biologia plantarum
an international journal for experimental botany

BIOLOGIA PLANTARUM (2021) 65: 273-282 DOI: 10.32615/bp.2021.033

Transcriptome analysis of developing castor bean seeds and identification of ricinoleic acid biosynthesis genes

Z.-T. WU[1,2], F. XU[1], L.-L. YU[1],Y. OUYANG[1], and X.-X. GENG[1,*]

[1] *Applied Biotechnology Center, Wuhan University of Bioengineering, Wuhan 430415, P.R. China*
[2] *School of Food and Biological Engineering Hubei University of Technology, Wuhan 430064, P.R. China*

[*]Corresponding author: E-mail: creamxinxin111@sina.com

Abstract

Ricinoleic acid is a kind of unsaturated fatty acids in oil of castor bean (*Ricinus communis*) seeds with wide application value. However, there is little transcriptomic information on genes related to ricinoleic acid biosynthesis in castor beans. To better understand the regulation mechanism of ricinoleic acid biosynthesis, immature seeds at three developmental stages (S1, S2, and S3 corresponding to 15, 30, and 45 d after pollination) were collected. The results indicated that the accumulation of castor bean oil and ricinoleic acid increased gradually during seed development, and reached the maximum value at the late stages of seed development (45 d after pollination). Furthermore, RNA sequencing was conducted to analyze the transcriptome of the developing seeds at three developmental stages. Totals of 9 875 differentially expressed genes (DEGs) were identified among the three time points. Based on the annotation information, 49 DEGs related to lipid biosynthesis were screened among all DEGs. Through cluster analysis of the 49 DEGs, ten genes with increasing FPKM values from seed development stages S1 to S3 were selected as candidate key enzymes, since they showed similar patterns of increase with castor bean oil accumulation and ricinoleic acid biosynthesis during seed development. The transcriptomic data of the 10 candidate key enzyme genes was further validated by RT-qPCR. Ultimately, a putative model of key genes correlated with ricinoleic acid accumulation was built. Our study identified a series of key genes and revealed the proposed molecular mechanism of ricinoleic acid accumulation in castor bean seeds through the transcriptional analysis. It broadens our knowledge of ricinoleic acid biosynthesis and castor bean oil accumulation and also provides a theoretical foundation for the genetic engineering key genes that can improve the ricinoleic acid production in castor beans as well as in other plants.

Keywords: castor bean oil, cluster analysis, differentially expressed genes, ricinoleic acid, *Ricinus communis*, transcriptome sequencing.

Introduction

Castor bean (*Ricinus communis* L.) which belongs to the family *Euphorbiaceae*, is a kind of oil crop with very high economic value. Its seeds contain more than 50 % of an unusual oil with many industrial uses. The oil is particularly rich in ricinoleic acid (~ 80 %), which is a high-value hydroxy fatty acid that is emerging as a raw material for high-grade lubricating oil production.

However, the oil content of different castor bean

Received 3 February 2021, *last revision* 13 May 2021, *accepted* 28 May 2021.

Abbreviations: ACBP - acyl CoA-binding protein; ACCase - acetyl-coenzyme A carboxylase; ACP - acyl carrier protein; *BLAST* - Basic Local Alignment Search Tool; *COG* - Cluster of Orthologous Groups of proteins; DAP - days after pollination; DAG - diacylglycerol; DGAT - diacylglycerol acyltransferase; DEG - differentially expressed gene; EAR - enoyl-ACP reductase; FAT - fatty acyl-ACP thioesterase; FAH12 - oleoyl-12-hydroxylase; FPKM - fragments per kilobase of exon per million fragments mapped; *GO* - Gene Ontology; GPAT - glycerol-3-phosphate acyltransferase; KAR - ketoacyl-ACP reductase; KAS - ketoacyl-ACP synthase; *KEGG* - Kyoto Encyclopedia of Genes and Genomes; LACS - long-chain acyl-CoA synthetase; LPAAT - lysophosphatidic acid acyltransferase; LPCAT - lysophosphatidylcholine acyltransferase; NR - non-redundant protein; OLE - oleosin; PLA2 - phospholipase A2; PAP - phosphatidic acid phosphatase; PDAT - phosphatidylcholine: diacylglycerol transferase; PDCT - phosphatidylcholine: diacylglycerol choline phosphotransferase; Pfam - protein family; PLC - phospholipase C; PLD - phospholipase D; PXG - peroxygenase; RT-qPCR - reverse transcription quantitative PCR; RRR - triricinolein; SAD - stearoyl-ACP desaturase; SGS - second-generation sequencing; *Swiss-Prot* - the manually annotated and reviewed protein sequence database; TAG - triacylglycerol.

Acknowledgements: This work was supported by the National Natural Science Foundation of China (31801275, 31400242, 31900242), and the Science and Technology Research Project, Science and Technology Department of Hubei Province, China (2018CFB310).

Conflict of interest: The authors declare that they have no conflict of interest.

cultivars varies greatly, and a knowledge-based molecular breeding system has not been established. The yield of castor bean oil has always been relatively low because of inadequate traditional planting methods, which hampered the broader use of castor bean oil. Although the castor oil biosynthetic pathway has been elucidated by Broun and Somerville (1997), Lin *et al.* (2007), and Van Erp *et al.* (2011), there is a lack of molecular research on improving the castor bean oil yield.

The first stage of oleic acid synthesis in plastids is catalyzed by the fatty acid synthetase complex, encompassing acetyl-coenzyme A carboxylase (ACCase) (Sasaki *et al.* 2004), acyl carrier protein (ACP) (Zhou *et al.* 1999), ketoacyl-ACP synthase (KAS) (Puyaubert *et al.* 2005), ketoacyl-ACP reductase (KAR) (Puyaubert *et al.* 2005), enoyl-ACP reductase (EAR) (Saito *et al.* 2008), stearoyl-ACP desaturase (SAD) (Lindqvist *et al.* 1996), and fatty acyl-ACP thioesterase (FAT) (Slabas *et al.* 2002). In the second step, oleic acid is converted to ricinoleoyl-CoA in the cytoplasm by phospholipase A2 (PLA2) (Chen *et al.* 2015, Bates *et al.* 2016), oleoyl-12-hydroxylase (FAH12) (Bafor *et al.* 1991, Zhou *et al.* 2013), long-chain acyl-CoA synthetase (LACS) (Xu *et al.* 2018a), acyl CoA-binding protein (ACBP) (Lung *et al.* 2016), and lysophosphatidylcholine acyltransferase (LPCAT) (Broun *et al.* 1997). Finally, through the Kennedy pathway and acyl editing, the ricinoleic acid is assembled into the triacylglycerol triricinolein (RRR) (Bates *et al.* 2011). The key enzymes of the Kennedy pathway include glycerol-3-phosphate acyltransferase (GPAT), lysophosphatidic acid acyltransferase (LPAAT), phosphatidic acid phosphatase (PAP), and acyl-CoA: diacylglycerol acyltransferase (DGAT) (Weiss *et al.* 1960). Two pathways were found in acyl editing. In the first pathway, ricinoleoyl-PC is directly converted into RRR by phosphatidylcholine: diacylglycerol transferase (PDAT) (Hu *et al.* 2012, Lu *et al.* 2009). The second pathway proceeds from ricinoleoyl-PC to diacylglycerol (DAG) and then triacylglycerol (TAG), catalyzed by phospholipase C (PLC), phospholipase D (PLD), PAP, and phosphatidylcholine: diacylglycerol choline phosphotransferase (PDCT). In oleaginous plants, the majority of TAG is accumulated in oil bodies (Huang *et al.* 2018). The last stage encompasses the assembly of TAGs and oil body proteins to form oil bodies by oleosin (OLE) and peroxygenase (PXG) (Xu *et al.* 2018a).

Illumina RNA sequencing has promptly spread into the most important and commonly used approach for exploring functional genes, screening DEGs, or identifying new genes, in both animals and plants (Du *et al.* 2017). In plants including *Arabidopsis* (Weng *et al.* 2008), rice, and maize, RNA transcriptome profiling is also widely used to provide valuable data on new genes, highly expressed genes, simple sequence repeats, DEGs, *etc*. In *R. communis* and other oil crops, recently, several kinds of research have been reported to reveal the castor bean oil accumulation (Chandrasekaran *et al.* 2014), and ricinoleic acid biosynthesis (Tian *et al.* 2019, Wang *et al.* 2019) through transcriptome approach.

In this study, we aimed to better understand the regulation mechanism of ricinoleic acid biosynthesis and castor bean oil accumulation by transcriptional profiling. According to the whole development process of castor bean seeds, the seeds of a high oil-content cultivar of castor bean were harvested at three developmental stages (15, 30, and 45 d after pollination - DAP) for ricinoleic acid analysis. Additionally, the developing castor bean seeds transcriptome profiles of developing seeds and candidate key enzyme genes were summarized by second-generation sequencing (SGS). Furthermore, a pool of DEGs encoding key enzymes that possibly regulate ricinoleic acid accumulation were identified and the verification of their expression profiles with related genes by real-time reverse transcription quantitative PCR (RT-qPCR) was analyzed. Finally, a putative model of ricinoleic acid biosynthesis and castor bean oil accumulation was built. The results can provide a theoretical foundation for the genetic engineering and key genes that can improve the ricinoleic acid production in castor bean as well as other plants.

Materials and methods

Plants and cultivation: The castor bean (*Ricinus communis* L.) line 93-10 with high oil content (more than 60 %) used in this study was obtained from the National Infrastructure for Crop Germplasm Resources (Wuhan, China). This material has been deposited in a publicly available herbarium of National Infrastructure for Crop Germplasm Resources (OCRI, CAAS, Wuhan, China), with deposition No. 1115. As *Ricinus* spp. are not endangered, collection of samples for scientific purposes was permitted by local legislation. Professor Xinchu Yan and Research associate Lijun Wang (OCRI, CAAS) participated in the identification of specimens. Castor bean seeds were sown and planted in the Yangluo experimental field of the OCRI, CAAS, under standard field conditions, from March to September 2018. The plot was 24 m^2 in area, two-row, with row length 8 m, row spacing 1.5 m, plant spacing 0.8 m, density about 12 000 plants per 10 000 m^2. Protection lines were set around the test area. The test sites should be representative with consistent previous crops, the total area should be not less than 1 200 m^2. Nine immature seed groups [three time points (15, 30, and 45 d after pollination, DAP) × three biological replicates] were collected, and then immediately frozen and stored at -80 °C. S1, S2, and S3 refer to the developmental stage 15 DAP, 30 DAP, and 45 DAP, respectively.

Analysis of the oil and ricinoleic acid content: The immature seeds at three time points were harvested. After removing the seed coat, the seeds were weighed in the fresh state, dried in the oven at 60 °C for 48 h, and weighed again in the dry state. The measurement of the castor bean oil content was according to Tian *et al.* (2019). The resulting lipid mass was used to calculate the oil content as the percentage of oil in the dry mass of seeds (m/m). The ricinoleic acid content was measured by gas chromatography. Heptadecanoic acid (C17:0) (≥ 99 %, *GLPBIO*, Montclair, USA) was chosen as the internal standard and the details of ricinoleic acid content

measurement were referred to the method of Pan *et al.* (2013). The relative percentage of ricinoleic acid was calculated from its peak area.

Total RNA extraction and library construction: According to the manufacturer's instructions, the total RNA of immature seeds (S1 - S3) was isolated in three replicates using the *TRIzol* reagent (*Invitrogen*, Carlsbad, CA, USA). A *NanoDrop 2000* UV-Vis spectrophotometer (*NanoDrop*, Wilmington, DE, USA) and an *Agilent 2100* bioanalyzer (*Agilent*, Palo Alto, CA, USA) were used to assess the quality and quantity of the extracted RNA. For *Illumina* RNA sequencing, 9 libraries (S1-1, S1-2, S1-3, S2-1, S2-2, S2-3, S3-1, S3-2, and S3-3) were constructed from each RNA sample and then sequenced using the *Illumina HiSeq™ 4000* sequencing platform (*Illumina*, San Diego, CA, USA) at *Biomarker Technologies Corporation* (Beijing, China).

Bioinformatics analysis: *FastQC* (*v. 0.11.8*) was used for quality-checking the raw RNA-seq reads to eliminate the low-quality reads and adaptor sequences (Zhu *et al.* 2018). The high-quality clean reads were then mapped to the published *R. communis* genome sequences (Chan *et al.* 2010). *Bowtie* (*v. 2.2.3*) software was used to adjust the reference genome (Langmead *et al.* 2012). The paired-end clean reads were aligned to the reference genome using *TopHat* (*v. 2.0.12*), a software used to detect all transcripts *via* RT-qPCR (Trapnell *et al.* 2009, 2012). *GATK* (*v. 3.8*) was used to identify single base mismatches between sequenced samples and the reference genome, and identify potential single nucleotide polymorphism (SNP) and InDel loci (McKenna *et al.* 2018). *SnpEff* (*v. 3.6c*) was used for annotating variations (SNP, InDel) and predicting the effects of variations (Cingolani *et al.* 2012).

Differential expression analysis and functional annotation: For analyzing the differential expression of genes, software *TopHat* and *Cufflinks* were used to blast the RNA sequencing data against *R. communis* genome sequences. The abundance of gene transcripts was calculated using the fragments per kilobase of exon per million fragments mapped (FPKM) method. DEG screening was conducted using *DESeq* to analyze biological duplicates. During the DEG screening, DEG fold change was ≥2 and false discovery rate (FDR) < 0.01 was considered that DEG was significantly differentially expressed between the control and treatment groups. The FPKM values of DEGs of three time points were first normalized *via* $\log_2$ and then used for clustering with their expression patterns.

All DEG sequences were blasted by the Gene Ontology (*GO*), Cluster of Orthologous Group (*COG*), Protein family (*Pfam*), Kyoto Encyclopedia of Genes and Genomes (*KEGG*), the manually annotated and reviewed protein sequence database (*SwissProt*), and non-redundant (*NR*) databases using *BLAST* (*v. 2.2.26*) to gain the functional annotation information. The annotated information of all DEGs during seed development was analyzed and then used to screen out the genes encoding key enzymes related to the ricinoleic acid biosynthesis pathways.

RT-qPCR analysis: The transcriptions of a total of 10 genes at three time points (S1, S2, and S3) were performed by RT-qPCR analysis to verify the *Illumina* RNA sequencing data. RNA samples for RNA-seq were reverse-transcribed using the *PrimeScript™* RT reagent kit with *gDNA Eraser* (*Takara*, Tokyo, Japan). The *actin* gene of *R. communis* was used as the internal standard (Wang *et al.* 2016). Primers were designed by *Primer 5.0* software (*Premier*, Quebec, Canada). All gene primer sequences including the *actin* gene are listed in Table 1 Suppl. RT-qPCR used *SYBR Green RT-PCR Master Mix* kit (*Kapa*, Shanghai, China) to operate according to the manufacturer's instructions, with a temperature program (95 °C for 5 min, 40 cycles of 95 °C for 15 s, 60 °C for 15 s, and 72 °C for 32 s) in an *ABI PRISM7500* sequence detection system (*Applied Biosystems*, Foster City, CA, USA). All reactions were performed in three technical replicates. Gene relative expressions were calculated using the $2^{-\Delta\Delta Ct}$ comparative threshold cycle (Ct) method (Schmittgen and Livak 2008). Each sample was performed with three biological replicates.

Results

As displayed in Fig. 1, the castor bean oil and ricinoleic acid accumulation in *R. communis* seeds at three developmental stages (15, 30, and 45 DAP) were analyzed. The seed oil content quickly increased from S1 (15 DAP) to S2 (30 DAP) (to 53.8 %), and then gradually increased to 61.7 % at S3 (45 DAP). The ricinoleic acid content showed a similar trend, increasing rapidly to 64.5 % from S1 to S2, and reached a maximum of 74.8 % at S3 (Fig. 1).

A total of 22 691 752 (S1), 23 393 272 (S2), and 22 069 301 (S3) clean reads after eliminating the low-quality sequences were acquired from the corresponding three libraries. Almost 19 732 748 (86.96 %, S1), 19 449 166 (83.14 %, S2), and 18 160 828 (82.29 %, S3) clean reads were successfully matched to the *R. communis* reference genome at S1, S2, and S3, respectively. Among the above-mapped data, 80.77, 71.59, and 76.22 % of the total mapped reads were unique mapping reads at S1, S2, and S3, respectively.

Reads from mature mRNAs should be aligned to exon regions. The comparison of reads to introns is due to the retention of mRNA precursors and introns with variable splicing. Some reads matched to intergenic regions due to incomplete genome annotation. As shown in Fig. 1 Suppl., most of the reads were aligned to exon regions, 87.56 % (S1, Fig. 1*A* Suppl.), 87.09 % (S2, Fig. 1*B* Suppl.) and 87.47 % (S3, Fig. 1*C* Suppl.) in each library.

Overall, 19 442, 19 143, and 18 941 expressed genes were identified at the S1, S2, and S3 stages, respectively (Fig. 2*A*). In all, there were 20 423 genes identified during the entire seed development (Fig. 2*B*). Among these, 17 872 genes were continuously expressed from S1 to S3, while conversely, 1 192 genes (529 in S1, 296 in S2, and 367 in S3) were uniquely expressed at the

Table 1. Functional annotations of DEGs between different stages.

DEGs	Total	*COG*	*eggNOG*	*NR*	*Pfam*	*Swiss-Prot*	*GO*	*KEGG*
S1 *vs*. S2	4240	1798	4092	4194	3579	3233	3277	1543
S1 *vs*. S3	7807	3223	7540	7687	6572	5725	5886	2893
S2 *vs*. S3	4936	2002	4785	4866	4198	3670	3691	1739

unique developmental stage. At stage 1, it owned the highest abundance of developmental stage-specific genes (529), which suggested that a considerable number of genes regulating oil accumulation are concentrated in S1. Furthermore, serious genes (657 in S1 and S2, 318 in S2 and S3, 384 in S1 and S3) were found to be expressed simultaneously at two stages. (Fig. 2*B*).

To better reflect the difference between the samples and reference genome, we identified a total of 130 461 polymorphisms (97 932 SNPs and 32 529 InDels) at the three time points. SNP sites can be divided into two types according to the different ways of base substitution. For all samples in this study, most of the SNPs were the type of transition (more than 60 %). The annotated information showed that the top locations for both SNPs and InDels were intergenic, upstream, and downstream genes.

A total of 9 875 (4 955 up- and 4 920 down-regulated) genes were differentially expressed during *R. communis* seed development. All the identified differentially expressed genes (DEGs) were searched against six databases (*GO*, *COG*, *Pfam*, *KEGG*, *SwissProt*, and *NR*) to obtain their annotated information using *BLAST* software (Altschul *et al.* 1997).

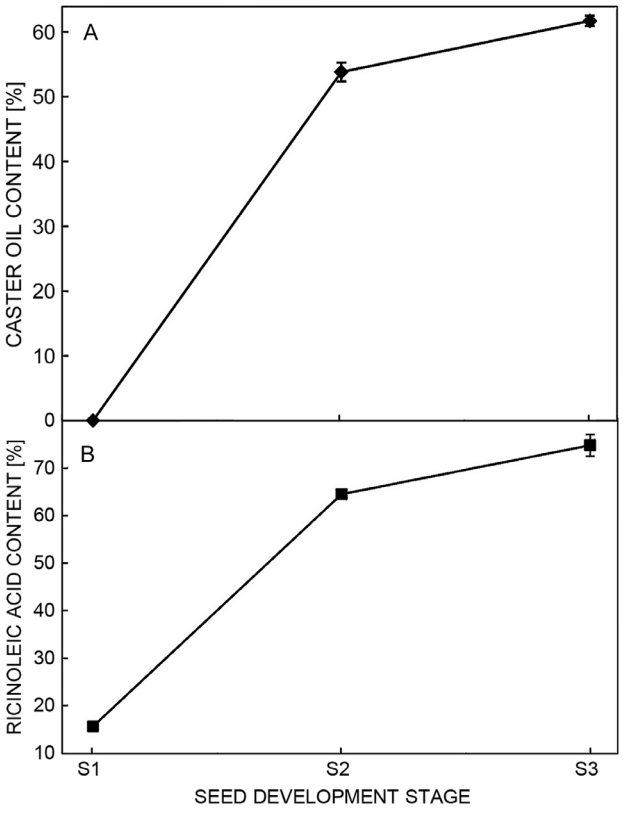

Fig. 1. The castor bean oil content (*A*) and ricinoleic acid ($C_{18:1}$-OH) content (*B*) of developing plants at 15 d after pollination (S1), 30 d after pollination (S2), and 45 d after pollination (S3). Means ± SDs, $n = 3$.

Next, the number of DEG were analyzed between any two libraries (S1 *vs*. S2, S2 *vs*. S3, and S1 *vs*. S3). In the above three paired comparison groups, the number of up-regulated DEGs was higher than the number of down-regulated DEGs (Fig. 3). The smallest number of DEGs (4 240) was detected for the comparison of S1 with S2, encompassing 2 323 up- and 1 917 down-regulated genes. The DEGs number for the comparison of S2 with S3 increased slightly to 4 936, including 2 613 up- and 2 323 down-regulated genes. The largest number of DEGs (7 807) were found for the comparison of S1 and S3, among which 3 930 genes were up- and 3 877 down-regulated (Fig. 3). Heat maps from the hierarchical clustering of DEGs between S1 *vs*. S2, S2 *vs*. S3, and S1 *vs*. S3 paired groups were shown in Fig. 2 Suppl.

To functionally annotate the castor bean transcriptome, 4 240 (S1 *vs*. S2), 7 807 (S1 *vs*. S3), and 4 936 (S2 *vs*. S3) DEGs were applied in the *GO*, *COG*, *Pfam*, *KEGG*, *Swiss-Prot*, and *NR* databases using *BLAST*, respectively (Table 1).

For the *GO* classification analysis of DEGs, all DEGs were assigned to three main categories and then divided into 42 (S1 *vs*. S2) 43 (S1 *vs*. S3), and 43 (S2 *vs*. S3) sub-categories (Fig. 3 Suppl.). Among the different functions, *GO* terms such as phospholipase A2 activity (GO:0004623) and phospholipase C activity (GO:0004629) showed ricinoleic acid biosynthesis function.

Meanwhile, *KEGG* pathway annotation was successful for 1 543 (S1 *vs*. S2), 2 893 (S2 *vs*. S3), and 1 739 (S1 *vs*. S3) DEGs, involved in 119, 125, and 125 different pathways, respectively (Fig. 4 Suppl.). For S1 *vs*. S2 pair group, 6 *KEGG* pathways with a Q value ≤ 0.05 (biosynthesis of amino acids, carbon fixation in photosynthetic organisms, photosynthesis-antenna proteins, carbon metabolism, glycine, serine and threonine metabolism, and cysteine and methionine metabolism) were notably enriched (Table 2). For the paired groups comprising S1 and S3, as well as S2 and S3, 8 and 9 pathways were found to be remarkably enriched, respectively. Compared with earlier stages, pathways related to glyoxylate and dicarboxylate metabolism, fatty acid metabolism, ribosomes, fatty acid biosynthesis, synthesis and degradation of ketone bodies, pyruvate metabolism, fatty acid degradation, valine, leucine and isoleucine degradation, and peroxisomes were found at later stages (Table 2).

The key enzymes in oil synthesis include ACCase, ACP, KAS, KAR, EAR, SAD, FAT, LACS, ACBP, LPCAT, FAH12, PLA2, GPAT, LPAAT, PAP, DGAT,

Table 2. Significantly enriched pathways in different comparison groups.

Group	Pathway_term	Q-value	Gene_number
S1 *vs*. S2	biosynthesis of amino acids	4.20E-07	84
	carbon fixation in photosynthetic organisms	5.70E-07	41
	photosynthesis - antenna proteins	5.65E-06	14
	carbon metabolism	1.31E-05	90
	glycine, serine, and threonine metabolism	1.04E-03	28
	cysteine and methionine metabolism	3.29E-02	32
S1 *vs*. S3	biosynthesis of amino acids	4.36E-08	133
	glyoxylate and dicarboxylate metabolism	3.32E-05	44
	fatty acid metabolism	1.13E-03	49
	ribosome	1.83E-03	150
	carbon fixation in photosynthetic organisms	3.88E-03	51
	carbon metabolism	1.42E-02	132
	photosynthesis - antenna proteins	2.67E-02	14
	fatty acid biosynthesis	3.78E-02	29
S2 *vs*. S3	fatty acid metabolism	3.99E-05	38
	fatty acid biosynthesis	3.05E-03	23
	synthesis and degradation of ketone bodies	6.67E-03	7
	biosynthesis of amino acids	9.01E-03	77
	pyruvate metabolism	1.47E-02	36
	fatty acid degradation	1.56E-02	21
	photosynthesis - antenna proteins	2.94E-02	11
	valine, leucine and isoleucine degradation	4.42E-02	23
	peroxisome	4.99E-02	32

PDCT, PDAT, PLD, PLC2, OLE, and PXG. Based on the annotation information, 49 key enzyme DEGs associated with ricinoleic acid biosynthesis were identified (Table 2 Suppl.). Derived from the normalized FPKM values of DEGs among three developmental stages, hierarchical cluster analysis was executed. All 49 key enzymes associated with ricinoleic acid biosynthesis were classified into three clusters (Fig. 5 Suppl.). Cluster I (a rising expression model) was composed of 10 genes, with the increasing $\log_2$(FPKM+1) values of DEGs from S1 to S3 (Fig. 5*A* Suppl.). Cluster II (a bell-shaped pattern) possessed 10 genes with the $\log_2$(FPKM+1) values of

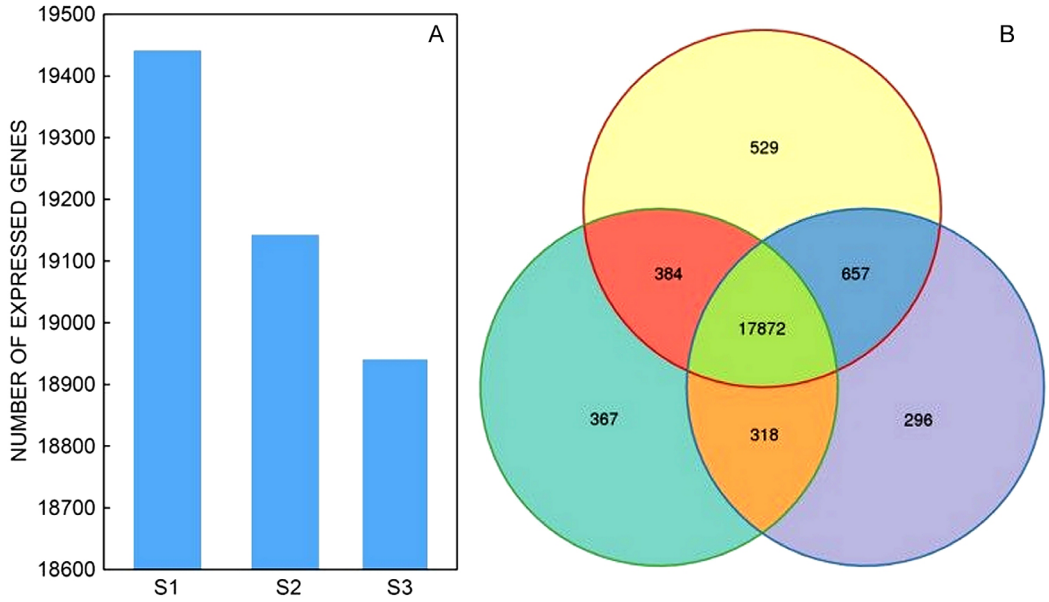

Fig. 2. Transcriptome analysis of *R. communis* seed development at 15 d after pollination (S1), 30 d after pollination (S2), and 45 d after pollination (S3). *A* - Number of expressed genes at S1, S2, and S3, and *B* - their relationships presented in the form of a Venn diagram at S1, S2, and S3.

Table 3. Key lipid-related genes differentially expressed in developing seeds of castor bean.

Gene name	Gene ID	Pathway description	Enzymes	FPKM S1	S2	S3
OLE	29794.t000071	lipid storage	oleosin	300.13	1184.99	4728.43
OLE	30147.t000162	lipid storage	oleosin	209.82	1114.64	2037.97
PXG	30008.t000036	lipid oxidation	peroxygenase	181.32	1240.02	1505.61
LACS7	29844.t000211	fatty acid biosynthesis	long chain acyl-CoA synthetase	7.64	20.13	522.62
PLA2	30142.t000005	glycerolipid metabolism	phospholipase A2	7.74	30.86	224.60
PXG4	29673.t000033	lipid oxidation	peroxygenase	11.30	34.00	54.65
PLC	29756.t000030	glycerolipid metabolism	phospholipase C	0.14	9.27	28.64
ACCase	30174.t000396	fatty acid biosynthesis	acetyl-coenzyme A carboxylase carboxyl transferase	1.47	6.50	18.38
DGAT	29801.t000101	glycerolipid metabolism	diacylglycerol P-acyltransferase	1.18	4.37	9.76
PDAT	29637.t000044	glycerolipid metabolism	phospholipid:diacylglycerol acyltransferase	1.71	5.69	7.82

DEGs raising from S1 to S2 first and then declining values from S2 to S3 (Fig. 5*B* Suppl.). Cluster III was composed of 29 genes with decreasing expression patterns, whereby the $\log_2$(FPKM+1) values of DEGs dropped from S1 to S3 (Fig. 5*C* Suppl.).

According to the analysis results of castor bean oil and ricinoleic acid content during castor bean seed development, we found that the content increased from S1 to S2 and then to S3. On this basis, we screened the key enzyme DEGs with high FPKM values at S3. Ten genes of cluster I with similar expression patterns were selected as the candidate key genes related to seed oil accumulation in *R. communis* (Table 3). The 10 genes included two lipid storage genes (*OLEs 29794.t000071* and *30147.t000162*), two lipid oxidation genes (*PXG 30008.t000036* and *PXG4 29673.t000033*), two fatty acid biosynthesis genes (*LACS7 29844.t000211* and *ACCase-α 30174.t000396*), as well as four glycerolipid metabolism genes (*PLA2 30142.t000005*, *PLC 29756.t000030*, *DGAT 29801.t000101*, and *PDAT 29637.t000044*) (Table 3).

The expression changes of 10 key candidate genes were confirmed by RT-qPCR analysis to validate the *Illumina* RNA sequencing data (Fig. 4). Moreover, a correlation analysis between RT-qPCR and *Illumina* RNA sequencing at the three time points showed a significant correlation ($R = 0.852$, $P < 0.05$). The results suggested that the data by RNA-seq were consistent with RT-qPCR analysis during castor bean seed development.

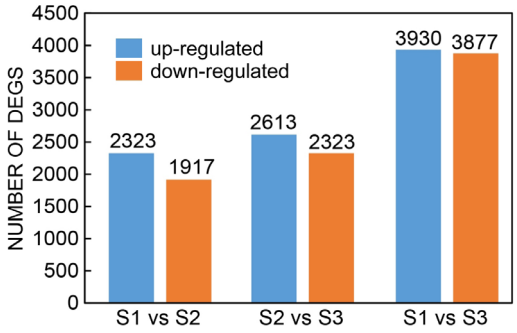

Fig. 3. Analysis of differential genes (including up-regulated and down-regulated DEGs) expression in paired comparison groups (S1 *vs*. S2, S2 *vs*. S3, and S1 *vs*. S3) among three development stages.

Discussion

Ricinoleic acid is a highly valuable unusual hydroxy fatty acid with wide industrial value, but its yield is not too high. Therefore, the identification of the ricinoleic acid accumulation mechanism may provide a theoretical foundation for the genetic engineering and key genes that can improve the ricinoleic acid production in castor bean as well as other plants.

In this study, transcriptome sequencing was performed to comprehensively analyze the transcriptome profile and identify the DEGs in developing seeds from different oil-accumulation stages.

Through analyzing the DEG number between different pair groups, we found that, from S1 to S2, a higher number of up-regulated DEGs than down-regulated DEGs were detected, while from S2 to S3, a higher number of down-regulated DEGs were identified, indicating in the early stage more positive control of gene expression and then in the late stage more negative regulation during the ricinoleic acid accumulation. It was found that the number of DEGs in the early stage (S1 *vs*. S2) was relatively lower than in the late stage (S2 *vs*. S3 and S1 *vs*. S3), and most of these were down-regulated. This result indicated that castor bean oil accumulation is mainly the result of late key gene regulation. The gene regulation in the early stage was not obvious. Six DEGs could be found both from S1 *vs*. S2 group and the S2 *vs*. S3 group, indicating that these six genes played stable roles in ricinoleic acid biosynthesis and regulation during the seed maturation.

In this study, 10 differentially expressed lipid-related

genes were identified as likely playing important roles in oil accumulation during seed maturation of *R. communis*. In the Kennedy pathway, DGAT, which was the last catalytic enzyme for TAG formation, has a dominant role in adding the hydroxyl fatty acyl into TAGs (Xu *et al.* 2018a). One *DGAT*-encoding gene (*29801.t000101*), found in this study during castor bean seed development, may act a pivotal part in transferring hydroxyl fatty acids from PC to TAG in *R. communis* (Table 3). The final enzyme catalyzing TAG formation in the acyl editing pathway, PDAT, showed high expression at S3 during castor bean seed development in this study (Table 3). Cagliari *et al.* (2010) and Troncoso-

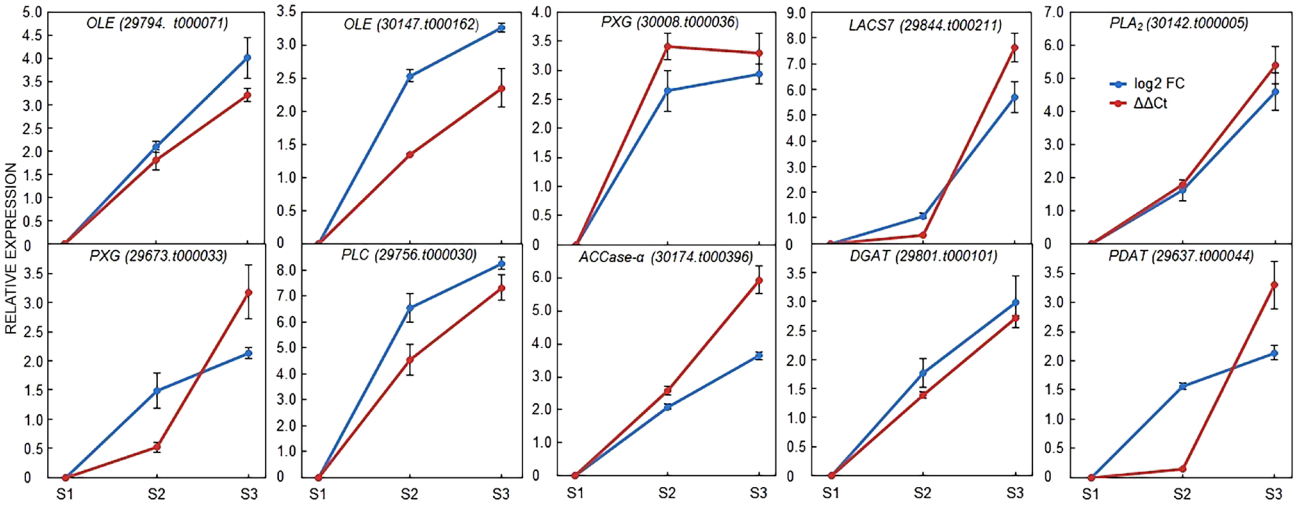

Fig. 4. The expression changes of 10 key differentially expressed lipid-related genes confirmed by RT-qPCR analysis to validate the *Illumina* RNA sequencing data. The comparative log_2FPKM and ΔΔCt values at stage S1 were used as the control for normalization. Means ± SDs, $n = 3$ (biological replicates).

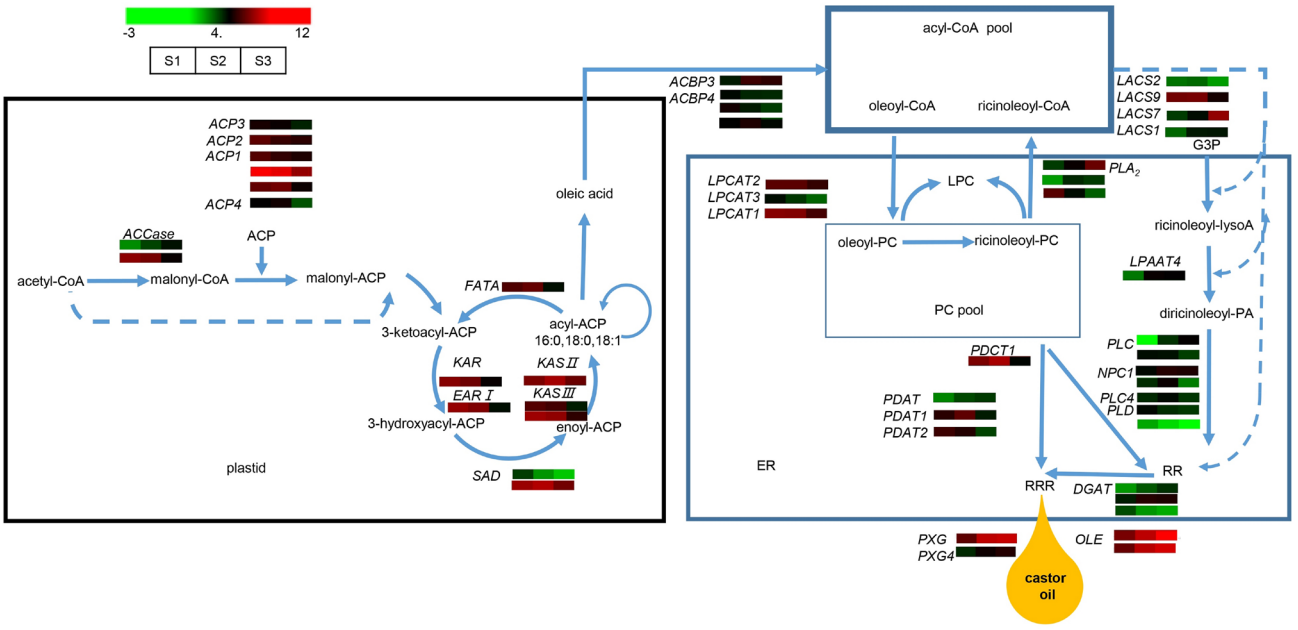

Fig. 5. Theoretical model of gene expression changes in the ricinoleic acid biosynthetic pathway during seed development based on the transcriptomic data obtained in this study. The expressions (represented by Log_2FPKM) of the 49 key enzyme-encoding genes in developing seeds of *R. communis* at different development stages (S1 - S3) are highlighted in color scales (*green* to *red*). *ACBP* - acyl CoA-binding protein, *ACCase* - acetyl-CoA Carboxylase, *ACP* - acyl carrier protein, *DGAT* - diacylglycerol acyltransferase, *EAR* - enoyl-ACP reductase, ER - endoplasmic reticulum, *FATA* - acyl-ACP thioesterase FatA, G3P - sn-glycerol-3-phosphate, *KAR* - ketoacyl-ACP reductase, *KAS* - ketoacyl-ACP synthase, *LACS* - long-chain acyl-CoA synthase, *LPAAT* - acyl-CoA:lysophosphatidic acid acyltransferase, *LPCAT* - lysophosphatidylcholine acyltransferase, *NPC* - non-specific phospholipase C, *OLE* - oleosin, *PDAT* - phospholipid:diacylglycerol acyltransferase, *PDCT* - phosphatidylcholine:diacylglycerol choline phosphotransferase, *PLA_2* - phospholipase A_2, *PLC* - phospholipase C, *PLD* - phospholipase D, *PXG* - peroxygenase, RR - 1,2- diricinoleoyl-sn-glycerol, RRR - triricinolein, *SAD* - stearoyl-ACP desaturase. This model was developed based on the transcriptome data of this study and information from Bafor *et al.* (1991), Cagliari *et al.* (2010), Chandrasekaran *et al.* (2013), Tian *et al.* (2019), and Wang *et al.* (2019).

Ponce *et al.* (2011) also reported that *DGAT2* played a vital role in the enrichment of ricinoleic acid in *R. communis*. Furthermore, in other plants which can produce ricinoleic acid, *DGAT* also possesses a significant effect in ricinoleic acid-enriched TAGs enrichment. In *Physaria fendleri*, *DGAT1*, *DGAT2,* and *PDAT2* were all conducive primarily to hydroxy fatty acid enrichment (Horn *et al.* 2016), whereas, in *Hiptake benghalensis, DGAT2* and *PDAT2* were found to own high expressions in the process of ricinoleic acid biosynthesis (Tian *et al.* 2019).

LACS might participate in transferring modified fatty acids from PC to the acyl-CoA pool together with *PLA2*, which is later for producing TAGs by some enzymes such as *DGAT* in the Kennedy pathway (Xu *et al.* 2018b). In this study, a *LACS7* gene (*29844.t000211*), which had high expression during seed development, may contribute to ricinoleic acid accumulation (Table 3). Nevertheless, in *H. benghalensis, LACS8* has been reported to have the same effect (Tian *et al.* 2019). At the same time, *LACS9* as the dominant LACS isoform was revealed to act a vital role in oil synthesis in *R. communis* and *P. fendleri* (Horn *et al.* 2016). During castor bean seed development, *PLA2* can help to accumulate the hydroxy fatty acid ricinoleate in TAGs (Chen *et al.* 2015). In this study, a *PLA2* gene (*30142.t000005*) showed a high expression at the S3 stage during seed development, which may have an important contribution to the accumulation of ricinoleic acid.

OLEs are the main oil-body proteins, with steroleosins and caleosins constituting the other two major classes of proteins associated with oil bodies (Badami *et al.* 1970). In this study, two *OLE* genes (*29794.t000071* and *30147.t000162*) showed the highest expression among all selected candidate key DEGs. Tian *et al.* (2019) reported that high expression of *OLE3* during seed development in *H. benghalensis* participated in the accumulation of ricinoleic acid in oil-body formation. *OLE* genes had been reported to contribute to monoacylglycerol acyltransferase and *PLA2* bifunctional enzyme activity in peanuts (Parthibane *et al.* 2012). The *PXG* gene encodes a caleosin, another kind of oil body (Hanano *et al.* 2006). In this study, two *PXG* genes (*30008.t000036* and *29673.t000033*) showed a high expression at S3 and may possess a part in TAG storage. In previous studies, a similar function of *PXG* was observed in *H. benghalensis*, but not in *R. communis* (Troncoso-Ponce *et al.* 2011) or *P. fendleri* (Horn *et al.* 2016).

Lin *et al.* (2007) found that *PLC2* catalyzes the transformation of 2-oleoyl-PC into 1-acyl-2-oleoyl-sn-glycerol. *PLC* and/or *PLD*, together with *PAPase*, may also relate to the transformation of PtdCho to DAG (Sasaki *et al.* 2004). In this study, a *PLC* gene (*29756.t000030*) showed high expressions at the S3 stage during seed development and may play a role in converting ricinoleoyl-PC into 1,2-diricinoleoyl-sn-glycerol.

In this study, an *ACCase-α* gene (*30174.t000396*) also showed a high expression at the S3 stage during seed development. The synthesis of oil in castor bean seeds starts from the production of malonyl-CoA by *ACCase*, which is a key regulatory step of fatty acid synthesis and oil formation in seeds, and the encoding gene affects the entire life process of plants. *ACCase* is present in the cytosol and plays the role of transferring a carboxyl group in the reaction process (Sasaki *et al.* 2004). The *ACCase*-catalyzed reaction is a rate-limiting step in the fatty acid synthesis. Therefore, overexpression of the *ACCase* gene in plants, especially in oil crops such as soybean, *Brassica napus*, and sesame, can improve their ability to synthesize fatty acids, which has been confirmed by many studies (Roesler *et al.* 1997).

To illustrate the mechanism of ricinoleic acid biosynthesis, based on our transcriptome data and previous studies, a theoretical model of 49 key genes expression changes the ricinoleic acid biosynthetic pathway during seed development was constructed based on our transcriptomic data (Fig. 5).

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.: Gapped BLAST and PSIBLAST: a new generation of protein database search programs. - Nucl. Acids Res. **25**: 3389-3402, 1997.

Badami, R.C., Kudari, S.M.: Analysis of *Hiptage madablota* seed oil. - J. Sci. Food Agr. **21**: 248-249, 1970.

Bafor, M., Smith, M.A., Jonsson, L., Stobart, K., Stymne, S.: Ricinoleic acid biosynthesis and triacylglycerol assembly in microsomal preparations from developing castor bean (*Ricinus communis*) endosperm. - Biochem. J. **280**: 507-514, 1991.

Bates, P.D.: Understanding the control of acyl flux through the lipid metabolic network of plant oil biosynthesis. - Biochim. Biophys. Acta **1861**: 1214-1225, 2016.

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

Broun, R., Somerville, C.: Accumulation of ricinoleic, lesquerolic, and densipolic acids in seeds of transgenic in *Arabidopsis* plants that express a fatty acyl hydroxylase cDNA from castor bean. - Plant Physiol. **113**: 933-942, 1997.

Cagliari, A., Margis-Pinheiro, M., Loss, G., Mastroberti, A.A., Mariath, J.E.A., Margis, R.: Identification and expression analysis of castor bean (*Ricinus communis*) genes encoding enzymes from the triacylglycerol biosynthesis pathway. - Plant Sci. **179**: 499-500. 2010.

Chan, A.P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., Melake-Berhan, A., Jones, K.M., Redman, J., Chen, G., Cahoon, E.B., Gedil, M., Stanke, M., Haas, B.J., Wortman, J.R., Fraser-Liggett, C.M., Ravel, J., Rabinowicz, P.D.: Draft genome sequence of the oilseed species *Ricinus communis*. - Nat. Biotechnol. **28**: 951-956, 2010.

Chandrasekaran, U., Liu, A.Z.: Seed filling and fatty acid changes in developing seeds of castor bean (*Ricinus communis* L.). - Aust. J. Crop Sci. **11**: 1761-1765, 2013.

Chandrasekaran, U., Xu, W., Liu, A.Z.: Transcriptome profiling identifies ABA mediated regulatory changes towards storage filling in developing seeds of castor bean (*Ricinus communis* L.). - Cell Biosci. **4**: 33, 2014.

Chen, G., Woodfield, H., Pan, X., Harwood, J.L., Weselake, R.J.: Acyl-trafficking during plant oil accumulation. - Lipids **50**: 1057-1068, 2015.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M.: A program for

annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. - Fly **6**: 80-92, 2012.

Du, H., Yu, Y., Ma, Y., Gao, Q., Cao, Y., Chen, Z., Ma, B., Qi, M., Li, Y., Zhao, X., Wang, J., Liu, K., Qin, P., Yang, X., Zhu, L., Li, S., Liang, C.: Sequencing and *de novo* assembly of a near complete indica rice genome. - Nat. Commun. **8**: 15324, 2017.

Hanano, A., Burcklen, M., Flenet, M., Ivancich, A., Louwagie, M., Garin, J., Blée, E.: Plant seed peroxygenase is an original heme-oxygenase with an EF-hand calcium binding motif. - J. biol. Chem. **281**: 33140-33151, 2006.

Horn, P.J., Liu, J., Cocuron, J.C., McGlew, K., Thrower, N.A., Larson, M., Lu, C., Alonso, A., Ohlrogge, J.: Identification of multiple lipid genes with modifications in expression and sequence associated with the evolution of hydroxy fatty acid accumulation in *Physaria fendleri*. - Plant J. **86**: 322-348, 2016.

Hu, Z., Ren, Z., Lu, C.: The phosphatidylcholine diacylglycerolcholinephosphotransferase is required for efficient hydroxy fatty acid accumulation in transgenic *Arabidopsis*. - Plant Physiol. **158**: 1944-1954, 2012.

Huang, A.H.C.: Plant lipid droplets and their associated proteins: potential for rapid advances. - Plant Physiol. **176**: 1894-1918, 2018.

Langmead, B., Salzberg, S.L.: Fast gapped-read alignment with bowtie 2. - Nat. Methods **9**: 357, 2012.

Lin, J., Arcinas, A.: Regiospecific analysis of diricinoleoylacylglycerols in castor (*Ricinus communis* L.) oil by electrospray ionization-mass spectrometry. - J. Agr. Food Chem. **55**: 2209-2216, 2007.

Lindqvist, Y., Huang, W., Schneider G., Shanklin J.: Crystal structure of delta9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. - EMBO J. **15**: 4081-4092, 1996.

Lu, C., Xin, Z., Ren, Z., Miquel, M., Browse, J.: An enzyme regulating triacylglycerol composition is encoded by the *ROD1* gene of *Arabidopsis*. - Proc. nat. Acad. Sci. USA **106**: 18837-18842, 2009.

Lung, S.C., Chye, M.L.: The binding versatility of plant acyl-CoA-binding proteins and their significance in lipid metabolism. - Biochim. biophys. Acta **1861**: 1409-1421, 2016.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., Depristo, M.A.: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. - Genome Res. **20**: 1297-1303, 2018.

Pan, X., Siloto, R.M., Wickramarathna, A.D., Mietkiewska, E., Weselake, R.J.: Identification of a pair of phospholipid:diacylglycerol acyltransferases from developing flax (*Linum usitatissimum* L.) seed catalyzing the selective production of trilinolenin. - J. biol. Chem. **288**: 24173-24188, 2013.

Parthibane, V., Rajakumari, S., Venkateshwari, V., Iyappan, R., Rajasekharan, R.: Oleosin is bifunctional enzyme that has both monoacylglycerol acyltransferase and phospholipase activities. - J. biol. Chem. **287**: 1946-1954, 2012.

Puyaubert, J., Dieryck, W., Costaglioli, P., Chevalier, S., Breton, A., Lessire, R.: Temporal gene expression of 3-ketoacyl-CoA reduetase is different in high and in low erucic acid. *Brassica napus* cultivars during seed development. - Biochim. biophys. Acta **1687**: 152-163, 2005.

Roesler, K., Shintani, D., Savage, L., Boddupalli, S., Ohlrogge, J.: Targeting of the *Arabidopsis* homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. - Plant Physiol. **113**: 75-81, 1997.

Saito, J., Yamada, M., Watanabe, T., Iida, M., Kitagawa, H., Takahata, S., Lida, M., Kitagawa, H., Takahata, S., Ozawa, T., Takeuchi, Y., Ohsawa, F.: Crystal structure of enoyl-acyl carrier protein reductase (FabK) from Streptococcus pneumoniae reveals the binding mode of an inhibitor. - Protein Sci. **17**: 691-699, 2008.

Sasaki, Y., Nagano, Y.: Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. - Biosci. Biotech. Biochem. **68**: 1175-1184, 2004.

Schmittgen, T.D., Livak, K.J.: Analyzing real-time PCR data by the comparative CT method. - Nat. Protoc. **3**: 1101-1108, 2008.

Slabas, A.R., White, A., O'hara, P., Fawcett, T.: Investigations into the regulation of lipid biosynthesis in *Brassica napus* using antisense down-regulation. - Biochem. Soc. Trans. **30**: 1056-1059, 2002.

Tian, B., Lu, T., Xu, Y., Wang, R., Chen, G.: Identification of genes associated with ricinoleic acid accumulation in *Hiptage benghalensis* via transcriptome analysis. - Biotechnol. Biofuels **12**: 16, 2019.

Trapnell, C., Pachter, L., Salzberg, S.L.: TopHat: discovering splice junctions with RNA-Seq. - Bioinformatics **25**: 1105-1111, 2009.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L.: Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. - Nat. Protoc. **7**: 562-578, 2012.

Troncoso-Ponce, M.A., Kilaru, A., Cao, X., Durrett, T.P., Fan, J., Jensen, J.K, Thrower, N.A., Pauly, M., Wilkerson, C., Ohlrogge, J.B.: Comparative deep transcriptional profiling of four developing oilseeds. - Plant J. **68**: 1014-1027, 2011.

Van Erp, H., Bates, P.D., Burgal, J., Shockey, J., Browse, J.: Castor phospholipid:diacylglycerol acyltransferase facilitates efficient metabolism of hydroxyl fatty acids in transgenic *Arabidopsis*. - Plant Physiol. **15**: 683-693, 2011.

Wang, L., Jiang, X., Wang, L., Wang, W., Fu, C., Yan, X., Geng, X.: A survey of transcriptome complexity using PacBio single-molecule real-time analysis combined with Illumina RNA sequencing for a better understanding of ricinoleic acid biosynthesis in *Ricinus communis*. - BMC Genomics **20**: 456, 2019.

Wang, X., Wang, L., Yan, X., Wang, L., Tan, M., Geng, X., Wei, W.: Transcriptome analysis of the germinated seeds identifies low-temperature responsive genes involved in germination process in *Ricinus communis*. - Acta Physiol. Plant. **38**: 6, 2016.

Weiss, S.B., Kennedy, E.P., Kiyasu, J.Y.: The enzymatic synthesis of triglycerides. - J. biol. Chem. **235**: 40-44, 1960.

Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., Lei, C., Zhang, X., Cheng, Z., Guo, X., Wang, J., Jiang, L., Zhai, H., Wan, J.: Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. - Cell Res. **18**: 1199-1209, 2008.

Xu, Y., Caldo, K.M.P., Pal-Nath, D., Ozga, J., Lemieux, M.J., Weselake, R.J., Chen, G.: Properties and biotechnological applications of acyl-CoA: diacylglycerol acyltransferase and phospholipid: diacylglycerol acyltransferase from terrestrial plants and microalgae. - Lipids **53**: 663-688, 2018b.

Xu, Y., Holic, R., Li, D., Pan, X., Weselake, R.J.: Substrate preferences of long-chain acyl-CoA synthetase and diacylglycerol acyltransferase contribute to enrichment of flax seed oil with α-linolenic acid. - Biochem. J. **475**: 1473-1489, 2018a.

Zhou, P., Florova, Q., Reynolds, K.A.: Polyketide synthase acyl carrier protein (ACP) as a substrate and a catalyst for malonyl

ACP biosynthesis. - Chem. Biol. **6**: 577-584, 1999.
Zhou, X.R., Singh, S.P., Green, A.G.: Characterization of the *FAD2* gene family from *Hiptage benghalensis*: a ricinoleic acid accumulating plant. - Phytochemistry **92**: 42-48, 2013.
Zhu, C., Li, X., Zheng, J.: Transcriptome profiling using Illumina- and SMRT based RNA-seq of hot pepper for in-depth understanding of genes involved in CMV infection. - Gene **666**: 123-133, 2018.