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

Ziting Wu

Wuhan Institute of Bioengineering

Qin Yang

Wuhan Institute of Bioengineering

Fei Xu

Wuhan Institute of Bioengineering

Yi Ouyang

Wuhan Institute of Bioengineering

Xinxin Geng (✉ creamxinxin111@sina.com)

Wuhan Institute of Bioengineering

Research article

Keywords: Ricinus communis, transcriptome sequencing, castor oil, ricinoleic acid, key enzyme-encoding genes

Posted Date: August 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-38573/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background Ricinoleic acid is a kind of unsaturated fatty acid in castor oil with wide application value. It is sourced from *R. communis* seed oil, where it is present in large amounts. However, there is little transcriptomic information on genes related to ricinoleic acid biosynthesis in castor bean.

Results In order to better understand the regulation mechanism of ricinoleic acid biosynthesis, immature seeds at three developmental stages (15, 30, and 45 days after pollination) were collected. The results indicated that the accumulation of castor oil and ricinoleic acid increased gradually during seed development, and reached the maximum value at the late stages of seed development (45 days after pollination). Furthermore, RNA sequencing was conducted to analyze the transcriptome of the developing seeds at three developing stages. Totals of 9,875 differentially expressed genes 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 oil accumulation and ricinoleic acid biosynthesis during seed development. The transcriptomic data of the 10 candidate key enzyme genes was further validated by qRT-PCR. Ultimately, a putative model of key genes correlated with ricinoleic acid accumulation was built.

Conclusion Our study identified a series of key genes and revealed the proposed molecular mechanism of ricinoleic acid accumulation in castor seeds through the transcriptional analysis. It broadens our knowledge of ricinoleic acid biosynthesis and castor oil accumulation and also provides a theoretical foundation for the genetic engineering key genes that can improve the ricinoleic acid production in castor bean as well as other plants.

Background

Castor bean (*Ricinus communis* L.) which belongs to the spurge family Euphorbiaceae, is a kind of oil crops 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 emerging as a raw material for high-grade lubricating oil production.

However, the oil content of different castor bean varieties varies greatly, and a knowledge-based molecular breeding system has not been established. The yield of castor oil has always been relatively low because of inadequate traditional planting methods, which hampered the broader use of castor oil. Although the castor oil biosynthetic pathway has been elucidated by Lin et al [1], as well as Broun et al [2], among others [3, 4], there is a lack of molecular research on improving the castor 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) [5], acyl carrier protein (ACP) [6], ketoacyl-ACP synthase (KAS) [7], ketoacyl-ACP reductase (KAR) [7], enoyl-ACP reductase (EAR) [8], stearoyl-ACP desaturase (SAD) [9], and fatty acyl-ACP thioesterase (FAT) [10]. In the second step, oleic acid is converted to ricinoleoyl-CoA in the cytoplasm by phospholipase A2 (PLA2) [11–12], oleoyl-12-hydroxylase (FAH12) [13–14], long chain acyl-CoA synthetase (LACS) [15], acyl CoA-binding protein (ACBP) [16] and lysophosphatidylcholine

acyltransferase (LPCAT) [17]. Finally, through the Kennedy pathway and acyl editing, the ricinoleic acid is assembled into the triacylglycerol triricinolein (RRR) [18]. The key enzymes of Kennedy pathway include glycerol-3-phosphate acyltransferase (GPAT), lysophosphatidic acid acyltransferase (LPAAT), phosphatidic acid phosphatase (PAP), and acyl-CoA: diacylglycerol acyltransferase (DGAT) [19]. Two pathways were found in acyl editing proceeds via. In one, ricinoleoyl-PC is directly converted into RRR by phosphatidylcholine: diacylglycerol transferase (PDAT) [20, 21]. 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 [22]. The last stage encompasses the assembly of TAGs and oil body proteins to form oil bodies by oleosin (OLE) and peroxygenase (PXG) [23].

Illumina RNA sequencing has promptly spread into the most important and commonly used approach for exploring functional genes, screening differentially expressed genes, or identifying new genes, in both animals and plants [24–25]. In plants, RNA transcriptome profiling is also widely used to provide valuable data on new genes, highly expressed genes, SSR, differentially expressed genes and etc.. Many such studies were done in *Arabidopsis* [26], rice [27], maize [28] and *R. communis*. Chandrasekaran et al. (2014) [29] identified ABA mediated regulatory changes in the process of oil accumulation in developing castor seeds through transcriptome profiling. Han et al. (2020) [30] found autophagy-related genes in castor bean by RNA-seq. Tian et al. (2019) revealed a series of genes related to biosynthesis via transcriptome analysis in *Hiptage benghalensis* [31].

In this study, we aimed to better understand the regulation mechanism of ricinoleic acid biosynthesis and castor oil accumulation by transcriptional profiling. A high oil-content variety of castor bean was chosen and seeds were harvested at three developmental stages [15, 30, and 45 days after pollination (DAP)] for ricinoleic acid analysis. Additionally, the developing castor seeds transcriptome profiles of developing seeds and candidate key enzyme genes were summarized by second-generation sequencing (SGS). Furthermore, a pool of differentially expressed genes encoding key enzymes that possibly regulate ricinoleic acid accumulation were identified and the verification of their expression profiles with related genes by quantitative real-time reverse transcription PCR (qRT-PCR) was analyzed. Finally, a putative model of ricinoleic acid biosynthesis and castor oil accumulation was built. The results 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.

Results

Castor oil and ricinoleic acid accumulation during seed development in *R. communis* seeds

The oil content of mature seeds from 213 *R. communis* varieties was analyzed, and the strain 93 – 10 with the highest seed oil content (about 69.2%) was selected. As displayed in Fig. 1, the castor 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 stage 1 (15 DAP, S1) to stage 2 (30 DAP, S2) (53.8%), and then gradually increased to 61.7% at stage 3 (45 DAP, S3). The ricinoleic acid content showed a similar trend, increasing rapidly to 64.5% from S1 to S2, and reached the maximum of 74.8% at S3 (Fig. 1).

Transcriptome sequencing

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 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 (Additional file 1: Table S1).

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. 2, most of the reads were aligned to exon regions, 87.56% (S1, Fig. 2a), 87.09% (S2, Fig. 2b) and 87.56% (S3, Fig. 2c) in each library.

Global gene expression and SNP/InDel analysis

Overall, 19,442, 19,143, and 18,941 expressed genes were identified at the S1, S2 and S3 stage, respectively (Fig. 3a). In all, there were 20,423 genes were identified during the entire seed development (Fig. 3b). Among these, 17,872 genes were continuously expressed from S1 to S3, while conversely, 1,192 genes (529 in S1, 367 in S2 and 296 in S3) were uniquely expressed at 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 S3, 318 in S2 and S3, 384 in S1 and S3) were found 657 to be expressed simultaneously at two stages. (Fig. 3b).

We identified a total of 130,461 polymorphisms (97,932 SNPs and 32,529 InDels) at the three time points. The annotated information showed that the top locations for both SNPs (Fig. 4a) and InDels (Fig. 4b) were intergenic, upstream and downstream of genes.

Identification of Differentially expressed genes related to ricinoleic acid biosynthesis during *R. communis* seed development

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, 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) database, to obtain their annotated information using Basic Local Alignment Search Tool (BLAST) software [32].

Next, the number of DEG were analyzed between any two libraries (S1 vs. S2, S2 vs. S3 and S1 vs. S3). In above three paired comparison groups, of the number of up-regulated DEGs was all higher than the number of down-regulated DEGs (Fig. 5). 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 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. 5). Heat maps from the hierarchical clustering of DEGs between S1 vs. S2 (Fig. 6a), S2 vs. S3 (Fig. 6b) and S1 vs. S3 (Fig. 6c) paired groups were shown in Fig. 6.

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 a blast with the GO, COG, Pfam, KEGG, Swiss-Prot and NR databases using BLAST, respectively. All annotated information of DEGs is listed in Additional file 1: Table S2-S4.

For the GO classification analysis of DEGs, all DEGs were assigned to three main categories and then divided into 42 (S1 vs. S2) (Fig. 7a), 43 (S1 vs. S3) (Fig. 7b) and 43 (S2 vs. S3) (Fig. 7c) sub-categories. 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) (Fig. 8a), 2,893 (S2 vs. S3) (Fig. 8b) and 1,739 (S1 vs. S3) (Fig. 8c) DEGs, involved in 119, 125 and 125 different pathways, respectively. 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).

Table 1
Functional annotations of DEGs between different stages.

DEG Set	Total	COG	eggNOG	NR	Pfam	Swiss-Prot	GO	KEGG
S1 vs. S2	4,240	1,798	4,092	4,194	3,579	3,233	3,277	1,543
S1 vs. S3	7,807	3,223	7,540	7,687	6,572	5,725	5,886	2,893
S2 vs. S3	4,936	2,002	4,785	4,866	4,198	3,670	3,691	1,739

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

The key enzymes in oil synthesis include *ACCase*, *ACP*, *KAS*, *KAR*, *EAR*, *SAD*, *FAT*, *LACS*, *ACBP*, *LPCAT*, *FAH12*, *PLA2*, *GPAT*, *LPAAT*, *PAP*, *DGAT*, *PDCT*, *PDAT*, *PLD*, *PLC2*, *OLE* and *PXG*. Based on the annotation information, 49 key enzyme DEGs associated with ricinoleic acid biosynthesis were identified (Additional file 1: Table S5). Derived from the normalized Fragments per Kilobase of transcript per Million mapped reads (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. 9). Cluster I (a rising expression model) was composed of 10 genes, with the increasing $\log_2(\text{FPKM} + 1)$ values of DEGs from S1

toS3(Fig. 9a). Cluster II (a bell-shaped pattern) possessed 10 genes with the rising $\log_2(\text{FPKM} + 1)$ values of DEGs raised from S1 to S2 first and then declining values from S2 to S3 (Fig. 9b). Cluster III was composed of 29 genes with decreasing expression patterns, whereby the $\log_2(\text{FPKM} + 1)$ values of DEGs dropped from S1 to S3 (Fig. 9c).

Selection of differentially expressed oil-related genes in *R. communis* and qRT-PCR confirmation

According to the analysis results of castor oil and ricinoleic acid content during castor seed development, we found that the contents 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 29673.t000033), two fatty acid biosynthesis genes (*LACS7* 29844.t000211 and *ACCCase-α* 30174.t000396), as well as four glycerolipid metabolism genes (*PLA2s* 30142.t000005, *PLC* 29756.t000030, *DGAT* 29801.t000101 and *PDAT* 29637.t000044) (Table 3).

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

Symbol	Gene ID	Pathway Description	Enzymes	FPKM		
				S1	S2	S3
<i>OLE</i>	29794.t000071	lipid storage	<i>Oleosin</i>	300.13	1184.99	4728.43
<i>OLE</i>	30147.t000162	lipid storage	<i>Oleosin</i>	209.82	1114.64	2037.97
<i>PXG</i>	30008.t000036	lipid oxidation	<i>Peroxygenase</i>	181.32	1240.02	1505.61
<i>LACS7</i>	29844.t000211	Fatty acid biosynthesis	<i>Long chain acyl-CoA synthetase</i>	7.64	20.13	522.62
<i>PLA2</i>	30142.t000005	Glycerolipid metabolism	<i>Phospholipase A2</i>	7.74	30.86	224.60
<i>PXG4</i>	29673.t000033	lipid oxidation	<i>Peroxygenase</i>	11.30	34.00	54.65
<i>PLC</i>	29756.t000030	Glycerolipid metabolism	<i>Phospholipase C</i>	0.14	9.27	28.64
<i>ACCCase</i>	30174.t000396	Fatty acid biosynthesis	<i>Acetyl-coenzyme A carboxylase carboxyl transferase</i>	1.47	6.50	18.38
<i>DGAT</i>	29801.t000101	Glycerolipid metabolism	<i>Diacylglycerol P-acyltransferase</i>	1.18	4.37	9.76
<i>PDAT</i>	29637.t000044	Glycerolipid metabolism	<i>Phospholipid:diacylglycerol acyltransferase</i>	1.71	5.69	7.82

The expression changes of 10 key candidate genes were confirmed by qRT-PCR analysis to validate the Illumina RNA sequencing data (Fig. 10). The results suggested that the data by RNA-seq were consistent with

qRT-PCR analysis during castor seed development.

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.

Gene expression analysis between different seed maturation stages

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 upregulated 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 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 S2 vs. S3 group, indicating that these six genes played stable roles in ricinoleic acid biosynthesis and regulation during the seed maturation.

Expression of oil-related genes during seed development after flowering

In this study, 20 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 enzymes for TAG formation, has a dominant role in adding the hydroxyl fatty acyl into TAGs [23]. One *DGAT*-encoding gene (*29801.t000101*), found in this study during castor 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 levels at S3 during castor seed development in this study (Table 3). Cagliari et al. (2010) also reported that *DGAT2* played a vital role in the enrichment of ricinoleic acid in *R. communis* [33]. Furthermore, in other plants which can produce ricinoleic acid, *DGAT* also possesses a significant affect in ricinoleic acid-enriched TAGs enrichment. In *P. fendleri*, *DGAT1*, *DGAT2* and *PDAT2* were all conducive primarily to hydroxy fatty acid enrichment [34], whereas in *H. benghalensis*, *DGAT2* and *PDAT2* were found to own high expression levels in the process of ricinoleic acid biosynthesis [31].

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* and etc., in the Kennedy pathway [15]. In this study, a *LACS7* gene (*29844.t000211*), had high expression levels during seed development, may contribute

to ricinoleic acid accumulation (Table 3). Nevertheless, in *H. benghalensis*, *LACS8* has been reported to have the same effect [31]. At the same time, *LACS9* as the dominant LACS isoform was revealed to act an vital role in oil synthesis in *R. communis* and *P. fendleri* [34]. During castor seed development, *PLA2* can help to accumulate a high level of the hydroxy fatty acid ricinoleate in TAGs [35]. In this study, a *PLA2* gene (*30142.t000005*) showed a high expressed level at S3 during seed development, may have 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 [36]. In this study, two *OLE* genes (*29794.t000071* and *30147.t000162*) showed the highest expression level among all selected candidate key DEGs. Tian et al. reported that high expression of *OLE3* during seed development in *H. benghalensis* was participated in the accumulation of ricinoleic acid in oil-body formation [31]. *OLE* genes had been reported to contribute to monoacylglycerol acyltransferase and *PLA2* bifunctional enzyme activity in peanuts [37]. The *PXG* gene encodes a caleosin, another kind of oil body [38]. In this study, two *PXG* genes (*30008.t000036* and *29673.t000033*) showed a high expression level at S3 during seed development 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* [33] or *P. fendleri* [34].

Lin et al. (2007) found *PLC2* catalyzes the transformation of 2-oleoyl-PC into 1-acyl-2-oleoyl-sn-glycerol [39]. *PLC* and/or *PLD*, together with *PAPase*, may also relate to the transformation of PtdCho to DAG [40]. In this study, a *PLC* gene (*29756.t000030*) showed high expression levels at S3 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*) showed a high expressed level at S3 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 [40]. The *ACCase*-catalyzed reaction is a rate-limiting step in 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 [41–42].

To deep illustrate the mechanism of ricinoleic acid biosynthesis, a theoretical model of gene expression changes during seed development was constructed based on our transcriptomic data (Fig. 11).

Conclusions

In this study, the mechanism and regulation of ricinoleic acid biosynthesis in *R. communis* was revealed at the transcriptome level. A series of genes encoding lipid-related enzymes were screened, and their expression profiles during ricinoleic acid accumulation were obtained through bioinformatics analysis. Based on the results of this study and previous reports, a theoretical model of oil accumulation in developing seeds of *R. communis* was constructed. The findings of this study broaden our understanding of ricinoleic acid biosynthesis and regulation at the level of molecular biology, laying a solid basis for improving the ricinoleic acid accumulation capacity of important oil crops.

Methods

Plant material

The castor line (*R. communis* L.) 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, National Infrastructure for Crop Germplasm Resources (OCRI, CAAS, Wuhan, China), with the 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 seeds were sown and planted in the Yangluo experimental field of the OCRI, CAAS, under standard field conditions, from March to September of 2018. Nine immature seeds [three time points (15, 30 and 45 DAP) × three biological replicates] were collected. Seed sample at different developmental stages were immediately frozen and stored at -80 °C.

Analysis of the oil and ricinoleic acid content

Ten immature seeds of castor line 93 – 10 during seed development at three time points were harvested with three biological replicates. After removing the seed coat, the seeds were weighed in the fresh state, dried in the oven at 60 °C for 48 h, and weighted again in the dry state. The water content of seeds was calculated by subtracting the dry weight from the fresh weight of the seeds. To measure the castor oil content the dry seeds were ground into a powder with a mortar and pestle, wrapped in filter paper, and extracted with n-hexane in a Soxhlet extractor at 85 °C for 12 hours. After extraction, the n-hexane was removed in a rotary evaporator, followed by drying in an oven at 68 °C for 1 hour. The resulting lipid mass was used to calculate the oil content as the percentage of oil in dry seeds (w/w). The ricinoleic acid content was measured by gas chromatography. Heptadecanoic acid (C17:0) (≥ 99%, GLPBIO, USA) was chosen as the internal standard, following the method of Pan et al with minor modification [31, 43]. About 15 mg of seed samples were methylated to generate fatty acid methyl esters with 3 N methanolic HCl at 80 °C for 2 h, which were extracted with n-hexane, dried in nitrogen and suspended in 1.5 mL dichloromethane. The analysis was carried out by an Agilent 6890 N gas chromatograph, was and then equipped with a hydrogen flame ionization detector and a DB-WAX capillary column (30 m × 0.32 mm × 0.15 µm) (Agilent, USA). The temperature program was 200 °C keep for 25 min, 5 °C per min until to 220 °C for 20 min. Then 1 µL of injection volume, 250 °C of injector temperature, and a split injection mode with a split ratio of 30:1 was used. Helium was used as carrier gas with a flow rate of 1.5 mL per min. Fatty acids meet fatty acid methyl ester standards (Sigma-Aldrich, USA). The relative percentage of ricinoleic acid was calculated from its peak area.

Total RNA extraction and library construction

According to the manufacturer's instructions, 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, California, 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 samples, and then sequenced using the Illumina HiSeq™ 4000 sequencing platform (Illumina Inc., San Diego, CA, USA) at Biomarker Technologies Corporation (Beijing, China).

Bioinformatics analysis

FastQC (v0.11.8) was used for quality-checking the raw RNA-seq reads to eliminate the low-quality reads and adaptor sequences [44]. The high-quality clean reads were then mapped to the published *R. communis* genome sequences [45]. Bowtie (v2.2.3) software was used to adjust the reference genome [46]. The paired-end clean reads were aligned to the reference genome using TopHat (v2.0.12) [47]. Cufflinks (v2.1.1) was a software used to detect all transcripts via qRT-PCR [48]. GATK [49] was used to identify single base mismatches between sequenced samples and the reference genome, and identify potential SNP loci. SnpEff [50] was used for annotating variations (SNP, InDel) and predicting the effects of variations.

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 FPKM method. DEG screening was conducted using DESeq to analyze biological duplicates. During the DEG screening, DEG fold change was ≥ 8 and a FDR < 0.001 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 blast by the GO, COG, KEGG pathway, SwissProt and NR databases using BLAST software 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.

QRT-PCR analysis

The transcript levels of a total of 10 genes at three time points (S1, S2 and S3) were performed by qRT-PCR analysis to verify the Illumina RNA sequencing data. Total RNA of developing seeds at S1-S3 was isolated and 1 μg was reverse-transcribed using the PrimeScript™ RT reagent Kit with gDNA Eraser (Takara, Japan). The actin gene of *R. communis* was used as the internal standard [51]. Primers were designed by Primer 5.0 software (Premier, Canada). All gene primer sequences including actin gene are listed in Table S6.. QRT-PCR used SYBR Green qRT-PCR Master Mix kit (Kapa, 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, USA). All reactions (20 μl) were performed in three technical replicates. Gene relative expression levels were calculated using the $2^{-\Delta\Delta\text{Ct}}$ comparative threshold cycle (Ct) method [52].

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 **DAGT**: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 **QRT-PCR**:Quantitative real-time reverse transcription PCR
RRR:Triricinolein **SAD**:stearoyl-ACP desaturase **SGS**:Second-generation sequencing **Swiss-Prot**:The manually
annotated and reviewed protein sequence database **TAG**:Triacylglycerol

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Data for this study will be available upon publication of the manuscript; until then, the raw data can be made available to researchers upon reasonable request to the corresponding author.

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by National Natural Science Foundation of China (No.31801275), Science and Technology Research Project, Science and Technology Department of Hubei Province (2018CFB310). The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

Author Contributions

X.G. conceived and designed the experiments; Z.W., Q.Y. and Y.OY performed the experiments; X.G. and F.X. analyzed the data; X.G. wrote the manuscript. All authors have read the final version of the manuscript and agreed for it to be published.

Acknowledgments

We thank Lijun Wang (Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, China) for providing castor beans and Biomarker Technologies Co. (Beijing, China) for providing technical support.

References

1. Lin J and Arcinas A. Regiospecific analysis of diricinoleoylacylglycerols in castor (*Ricinus communis*) oil by electrospray ionization-mass spectrometry. *J Agric Food Chem*. 2007;55: 2209-2216.
2. Broun R and Somerville C. Accumulation of ricinoleic, lesquerolic, and densipolic acids in seeds of transgenic *Arabidopsis* plants that express a fatty acyl hydroxylase cDNA from castor bean. *Plant physiol*. 1997;113:933-942.
3. van Erp H, Bates PD, Burgal J, Shockey J, Browse J. Castor phospholipid:diacylglycerol acyltransferase facilitates efficient metabolism of hydroxyl fatty acids in transgenic *Arabidopsis*. *Plant physiol*. 2011;15:683-693.
4. Tsuboi S, Osafune T, Tsugeki R, Nishimura M, Yamada M. Nonspecific Lipid Transfer Protein in Castor Bean Cotyledon Cells: Subcellular Localization and a Possible Role in Lipid Metabolism. *J Biochem*. 1992;111:500-508.
5. Sasaki Y and Nagano Y. Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Biosci BioWchnoL Biochem*. 2004;68:1175-1184.
6. Zhou P, Florova Q, Reynolds KA. Polyketide synthase acyl carrier protein (ACP) as a substrate and a catalyst for malonyl ACP biosynthesis. *Chem Biol*. 1999;6: 577-584.
7. Puyaubert J, Dieryck W, Costaglioli P, Chevalier S, Breton A, Lessire R. Temporal gene expression of 3-ketoacyl-CoA reductase is different in high and in low erucic acid. *Brassica napus* cultivars during seed development. *Biochim Biophys Acta*. 2005;1687:152-163.
8. Saito J, Yamada M, Watanabe T, Iida M, Kitagawa H, Takahata S, et al. Crystal structure of enoyl-acyl carrier protein reductase (FabK) from *Streptococcus pneumoniae* reveals the binding mode of an inhibitor. *Protein Sci*. 2008;17:691-699.
9. 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 diiron proteins. *Eeboj*. 1996;15:4081-4092.
10. Slabas AR, White AOE, Fawcett T. Investigations into the regulation of lipid biosynthesis in *Brassica napus* using antisense down-regulation. *Biochem Soc Trans*. 2002;30:1056-1059.
11. Chen G, Woodfield HK, Pan X, Harwood JL, Weselake RJ. Acyl-trafficking during plant oil accumulation. *Lipids*. 2015;50:1-12.
12. Bates PD. Understanding the control of acyl flux through the lipid metabolic network of plant oil biosynthesis. *Biochim Biophys Acta Mol Cell Biol Lipids*. 2016;1861:1214-1225.
13. Bafor M, Smith MA, Jonsson L, Stobart K, Stymne S. Ricinoleic acid biosynthesis and triacylglycerol assembly in microsomal preparations from developing castor bean (*Ricinus communis*) endosperm. *Biochem J*. 1991;280:507-514.
14. Zhou XR, Singh SP, Green AG. Characterization of the *FAD2* gene family from *Hiptage benghalensis*: a ricinoleic acid accumulating plant. *Phytochemistry*. 2013;92:42-48.
15. Xu Y, Holic R, Li D, Pan X, Mietkiewska E, Chen G, et al. Substrate preferences of long-chain acyl-CoA synthetase and diacylglycerol acyltransferase contribute to enrichment of flax seed oil with α -linolenic acid. *Biochem J*. 2018;475:1473-1489.
16. Lung SC and Chye ML. The binding versatility of plant acyl-CoA-binding proteins and their significance in lipid metabolism. *Biochim Biophys Acta Mol Cell Biol Lipids*. 2016;1861:1409-1421.

17. Broun P and Somerville C. Accumulation of ricinoleic, lesquerolic, and densipolic acids in seeds of transgenic *Arabidopsis* plants that express a fatty acyl hydroxylase cDNA from castor bean. *Plant Physiology*. 1997;113:933-942.
18. Bates PD and 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*, 2011;68:387–399.
19. Weiss SB, Kennedy EP, Kiyasu JY. The enzymatic synthesis of triglycerides. *J Biol Chem*. 1960;235:40-44.
20. Hu Z, Ren Z, Lu C. The phosphatidylcholine diacylglycerolcholinephospho transferase is required for efficient hydroxy fatty acid accumulation in transgenic *Arabidopsis*. *Plant Physiol*. 2012;158:1944-1954.
21. Lu C, Xin Z, Ren Z, Miquel M, Browse J. An enzyme regulating triacylglycerol composition is encoded by the *ROD1* gene of *Arabidopsis*. *Proc Natl Acad Sci USA*. 2009;106:18837-18842.
22. Huang AHC. Plant lipid droplets and their associated proteins: potential for rapid advances. *Plant Physiol*. 2018;176:1894-1918.
23. Xu Y, Caldo KMP, Pal-Nath D, Ozga J, Lemieux MJ, Weselake RJ, et al. Properties and biotechnological applications of acyl-CoA: diacylglycerol acyltransferase and phospholipid: diacylglycerol acyltransferase from terrestrial plants and microalgae. 2018;53:663-688.
24. Du H, Yu Y, Ma Y, Gao Q, Cao Y, Chen Z, et al. Sequencing and de novo assembly of a near complete indica rice genome. *Nat Commun*. 2017;8:15324.
25. Madeleine C, Andrey AY, Julian JA, Colin EA, Pawel H, Kathryn RE. Correction to: De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC Genomics*. 2018;19:448.
26. Weng J, Gu S, Wan X, Gao H, Guo T, Su N, et al. Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res*. 2008;18:1199-1209.
27. Depre C, Powell SR, Wang X. The role of the ubiquitin-proteasome pathway in cardiovascular disease. *Cardiovasc Res*. 2010; 85:251-252.
28. Stangl K, Stangl V. The ubiquitin-proteasome pathway and endothelial (dys) function. *Cardiovasc Res*. 2010;85:281-290.
29. Chandrasekaran U, Xu W, Liu AZ. Transcriptome profiling identifies ABA mediated regulatory changes towards storage filling in developing seeds of castor bean (*Ricinus communis*). *Cell and bioscience*, 2014;4:33.
30. Han B, Xu H, Feng Y et al. Genomic Characterization and Expressional Profiles of Autophagy-Related Genes (ATGs) in Oilseed Crop Castor Bean (*Ricinus communis*). *J. Mol. Sci*. 2020;21:562.
31. 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*. 2019;12:16.
32. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSIBLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–3402.
33. Troncoso-Ponce MA, Kilaru A, Cao X, Durrett TP, Fan J, Jensen JK, et al. Comparative deep transcriptional profiling of four developing oilseeds. *Plant J*. 2011;68:1014–1027.

34. Horn PJ, Liu J, Cocuron JC, McGlew K, Thrower NA, Larson M, et al. 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. 2016;86:322–348.
35. Chen et al. Acyl-Trafficking during plant oil accumulation. Lipids. 2015;50:1057–1068.
36. Badami RC, Kudari SM. Analysis of Hiptage madablota seed oil. J Sci Food Agric. 1970;21:248–249.
37. 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. 2012;287:1946–1954.
38. Hanano A, Burcklen M, Flenet M, Ivancich A, Louwagie M, Garin J, et al. Plant seed peroxygenase is an original heme-oxygenase with an EF-hand calcium binding motif. J Biol Chem. 2006;281:33140-33151.
39. Chen GQ, Lin JT, Lu C. Hydroxy fatty acid synthesis and lipid gene expression during seed development in *Lesquerella fendleri*. Ind Crops Prod. 2011;34:1286–1292.
40. Sasaki Y, Nagano Y. Plant acetyl-CoA carboxylase—structure—biosynthesis—regulation—and gene manipulation for plant breeding—Biosci BioWchnoL Biochem. 2004;68(6):1175-1184.
41. Roesler K, Shintani D, Savage L, Boddupalli S and Ohlrogge J. Targeting of the Arabidopsis Homomeric Acetyl-Coenzyme A Carboxylase to Plastids of Rapeseeds. Plant Physiol. 1997;1(13):75-81.
42. Thelen JJ and Ohlrogge JB. Both antisense and sense expression of biotin carboxyl carrier protein isoform 2 inactivates the plastid acetyl-coenzyme A carboxylase in Arabidopsis thaliana. Plant J. 2002;32:419-431.
43. Pan X, Siloto RM, Wickramarathna AD, Mietkiewska E, Weselake RJ. Identification of a pair of phospholipid:diacylglycerol acyltransferases from developing flax (*Linum usitatissimum* L.) seed catalyzing the selective production of trilinolenin. J Biol Chem. 2013;288:24173–24188.
44. 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. 2018;666:123–133.
45. Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, et al. Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol. 2010;28:951–956.
46. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9:357.
47. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105–1111.
48. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. Nat Protoc. 2012;7:562.
49. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 2010;20:1297-1303.
50. Cingolani P, Platts A, Le L W, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly. 2012;6:80.
51. Wang X, Wang L, Yan X, Wang L, Tan M, Geng X, et al. Transcriptome analysis of the germinated seeds identifies low-temperature responsive genes involved in germination process in *Ricinus communis*. Acta Physiol Plant. 2016;38:6.
52. Chmitten TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. Nat Protoc. 2008;3:1101–1108.

Description Of Supplementary Files

- Table S1.** Statistics of the comparison of sequencing data with the reference genome.
- Table S2.** List of all annotated information of DEGs between S1 and S2.
- Table S3.** List of all annotated information of DEGs between S2 and S3.
- Table S4.** List of all annotated information of DEGs between S1 and S3.
- Table S5.** List of 49 key enzymes associated with lipid biosynthesis.
- Table S6.** Primers used for quantitative real-time PCR.

Figures

Figure 1

The castor oil content (A) and ricinoleic acid (C18:1-OH) (B) content of developing castor bean (mean±SD, n=3).

Figure 2

Statistics of mapped reads in different regions of the genome for three the libraries S1 (A), S2 (B), and S3 (C).

Figure 3

Transcriptome analysis of *R. communis* seed development. (A) Number of expressed genes at each developmental stage, and (B) their relationships presented in the form of a Venn diagram.

Figure 4

The annotation of all SNPs (A) and InDels (B).

Figure 5

Analysis of differential genes expression in paired comparison groups among the three development stages.

Figure 6

Hierarchical clustering analysis of intersectional DEGs in the three paired groups (S1 vs. S2, S1 vs. S3, and S2 vs. S3) based on data of log2Ratio from one gene. The color bar ranging from blue to red indicates low to high transcriptional expression level.

Figure 7

GO classifications of differentially expressed genes (DEGs) in the three paired groups (S1 vs. S2, S2 vs. S3, S1 vs. S3). DEGs are divided into three main categories: biological process, molecular function and cellular component by GO analysis.

Figure 8

KEGG pathway categories of differentially expressed genes (DEGs) in the three paired groups (S1 vs. S2, S2 vs. S3, and S1 vs. S3).

Figure 9

Three clusters with different gene expression patterns for all key enzyme genes. (A) Cluster I: rising pattern; (B) Cluster II: bell-shaped pattern; (C) Cluster III: decreasing pattern.

Figure 10

The expression changes of 10 key differentially expressed lipid-related genes were confirmed by qRT-PCR analysis to validate the Illumina RNA sequencing data. The comparative log2FC and $\Delta\Delta C_t$ values at stage S1 were used as the control for normalization. The results represent the means of three biological replicates (mean $\pm$ SD, n=3).