MCFA accumulation and TAG production, with oil content of mature endosperm reaching almost 65% (Gunstone et al. 2007). Following analysis of the coconut endosperm TAG content and composition it was hypothesized that the genes of interest, those involved in the MCFA lipid synthesis pathways, were being highly expressed at this developmental stage as the oil composition reflects active accumulation of MCFAs.

A transcriptomic approach was undertaken for the identification of genes involved in the metabolic processes of oil production. Illumina RNA-Seq technology was used to generate 473,728,278 short reads (236,864,139 pairs). The quality control analysis of the raw reads using BioKanga next generation sequencing did not show any concerns with the data quality (Supplementary Fig. S1) (Sayer et al. 2004, Ruffalo et al. 2012). Due to the scarcity of reference sequences for coconut palm available in the NCBI and other databases, *de novo* assembly was performed to create a platform of sequence information for the discovery of novel functional genes from coconut palm. After filtering (Supplementary Table S1), there were 271,312,922 reads (135,656,461 pairs, mean length of 100 bp; Supplementary Fig. S2), which contained no ambiguous bases. These filtered reads were used for the assembly of the *de novo* transcriptome using BioKanga v3.8.4 (Supplementary Table S2). The resulting assembly contained 838,319 contigs (total length: 278,262,217 bp). Scaffolding of the *de novo* assembly was constructed using BioKanga with a fragment length of 100–750 bp (Supplementary Table S3), with the final assembly containing 219,807 contigs (total length: 204,112,078 bp). Aligning reads using BioKanga (Supplementary Table S4) and allowing up to four substitutions, showed 90,611,882 pairs of 100 bp that uniquely mapped back to the *de novo* transcriptome (45,572,443 sense, 45,039,439 antisense), at approximately 32.6× coverage.

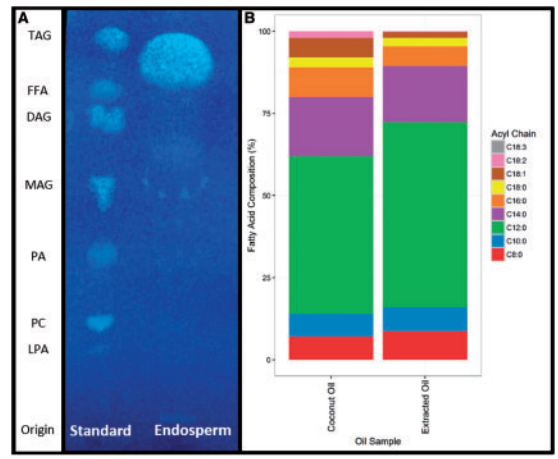

Fig. 2 Analysis of coconut oil composition. Extracted coconut oil from developing *Cocos nucifera* endosperm was analyzed by thin-layer chromatography (TLC) (A). The fatty acid composition of triacylglycerol (TAG) was compared between commercial coconut oil (Murray River Organics, Australia) and oil extracted from freshly harvested developing coconut endosperm (B). LPA, lysophosphatidic acid; PC, phosphatidylcholine; PA, phosphatidic acid; MAG, monoacylglycerol; DAG, diacylglycerol; FFA, free fatty acid; TAG, triacylglycerol.

Analysis of the *de novo* transcriptome assembly by the core eukaryotic genes mapping approach (CEGMA) (Parra et al. 2007) showed that 96% (238) of the 248 conserved eukaryotic transcripts completely matched the *de novo* transcriptome, with 99.6% (247) of transcripts at least partially matching (Supplementary Table S5). Further analysis using plantae

benchmarking universal single-copy orthologues (BUSCO) reference (Simão et al. 2015) revealed that 93.8% of the plantae reference sequences were identified in the assembled transcriptome (Supplementary Table S6), either as complete or fragmented sequences.

Following the assembly and quality analysis of the *de novo* coconut endosperm transcriptome, the transcriptome was initially used as a BLAST sequence database for the identification and confirmation of previously discovered and characterized genes from coconut, including the thioesterases (Jing et al. 2011, Yuan et al. 2013) and *LPAAT* (Knutzon et al. 1995). The identified transcripts from the assembled transcriptome were then aligned against the published sequences. In all cases, there was high consensus with the previously characterized genes and the sequence identified from the transcriptome (Supplementary Table S7). Importantly, both the identification of these sequences within the assembled *de novo* transcriptome and their high sequence consensus provided further confidence in the transcriptome being a powerful and reliable tool for the discovery of new genes of interest.

Identification of genes involved in lipid biosynthesis

The assembled coconut endosperm transcriptome was used as a BLAST reference to identify candidate transcripts of genes that are involved in lipid synthesis and assembly pathways. Of particular interest were acyltransferases as they are crucial to both the assembly and fatty acid composition of plant oils. BLAST analysis of the assembled coconut transcriptome revealed many candidate sequences. These candidate sequences were analyzed via the Transeq online tool (Rice et al. 2000) to identify open reading frames, which were then used for analysis via multiple sequence alignment (VectorNTI Advance® 11.5.4; Invitrogen, USA), from which a phylogenetic tree was generated (Supplementary Fig. S3). It was observed that each gene class forms individual clusters, with the candidate sequences located within their respective cluster. The identification of the *CnGPAT9* has been described previously [NCBI: KX235871; (Reynolds et al. 2017)].

Identification and characterization of functional *Cocos nucifera DGAT1*

The translated protein sequences of the *Cn*DGAT1 candidates identified (Supplementary Table S8) were compared to the *A.thaliana* and oil palm homologs through multiple sequence alignments to compare sequence homology (Supplementary Fig. S4), which revealed that the CnDGAT1 candidate sequences were 56–60% identical to the *At*DGAT1 homolog. It was also identified that the *Cn*DGAT1_V1 sequence was highly similar to the *Eg*DGAT1_V1 sequence, with 90% homology. The translated protein sequences were analyzed through BLAST, which confirmed the presence of the membrane-bound acyltransferase (MBOAT) domain, a characteristic feature of DGAT1 required for functionality (Turchetto-Zolet et al. 2016).

The function of the three *Cn*DGATs were tested through transient expression studies for their ability to increase TAG production, using the *N.benthamiana* infiltration assay. It has previously been shown that DGAT1 functions synergistically with WRI1 to further increase the TAG content in leaves (Vanhercke et al. 2013). This concept was explored here, with each DGAT being co-infiltrated with *At*WRI1 to both amplify increases in TAG content and evaluate gene functionality. It was observed that one of the isolated coconut DGAT1 candidates (*Cn*DGAT1_V1) was functional in *N.benthamiana*, based on the significantly increased TAG ($P = 0.0004$) and TFA ($P = 0.03$; data not shown) contents compared to the p19 control infiltrated leaf samples (**Fig. 3A**). Changes in the fatty acid profile provided further confirmation that the *Cn*DGAT1_V1 candidate was functional (**Fig. 3B**), with significant increases observed for both C16:0 ($P = 0.0008$) and C18:1 ($P = 0.03$) and also a significant decrease in the C18:3 content ($P = 0.005$), when compared to the p19 alone control.

Functionality was also confirmed by yeast complementation, using the *Saccharomyces cerevisiae* quadruple mutant (*are1*, *are2*, *dga1* and *lro1*) strain H1246 (Aymé et al. 2014). After 5 d incubating at 28°C on media complemented with oleic acid (C18:1), yeast growth was observed for *CnDGAT1*_V1 (pKR.B1) (**Fig. 3C**), and hence it could be confirmed that the *CnDGAT1*_V1 encodes a functional DGAT1.

Following confirmation of functionality for *Cn*DGAT1_V1, further experiments were conducted to investigate substrate specificity. The *Cn*DGAT1 fatty acid substrate specificity was determined through analyzing both the fatty acid composition of TAG and the TAG content of *N.benthamiana* infiltrated leaves, following expression with and without the co-expression of different type B thioesterases (**Fig. 4**). In most *FATB* combinations the fatty acid composition of TAG was not significantly different with the absence or presence of *CnDGAT1* expression, except for combinations involving *CnFATB3* or *UcFATB* expression. The total increase in MCFA accumulation was more pronounced in samples expressing *UcFATB*, with a more significant increase in C12:0 ($P < 0.0001$) than compared to C14:0 ($P = 0.0003$). Although the addition of *AtWRI1*+*CnDGAT1* led to a significant increase in oil production in all combinations, the most significant increase was associated with the expression of *UcFATB* ($P < 0.001$). Based on this data the fatty acid substrate preference of *CnDGAT1* is C12:0 > C14:0 > C16:0.

Identification and characterization of functional *Cocos nucifera PDAT*

Based on observations from previous studies, it was hypothesized that a functional PDAT from coconut would be a good candidate for improving the MCFA content of TAG by utilizing the MCFA in PC for incorporation into TAG, through the PC-DAG pathway, and also increase total TAG production (Dahlqvist et al. 2000, Ståhl et al. 2004, Zhang et al. 2009, J. Fan et al. 2013, Reynolds et al. 2017).

The translated protein sequences of the *Cn*PDAT candidates identified (Supplementary Table S8) were examined through multiple sequence alignments with *At*PDAT and the predicted *Eg*PDAT sequences (Supplementary Fig. S5). This sequence alignment revealed that all of the products were highly similar

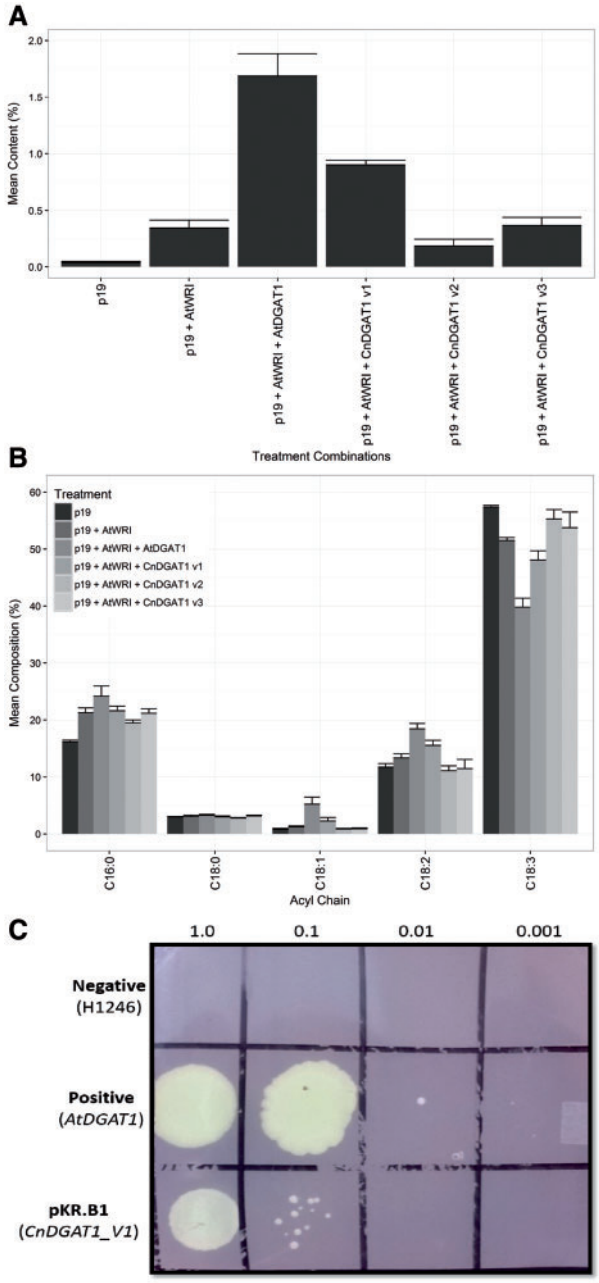

Fig. 3 (A) Examining *Cocos nucifera DGAT1* candidates for functionality through the quantification of changes in the accumulation of triacylglycerols (TAG), using the *Nicotiana benthamiana* transient expression assay. Error bars are representative of standard deviation ($n = 4$). (B) Analysis of major fatty acids in the total fatty acid profile from the *N.benthamiana* transient expression assay, associated with the testing of different diacylglycerol acyltransferase (*DGAT1*) candidates. Error bars are representative of standard deviation ($n = 4$). (C) Yeast (*Saccharomyces cerevisiae*) complementation assay of quadruple mutant strain H1246, investigating the functionality of DGAT1 through restoration of function. Transformants were fed C18:1Δ9 free fatty acids. Yeast transformants were resuspended in sterile water and diluted to OD600 = 1. Samples were further diluted in four consecutive dilutions, each at 1:10. From each dilution 5 µL was spotted on to YPD+FFA plates. The plates were incubated at 28°C for 3 d. Negative control, H1246::pYES2; Positive control, H1246::*AtDGAT1*; pKR.B1, H1246::*CnDGAT1*_V1. At, *Arabidopsis thaliana*; Cn, *Cocos nucifera*.

at both the N and C termini, but varied significantly through the middle of the proteins with variation in the length of deletions. It may be that the smaller products are alternative splice variants of the *Cn*PDAT_V1, which appears to be the full length transcript based on sequence homology with both the *At*PDAT (75%) and *Eg*PDAT (96%). The full length protein sequence was analyzed through BLAST, which confirmed the presence of the abhydrolase domain, a characteristic feature of *PDAT*. Based on the results of analyzing sequence homology and features, it provided confidence that the *CnPDAT*_V1 should encode a functional PDAT.

The PDAT function was tested using transient *N.benthamiana* infiltration assays. When *At*PDAT and *Cn*PDAT were tested individually for their ability to increase TAG production (**Fig. 5A**), the expression of either PDAT was associated with a significant increase in TAG compared to the p19 control ($P = 0.05$). In scenarios following the addition of *U.californica* 12:0-ACP thioesterase (*UcFATB*) and other acyltransferases (*Cn*LPAAT±*Cn*GPAT9), the further addition of *Cn*PDAT resulted in further increases in TAG production ($P = 0.002$). The co-expression of *Cn*PDAT also recovered the chlorotic phenotype previously associated with a build-up of MCFA-containing PC lipids (Reynolds et al. 2015). In contrast, a TAG increase was not observed in the scenario of *At*PDAT expression ($P = 0.3$), which is suspected to be a result of *At*PDAT not being capable of utilizing C12:0 as a substrate for TAG incorporation.

Through analysis of the fatty acid composition (**Fig. 5B**) of TAG, it was also demonstrated that in treatments of *UcFATB* expression, that the C12:0 content of TAG was significantly improved following the co-expression of *Cn*PDAT when compared to both *UcFATB* only and *UcFATB*+*At*PDAT ($P = 0.03$). However, when comparing treatment combinations of *UcFATB*+*Cn*LPAAT+*Cn*GPAT9, there was no significant difference in the C12:0 content of TAG ($P = 0.06$) following the addition of *Cn*PDAT. Although there was no difference in the C12:0 composition following the addition of *Cn*PDAT, the TAG content was significantly higher, suggesting that more TAG was being produced by the PC-DAG pathway. This concept was reaffirmed by the phospholipid fatty acid composition, where it's clear that C12:0-containing PC lipids were being used for TAG assembly. Analysis of the TFA composition (data not shown) revealed that the highest level of MCFA accumulation was achieved where the combination of *Cn*LPAAT+*Cn*GPAT9+*Cn*PDAT was being used for assembly ($P = 0.01$).

Identification and characterization of functional *Cocos nucifera* oleosin in leaf transient expression assays

The full length transcripts of *CnOleosin* candidates identified from the assembled transcriptome (Supplementary Table S8) were examined through multiple sequence alignment to determine similarity with *A.thaliana* homologs (Supplementary Fig. S6). Further examination of the protein sequences revealed that only one of the candidate sequences contained the characteristic proline-knot that defines an oleosin (Hsieh and Huang

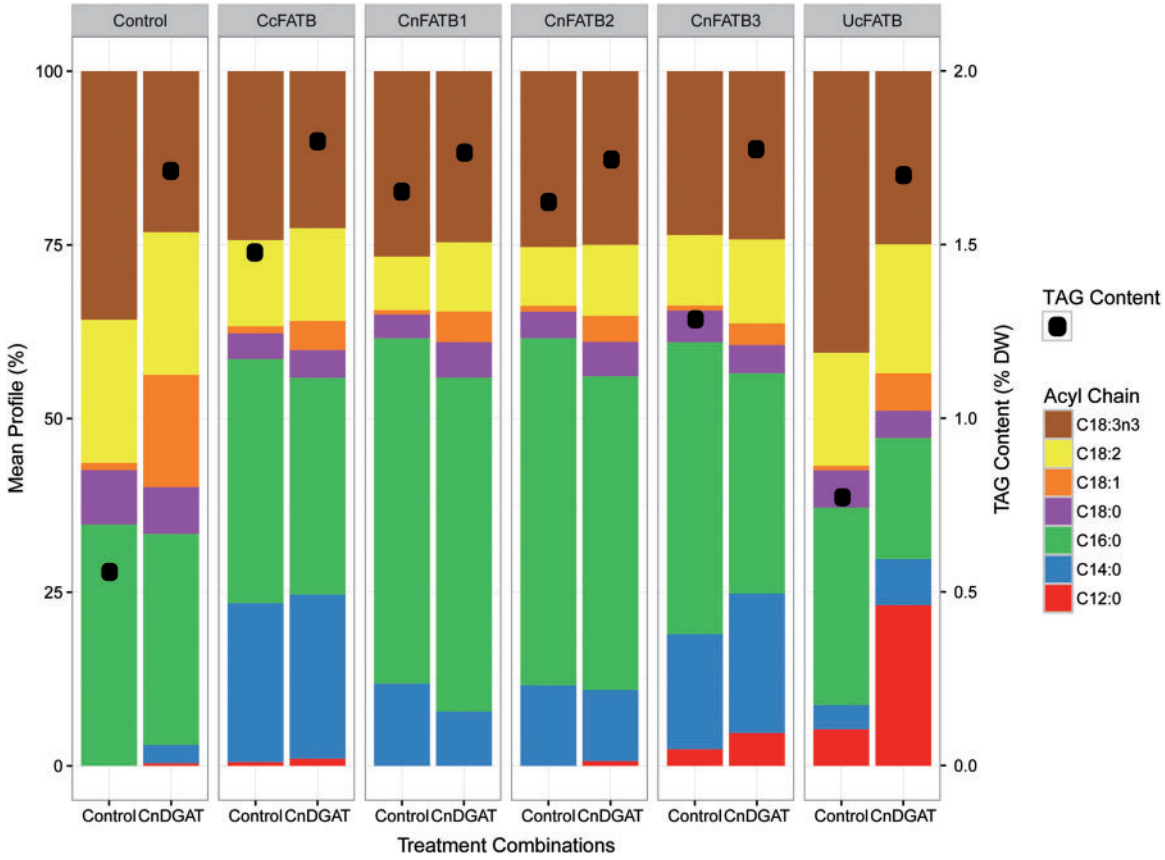

Fig. 4 Investigating the substrate specificity of *Cocos nucifera DGAT1* candidates for functionality through analyzing the changes in both the total triacylglycerol (TAG; block dots) content and the TAG fatty acid profile, using the *Nicotiana benthamiana* transient expression assay ($n = 4$). The results were grouped by the corresponding class B acyl-ACP thioesterase (*FATB*), where control refers to the expression of p19 only. On the x-axis, the treatments were grouped into respective diacylglycerol acyltransferase (*DGAT*) expression. Control refers to samples that were infiltrated with the corresponding *FATB* only (Control = p19 only). The annotation of CnDGAT represents those treatments that exhibited *CnDGAT1+AtWRI1* co-expression. At, *Arabidopsis thaliana*; Cc, *Cinnamomum camphora*; Cn, *Cocos nucifera*; Uc, *Umbellularia californica*.

2004), *Cn*Oleosin_V4. This candidate sequence was further analyzed using online protein property analysis tools (Lear and Cobb 2016), which revealed that the sequence was 49.3% hydrophobic and contained a 75 amino acid central hydrophobic region (Supplementary Fig. S7), a characteristic feature of oleosin which enables the formation of the hairpin-like structure that interacts with lipid droplets (Li and Fan 2009, Maurer et al. 2013). The *Cn*Oleosin_V4 protein had highest similarity with *At*OLEO4 and *At*OLEO5 with the sequences being 35.6% and 36.6% identical, respectively. Based on sequence homology and the peptide characteristics identified it was predicted that the *Cn*Oleosin_V4 should be functional.

To perform their function, oleosin proteins first localize in the endoplasmic reticulum (ER) where they then wrap lipid droplets as they emerge and later detach from the ER (Hsieh and Huang 2004, Huang and Huang 2017). We therefore investigated the subcellular-localization of *Cn*Oleosin-mCherry in a control background (p19 alone) or in tissue producing increased levels of TAG (p19+*At*WRI1+*At*DGAT1) by confocal microscopy (**Fig. 6**). In tissue expressing p19, lipid droplets were scarce and small (**Fig. 6A–D**). When expressed in this background, *Cn*Oleosin-mCherry localized to the ER and to these small lipid droplets (**Fig. 6E–H**), as expected. In tissue expressing p19+*At*WRI1+*At*DGAT1, lipid droplets were both more abundant and larger (**Fig. 6I–L**). In this background, *Cn*Oleosin-mCherry localized in the ER and in these larger droplets, where it was found to coat the lipid droplets (**Fig. 6M–T**). These findings suggest that *CnOleosin* encodes a functional oleosin.

From the same plants, leaves were harvested for quantification and analysis of TAG accumulation (**Fig. 7A**). As visually observed from confocal microscopy, the expression of *Cn*Oleosin alone did not produce a significant increase in TAG production when compared to the p19 only control ($P = 0.08$). However, in a high oil background of *At*WRI1+*At*DGAT1 expression the additional expression of *Cn*Oleosin resulted in a significant increase of TAG accumulation ($P = 0.002$).

The *Cn*Oleosin was also examined for the function of lipid droplet stabilization and protection, via analyzing the incorporation of [^{14}C] into infiltrated *N.benthamiana* leaves (**Fig. 7B**). In the control sample, expression of p19 alone, the [^{14}C] acetate incorporation into TAG did not exceed 500 disintegrations per minute (dpm) during the 15 min time course.

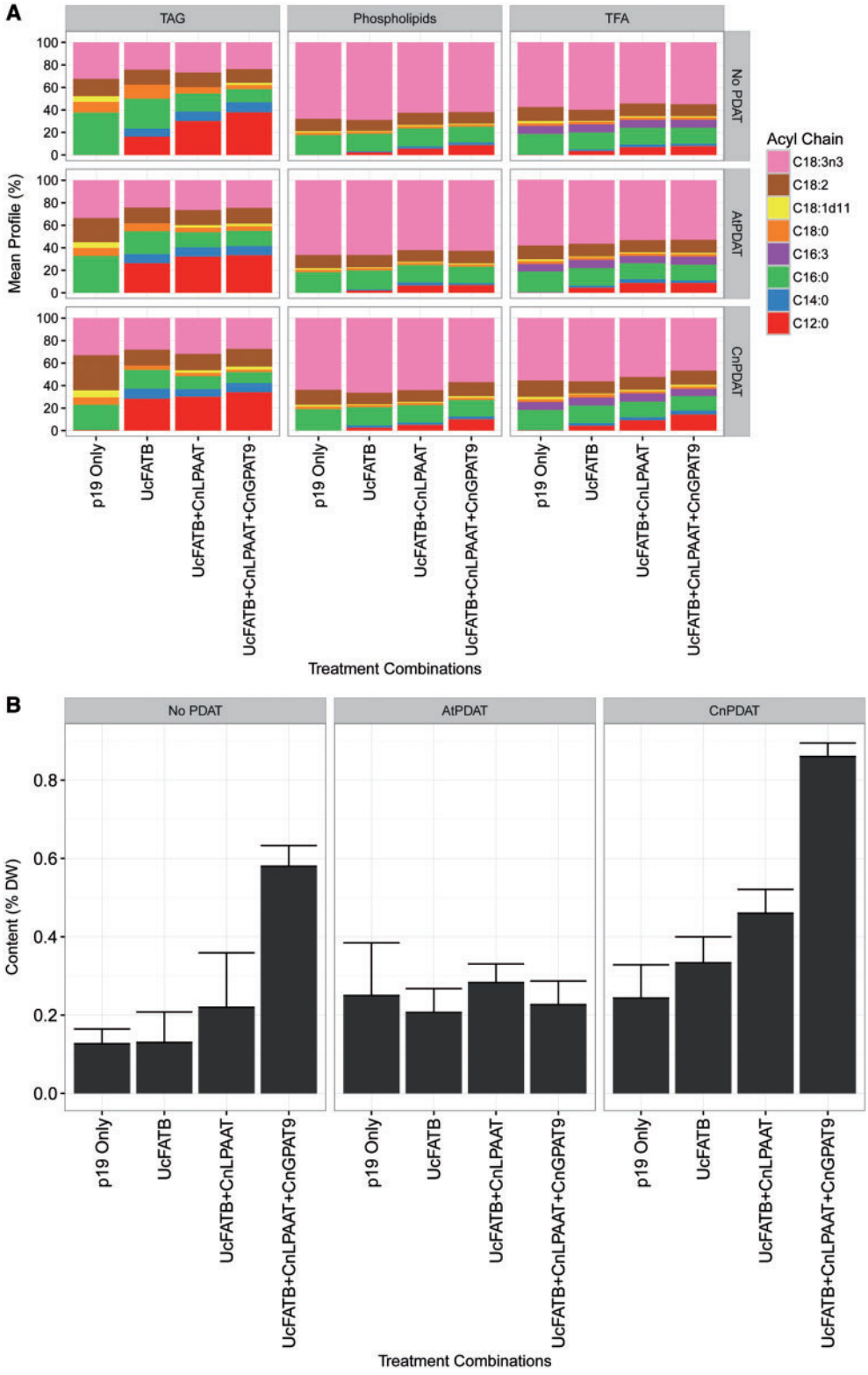

Fig. 5 (A) Investigating the ability of phosphatidylcholine-diacylglycerol acyltransferase (*PDAT*) to increase production of triacylglycerol (TAG) in the leaves of *Nicotiana benthamiana*, following 5 d of expression. Error bars are representative of standard deviation ($n = 3$). Treatments were grouped by their respective *PDAT* expression including no *PDAT*, *PDAT* from *Arabidopsis thaliana* (*AtPDAT*) and *Cocos nucifera* (*CnPDAT*). (B) Investigating the changes in fatty acid composition of different lipid groups associated with the overexpression of *PDAT*, using the *N.benthamiana* transient expression assay. The lipid pools analyzed included were TAG, phospholipids and total fatty acids (TFA). Treatments were grouped by their respective *PDAT* expression including no *PDAT*, *PDAT* from *Arabidopsis thaliana* (*AtPDAT*) and from *C.nucifera* (*CnPDAT*) ($n = 3$).

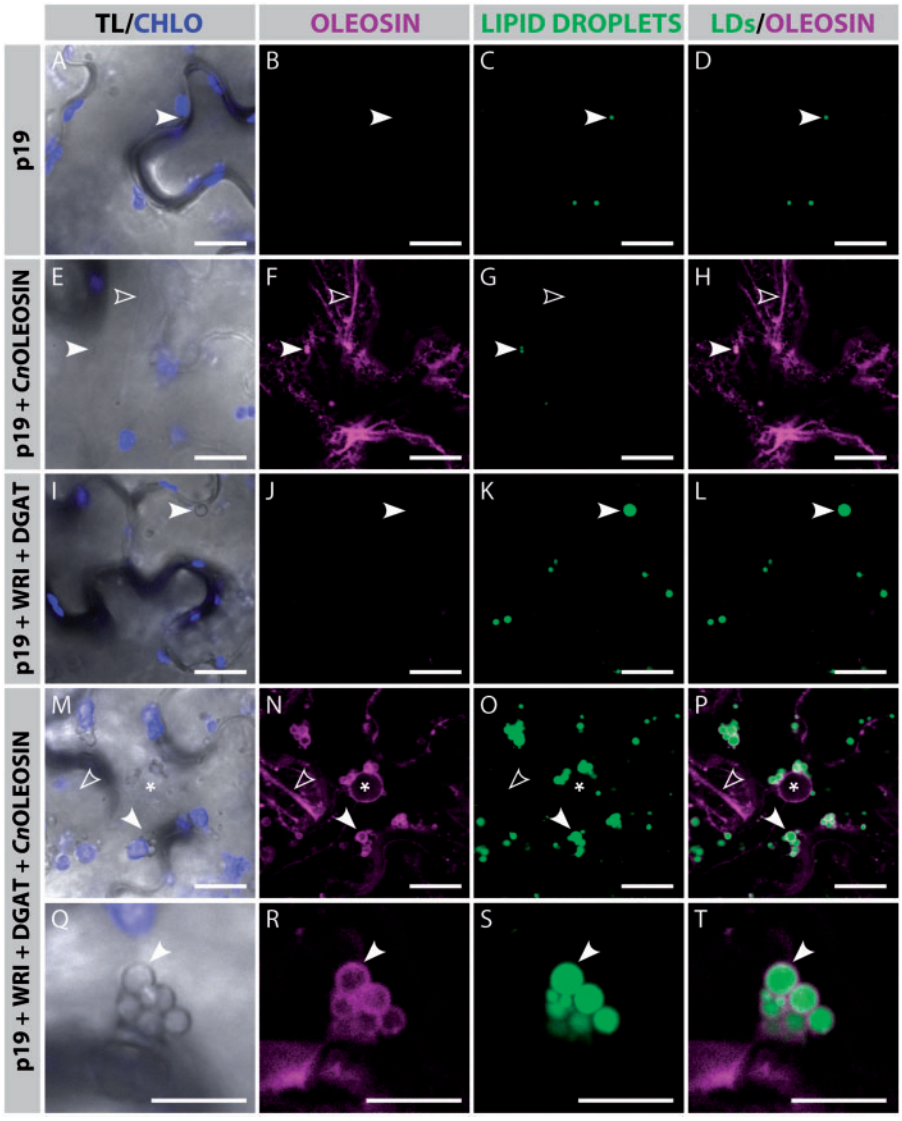

Fig. 6 Confocal microscopy showing the localization of CnOLEOSIN-mCherry in epidermal cells of *Nicotiana benthamiana* leaves, 3 d post-infiltration (DPI). Lipid droplets were stained with BODIPY-505/515 and are shown in green, chloroplast autofluorescence in blue and *Cn*OLEOSIN-mCherry in magenta. In all panels, white arrowheads point at lipid droplets, while empty arrowheads highlight *Cn*OLEOSIN signal in the endoplasmic reticulum (ER), and stars highlight the nucleus of the cell (M–P). In leaves infiltrated with p19 only (A–D), the lipid droplets were very small. In leaves infiltrated with p19+*Cn*OLEOSIN-mCherry (E–H), *Cn*OLEOSIN localized to the ER and lipid droplets, and the size of the lipid droplets was not increased compared to p19 alone. In leaves infiltrated with p19+*At*WRI1+*At*DGAT1 (I–L), the lipid droplets increased in both size and abundance. Finally, in leaves infiltrated with p19+*At*WRI1+*At*DGAT1+*Cn*OLEOSIN-mCherry (M–T), *Cn*OLEOSIN was still detected in the ER (M–P) and was also found to coat the lipid droplets (Q–T). TL is transmitted light. Scale bars represent 20 µm except the lower row where they represent 10 µm. At, *Arabidopsis thaliana*; Cn, *Cocos nucifera*.

There was no significant difference in the amount of [^{14}C] acetate incorporation following the co-expression of *CnOleosin* alone ($P = 0.07$). In leaf samples exhibiting the expression of p19+*AtWRI1*+*AtDGAT1*, TAG accumulation continued to occur in the first 10 min of the assay, followed by rapid degradation. Although the additional expression of *CnOleosin* did not result in further increasing the production of TAG, its expression enabled a higher final TAG content through reducing TAG turnover. The demonstrated ability of this protein to both increase TAG accumulation and improve TAG stability, along with its physical co-localization with lipid droplets provide confirmation of a functional *oleosin*.

Identification and characterization of functional *Cocos nucifera FATA* in transient leaf assays

Although MCFA is the main constituent of coconut oil, long chain fatty acids are also an important component representing 15% of the total fatty acid (TFA) composition (Arkcoll 1988, Laureles et al. 2002, Kumar 2011, Dussert et al. 2013). The acyl-ACP thioesterase class A (*FATA*) therefore needs to be considered as playing an essential role in the final

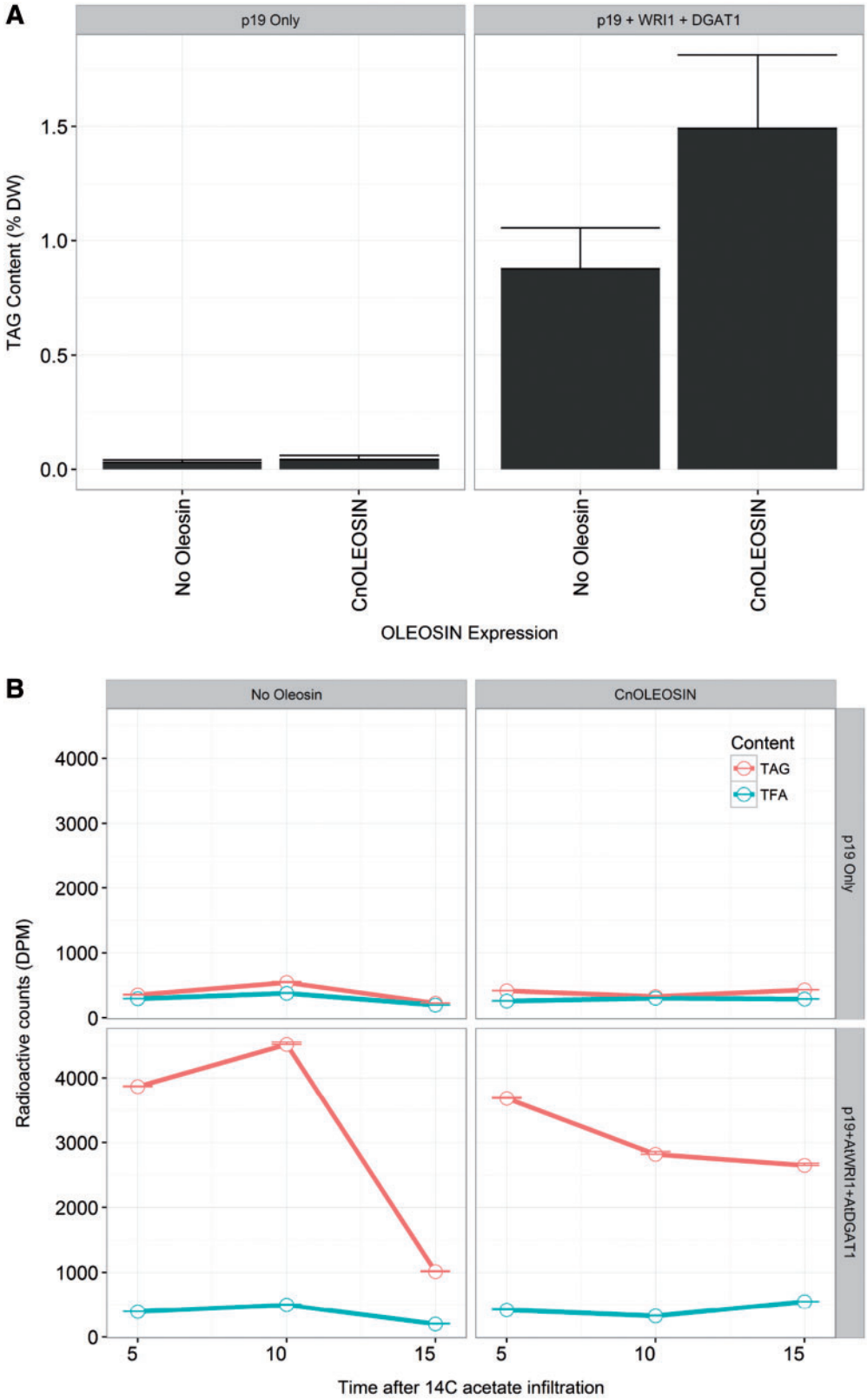

Fig. 7 (A) Investigating the ability of *CnOLEOSIN* to enable greater accumulation of triacylglycerol (TAG) in the leaves of *Nicotiana benthamiana*, following 5 d of expression. The *WRI1* and *DGAT1* genes were both from *Arabidopsis thaliana*. Error bars are representative of standard deviation ($n = 6$). (B) Investigating the ability of *CnOLEOSIN* in protecting lipid droplets by minimizing the turnover of TAG, by measuring the incorporation of [14C] labelled acetate into TAG. Error bars are representative of standard error ($n = 3$). At, *Arabidopsis thaliana*; Cn, *Cocos nucifera*.

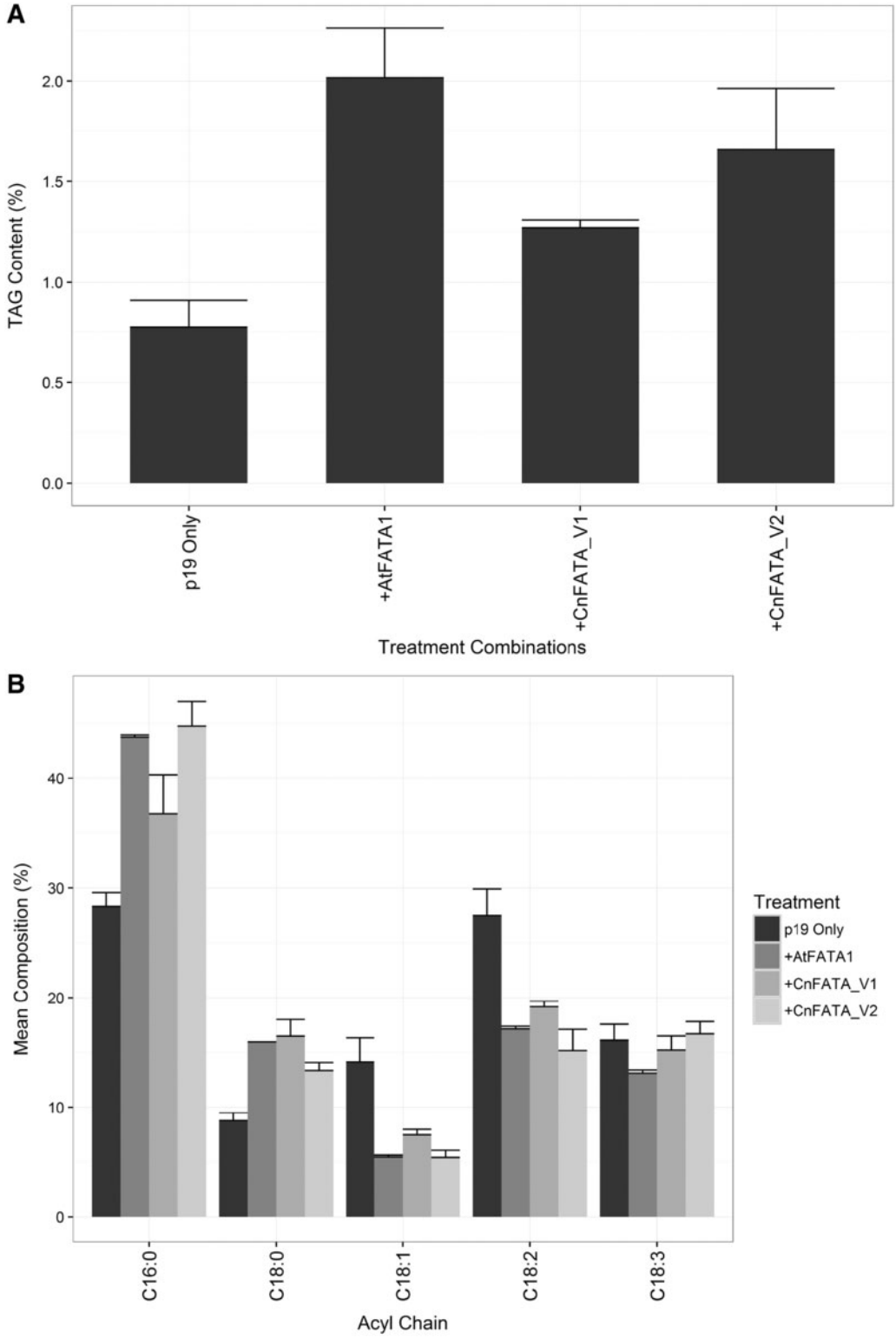

Fig. 8 (A) Examining *Cocos nucifera* class A fatty acid thioesterase (*FATA*) candidates for functionality through the quantification of changes in the accumulation of triacylglycerols (TAG) content, using the *Nicotiana benthamiana* transient expression assay. The *Arabidopsis thaliana* homolog (*AtFATA1*) was used as a positive control. The sample 'p19 Only' refers to the background infiltration of *AtWRI1* and *AtDGAT1*. Error bars are representative of standard deviation ($n = 4$). (B) Analysis of the major fatty acids in the TAG fatty acid composition, associated with the testing different *FATA* candidates, using the *N.benthamiana* transient expression assay. All treatments were performed in the background of *AtWRI1* and *AtDGAT1* co-expression. The *A.thaliana* homolog (*AtFATA1*) was used as a positive control. Error bars are representative of standard deviation ($n = 4$). At, *Arabidopsis thaliana*; Cn, *Cocos nucifera*.

determination of the composition of coconut oil. The full length transcripts for *Cn*FATA candidates identified from the assembled transcriptome (Supplementary Table S8) were then examined through multiple sequence alignment with the *A.thaliana* homolog to determine sequence homology (Supplementary Fig. S8). Further examination of the protein sequences revealed that both of the candidate sequences contained the acyl-ACP thioesterase domain. The sequence similarity of *At*FATA with the identified coconut candidates *Cn*FATA_V1 (65%) and *Cn*FATA_V2 (50%) provided further evidence that they may be functional FATA proteins.

The FATA function was tested through a transient expression study for its ability to not only increase TAG production but also modify the fatty acid composition of TAG using the *N.benthamiana* infiltration assay. The *CnFATA* candidates were co-infiltrated with *AtWRI1* and *AtDGAT1* to model an increased oil background. The *A.thaliana* homolog (*AtFATA1*: NM_113415) was used as the positive control for direct comparison and to enable analysis of functionality. With the co-expression of both *CnFATA* candidates a significant increase in TAG production (**Fig. 8A**) was achieved ($P < 0.0005$), although TAG production was significantly higher with the co-expression of *CnFATA*_V2 when compared to *CnFATA*_V1 ($P = 0.047$). It was also determined that the TAG content following *CnFATA*_V2 expression was similar to the TAG content achieved with *AtFATA1* co-expression ($P = 0.1$).

For each combination the fatty acid composition of TAG was also analyzed (**Fig. 8B**). The changes in the fatty acid composition following expression of either *CnFATA* candidate were similar to that observed in the positive control with *AtFATA1* overexpression. In all three scenarios, compared to the *AtWRI1* + *AtDGAT1* control, a significant increase was observed for both C16:0 and C18:0 ($P \leq 0.005$), as well as significant decreases in both C18:1 and C18:2 ($P \leq 0.005$). Together these results confirm the functionality of both *Cn*FATA candidates, although *Cn*FATA_V2 outperforms *Cn*FATA_V1 in terms of ability to increase TAG production.

Discussion

Deep sequencing strategies were undertaken for the *de novo* assembly of a transcriptome from developing endosperm of *C.nucifera* (coconut palm), which has enabled the identification of functional genes that are involved in the metabolic processes of TAG biosynthesis. This transcriptome has been successfully utilized for the identification of candidate genes involved in fatty acid profile determination (*FATA*) and that are essential to both the assembly (*GPAT9*, *DGAT1* and *PDAT*) and protection (*oleosin*) of TAG, which were then examined using heterologous expression assays. Due to the requirement of these genes for increased TAG production, they are likely to be the targets of future metabolic engineering studies targeted towards increasing vegetable oil yields. The further understanding of how coconut palm achieves such high oil yields, at 67% of the seed weight (Padolina et al. 1987, Arkcoll 1988), may be crucial to future directions of increasing global vegetable oil production. As it has been demonstrated in this study, the assembled transcriptome now provides a platform for further discovery of important lipid synthesis related genes.

For the first time, this study identified and confirmed functionality for a *DGAT1* from coconut palm. The transient *N.benthamiana* assay demonstrated that the *At*DGAT1 outperformed the *Cn*DGAT1v1 in respect to ability to increase TAG and TFA contents. It was also observed in the yeast complementation assay that the positive control (*AtDGAT1*) had a much faster growth rate compared to *CnDGAT1*_V1. The reduced growth rate of *CnDGAT1*_V1 may be attributed to its reduced capacity of utilizing C18:1 for TAG assembly, evident by the lower C18:1 content ($P = 0.01$) in *N.benthamiana* leaves compared to *AtDGAT1*. Both of these results could be explained by the differences in substrate specificities, and hence the *Cn*DGAT1v1 may require an increased flux of MCFA in order to perform optimally (Reynolds et al. 2017). It would be expected that the *Cn*DGAT1 would exhibit a preference for MCFA based on the high MCFA content of coconut oil (Padolina et al. 1987, Arkcoll 1988, Laureles et al. 2002). The fatty acid substrate specificity of *CnDGAT1* was investigated through utilization of the transient *N.benthamiana* infiltration assay. TAG fatty acid compositions were examined following the expression of *CnDGAT1* with and without the co-expression of different class B acyl-ACP thioesterases. In combinations involving *CnFATB3* or *UcFATB* expression, significant increases in TAG content and MCFA incorporation were observed following the addition of *CnDGAT1*. Fatty acid substrate preference of *CnDGAT1* was determined on significant increases in fatty acid composition increase and hence described as C12:0 > C14:0 > C16:0. This fatty acid substrate preference could be validated through the preparation of microsomes from a transformed yeast mutant strain H1246 expressing the *CnDGAT1*, followed by subsequent radiolabeled feeding of DAG and different non-labelled acyl-CoAs. This would allow accurate quantification of fatty acid utilization and hence preference of *CnDGAT1* for TAG production (Liu et al. 2012, Zhou et al. 2013, Aznar-Moreno et al. 2015). This system can be used to reveal the biochemistry of TAG assembly in coconut endosperm. Investigating the ability of *CnDGAT1* to efficiently utilize C8:0 or C10:0 acyl-CoAs for TAG assembly may reveal that other DGATs are important in determining the composition of coconut oil.

With the high abundance of C18:3 in the fatty acid profile of the *N.benthamiana* leaves, it was reasoned that the major metabolic pathway in these leaves for TAG production is the PC-DAG pathway. In this study, the first *PDAT* from coconut palm was identified and investigated for functionality. It was demonstrated that TAG accumulation can be increased by the overexpression of *Cn*PDAT. It was also previously hypothesized that the expression of a substrate-specific PDAT may assist in the recovery of the chlorotic phenotype observed following the build-up of MCFA-containing PC lipids associated with the expression of *FATB*+*CnGPAT9*+*CnLPAAT* (Reynolds et al. 2017), through the conversion of these PC lipids into TAG. In previous work that investigated the ability of *Eg*DGAT1 to use C12:0 for assembly (Reynolds et al. 2017), it eliminated the accumulation

of C12: 0 in phospholipids as TAG assembly occurred by the DAG-derived pathway before it could be converted to PC (Lung and Weselake 2006, Yen et al. 2008, Turchetto-Zolet et al. 2011, Bates et al. 2012). Similarly, the overexpression of *Cn*PDAT was able to recover the chlorotic phenotype previously observed (data not shown). However, in the *PDAT* context of this study, there is still C12:0 in phospholipids as the MCFA-containing PC lipids are being used as a substrate by PDAT for TAG assembly.

TAGs are present in subcellular spherical lipid droplets of approximately 0.5–2 μm in diameter, surrounded by a matrix of phospholipids and structural proteins, for example oleosin (Hsieh and Huang 2004, Chapman and Ohlrogge 2012, Murphy 2012). In this study, a new *oleosin* gene from coconut palm was identified and functionally confirmed. The visualization of mCherry fluorescence revealed that *Cn*Oleosin expression was localized in the ER and as a surrounding layer of lipid droplets. This result is consistent with the understanding that the oleosin protein is contained within the ER and becomes part of the phospholipid layer surrounding lipid droplets (Hsieh and Huang 2004, Huang and Huang 2017). This demonstrates that the expression of *oleosin*, or other lipid droplet associated proteins, will be important for future studies that aim to both increase oil production and stabilize the synthesized oil by providing protection against degradation (Chapman et al. 2012).

The specificity of acyl-ACP thioesterases play an important role in determining the fatty acid composition of plant oils (Salas and Ohlrogge 2002). The relative activities of the acyl-ACP thioesterases and acyl-ACP desaturases are responsible for the export of a variety of FAs exported from the plastid, including saturated, mono- and poly-unsaturated FAs (Bates et al. 2013). Here, *FATA*s from coconut palm are first described, showing that *CnFATA* overexpression resulted in a TAG profile predominantly composed of saturated FAs (60%) compared to a more unsaturated TAG profile in the p19+*AtWRI1*+*AtDGAT1* control (40%). This indicates that the increased flow of saturated FAs are efficiently utilized for TAG assembly, assisted by the additional pull capability created by *AtDGAT1* overexpression (Vanhercke et al. 2013, El Tahchy et al. 2017). Therefore, the overexpression of thioesterases could be used to enable not only fatty acid profile modifications but to assist in increasing oil yields.

Currently, crop sources of MCFA-enriched oils are supplied predominantly by the production of coconut palm and oil palm (both palm oil and palm kernel oil which are derived from the mesocarp and endosperm, respectively) (Arkcoll 1988). With the continued growth in the world population there will be a correlating increased demand for both food and industrial products derived from vegetable oils. Methods of increasing global oil production have previously been proposed including the expansion of agricultural land, and increasing the yields of current oilseed crops through conventional breeding or genetic technology approaches (Kinney 1997, Kempken and Jung 2009, Baud and Lepiniec 2010, Bates et al. 2013, Vanhercke et al. 2014b, Barcelos et al. 2015). The expansion of land usage for vegetable oil production would be associated with various negative consequences, such as having less available land for the production of other food crops, such as wheat (Gunstone et al. 2007, Murphy 2007). With these reasons considered, it provides motives for the transition of MCFA-enriched oils into new crops, whether that may be through traditional oilseed crops or the further development of leaf oil production. In this study, the components needed for the development of a MCFA-enriched oil have been identified, although it remains to be assembled into a complete pathway as previously demonstrated (Reynolds et al. 2017).

Materials and Methods

Plant material and harvest

Coconut (*C.nucifera*) fruits were harvested at approximately 3 months old, from coconut palm trees in Parramatta Park, Cairns, Queensland, Australia. The harvested coconuts were transported to Canberra at 4°C. For optimal RNA integrity it was important that the coconut seed coat was not cracked, to ensure that there was no oxidative degradation of the coconut endosperm material.

RNA extraction using cetyl trimethylammonium bromide

The fresh coconuts were cracked open using a hammer. The endosperm flesh was immediately harvested and frozen in liquid nitrogen. Samples were stored at −80°C before further processing, for long-term preservation of RNA integrity.

For the rapid cetyl trimethylammonium bromide (CTAB)-based procedure, 900 μL of extraction buffer (2% CTAB w/v, 2.5% PVP-40 w/v, 2M NaCl, 100 mM Tris–HCl pH 8.0, 25 mM EDTA pH 8.0 and 2% of β-mercaptoethanol added just before use) was heated at 65°C in a microcentrifuge tube. Samples (150 mg) were powdered in liquid nitrogen using a mortar and pestle, then added to the extraction buffer and incubated at 65°C for 10 min. An equal volume of chloroform:isoamyl alcohol (24:1, v/v) was added and the tube was inverted vigorously and centrifuged at 11,000×*g* for 10 min at 4°C. The supernatant was recovered and a second extraction with chloroform:isoamyl alcohol was performed. The supernatant was transferred to new microcentrifuge tube and LiCl (3 M final concentration) was added. The mixture was incubated on ice for 30 min and RNA was selectively pelleted after centrifugation at 21,000×*g* for 20 min at 4°C. The pellet was resuspended in 500 μL of SSTE buffer (10 mM Tris–HCl pH 8.0, 1 mM EDTA pH 8.0, 1% SDS and 1 M NaCl) pre-heated at 65°C, an equal volume of chloroform:isoamyl alcohol was added and the mixture was centrifuged at 11,000×*g* for 10 min at 4°C. The supernatant was transferred to new microcentrifuge tube and the RNA was precipitated overnight at −20°C with 0.7 volume of cold isopropanol and then centrifuged at 21,000×*g* for 15 min at 4°C. The pellet was washed with ethanol (70%), dried and resuspended in DEPC-water. The RNA yield and quality was then analyzed with a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Australia).

Coconut endosperm transcriptome assembly and gene amplification

The extracted RNA samples ($n = 2$) from the developing coconut endosperm were sent to a commercial sequencing provider, Australian Genome Research Facility (AGRF: Melbourne, VIC, Australia) for sequencing using a single lane of Illumina HiSeq 2500 (access: https://doi.org/10.4225/08/5ab330902ddb8). QC and read filtering was performed using BioKanga (v3.8.4; NGSQC and filter), discarding reads that overlapped less than 50% with any other read or contained ambiguous bases. Assembly was performed using BioKanga (v3.8.4; assemb) using default parameters, with an initial overlap of 150 bp with a final overlap of 35 bp (PE) and 25 bp (SE). Scaffolding was performed using BioKanga (v3.8.4; scaffold) using default parameters with an insert length of 100–750 bp. The scaffolded *de novo* assembly was assessed using CEGMA (Parra et al. 2007, Parra et al. 2009) and BUSCO (Simão et al. 2015). Back alignment was performed using BioKanga (v3.8.4; align) using 4% of read lengths as substitutions and 1% of read length as ambiguous bases,

with any orphans treated as single ended reads (Supplementary Tables S1–S4).

The *de novo* assembled transcriptome was used as a BLAST platform using previously characterized genes to identify similar genes from *C.nucifera*. Identified candidate sequences were analyzed using the online Transeq tool (European Molecular Biology Laboratory-European Bioinformatics Institute 2018) to enable identification of the full length open reading frames. Following the identification of candidate genes, cDNA was synthesized from RNA (First Strand cDNA Synthesis; New England Biolabs, USA). Primers were then designed for full length amplification of the transcripts using Phusion-PCR (Phusion® High-Fidelity DNA Polymerase; NEB, Ipswich, MA, USA).

Syringe-mediated *Agrobacterium* infiltration of *Nicotiana benthamiana* leaves

Nicotiana benthamiana plants were grown at 24°C in a growth room with 16 h days (250 μmol $m^{-2}s^{-1}$). The infiltrations were performed using 5-week-old plants. Transient expression in *N.benthamiana* leaves was performed as previously described (Wood et al. 2009), with some minor modifications. *Agrobacterium tumefaciens* (AGL1) cultures containing the gene coding for the p19 viral suppressor protein and the chimeric gene(s) of interest were mixed such that the final OD_{600} of each culture was equal to 0.1 prior to infiltrations. The bacterial pellets were resuspended in an infiltration buffer containing 10 mM MES pH 5.7, 10 mM $MgCl_2$ and 100 μM acetosyringone. The samples being compared were randomly located on leaves, with a p19 control infiltrated for each plant. After infiltrations, the *N.benthamiana* plants were grown for a further 5 d before leaves were harvested, snap frozen in liquid nitrogen and then freeze-dried (Flexi-Dry MP: FTS Systems, Stone Ridge, NY, USA), typically overnight before processing. For samples that were used for confocal microscopy, fresh disc samples were harvested 3 d post-infiltration and processed immediately.

Bligh and dyer total lipid extractions

Total lipid extraction protocol was followed for the extraction of lipids from freeze-dried *N.benthamiana* leaves. Freeze-dried leaf tissue samples were first weighed using an analytical balance (AE100: Mettler, Port Melbourne, VIC, Australia) and then ground to powder in a microcentrifuge tube containing a metallic ball using a Reicht tissue lyser (QIAGEN, Germantown, MD, USA) for 3 min at 20 Hz. An internal standard of triheptadecanoin (TAG 51:0; Sigma, USA) was added for the quantification of both TFA and TAG contents. Chloroform:methanol (2:1, v/v) was added and mixed for a further 3 min on the tissue lyser before the addition of 0.1 M KCl (1:3, v/v). The sample was then mixed for a further 3 min before centrifugation (5 min at 14,000×*g*), after which the lower lipid phase was collected. The remaining phase was washed once with chloroform, and the lower phase extracted and pooled with the earlier extract. Lipid phase solvent was then evaporated completely using nitrogen gas flow and the lipids were then resuspended in 2 μL chloroform per mg of the leaf original dry weight. Lipid extractions from *C.nucifera* endosperm were performed as described above.

Thin-layer chromatography

Single phase TLC. TAG and polar lipids were fractionated by TLC (Silica gel 60, MERCK, Darmstadt, Hesse, Germany) in a solvent system of hexane:diethyl-ether:acetic acid (70:30:1, v/v/v), visualized by spraying primuline (Sigma, USA) 5 mg/100 mL acetone:water (80:20, v/v) and exposing plate under UV light at a wavelength of 254 nm ('UV Cabinet 4': CAMAG, Muttenz, Switzerland). All TLC analyses were performed in glass TLC developing tanks (Sigma, USA). TLC analysis was primarily used for the identification of TAG species from lipid extraction samples, enabling the determination of the total TAG content of each sample. The TAG fractions were scraped from the TLC plates and collected for transesterification.

Dual-phase TLC. Total lipids were first fractionated by TLC in a solvent system of chloroform:methanol:acetic acid:water (68:22:6:4, v/v/v/v), which was run until the solvent front was approximately halfway through the plate. The lipids were then further separated in a second solvent system of hexane:diethyl-ether:acetic acid (70:30:1, v/v/v), which was run until the solvent front had run through the entire length of the plate. The utilization of the dual-phase TLC for the analysis of lipid extractions enables the total separation of the lipid classes.

Transesterification of lipids by methylation using acidified methanol

Fatty acid methyl esters (FAMEs) of total lipids were produced by incubating the dry leaf sample or extracted lipid in 1 N methanolic-HCl (Supelco, Bellefonte, PA, USA) at 80°C for 3 h. Following the incubation phase, 500 μL of 0.1 M NaCl and 300 μL of hexane were added to the sample, and mixed via shaking with a 'Multi Tube Vortex Mixer' (Ratex Instruments, Australia) for 5 min. The hexane, containing FAMEs, fraction was then separated by centrifugation at 1,700×*g* for 2 min. The hexane phase was then transferred to inserts for analysis via gas chromatography-flame ionization detection (GC-FID).

Gas chromatography-flame ionization detection

FAMEs were analyzed by an Agilent 7890A gas chromatograph coupled with flame ionization detector (GC-FID; Agilent Technologies, Palo Alto, CA, USA), on a BPX70 column (30 m, 0.25 mm inner diameter, 0.25 μm film thickness, SGE) essentially as described previously (Zhou et al. 2011), except for the column temperature program. The column temperature was programmed at an initial temperature of 100°C holding for 3 min, ramping to 240°C at a rate of 7°C/min and holding for 1 min. NuChek GLC-426 was used as the external reference standard. Peaks were integrated with Agilent Technologies ChemStation software (Rev B.04.03). All analyses of calculating statistical significance were performed using a student's *t*-test, assuming equal variances within individual treatments.

Staining and imaging of lipid droplets

Fresh leaf discs were harvested from plants 3 d post-infiltration. The leaf discs were cut in half, with one half being stained with 2 μg/mL BODIPY 505/515 (ThermoFisher Scientific, Scoresby, VIC, Australia) in 50 mM PIPES pH 7.0 for 10 min and washed in 50 mM PIPES pH 7.0. The other half was used as a negative control for the BODIPY stain and was immediately placed in 50 mM PIPES pH 7.0. Leaf discs were then imaged using a Leica, Wetzlar, Hesse, Germany SP8 confocal laser-scanning microscope (Leica microsystems) equipped with a 20× water immersion objective [NA = 0.75] and the LasX software (version 3.3.0). Imaging was done sequentially. In the first track, signals from BODIPY (excitation 505 nm; emission 510–540 nm) and chloroplasts (excitation 650 nm; emission 670–690 nm) were collected, together with the transmitted light. In the second track, mCherry (excitation 587 nm; emission 600–630 nm) was imaged alone.

In vitro [^{14}C] acetate feeding and pulse chase assays

[^{14}C] acetate solution (1 μci/μL) was made in the same infiltration buffer used to infiltrate the genes in *N.benthamiana*. [^{14}C] acetate was infiltrated into the same leaves at 4 d post-infiltration of the genes. At 3-h post-infiltration of [^{14}C] acetate, leaf discs were pooled and lipid biosynthesis was stopped with the addition of 100 μL chloroform. Total lipid extraction and TLC fractionation were performed as described above. The TLC plate was exposed to phosphor imaging screens overnight and analyzed by a Fujifilm FLA-5000 Phosphorimager (GE Healthcare Life Sciences, Pittsburgh, PA, USA). Radiolabeled lipid spots were measured using a Beckman-Coulter Ready Safe liquid scintillation cocktail and Beckman-Coulter LS 6500 Multipurpose Scintillation Counter (Beckman Coulter, Brea, CA, USA).

DGAT assay in *Saccharomyces cerevisiae* H1246

Saccharomyces cerevisiae strain H1246 (Sandager et al. 2002) is completely devoid of *DGAT* activity and lacks TAG and sterol esters as a result of knockout mutations in four genes (*DGA1*, *LRO1*, *ARE1* and *ARE2*). The addition of free fatty acid (1 mM C18:1$^{\Delta 9}$) to H1246 growth media is toxic in the absence of *DGAT* activity. Growth on such media can therefore be used as an

indicator or selection for the presence of DGAT activity in this yeast strain. A functional DGAT will restore the capability of the mutant yeast strain to accumulate TAG, and hence restores growth. *Saccharomyces cerevisiae* H1246 was transformed with the pYES2 construct (negative control), a construct encoding *A.thaliana DGAT1* in pYES2 (positive control), or a construct encoding *C.nucifera DGAT1* in pYES2 (designated pKR.B1). Transformants were fed C18:1$^{\Delta 9}$ free fatty acids.

Yeast transformants were resuspended in sterile water and diluted to OD_{600} = 1. Samples were further diluted in four consecutive dilutions, each at 1:10. From each dilution 5 µL was spotted on each of the plates (YPD and YPD+FFA). The plates were incubated for 3 d at 28°C, before scoring growth.

Yeast media details. **Liquid media:** Synthetic defined minimal dropout medium lacking uracil (SD Medium-URA: MP Biomedicals, Australia). Media contains 0.2% yeast nitrogen base, 0.5% ammonium sulfate and 2% glucose. Media was sterilized by autoclaving at 121°C for 15 min.

Yeast growth media (YPD): Synthetic defined minimal dropout medium lacking uracil (SDA Medium-URA: MP Biomedicals, Seven Hills, NSW, Australia). Media contains 0.2% yeast nitrogen base, 0.5% ammonium sulfate, 2% glucose and 1.7% agar. The media was sterilized by autoclaving at 121°C for 15 min, then allowed to cool. Following the addition of 0.01% NP-40 (Sigma, St. Louis, MO, USA), 40 mL of agar media was poured into each plate. Plates were stored at 4°C until required.

Functionality media (YPD+FFA): The media was prepared similar to the standard growth media. Instead of glucose, the media contained 2% galactose and 1% raffinose. Before pouring plates, 1 mM C18:1$^{\Delta 9}$ was also added to the media. Plates were stored at 4°C until required.

Accession Numbers

CnDGAT1: MH158653; CnPDAT: MH158654; CnFATA.1: MH158655; CnFATA.2: MH158656; CnOleosin: MH158657.

Supplementary Data

Supplementary data are available at PCP online.

Acknowledgments

Thanks to Katharina Schulte (James Cook University, Townsville, Queensland, Australia) who provided the developing coconut fruits which were used in this study. We would like to thank Qing Liu and Madeline Mitchell for their time and critical review of the manuscript before submission.

Disclosures

The authors have no conflicts of interest to declare.

References

Arkcoll, D. (1988) Lauric oil resources. *Econ. Bot.* 42: 195–205.

Aymé, L., Baud, S., Dubreucq, B., Joffre, F. and Chardot, T. (2014) Function and localization of the *Arabidopsis thaliana* diacylglycerol acyltransferase DGAT2 expressed in yeast. *PLoS One* 9: e92237.

Aymé, L., Jolivet, P., Nicaud, J.-M. and Chardot, T. (2015) Molecular characterization of the *Elaeis guineensis* medium-chain fatty acid diacylglycerol acyltransferase DGAT1-1 by heterologous expression in yarrowia lipolytica. *PLoS One* 10: e0143113.

Aznar-Moreno, J., Denolf, P., Van Audenhove, K., De Bodt, S., Engelen, S., Fahy, D., et al. (2015) Type 1 diacylglycerol acyltransferases of *Brassica napus* preferentially incorporate oleic acid into triacylglycerol. *J. Exp. Bot.* 66: 6497–6506.

Barcelos, E., de Almeida Rios, S., Cunha, R,N.V., Lopes, R., Motoike, S.Y., et al. (2015) Oil palm natural diversity and the potential for yield improvement. *Front. Plant Sci.* 6: 1–16.

Basiron, Y. (2007) Palm oil production through sustainable plantations. *Eur. J. Lipid Sci. Technol.* 109: 289–295.

Bates, P.D. and Browse, J. (2011) The pathway of triacylglycerol synthesis through phosphatidylcholine in Arabidopsis produces a bottleneck for the accumulation of unusual fatty acids in transgenic seeds. *Plant J.* 68: 387–399.

Bates, P.D., Fatihi, A., Snapp, A.R., Carlsson, A.S., Browse, J. and Lu, C. (2012) Acyl editing and headgroup exchange are the major mechanisms that direct polyunsaturated fatty acid flux into triacylglycerols. *Plant Physiol.* 160: 1530–1539.

Bates, P.D., Stymne, S. and Ohlrogge, J. (2013) Biochemical pathways in seed oil synthesis. *Curr. Opin. Plant Biol.* 16: 358–364.

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

Carter, C., Finley, W., Fry, J., Jackson, D. and Willis, L. (2007) Palm oil markets and future supply. *Eur. J. Lipid Sci. Technol.* 109: 307–314.

Chapman, K.D., Dyer, J.M. and Mullen, R.T. (2012) Biogenesis and functions of lipid droplets in plants: thematic review series: lipid droplet synthesis and metabolism: from yeast to man. *J. Lipid Res.* 53: 215–226.

Chapman, K.D. and Ohlrogge, J.B. (2012) Compartmentation of triacylglycerol accumulation in plants. *J. Biol. Chem.* 287: 2288–2294.

Dahlqvist, A., Ståhl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., et al. (2000) Phospholipid: diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proc. Natl. Acad. Sci. USA* 97: 6487–6492.

Dehesh, K., Jones, A., Knutzon, D.S. and Voelker, T.A. (1996) Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from *Cuphea hookeriana*. *Plant J.* 9: 167–172.

Durrett, T.P., McClosky, D.D., Tumaney, A.W., Elzinga, D.A., Ohlrogge, J. and Pollard, M. (2010) A distinct DGAT with sn-3 acetyltransferase activity that synthesizes unusual, reduced-viscosity oils in *Euonymus* and transgenic seeds. *Proc. Natl. Acad. Sci. USA* 107: 9464–9469.

Dussert, S., Guerin, C., Andersson, M., Joet, T., Tranbarger, T.J., Pizot, M., et al. (2013) Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. *Plant Physiol.* 162: 1337–1358.

El Tahchy, A., Reynolds, K.B., Petrie, J.R., Singh, S.P. and Vanhercke, T. (2017) Thioesterase overexpression in *Nicotiana benthamiana* leaf increases the fatty acid flux into triacylglycerol. *FEBS Lett.* 591: 448–456.

Fan, H., Xiao, Y., Yang, Y., Xia, W., Mason, A.S., Xia, Z., et al. (2013a) RNA-Seq analysis of *Cocos nucifera*: transcriptome sequencing and de novo assembly for subsequent functional genomics approaches. *PLoS One* 8: e59997.

Fan, J., Yan, C., Zhang, X. and Xu, C. (2013b) Dual role for phospholipid: diacylglycerol acyltransferase: enhancing fatty acid synthesis and diverting fatty acids from membrane lipids to triacylglycerol in arabidopsis leaves. *Plant Cell.* 25: 3506–3518.

Frentzen, M. (1998) Acyltransferases from basic science to modified seed oils. *Lipid/Fett* 100: 161–166.

Gunstone, F., Harwood, J., Dijkstra, A., Murphy, D., Quinn, P., Erhan, S., et al. (2007) The Lipid Handbook with CD-ROM. CRC Press, Boca Raton, FL.

Hsieh, K. and Huang, A.H.C. (2004) Endoplasmic reticulum, oleosins, and oils in seeds and tapetum cells. *Plant Physiol.* 136: 3427–3434.

Huang, C.Y. and Huang, A.H.C. (2017) Motifs in oleosin target the cytosolic side of endoplasmic reticulum and budding lipid droplet. *Plant Physiol*. 174: 2248–2260.

Huang, Y.-Y., Lee, C.-P., Fu, J.L., Chang, B.C.-H., Matzke, A.J.M. and Matzke, M. (2014) De novo transcriptome sequence assembly from coconut leaves and seeds with a focus on factors involved in RNA-directed DNA methylation. *G3* 4: 2147–2157.

Huang, Y.-Y., Matzke, A.J. and Matzke, M. (2013) Complete sequence and comparative analysis of the chloroplast genome of coconut palm (Cocos nucifera). *PLoS One* 8: e74736.

Iskandarov, U., Silva, J.E., Kim, H.J., Andersson, M., Cahoon, R.E., Mockaitis, K., et al. (2017) A specialized diacylglycerol acyltransferase contributes to the extreme medium-chain fatty acid content of *Cuphea* seed oil. *Plant Physiol*. 174: 97.

Jing, F., Cantu, D., Tvaruzkova, J., Chipman, J., Nikolau, B., Yandeau-Nelson, M., et al. (2011) Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. *BMC Biochem*. 12: 44.

Kempken, F. and Jung, C. (2009) Genetic Modification of Plants. Springer, Dordrecht.

Kim, H.J., Silva, J.E., Iskandarov, U., Andersson, M., Cahoon, R.E., Mockaitis, K., et al. (2015a) Structurally divergent lysophosphatidic acid acyltransferases with high selectivity for saturated medium chain fatty acids from Cuphea seeds. *Plant J*. 84: 1021–1033.

Kim, H.J., Silva, J.E., Vu, H.S., Mockaitis, K., Nam, J.-W. and Cahoon, E.B. (2015b) Toward production of jet fuel functionality in oilseeds: identification of FatB acyl-acyl carrier protein thioesterases and evaluation of combinatorial expression strategies in Camelina seeds. *J. Exp. Bot*. 66: 4251–4265.

Kinney, A.J. (1997) Genetic engineering of oilseeds for desired traits. *In* Genetic Engineering: Principles and Methods. Edited by Setlow, J.K. pp. 149–166. Springer US, Boston, MA.

Knutzon, D.S., Hayes, T.R., Wyrick, A., Xiong, H., Maelor Davies, H. and Voelker, T.A. (1999) Lysophosphatidic acid acyltransferase from coconut endosperm mediates the insertion of laurate at the sn-2 position of triacylglycerols in lauric rapeseed oil and can increase total laurate levels. *Plant Physiol*. 120: 739–746.

Knutzon, D.S., Lardizabal, K.D., Nelsen, J.S., Bleibaum, J.L., Davies, H.M. and Metz, J.G. (1995) Cloning of a coconut endosperm cDNA encoding a 1-acyl-sn-glycerol-3-phosphate acyltransferase that accepts medium-chain-length substrates. *Plant Physiol*. 109: 999–1006.

Kumar, S.N. (2011) Variability in coconut (*Cocos nucifera* L.) germplasm and hybrids for fatty acid profile of oil. *J. Agric. Food Chem*. 59: 13050–13058.

Laureles, L.R., Rodriguez, F.M., Reaño, C.E., Santos, G.A., Laurena, A.C. and Mendoza, E.M.T. (2002) Variability in fatty acid and triacylglycerol composition of the oil of coconut (*Cocos nucifera* L.) hybrids and their parentals. *J. Agric. Food Chem*. 50: 1581–1586.

Lear, S. and Cobb, S.L. (2016) Pep-Calc.com: a set of web utilities for the calculation of peptide and peptoid properties and automatic mass spectral peak assignment. *Journal of computer-aided molecular design*. 30: 271–277.

Li, D. and Fan, Y. (2009) Cloning, characterisation, and expression analysis of an oleosin gene in coconut (*Cocos nucifera* L.) pulp. *J. Hortic. Sci. Biotechnol*. 84: 483–488.

Liang, Y., Yuan, Y., Liu, T., Mao, W., Zheng, Y. and Li, D. (2014) Identification and computational annotation of genes differentially expressed in pulp development of *Cocos nucifera* L. by suppression subtractive hybridization. *BMC Plant Biol*. 14: 1–17.

Liu, Q., Siloto, R.M., Lehner, R., Stone, S.J. and Weselake, R.J. (2012) Acyl-CoA: diacylglycerol acyltransferase: molecular biology, biochemistry and biotechnology. *Prog. Lipid Res*. 51: 350–377.

Lung, S.C. and Weselake, R.J. (2006) Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis. *Lipids* 41: 1073–1088.

Maurer, S., Waschatko, G., Schach, D., Zielbauer, B.I., Dahl, J., Weidner, T., et al. (2013) The role of intact oleosin for stabilization and function of oleosomes. *J. Phys. Chem. B* 117: 13872–13883.

Murphy, D.J. (2007) Future prospects for oil palm in the 21st century: biological and related challenges. *Eur. J. Lipid Sci. Technol*. 109: 296–306.

Murphy, D.J. (2012) The dynamic roles of intracellular lipid droplets: from archaea to mammals. *Protoplasma* 249: 541–585.

Nejat, N., Cahill, D.M., Vadamalai, G., Ziemann, M., Rookes, J. and Naderali, N. (2015) Transcriptomics-based analysis using RNA-Seq of the coconut (*Cocos nucifera*) leaf in response to yellow decline phytoplasma infection. *Mol. Genet. Genomics* 290: 1899–1910.

OECD (2017) Oilseeds and Oilseed Products, in OECD-FAO Agricultural Outlook 2017-2026, OECD Publishing, Paris, https://doi.org/10.1787/agr_outlook-2017-8-en.

Oo, K. and Stumpf, P. (1979) Fatty acid biosynthesis in the developing endosperm of *Cocos nucifera*. *Lipids* 14: 132–143.

Padolina, W., Lucas, L. and Torres, L. (1987) Chemical and physical properties of coconut oil. *Philipp. J. Coconut Stud*. 12: 4–17.

Parra, G., Bradnam, K. and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23: 1061–1067.

Parra, G., Bradnam, K., Ning, Z., Keane, T. and Korf, I. (2009) Assessing the gene space in draft genomes. *Nucleic Acids Res*. 37: 289–297.

Reynolds, K., Taylor, M., Zhou, X.-R., Vanhercke, T., Wood, C., Blanchard, C., et al. (2015) Metabolic engineering of medium-chain fatty acid biosynthesis in *Nicotiana benthamiana* plant leaf lipids. *Front. Plant Sci*. 6, 10.3389/fpls.2015.00164.

Reynolds, K.B., Taylor, M.C., Cullerne, D.P., Blanchard, C.L., Wood, C.C., Singh, S.P., et al. (2017) A reconfigured Kennedy pathway which promotes efficient accumulation of medium-chain fatty acids in leaf oils. *Plant Biotechnol. J*. 15: 1397–1408.

Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* 16: 276–277.

Ruffalo, M., Koyutürk, M., Ray, S. and LaFramboise, T. (2012) Accurate estimation of short read mapping quality for next-generation genome sequencing. *Bioinformatics* 28: i349–i355.

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

Sandager, L., Gustavsson, M.H., Ståhl, U., Dahlqvist, A., Wiberg, E., Banas, A., et al. (2002) Storage lipid synthesis is non-essential in yeast. *J. Biol. Chem*. 277: 6478–6482.

Sayer, D., Goodridge, D. and Christiansen, F. (2004) Assign 2.0: software for the analysis of Phred quality values for quality control of HLA sequencing-based typing. *Tissue Antigens* 64: 556–565.

Shockey, J., Regmi, A., Cotton, K., Adhikari, N., Browse, J. and Bates, P.D. (2016) Identification of Arabidopsis GPAT9 (At5g60620) as an essential gene involved in triacylglycerol biosynthesis. *Plant Physiol*. 170: 163–179.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.

Singer, S.D., Chen, G., Mietkiewska, E., Tomasi, P., Jayawardhane, K., Dyer, J.M., et al. (2016) Arabidopsis GPAT9 contributes to synthesis of intracellular glycerolipids but not surface lipids. *J. Exp. Bot*. 67: 4627–4638.

Ståhl, U., Carlsson, A.S., Lenman, M., Dahlqvist, A., Huang, B., Banaś, W., et al. (2004) Cloning and functional characterization of a phospholipid: diacylglycerol acyltransferase from Arabidopsis. *Plant Physiol*. 135: 1324–1335.

Turchetto-Zolet, A.C., Christoff, A.P., Kulcheski, F.R., Loss-Morais, G., Margis, R. and Margis-Pinheiro, M. (2016) Diversity and evolution of plant diacylglycerol acyltransferase (DGATs) unveiled by phylogenetic, gene structure and expression analyses. *Genet. Mol. Biol*. 39: 524–538.

Turchetto-Zolet, A.C., Maraschin, F.S., de Morais, G.L., Cagliari, A., Andrade, C.M., Margis-Pinheiro, M., et al. (2011) Evolutionary view of acyl-CoA diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis. *BMC Evol. Biol*. 11: 263.

Vanhercke, T., El Tahchy, A., Liu, Q., Zhou, X.R., Shrestha, P., Divi, U.K., et al. (2014a) Metabolic engineering of biomass for high energy density: oilseed-like triacylglycerol yields from plant leaves. *Plant Biotechnol. J.* 12: 231–239.

Vanhercke, T., El Tahchy, A., Shrestha, P., Zhou, X.-R., Singh, S.P. and Petrie, J.R. (2013) Synergistic effect of WRI1 and DGAT1 coexpression on triacylglycerol biosynthesis in plants. *FEBS Lett*. 587: 364–369.

Vanhercke, T., Petrie, J.R. and Singh, S.P. (2014b) Energy densification in vegetative biomass through metabolic engineering. *Biocatal. Agric. Biotechnol*. 3: 75–80.

Wiberg, E., Banas, A. and Stymne, S. (1997) Fatty acid distribution and lipid metabolism in developing seeds of laurate-producing rape (*Brassica napus* L.). *Planta* 203: 341–348.

Wood, C.C., Petrie, J.R., Shrestha, P., Mansour, M.P., Nichols, P.D., Green, A.G., et al. (2009) A leaf-based assay using interchangeable design principles to rapidly assemble multistep recombinant pathways. *Plant Biotechnol. J.* 7: 914–924.

Yen, C.-L.E., Stone, S.J., Koliwad, S., Harris, C. and Farese, R.V. (2008) DGAT enzymes and triacylglycerol biosynthesis. *J. Lipid Res.* 49: 2283–2301.

Yuan, Y., Chen, Y., Yan, S., Liang, Y., Zheng, Y. and Dongdong, L. (2013) Molecular cloning and characterisation of an acyl carrier protein thioesterase gene (CocoFatB1) expressed in the endosperm of coconut (*Cocos nucifera*) and its heterologous expression in Nicotiana tabacum to engineer the accumulation of different fatty acids. *Funct. Plant Biol*. 41: 80–86.

Zhang, M., Fan, J., Taylor, D.C. and Ohlrogge, J.B. (2009) 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.

Zhou, X.R., Green, A.G. and Singh, S.P. (2011) *Caenorhabditis elegans* delta12-desaturase FAT-2 is a bifunctional desaturase able to desaturate a diverse range of fatty acid substrates at the Delta12 and Delta15 positions. *J. Biol. Chem.* 286: 43644–43650.

Zhou, X.R., Shrestha, P., Yin, F., Petrie, J.R. and Singh, S.P. (2013) AtDGAT2 is a functional acyl-CoA: diacylglycerol acyltransferase and displays different acyl-CoA substrate preferences than AtDGAT1. *FEBS Lett*. 587: 2371–2376.