Original Article
Food Science and Technology

ISSN 0101-2061 (Print)
ISSN 1678-457X (Online)

DOI: https://doi.org/10.1590/fst.93721



# Gene screening for fatty acid synthesis of flax based on transcriptome sequencing

Shuyan WANG[1+*], Jun XU[2+], Zhuo WANG[3], Zhiwei LI[1], Liuxi YI[1], Lijun YAO[1], Xia WANG[1]

**Abstract**

To understand differential expression of genes (DEGs) based on transcriptome and transcription features of fatty acid desaturase genes (*fad*s) in development of oil flax capsules. Three varieties of flax with high, medium and low linolenic acid (C18:3) content was investigated. The differential expressed genes (DEGs) of seed developing periods (15 d and 30 d after flowering, DAF) were compared with that of the flowering day based on RNA sequencing technology. DEGs related to fatty acid accumulation were distinguished and compared with Gene Ontolygy (GO) database and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. In the synthesis of unsaturated fatty acids and α-linolenic acid metabolism, DEGs acted in multiple processes and signaling pathways. Lus_GLEAN_10046787, Lus_GLEAN_10024616, Lus_GLEAN_10024710, Lus_GLEAN_10046516, scaffold280_unigene_193 and scaffold272_unigene_19 participated in 23, 11, 12, 12, 12, 13 processes in GO description, respectively. And they acted in fatty acid degradation (ath00071), metabolism (ath01212), biosynthesis (ath00061) and elongation (ath00062) in the KEGG signaling pathway. The scaffold96_unigene_47 was differentially expressed between FR46 and FR48 with different linolenic acid contents. The DEGs can be utilized describe the structure and transcription expression features, which contribute to the selection and utilization of flax breeding lines.

**Keywords:** flax; gene sequencing; fatty acid; transcriptome sequencing.

**Practical Application:** DEGs can be utilized describe structure and transcription expression feature.

## 1 Introduction

Flax (*Linum usitatissimum L.*), as an annual herb of *Linaceae*, is an ancient bast fiber crop and oil crop (Ander et al., 2004; Ma et al., 2017; Oomah, 2001; Thambugala et al., 2013; Thambugala & Cloutier, 2014; Wiesenfeld et al., 2003; Zhang et al., 2018). Flax is cold-resistant, drought-tolerant and has rich nutritional value (Ander et al., 2004; Dyer et al., 2008; Damude & Kinney, 2008; Dash et al., 2017; Wang et al., 2012; Dang et al., 2010). The flax genome sequence is based on the Center for Crop Development (CDC) Bethune variety, which is a highly related hybrid and widely cultivated in Canada (Rowland et al., 2002). Genomic resources in flax were assessed and selected (Dash et al., 2014; 2017) to accelerate its varietal improvement program. RNA sequencing in flax can be performed to identify the differentially-expressed genes (DEGs) related to edaphic stress (Wu & Zhou, 2018), pathogenic fungus (Galindo-González & Deyholos, 2016) flax lodging resistance. fiber development in flax (Zhang & Deyholos, 2016; Gorshkov et al., 2017). Real-time fluorescence quantitative PCR can be utilized to verify the authenticity of DEGs (Li et al., 2016; Wu & Zhou, 2018). High-throughput RNA sequencing accelerates the discovery and exploration of DEGs (Wang et al., 2009).

In previous studies, there was a significant correlation between the expression patterns of *fad2a* and *fad3a* genes and flax fatty acid synthesis (Wang et al., 2016; Thambugala & Cloutier, 2014; Wang et al., 2019). *Fad2* and *fad3* genes have different subtypes (Vrinten et al., 2005). The differential expression of these genes in flax varieties with different fatty acid contents has been studied by real-time fluorescence quantitative analysis (Banik et al., 2011; Abuelnaga et al., 2021). In this study, different varieties of fatty acids were selected to study the genes related to the synthesis of unsaturated fatty acids.

## 2 Materials and methods

### *2.1 Growing location and sample collection*

The test materials were grown in the teaching farm of Inner Mongolia Agricultural University on April 28, 2018. Farm soil texture is light soil. The annual precipitation ranges from 335.2 to 534.6 mm. From June to July 2018, flax capsules of three varieties with FR48, NeiYa No.7 (symbol as NY) and FR46 were randomly collected representing high, medium and low linolenic acid genotypes, respectively. Transcriptome sequencing samples were capsules of Neiya No.7 in flowering day (or 0 d), 15 d after flowering day (DAF) and 30 d DAF, and capsules of FR48 15DAF and that of FR46 15 DAF. A total of 15 samples were sent to Beijing Novogenes Technology Co., Ltd for RNA sequencing.

*Received 27 Sept., 2021*
*Accepted 03 Nov., 2021*
[1] *Agricultural College, Inner Mongolia Agricultural University, Inner Mongolia, China*
[2] *College of Grassland Resources and Environment, Inner Mongolia Agricultural University, Inner Mongolia, China*
[3] *Inner Mongolia Grass Digital Ecological Industry Co., LTD, Inner Mongolia, China*
*Corresponding author: wangsci69@sina.com
[+] *Shuyan Wang and Jun Xu are both the co-first author of this study.*

### *2.2 Transcriptome sequencing*

*RNA extraction*

RNA was extracted by the standard method, and then subjected to quality control (including agarose gel electrophoresis: RNA integrity and DNA contamination of samples were analyzed; Determination of RNA purity (od260/280 and od260/230 ratios); Agilent 2100 bioanalyzer: accurate detection of RNA integrity.

*Library construction and quality control*

The first strand of cDNA was synthesized using fragmented mRNA as a template and random oligonucleotides as primers, and then RNA strand was degraded by RNaseH and the second strand was synthesized using dNTPs under the DNA polymerase I system. AMPure XP beads were used to screen cDNA around 250-300 bp after the purification of double-stranded cDNA through terminal repair. AMPure XP beads were prepared for PCR amplification and the PCR products were purified to obtain the library. Accurate quantification of the effective concentration of the library (the effective concentration of the library was higher than 2 nM) was conducted by qRT-PCR to ensure the quality of the library.

*Computer sequencing*

After library inspection was qualified, Illumina sequencing was conducted after pooling different libraries. The basic principle of sequencing was sequencing followed by synthesis. In the sequencing of the flow cell with four kinds of fluorescence-labeled dNTP, primers for amplification, DNA polymerase and joint in each sequence complementary chain clusters, each to join a fluorescence-labeled dNTP could release the corresponding fluorescence, sequencing machine by capturing the fluorescent signal, and through the computer software to convert optical signals to sequencing peak, thereby gaining fragment awaiting testing sequence information.

*Sequencing analysis*

Sequencing fragments were randomly interrupted by mRNA. To determine which genes transcribed these fragments, the quality-controlled clean reads should be matched to the reference genome. The HISAT2 alignment software, developed by a laboratory at Johns Hopkins university, USES global and local search methods and a large number of FM indexes were employed to cover the entire genome.

*Gene expression quantification*

Subread featureCounts of software tools were utilized for quantitative analysis of gene expression levels, featureCounts mainly utilized - Q 10 - B - C parameters, respectively, to filter out than quality value less than 10 reads, reads the paired comparison, comparing to the genome of multiple region reads. The meta-features are primarily the genes corresponding to the annotation file. By sorting the features on each chromosome according to the initial position, the chromosomes are divided into 128 kb unoverlapping bins, and allocated to the corresponding bin according to the initial position of the features, where the solid lines represent the features. The reads searched in the figure are first compared with bins in the genome, then with blocks in bin, and finally with features in blocks. The final read localization result is located in the first blocks in the ith bin, and there is overlap between read and two features in blocks, namely junction read. The transcription factor families of genes are annotated by the transcriptional factor database (AnimalTFDB/PlantTFDB) (for species included in the database) and predicted by the protein domain databases such as Pfam/SUPERFAMILY.

## 3 Results

### *3.1 Transcriptome sequencing and assembly*

*Splicing of RNA-seq*

After the removal of low-quality reads ($Q_{phred}$ <= 20 base number accounts for more than 50% of the read length), the proportion of available data reached more than 95%, where 40 to 60 million clear reads were generated from each transcriptome library. GC pct represents GC content. GC and AT contents of each sequencing reading each location should be equal, and the process is basically stable and unchanged, showing a horizontal line. Only single chain information is retained, and AT and GC are separated. GC PCT values fluctuate around 50% among all samples. In line with quality control requirements, further analysis can be carried out (Table 1).

### *3.2 Quantitative analysis*

The read count ranged from 36550 (FR46_I) to 309186(NY30_II) in 15 samples. However, regarding the influence of sequencing depth and gene length, the RNA-seq was generally not expressed by reads count, but by FPKM, referring to total exon fragments per kilobase sequence per million mapped reads, and the sequencing depth and gene length were successively corrected (Du H.2016). When FPKM>1, falls into the category of gene expression (Figure 1). FPKM values of 15, DAF of three flax varieties were higher than that of NY0 and NY30.

### *3.3 Principle component analysis*

As illustrated in Figure 2, the developing stages of flax capsules contributed 65.82% of principle component in the transcription data. Three biological repetitions of 3 stages (0 d,15 d and 30 d) clusters were identified. The 3 flax varieties of 15DAF gathered in 0-50/-50-0 area of the coordinate axis, genetic backgrounds made distinct area difference from the biological repetitions.

### *3.4 DEGs*

According to the KEGG signaling pathways in each group, 5 signaling pathways related to desaturatase were selected. GO gene function was annotated for the genes in the signaling pathway. A total of 23 related genes were distributed in 5 KEGG signaling pathways, which controlled the synthesis and metabolic function of fatty acids, as illustrated in Schedule 1.

**Table 1**. Summary of data production.

| Sample | Raw_reads | Clean_reads | Clean_bases | Clean reads rate(%) | Q20 (%) | Q30 (%) | GC pct(%) |
|---|---|---|---|---|---|---|---|
| NY0_I | 53979422 | 52615328 | 7.89G | 97.47 | 97.96 | 93.49 | 47.03 |
| NY0_II | 49215492 | 48246352 | 7.24G | 98.03 | 97.82 | 93.86 | 48.73 |
| NY0_III | 61943702 | 60433436 | 9.07G | 97.56 | 97.23 | 92.43 | 48.6 |
| NY15_I | 53194756 | 51737210 | 7.76G | 97.26 | 97.5 | 93.21 | 51.47 |
| NY15_II | 54878984 | 53502538 | 8.03G | 97.49 | 97.74 | 93.71 | 51.28 |
| NY15_III | 42369422 | 41038486 | 6.16G | 96.86 | 97.15 | 92.48 | 51.54 |
| NY30_I | 59028088 | 57079880 | 8.56G | 96.7 | 97.23 | 92.44 | 49.32 |
| NY30_II | 58528158 | 57309504 | 8.6G | 97.92 | 96.99 | 92.11 | 51.87 |
| NY30_III | 53100994 | 51489172 | 7.72G | 96.96 | 97.43 | 93.09 | 50.68 |
| FR46_I | 53693140 | 52489968 | 7.87G | 97.76 | 96.57 | 91.1 | 52.16 |
| FR46_II | 53903250 | 52290664 | 7.84G | 97.01 | 97.82 | 93.9 | 51.52 |
| FR46_III | 56331300 | 54947920 | 8.24G | 97.54 | 97.44 | 93.12 | 51.47 |
| FR48_I | 59054448 | 56672616 | 8.5G | 95.97 | 97.54 | 93.31 | 51.52 |
| FR48_II | 49132348 | 47554884 | 7.13G | 96.79 | 94.93 | 87.29 | 51.95 |
| FR48_III | 49397972 | 48182392 | 7.23G | 97.54 | 97.77 | 93.85 | 51.63 |

Note: NY0 represents 'Neiya No.7'flowering day (or 0 d); NY15, 'Neiya No.7' 15 d after flowering day; NY30, 'Neiya No.7' 30 d after flowering day; FR46, 'FR46' 15 d after flowering day; FR48, 'FR48' 15 d after flowering day; I-III represents 3 repetitions; raw_reads: the number of reads in the original data; clean_reads: the number of reads after the original data is filtered; clean_bases: the number of bases after the original data is filtered; clean reads rate: the number of reads after the original data is filtered. Percentage of reads in Q20: the percentage of bases with a Phred value greater than 20 as a percentage of total bases; Q30: the percentage of bases with a Phred value greater than 30 as a percentage of total bases; GC_pct: clean reads where G and C account for four bases Percentage of base.

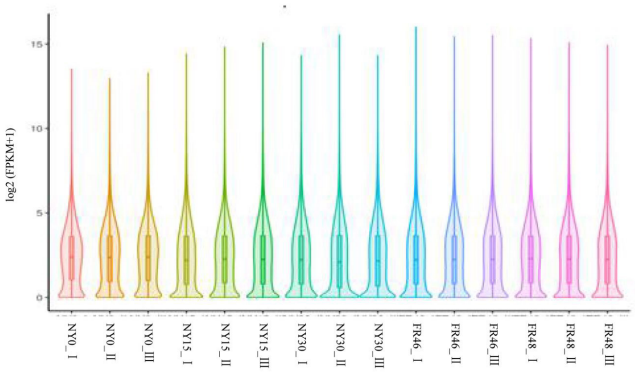


**Figure 1**. Comparison of three repeated FPKM values of flax varieties in different periods. Note: The abscissa shows the flax capsule samples names for each period, and the ordinate shows the log2 (FPKM+1) of gene expression level. NY0 represents 'Neiya No.7'flowering day (or 0 d); NY15, 'Neiya No.7' 15 d after flowering day; NY30, 'Neiya No.7' 30 d after flowering day; FR46, 'FR46' 15 d after flowering day; FR48, 'FR48' 15 d after flowering day; I-III represents 3 repetitions.

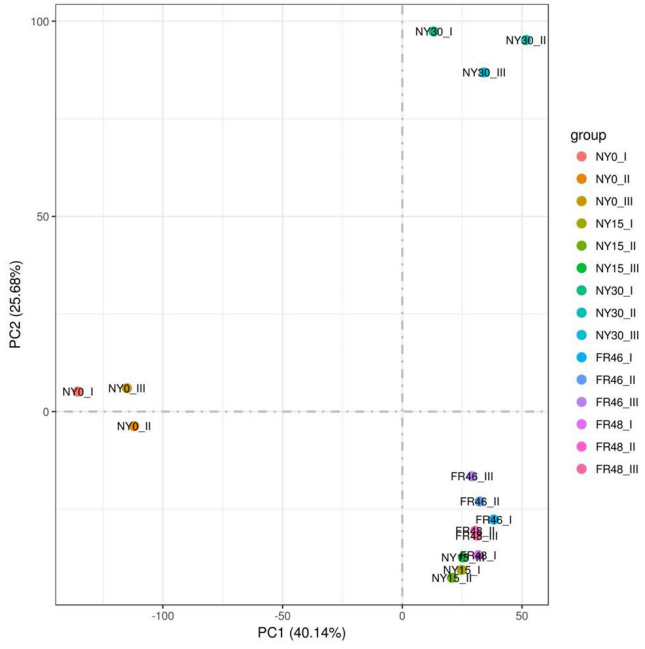


**Figure 2**. Principle component analysis (PCA) of transcriptomes of flax capsule samples.

*KEGG signaling pathway enrichment*

After collection, KEGG signaling pathway enrichment was conducted for all genes in (Schedule 1). The *P* value uniformly was calculated by -log10(padj). These genes could be divided into 17 categories. DEGs belonged to a total of 15 categories. Genes related to fatty acid metabolism accounted for 68.42%, 36.84% for scaffold103_unig ene_85 and Lus_GLEAN_10023689 genes related to the biosynthetic of unsaturated fatty acids, 36.84% for genes related to fatty acid degradation, 26.32% for genes related to alpha-linolenic acid metabolism, 15.79% for propionic acid metabolism related genes, 15.79% for genes related to fatty acid biosynthesis All the KEGG signaling pathways showed significant enrichment (all $P < 0.05$). The interaction related genes of plant-pathogens accounted for 15.79%. Genes related to carbon metabolism accounted for 5.26% and it was scaffold156_unigene_193 (Figure 3).

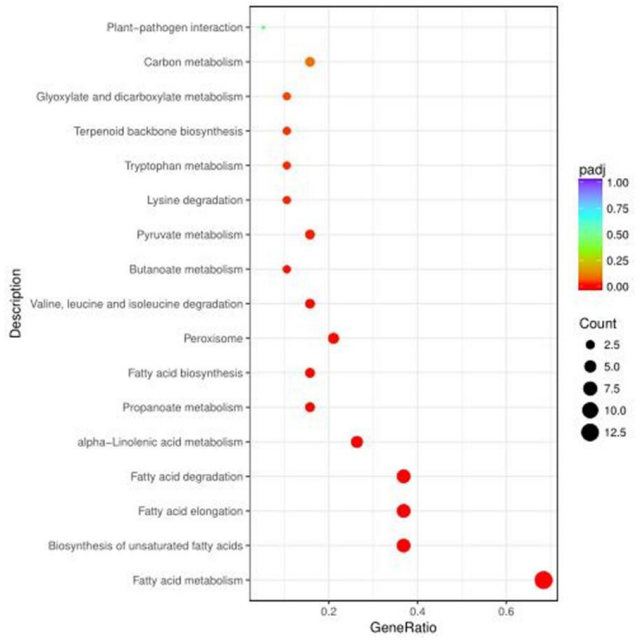


**Figure 3**. The ratio of annotated DEGs to the total number of DEGs on the KEGG signaling pathway and the corresponding cellular pathway map. Note: Padj is the *P*-value corrected by the multiple hypothesis test, Count is the number of differential genes annotated to the KEGG pathway number, Gene ratio is the ratio of the number of differential genes on the pathway number to the total number of differential genes, and the ordinate is the pathway annotation.

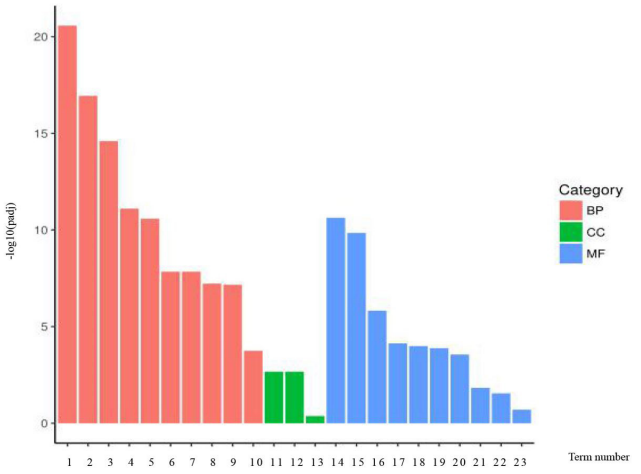


**Figure 4**. GO function enrichment map. Note: the abscissa showed the GO term number, 1. fatty acid metabolic process; 2.monocarboxylic acid metabolic process; 3.cellular lipid metabolic process; 4.fatty acid biosynthetic process; 5.monocarboxylic acid biosynthetic process; 6.organic acid biosynthetic process; 7.carboxylic acid biosynthetic process; 8.lipid biosynthetic process; 9.small molecule biosynthetic process; 10.fatty acid beta-oxidation; 11.peroxisome; 12.microbody; 13.catalytic complex; 14. transferase activity, transferring acyl groups other than amino-acyl groups; 15. transferase activity, transferring acyl groups; 16. oxidoreductase activity, acting on the CH-CH group of donors; 17. 3-hydroxyacyl-CoA dehydrogenase activity; 18. acyl-CoA oxidase activity; 19. oxidoreductase activity, acting on the CH-CH group of donors, oxygen as acceptor; 20. oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water; 21. oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor; 22.oxidoreductase activity, acting on CH-OH group of donors; 23.ligase activity

*GO annotation*

According the GO annotation, the existing 23 genes were divided into three categories of celluar component (CC), molecular function (MF) and biological processes (BP) (Figure 4) In the biological process, there were 24 functions, and 21 functions with significant differences (all $P < 0.05$). In the cell component category, there were 3 functions, and 2 cell components were significantly different (both $P < 0.05$), peroxidase body and microbody cell components, respectively. Among the types of 'molecular functions', there were 10 functions in total, and 9 functions (all $P<0.05$) were significantly related to the activity of transenzyme, acyl group. Oxidoreductase activity of CH-CH group acting on donor; 3- hydroxy acyl-CoA dehydrogenase activity; Acyl coA oxidase activity; And three functional types related to REDOX enzymes, respectively used for CH-CH group, CH-OH group, and for the oxidation-reduction of CH-OH group when NAD and NADP were the receptors (Schedule 2).

### *3.5 Cluster analysis of desaturation of fatty acids and related genes*

The metabolism of fatty acids included biosethogenesis of fatty acids (KEGG ID: ath00061), prolongation of fatty acids (KEGG ID: ath00062), desaturation of fatty acids (KEGG ID: ath01040), and degradation of fatty acids (KEGG ID: ath00071). First, through the citric acid cycle and pyruvate metabolism in the cell fluid, acetyl-acp and malonyl-acp were formed as the substrates for fatty acid synthesis. Then, acetyl-acp prolonged 16 carbon fatty acids in mitochondria, malonyl-ACP prolonged 18 carbon fatty acids in endoplasmic reticulum, and fatty acid biosynthesis (KEGG ID: ath00061) and prolongation process (KEGG ID: ath00062) were conducted. The fatty acids were then desaturated in the endoplasmic reticulum (KEGG ID: ath01040). Finally, saturated fatty acids were broken down in mitochondria (KEGG ID: ath00071) (Figure 5).

*Cluster analysis in the fatty acid desaturation pathway*

Cluster analysis was conducted on the related genes of the fatty acid desaturation pathway (KEGG ID: ath01040), and the genes with similar expression patterns were grouped together, which may share common function and signaling pathway, and these genes were divided into three categories (Figure 6). In the medium fatty acid variety (NeiYa No.7), the genes of the first and second types were up-regulated, whereas the genes of the third type were down-regulated with significant differences (all $P < 0.05$). In NY15, scaffold31_unigene_152 was significantly up-regulated, but other genes were significantly down-regulated (all $P < 0.05$). In NY30, the first type of genes were significantly

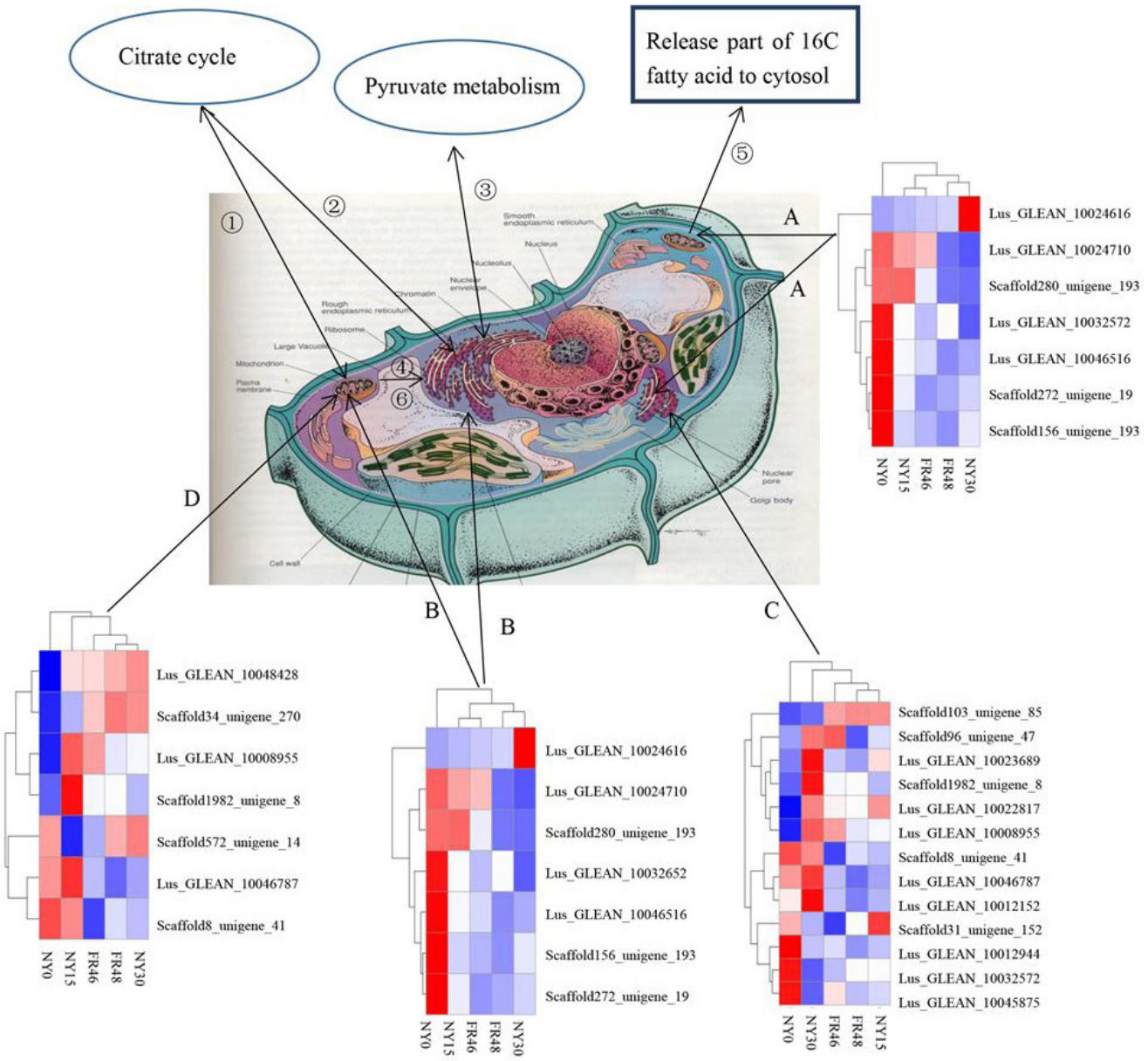


**Figure 5**. Fatty acid metabolism map. Note: ①,②,③ are the citric acid cycle and pyruvate metabolism for the biosynthesis of fatty acids to provide acetyl-ACP, malonyl-ACP and malonyl-CoA; ④ 16 carbon fatty acids are released into the endoplasmic reticulum The 18 carbon is extended and subjected to a desaturation process. ⑤ The C16 saturated fatty acid released into the cytosol is transported to the mitochondria for decomposition. A is a clustering map of genes related to fatty acid synthesis pathways in different groups, KEGG ID: ath00061; B is a clustering map of genes related to fatty acid elongation pathways in different groups, KEGG ID: ath00062; C is a cluster of genes related to fatty acid desaturation pathways in different groups. Class diagram, KEGG ID: ath00001040; D is a clustering map of genes associated with fatty acid degradation pathways in different groups, KEGG ID: ath00071. The abscissa of clustering map is the name of the cultivar group; the ordinate is the name of the gene related to fatty acid desaturation; red represents high expression; blue represents low expression.

down-regulated, the second type of genes were significantly up-regulated, and the third type of genes were significantly up-regulated except scaffold103_ unigene_85 (all $P < 0.05$), as illustrated in Figure 6.

*Cluster analysis of genes in fatty acid metabolism pathway*

Genes related to the synthesis, extension and decomposition of fatty acids occurred in mitochondria, which were divided into two categories (Figure 7). In NeiYa No.7, 0DAF, the expression levels of Lus_GLEAN_10048428, scaffold34_unigene_270 and Lus_GLEAN_10008955, were significantly down-regulated (all $P < 0.05$). The expression levels of Lus_GLEAN_10032572, Lus_GLEAN_10046516, scaffold272_unigene_19, scaffold156_ unigene_193 were significantly up-regulated (all $P < 0.05$). In NY15, Lus_GLEAN_10032572, Lus_GLEAN_10046516, scaffold272_unigene_19, scaffold156_unigene_ 193 were down-regulated, and Lus_GLEAN_10048428, scaffold34_

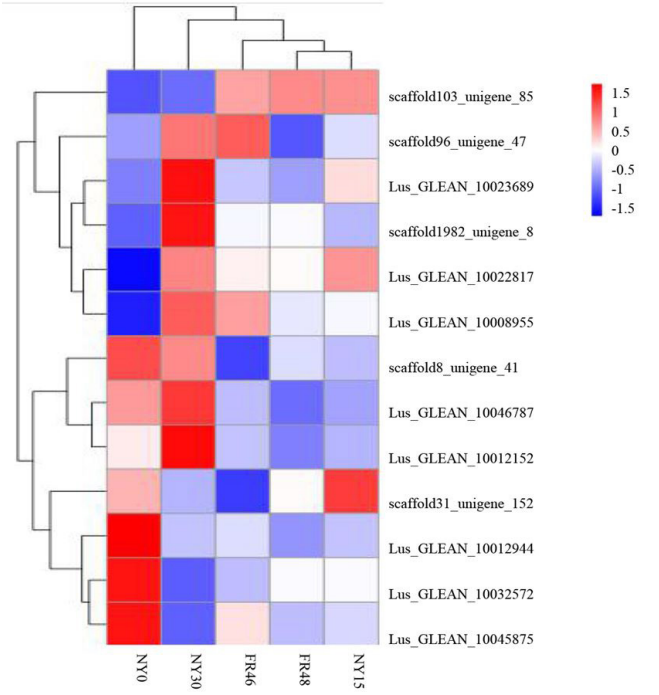


**Figure 6**. Cluster analysis of fatty acid desaturation related genes in different groups. Note: The abscissa is the name of the cultivar group; the ordinate is the name of the gene related to fatty acid desaturation; red represents high expression; blue represents low expression.

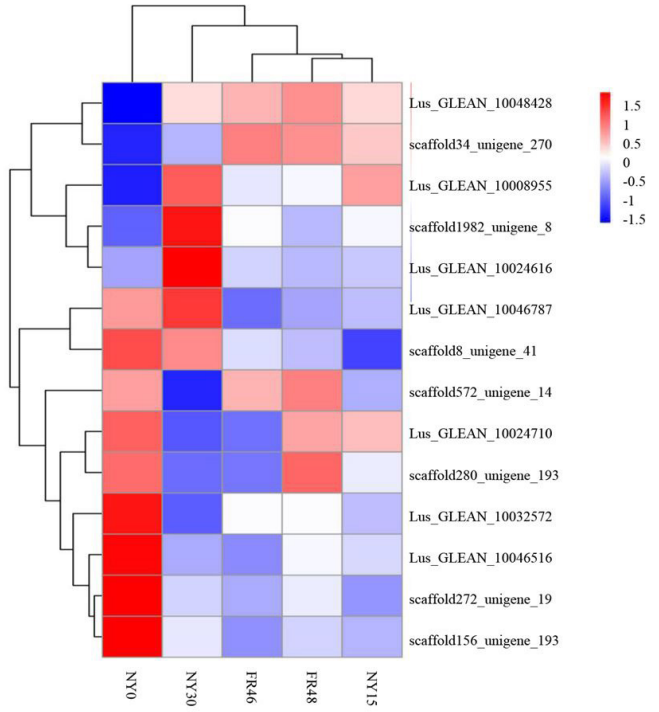


**Figure 7**. Prolongation of fatty acid synthesis and clustering of partially decomposed genes in different groups.

unigene_270 showed up-regulated expression. Because the oil content in NY30 period was the highest, and scaffold1982_unigene_8, Lus_GLEAN_10046787 and Lus_GLEAN_10008955 were genes. Scaffold34_unigene_270 and Lus_G LEAN_10008955 were distributed in the fatty acid degradation pathway (Schedule 1). In the KEGG functional analysis table, scaffold1982_unigene_8 and Lus_GLEAN_10008955 were related to fatty acid metabolism, biosynthetic of unsaturated fatty acid, fatty acid degradation, alpha-linolenic acid metabolism and peroxidase body. In NY30, Lus_GLEAN_10008955, scaffold1982_unigene_8, Lus_GLEAN_10024616, Lus_GLEAN_10046787 in the molecular functional components of GO function enrichment table (Schedule 2), scaffold1982_unigene_8 had oxidation-reductase activity and acted on CH-CH groups for donors.

For three varieties with high, medium and low linolenic contents in the same period, Lus_GLEAN_10048428 and scaffold34_unigene_270 of the first group of genes with high fatty acid content (FR48) showed a trend of high expression, whereas the others showed a trend of low expression. Only Lus_GLEAN_10008955 gene was increased in the varieties with low linolenic acid content, and distributed in the fatty acid desaturation and degradation pathways (Schedule 1). In the second type of genes, the expression of scaffold572_unigene_14 in the varieties (FR48) with high fatty acid content was increased.

In NeiYa No.7, the expression levels of scaffold572_unigene_14, Lus_GLEAN_10024710 and scaffold280_unigene_193 genes were increased. The expression levels of Lus_GLEAN_10024710 and Lus_GLEAN_10008955 genes were up-regulated in the low fatty acid content variety (FR46), located in both the biosynthetic pathway of fatty acid and the prolongation pathway of fatty acid (Schedule 2). In the KEGG signaling pathway, scaffold572_unigene_14 was related to the degradation of leucine and isoleucine, butyric acid metabolism and pyruvate metabolism, *etc*. Scaffold572_unigene_14 gene participated in the process of increasing transferase activity and transferring acyl group. In the variety (Neiya No.7) with moderate fatty acid content, both genes were present. At 15DAF, the synthesis of fatty acid should be increased (Banik et al., 2011).

### *Cluster analysis of genes related to desaturation and degradation of fatty acids in endoplasmic reticulum*

The clusters of genes related to the synthesis, extension and decomposition of fatty acids in the endoplasmic reticulum were divided into two categories (Figure 8). For 15DAF of NeiYa No.7, the expression levels of the first group of genes scaffold31_unigene_152, Lus_GLEAN_10024710, scaffold280_unigene_193 were up-regulated, whereas those of the other genes were down-regulated. The Lus_GLEAN_10024710 gene was located in the biosynthesis and extension pathway of fatty acids. In the NY30 phase, the content of fatty acid was the highest in four phases. The expression levels of scaffold8_unigene_41, Lus_GLEAN_10046787 and Lus_GLEAN_10012152 genes in the first group of genes were higher compared with that in the 15DAF. In the second category, the expression levels of Lus_GLEAN_10023689, scaffold19 82_unigene_8 and Lus_GLEAN_10024616 genes were up-regulated. Lus_GLEAN_10024616 genes were also located in the fatty acid lengthening pathway (all $P < 0.05$) (Schedule 1). In the GO functional analysis (Schedule 2), peroxisome and microbodies were involved in 18 processes.

### *3.6 Venns cluster of genes related to fatty acid desaturation among different groups*

The four groups of NY0 vs NY15, NY0 vs NY30, NY0 vs FR46, NY0 vs FR48 for the analysis of venn of DEGs (Figure 9). NY0 vs FR46 had two DEGs of scaffold8_unigene_41 and scaffold96_unigene_47. One DEG of caffold1982_unigene_8 was found in NY0 vs NY30, Lus_GLEAN_10008955 in NY0 vs NY30 and NY0 vs FR48, Lus_GLEAN_10046787 in NY0 vs FR46 and NY0 vs FR48, scaffold103_ unigene_85 in NY0 vs NY15, NY0 vs FR46 and NY0 vs FR48 groups, and Lus_GLEAN_10023689 was found in NY0 vs NY15, NY0 vs FR46 and NY0 vs NY30. Four DEGs of Lus_GLEAN_10012944, Lus_GLEAN_10045 875, Lus_GLEAN_10022817 and Lus_GLEAN_10032572 were identified in NY0 vs NY15, NY0 vs NY30, NY0 vs FR46 and NY0 vs FR48. Only three genes including Lus_GLEAN_10046787, scaffold103_unigene_85 and Lus_GLEAN_10008955 were associated with fatty acid desaturization (Schedule 1).

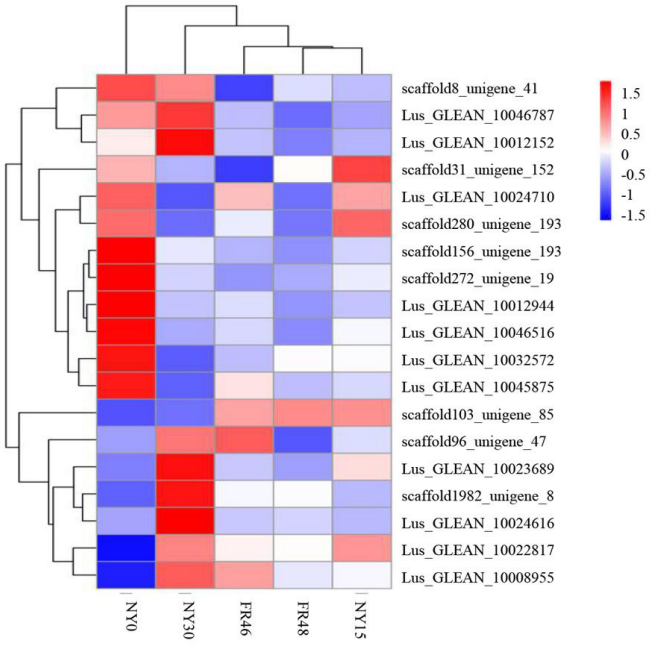


**Figure 8**. Clustering of genes related to fatty acid elongation and saturated fatty acid desaturation.

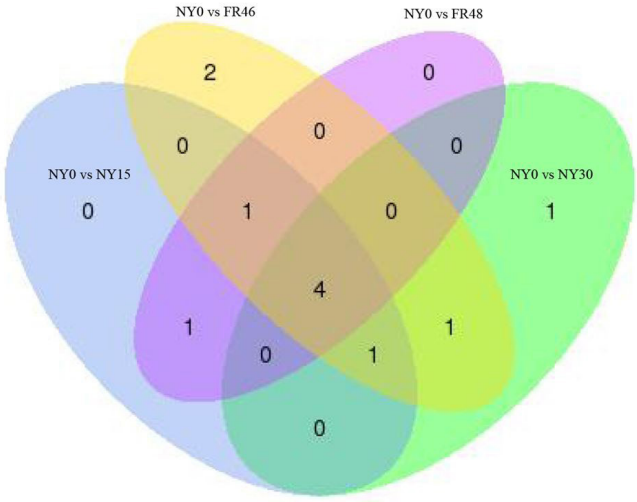


**Figure 9**. 26 DEGs related to *fads* Venn diagram. Note: Different colors represent different groups.

## 4 Discussion

In this study, RNA-seq technology was applied to determine different varieties of flax after flowering. Comparative analysis of transcriptome at 3 time points in the early, middle and late stages of capsule demonstrates the overall pattern of gene expression over the development time, follows the tendency of initial up-regulation and subsequent down-regulation, which was consistent with previous studies (Banik et al., 2011; Wang et al., 2019). DEGs of 32.7 thousand are detected in most samples on average. Transcription level may be influenced by plant physiological state, nutrition and climate (Altiner et al., 2021). However, PCA reveals the distinct replicate and variety groups among flax samples. Wang et al. (2019) has identified the difference in the transcription levels between *fad2* and *fad3* genes.

Genes related to fatty acid are more than 23 DEGs in flax. Regulating genes, such as transcriptional factors and miRNA, have not been discussed (Zhou et al., 2016; Yu et al., 2016; Yan et al., 2021).

Based on the transcriptome of flax, 23 genes related the unsaturated fatty acid synthesis and alpha-linolenic acid metabolism were selected on GO and KEGG database. More than one product and function was annotated for each DEG. The gene Lus_GLEAN_10046787 and Lus_GLEAN_10008955, Lus_GLEAN_10048428 may participate 7 processes, scaffold572_unigene_14, scaffold8_unigene_41, scaffold34_unigene_270 and scaffold1982_unigene_8 may participate 2 processes in GO annotation and in KEGG pathway of fatty acid degradation. Lus_GLEAN_10024616, Lus_GLEAN_10024710, Lus_GLEAN_10046516 and scaffold280_unigene_193 may participate 12 processes respectively, scaffold272_unigene_19 may participate 13 processes, scaffold156_unigene_193 participate 11 processes, Lus_GLEAN_10024616 participate 10 processes, Lus_GLEAN_10032572 participate 1 process in GO annotation and in KEGG pathway of fatty acid biosynthesis. In GO annotation and in KEGG pathway of biosynthesis of unsaturated fatty acids, Lus_GLEAN_10046787 may participate in 26 processes, Lus_GLEAN_10008955 in 22 processes, scaffold103_unigene_85 in 4 processes, scaffold8_unigene_41 in 2 processes, Lus_GLEAN_10032572, Lus_GLEAN_10023689 and scaffold1982_unigene_8 in 1 process, respectively. Lus_GLEAN_10024710 may participate in 12 processes in GO annotation and in KEGG pathway of fatty acid elongation. Lus_GLEAN_10024616, scaffold156_unigene_193 and Lus_GLEAN_10046516 may participate 11 processes respectively, scaffold272_unigene_19 may participate in 13 processes, Lus_GLEAN_10032572 may participate 1 process in GO annotation and in KEGG pathway of fatty acid elongation. The gene scaffold572_unigene_14 may participate 3 processes, scaffold103_unigene_85 may participate 4 processes, Lus_GLEAN_10032631 may participate 11 processes,

Lus_GLEAN_10046787 and Lus_GLEAN_10008955 may participate 23 processes respectively, Lus_GLEAN_10023689, Lus_GLEAN_10032572 and scaffold1982_unigene_8 may participate 1 process respectively, scaffold8_unigene_41and scaffold34_unigene_270 may participate 2 processes respectively, Lus_GLEAN_10019463 may participate 12 processes, Lus_GLEAN_10048428 may participate 6 processes, novel.1297 may participate 7 processes, scaffold103_unigene_85 may participate 4 processes, in GO annotation and in KEGG pathway of fatty acid metabolism, showing significant enrichment on GO annotation. Some DEGs participate in 2 or more pathways in KEGG.

Lus_GLEAN_10046787 act in fatty acid degradation (ath00071) and metabolism (ath01212) in KEGG signaling pathway. Lus_GLEAN_10024710, Lus_GLEAN_10046516 and scaffold280_unigene_193 act in Fatty acid biosynthesis (ath00061) and fatty acid elongation (ath00062) in KEGG pathway respectively. Cluster between DEGs and samples in KEGG pathway reveal the features of DEGs.

Lus_GLEAN_10008955 is involved in the synthesis of unsaturated fatty acids, the metabolism of alpha-linolenic acid and the degradation of fatty acids in peroxisome, which can increase the content of unsaturated fatty acids. Scaffold280_unigene_193 genes play a role in the biosynthetic and lengthening processes of fatty acids. Scaffold96_unigene_47 is related to the desaturation of plant fatty acids, differentially expressed between FR46 and FR48.

### Genes located in cell organells play a vital role in fatty acid metabolism

The FAS complex is first synthesized from 18:0-ACP, however, in most species, only a limited amount of stearic acid (18:0) is exported from the plastid, so the fatty acid rarely accumulates in the seed. First, some 16:0-ACP are released from FAS before being converted to 18:0-ACP, resulting in palmitic acid ester (16:0) being exported to the endoplasmic reticulum. Then, molecules extending to 18:0-ACP are effectively desaturated by the AAD1 family's matrix D9 stearyl-acp desaturating enzyme, so oleate (18:1) is preferentially exported to the endoplasmic reticulum to form monounsaturated fatty acids (MUFA). The seeds of Theobroma cacao are an exception because stearic acid (18:0) is effectively exported to ER (endoplasmic reticulum) and incorporated into TAG (triglyceride), giving cocoa butter unique solidification and melting properties (Patel et al., 1994; Griffiths et al., 1991; 1993). This part of oleate (18:1), under the action of a series of enzymes, generates phosphatidylcholine (PC), the sole substrate of FAD (Wu & Zhou, 2018). In the endoplasmic reticulum, oleoyl desaturase (FAD2) catalyzed the dehydrogenation of monounsaturated fatty acids (oleic acids) to linoleic acid, and linoleic acid was catalyzed to linolenic acid by linoleoyl desaturase (FAD3). On the plastid membrane, PC generates linoleic acid under the catalysis of FAD6 (plastid isoenzyme of FAD2), and then linolenic acid under the catalysis of FAD7 (plastid isoenzyme of FAD3) (Ohlrogge & Browse, 1995).

In plastids, C16-ACP and C18-ACP were catalyzed by fatty acid synthase (FAS). Then, in the cytoplasm, the sad gene is responsible for inserting a double bond between the 9$^{th}$ and 10$^{th}$ C atoms, causing the dehydrogenation of stearyl acyl ACP to oleo-acp, which an important gene in determining the synthesis of unsaturated fatty acids in plants (Ohlrogge & Jaworski, 1997). First, only *sad* gene was found in flax (Li, 2018). Singh et al. (1994) reported that cDNA sequences encoding sad protein have been isolated and identified in AC McDuff species (Fofana et al., 2004). Later, two parallel homologous genes, *sad1* and *sad2* genes, were found in plants with different expression in the promoter region of *sad* gene subtype (Jain et al., 1999).

*Fad3* gene is the main control gene of alpha-linolenic acid in flax seeds. *Fad3* genes compared to the substrate in Ω-3 position inserted into the double bond formation LIN, it applies to speak to the endoplasmic reticulum and only applies to (n-6) (Ohlrogge & Browse, 1995). In flax, *fad3a* and *fad3b* have been cloned successfully in varieties of Normandy, and both of these mutants are caused by point mutation, and have characteristics of early termination codon in line 593-708 (Vrinten et al., 2005). Although there are mutation points in 593-708 lines, 2%-3% of alpha-linolenic acid can be produced (Vrinten et al., 2005). *Fad3c* was extracted from a cDNA library constructed from capsules of AC McDuff with moderate linolenic acid content on day 12 and sequenced as a primer (Banik et al., 2011). Vrinten and Banik found three subunits of *fad3* in the flax genome, namely fad3a, fad3b and fad3c, and proved that fad3c was not significantly related to the accumulation of alpha-linolenic acid (Vrinten et al., 2005; Banik et al., 2011).

*Fad4* and *fad6* mainly catalyze the dehydrogenation of carbon-carbon single bonds at certain positions in the chain to form double bonds. Cytochrome b5 organism is a kind of important REDOX proteins, in the process of fatty acids to saturated, need cytochrome b5 and cytochrome b5 reductase (Patel et al., 2004), and *fad6* and cytochrome b5 REDOX reaction, can improve the conversion rate of the substrate (Huang et al., 2008). Fad7 and fad8 are located in plasmids and used to catalyze the formation of alpha-linolenic acid from linoleic acid at 16:2 (n-6) and 18:2 (n-6) positions (Browse et al., 1993; Song, 2016). *Fad5* is responsible for introducing the double bond into the 16C acyl chain of esterified galactosyl diacyl glycerol (MGD). In *arabidopsis* mutants, the ratio of 18/16c fatty acid is increased,regulatory response of *fad5* deficiency (Kunst et al., 1989). When portulaca oleracea is subject to cold stress, the expression of *fad8* gene is increased (Teixeira et al., 2010). However, the expression of *fad7* is increased with plant injury. The genes would be explored in flax resources once its identity is verified.

## Conflict of interest

None.

## Availability of data and material

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

## Funding

This work was supported by the Natural science foundation of Inner Mongolia autonomous region, Study on transcriptional expression of related genes of lenolenic acid quality in oil flax, (No. 2017MS0346); The nature conservancy (TNC) Execute sub-projects in china, Sustainable agriculture demonstration

project---Research on key technologies of sustainable arid land farming, (P104710).

## Author contributions

Shuyan Wang was the experimental designer and author, executor of this study, the completion of the first draft of the paper writing; Jun Xu finished material planting, participate in data analysis and document proof reading; Zhuo Wang was responsible of definition of intellectual content and manucscipt. Zhiwei Li and Liuxi Yi participate in statistical analysis and manuscript preparation. Lijun Yao and Xia Wang participate in experimental studies and literature research. All the authors read and agree with the final text.

## References

Abuelnaga, A. S. M., El-Razik, K. A. E. A., Ata, N. S., Hedia, R. H., Elgabry, E. A., Soliman, M. M. H., & Marie, H. S. A. W. (2021). Bacteriological assessment and multiplex-PCR test for the detection of meat adulteration of different animal species. *Food Science and Technology*, 41(Suppl. 1), 98-104. http://dx.doi.org/10.1590/fst.11520.

Altiner, D. D., Altunatmaz, S. S., Sabuncu, M., Aksu, F., & Sahan, Y. (2021). In-vitro bioaccessibility of antioxidant properties of bee pollen in Turkey. *Food Science and Technology*, 41(Suppl. 1), 133-141. http://dx.doi.org/10.1590/fst.10220.

Ander, B. P., Weber, A. R., Rampersad, P. P., Gilchrist, J. S. C., Pierce, G. N., & Lukas, A. (2004). Dietary flaxseed protects against ventricular fibrillation induced by ischemia-reperfusion in normal and hypercholesterolemic rabbits. *The Journal of Nutrition*, 134(12), 3250-3256. http://dx.doi.org/10.1093/jn/134.12.3250. PMid:15570021.

Banik, M., Duguid, S., & Cloutier, S. (2011). Transcript profiling and genen characterization of three fatty acid desaturase genes in high, moderate, and low linolenic acid genotypes of flaxes (Linum usitatissimum L.) and their role in linolenic acid accumulation. *Genome*, 54(6), 471-483. http://dx.doi.org/10.1139/g11-013. PMid:21627464.

Browse, J., McConn, M., James, D. Jr., & Miquel, M. (1993). Mutants of Arabidopsis deficient in the synthesis of α-linolenmate: biochemical and genetic characterization of the endoplasmic reticulum linoleoyl desasturase. *The Journal of Biological Chemistry*, 268(22), 16345-16351. http://dx.doi.org/10.1016/S0021-9258(19)85427-3. PMid:8102138.

Damude, H. G., & Kinney, A. J. (2008). Engineering oilseeds to produce nutritional fatty acids. *Physiologia Plantarum*, 132(1), 1-10. PMid:18251865.

Dang Z., Zhao R., & Wang M. (2010). Visual analysis of flax research from an international perspective. *Hemp Science in China*, 32(6), 305-313. [In chinese]

Dash, P. K., Cao, Y., Jailani, A. K., Gupta, P., Venglat, P., Xiang, D., Rai, R., Sharma, R., Thirunavukkarasu, N., Abdin, M. Z., Yadava, D. K., Singh, N. K., Singh, J., Selvaraj, G., Deyholos, M., Kumar, P. A., & Datla, R. (2014). Genome-wide analysis of drought induced gene expression changes in flax (*Linum usitatissimum*). *GM Crops & Food: Biotechnology in Agriculture and the Food Chain*, 5(2), 106-119. http://dx.doi.org/10.4161/gmcr.29742. PMid:25072186.

Dash, P. K., Rai, R., Mahato, A. K., Gaikwad, K., & Singh, N. K. (2017). Transcriptome landscape at different developmental stages of a drought tolerant cultivar of flax (*Linum usitatissimum*). *Frontiers in Chemistry*, 5, 82. http://dx.doi.org/10.3389/fchem.2017.00082. PMid:29170736.

Dyer, J. M., Stymne, S., Green, A. G., & Carlsson, A. S. (2008). High-value oils from plants. *The Plant Journal*, 54(4), 640-655. http://dx.doi.org/10.1111/j.1365-313X.2008.03430.x. PMid:18476869.

Fofana, B., Duguid, S., & Cloutier, S. (2004). Cloning of fatty acid biosynthetic genes β-Ketoacyl CoA synthase, fatty acid elongase, stearoyl-ACP desaturase, and fatty acid desaturase and analysis of expression in the early developmental stages of flax (*Linum usitatissimum L.*) seeds. *Plant Science*, 166(6), 1487-1496. http://dx.doi.org/10.1016/j.plantsci.2004.01.025.

Galindo-González, L., & Deyholos, M. K. (2016). RNA-seq transcriptome response of Flax (*Linum usitatissimum L.*) to the pathogenic fungus Fusarium oxysporum f. sp. lini. *Frontiers in Plant Science*, 7, 1766. http://dx.doi.org/10.3389/fpls.2016.01766. PMid:27933082.

Gorshkov, O., Mokshina, N., Gorshkov, V., Chemikosova, S., Gogolev, Y., & Gorshkov, T. (2017). Transcriptome portrait of cellulose-enriched flax fibres at advanced stage of specialization. *Plant Molecular Biology*, 93(4-5), 431-449. http://dx.doi.org/10.1007/s11103-016-0571-7. PMid:27981388.

Griffiths, G., & Harwood, J. L. (1991). The regulation of triacylglycerol biosynthesis in cocoa (Theobroma cacao) L. *Planta*, 184, 279-284. http://dx.doi.org/10.1007/BF00197958. PMid:24194081.

Griffiths, G., Walsh, M. C., & Harwood, J. L. (1993). Acyl-thioesterase activity in developing seeds of cocoa. *Phytochemistry*, 32(6), 1403-1405. http://dx.doi.org/10.1016/0031-9422(93)85146-I.

Huang, B. Y., Zhang. X. Y., Miao, L. J., Yan, M., Hai, Y., Yi, M. L., Xu, J., & Chen, Z. K. (2008). Transformation of peanut FAD2 gene RNAi vector and analysis of transgenic seed fatty acid. *Chinese Journal of Oil Crops*, 30, 290-293. [In Chinese]

Jain, R. K., Thompson, R. G., Taylor, D. C., MacKenzie, S. L., Mchughen, A., Rowland, G. G., Tenaschuk, D., & Coffey, M. (1999). Isolation and characterization of two promoters from linseed for genetic engineering. *Crop Science*, 49(12), 1867-1878. http://dx.doi.org/10.2135/cropsci1999.3961696x.

Kunst, L., Browse, J., & Somerville, C. (1989). A mutant of Arabidopsis deficient in desaturation of palmitic acid in leaf lipids. *Plant Physiology*, 90(3), 943-947. http://dx.doi.org/10.1104/pp.90.3.943. PMid:16666902.

Li, M., Dong, X. J., Peng, J. Q., Xu, W., Ren, R., Liu, J., Cao, F., & Liu, Z. (2016). De novo transcriptome sequencing and gene expression analysis reveal potential mechanisms of seed abortion in dove tree (Davidia involucrata Baill.). *BMC Plant Biology*, 16(1), 82. http://dx.doi.org/10.1186/s12870-016-0772-x. PMid:27068221.

Li, W. J. (2018). Til Δ 9 stearyl ACP dehydrogenase (SAD2) gene expression analysis. *Journal of Gansu Agricultural Science and Technology*, (10): 13-16. [In Chinese]

Ma, L. X., Wang, S. Y., Wang, Z., Han, B., & Mi, H. (2017). Agronomic traits and quality evaluation of flax varieties from china and abroad. *Journal of North Agricultural Sciences*, 45(3), 19-22. [In Chinese]

Ohlrogge, J. B., & Jaworski, J. G. (1997). Regulation of fatty acid synthesis. *Annual Review of Plant Physiology and Plant Molecular Biology*, 48(1), 109-136. http://dx.doi.org/10.1146/annurev.arplant.48.1.109. PMid:15012259.

Ohlrogge, J., & Browse, J. (1995). Lipids biosynthesis. *The Plant Cell*, 7(7), 957-970. PMid:7640528.

Oomah, B. D. (2001). Flaxseed as a functional food source. *Journal of the Science of Food and Agriculture*, 81(9), 889-894. http://dx.doi.org/10.1002/jsfa.898.

Patel, M., Jung, S., Moore, K., Powell, G., Ainsworth, C., & Abbott, A. (2004). High-oleate peanut mutants result from a mite insertion into

the FAD2 gene. *Theoretical and Applied Genetics*, 108(8), 1492-1502. http://dx.doi.org/10.1007/s00122-004-1590-3. PMid:14968307.

Patel, V. K., Shanklin, J., & rtek, D. B. (1994). Changes in fatty-acid composition, stearoyl-acyl carrier protein desaturase expression in developing Theobroma cacao L. Embryos. *Planta*, 193(1), 83-88. http://dx.doi.org/10.1007/BF00191610.

Rowland, G. G., Hormis, Y. A., & Rashid, K. Y. (2002). CDC bethune flax. *Canadian Journal of Plant Science*, 82(1), 101-102. http://dx.doi.org/10.4141/P01-066.

Singh, S., Mckinney, S., & Green, A. (1994). Sequence of a cDNA from Linum usitatissimum encoding the Stearoyl-Acyl carrier protein desaturase. *Plant Physiology*, 104(3), 1075. http://dx.doi.org/10.1104/pp.104.3.1075. PMid:7513084.

Song, S. X. (2016). *Cloning and expression analysis of Paeonia ostii fatty acid desaturatase gene PoFAD_2* (Dissertation). Shandong Agricultural University, Shangdong Province.

Teixeira, M. C., Carvalho, L. S., & Brodelius, M. (2010). ω-3 fatty acid desaturases genes isolated from purslane (Portulaca oleracea L.): expression in differen tissues and response to cold and wound stress. *Journal of Agricultural and Food Chemistry*, 58(3), 1870-1877. http://dx.doi.org/10.1021/jf902684v. PMid:20070085.

Thambugala, D., & Cloutier, S. (2014). Fatty acid composition and desaturase gene expression in flax (Linum usitatissimum L.). *Journal of Applied Genetics*, 55(4), 423-432. http://dx.doi.org/10.1007/s13353-014-0222-0. PMid:24871199.

Thambugala, D., Duguid, S., Loewen, E., Rowland, G., Booker, H., You, F. M., & Cloutier, S. (2013). Genetic variation of six desaturase genes in flax and their impact on fatty acid composition. *Theoretical and Applied Genetics*, 126(10), 2627-2641. http://dx.doi.org/10.1007/s00122-013-2161-2. PMid:23928861.

Vrinten, P., Hu, Z., Munchinsky, M. A., Rowland, G., & Qiu, X. (2005). Two FAD3 desaturase genes control the level of linolenic acid in flaxes seed. *Plant Physiology*, 139(1), 79-87. http://dx.doi.org/10.1104/pp.105.064451. PMid:16113219.

Wang, S. Y., Han, B., Zhou, S. M., & Xu, J. (2016). Fatty acid content of flax and differential expression of related genes. *Chinese Journal of Oil Crops*, 38(6), 771-777. [In Chinese].

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63. http://dx.doi.org/10.1038/nrg2484. PMid:19015660.

Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., Yang, L., Hawkins, S., Neutelings, G., Datla, R., Lambert, G., Galbraith, D. W., Grassa, C. J., Geraldes, A., Cronk, Q. C., Cullis, C., Dash, P. K., Kumar, P. A., Cloutier, S., Sharpe, A. G., Wong, G. K. S., Wang, J., & Deyholos, M. K. (2012). The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads. *The Plant Journal*, 72(3), 461-473. http://dx.doi.org/10.1111/j.1365-313X.2012.05093.x. PMid:22757964.

Wang, Z., Wang, S. Y., Han, B., Xu, J., Yu, D. Y., & Zhang, G. H. (2019). Analysis of the expression of fatty acid desaturase gene in oil flax. *Molecular Plant Breeding*, 4, 28. [In Chinese]

Wiesenfeld, P. W., Babu, U. S., Collins, T. F. X., Sprando, R., O'Donnell, M. W., Flynn, T. J., Black, T., & Olejnik, N. (2003). Flaxseed increased α-linolenic and eicosapentaenoic acid and decreased arachidonic acid in serum and tissues of rat dams and offspring. *Food and Chemical Toxicology*, 41(6), 841-855. http://dx.doi.org/10.1016/S0278-6915(03)00035-8. PMid:12738189.

Wu, C. T., & Zhou, B. L. (2018). Characterization of a sterile dwarf mutant and the cloning of zeaxanthin epoxidase in Asian cotton (Gossypium arboreum L.). *Plant Growth Regulation*, 85(1), 57-72. http://dx.doi.org/10.1007/s10725-018-0373-9.

Yan, L. P., Luo, W. J., Xiao, X. H., Luo, T., & Liu, H. (2021). A meta-analysis: association between Beta-2 adrenergic receptor Arg16Gly polymorphism and asthma in China. *Food Science and Technology*, 41(3), 564-569. http://dx.doi.org/10.1590/fst.18520.

Yu, Y., Wu, G. W., Yuan, H. M., Cheng, L., Zhao, D. S., Huang, W. G., Zhang, S. Q., Zhang, L. G., Chen, H. Y., Zhang, J., & Guan, F. Z. (2016). Identification and characterization of miRNAs and targets in flax (Linum usitatissimum) under saline, alkaline, and saline-alkaline stresses. *BMC Plant Biology*, 16(1), 124. http://dx.doi.org/10.1186/s12870-016-0808-2. PMid:27234464.

Zhang, N., & Deyholos, M. K. (2016). RNASeq analysis of the shoot apex of flax (*Linum usitatissimum*) to identify phloem fiber specification genes. *Frontiers in Plant Science*, 7, 950. http://dx.doi.org/10.3389/fpls.2016.00950. PMid:27446177.

Zhang, X., Xu, L. Q., Wang, Q. F., Li, Q. P., Wang, S. F., & Han, Y. Z. (2018). Research progress of flax for different uses. *Journal of Agricultural Sciences of Northeast China*, 43(5), 16-20. [In Chinese]

Zhou, B., Fan, P. Z., Li, Y. H., Yan, H. F., & Xu, Q. J. (2016). Exploring miRNAs involved in blue/UV-A light response in Brassica rapa reveals special regulatory mode during seedling development. *BMC Plant Biology*, 16(1), 111. http://dx.doi.org/10.1186/s12870-016-0799-z. PMid:27160188.

## Supplementary Material

Supplementary material accompanies this paper.

**Schedule 1.** KEGG function-related pathway and GO-related functional gene statistics.

**Schedule 2.** significant enrichment of related genes on GO annotation.

This material is available as part of the online article from https://www.scielo.br/j/CTA