PHILIPP AGRIC SCIENTIST
Vol. 102 Special Issue
December 2019

ISSN 0031-7454

Transcriptome Analysis of 'Philippine Lono Tall' Coconut (*Cocos nucifera* L.) Endosperm Reveals Differential Expression of Genes Involved in Oil Biosynthesis

Cecille Ann L. Osio[1*], Genevieve Mae B. Aquino[1,2*], Wilson F. Aala Jr.[3*], Jickerson P. Lado[4], Bernabeth Jo T. Tendero[1,2], Cristeta A. Cueto[5], Maria Genaleen Q. Diaz[1,4†], and Rita P. Laude[1,4†]

[1] Philippine Genome Center – Program for Agriculture, Livestock, Fisheries, and Forestry, University of the Philippines Los Baños, Philippines
[2] Office of the Vice-Chancellor for Research and Extension, University of the Philippines Los Baños, Philippines
[3] Philippine Genome Center Mindanao Satellite Facility, University of the Philippines Mindanao, Philippines
[4] Genetics and Molecular Biology Division, Institute of Biological Sciences, College of Arts and Sciences, University of the Philippines Los Baños, Philippines
[5] Philippine Coconut Authority – Albay Research Center, Philippines

*These authors contributed equally to this work.
†Correspondence: mqdiaz@up.edu.ph, ritalaude50@gmail.com

The 'Philippine Lono Tall' (PLNT) is a variant of the more common 'Philippine Laguna Tall' (LAGT), which produces fruits with soft endosperm and reported higher fat content. To understand patterns of fatty acid (FA) and oil accumulation in LAGT and PLNT fruits, transcriptomes of 6–7 month-old endosperm samples were analyzed by RNA-Seq. Quantitative PCR was performed to analyze the differential expression of selected genes related to oil biosynthesis. Further, oil samples from the PLNT endosperm were analyzed to determine their FA composition across developmental stages. A total of 416,488 contigs were *de novo* assembled, including 15,497 (14,356 upregulated and 1,141 downregulated) differentially expressed contigs. Several putative unigenes related to cell membrane and wall biogenesis, endosperm development, and oil biosynthesis and accumulation were identified among the assembled contigs. This first report of the complete ontogenetic FA profile revealed that medium chain fatty acids are the main components of oil from the PLNT endosperm. This pilot study is the first to suggest a molecular basis for the unique 'Lono' phenotype.

Keywords: coconut, fatty acids, oil biosynthesis, 'Philippine Laguna Tall', 'Philippine Lono Tall', quantitative real-time PCR, RNA sequencing, transcriptome

Abbreviations: C6:0 – caproic acid, C8:0 – caprylic acid, C10:0 – capric acid, C12:0 – lauric acid, C14:0 – myristic acid, C16:0 – palmitic acid, C16:1$^{\Delta 9}$ – palmitoleic acid, C18:0 – stearic acid, C18:1$^{\Delta 9}$ – oleic acid, C18:2$^{\Delta 9,12}$ – linoleic acid, C18:3$^{\Delta 9,12,15}$ – linolenic acid, C20:0 – arachidic acid, DEC – differentially expressed contigs, DEG – differentially expressed genes, FA – fatty acids, FAS – fatty acid synthase, LAGT – 'Philippine Laguna Tall', LCB – long-chain bases, LCFA – long chain fatty acids, MAP – months after pollination, MCFA – medium chain fatty acids, PCA – Philippine Coconut Authority, PLNT – 'Philippine Lono Tall', RNA-Seq – RNA sequencing, RT-qPCR – quantitative real-time PCR, SFA – saturated fatty acids, T_a – annealing temperature, TAG – triacylglycerols, VLCFA – very long chain fatty acids

INTRODUCTION

The 'Philippine Lono Tall' (PLNT) variant of the more common 'Philippine Laguna Tall' (LAGT) variety is a high-value, soft-endosperm phenotype. Zuniga et al. (1970) and Santos et al. (1984) classified PLNT as a member of the tall *typica* group. The earliest known origin of PLNT is La Union, Philippines (Padolina 1985); other planting sites include the University of the Philippines in Los Baños, Laguna and the Philippine Coconut Authority – Albay Research Center (PCA-ARC), Camalig, Albay, and Sorsogon in Bicol. International conservation sites for PLNT include the Central Plantation Crops Research Institute (CPCRI) in Kasaragod, Kerala, India and the Coconut Research Station of Tamil Nadu Agricultural University in Aliyarnagar, India. The PLNT accession

IND 050 at CPCRI was deposited in 1960 under the code "PHL" (Batugal et al. 2010).

Fruits with the PLNT phenotype usually arise from LAGT palms, although some PLNT fruits have been observed in other cultivars. Ocular inspection cannot distinguish LAGT from PLNT endosperm samples and sensory evaluation is required to determine the PLNT variant (Rillo 2014). The molecular detection of the PLNT phenotype has not been reported to date. Spontaneous termination of fruit development and limited *in situ* embryo germination impede the analysis of PLNT inheritance. Rillo (2014) observed the PLNT phenotype in 3% of the fruit produced by Lono-bearing coconut palms from 1998 to 2001. Proximate analysis of PLNT and LAGT endosperm samples revealed an average fat content of 71.80% and 54.83%, respectively.

Coconut oil is a triacylglycerol (TAG) ester composed of three fatty acid (FA) chains, with lengths from C6:0 to C20:0 (Bezard et al. 1971). Dried mature coconut kernel called copra is crushed to produce the oil, which is mainly used for food production or for manufacturing toiletries, surfactants, and other industrial products. According to Codex Alimentarius Standard for Named Vegetable Oils (CXS 210-1999), coconut oil primarily contains approximately 45.1%–53.2% lauric acid (C12:0), which is comparable to palm kernel oil (45%–55%). By contrast, oleic (C18:1$^{\Delta 9}$) or linoleic acid (C18:2$^{\Delta 9,12}$) are the primary FA in other oil-producing crops like almond, corn, pistachio, peanut, sesame, soybean, canola, and sunflower (FAO/WHO 2007). CPCRI (2013) reported an annual average copra yield of >25 kg·palm^{-1} from PLNT palms (Supplementary Table 1). Gas chromatography showed 93.2% saturated FA and 6.8% unsaturated FA from copra based on 10,000 PLNT fruits sampled over a three-year period (Kumar 2007, 2011; Kumar and Balakrishnan 2009).

High-performance liquid chromatography (HPLC) revealed that LAGT coconut contains 90.86% saturated and 9.15% unsaturated FA (Laureles et al. 2002), while PLNT coconut has 96.90% saturated and 3.10% unsaturated FA (Rodriguez 2013). The saturated medium chain FA (MCFA) composition is the same for LAGT and PLNT, except for different levels of lauric acid (C12:0) in LAGT (49.70%) and PLNT (52.10%). Concentrations of saturated long chain FA (LCFA) were lower in LAGT (26.41%) than in PLNT (35.90%), and stearic acid (C18:0; 6.10%) was only observed in the latter. Unsaturated oleic acid (C18:1$^{\Delta 9}$; 6.02%) and linoleic acid (C18:2$^{\Delta 9,12}$; 3.13%) were detected in LAGT, whereas PLNT only contained C18:1$^{\Delta 9,12}$ (0.90%).

FA biosynthesis (Fig. 1) is a repetitive cycle of condensation, reduction of the carbonyl group, dehydration, and the reduction of the carbon-carbon double bond (Dussert et al. 2013). The fatty acid synthase (FAS) complex catalyzes all enzymatic processes during FA synthesis to produce a palmitoyl-ACP complex (C16:0) that is cleaved by acyl-ACP thioesterases (Reshef et al. 2003). Developing seeds of plants like *Arabidopsis thaliana* (L.) Heynh. have low levels of oleoyl-ACP thioesterase (*FatA*) transcription (Moreno-Pérez et al. 2011), whereas high palmitoyl-ACP thioesterase (*FatB*) transcription was observed in developing tissues, such as the coffee (*Coffea arabica* L.) endosperm (Joët et al. 2009). The upregulation of *FatB* apparently increases C16:0 levels in oil bodies (Dussert et al. 2013).

Additional genes are required for FA elongation and desaturation to produce very long chain fatty acids (VLCFA) that accumulate in embryonic cells as seed oil storage compounds (Haslam and Kunst 2013). The 3-ketoacyl-CoA synthases utilized for VLCFA synthesis are encoded by the multigenic *ELO* gene family, such as an *Arabidopsis* elongation of FA 3 (*ELO3*)-like gene (Quist et al. 2009). VLCFA are used to produce sphingolipids, which control various cellular processes (Haslam and Kunst 2013). Sphingolipid structures are formed via the desaturation of long-chain bases (LCB) by sphingoid desaturases like delta(8)-fatty-acid desaturase 2 (d8-DES2), which produces a *cis* or *trans* delta(8) double bond (Chen et al. 2012). However, the metabolic and physiological significance of LCB unsaturation is unknown.

Free FA are used to synthesize and maintain structural components in organellar membranes or to assemble TAG in the endoplasmic reticulum (Dussert et al. 2013). Intracellular TAG are deposited in oil bodies or oleosomes (Siloto et al. 2006). Oleosome size and stability are maintained by oleosins, hydrophobic proteins in the phospholipid coat that facilitate oil body biogenesis and stabilization (Barre et al. 2008). Oleosin (*OLE*) genes have been characterized in African oil palm (*Elaeis guineensis* Jacq.) (Jamek et al. 2010) and coconut (Regalado et al. 2008; Vargas et al. 2017; Reynolds et al. 2019). The roles of oleosins in seed development and dormancy have not been fully explored.

Given the consistently high demand for coconut oil in the international market, coconut cultivar improvement and hybrid development would benefit from a detailed understanding of the major components of coconut oil as well as the underlying biosynthetic pathways for oil biosynthesis. Putative genes and transcription factors involved in oil biosynthesis have been identified *in silico* from the Hainan Tall coconut transcriptome (Fan et al. 2013). Similarly, RNA-Seq analysis has been used to compare transcriptomes of fruit and seed tissues in oil palm (Dussert et al. 2013) and analyze differential expression between the oil palm and date palm (*Phoenix dactylifera* L.) mesocarp (Bourgis et al. 2011).

The high FA and fat contents of PLNT copra makes it a promising variant for improved coconut oil

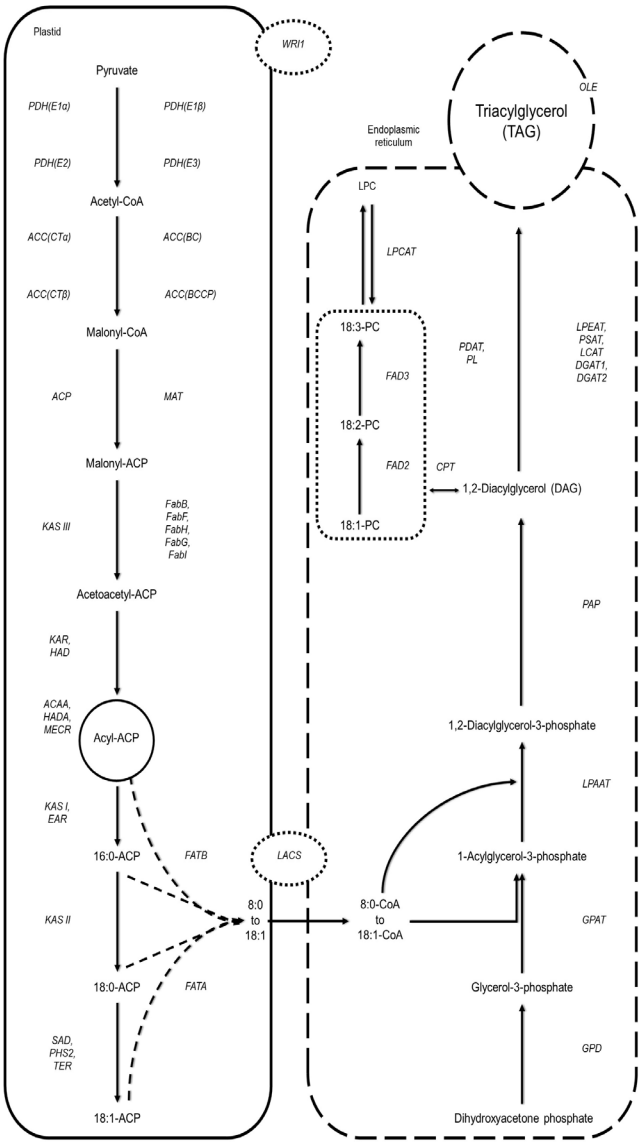

Fig 1. A simplified lipid biosynthesis pathway showing the genes related to triacylglycerol (TAG) production in palms (modified from Dussert et al. 2013). *WRI1,* ethylene-responsive transcription factor WRI1; *PDH(E1α/E1β/E2/E3),* pyruvate dehydrogenase E1 component subunit alpha/E1 component subunit beta/E2/E3; *ACC-CTα/CTβ/BC/BCCP*, acetyl-CoA carboxylase carboxyl transferase subunit alpha/beta/biotin carboxylase/biotin carboxyl carrier protein; *ACP*, acyl carrier protein; *MAT,* malonyl-CoA:ACP malonyltransferase; *FabB/FabF/FabH,* 3-oxoacyl-ACP synthase I/II/III; *FabG*, 3-oxoacyl-ACP reductase; *FabI*, enoyl-ACP reductase; *KAS I/II/III*, 3-ketoacyl-CoA synthase I/II/III; *KAR,* ketoacyl-ACP reductase; *ACAA*, 3-ketoacyl-CoA thiolase; *HADA*, enoyl-CoA hydratase; *MECR*, trans-2-enoyl-CoA reductase; *HAD*, hydroxyacyl-ACP dehydrase; *EAR*, enoyl-ACP reductase; *SAD*, stearate desaturase; *PHS2*, very-long chain (3R)-3-hydroacyl-CoA dehydratase 2; *TER*, very-long-chain enoyl-CoA reductase; *LACS*, long chain acyl-CoA synthetase; *GPD*, glycerol-3-phosphate dehydrogenase; *GPAT*, glycerol-3-phosphate acyltransferase; *LPAAT*, 1-acyl-sn-glycerol-3-phosphate acyltransferase; *PAP*, phosphatidate phosphatase; *LPEAT*, lysophospholipid acyltransferase; *PSAT*, phospholipid-sterol-O-acyltransferase; *LCAT*, lecithin-cholesterol acyltransferase; *DGAT1/2*, diacylglycerol-O-acyltransferase 1/2; *CPT*, diacylglycerol cholinephosphotransferase; *FAD2*, omega-6 fatty acid desaturase 2; *FAD3*, omega-3-fatty acid desaturase; *LPCAT*, 1-acylglycerol-3-phosphate-O-acyltransferase; *PDAT*, phospholipid:diacylglycerol acyltransferase; *PL*, phospholipase; *OLE*, oleosin.

production. The potential of PLNT remains unexplored because of the low probability of obtaining true Lono fruits and the limited success of its propagation via embryo rescue and tissue culture. Genomic and transcriptomic data could elucidate the molecular events that cause the PLNT phenotype. Transcriptomics can compare the expression profiles between several cultivars or various developmental stages. RNA-Seq analysis of the PLNT endosperm could facilitate the discovery of novel genes and gene networks underlying oil biosynthesis.

The present study aimed to differentiate the expression of genes for oil biosynthesis in LAGT and PLNT endosperm samples using next-generation sequencing of RNA (RNA-Seq). Differentially expressed oil biosynthesis genes in PLNT and LAGT were profiled by real-time quantitative PCR (RT-qPCR) of endosperm samples at 6 to 10 months after pollination (MAP).

MATERIALS AND METHODS

Plant Materials

A total of 10 fruits each were collected from PLNT and LAGT coconut palms at the PCA Albay Research Center (PCA-ARC) in Guinobatan City and from LAGT coconut palms at the PCA Zamboanga Research Center (PCA-ZRC) in Zamboanga City. PLNT fruits could only be identified using sensory tests. The collected fruits were aged 6, 7, 8, 9, and 10 MAP. The endosperm was isolated from the shell and ground in liquid nitrogen before indefinite storage at -80°C.

De novo Sequencing, Assembly, and Annotation

Total RNA was isolated from the endosperms of two 6–7 MAP PLNT and LAGT fruits using the Li and Trick (2008) method for the Modified RNA Isolation from Germinating Seeds. Eluted RNA was treated

with Turbo™ DNase (Invitrogen, Carlsbad, CA, USA) for one hour at 37°C. The solution was purified using the RNA cleanup protocol of the RNeasy Plant Mini Kit (Qiagen, Chatsworth, CA, USA). Total RNA quality was visualized by agarose gel electrophoresis. Spectrophotometric absorbance readings for the 260 and 280 wavelengths (A260/A280) were also obtained using a Nanodrop® spectrophotometer. Quantification of RNA samples were done using Qubit® RNA High Sensitivity (HS) Assay Kit (Molecular Probes, Eugene, OR, USA) and Qubit® 2 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). The RNA integrity number (RIN) was also obtained for each sample using Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., California, USA). After library preparation using the TruSeq® Stranded Total RNA kit, the extracted RNA was sequenced using the Illumina® HiSeq 2000 and NovaSeq 6000 platforms (Illumina, San Diego, CA, USA).

The raw sequences were deposited at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject accession number: PRJNA511560. Contigs were generated and unigenes were assembled by a *de novo* assembly algorithm based on de Bruijn graphs (Li et al. 2010) as implemented in the stand-alone CLC Genomic Workbench 9.0.1 (CLC Bio, Aarhus, Denmark; https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/). The above-mentioned unigenes were merged and used as the reference transcriptome for subsequent read mapping and abundance estimation of transcripts for LAGT and PLNT. A local database was created from all plant non-redundant (nr) protein sequences in NCBI using Blast2GO software (Conesa et al. 2005). The assembled contigs were queried against the local database with an expectation value cut-off of $1x10^{-5}$. The tabulated BLAST result was used as the expression data annotation file.

Statistical Analysis of Read Abundance

Differential expression analysis was performed using the RNA-Seq Analysis tool of the CLC Genomic Workbench with default parameters. Pairwise comparisons were performed using the Baggerley et al.'s test, a β-binomial sampling model comparing the proportions of counts in a group of samples against another that is suited to cases where replicates are available in the groups (Baggerley et al. 2003). Contigs were considered to be differentially expressed when the false discovery rate *p*-values (FDR *p*-val) were less than 0.05. Additionally, the absolute value of the log2 fold change for the differentially expressed contigs (DEC) should be greater than 3, with differences greater than 5. The data were reanalyzed using stricter parameters (FDR *p*-val < 0.01) however the results did not significantly change; hence, default parameters were used for RNA-Seq analysis. The expression values of the filtered dataset were normalized using a scaling method with a trimming percentage cutoff of 5%. The identified DEC were clustered using the Euclidean distance measure with the assumption of complete linkage.

Real-time Quantitative PCR of Selected Differentially Expressed Contigs

Candidate genes for preliminary RT-qPCR experiments were selected from literature and from the DEC of LAGT vs. PLNT transcriptome comparison. Primer pairs for each gene of interest (Table 1) were designed using Primer-BLAST (Ye et al. 2012). RT-qPCR of 100 ng cDNA templates of PLNT and LAGT at 6–10 MAP was performed using Step One® Real-Time PCR (Applied Biosystems, Austin, TX, USA) with the PowerUp™ SYBR® Green Master (Thermo Fisher Scientific) Mix kit. The amplification profile consisted of an initial denaturation step at 95°C for 10 min, followed by 35 cycles of denaturation at 95°C for 45 sec, annealing at optimum temperature for 30 sec, and elongation at 72°C for 60 sec, and a final elongation step at 72°C for 10 min. Relative quantification (RQ) of the candidate genes was based on the comparative ΔC_T ($\Delta\Delta C_T$) method. Actin was used as the internal control gene (dela Cruz et al. 2011). Student's *t*-test was performed to determine if the RQ values observed from PLNT were significantly different from RQ values of LAGT across developmental stages for each gene (α = 0.05) using the stats package in R version 3.6.0 (R Core Team 2019; https://www.R-project.org).

Table 1. Primers designed for the amplification of candidate genes related to oil biosynthesis in coconut endosperm.

Gene ID	Forward Primer	Reverse Primer	Ta (°C)
d8-DES2	TGCCTCTCAAACACCAAGTTGA	ACTTCTCGGCGAGCGTCTA	60.6
ELO3	GCCCAGGCCCCCTCA	CCACCGACACGGCGAAG	60.8
FatA	CTTATATCGGATGGGTCCTTGAA	TGCTGGCATTCTCTCCTGTAATC	59.0
FatB	AATTGCCGGACTGGAGCAT	TCCGCGGCCGAAAAG	59.0
OLE16k-Da-isoform1	AAGAAGCCAGTGGCAGTGT	TGATAAAGAGCGACGAGGGAG	59.0
OLE16k-Da-isoform2	GTTAGCACAGCACAGCATTA	GACAAAGTGAAGTTGGAGGG	59.0
OLE18kDa	CGAAGGCCTGAAGGGTCTC	GGGAGATGTTTTCATGTCCTGC	59.0
OLE5	CTTGACGTAGCTCACCACCC	TTTGATCACCGATGGCGGA	59.0
ACTIN	CGTGCCCGCAATGTATGTT	TGTGCGGCCACTAGCATAGA	59.0

Oil Isolation and FA Analysis

Solid endosperm comprising three replicate samples were pulverized in liquid nitrogen using a mortar and pestle, transferred to a 50 mL tube, and mixed with HPLC-quality water in a 1:1 volume ratio following the protocol of Raghavendra and Raghavarao (2010) for coconut oil extraction using the "chilling and thawing" method with modifications. The resulting homogenate was vigorously mixed using a vortex mixer and spun at 6,000 ×g for 20 min. The upper creamy layer was transferred to 15 mL tubes. Centrifugation of the homogenate was repeated until no upper creamy layer can be obtained. The tubes were stored at 4°C for 24 hours. The samples were placed inside a water bath at 50°C for one hour. The top oily layer (>1 mL) was transferred to new 2 mL tube and stored at 4°C. The FA profile of the oil samples was analyzed using AOAC standard protocols (AOAC 969.33 and AOAC 963.22, 16th edition) by the Central Analytical Services Laboratory, National Institute of Molecular Biology and Biotechnology (CASL, BIOTECH), UPLB. Gas chromatography (GC) was performed through a 1.10 mm wide × 2.10 mm long column with Chromosorb W packing material. The following temperature program was used: 90°C for 6 min; increased to 180°C at 5°C·min^{-1} and held for 2 min; increased to 220°C at 10°C·min^{-1} and held for 15 min. The injector and detector temperatures were set at 260°C for three replicates. Each FA was identified against a virgin coconut fatty acid methyl ester and quantified using area normalization. Results were expressed as percent of individual FA.

The FA composition of LAGT and PLNT at 7–12 MAP were statistically analyzed using the stats package in R version 3.6.0 (R Core Team 2019; https://www.R-project.org). The Wilk-Shapiro test was conducted to check for normal distribution of the data (Shapiro and Wilk 1965) while Mauchly's sphericity test was performed to validate the homogeneity of variance (Mauchly 1940). Three tests were performed to identify significant differences: (1) between the overall percentage of FA, (2) within the FA profile across developmental stages via Hotelling's t-squared statistic (Hotelling 1931), and (3) among the percent composition of each FA across developmental stages. Statistical analysis using Student's *t*-test (p-val < 0.05) was also done to determine if there are significant differences between FA composition of LAGT and PLNT for each developmental stage.

RESULTS AND DISCUSSION

Transcriptome Assembly and Annotation in "Lono"

After quality assessment, adapter and poly-A tail trimming, and filtering of low quality reads, a total of 263,251,316 reads from 'Laguna Tall' and 132,259,706 reads from 'Lono were obtained (Table 2). *De novo* assembly of LAGT and PLNT solid endosperm transcriptomes at 6–7 MAP yielded 285,037 and 267,827 contigs, respectively. The average contig lengths of LAGT and PLNT transcriptomes were 723 bp and 398 bp, respectively. The two sets of reads were pooled to generate the contigs and then assembled into a non-redundant unigene set. A total of 416,488 unigenes made up of 266,653,479 bases, with an average length of 640 bp and median length of 376 bp, were generated from the merged LAGT and PLNT reads. Among these, 58,428 (14%) were over 1,000 bp long. The assembled unigenes were of sufficient length (N50 = 874) and mapped to a high percentage of the original reads (92.60% mapped reads).

Table 2. Summary statistics of the reads, contigs, and unigenes of 6–7 months old 'Laguna Tall', 'Lono', and merged endosperm transcriptomes.

Statistics	Laguna Tall	Lono	Merged
Total reads	263,251,316	132,259,706	395,511,022
Total length of reads	26,724,380,103	13,389,253,978	40,113,634,081
Mean Q30 (%)	95.0	94.0	95.0
Number of contigs	283,145	265,349	413,678
Total length of contigs (bp)	201,141,791	105,509,561	255,595,263
Mean contig length (bp)	710	398	618
N50 contig length (bp)	1,039	401	817
Contig GC content (%)	39.2	41.1	39.4
Number of unigenes	285,037	267,827	416,488
Total length of unigenes (bp)	209,928,010	108,731,964	266,653,479
Mean unigene length (bp)	736	406	640
N50 unigene length (bp)	1,108	410	874
Unigene GC content (%)	38.7	40.6	38.9

Read Mapping and Differential Expression Analysis

A high percentage of the LAGT and PLNT reads successfully mapped back to the reference unigene set which shows the efficiency of the assembly accounting for a good majority of the sequencing reads (Table 3). The reference was able to align a high percentage of the LAGT and PLNT reads with over 85.75% and 88.55% mapping, respectively. The uniquely mapping reads of LAGT were slightly higher than the PLNT reads at 83.26% and 83.04%, respectively. There were only 14.25% and 11.45% unmapped reads for LAGT and PLNT, respectively.

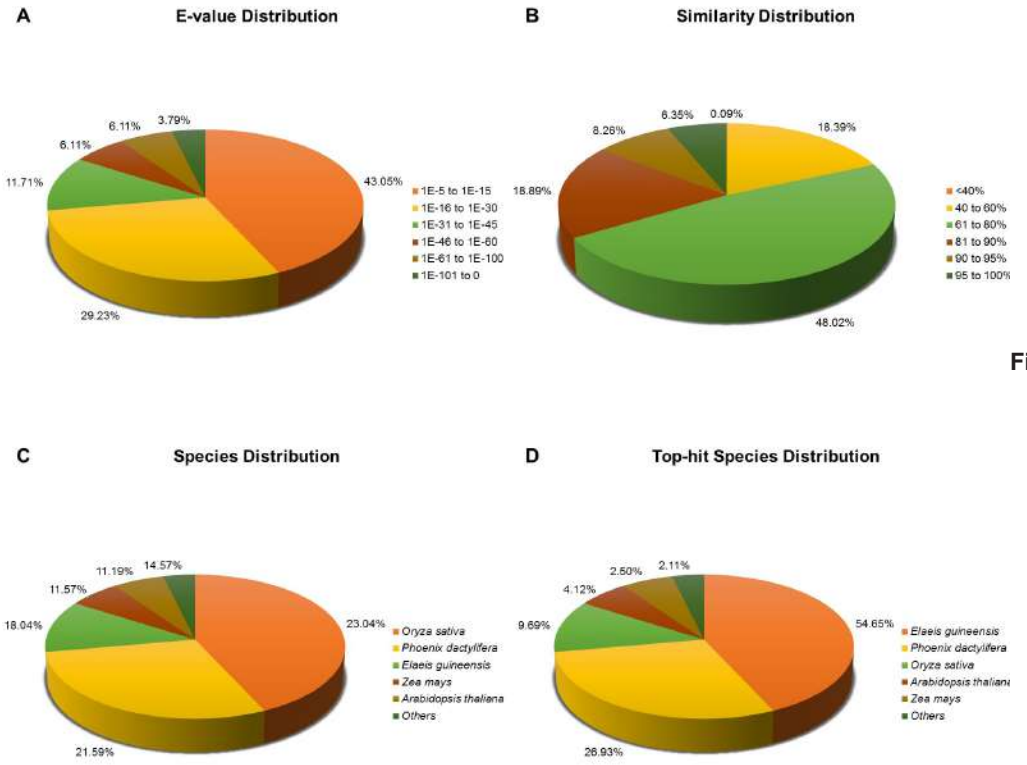

Fig 2. Characterization of the 15,497 differentially expressed contigs in 'Laguna Tall' and 'Lono' solid endosperms using the NCBI non-redundant version 5 (nr_v5) database. (A) E-value distribution with a cutoff of 1E-5. (B) Similarity distribution above 40%. (C) Species distribution of BLAST hits. (D) Distribution of top 5 species with BLAST hits.

Table 3. RNA-Seq mapping statistics for 'Laguna Tall' (LAGT) and 'Lono' (PLNT) reads mapped to the reference unigene set.

Read statistics	Laguna Tall	Lono
Counted fragments	125,102,115 (85.75%)	117,111,723 (88.55%)
Unique	121,464,733 (83.26%)	109,827,385 (83.04%)
Non-specific	3,637,382 (2.49%)	7,284,338 (5.51%)
Uncounted fragments	20,784,583 (14.25%)	15,147,983 (11.45%)
Total fragments	145,886,698	132,259,706

Among the assembled unigenes, 15,497 were differentially expressed between LAGT and PLNT at 6–7 MAP (Fig. 2). Based on fold change, 14,356 (92.64%) of the DEC were upregulated and 1,141 (7.36%) were downregulated in PLNT. Among the DEC, 2,280 (14.71%) showed significant similarity to known proteins in the NCBI nr protein database. The remaining 13,217 (85.29%) with missing annotations may be attributed to the short query sequences and limited genomic information in coconut. A total of 436 unigenes (19.12%) had BLAST hits that significantly matched with the aligned proteins (E-value < $1x10^{-45}$). Approximately 1,000 unigenes (43.86%) showed greater than 75% similarity to the protein entries. The *de novo* assembled *Cocos nucifera* (L.) contigs had high sequence homology with transcripts from other monocot species, namely, *E. guineensis* (54.65%) and *P. dactylifera* (26.93%). Although the genomes of different coconut varieties have been recently published (Fan et al. 2013; Lantican et al. 2019), much of the sequence annotation is incomplete. The predominance of homologous genes from species other than *C. nucifera* may be explained by the more comprehensive annotation of the complete *E. guineensis* (Singh et al. 2013) and *P. dactylifera* (Al-Dous et al. 2011) genomes. From the 2,280 identified differentially expressed genes (DEG), only 1,759 (77.15%) DEG were successfully mapped and only 1,268 (55.61%) DEG were successfully annotated. The functionally annotated DEG were classified into three gene ontology (GO) categories - molecular function, biological process, and cellular component (Fig. 3). Each category is composed of level 5 GO annotation terms.

PLNT has a higher number of upregulated DEG than LAGT (Fig. 4). The most upregulated DEG (Table 4) was annotated as cocosin 1, which has been associated with nutrient reservoir activity during seed development (Caldo et al. 2015). Several upregulated DEG encoded integral membrane proteins, such as "fasciclin-like arabinogalactan protein 2" (Johnson et al. 2003) and "formin-like protein 16" (Kawahara et al. 2013). Coconut fruit undergoes nuclear endosperm development and cellularization occurs at 7 MAP to form the solid endosperm (Berger 1999). Integral membrane proteins are crucial in building endosperm cells; the upregulation of these proteins in PLNT implies an early buildup of transcripts for cell biogenesis.

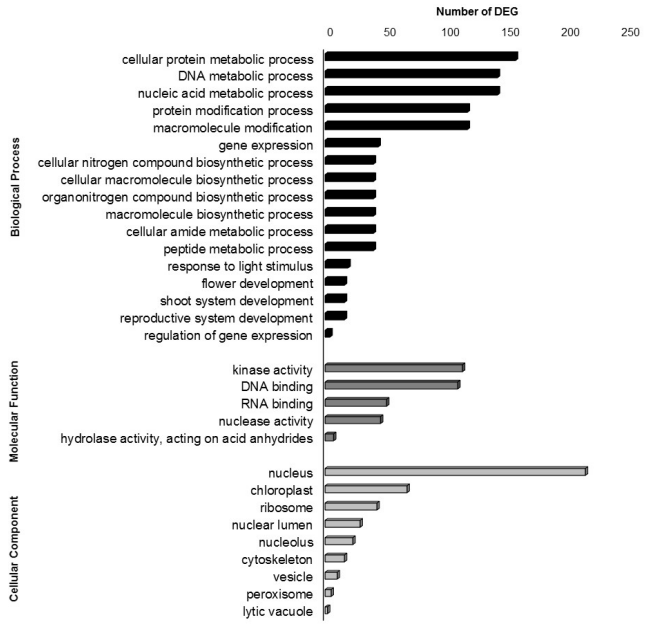

Fig 3. Distribution of the top gene ontology (GO) terms under the biological process, molecular function, and cellular component categories for the identified DEG in 'Laguna Tall' and 'Lono'.

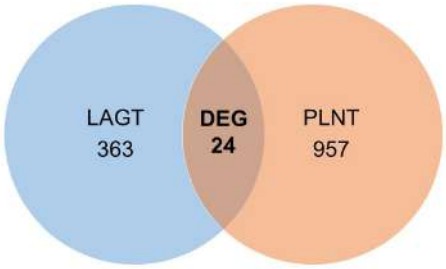

Fig 4. Venn diagram showing the distribution of upregulated differentially expressed genes (DEG) in 'Laguna Tall' and 'Lono' endosperm transcriptome.

The upregulated cytosolic "embryonic cell protein 63" was previously associated with BHLH109-mediated regulation of somatic embryogenesis (Nowak and Gaj 2016), whereas the chloroplastic "adenine nucleotide alpha hydrolases-like superfamily protein" has been linked to embryo development ending in seed dormancy (Apuya et al. 2002). Several of the most downregulated DEG encode for membrane proteins, such as "ABC transporter G family member 31" (Verrier et al. 2008) and "syntaxin-related protein KNOLLE" (Lukowitz et al. 1996). Notable downregulated DEG associated with cell division and the cell cycle include "syntaxin-related protein KNOLLE" (Lukowitz et al. 1996), "cyclin-B2-2 isoform X2" (Al-Mssallem et al. 2013), and "cyclin-dependent kinase B2-1" (Lee et al. 2003).

The differential expression of genes involved in oil biosynthesis was also revealed by comparing the LAGT and PLNT transcriptomes. The oil biosynthesis genes identified in the present analysis (Table 5) were compared with genes previously identified in 'Hainan Tall' by Fan et al. (2013) and in LAGT by Punzalan et al. (2019). Contrasting gene expression patterns during endosperm development may account for observed phenotypic differences between PLNT and LAGT in terms of the FA composition and oil yield.

Real-Time Quantitative PCR of Selected Differentially Expressed Contigs

Based on the results of RNA-Seq analysis of LAGT and PLNT transcriptomes and a review of related literature, 8 genes related to lipid biosynthesis were identified for preliminary analysis via RT-qPCR experiments. The expression profiles of *d8-DES2, ELO3, FatB,* and *OLE5* were revealed to be significantly different between LAGT and PLNT across developmental stages (Fig. 5).

Delta(8)-desaturase 2 (d8-DES2) and Elongation of fatty acids 3 (ELO3). Increased transcription of *d8-DES2* was observed in LAGT and PLNT at 6 MAP (Fig. 5A). The transcript levels differed greatly between the two variants at 7 MAP, with significantly higher expression in LAGT. Subsequent transcript levels at 8–10 MAP decreased in both samples. The abundance of *ELO3* transcripts in LAGT and PLNT at 6–7 MAP was observed to suddenly decrease at 7–8 MAP then plateau at almost half the original levels at 9–10 MAP with significantly lower expression levels in PLNT (Fig. 5B). These trends of gene expression in LAGT and PLNT are observed for both computational RNA-Seq analysis and experimental FA profiling. The differential gene expression may account for the low levels of unsaturated FA in PLNT at 10 MAP.

The sudden increase of *ELO3* transcript level in LAGT at 7 MAP implies that unsaturated FA accumulation could require other enzymes for FA elongation because the increased transcriptional activity does not coincide to any visible differences between LAGT and PLNT in terms of LCFA accumulation. The observed average amounts of C16:0-C18:x FA at 10 MAP may be explained by the low to intermediate expression levels of *ELO3* in LAGT and PLNT after 7 MAP. The spike in transcript abundance of *ELO3* in LAGT at 7 MAP could indicate that other FA elongation enzymes (e.g., VLCOR, KAS, LACS) are equally being utilized during unsaturated FA accumulation.

FA elongation and desaturation are critical points in oil biosynthesis. The functions of plant *ELO3* homologs have not been well-studied, although an *Arabidopsis* homolog has been reported to control VLCFA composition and influence cell signaling (Quist et al. 2009). The functional importance of LCB delta(8) unsaturation in plant development is likewise uncertain, although desaturation is required to produce the wide array of sphingolipid structures (Chen et al. 2012). The moderate levels of saturated and unsaturated

Table 4. List of the top 50 differentially expressed genes in the 'Laguna Tall' vs. 'Lono' transcriptome comparison at 6–7 months after pollination based on log2 transformed values.

Description	Difference	Fold Change
Upregulated Differentially Expressed Genes		
cocosin 1	14.44	5.81
fasciclin-like arabinogalactan protein 2	10.79	4.12
hypothetical protein OsI_34071	9.11	3.01
embryonic cell protein 63	9.62	3.40
alpha-galactosidase 2	12.05	13.05
uncharacterized protein LOC103700431, partial	8.78	3.77
adenine nucleotide alpha hydrolases-like superfamily protein	9.03	4.49
egg cell-secreted protein 1.2-like	11.59	∞
uncharacterized protein LOC103715169	7.99	3.23
egg cell-secreted protein 1.2-like	9.86	10.86
eukaryotic aspartyl protease family protein	7.77	3.59
egg cell-secreted protein 1.2-like	7.85	3.80
hypothetical protein OsI_34071	7.07	3.05
formin-like protein 16	7.06	3.04
G-type lectin S-receptor-like serine/threonine-protein kinase SD2-5	8.32	5.16
G-type lectin S-receptor-like serine/threonine-protein kinase SD2-5	8.32	5.16
polypyrimidine tract-binding-like protein	8.28	5.14
zinc finger CCCH domain-containing protein 18 isoform X2	8.25	5.12
gibberellin-regulated family protein	7.64	3.96
pectin acetylesterase 5	7.37	3.63
uncharacterized protein LOC105048222	6.85	3.06
putative polyprotein	7.15	3.38
ABC transporter G family member 36-like	6.77	3.04
retrotransposon protein, putative, Ty3-gypsy subclass	9.09	10.09
enolase-phosphatase E1	7.43	3.88
Downregulated Differentially Expressed Genes		
ABC transporter G family member 31	-9.61	-3.60
cotton fiber (DUF761)	-9.16	-3.56
syntaxin-related protein KNOLLE	-11.79	-∞
ribonucleoside-diphosphate reductase small chain	-8.62	-3.72
serine carboxypeptidase-like 40	-8.89	-4.83
phosphoinositide binding	-9.81	-10.81
cyclin-B2-2 isoform X2	-7.47	-3.25
probable serine/threonine-protein kinase PBL25 isoform X1	-7.34	-3.45
ATP-dependent DNA helicase SRS2-like protein At4g25120 isoform X2	-8.61	-6.43
cyclin-dependent kinase B2-1	-7.59	-3.94
high mobility group B protein 7	-8.56	-6.40
chaperone protein dnaJ 6	-8.00	-5.00
calmodulin-binding transcription activator 1 isoform X1	-7.97	-4.99
probable pectinesterase 68	-7.36	-3.85
protein argonaute 16	-6.82	-3.27

Cont. Table 4

Description	Difference	Fold Change
remodeling and spacing factor 1-like	-7.80	-4.90
metal tolerance protein 1	-7.21	-3.79
transmembrane protein	-7.77	-4.88
ATP-dependent DNA helicase SRS2-like protein At4g25120 isoform X1	-7.41	-4.19
copper transporter 1	-8.11	-6.12
uncharacterized protein LOC105056672	-6.47	-3.04
ubiquitin-conjugating enzyme E2 20	-6.78	-3.42
TIMELESS-interacting protein	-7.41	-4.71
origin of replication complex subunit 6	-6.41	-3.14
B3 domain-containing protein Os06g0194400	-6.79	-3.63

NOTE: Positive values indicate higher expression in PLNT while negative values indicate higher expression in LAGT.

Table 5. List of genes involved in lipid biosynthesis identified from three available *Cocos nucifera* Tall cultivar transcriptome assemblies.

Pathway	Hainan Tall (Fan et al. 2013)	Laguna Tall (Punzalan et al. 2019)	Lono Tall
Fatty acid biosynthesis and triacylglycerol assembly			
acetyl CoA-ACP carboxylase	+	+	
β-ketoacyl-ACP synthase	+	+	
β-ketoacyl-ACP reductase	+		
3-ketoacyl-CoA synthase			+
enoyl-ACP reductase	+		+
fatty acyl-ACP thioesterase	+		
β-hydroxyacyl-ACP dehydratase	+	+	+
3-oxoacyl-ACP-synthase			+
oleoyl-ACP hydrolase		+	
oleoyl-ACP thioesterase			+
palmitoyl-ACP thioesterase			+
stearyl-ACP desaturase	+		
ACP S-malonyltransferase		+	
acetyl-Coa carboxylase caboxyltransferase			+
oxidoreductases	+		+
fatty acid desaturase	+		+
fatty-acyl-CoA synthase		+	+
cytochrome P450	+		+
glycerol-3-phosphate 1-O-acyltransferase		+	+
glycerol-3-phosphate dehydrogenase	+		+
diacylglycerol acyltransferase		+	+
1-acylglycerol-3-phosphate O-acyltransferase		+	+
1-acyl-sn-glycerol-3-phosphate acyltransferase			+
phosphatidate phosphatase		+	+
oleosin		+	+

Cont. Table 5

Pathway	Hainan Tall (Fan et al. 2013)	Laguna Tall (Punzalan et al. 2019)	Lono Tall
Fatty acid metabolism			
long-chain-fatty-acid-CoA synthetase	+		
long-chain-fatty-acid--CoA ligase		+	+
acyl-CoA oxidase	+		
acyl-CoA C-acetyltransferase	+		
acyl-CoA dehydrogenase	+		+
acyl-CoA N-acyltransferase			+
alcohol dehydrogenase	+		+
aldehyde dehydrogenase	+		+
enoyl-CoA hydratase	+		+
enoyl-ACP reductase		+	
lysophospholipid acyltransferase			+
mitochondrial carnitine/acylcarnitine carrier-like protein			+
glyoxysomal fatty acid β-oxidation multifunctional protein			+
monogalactosyldiacylglycerol synthase 1			+
Fatty acid elongation			
quinone reductase	+		+
lecithin-cholesterol acyltransferase-like 4			+
trans-2-enoyl-CoA reductase	+		+
palmitoyl-protein thioesterase	+		+
elongation of fatty acids protein 3			+
Citrate cycle			
citrate synthase	+		+
malate dehydrogenase	+		+
isocitrate dehydrogenase	+		+
pyruvate dehydrogenase	+	+	+
oxoglutarate dehydrogenase	+		+
fatty acid synthase		+	
succinate dehydrogenase	+		+
phosphoenolpyruvate carboxykinase	+		+
2-oxoglutarate dehydrogenase			+
fumarate hydratase	+		+
succinate-CoA ligase	+		+
aconitate hydratase	+		+

NOTE: + means the gene is present in the corresponding transcriptome data.

MCFAs and LCFA in LAGT may be attributed to the increased abundance of *d8-DES2* and *ELO3*, which are both involved in VLCFA biosynthesis. Variations in the temporal expression of *d8-DES2* and *ELO3* imply that the synthesis of VLCFA may be triggered earlier in PLNT than in LAGT.

Acyl-ACP thioesterases (FATA, FATB). The expression profiles of oleoyl-ACP thioesterase (*FatA*) in LAGT and PLNT endosperms steadily decreased from 6–11 MAP (Fig. 5C). A significant decline in *FatA* transcript abundance was observed in the LAGT endosperm between 6 and 7 MAP. *FatB* transcription was higher in PLNT than LAGT at 6 MAP, but the amount of PLNT *FatB* transcripts was relatively lower at the subsequent developmental stages (Fig. 5D). By contrast, relatively constant *FatB* abundance was detected in LAGT

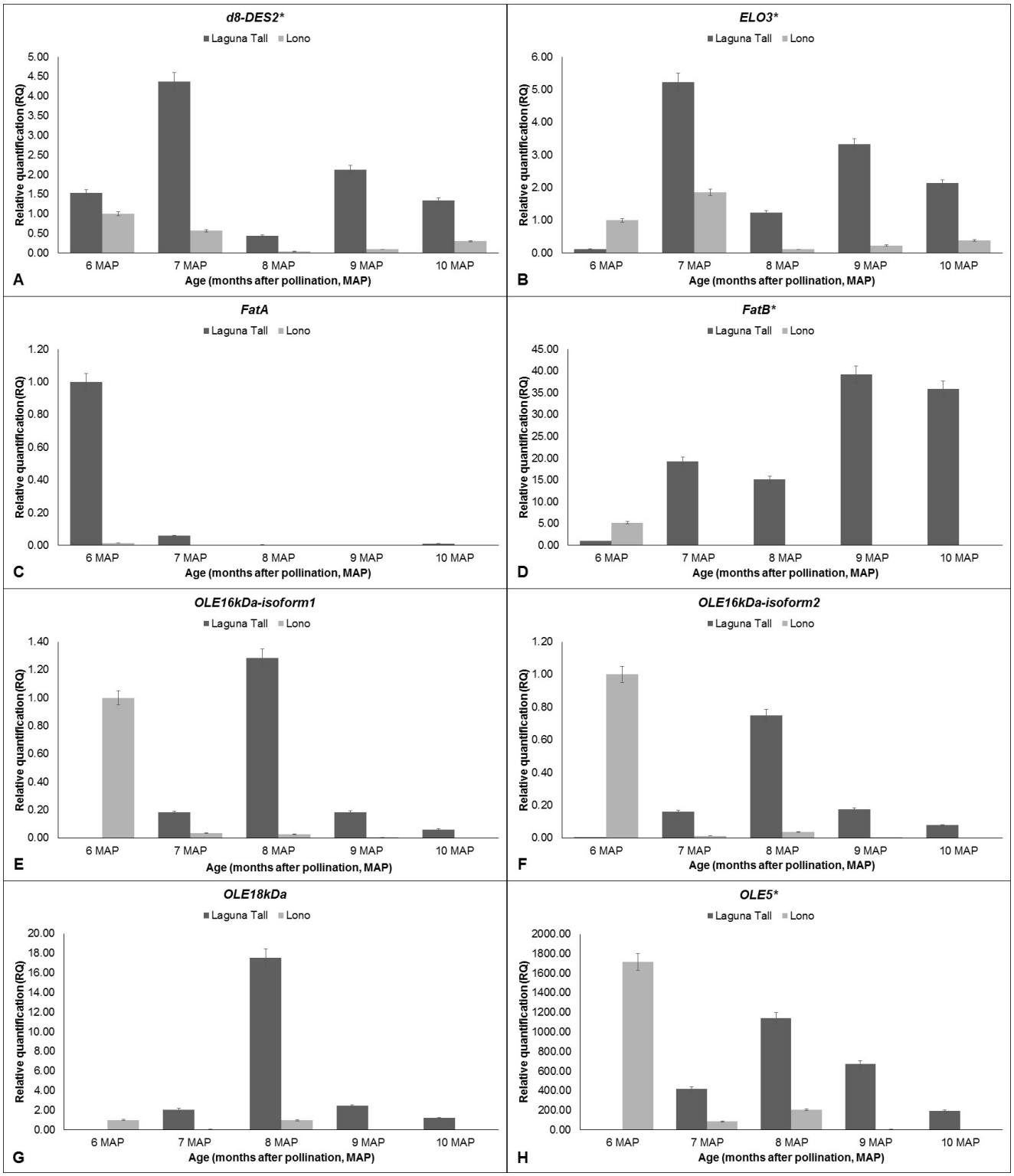

Fig 5. Relative quantification values for (A) *d8-DES2*, (B) *ELO3*, (C) *FatA*, (D) *FatB*, (E) *OLE16kDa* isoform1, (F) *OLE16kDa* isoform2, (G) *OLE18kDa*, and (H) *OLE5* based on real-time quantitative PCR of 'Laguna Tall' and 'Lono' coconut endosperms aged 6–10 months after pollination (MAP). *Genes with significantly different expression levels between 'Laguna Tall' and 'Lono' (α=0.05).

endosperms from 7–10 MAP.

FatA cleaves oleoyl-ACP (18:1-ACP) to produce unsaturated oleic acid (Dussert et al. 2013). The lower levels of *FatA* expression at 7 MAP may explain the reduced amount of free oleic acid due to decreased hydrolysis by oleoyl-ACP. The amount of oleic acid was previously reported to decrease from 23% to 10% throughout endosperm development until only 6% oleic acid is observed in the coconut endosperm at 11 MAP (Kumar and Balakrishnan 2009). *FatA* transcription similarly decreases in both PLNT and LAGT endosperms after the initial accumulation phase at 6–7 MAP.

Palmitoyl-ACP thioesterases (FatB) cleave palmitoyl-ACP (16:0-ACP) to produce free palmitic acid and regulate the premature termination of FA synthesis to continue producing other saturated FA. The *FatB3* paralog encodes the enzyme that catalyzes the reaction producing saturated MCFA (Dussert et al. 2013). The early transcription of *FatB* in PLNT at 6 MAP may compensate for the subsequent decrease in transcription at later developmental stages. Hypothetically, *FatB* transcription in PLNT endosperms could occur before 6 MAP, but the coconut endosperm is still loose and watery at 4–5 MAP.

Oleosins (OLE5, OLE16kDa isoform1 and isoform2, OLE18kDa). The expression of *OLE* isoforms was observed earlier in PLNT at 6 MAP, but transcript levels were significantly lower at later stages of development (Fig. 5E–5H). By contrast, oleosin transcripts in LAGT began to accumulate at approximately 7–8 MAP and their expression persisted at 10 MAP. The onset of oleosin accumulation in PLNT could not be determined because endosperm samples younger than 6 MAP were unavailable. Although oleosin transcription in PLNT may be triggered before 6 MAP, the coconut endosperm is predominantly liquid prior to this stage, thus, making RNA isolation extremely difficult.

OLE5 showed higher levels of transcript abundance in PLNT at 6 MAP, although LAGT did exhibit peak *OLE5* transcript activity at 8 MAP (Fig. 5G). The relatively early increase of *OLE5* abundance in PLNT implies that a crucial threshold of oleosin accumulation may be required before the irreversible transition to seed dormancy. Oleosins are a highly diverse group of oil storage proteins; results suggest that *OLE5* and *OLE18kDa* have overlapping functions in the PLNT endosperm. Four oleosin sequences identified in the present study had high sequence similarity with the coding and protein sequences of *C. nucifera* oleosin genes available in the GenBank database (Supplementary Table 2). Nucleotide BLAST revealed that the sequences had high homology with partial coding sequences of isoforms OLE500a (99% and 79% identity), OLE500c (99% identity), and OLE300a (96% identity) mRNAs (Regalado et al. 2008), complete coding sequences of isoforms 500a (99% and 79% identity), 500c (99% identity), and 300a (96% identity) mRNAs (Vargas et al. 2017), and complete coding sequence of oleosin mRNA (93% identity) (Reynolds et al. 2019). Likewise, protein BLAST showed homology with the same oleosin isoforms. Three contigs had the highest amino acid sequence similarity with partial sequence of oleosin isoform OLE500a (Regalado et al. 2008), oleosin isoform 500c (Vargas et al. 2017), and oleosin (Reynolds et al. 2019) with 100%, 100%, and 89% identities, respectively.

Ontogenetic FA Profile of the PLNT Solid Endosperm

Samples of the PLNT solid endosperm were observed to contain 11 different FAs, with C12:0 as the major component (48.12%), followed by C14:0 (20.76%), C16:0 (9.62%), C18:$1^{\Delta 9}$ (6.91%), C8:0 (5.33%), and C10:0 (5.33%). C6:0 (0.17%), C16:$1^{\Delta 9}$ (non detectable, defined as ≤ 0.05%, ND), C18:0 (2.80%), C18:$2^{\Delta 9,12}$ (1.04%), and C20:0 (ND) were detected at low to very low levels (Table 6). The observed FA composition coincides with the range of total fatty acid percentages recorded by the CODEX STAN 210-1999 (FAO/WHO 2007). Across PLNT endosperm developmental stages, the saturated FA decreased from 10.42% at 7 MAP to 6.43% at 10 MAP. Saturated MCFA, which account for the bulk of observed FAs, continuously increased until 10 MAP (56.25%–63.48%). However, it declined as the fruit matures (60.85%). The composition of saturated LCFA was less varied, ranging from 32.21% to 34.70%. Overall, FA profile of PLNT showed no significant difference as endosperm development progressed. However, when the percent composition of each FA across development was analyzed, all FA types showed significant difference across maturity except for caproic acid and stearic acid.

Further comparison of the PLNT and LAGT FA profiles revealed significant changes in the caprylic acid and capric acid compositions between varieties for four developmental stages, i.e., 7, 8, 10, and 11 MAP vs. 8, 9, 10, and 11 MAP, respectively. The lauric acid composition was only significantly different between PLNT and LAGT at 10 MAP. Significant differences were also observed at 9 and 10 MAP for myristic acid. Stearic acid was observed to be similar between PLNT and LAGT throughout endosperm development. Oleic acid and linoleic acid significantly differed at 8 and 11 MAP vs. 10 and 11 MAP, respectively. Most changes in the FA composition were significant at 10 MAP, which suggests that the temporal collection of PLNT fruits at this stage is recommended.

Differential Expression of Oil Biosynthesis Genes May Account for Lono Endosperm Phenotype

The *in silico* RNA-Seq analysis and corresponding preliminary investigation of gene expression via

Table 6. Fatty acid composition in the solid endosperms of 'Philippine Lono Tall' and 'Philippine Laguna Tall' coconut (*Cocos nucifera* L.) cultivars across developmental stages (7–12 months after pollination).

Fatty Acid*	Philippine Lono Tall (MAP)						Philippine Laguna Tall (MAP)					
	7	8	9	10	11	12	7	8	9	10	11	12
Caproic acid (C6:0)	ND	ND	0.12	0.21	ND	0.18	ND	ND	0.19	0.39	0.38	0.29
Caprylic acid (C8:0)	5.34	5.41	6.04	5.19	4.45	5.56	4.29	6.49	6.49	7.81	7.54	5.92
Capric acid (C10:0)	5.47	4.52	5.13	7.78	4.23	4.86	4.20	5.32	5.95	6.15	5.79	4.92
Lauric acid (C12:0)	45.44	46.01	47.98	50.31	48.71	50.25	43.27	47.18	50.74	48.81	49.65	48.52
Myristic acid (C14:0)	20.46	21.26	20.21	21.16	21.10	20.36	19.86	19.96	18.23	19.60	19.79	20.69
Palmitic acid (C16:0)	10.23	10.57	9.49	8.00	10.22	9.21	12.45	9.65	8.41	8.93	8.61	9.55
Palmitoleic acid (C16:1)	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	2.53	ND
Stearic acid (C18:0)	2.63	2.87	2.36	3.05	2.73	3.16	5.40	2.96	2.64	2.84	4.02	2.80
Oleic acid (C18:1)	8.93	8.33	7.39	3.82	7.40	5.58	11.15	5.63	5.85	4.20	2.61	6.12
Linoleic acid (C18:2)	1.50	1.04	1.28	0.42	1.15	0.85	3.65	1.42	1.38	1.10	0.73	1.08
Arachidic acid (C20:0)	ND	ND	ND	0.06	ND	ND	ND	ND	ND	ND	ND	ND
Unsaturated FA	10.42	9.36	8.67	4.23	8.56	6.43	14.79	7.05	7.23	5.30	5.87	7.19
Saturated MCFA	56.25	55.93	59.27	63.48	57.39	60.85	51.76	58.99	63.36	63.16	63.35	59.64
Saturated LCFA	33.32	34.70	32.06	32.21	34.05	32.72	37.70	32.57	29.27	31.37	32.42	33.04

*NOTE: FA - fatty acids; MCFA - medium chain fatty acids; LCFA - long chain fatty acids; ND - Non Detectable, defined as ≤ 0.05%.

RT-qPCR in the present study reveals contrasting oil biosynthesis profiles for LAGT and PLNT. Both FA elongation and desaturation are critical points in oil biosynthesis. The moderate levels of saturated and unsaturated MCFA and LCFA in LAGT could be correlated with the increased transcription of *ELO3* and *d8-DES2*. Substrate-specific acyl-ACP thioesterases that terminate FA synthesis determine the available FA for TAG biosynthesis. Relatively higher *FatA* transcript levels coincide with the high amounts of unsaturated FA in the LAGT endosperm. By contrast, increased levels of C12:0 and C16:0 in the PLNT endosperm are accompanied by relatively higher amounts of *FatB3* transcription. Oleosins participate in seed oil body biogenesis and seed development. The sudden increase of *OLE5* transcription in PLNT at 7 MAP was not observed in the developing normal endosperm. The highest level of *OLE5* transcription in LAGT at 8 MAP was noticeably lower than that observed in the PLNT endosperm at the same stage.

Additional genes could still be verified using RT-qPCR to complete the ongoing work on PLNT. RNA-Seq analysis of subsequent developmental stages beyond 7 MAP would complement existing RT-qPCR data. The oil content in each developmental stage must be quantified and compared between LAGT and PLNT to clarify other differences between the two endosperm variants. The effect of oleosin overexpression on endosperm development could also be explored. The genetics of 'Lono' endosperm occurrence has yet to be established, and its mode of inheritance should be elucidated by further study.

SUPPLEMENTARY TABLE

Supplementary Table 1. Information for 'Philippine Lono Tall' registered at the Central Plantation Crops Research Institute in Kasaragod, Kerala, India (CPCRI 2013).

A. PLANT MORPHOLOGY			
Plant height	597.75 (SD=78.77)	Length of leaflet	124.50 cm (SD=14.77)
Total number of leaves	33.63 cm (SD=2.45)	Breadth of leaflet	5.76 cm (SD=0.60)
Length of petiole	116.87 cm (SD=8.87)	Girth of trunk	91.37 cm (SD=4.21)
Length of leaflet bearing portion	436.00 cm (SD=31.31)	Number of leaf scars in 1 m length	28.25 (SD=3.54)
Number of leaflets (right)	117.87 (SD=10.76)	Length of 10 internodes	47.37 (SD=6.16)
Number of leaflets (left)	116.38 (SD=11.48)		

Cont. Supplementary Table 1

B. REPRODUCTIVE BIOLOGY			
Age at first flowering	130.50 mo	No. of ♀ flowers/inflorescence	27.00 (SD=5.93)
Length of inflorescence	109.50 cm (SD=9.64)	Number of inflorescences/year	12
Length of spikelet bearing portion	61.38 cm (SD=2.83)	Length of male phase	20.00 days
Length of stalk	48.13 cm (SD=7.95)	Length of female phase	5.00 days
Length of spikelet	37.25 cm (SD=2.92)	Gap between ♂ and ♀ phases	3.00 days
No. of spikelets in inflorescence	37.63 (SD=2.97)	Intraspadix/interspadix overlapping	0.00/5.00 days
C. HARVEST DATA			
Age at first harvest	148.50 mo	Number of nuts/year	40.63 (SD=18.98)
Setting percentage	22.00%	Bunches with button	7.63 (SD=1.51)
Number of bunches/year	8.88 (SD=2.03)	Bunches with nuts	6.88 (SD=1.73)
D. FRUIT CHARACTERISTICS			
1. Tendernut		Weight of nut	795 g
Quantity of water	340.00 mL	Percentage of husk in whole fruit	47.64% (SD=2.59)
Total sugar content	5.32 g/100 mL	Thickness of kernel	1.60 cm (SD=0.21)
Potassium content	2450.00 ppm	Weight of kernel	338.00 g
Sodium content	29.64 ppm	Thickness of shell	0.30 cm (SD=0.02)
Amino acids	1.60 mg/100 mL	Copra content/nut	243.00 g/nut
2. Mature nut		Copra/palm/year	9.87 kg
Length of fruit	30.60 cm (SD=0.75)	Copra/ha	1.73 t
Breadth of fruit	19.00 cm (SD=0.60)	Oil content	64.50%
Weight of fruit	1509.00 g	Oil/ha	1.11 t
Thickness of husk	3.15 cm (SD=0.21)		
E. BIOCHEMICAL TRAITS			
Leaf protein content	3.69 mg/g FW	Leaf polyphenol content	5.53 mg/g FW

Supplementary Table 2. Nucleotide and protein sequence data of *Cocos nucifera* oleosin genes found in the 'Laguna Tall' and 'Lono' transcriptome comparison at 6–7 months after pollination.

Sequence ID	Description	E-value	Identity (%)	GenBank ID	Reference*
Nucleotide BLAST					
Contig 1065	oleosin isoform OLE500a mRNA, partial cds	0.00E+00	526/529 (99%)	FJ012156.1	1
	oleosin isoform 500a mRNA, complete cds	0.00E+00	395/396 (99%)	KU711778.1	2
Contig 2443	oleosin isoform OLE500c mRNA, partial cds	0.00E+00	501/504 (99%)	FJ012157.1	1
	oleosin isoform 500c mRNA, complete cds	3.00E-138	275/276 (99%)	KU711779.1	2
Contig 315158	oleosin isoform 300a mRNA, complete cds	5.00E-42	110/115 (96%)	KU711777.1	2
	oleosin isoform OLE300a mRNA, partial cds	5.00E-42	110/115 (96%)	FJ012155.1	1
	oleosin mRNA, complete cds	7.00E-170	392/423 (93%)	MH158654.1	3
Contig 93807	oleosin isoform 500a mRNA, complete cds	1.00E-08	85/107 (79%)	KU711778.1	2
	oleosin isoform OLE500a mRNA, partial cds	5.00E-07	82/104 (79%)	FJ012156.1	1

Cont. Supplementary Table 2

Sequence ID	Description	E-value	Identity (%)	GenBank ID	Reference*
	Protein BLAST				
Contig 1065	oleosin isoform 500a	5.00E-49	130/131 (99%)	AQT25673.1	2
	oleosin isoform OLE500a, partial	3.00E-39	73/73 (100%)	ACH91012.1	1
	oleosin isoform 500c	1.00E-30	101/128 (79%)	AQT25674.1	2
	oleosin isoform OLE500c, partial	2.00E-28	56/70 (80%)	ACH91013.1	1
Contig 2443	oleosin isoform 500c	4.00E-50	91/91 (100%)	AQT25674.1	2
	oleosin isoform OLE500c, partial	1.00E-40	76/76 (100%	ACH91013.1	1
	oleosin isoform 500a	3.00E-33	73/92 (79%)	AQT25673.1	2
	oleosin isoform OLE500a, partial	8.00E-33	63/79 (80%)	ACH91012.1	1
Contig 93807	oleosin	3.00E-76	124/140 (89%)	AZZ09171.1	3

*NOTE: 1 - Regalado et al. 2008, 2 - Vargas et al. 2017, 3 - Reynolds et al. 2019.

CONCLUSION

To aid in understanding the genetic and molecular mechanisms that give rise to the Lono endosperm phenotype, a pilot RNA-Seq survey of differentially expressed genes between the normal LAGT and the mutant PLNT was performed. The present study provides evidence of the early upregulation of oil biosynthesis genes such as *d8-DES2*, *ELO3*, *FatA*, *FatB*, *OLE16kDa* isoform1, *OLE16kDa* isoform2, *OLE18kDa*, and *OLE5* in the growing PLNT endosperm as compared with the more common LAGT variety. A molecular mechanism to explain the PLNT phenotype may be further elucidated by validating other differentially expressed genes that were identified by the present RNA-Seq analysis. Reference-guided assembly of the LAGT and PLNT transcriptomes using the recently-published Catigan Dwarf coconut genome (Lantican et al. 2019) is ongoing. An improved understanding of endosperm development in 'Lono' fruits would enable the design of appropriate propagation methods to compensate for the difficulty of establishing true-to-type PLNT breeding populations.

ACKNOWLEDGEMENTS

The authors thank the Department of Science and Technology (DOST) - Philippine Council for Agriculture, Aquatic, and Natural Resources Research and Development (PCAARRD) for the funding and implementation of the study. We also acknowledge the invaluable assistance of Dr. John E. Carlson of the Schatz Center of Tree Molecular Genetics at the Pennsylvania State University for facilitating the next-generation sequencing of some of our RNA samples. Lastly, we thank the Philippine Coconut Authority - Albay Research Center (PCA-ARC) for providing all of the PLNT fruits used in this study.

REFERENCES CITED

AL-DOUS EK, GEORGE B, AL-MAHMOUD ME, AL-JABER MY, WANG H, SALAMEH YM, AL-AZWANI EK, [...] MALEK JA. 2011. *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). Nature Biotech 29(6):521-527. doi: 10.1038/nbt.1860.

AL-MSSALLEM IS, HU S, ZHANG X, LIN Q, LIU W, TAN J, YU X, [...] YU J. 2013. Genome sequence of the date palm *Phoenix dactylifera* L. Nature Comm 4:2274. doi: 10.1038/ncomms3274.

APUYA NR, YADEGARI R, FISCHER RL, HARADA JJ, GOLDBERG RB. 2002. *RASPBERRY3* gene encodes a novel protein important for embryo development. Plant Physiol 129(2):691-705. doi: 10.1104/pp.004010.

BAGGERLEY KA, DENG L, MORRIS JS, ALDAZ CM. 2003. Differential expression in SAGE: accounting for normal between-library variation. Bioinformatics 19(12):1477-1483. doi: 10.1093/bioinformatics/btg173

BARRE A, SIMPLICIEN M, CASSAN G, BENOIST H, ROUGE P. 2018. Oil bodies (oleosomes): occurrence, structure, allergenicity. Rev Fr Allergol 58(8):574-580. doi: 10.1016/j.reval.2018.10.005.

BATUGAL PB, BOURDEIX R, OLIVER JT, GEORGE MLC. 2010. Catalogue of Conserved Coconut Germplasm. Serdang Selangor Darul Ehsan, Malaysia: International Coconut Genetic Resources Network

(COGENT), Bioversity International, Regional Office for Asia, the Pacific and Oceania. 397 p.

BERGER F. 1999. Endosperm development. Curr Opin Plant Biol 2:28-32. doi: 10.1016/S1369-5266(99)80006-5.

BOURGIS F, KILARU A, CAO X, NGANDO-EBONGUE GF, DRIRA N, OHLROGGE JB, ARONDEL V. 2011. Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. PNAS 108(30):12527-12532. doi: 10.1073/pnas.1106502108.

BEZARD J, BUGAUT M, CLEMENT G. 1971. Triglyceride composition of coconut oil. J Am Oil Chem Soc 48(3):134-139. doi:10.1007/bf02545736.

CALDO KMP, GARCIA RN, TECSON-MENDOZA EM. 2015. Biochemical and molecular characterization of two 11S globulin isoforms from coconut and their expression analysis during seed development. Int J Philipp Sci Technol 8:46-51.

[CPCRI] Central Plantation Crops Research Institute. 2013. Annual Report 2012-13. Kerala, India: Central Plantation Crops Research Institute. 142 p.

CHEN M, MARKHAM JE, CAHOON EB. 2012. Sphingolipid delta(8) unsaturation is important for glucosylceramide biosynthesis and low-temperature performance in Arabidopsis. Plant J 69(5):769-781.

CONESA A, GÔTZ S, GARCÌA-GÛMEZ JM, TEROL1 J, TALÛN M, ROBLES M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674-3676. doi: 10.1093/bioinformatics/bti610.

DELA CRUZ RY, LAUDE RP, DIAZ MGQ, LAURENA AC, MENDIORO MS, MENDOZA EMT. 2011. Gene for actin is a suitable internal reference for relative RT-PCR based expression analysis in normal and mutant 'Makapuno' endosperms of coconut (*Cocos nucifera* L.). Philipp Agric Scientist 94(2):118-123.

DUSSERT S, GUERIN C, ANDERSSON M, JOËT T, TRANBARGER TJ, PIZOT M, SARAH G, OMORE A, DURAND-GASSELIN T, MORCILLO F. 2013. Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. Plant Physiol 162(3):1337-1358. doi: 10.1104/pp.113.220525.

[FAO/WHO] Joint FAO/WHO Codex Alimentarius Commission. Codex alimentarius: Standard for named vegetable oils (CODEX STAN 210-1999, last amended 2019). Rome: World Health Organization: Food and Agriculture Organization of the United Nations, 2007. Retrieved January 31, 2019 from the World Wide Web: http://www.fao.org/fao-who-codexalimentarius/codex-texts/list-standards/en/.

FAN H, XIAO Y, YANG Y, XIA W, MASON AS, XIA Z, QIAO F, ZHAO S, TANG H. 2013. RNA-Seq analysis of *Cocos nucifera*: transcriptome sequencing and de novo assembly for subsequent functional genomics approaches. PLoS one 8(3):e59997. doi: 10.1371/journal.pone.0059997.

HASLAM TM, KUNST L. 2013. Extending the story of very-long-chain fatty acid elongation. Plant Sci 210:93-107. doi: 10.1016/j.plantsci.2013.05.008.

HOTELLING H. 1931. The generalization of Student's ratio. Ann Math Stat 2(3):360–378. doi: 10.1214/aoms/1177732979.

JAMEK SB, CHENG NG, GUAN CT. 2010. The isolation and amplification of full length cDNA of oleosins from oil palm (*Elaeis guineensis* Jacq.). Afr J Biotechnol 9(13):1859-1863. doi: 10.5897/AJB09.1248.

JOËT T, LAFFARGUE A, SALMONA J, DOULBEAU S, DESCROIX F, BERTRAND B, DE KOCHKO A, DUSSERT S. 2009. Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. New Phytol 182:146-162. doi: 10.1111/j.1469-8137.2008.02742.x.

JOHNSON KL, JONES BJ, BACIC A, SCHULTZ CJ. 2003. The fasciclin-like arabinogalactan proteins of Arabidopsis: a multigene family of putative cell adhesion molecules. Plant Physiol 133:1911-1925. doi: 10.1104/pp.103.031237.

KAWAHARA Y, DE LA BASTIDE M, HAMILTON JP, KANAMORI H, MCCOMBIE WR, OUYANG S, SCHWARTZ DC, TANAKA T, WU J, ZHOU S, CHILDS KL. 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice 6(1):4. doi: 10.1186/1939-8433-6-4.

KUMAR NS. 2007. Capillary gas chromatography method for fatty acid analysis of coconut oil. J Plantn Crops 35(1):23-27.

KUMAR NS. 2011. Variability in Coconut (*Cocos nucifera* L.) Germplasm and hybrids for fatty acid profile of oil. J Agril Food Chem 59(24): 13050-13058. doi: 10.1021/jf203182d.

KUMAR NS, BALAKRISHNAN A. 2009. Seasonal variations in fatty acid composition of oil in developing coconut. J Food Quality (USA) 32:158-176. doi: 10.1111/j.1745-4557.2009.00243.x.

LANTICAN DV, STRICKLER SR, CANAMA AO, GARDOCE RR, MUELLER LA, GALVEZ HF. 2019. *De novo* genome sequence assembly of dwarf coconut (*Cocos nucifera* L. 'Catigan Green Dwarf') provides insights into genomic variation between coconut types and related palm species. G3 9(8):2377-2393. doi: 10.1534/g3.119.400215.

LAURELES LR, RODRIGUEZ FM, REAÑO CE, SANTOS GA, LAURENA AC, MENDOZA EMT. 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(6):1581-1586. doi: 10.1021/jf010832w.

LEE J, DAS A, YAMAGUCHI M, HASHIMOTO J, TSUTSUMI N, UCHIMIYA H, UMEDA M. 2003. Cell cycle function of a rice B2-type cyclin interacting with a B-type cyclin-dependent kinase. Plant J 34(4):417-425. doi: 10.1046/j.1365-313X.2003.01736.x.

LI Z, TRICK HN. 2005. Rapid method for high-quality RNA isolation from seed endosperm containing high levels of starch. BioTechniques 38:872-876. doi: 10.2144/05386BM05.

LI R, ZHU H, RUAN J, QIAN W, FANG X, SHI Z, LI Y, LI S, SHAN G, KRISTIANSEN K, LI S, YANG H, WANG J, WANG J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res 20(2):265-272. doi: 10.1101/gr.097261.109.

LUKOWITZ W, MAYER U, JÜRGENS G. 1996. Cytokinesis in the Arabidopsis embryo involves the syntaxin-related KNOLLE gene product. Cell 84(1):61-71. doi: 10.1016/S0092-8674(00)80993-9.

MAUCHLY JW. 1940. Significance test for sphericity of a normal *n*-variate distribution. Ann Math Statist 11(2):204-209. doi: 10.1214/aoms/1177731915.

MORENO-PÉREZ A. J., VENEGAS-CALERÓN, M., VAISTIJ, F. E., SALAS, J. J., LARSON, T. R., GARCÉS, R., GRAHAM, I. A., MARTÍNEZ-FORCE, E. 2011. Reduced expression of FatA thioesterases in Arabidopsis affects the oil content and fatty acid composition of the seeds. Planta 235(3):629-39.

NOWAK K, GAJ MDJ. 2016. Stress-related function of *bHLH109* in somatic embryo induction in *Arabidopsis*. Plant Physiol 193:119-126. doi: 10.1016/j.jplph.2016.02.012.

PADOLINA WG. 1985. Contribution of biological research to the development of the coconut industry. JAOCS 62:206-210. doi: 10.1007/BF02541380.

PUNZALAN MR, BAUTISTA MA, EMMANUEL E, RIVERA R, RIVERA S, SALOMA CP. 2019. Transcriptome of the traditional coconut variety Laguna Tall. PJS Special Issue on Genomics 148(S1):153-164.

QUIST TM, SOKOLCHIK I, SHI H, JOLY RJ, BRESSAN RA, MAGGIO A, NARSIMHAN M, LI X. 2009. HOS3, an ELO-like gene, inhibits effects of ABA and implicates a S-1-P/ceramide control system for abiotic stress responses in *Arabidopsis thaliana*. Molecular Plant 2(1):138-151. doi: 10.1093/mp/ssn085.

R CORE TEAM. 2019. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

RAGHAVENDRA SN, RAGHAVARAO KSMS. 2010. Effect of different treatments for the destabilization of coconut milk emulsion. J Food Eng 97:341-347. doi: 10.1016/j.jfoodeng.2009.10.027.

REGALADO ES, ANGELES JGC, RODELAS AJD, GARCIA RN, LAURENA AC, TECSON-MENDOZA EM. 2008. Cloning and characterization of the oleosin cDNA isoforms from coconut (*Cocos nucifera* L.). Philipp Agric Scientist 91:287-294.

RESHEF L, OLSWANG Y, CASSUTO H, BLUM B, CRONIGER CM, KALHAN SC, TILGHMAN SM, HANSON RM. 2003. Glyceroneogenesis and the triglyceride/fatty acid cycle. J Biol Chem 278(30):413-430. doi: 10.1074/jbc.R300017200.

REYNOLDS KB, CULLERNE DP, EL TAHCHY A, ROLLAND V, BLANCHARD CL, WOOD CC, SINGH SP, PETRIE JR. 2019. Identification of genes involved in lipid biosynthesis through de novo transcriptome assembly from *Cocos nucifera* developing endosperm. Plant Cell Physiol 00(0):1-16. doi: 10.1093/pcp/pcy247.

RILLO EP. 2014. Lono: another soft endosperm coconut in the Philippines. Cocoinfo International 21(1):12-13.

RODRIGUEZ MJ. 2013. Terminal report. Development and optimization of village-level production of VCO, other plant oils, galactomannan and by-products. Quezon City, Philippines: Philippine Coconut Authority.

SANTOS GA, CANO SB, DELA CRUZ BV, ILGAN MC, BAHALA RT. 1984. Coconut germplasm collection in the Philippines. Phil J Coconut Studies 9:1-9

SHAPIRO SS, WILK MB. 1965. An analysis of variance test for normality (complete samples). Biometrika 52(3/4):591-611. doi: 10.2307/2333709.

SILOTO RM, FINDLAY K, LOPEZ-VILLALOBOS A, YEUNG EC, NYKIFORUK CL, MOLONEY MM. 2006. The accumulation of oleosins determines the size of seed oil bodies in *Arabidopsis*. Plant Cell 18(8):1961-1974. doi: 10.1105/tpc.106.041269.

SINGH R, ONG-ABDULLAH M, LOW ET, MANAF MA, ROSLI R, NOOKIAH R, OOI LCL, [...] SAMBANTHAMURTHI R. 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature 500(7462):335-339. doi: 10.1038/nature12309.

TRONCOSO-PONCE MA, KILARU A, CAO X, DURRETT TP, FAN J, JENSEN JK, THROWER NA, PAULY M, WILKERSON C, OHLROGGE JB. 2011. Comparative deep transcriptional profiling of four developing oilseeds. Plant J 68:1014-1027. doi: 10.1111/j.1365-313X.2011.04751.x.

VARGAS AG, CABANOS CS, GARCIA RN, MARUYAMA N, TECSON-MENDOZA EM. 2018. Cloning, molecular analysis, and developmental expression of 3 oleosin cDNA isoforms in coconut (*Cocos nucifera* L.). J Hortic Sci Biotech 93(3):255-263. doi: 10.1080/14620316.2017.1364145.

VERRIER PJ, BIRD D, BURLA B, DASSA E, FORESTIER C, GEISLER M, KLEIN M, KOLUKISAOGLU Ü, LEE Y, MARTINOIA E, MURPHY A. 2008. Plant ABC proteins – a unified nomenclature and updated inventory. Trend Plant Sci 13(4):151-159. doi: 10.1016/j.tplants.2008.02.001.

YE J, COULOURIS G, ZARETSKAYA I, CUTCUTACHE I, ROZEN S, MADDEN TL. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13(1):134. doi: 10.1186/1471-2105-13-134.

ZUNIGA LC, VILLEGAS LG, PENAFLORIDA G. 1970. Collection of coconut cultivars in the Philippines. In: Emata RG, editors. Coconut Production in the Philippines. Manila, Philippines: United Association of the Philippines. p. 9-24.