Liang et al. BMC Plant Biology 2014, 14:205
http://www.biomedcentral.com/1471-2229/14/205

BMC Plant Biology

RESEARCH ARTICLE Open Access

Identification and computational annotation of genes differentially expressed in pulp development of *Cocos nucifera* L. by suppression subtractive hybridization

Yuanxue Liang[1], Yijun Yuan[1], Tao Liu[2], Wei Mao[1], Yusheng Zheng[1] and Dongdong Li[1,3*]

Abstract

Background: Coconut (*Cocos nucifera* L.) is one of the world's most versatile, economically important tropical crops. Little is known about the physiological and molecular basis of coconut pulp (endosperm) development and only a few coconut genes and gene product sequences are available in public databases. This study identified genes that were differentially expressed during development of coconut pulp and functionally annotated these identified genes using bioinformatics analysis.

Results: Pulp from three different coconut developmental stages was collected. Four suppression subtractive hybridization (SSH) libraries were constructed (forward and reverse libraries A and B between stages 1 and 2, and C and D between stages 2 and 3), and identified sequences were computationally annotated using Blast2GO software. A total of 1272 clones were obtained for analysis from four SSH libraries with 63% showing similarity to known proteins. Pairwise comparing of stage-specific gene ontology ids from libraries B-D, A-C, B-C and A-D showed that 32 genes were continuously upregulated and seven downregulated; 28 were transiently upregulated and 23 downregulated. KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis showed that 1-acyl-sn-glycerol-3-phosphate acyltransferase (*LPAAT*), phospholipase D, acetyl-CoA carboxylase carboxyltransferase beta subunit, 3-hydroxyisobutyryl-CoA hydrolase-like and pyruvate dehydrogenase E1 β subunit were associated with fatty acid biosynthesis or metabolism. Triose phosphate isomerase, cellulose synthase and glucan 1,3-β-glucosidase were related to carbohydrate metabolism, and phosphoenolpyruvate carboxylase was related to both fatty acid and carbohydrate metabolism. Of 737 unigenes, 103 encoded enzymes were involved in fatty acid and carbohydrate biosynthesis and metabolism, and a number of transcription factors and other interesting genes with stage-specific expression were confirmed by real-time PCR, with validation of the SSH results as high as 66.6%. Based on determination of coconut endosperm fatty acids content by gas chromatography–mass spectrometry, a number of candidate genes in fatty acid anabolism were selected for further study.

Conclusion: Functional annotation of genes differentially expressed in coconut pulp development helped determine the molecular basis of coconut endosperm development. The SSH method identified genes related to fatty acids, carbohydrate and secondary metabolites. The results will be important for understanding gene functions and regulatory networks in coconut fruit.

Keywords: Coconut, Computational annotation, Differentially expressed genes, Fatty acids, Pulp development, Suppression subtractive hybridization

* Correspondence: liddfym@hotmail.com
[1]Department of Biotechnology, College of Materials and Chemical Engineering, Hainan University, Haikou, Hainan 570228, PR China
[3]Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresource, Hainan University, Haikou, Hainan 570228, PR China
Full list of author information is available at the end of the article

Background

Cocos nucifera L. (coconut) belongs to the Arecaceae family and is the only accepted species in the *Cocos* genus. Coconut is one of the world's most versatile economically important tropical crops and is well known for its commercial and industrial applications in tropical and subtropical areas, such as in food and beverages and as a source of wood and handicrafts [1]. The pulp (endosperm) of the coconut is initially suspended in the water phase of the coconut [2], which is a liquid endosperm. As development continues, the liquid endosperm is gradually deposited to the coconut inner wall, becoming the edible coconut pulp (solid endosperm). In mature coconut, the pulp (28% weight) is surrounded by a hard protective shell (12% weight), which is covered by a thick husk (35% weight) [3].

Oil extracted from coconut pulp is widely applied in cooking, soaps and cosmetics. Coconut is one of the few plants that store medium chain-length fatty acids (MCFAs) as the major portion of their energy reserves in the endosperm of seeds. In developed coconut fruit, more than 83.92% of the oil consists of MCFAs and long-chain fatty acids ($C_{12:0}$, $C_{14:0}$, $C_{16:0}$ and $C_{18:0}$), the majority of which is lauric acid ($C_{12:0}$) ranging from 47.48% to 50.5% [4]. Coconut also has more MCFAs than soybean [5], oil palm and safflower, and animal fats such as butter, tallow, fish oil and lard [6-10]. The special chemical composition of coconut oil makes it useful for a range of edible and nonedible purposes. In addition, coconut oil has unique features like pleasant odor, high resistance to rancidity, narrow melting temperature range and superior foam retention capacity for use in whipped toppings [11]. Coconut pulp also contains fiber, starch, sugars and sugar alcohols such as sucrose, glucose, fructose, mannitol, sorbitol, myoinositol and scylloinositiol [12], which are important for fatty acid and sugar biosynthesis and metabolism.

The composition of coconut endosperm has been widely reported [12-14], but the molecular basis is poorly understood, especially the dynamic expression of stage-specific genes and endosperm development. Most research has focused on evaluating genetic diversity using microsatellite markers [15-17]. Recently, to investigate gene expression profiles in endosperm development, cDNA libraries were constructed from coconut endosperm [18]. Transcriptome sequencing of spear leaves, young leaves and fruit flesh [19] and comparison to coconut chloroplast [20], and isolation of miRNAs (microRNAs) differentially expressed during solid endosperm development predicted the potential functional of some miRNAs [21]. A number of genes have been cloned and characterized including for *oleosin* [22],1-acyl-sn-glycerol-3-phosphate acyltransferase (*LPAAT*) [23], somatic embryogenesis receptor-like kinase gene [24], cyclin-dependent kinase [25] and the genes *CnFatB 1* [26], *CnFatB 2* and *CnFatB 3* [27]. Most previous research on the composition of coconut endosperm has included little molecular analysis. Even if numerous ESTs were obtained, few were annotated and many sequences were redundant. No published studies describe dynamic expression analysis of stage-specific genes and the related mechanisms of the development of coconut endosperm.

Suppression subtractive hybridization (SSH) is a powerful and popular strategy for screening subtracted cDNA. In a model system, SSH can be applied to enrich rare sequences by >1000-fold in a single round of subtractive hybridization [28]. At present, transcriptome sequencing technology can also obtain information about transcripts, but is tedious and inefficient and the identified sequences include many redundancies. This makes analysis of transcriptome data complex because of extensive background noise. SSH efficiently obtains many sequences with an average length of >500 bp. Therefore, SSH is an important alternative to molecular genetics and positional cloning for identifying genes that are differentially expressed in disease, development or specific tissues.

In this paper, differentially expressed genes at three coconut pulp developmental stages were characterized by SSH. Computational annotation was carried out using Blast2GO software. The efficiency of SSH was demonstrated by quantitation with real-time PCR and by content analysis of fatty acids.

Results

Samples collection and fatty acid analysis

Pulp from three different developmental stages of coconut was collected and total fatty acids were extracted and analyzed by gas chromatography–mass spectrometry (GC-MS) in three sets of parallel experiments. The mean percentage oil content of fresh coconut pulp by weight was 1.94% at stage 1, 8.73% at stage 2, and 10.03% at stage 3 (Figure 1A). This indicated a gradual accumulation in oil during coconut development. Various fatty acids with different relative content were detected at specific stages. In stage 1, the unsaturated fatty acid $C_{18:2}\Delta^{9,12}$ was the major fatty acid (34.1%), followed by $C_{12:0}$ (15.1%), $C_{14:0}$ (12.0%), $C_{16:0}$ (16.9%) and $C_{18:0}$ (16.6%) (Figure 1B). The saturated fatty acids $C_{12:0}$, $C_{14:0}$, $C_{16:0}$ and $C_{18:0}$ were the major fatty acids in stages 2 and 3. In stage 2, $C_{12:0}$ represented 42.2%, $C_{14:0}$ 20.0%, $C_{16:0}$ 11.8% and $C_{18:0}$ 12.4%. In stage 3, $C_{12:0}$ represented 46.9%, $C_{14:0}$ 20.9%, $C_{16:0}$ 11.2% and $C_{18:0}$ 5.0% (Figure 1B). Fatty acid contents from 10 g of coconut pulp from three different developmental stages were determined (Figure 1C). The saturated fatty acids $C_{12:0}$, $C_{14:0}$, $C_{16:0}$ and $C_{18:0}$ rapidly increased during development of coconut pulp from stage 1 to 3. Statistical analysis of corresponding fatty acid indicated that differences in content of each fatty acid between stages 2 and 1 were all significant ($P \leq 0.01$) except that $C_{18:2}\Delta^{9,12}$was significant at $0.01 \leq P \leq 0.05$. Between stages 3 and 2, the contents of $C_{10:0}$ and $C_{12:0}$ fatty

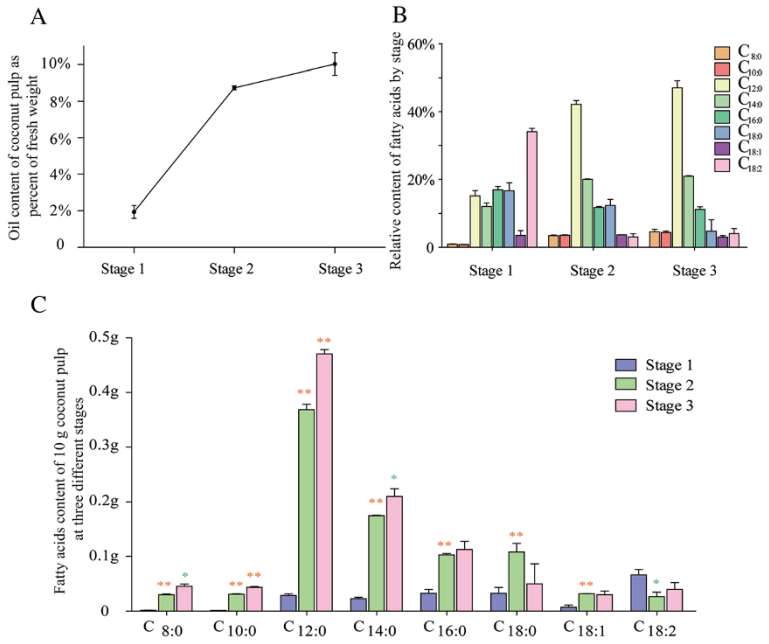

Figure 1 Content change in oil and fatty acids for stages 1, 2 and 3 of coconut pulp. (A) Oil content of coconut pulp as percent of fresh weight at stages 1, 2 and 3. Mean percentages were 1.94%, 8.73% and 10.03%. **(B)** Relative content of fatty acids by stage. In stage 1 unsaturated fatty acid $C_{18:2}\Delta^{9,12}$ (34.1%) was the major component, rather than $C_{12:0}$ as in stages 2 and 3. **(C)** Fatty acid content of 10 g coconut pulp at three developmental stages. Each fatty acid was analyzed for statistical significance between stages 1 and 2, and between stages 2 and 3. Significant differences according to Student's *t*-test: $^{*}P < 0.05$; $^{**}P < 0.01$.

acids were significantly different ($P \leq 0.01$), as were $C_{8:0}$ and $C_{14:0}$ ($0.01 \leq P \leq 0.05$). Other fatty acids showed no significance differences in content ($P > 0.05$). During the three developmental stages, $C_{8:0}$, $C_{10:0}$, $C_{12:0}$ and $C_{14:0}$ gradually accumulated while $C_{18:2}\Delta^{9,12}$ gradually decreased.

Construction and characterization of SSH libraries

Four SSH libraries (A, B, C and D) were generated from coconut pulp at the three different developmental stages (i.e. 1–3). Forward and reverse subtraction was performed for the libraries for stages 1 to 2 (libraries A and B), and stages 2 to 3 (libraries C and D). A total of 1272 clones were sequenced, yielding 1230 high-quality EST sequences with an average insert length of 500 bases after removing empty vectors. The lengths of all ESTs were more than the cutoff value of 100 bp (Table 1). The number of differentially expressed sequences that were randomly chosen for sequencing was 248–366 per library with 11 low-quality sequences (inserts < 100 bp) for library A, 9 for B, 8 for C and 4 for D; thus there were 244–345 high-quality sequences (insert length ≥ 100 bp) from A to D. The insertion length was 123–900 bp for library A, 141–862 bp for B, 110–838 bp for C and 106–825 bp for D. The average length for each library was 487–512 bp. Low-quality sequences, vector sequences and poly(A) ends were removed from each library using *Phred* and Crossmatch software. Preliminary analysis using the *Phrap* program isolated singlets and assembled overlapping sequences into contigs. Across the libraries, the number of singlets varied within 70–226 with average lengths of 484–522 bp. The number of contigs was 25–53 with average lengths of 629–726 bp. The number of unigenes (singlets and contigs) was 112–251 with average lengths of 503–574 bp. The single-sequence high-repetition redundancy for each library was within 2.6–7.4%. Size ranges and statistical values for clones, singlets, contigs, unigenes and redundancy are in Table 1. The sequences obtained in this research were deposited in the GenBank database and can be viewed on the NCBI website (http://www.ncbi.nlm.nih.gov) under accession numbers JZ533438 to JZ534322.

Species distribution of SSH EST BLAST top hits

BLAST top hits for sequences from the four SSH libraries retrieved from the NCBI database showed only 19 top hits with similarity to coconut sequences. Around 40% of top hits for coconut pulp developmental stage SSH EST sequences were homologous to coding sequences in oil palm, grape or rice, with >120 top hits per species (Figure 2). More than 60% of the BLAST hits were homologous to rice, grape, *Populus trichocarpa*, maize or soybean coding sequences (Additional file 1) with only 19 hits with similarity to coconut sequences.

Table 1 SSH library characteristics by clone abundance and length, proportion of singlets and contigs and related redundancy

	Clones				Singlets				Contigs				Unigenes				Redundancy %
library	No.	%	Mean (bp)	Size range (bp)	No.	%	Mean (bp)	Size range (bp)	No.	%	Mean (bp)	Size range (bp)	No.	%	Mean (bp)	Size range (bp)	
A	345	28.0	512	123-900	173	77.0	522	123-865	53	23.0	629	261-1401	226	30.7	546	123-1401	2.6
B	339	27.6	495	141-862	104	70.3	507	141-834	44	29.7	726	153-1620	148	20.1	572	141-1620	7.4
C	302	24.6	487	110-838	226	90.0	484	110-838	25	10.0	676	341-1541	251	34.0	503	110-1541	4.0
D	244	19.8	508	106-825	70	62.5	488	106-808	42	37.5	717	264-2295	112	15.2	574	106-2295	6.1
Total	1,230	100.0	500	106-900	573	77.7	500	106-865	164	22.3	685	153-2295	737	100.0	541	106-2295	—

Functional classification and gene ontology analysis of SSH EST sequences

After the removal of sequence impurities such as vector, short sequences and poly(A) ends, all sequences were analyzed by Blast2GO and aligned using BLASTX and the GenBank nonredundant database. Gene ontology (GO) mapping returned many terms. GO annotation assigned at least one GO term for 63.1% (776) of all SSH ESTs. GO annotations did not completely differ from each other, with many ESTs involved in different classes of functions and annotated with multiple GO terms; similarly, a single GO term was sometimes found to annotate several ESTs. Generally, sequences without BLAST hits (113) could not be annotated. The SSH ESTs (341) with BLAST hits that could not be annotated with GO terms were categorized as hypothetical or unknown proteins [29].

GO analysis describes biological processes, cellular components and molecular functions from genomes of several plants and animals [30]. Of complete GO terms for all SSH EST coconut sequences, 839 GO terms were in biological processes (Figure 3A), 1023 in cellular components (Figure 3B) and 547 in molecular functions (Figure 3C). GO terms and GO ids are shown in Additional file 2. In biological processes, the most highly represented categories were response to stress (GO:0006950, 14.8%), catabolic processes (GO:0009056, 12.8%) and carbohydrate metabolic processes (GO:0005975, 12.3%). Although numerically fewer, terms related to secondary metabolic processes (GO:0019748, 3.8%), lipid metabolic processes (GO:0006629, 5.8%) and ion transport (GO:0006811, 2.4%) were also found. In cellular components, the most representative categories were cytoplasmic membrane-bounded vesicles (GO:0016023, 22.3%), plasma membrane (GO:0005886, 13.4%) and plastids (GO:0009536, 11.1%). Cytosol (GO:0005829, 4.1%) and vacuoles (GO:0005773, 4.3%) were also represented. Lipid particle annotations were fewer (GO:0016023, 0.1%). Among molecular functions, the largest proportion of functionally assigned ESTs were in nutrient reservoir activity (GO:0045735, 40.6%), nucleotide binding (GO:0000166, 26.5%) and structural molecule activity (GO:0005198, 6.4%); the first two classes accounted for 67.1% of assignable GO terms.

Theoretically, SSH could obtain ESTs that were all stage-specific (or stage-associated) genes. We found that the EST sequences contained many differentially expressed genes but few stage-specific genes. To further analyze stage-specific genes, we compared GO ids from libraries A and B, and C and D for stage-specific GO ids (Figure 4 and Additional file 3). Among the 330 GO terms for library A and 311 for B, 203 were associated with downregulated and 184 with upregulated genes. In library A (stage 1), the most significant stage-specific GO terms were in cell division (GO:0051301), peroxidase reaction (GO:0006804), regulation of cell growth (GO:0001558), heme binding (GO:0020037), serine family amino acid metabolic processes (GO:0009069) and response to oxidative stress (GO:0006979). In libraries C and D, 360 and 251 GO terms were recovered, respectively, with 229 for downregulated and 120 for upregulated genes. For libraries B and C (stage 2), GO terms were mainly in nutrient reservoir activity (GO:0045735), acyl-carrier-protein biosynthetic (GO:0042967), fatty acid synthase complex (GO:0005835), response to heat (GO:0009408), cellulose synthase (GO:0016761), primary cell wall (GO:0009833) and carbohydrate metabolism (GO:0006011, GO:0005985 for sucrose and GO:0005982 for starch). In library D (stage 3), GO terms were associated with cell periphery synthesis

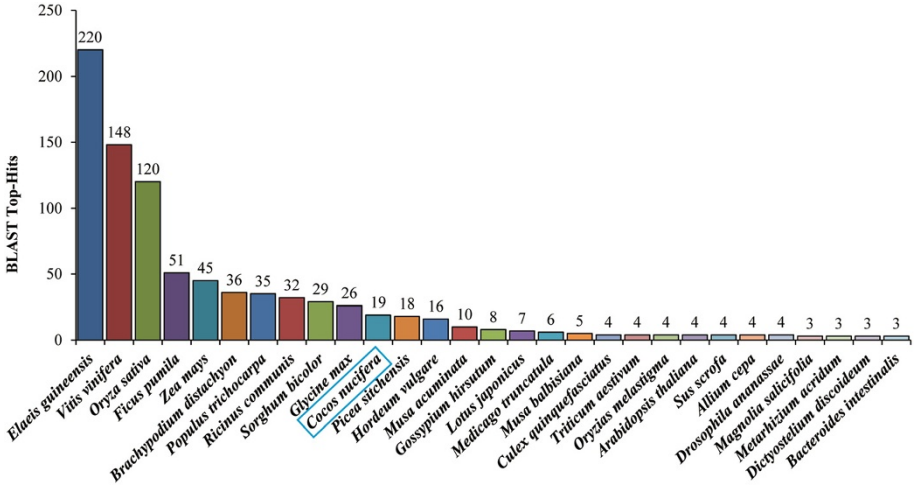

Figure 2 BLAST top hits by species distribution. Results of the highest similarity species with higher Max ident and lower e-values. Number of BLAST to -hits retrieved from NCBI databases. The majority of BLAST top hits were for oil palm, grapevine and rice; 19 were entries identified for coconut.

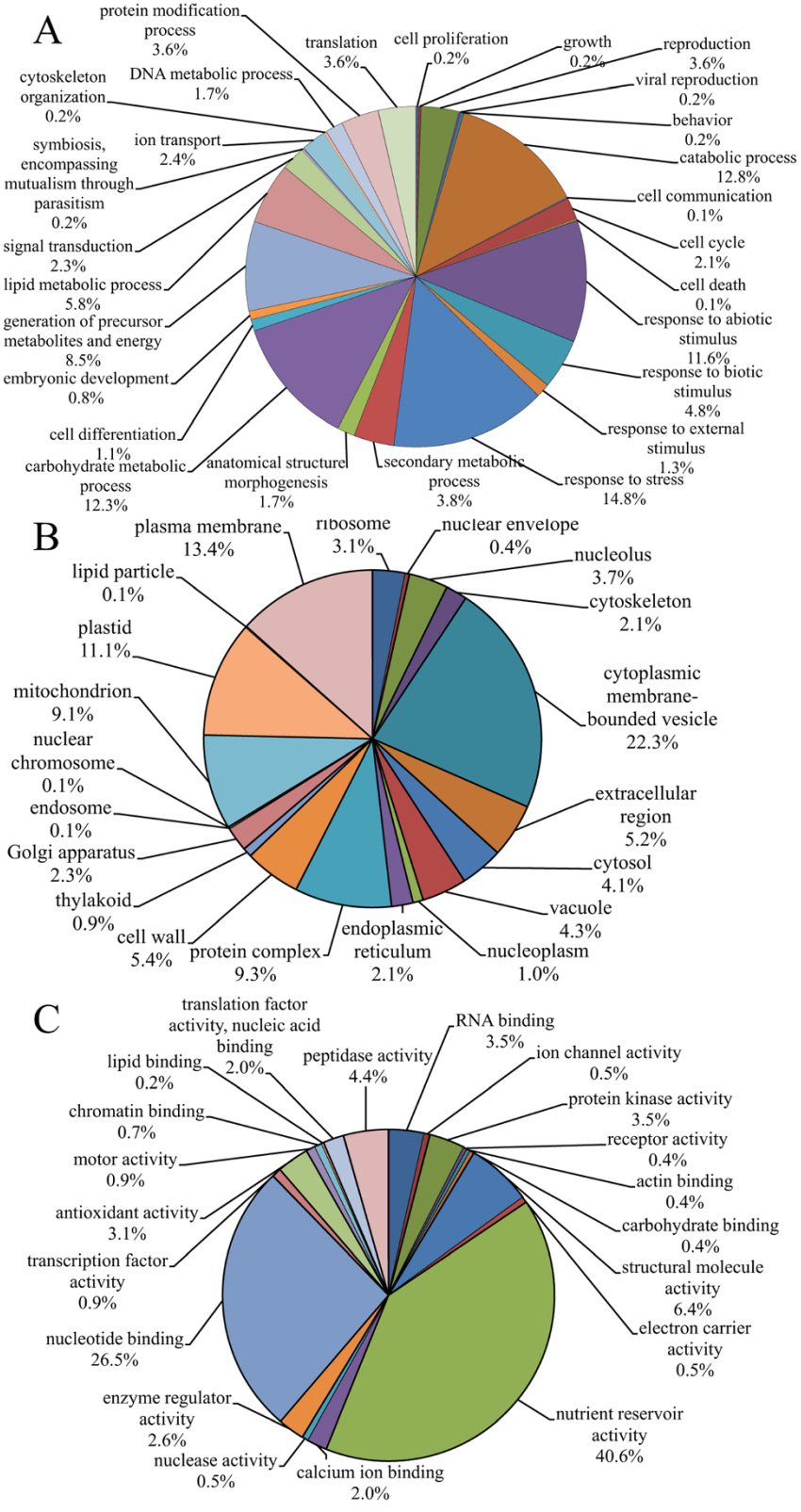

Figure 3 (See legend on next page.)

(See figure on previous page.)
Figure 3 Functional categorization of GO terms. Multilevel pie chart with the lowest node per branch from the annotations graph. GO terms were distributed into: **(A)** biological process, **(B)** cellular component, **(C)** molecular function. **A**: response to stress, catabolic processes and carbohydrate metabolic processes were the most common. **B**: cytoplasmic membrane-bounded vesicles, plasma membrane and plastids were the most common. **C**: Nutrient reservoir activity and nucleotide binding were most common.

(GO:0071944), nicotinamide adenine dinucleotide (NAD) transport (GO:0043132, GO:0051724), glycosyl bond hydrolase activity (GO:0016798, GO:0004553), defense response (GO:0006952, GO:0051707) and response to abiotic stimulus (GO:0009628).

Identification of continuously or transiently downregulated or upregulated genes

Genes showing continuous or transient downregulation or upregulation during the three developmental stages were obtained by comparing two SSH libraries. Comparing the GO ids of each pairwise forward and reverse SSH libraries for nonredundant stage-specific GO ids (Figure 4) yielded 203 for library A (330 GO ids in total), 184 for B (311 in total), 229 for C (360 in total) and 120 for D (251 in total). Pairwise comparisons of the nonredundant stage-specific GO ids, obtained above to identify stage-specific repeat GO ids, yielded 32 (Additional file 4) for libraries A-C (continuous downregulation), 28 (Additional file 4) for B-C (transient upregulation), 7 (Additional file 4) for B-D (continuous upregulation) and 23 (Additional file 4) for A-D (transient downregulation); with 148, 152, 116 and 124 repeat GO ids in pairwise comparisons of the total GO ids of A-C, B-C, B-D and A-D, respectively (Figure 5). In libraries A-C, stage-specific downregulated genes were mainly for 3-hydroxyisobutyryl-CoA hydrolase-like protein, acid phosphatase, ADP-ribosylation factor, 14-3-3-like protein, 2,3-bisphosphoglycerate-independent phosphoglyceratemutase, sarcosine oxidase, serine hydroxymethyl transferase, polynucleotide adeylyltransferase, multidrug-resistance protein ATP-binding cassette transporter and glutathione-S-transferase cla47. In libraries B-D, stage-specific upregulated genes were for alpha-tubulin, globulin precursor, 7-alpha-hydroxysteroid dehydrogenase, *LPAAT*, 7S globulin and triose phosphate isomerase cytosolic isoform. Among these genes, *LPAAT* was associated with fatty acid biosynthesis; and 7S globulin, triose phosphate isomerase cytosolic isoform, alpha-tubulin, globulin precursor and 7-alpha-hydroxysteroid dehydrogenase was related to sugar metabolism. In libraries B-C, stage-specific transiently upregulated or downregulated genes were mainly for oligopeptide transporter, phosphoenolpyruvate carboxylase (PEPC), ADP-ribosylation factor, enoyl-acylacyl carrier protein (ACP)-reductase protein, acetyl-CoA carboxylase carboxyl transferase beta subunit, nucleoside diphosphate kinase, cellulose synthase BoCesA6, glutaryl-CoA dehydrogenase, mannose-1-phosphate guanylyltransferase 3-like, bile acid:Na^+symporter family protein, mannose-1-phosphate guanylyltransferase 1-like, sucrose synthase 1 and UTP-glucose 1 phosphate uridylyltransferase. In libraries A-D, stage-specific transiently upregulated or downregulated genes were for beta-1,3-glucanase, pyruvate dehydrogenase E1 beta subunit, transport inhibitor response 1, V-type proton ATPase catalytic subunit A, voltage dependent anion channel protein, eukaryotic initiation factor iso-4 F subunit, casein kinase I-like, phospholipase D alpha 1 and uncharacterized protein LOC100820966.

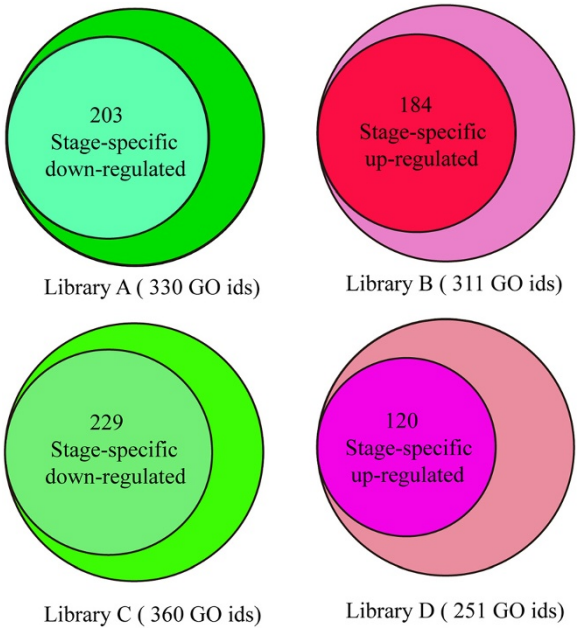

Figure 4 Venn diagrams for GO terms for libraries A, B, C and D. Libraries A, B, C and D contained 330, 311, 360 and 251 GO ids, respectively. Comparing libraries A and B showed 203 stage-specific downregulated GO ids and 184 upregulated. Comparing C and D libraries showed 229 stage-specific downregulated GO ids and 120 upregulated.

Real-time PCR and clustering map

Sequences that were chosen for clustering (Figure 6) were continuously or transiently upregulated or downregulated genes and fatty acid synthesis enzymes, transcription factors and other genes. Clustering was performed using quantitative real-time PCR (qRT-PCR) relative expression values. The results showed that 33 of 103 unigenes were upregulated in stage 2 versus stage 1,

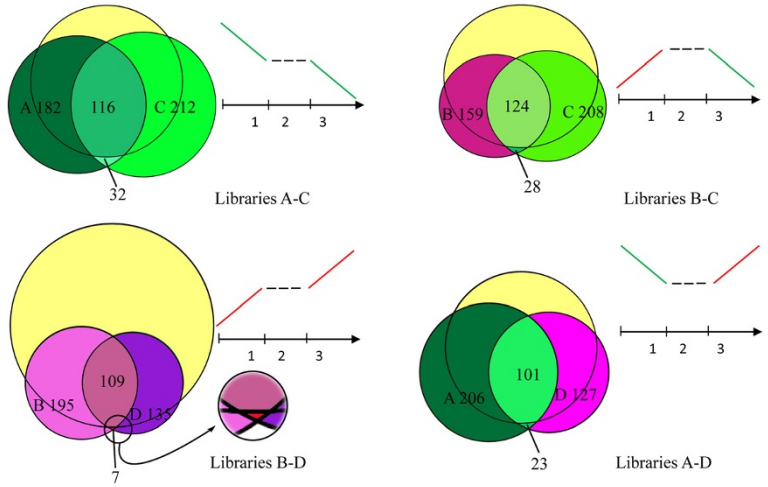

Figure 5 Venn diagrams of continuous or transiently upregulated and downregulated genes by GO ids. Pairwise comparison of stage-specific GO ids of libraries B-D, A-C, B-C and A-D found 7 continuously upregulated and 32 downregulated, 28 transiently upregulated and 23 downregulated.

58 of 103 unigenes were upregulated in stage 3 versus 2, and 18 of 103 unigenes were upregulated from stage 1 to 3 including genes for acyl-ACP thioesterase FatB2, long-chain acyl-CoA synthetase 4,MADS-box (GenBank: KF574389) and others (Additional file 5). Of 103 unigenes, 30 were downregulated from stage 1 to 3 including acetyl-COA carboxylase, fructose-1,6-bisphosphate and phosphoglyceromutase and other genes (Additional file 5).

KEGG pathways for fatty acid biosynthesis and metabolism and starch, sucrose and methane metabolism

A number of enzymes were associated with specific metabolic or biosynthetic pathways for carbohydrates, fatty acids or secondary metabolites. The most represented Kyoto Encyclopedia of Genes and Genomes (KEGG) maps that were associated with fatty acid or carbohydrate biosynthesis or metabolism were organized into a simple network (Additional file 6 and Additional file 7), and several enzymes were upregulated or downregulated during coconut endosperm development. Several components of carbon fixation in photosynthetic organisms (map:00710) showed modulated expression in the early stages of endosperm development (7–9 months after flowering [MAF]) and several components of starch and sucrose metabolism (map:00500) were modulated in the later stages (9–11 MAF). KEGG map analysis also indicated strong upregulation of most enzymes in carbon fixation in photosynthetic organisms (Map:00710), glycolysis and gluconeogenesis (Map:00010), and starch and sucrose metabolism (Additional file 7 and Table 2). Transcripts of enzymes involved in the synthesis of pyruvate from β-D-fructose-6-P such as 6-phosphofructokinase (EC:2.7.1.11), glyceraldehyde-3-phosphate dehydrogenase (EC:1.2.1.12) and phosphoglycerate kinase (EC:2.7.2.3) were upregulated at 7–9 MAF and showed a sustained level at 9–11 MAF (Table 2). Transcripts encoding enzymes related to carbon fixation from carbon dioxide such as aspartate transaminase (EC:2.6.1.1), phosphoglycerate kinase (EC:2.7.2.3) and PEPC (EC:4.1.1.31) were upregulated at 7–9 MAF (Table 2). Consistent with the accumulation of oil (Figure 1A) in coconut endosperm, which increased rapidly during stages 1 to 2 and more slowly during stages 2 to 3, several transcripts encoding enzymes in fatty acid biosynthesis (Map:00061) from glycerophospholipid metabolism (Map:00564) through pyruvate metabolism (Map:00620) and the citrate cycle (Map:00020) were upregulated (Table 3).

The genes for a series of enzymes were upregulated from stage 1 to 2, when lipid accumulation showed an increasing rate. Similar dynamics of specific transcript accumulation were recorded for several components of the fatty acid biosynthesis pathway such as stearoyl-acyl-carrier protein desaturase (EC:1.14.19.2) and acyl-ACP thioesterase FatB3 (EC:3.1.2.14). LPAAT (EC:2.3.1.51) is actively involved in the metabolism of glycerophospholipid and glycerolipid; PEPC (EC:4.1.1.31) is involved in pyruvate metabolism;isocitrate dehydrogenase (EC:1.1.1.41) is related to the citrate cycle; and glutaryl-dehydrogenase (EC:1.3.3.6) is related to fatty acid metabolism (Table 3). Previous studies showed that sugar concentration in coconut water steadily increases from about 1.5% in the early months of maturation to about 5.0–5.5% (at 8–9 MAF) and then slowly decreases to about 2% at full fruit maturity [14]. Consistent with the accumulation of carbohydrates during coconut water development, which increased strongly from stage 1 to stage 2 and gradually reduced from stage 2 to stage 3, several

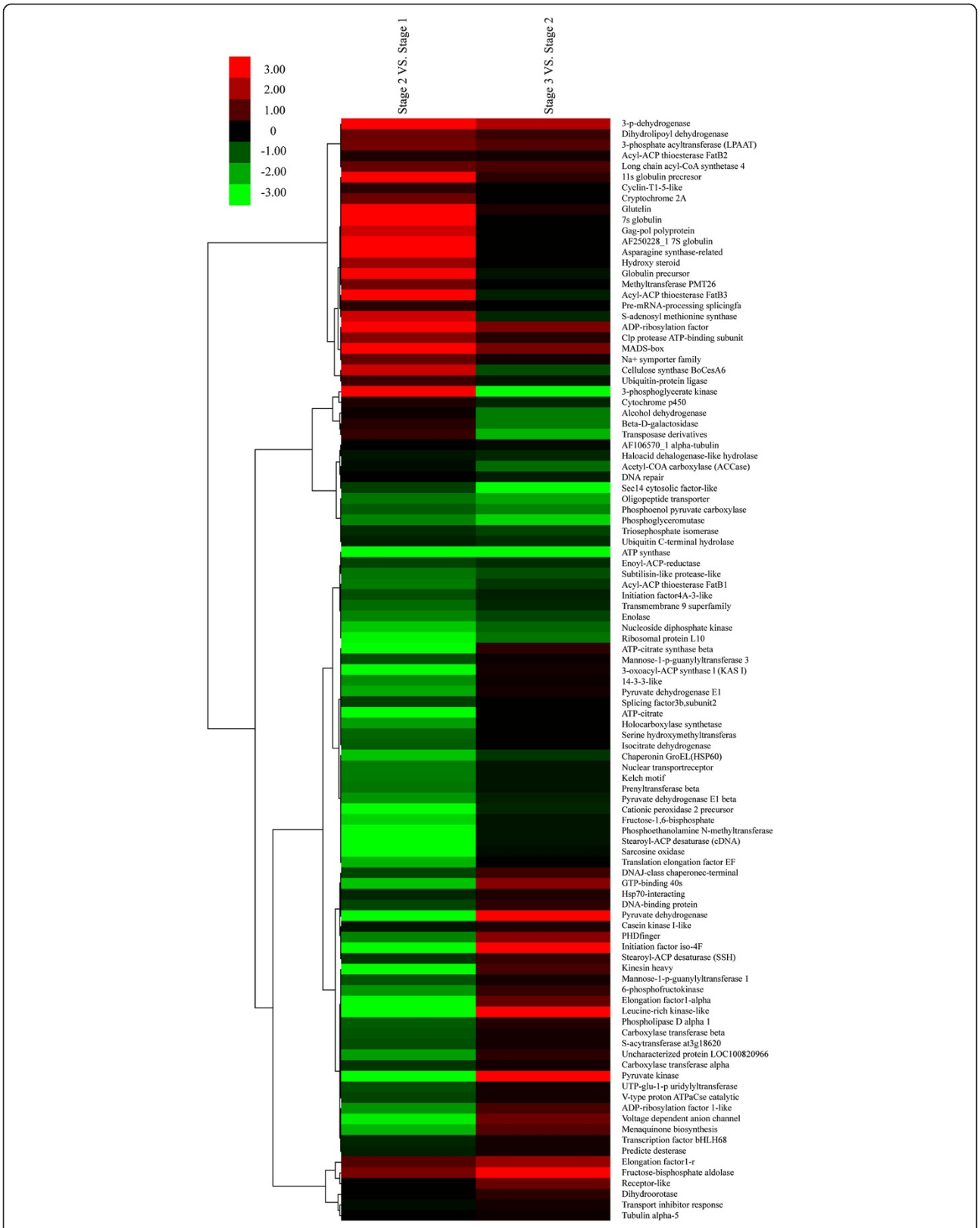

Figure 6 Clustering of gene sequences by real-time PCR values. Pairwise analysis of expression patterns for 103 selected sequences by Cluster software comparing three developmental stages (stage 2 versus stage 1; and stage 3 versus stage 2).

Table 2 Starch and sucrose biosynthesis and metabolism-related pathways from the KEGG pathway database

Enzyme	Enzyme Id	Library	KEGG metabolic pathways[a]								
			Fmm	Ansm	Aam	Ssm	Pgi	G	Ppp	Cfp	Cc
Alcohol dehydrogenase 1	EC:1.1.1.1	C						*			
Sorbitol dehydrogenase	EC:1.1.1.14	C	*								
Alcohol dehydrogenase	EC:1.1.1.22	C		*	*	*	*				
Malate glyoxysomal expressed	EC:1.1.1.37	A								*	*
Isocitrate dehydrogenase	EC:1.1.1.41	B									*
l-ascorbate oxidase homolog	EC:1.10.3.3	C			*						
Glyceraldehyde 3-phosphate dehydrogenase	EC:1.2.1.12	BCD						*			
Pyruvate dehydrogenase e1 component subunit beta	EC:1.2.4.1	AD						*			*
Dihydrolipoyl dehydrogenase-like	EC:1.8.1.4	C						*			*
ATP-citrate synthase beta chain protein 1-like	EC:2.3.3.8	C									*
Cellulose synthase catalytic subunit	EC:2.4.1.12	BC				*					
Sucrose synthase-like partial	EC:2.4.1.13	ABC				*					
Cellulose synthase catalytic subunit	EC:2.4.1.29	C				*					
Callose synthase	EC:2.4.1.34	C				*					
Aspartate transaminase	EC:2.6.1.1	B								*	
6-phosphofructokinase 3-like	EC:2.7.1.11	BCD	*					*	*		
Pyruvate kinase	EC:2.7.1.40	ACD						*		*	
Diphosphate-fructose-6-phosphate 1-phosphotransferase	EC:2.7.1.90	CD	*								
Phosphoglycerate kinase	EC:2.7.2.3	B						*		*	
GDP-mannose pyrophosphorylase	EC:2.7.7.13	BC	*	*							
GDP-mannose pyrophosphorylase	EC:2.7.7.22	B	*	*							
UDP-glucose pyrophosphorylase	EC:2.7.7.9	B		*		*	*				
Chitinase precursor	EC:3.2.1.14	C		*							
Probable polygalacturonase-like	EC:3.2.1.15	C				*	*				
Beta-amylase	EC:3.2.1.2	D				*					
Glucan 1,3-beta-glucosidase	EC:3.2.1.58	B				*					
Phosphoenolpyruvate carboxylase	EC:4.1.1.31	B								*	
Exosome complex exonuclease	EC:4.1.1.39	C								*	
Fructose-bisphosphatealdolase	EC:4.1.2.13	ABCD	*					*	*	*	
Enolase 2-like	EC:4.2.1.11	BCD						*			
Probable rhamnose biosynthetic enzyme 1	EC:4.2.1.76	C		*							
Triosephosphateisomerase	EC:5.3.1.1	BD	*					*		*	
Number of identified enzymes			**7**	**6**	**2**	**9**	**3**	**10**	**2**	**8**	**5**

[a]Fmm (map:00051): fructose and mannose metabolism; Ansm (map:00520): amino sugar and nucleotide sugar metabolism; Aam (map:00053): ascorbate and aldarate metabolism; Ssm (map:00500): starch and sucrose metabolism; Pgi (map:00040): pentose and glucuronate interconversions; G (map:00010): glycolysis/gluconeogenesis; Ppp (map:00030): pentose phosphate pathway; Cfp (map:00710): carbon fixation in photosynthetic organisms; Cc (map:00020): citrate cycle.

transcripts encoding enzymes were upregulated from stage 1 to stage 2 (Table 2), including enzymes in starch and sucrose metabolism (Map:00500), mannose metabolism (Map:00051), ascorbate and aldarate metabolism (Map:00053), and amino sugar and nucleotide sugar metabolism (Map:00520). Similar dynamics of specific transcript accumulation were recorded for a number of enzymes such as 6-phosphofructokinase 3-like (EC:2.7.1.11) and GDP-mannose pyrophosphorylase (EC:2.7.7.13 and EC:2.7.7.22) in fructose and mannose metabolism; enzyme UDP-glucose pyrophosphorylase (EC:2.7.7.9), which is involved in metabolism of amino sugars and nucleotide sugars; enzymes in starch and sucrose catalysis metabolism such as cellulose synthase (EC:2.4.1.12) and glucan 1,3-beta-glucosidase (EC:3.2.1.58); and enzymes involved in carbon fixation of photosynthetic organisms such as aspartate transaminase (EC:2.6.1.1), phosphoglycerate kinase (EC:2.7.2.3) and PEPC (EC:4.1.1.31)(Table 2).

Table 3 Enzymes of EST clones associated with fatty acid biosynthesis, metabolism and related KEGG pathways

Enzyme	Enzyme Id	Library	KEGG metabolic pathways[b]						
			Gpm	Py	Fb	Cc	Fm	Gm	Fe
Alcohol dehydrogenase	EC:1.1.1.1	C					*		
Malate dehydrogenase	EC:1.1.1.37	A		*		*			
Isocitrate dehydrogenase (NAD(+))	EC:1.1.1.41	B				*			
Stearoyl-acyl-carrier protein desaturase (Δ^9)	EC:1.14.19.2	B			*				
Pyruvate dehydrogenase e1 component subunit beta	EC:1.2.4.1	AD		*		*			
Glutaryl- dehydrogenase	EC:1.3.3.6	B					*		
Dihydrolipoyl dehydrogenase-like	EC:1.8.1.4	C		*		*			
Phosphoethanolamine n-methyltransferase	EC:2.1.1.103	A	*						
Lysophosphatidic acid acyltransferase (*LPAAT*)	EC:2.3.1.51	BD	*					*	
ATP-citrate synthase beta chain protein 1-like	EC:2.3.3.8	C				*			
Dihydroxyacetone kinase	EC:2.7.1.29	D						*	
Pyruvate kinase	EC:2.7.1.40	ACD		*					
Phosphatidylcholine 2-acylhydrolase	EC:3.1.1.4	A	*						
acyl-ACP thioesterase FatB3	EC:3.1.2.14	BCD			*				
Phospholipase D	EC:3.1.4.4	D	*						
Phosphoenolpyruvate carboxylase (PEPC)	EC:4.1.1.31	B		*					
3-hydroxyisobutyryl- hydrolase-like protein 1	EC:4.2.1.17	A					*		*
ATP-citrate synthase beta chain protein 1-like	EC:6.2.1.5	C				*			
Acetyl-carboxylase carboxyltransferase beta subunit	EC:6.4.1.2	BC		*	*				
Number of identified enzymes			**4**	**6**	**3**	**6**	**3**	**2**	**1**

[b]Gpm (map:00564): glycerophospholipid metabolism; Py (map:00620): pyruvate metabolism; Fb (map:00061): fatty acid biosynthesis; Cc (map:00020): citrate cycle; Fm (map:00071): fatty acid metabolism; Gm (map:00561): glycerolipid metabolism; Fe (map:00062): fatty acid elongation.

Discussion

This study identified 737 unigenes. However, only 19 BLAST hits for coconut were identified (Figure 2). These results indicated a shortage of molecular information on coconut in the NCBI database. Considering the importance of coconut to tropical regions and an increasing worldwide commercial interest in coconut products, the shortage of information on coconut genomic features is an obstacle to further research. Insights are needed into the genetic and molecular features that control coconut development and ripening, and metabolic pathways in oil and sugar metabolism. We constructed four SSH libraries and identified 737 nonredundant, specific sequences in three coconut pulp development stages. This information enriches the public coconut gene databases and provides fundamental molecular biology information for gene cloning and analysis of oil, sugar and other metabolic pathways.

In two reverse SSH libraries, an AGAMOUS-like (AGL) MADS-box transcription factor gene (GenBank: KF574389) (Figure 7) was continuously upregulated and was more than 22-fold higher in stage 2 than in stage 1; expression of this gene also increased by 2.5-fold during ripening (Figure 8). MADS-box genes are involved in controlling all major aspects of plant development, including development of male and female gametophytes, embryos, seeds, roots, flowers and fruits [31,32]. The coconut MADS-box gene we identified was derived from the coconut endosperm and was continuously upregulated during ripening. This expression pattern implied that the MADS-box gene was related to endosperm metabolism. However, once the flower organ has completed development in the coconut endosperm, only a few free cellular layers of endosperm deposit along the walls of the coconut and the edible endosperm accumulates other components such as carbohydrate, protein and oil. To the best of our knowledge, no research has investigated AGL-like genes in endosperm development in monocotyledons. These genes might improve endosperm weight in monocotyledon seeds without harming seed development. Our data suggested that the MADS-box gene identified in the three development stages of coconut endosperm was involved in endosperm development or metabolic pathways involved in coconut pulp components. Functional analysis of this gene might suggest a new strategy for improving coconut oil yield by increasing the weight of coconut pulp.

In the four SSH libraries, the *CnFatB3* (GenBank: JF338905) gene was the only ACP thioesterase gene identified. Plants have two classes of Fat genes, *FatA* and

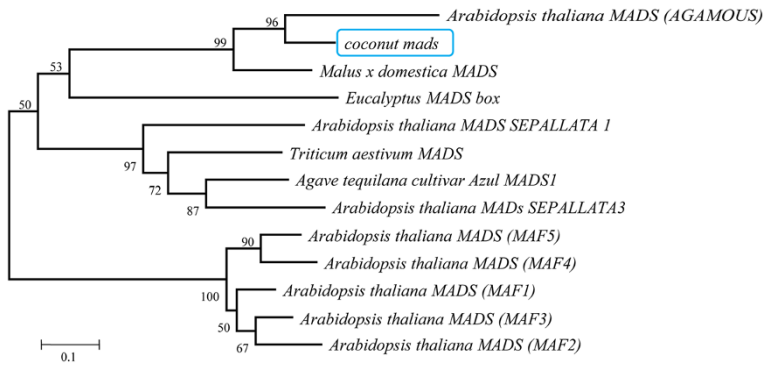

Figure 7 Neighbor-joining tree for coconut MADS-box generated with MEGA 4.0. Neighbor-joining tree showing coconut MADS-box with similarity to *Arabidopsis thaliana* AGAMOUS-like.

FatB, which differ in sequence identity and acyl-ACP substrate specificity [33]. Previously, FatA enzymes were found to predominantly hydrolyze 18:1-ACP with minor activity toward 18:0-ACP and 16:0-ACP. FatB enzymes usually hydrolyze saturated $C_{8:0}$-$C_{18:0}$ ACPs, preferentially 16:0-ACP; however, they also hydrolyze 18:1-ACP [34]. To date, three FatB-type fatty acyl-ACP thioesterases have been found in coconut pulp: *CnFatB 1* (GenBank:JF338903), *CnFatB 2* (GenBank:JF338904) and *CnFatB 3* (GenBank:JF338905). QRT-PCR analysis indicated that expression of *CnFatB 1* was continuously downregulated, and during coconut pulp ripening, *CnFatB 2* expression increased 1.4-fold and and *CnFatB 3* increased 11.5-fold. The hydrolytic specificities of the enzymes from the three *FatB* genes from coconut endosperm (*CnFatB 1*, *CnFatB 2* and *CnFatB 3*) and from oil palm (*EgFatB 1*) have been characterized using *in vitro* expression in *Escherichia coli* K27. CnFatB 1, CnFatB 2 and EgFatB 1 enzymes appear to be predominantly specific for 14:0 and 16:1 and CnFatB 3 enzyme is mainly specific for 12:0 and 14:1 [26]. In our previous research, heterogeneous expression of *CnFatB 1* showed additional specificity, mainly for 18:0-ACP and 16:0-ACP in transgenic tobacco seeds [25]. The results of our current study were partly consistent with earlier observations, suggesting that about 50% of total fatty acids in coconut pulp were saturated laurate ($C_{12:0}$), with the *CnFatB 3* expression level proportionally related to $C_{12:0}$ fatty acid accumulation (Figure 8).

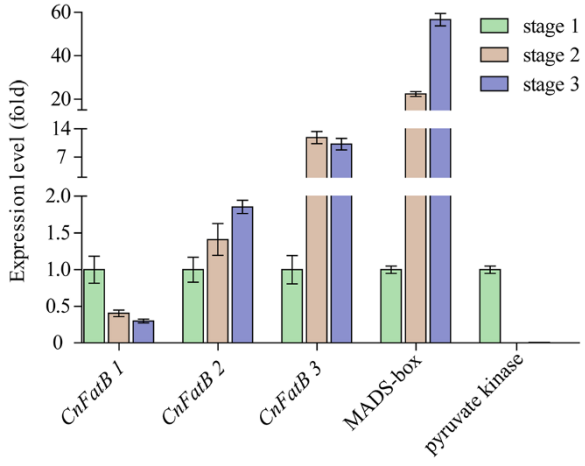

Figure 8 Expression levels of five genes during three coconut pulp development stages. Expression levels during three developmental stages for *CnFatB 1*, *CnFatB 2*, *CnFatB 3*, MADS-box and pyruvate kinase of coconut endosperm.

In the forward SSH library for stage 1 to 2, pyruvate kinase (PK) genes exhibited high expression in the early coconut endosperm (stage 1), but showed a significant decrease in stages 2 and 3 (Figure 8), and were barely detectable in maturing endosperm. Compared with stage 2, expression of PK genes was 10-fold higher in stage 3. The protein phylogeny of PK showed that PK was divided into cytosolic PK (PKc) and plastid PK (PKp) (Figure 9). The PK we identified was more similar to PKc than PKp. PKc is found in all eukaryotes; however, PKp is also found in plants [35]. Physical, immunological and kinetic analyses showed that the two PK isoenzymes (PKp and PKc) of vascular plants and green algae had markedly different characteristics [36,37]. PKc is suggested to control glycolysis to produce pyruvate in the plant cytosol [38] or generate the precursor pyruvate for biosynthetic pathways or mitochondria [39]. Tobacco plants (*Nicotiana tabacum*) lacking PKc in leaves were analyzed for growth, photosynthesis and respiration, and showed a number of conditional growth defects, mainly from altered regulation of sink-source metabolism due to a loss of PKc, specifically in leaves [40]. Decreased expression of PKc in potato tubers shows the importance of PKc in regulation of levels of pyruvate and related oxidase in heterotrophic plant tissue [41]. PKc downregulation by T-DNA insertion in rice results in dwarfism and panicle enclosure, suggesting that PKc affects the glycolytic pathway and changes monosaccharide metabolism and sugar transport in rice [42]. The pulp (solid

endosperm) of coconut is deposited from coconut water (liquid endosperm), and sugar content changes in coconut water might be consistent with changes in the endosperm. In previous studies, the sugar concentration of coconut endosperm from the inner region enclosing the water cavity through middle and outer regions to the testa decreased in a gradient [43,44]. This result implies that sugars accumulates at the first 9 MAF in the developing coconut and at 9–12 MAF, sugars transmigrate from coconut water to the solid endosperm or are catabolized. As expected, the expression of PKc isolated from coconut endosperm contrasted with the sugar concentration of coconut water, and this could be explained by PKc using sugar as its substrate, further implying that PKc was closely involved with sugar metabolism. However, no information is available on the effect of overexpression of PKc in monocotyledon endosperm, especially in an anoxic and liquid environment such as coconut endosperm. Therefore, future studies on coconut endosperm might provide new insights into the specific role of PKc isozymes in plants, complementing previous research. Further studies will be required to fully understand the function of the other plant PK isozymes.

GO classification (Figure 3) of differentially expressed ESTs showed that 44.0% were in biological processes involved in stress response (14.8%), abiotic/biotic stimulus (11.6/4.8%) and catabolism (12.8%). Furthermore, 46.8% of ESTs in cellular components were involved in cytoplasmic membrane-bounded vesicles (22.3%), plasma membrane (13.4%) or plastids (11.1%). Among molecular functions, 73.5% of functionally assigned ESTs were involved in nutrient reservoir activity (40.6%), nucleotide binding (26.5%) or structural molecule activity (6.4%). Combining the three GO major subcategories, specific expression sequences were mainly involved in the stress response of the nutrient reservoir and abiotic/biotic stimulus. Results for stage-specific genes (Figure 4 and Additional file 3) indicated that in stage 1, these genes mainly affected regulation of cell growth and cell division and response to oxidative stress and biotic or abiotic stimulus. In stage 2, the plants mainly synthesized enzymes in fatty acid synthesis, cellulose synthase and carbohydrate metabolism genes that respond to heat and ammonium ion stress and jasmonic acid stimulus. In stage 3, the major processes were in defense response, cell periphery synthesis and glycosyl bond hydrolase activity. Throughout all processes and functions, the stress response was prominent.

The coconut endosperm is exposed to a special nutrient-water environment, so stress is mainly abiotic. Coconut water has high salinity, with especially high concentrations of potassium, malate and chloride. Thus, coconut endosperm growth occurs in high salinity, anoxia and water, with hyperosmotic stress. Growth might also occur in a high-temperature environment. Several genes were found in our SSH libraries that were involved in response to biotic stress (40), abiotic stress (97), or endogenous stimulus (117) and nutrient reservoir activity (222). Several important genes involved in stress have been identified, such as glyceraldehyde-3-phosphate dehydrogenase, which responds to heat and mediates reactive oxygen in plants [45]. Nucleoside diphosphate kinase responds to drought [46], cold [47] and salt stress [48]. Glutathione-S-transferase responds to salt and osmotic stress [49,50] and possesses antioxidant capability [50,51]. Enolase responds to salt and hyperosmotic stress [52]; and β-1,3-glucanase responds to abscisic acid, cold, auxin, ethylene, wounding and light–dark cycles during ripening in banana fruit [53]. Fructose bisphosphatealdolase responds to anoxic environments [54]; temperature-induced lipocalin responds to heat stress [55,56]; and heat shock protein responds to a range of environmental stresses, including oxidative [57], salt and osmotic stresses [49]. The vacuolar H^+-ATPase catalytic subunit responds to salt and osmotic stress [58,59] and 1-cys peroxiredoxin responds to overoxidation [60] and unknown proteins.

Characterization of the action of these genes during coconut endosperm development will improve our understanding of the adaptive stress mechanisms that operate in coconut and enhance our ability to control abiotic damage such as from flooding, drought, high salinity and high temperature, which cause major crop losses worldwide [61]. Stress-response genes in coconut endosperm were mainly involved in abiotic stimulus such as

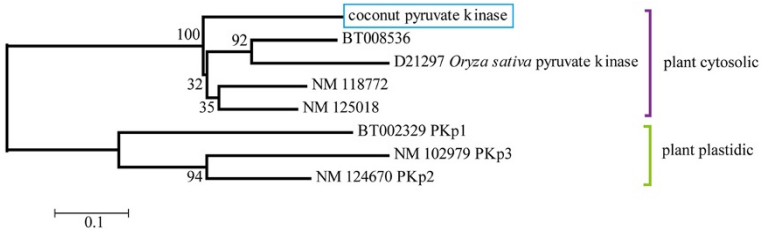

Figure 9 Neighbor-joining tree f generated with MEGA 4.0 or pyruvate kinases [70]. Neighbor-joining tree showing coconut pyruvate kinase belongs to the plant cytosolic pyruvate kinase group.

high salinity, anoxia, heat and hyperosmotic stress. Because of the special growth environment of coconut, we identified genes for accurate information about abiotic stress-response mechanisms as well as unknown stress genes. These genes could be applied to the genetic improvement of other crops by combining traditional and molecular breeding.

Conclusion

In this project, 737 unigenes differentially expressed during three developmental stages of coconut pulp were identified by SSH. All genes were annotated by Blast2GO and validated by RT-PCR. According to Blast2GO analysis, several enzymes were considered related to fatty acids, carbohydrates, transcription factors, stresses responses and secondary metabolites. Through pairwise comparison of those stage-specific GO ids, many genes continuously or transiently upregulated or downregulated were characterized. Comprehensive analysis of the results showed a number of genes had key roles in coconut pulp development.

Comparison of the stage-specific GO ids with the functional annotation of genes differentially expressed during coconut pulp development helped identify the crucial genes and led to determination of the molecular basis of coconut pulp development. This study will have great importance in understanding gene functions and regulatory networks in coconut fruit.

Methods

Plant material

C. nucifera L. pulp was collected from plants grown in fields at the Coconut Research Institute, Chinese Agricultural Academy of Tropical Crops, Hainan, P.R. China. Pulp was collected after fertilization, about 7 MAF (stage 1, when pulp was tender and thin), 9 MAF (stage 2, with the pulp hardening and thickening) and 11 MAF (stage 3, when pulp had completed hardening and little liquid was present). After collection, pulp was wrapped in aluminum foil, immediately frozen in liquid nitrogen and then stored at −70°C.

RNA extraction

Instruments and reagents used for RNA extraction were treated with 0.1% diethylpyrocarbonate. Total RNA was extracted from coconut pulp using hexadecyltrimethyl ammonium bromide-lithium chloride (CTAB-LiCl) methods according to Li and Fan [62]. Total RNA quality was determined by both ultraviolet spectrophotometry and gel electrophoresis on formaldehyde denaturing agarose gels.

cDNA synthesis and SSH library construction

For all samples, double-stranded cDNAs were synthesized from 1 μg of total RNA using SMARTer PCR cDNA synthesis kits (Clontech, Palo Alto, CA). Four subtractive hybridization libraries were constructed using PCR-Select cDNA Subtraction kits (Clontech). Libraries were forward and reverse between stages 1 and 2 (libraries A and B), and between stage 2 and 3 (libraries C and D). Subtracted cDNA was cloned into the PMD19-T simple vector (TaKaRa) and used to transform Trans1-T1 phage-resistant chemically competent cells (Trans). Plasmid DNA was obtained by Axyprep plasmid miniprep kits (AXYGEN).

Sequencing and bioinformatics analysis

From each SSH library, 248–366 cDNA clones were randomly selected and 1230 recombinant bacterial colonies were placed into 96-well plates with wells containing 1 mL of LB broth with 60 μg/mL ampicillin and cultured overnight at 37°C. All cloned products were sequenced using sequencing primer M13F (TGTAAAACGACGGCCAGT) in an ABI 3730XI DNA analyzer. All nucleotide sequences retrieved by SSH had low-quality regions (error rates of traces of sequencing results >0.05) were removed and base-calling of automated sequencer traces used *Phred* [63] (http://www.phrap.org/consed/consed.html#howToGet). Crossmatch (version 0.990329) (same website as *Phred*) software was used to remove short sequences (length of inserted sequences <100 bp) and poly(A) ends and screen for vector contamination and removal of vector sequences, with grouping for ≥96% sequence similarity to vector sequences in the GenBank database in any window of 150 bases. *Phrap* (version 1.080812) (website as for *Phred*) was used to assemble clean ESTs for unigene sequences. Unigene ORF searches used GETORF software (http://embossgui.sourceforge.net/demo/getorf.html). Multiple sequence alignment was carried out using the BLASTX algorithm from the National Center for Biotechnology Information nonredundant (NCBI nr) protein database. GO annotation of unigenes was performed using Blast2GO [64] (version 2.6.4) (http://www.blast2go.de). Sequences were annotated using BLASTX (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM) search in the GenBank NCBI nr database. BLASTX results with an e-value cutoff of $1e^{-6}$ and a minimum similarity of 55% were annotated with GO terms. We also used NCBI nr/nt (ftp://ftp.ncbi.nih.gov/blast/db/FASTA/; 20111209) and Swissport databases (http://www.uniprot.org/downloads; release 2011_12) to align unigenes from SSH; and the KEGG (ftp://ftp.genome.jp/pub/kegg/; release 59.0) database to sort and map data by alignment results and make pathway maps.

Extraction of fatty acids and GC-MS analysis

An extraction solvent system of chloroform-methanol (2:1 by volume) was used with modification of the Folch [65], Bligh [66] and Iverson [67] methods. Methyl esterification was by adding 1 mL of 14% BF_3-methanol solvent with

incubation at 80°C for 15 min. Dried samples were dissolved in 1 mL of n-hexane. GC-MS analysis for fatty acid methyl esters used a Hewlett-Packard Ultra-GC instrument. An HP-FFAP capillary column (30 m × 0.25 mm, 0.25 μm film thickness) was used to separate fatty acid methyl esters. The oven temperature was initially maintained at 150°C for 1 min, then increased at 8°C/min to 250°C, and then to 250°C and maintained for 5 min. The split ratio was 1:30, and the carrier gas was helium at a flow rate of 1.0 mL/min in constant flow mode. The injector was at 250°C and the detector at 230°C. The MS was operated in electron impact mode at 70 eV in a scan range of 10–500 aum. The injection sample volume was 1.0 μL.

Real-time PCR analysis

QRT-PCR was performed to validate the expression pattern of genes isolated by SSH and characterized by Blast2GO. From the entire dataset of nonredundant sequences, 103 gene sequences encoded enzymes related to fatty acid and carbohydrate biosynthesis and metabolism, transcription factors and other stage-specific genes. Genes were selected according to length (>100 bp) and e-value ($>1e^{-6}$). CDNAs were synthesized from 1 μg of total RNA from coconut pulp collected at three developmental stages (7, 9 and 11 MAF) from samples used for SSH library construction using FastQuantRT kits (with DNase) (Tiangen, Beijing, China).

For selected gene sequences, specific primer pairs (Additional file 8) were designed by Primer Premier (http://www.premierbiosoft.com; version 5.00) according to EST sequences and tested for activity at 60°C by conventional PCR. QRT-PCR was performed with three replicates for each gene of each cDNA sample with an ABI 7500 Real-Time System (Applied Biosystems) and resulting data were averaged. Relative expression was quantified using the comparative threshold cycle (Ct) method ($2^{-\Delta\Delta Ct}$ method) [68]. Expression values at the three developmental stages were determined by comparing pulp at stages 1 and 2, and comparing stages 2 and 3. Two series of ratios were analyzed using Cluster 3.0 software (for Windows; http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm#ctv) [69]. All data were adjusted by log transformation and then hierarchical clustering was performed using the average linkage method.

Phylogenetic analysis

The phylogenetic tree was constructed by the use of a neighbor-joining method in MEGA4 [70]. Bootstrap with 1000 replicates was used to establish the confidence limit of the tree branches. Default program parameters were used.

Availability of supporting data

The data sets supporting the results of this article are included within the article and its additional files.

Additional files

Additional file 1: Species distribution of BLAST hits for sequences from libraries A, B, C and D.

Additional file 2: List of GO functional classifications with GO terms and GO ids for all SSH ESTs.

Additional file 3: Mainly stage-specific GOs, GO numbers and GO terms.

Additional file 4: Stage-specific gene GOs for libraries A-C, B-C, B-D and A-D.

Additional file 5: Ratio of expression quantity between two adjacent stages.

Additional file 6: KEGG map. KEGG pathways for fatty acid biosynthesis and related metabolism. Different box colors represent different enzymes.

Additional file 7: KEGG map. KEGG pathways for starch and sucrose and related metabolic pathways. Different box colors represent different enzymes.

Additional file 8: Primers for real-time PCR.

Abbreviations

CnFatB: *Cocos nucifera* acyl-ACP thioesterase type B; CTAB-LiCl: hexadecyltrimethyl ammonium bromide-lithium chloride; GC-MS: Gas chromatography–mass spectrometry; GO: Gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; *LPAAT*: 1-acyl-sn-glycerol-3-phosphate acyltransferase; MAF: Months after flowering; MCFAs: Medium chain-length fatty acids; miRNAs: microRNAs; NAD: Nicotinamide adenine dinucleotide; PEPC: Phosphoenolpyruvate carboxylase; PK: Pyruvate kinase; SSH: Suppression subtractive hybridization.

Competing interests

The authors declare that they have no competing interests.

Authors' contribution

DL contributed substantially to the experimental design, conceived of the study and revised the article. YZ directed discussions and gave substantial suggestions to the paper writing and language organization and helped annotate all sequences. DL and YZ collected coconut pulp material. TL carried out computational annotation of all obtained gene sequences. YY and WM extracted and detected coconut pulp fatty acids and did statistical analyses of data. YY also helped extract total RNAs from coconut pulp. YL carried out almost all experiments, drafted the manuscript, comprehensively analyzed data from all experimental results, drew figures and tables and participated in the bioinformatics analysis process. All authors read and approved the manuscript.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (NSFC) (grant Nos. 31060259, 31160171 and 31360476), and partially supported by the State Key Subject of Botany at Hainan University (grant No. 071001).

Author details

[1]Department of Biotechnology, College of Materials and Chemical Engineering, Hainan University, Haikou, Hainan 570228, PR China. [2]Annoroad Gene Technology Co. Ltd, Beijing 100176, PR China. [3]Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresource, Hainan University, Haikou, Hainan 570228, PR China.

Received: 28 May 2014 Accepted: 22 July 2014
Published: 2 August 2014

References

1. Parfene G, Horincar V, Tyagi AK, Malik A, Bahrim G: **Production of medium chain saturated fatty acids with enhanced antimicrobial activity from crude coconut fat by solid state cultivation of Yarrowia lipolytica.** *Food Chem* 2013, **136**(3–4):1345–1349.
2. Silva PSA, Bamunuarachchi A: **Manufacture of carbonated tender coconut water and development of a process for the utilization of coconut flesh.** *Asian J Food Agro-Industry* 2009, **2**(2):210–213.
3. Van Dam JE, van den Oever MJ, Teunissen W, Keijsers ER, Peralta AG: **Process for production of high density/high performance binderless boards from whole coconut husk: Part 1: Lignin as intrinsic thermosetting binder resin.** *Ind Crops Prod* 2004, **19**(3):207–216.
4. Laureles LR, Rodriguez FM, Reaño CE, Santos GA, Laurena AC, Mendoza EMT: **Variability in fatty acid and triacylglycerol composition of the oil of coconut (*Cocos nucifera* L.) hybrids and their parentals.** *J Agric Food Chem* 2002, **50**(6):1581–1586.
5. Olivares-Carrillo P, Quesada-Medina J: **Thermal decomposition of fatty acid chains during the supercritical methanol transesterification of soybean oil to biodiesel.** *J Supercritical Fluids* 2012, **72**:52–58.
6. Rossell J, King B, Downes M: **Composition of oil.** *J Am Oil Chem Soc* 1985, **62**(2):221–230.
7. Padolina W, Lucas L, Torres L: **Chemical and physical properties of coconut oil.** *Philipp J Coconut Stud* 1987, **12**:4–15.
8. Kintanar Q, Castro J: **Is coconut oil hypercholesterolemic and atherogenic? A focused review of the literature.** *Trans Nat Acad Science and Techn (Phil)* 1988, **10**:371–414.
9. Ide T, Kobayashi H, Ashakumary L, Rouyer IA, Takahashi Y, Aoyama T, Hashimoto T, Mizugaki M: **Comparative effects of perilla and fish oils on the activity and gene expression of fatty acid oxidation enzymes in rat liver.** *Biochim Biophys Acta* 2000, **1485**(1):23–35.
10. Bell JG, Henderson RJ, Tocher DR, McGhee F, Dick JR, Porter A, Smullen RP, Sargent JR: **Substituting fish oil with crude palm oil in the diet of Atlantic salmon (Salmo salar) affects muscle fatty acid composition and hepatic fatty acid metabolism.** *J Nutr* 2002, **132**(2):222–230.
11. Man YBC, Marina AM: **Medium chain triacylglycerol**. In *Nutraceutical and specialty lipids and their Co-products*. Boca Raton, New York, USA: F. Shahidi, Taylor & Francis, CRC; 2006:27–56.
12. Yong JW, Ge L, Ng YF, Tan SN: **The chemical composition and biological properties of coconut (*Cocos nucifera* L.) water.** *Molecules* 2009, **14**(12):5144–5164.
13. Laureles LR, Rodriguez FM, Reaño CE, Santos GA, Laurena AC, Mendoza EMT: **Variability in fatty acid and triacylglycerol composition of the oil of coconut (*Cocos nucifera* L.) hybrids and their parentals.** *J Agric Food Chem* 2002, **50**(6):1581–1586.
14. Jackson JC, Gordon A, Wizzard G, McCook K, Rolle R: **Changes in chemical composition of coconut (*Cocos nucifera*) water during maturation of the fruit.** *J Sci Food Agric* 2004, **84**(9):1049–1052.
15. Teulat B, Aldam C, Trehin R, Lebrun P, Barker JHA, Arnold GM, Karp A, Baudouin L, Rognon F: **An analysis of genetic diversity in coconut (*Cocos nucifera*) populations from across the geographic range using sequence-tagged microsatellites (SSRs) and AFLPs.** *Theor Appl Genet* 2000, **100**(5):764–771.
16. Meerow AW, Wisser RJ, Brown SJ, Kuhn DN, Schnell RJ, Broschat TK: **Analysis of genetic diversity and population structure within Florida coconut (*Cocos nucifera* L.) germplasm using microsatellite DNA, with special emphasis on the Fiji Dwarf cultivar.** *Theor Appl Genet* 2003, **106**(4):715–726.
17. Yong X, Yi L, YaoDong Y, HaiKuo F, Wei X, Mason A, SongLin Z, Sager R, Fei Q: **Development of microsatellite markers in *Cocos nucifera* and their application in evaluating the level of genetic diversity of *Cocos nucifera*.** *Plant Omics* 2013, **6**(3):193–200.
18. D-d L, Fan Y-m: **Construction and Characterization of a cDNA Library from the Pulp of Coconut (*Cocos nucifera* L.).** *Agri Sci China* 2008, **7**(9):1071–1076.
19. Fan H, Xiao Y, Yang Y, Xia W, Mason AS, Xia Z, Qiao F, Zhao S, Tang H: **RNA-Seq Analysis of *Cocos nucifera*: Transcriptome Sequencing and De Novo Assembly for Subsequent Functional Genomics Approaches.** *PLoS One* 2013, **8**(3):e59997.
20. Huang Y-Y, Matzke AJ, Matzke M: **Complete Sequence and Comparative Analysis of the Chloroplast Genome of Coconut Palm (*Cocos nucifera*).** *PLoS One* 2013, **8**(8):e74736.
21. Li D, Zheng Y, Wan L, Zhu X, Wang Z: **Differentially expressed microRNAs during solid endosperm development in coconut (*Cocos nucifera* L.).** *Sci Hortic* 2009, **122**(4):666–669.
22. Li D, Fan Y: **Cloning, characterisation, and expression analysis of an oleosin gene in coconut (*Cocos nucifera* L.) pulp.** *J Horticultural Sci Biotechnol* 2009, **84**(5):483.
23. Xu L, Ye R, Zheng Y, Wang Z, Zhou P, Lin Y, Li D: **Isolation of the endosperm-specific *LPAAT* gene promoter from coconut (*Cocos nucifera* L.) and its functional analysis in transgenic rice plants.** *Plant Cell Rep* 2010, **29**(9):1061–1068.
24. Pérez-Núñez M, Souza R, Sáenz L, Chan J, Zúñiga-Aguilar J, Oropeza C: **Detection of a SERK-like gene in coconut and analysis of its expression during the formation of embryogenic callus and somatic embryos.** *Plant Cell Rep* 2009, **28**(1):11–19.
25. Montero-Cortés M, Rodríguez-Paredes F, Burgeff C, Pérez-Nuñez T, Córdova I, Oropeza C, Verdeil J-L, Sáenz L: **Characterisation of a cyclin-dependent kinase (CDKA) gene expressed during somatic embryogenesis of coconut palm.** *Cell Tissue Organ Cult* 2010, **102**(2):251–258.
26. Yuan Y, Chen Y, Yan S, Liang Y, Zheng Y, Dongdong L: **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* 2013, http://dx.doi.org/10.1071/FP13050.
27. Jing F, Cantu D, Tvaruzkova J, Chipman J, Nikolau B, Yandeau-Nelson M, Reilly P: **Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity.** *BMC Biochem* 2011, **12**(1):44.
28. Diatchenko L, Lau Y, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED: **Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries.** *Proc Natl Acad Sci* 1996, **93**(12):6025–6030.
29. Low E-T, Alias H, Boon S-H, Shariff E, Tan C-Y, Ooi L, Cheah S-C, Raha A-R, Wan K-L, Singh R: **Oil palm (*Elaeis guineensis* Jacq.) tissue culture ESTs: identifying genes associated with callogenesis and embryogenesis.** *BMC Plant Biol* 2008, **8**(1):62.
30. Lopez C, Jorge V, Piégu B, Mba C, Cortes D, Restrepo S, Soto M, Laudié M, Berger C, Cooke R, Delseny M, Tohme J, Verdier V: **A unigene catalogue of 5700 expressed genes in cassava.** *Plant Mol Biol* 2004, **56**(4):541–554.
31. Gramzow L, Theissen G: **A hitchhiker's guide to the MADS world of plants.** *Genome Biol* 2010, **11**(6):214.
32. Becker A, Theissen G: **The major clades of MADS-box genes and their role in the development and evolution of flowering plants.** *Mol Phylogenet Evol* 2003, **29**(3):464–489.
33. Jones A, Davies HM, Voelker TA: **Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases.** *Plant Cell Online* 1995, **7**(3):359–371.
34. Ghosh SK, Bhattacharjee A, Jha JK, Mondal AK, Maiti MK, Basu A, Ghosh D, Ghosh S, Sen SK: **Characterization and cloning of a stearoyl/oleoyl specific fatty acyl-acyl carrier protein thioesterase from the seeds of Madhuca longifolia (latifolia).** *Plant Physiol Biochem* 2007, **45**(12):887–897.
35. Gottlob-McHugh SG, Sangwan RS, Blakeley SD, Vanlerberghe GC, Ko K, Turpin DH, Plaxton WC, Miki BL, Dennis DT: **Normal growth of transgenic tobacco plants in the absence of cytosolic pyruvate kinase.** *Plant Physiol* 1992, **100**(2):820–825.
36. Ambasht P, Kayastha AM: **Plant pyruvate kinase.** *Biol Plant* 2002, **45**(1):1–10.
37. Muñoz E, Ponce E: **Pyruvate kinase: current status of regulatory and functional properties.** *Biochem Mol Biol* 2003, **135**(2):197–218.
38. Hatzfeld WD, Stitt M: **Regulation of glycolysis in heterotrophic cell suspension cultures of Chenopodium rubrum in response to proton fluxes at the plasmalemma.** *Physiol Plant* 1991, **81**(1):103–110.
39. Knowles VL, McHugh SG, Hu Z, Dennis DT, Miki BL, Plaxton WC: **Altered growth of transgenic tobacco lacking leaf cytosolic pyruvate kinase.** *Plant Physiol* 1998, **116**(1):45–51.
40. Grodzinski B, Jiao J, Knowles VL, Plaxton WC: **Photosynthesis and carbon partitioning in transgenic tobacco plants deficient in leaf cytosolic pyruvate kinase.** *Plant Physiol* 1999, **120**(3):887–896.
41. Oliver SN, Lunn JE, Urbanczyk-Wochniak E, Lytovchenko A, van Dongen JT, Faix B, Schmälzlin E, Fernie AR, Geigenberger P: **Decreased expression of cytosolic pyruvate kinase in potato tubers leads to a decline in pyruvate**

resulting in an in vivo repression of the alternative oxidase. *Plant Physiol* 2008, **148**(3):1640–1654.
42. Zhang Y, Xiao W, Luo L, Pang J, Rong W, He C: **Downregulation of OsPK1, a cytosolic pyruvate kinase, by T-DNA insertion causes dwarfism and panicle enclosure in rice.** *Planta* 2012, **235**(1):25–38.
43. Jayalekshmy A, Arumughan C, Narayanan C, Mathew A: **Changes in the chemical composition of coconut water during maturation.** *J Food Sci Tech* 1986, **23**(4):203–207.
44. Balachandran C, Arumughan C, Mathew A: **Distribution of major chemical constituents and fatty acids in different regions of coconut endosperm.** *J Am Oil Chem Soc* 1985, **62**(11):1583–1586.
45. Laino P, Shelton D, Finnie C, De Leonardis AM, Mastrangelo AM, Svensson B, Lafiandra D, Masci S: **Comparative proteome analysis of metabolic proteins from seeds of durum wheat (cv. Svevo) subjected to heat stress.** *Proteomics* 2010, **10**(12):2359–2368.
46. Hajheidari M, Abdollahian-Noghabi M, Askari H, Heidari M, Sadeghian SY, Ober ES, Hosseini Salekdeh G: **Proteome analysis of sugar beet leaves under drought stress.** *Proteomics* 2005, **5**(4):950–960.
47. Imin N, Kerim T, Rolfe BG, Weinman JJ: **Effect of early cold stress on the maturation of rice anthers.** *Proteomics* 2004, **4**(7):1873–1882.
48. Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D, Bohnert HJ: **Gene expression profiles during the initial phase of salt stress in rice.** *Plant Cell Online* 2001, **13**(4):889–905.
49. Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T: **Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray.** *Plant J* 2002, **31**(3):279–292.
50. Kobayashi M, Ishizuka T, Katayama M, Kanehisa M, Bhattacharyya-Pakrasi M, Pakrasi HB, Ikeuchi M: **Response to oxidative stress involves a novel peroxiredoxin gene in the unicellular cyanobacterium Synechocystis sp. PCC 6803.** *Plant Cell Physiol* 2004, **45**(3):290–299.
51. Dixon DP, Davis BG, Edwards R: **Functional Divergence in the Glutathione Transferase Superfamily in Plants identification of two classes with putative functions in redox homeostasis in *arabidopsis thaliana*.** *J Biol Chem* 2002, **277**(34):30859–30869.
52. Ndimba BK, Chivasa S, Simon WJ, Slabas AR: **Identification of Arabidopsis salt and osmotic stress responsive proteins using two-dimensional difference gel electrophoresis and mass spectrometry.** *Proteomics* 2005, **5**(16):4185–4196.
53. Choudhury SR, Roy S, Singh SK, Sengupta DN: **Molecular characterization and differential expression of β-1, 3-glucanase during ripening in banana fruit in response to ethylene, auxin, ABA, wounding, cold and light–dark cycles.** *Plant Cell Rep* 2010, **29**(8):813–828.
54. Dawson NJ, Biggar KK, Storey KB: **Characterization of Fructose-1, 6-Bisphosphate Aldolase during Anoxia in the Tolerant Turtle, Trachemys scripta elegans: An Assessment of Enzyme Activity, Expression and Structure.** *PLoS One* 2013, **8**(7):e68830.
55. Charron J-BF, Ouellet F, Pelletier M, Danyluk J, Chauve C, Sarhan F: **Identification, expression, and evolutionary analyses of plant lipocalins.** *Plant Physiol* 2005, **139**(4):2017–2028.
56. CHI WT, FUNG RW, LIU HC, HSU CC, CHARNG YY: **Temperature-induced lipocalin is required for basal and acquired thermotolerance in Arabidopsis.** *Plant Cell Environ* 2009, **32**(7):917–927.
57. Wang W, Vinocur B, Shoseyov O, Altman A: **Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response.** *Trends Plant Sci* 2004, **9**(5):244–252.
58. Fukuda A, Chiba K, Maeda M, Nakamura A, Maeshima M, Tanaka Y: **Effect of salt and osmotic stresses on the expression of genes for the vacuolar H$^+$-pyrophosphatase, H$^+$-ATPase subunit A, and Na$^+$/H$^+$ antiporter from barley.** *J Exp Bot* 2004, **55**(397):585–594.
59. Silva P, Gerós H: **Regulation by salt of vacuolar H$^+$-ATPase and H$^+$-pyrophosphatase activities and Na$^+$/H$^+$ exchange.** *Plant Signal Behav* 2009, **4**(8):718–726.
60. Kim SY, Paeng SK, Nawkar GM, Maibam P, Lee ES, Kim K-S, Lee DH, Park D-J, Kang SB, Kim MR: **The 1-Cys peroxiredoxin, a regulator of seed dormancy, functions as a molecular chaperone under oxidative stress conditions.** *Plant Sci* 2011, **181**(2):119–124.
61. Ouyang B, Yang T, Li H, Zhang L, Zhang Y, Zhang J, Fei Z, Ye Z: **Identification of early salt stress response genes in tomato root by suppression subtractive hybridization and microarray analysis.** *J Exp Bot* 2007, **58**(3):507–520.
62. Li D, Fan Y: **Extraction and quality analysis of total RNA from coconut pulp tissue.** *Mol Plant Breed* 2007, **5**:883–886.
63. Sayer D, Goodridge D, Christiansen F: **Assign 2.0: software for the analysis of Phred quality values for quality control of HLA sequencing-based typing.** *Tissue Antigens* 2004, **64**(5):556–565.
64. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674–3676.
65. Folch BJ, Lees M, GH S s: **A Simple Method for the Isolation and Purification of Total Lipids from Animal Tissues.** *J Biol Chem* 1956, **226**:497–509.
66. Bligh E, Dyer WJ: **A rapid method of total lipid extraction and purification.** *Can J Biochem Physiol* 1959, **37**(8):911–917.
67. Iverson SJ, Lang SL, Cooper MH: **Comparison of the Bligh and Dyer and Folch methods for total lipid determination in a broad range of marine tissue.** *Lipids* 2001, **36**(11):1283–1287.
68. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ Method.** *Methods* 2001, **25**(4):402–408.
69. De Hoon MJL, Imoto S, Nolan J, Miyano S: **Open source clustering software.** *Bioinformatics* 2004, **20**(9):1453–1454.
70. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**(8):1596–1599.

doi:10.1186/s12870-014-0205-7
Cite this article as: Liang *et al.*: Identification and computational annotation of genes differentially expressed in pulp development of *Cocos nucifera* L. by suppression subtractive hybridization. *BMC Plant Biology* 2014 **14**:205.