Transcriptional Profiles of Long Non-coding RNAs Involved in Fatty Acid Biosynthesis at Filling Stage of Soybean Pods

Bohan Ma

Crop germplasm Resources Innovation Laboratory, College of Life Sciences, Jilin Agricultural University, Changchun 130118

Yue Li

Crop germplasm Resources Innovation Laboratory, College of Life Sciences, Jilin Agricultural University, Changchun 130118

Mohamed Khalifa

Plant Biotechnology Centre, College of Agronomy, Jilin Agricultural University, Jilin, Changchun 130118

Meng Teng

Crop germplasm Resources Innovation Laboratory, College of Life Sciences, Jilin Agricultural University, Changchun 130118

Aijing Zhang

Crop germplasm Resources Innovation Laboratory, College of Life Sciences, Jilin Agricultural University, Changchun 130118

Qiuzhu Zhao

Crop germplasm Resources Innovation Laboratory, College of Life Sciences, Jilin Agricultural University, Changchun 130118

Abraham Lamboro

Plant Biotechnology Centre, College of Agronomy, Jilin Agricultural University, Jilin, Changchun 130118

Dan Yao (✉ jlauyd1977@163.com)

Crop germplasm Resources Innovation Laboratory, College of Life Sciences, Jilin Agricultural University, Changchun 130118

Jun Zhang

Plant Biotechnology Centre, College of Agronomy, Jilin Agricultural University, Jilin, Changchun 130118

Research Article

Keywords: soybean (*Glycine max* L.), high-throughput sequencing technology, lncRNAs, mRNAs, fatty acid synthesis

Posted Date: September 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-853545/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Long non-coding RNAs (lncRNAs) are non-coding RNAs of more than 200 nucleotides. To date, the roles of lncRNAs in soybean fatty acid synthesis have not been fully studied. Here, the low-linolenic acid 'mutant 72' (MT72) and the wild-type control 'JiNong 18'(JN18) were used. Based on the previously published data on lncRNAs related RNA-seq in young pods of soybean 30 d to 40 d after flowering, lncRNAs and mRNAs from soybean pods 50 d after flowering were identified using high-throughput sequencing. The possible target genes of lncRNAs were predicted, and the functions related to fatty acid synthesis were analyzed using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG).

Results: A total of 10,500 lncRNAs and 54,370 mRNAs were identified, and 115 possible target genes of 1,805 differentially expressed lncRNAs were found to be involved in fatty acid synthesis. A network of lncRNAs and mRNAs was constructed, and a total of 604 lncRNAs and 1,484 mRNAs had regulatory relationships. Among them, 115 target genes of 77 lncRNAs were directly or indirectly involved in fatty acid biosynthesis.

Conclusions: The function related to fatty acid synthesis was predicted by differential expression of the target gene mRNAs interacting with lncRNAs. In conclusion, our results provide a theoretical basis for studies on fatty acid synthesis of lncRNAs in soybean.

Background

Soybean oil is one of the most widely consumed edible vegetable oils [1]. It pre-dominantly comprises fatty acids and glycerol. The main components of soybean oil are fatty acids, accounting for more than 90% of the total fat and oil contents. The main components of fatty acids are oleic acid, linoleic acid, linolenic acid, palm oil, and stearic acid. The first three constitute more than 80% of the total fats and are unsaturated fatty acids. Of these, linoleic acid is the most abundant, followed by oleic acid, and finally by linolenic acid, which accounts for approximately 9% of the total fat content. Palmitic and stearic acid are saturated fatty acids that account for 11% and 3%, respectively, of the total fat content. The composition of fatty acids is closely related to the quality and nutritional value of vegetable oil and food. Therefore, effective measures to improve the quality of oil crops include regulating the fatty acid metabolism pathway, increasing the fat content, and modifying the fatty acid composition of plant seeds via genetic engineering. These enable us to obtain soybean oil with a higher nutritional value and wider industrial application [2].

Non-coding RNAs (ncRNAs) are naturally occurring non-protein-coding RNA molecules [3]. With the development of high-throughput sequencing technology in recent years, more and more novel ncRNAs have begun to gradually emerge, and their regulatory mechanism is currently one of the most researched topics in the field of life science. In recent years, circular RNAs (circRNAs), long non-coding RNAs (lncRNAs), microRNAs (miRNAs) have been widely studied. LncRNAs are a class of ncRNAs that are

longer than 200 nt, have a promoter substructure and polyA tail, and show tissue and space-time specificity [4]. At present, a large number of lncRNAs has been identified in soybean (*Glycine max* L.) [5], Arabidopsis (*Arabidopsis thaliana* L.) [6], wheat (*Triticum aestivum* L.) [7], rice (*Oryza sativa* L.) [8], maize (*Zea mays* L.) [9], and tomato (*Solanum lycopersicum* L.) [10], among others. These RNA molecules have been increasingly identified in plants, and studies have shown that they play important roles in plant growth and development, in response to abiotic and biological stresses, and in fatty acid synthesis [11,12]. For example, Ma et al. found that during seed development at 15 and 35 d after anthesis, the MSTRG.13500.1 and MSTRG.9304.1 lncRNAs and their *cis*-acting target genes may play key roles in peanut (*Arachis hypogaea* L.) seed development [13]. Garima et al. identified and analyzed some lncRNAs in *Brassica juncea* (Indian mustard), and found that they played a potential role in the response to high temperature and drought stress [14]. Most importantly, Yin et al. identified and analyzed lncRNAs in two peony varieties with different fatty acid contents and predicted that a total of 39 lncRNAs are related to lipid metabolism [15]. These results laid a foundation for the study of fatty acid synthesis of lncRNAs in soybean of oil crops.

However, no studies are available in the literature on the involvement of lncRNAs in the synthesis of soybean oil. In this study, lncRNAs and protein-coding target mRNAs were identified by analyzing the expression of different transcripts in soybean pods at the filling stage (50 d after soybean flowering). Furthermore, the function of the potential protein-coding target genes of the lncRNAs was predicted. Combined with previous data, this study focused on the analysis of protein-coding RNAs related to fatty acid synthesis in soybean pods [16]. This study provides not only new information for the identification of soybean ncRNAs, but also a theoretical basis for the functional research of lncRNAs in soybean and other oilseed-bearing crops.

Results

Identification and expression patterns of lncRNAs at 50d after flowering in soybean

A number of lncRNAs were previously identified in young soybean pods 30-40 d after flowering [16]. On the basis of this information, additional lncRNAs were further identified in soybean pods 50 d after flowering. Six cDNA libraries were constructed including 'MT72' (libraries MT_50d) and 'JN18' (libraries JN_50d), both 50 d after flowering (Additional file 1: Table S1). A total of 10,500 lncRNAs and 54,370 target mRNAs were identified, and 7,520 lncRNAs were identified in MT_50d as well as JN_50d, accounting for 71.6% of the total (Fig. 1a, Additional file 2: Table S2). Then, to further explore the transcription and genetic characteristics of the identified lncRNAs, the mRNAs encoded by soybean proteins were compared and analyzed. Based on the location of the corresponding protein-coding genes, it was observed that 5,129 (48.8%) lncRNAs were mainly produced in intron regions, accounting for the largest number, whereas 196 (1.9%) antisense lncRNAs were detected, accounting for the smallest number (Fig. 1b). By comparing the number of exons, it was found that most lncRNAs had fewer exons

than mRNAs. In particular, 8,527 (81.2%) lncRNAs contained two exons, accounting for the highest proportion. However, the number of mRNAs containing exons was relatively average, and most mRNAs contained approximately two exons (Fig. 1c). In addition, the length and chromosome distribution of the two RNAs types were compared. When the length distribution of lncRNAs was less than 1kb, the proportion of lncRNAs was higher than that of mRNAs, and it subsequently decreased (Fig. 1d). Chromosome distribution results showed that different types of lncRNAs were also associated with different chromosomal regions. Denser lines indicate an increased lncRNA distribution (Fig. 1e-f).

Differentially expressed lncRNAs and mRNAs in five groups at different stages after soybean flowering

Previously published data on the differential expression of lncRNAs and mRNAs in soybean 30 d and 40 d after flowering were combined with the newly identified data 50 d after flowering to conduct a comprehensive analysis. The criteria for identifying differentially expressed lncRNAs and mRNAs were $|\log_2(\text{fold change})| \text{ value} > 1$ and a statistically significant p value of ≤ 0.05 . In total, 1,805 differentially expressed lncRNAs were identified in five different comparison groups (Additional file 3: Table S3). Among them, 336, 489, 44, 485 and 451 were identified in JN_50d versus JN_30d ('JN18' 50 d after flowering versus 'JN18' 30 d after flowering), JN_50d versus JN_40d ('JN18' 50 d after flowering versus 'JN18' 40 d after flowering), MT_50d versus JN_50d ('MT72' 50 d after flowering versus 'JN18' 50 d after flowering), MT_50d versus MT_30d ('MT72' 50 d after flowering versus 'MT72' 30 d after flowering), and MT_50d versus MT_40d ('MT72' 50 d after flowering versus 'MT72' 40 d after flowering) (Fig. 2a). Additionally, among the 1,805 differentially expressed lncRNAs identified in the five groups, 1,017 were upregulated and 788 were down-regulated (Fig. 2b). Volcano maps were drawn to further analyze their overall distribution. Most lncRNAs presented different multiples of differences (Fig. 3). Then a systematic cluster analysis was performed for all 1,805 differentially expressed lncRNAs (Fig. 4). Most of them had specific expression patterns in different materials at different stages.

Also, 39,313 target mRNAs were found to be differentially expressed in five comparison groups, among them, 11,127 were up-regulated and 28,186 were down-regulated (Fig. 5a-b, Additional file 4: Table S4). Cluster analysis was conducted on the 39,313 differentially expressed target mRNAs, and it was found that most mRNAs showed specific expression patterns in the same period in different materials (Fig. 6).

Functional analysis of differentially expressed genes

To further explore the potential functions of lncRNAs, GO and KEGG analysis were conducted on the potential target mRNAs of differentially expressed lncRNAs. A total of 322 target mRNAs were found to be related to fatty acid synthesis (Additional file 5: Table S5). The GO term indicated that the main functions of the target mRNAs of differentially expressed lncRNAs included oxidation-reduction process (GO:0055114), integral component of membrane (GO:0016021), ATP binding (GO:0005524), metal ion

binding (GO:0046872) (Fig. 7). More importantly, the 20 most significant GO terms showed that the functions related to fatty acid synthesis of the target mRNAs included fatty acid biosynthetic process (GO:0006633), unsaturated fatty acid biosynthetic process (GO:0006636), and long-chain fatty-acyl-CoA metabolic process (GO:0035336). In particular, the XM_003536636.3, XM_003521793.3, XM_003539327.2 and XM_003551789.3 genes have the potential functions related to fatty acid biosynthetic process (GO:0006633). The XM_003534019.3 gene has the potential functions related to fatty acid transporter (GO:0015908) and fatty acid transporter activity (GO:0015245). The XM_003540190.3 gene has the potential functions related to fatty-acyl-CoA reductase (alcohol-forming) activity (GO:0080019), lipid metabolic process (GO:0006629) and long-chain fatty-acyl-CoA metabolic process (GO:0035336).

Subsequently, KEGG analysis showed that the 20 most significant metabolic pathways in each comparison group mainly included photosynthesis (ko00195), carbon fixation in photosynthetic organisms (ko00710), tryptophan metabolism (ko00380), and plant hormone signal transduction (ko04075) (Fig. 8). Among them, six metabolic pathways may be related to soybean fatty acid synthesis, including fatty acid degradation (ko00071), fatty acid elongation (ko00062), fatty acid biosynthesis (ko00061), biosynthesis of unsaturated fatty acids (ko01040), alpha-Linolenic acid metabolism (ko00592), and linoleic acid metabolism (ko00591) (Fig. 9). This indicates that some mRNA-encoding proteins related to lncRNAs are likely to regulate lipid synthesis by influencing the metabolic pathways of fatty acids in soybean. Through KEGG pathway analysis, it was shown that the XM_003552194.3 and XM_003544690.3 target genes may be involved in fatty acid biosynthesis (ko00061) and degradation (ko00071). The XM_003518957.3 gene has the potential functions related to alpha-Linolenic acid metabolism (ko00592).

Validation of differentially expressed lncRNAs and target genes

To further verify the RNA-seq data, 9 differentially expressed lncRNAs and 15 target mRNAs were randomly selected using qRT-PCR. The results of 7 lncRNAs and 12 target mRNAs were consistent with high-throughput sequencing data (Fig. 10, Additional file 6: Table S6). For example, the expression patterns of lncRNAs MSTRG. 11445.3 and MSTRG. 5822.1 were up-regulated in qPCR, confirming the RNA-seq results. The expression patterns of target mRNAs XM_003552194.3 and XM_003544690.3 showed a downregulation trend, which also confirmed the RNA-Seq results. This indicates that using RNA-seq to detect the expression of lncRNAs is highly reliable.

Construction of lncRNAs and mRNAs network, and prediction analysis of lncRNAs functions related to fatty acid synthesis

Genes or proteins in the same expression module can be modulated together. Therefore, using computational construction of coding and non-coding genes, it would be possible to obtain co-expression networks to deduce the potential biological functions of lncRNAs. The targets of 1,805 differentially expressed lncRNAs were predicted. A network was constructed for each comparison using the identified differentially expressed genes. In total, 604 lncRNAs and 1,484 mRNAs were found to be related (Additional file 7: Table S7). The lncRNAs-mRNAs network was constructed using Cytoscape (version 3.7.0, [http:// www. cytoscape. org/](http://www.cytoscape.org/)) (Fig. 11).

GO and KEGG pathway analysis were performed for the target mRNAs of the lncRNAs. It was found that a total of 77 lncRNAs may regulate 115 target genes to participate in fatty acid synthesis and, therefore, they have the potential to influence oil synthesis in soybean (Additional file 8: Table S8). In particular, XM_003536636.3, NM_001250756.1, NM_001248321.1, and NM_001251763.1 may have potential functions including fatty acid biosynthetic process (GO:0006633), lipid metabolic process (GO:0006629), and fatty acid metabolic process (GO:0006631). In addition, XM_006588443.1, NM_001251413.1, and XM_003524392.3 may also be involved in biosynthesis of unsaturated fatty acids (ko01040), fatty acid biosynthesis (ko00061), and fatty acid degradation (ko00071).

Discussion

In recent years, identifying the species and functions of ncRNAs in plants and elucidating their biosynthetic pathways have become basic scientific challenges in this field of research. The rapid development of high-throughput sequencing technology and the progress of bioinformatics introduced revolutionary methods to predict new ncRNAs in plants. At present, lncRNAs are attracting increasing attention in various research fields. Although the research on plant lncRNAs is obviously lagging behind compared with that on animal lncRNAs, the regulatory role of lncRNAs in plants is very important. To date, a large number of lncRNAs have been identified in a variety of plants species. Here, lncRNAs in pods of the low-linolenic acid mutant 'MT72' and wild-type soybean 'JN18' were systematically identified. Data related to lncRNAs in young pods 30 and 40 d after flowering have been previously for two soybean groups [16]. A total of 39,324 lncRNAs were identified, of which 561 were differentially expressed. On this basis, 10,500 lncRNAs were further identified in soybean pods 50 d after flowering. Overall, 49,824 lncRNAs were identified in the three stages of soybean bulging. To compare these results with the previous ones, five groups were established, as follows: JN_50d versus JN_30d, JN_50d versus JN_40d, MT_50d versus JN_50d, MT_50d versus MT_30d, and MT_50d versus MT_40d. The results obtained from lncRNAs at 50 d after flowering were analyzed. In the five comparison groups, a total of 1,805 differentially expressed lncRNAs were identified, of which 1,017 were up-regulated and 788 were down-regulated.

Prior to this study, several lncRNAs were identified and their functions predicted in plants. In a study by Zhang et al., two soybean genotypes with different levels of phosphorus deficiency were sequenced and 4,166 lncRNAs were identified, of which 525 were differentially expressed. A large number of differentially expressed lncRNAs were found to be involved in phosphate-related biological functions [17]. In a study by

Ma et al., the expression profiles of lncRNAs in the seeds of two peanut recombinant inbred lines (RIL8) were studied 15 and 35 d after flowering. 11 lncRNAs and their *cis*-acting target genes may play a key role in peanut seed development [18]. In another study, maize samples from three different tissues at four developmental stages were analyzed and 3,488 lncRNAs were identified. Through the function prediction of target genes, it was found that the targets of many lncRNAs might be related to drought stress [19]. These reports revealed that lncRNAs play important roles in plant growth and development, fruit ripening, abiotic stress response, and other biological processes. However, the role of lncRNAs in fatty acid synthesis in plants has been rarely reported. In a study by Yin et al. (2018) the expression profiles of lncRNAs in two different varieties of peony seeds were analyzed and compared and it was predicted that 39 target genes of lncRNAs regulate fatty acid synthesis and lipid metabolism [15]. In another study by Shen et al. (2018) genome-wide identification of lncRNAs—especially of those related to lipid metabolism—was conducted in *Brassica napus*, and it was predicted that some lncRNAs regulate lipid biosynthesis [20]. Furthermore, in a study by Chen et al. (2021) the transcriptome data of several *Arabidopsis* strains under long-term and short-term drought treatment were compared and it was revealed that some lncRNAs might be related to fatty acid synthesis pathway through function prediction of target genes [21]. However, the function of lncRNAs in soybean fatty acid synthesis has not been reported yet. Based on previous data, the present study further analyzed the lncRNAs related to fatty acid synthesis and possible target mRNA genes in the soybean pods of mutant 'MT72' and wild-type 'JN18' 50 d after flowering. The results suggested that a total of 322 target mRNAs are directly or indirectly related to soybean fatty acid synthesis.

Most of these mRNAs are key enzyme genes involved in the fatty acid synthesis. For example, fatty acid desaturases (FADs) are key enzymes in soybean fatty acids synthesis [22]. At present, an increasing number of studies have been conducted on the role of the FAD family genes in vegetable lipid synthesis. A total of seven mRNAs were identified in this family: *GmFAD2-1*, *GmFAD3a*, *GmFAD8*, *GmFAD8-1*, *GmFAD3b*, *GmFAD7-1*, *GmFAD3-2b*, and *GmFAD7*, were identified. These genes may participate in the biological processes such as soybean fatty acid biosynthetic process and lipid metabolic process. In addition, *GmACCB-1* was also identified. ACCase is an enzyme involved in the conversion of acetyl-CoA to malonyl-CoA, which is the first reaction occurring in fatty acid biosynthesis [23]. And, *GmLOX7*, *GmLOX1* and *GmLOX10*, were identified. Soybean lipoxygenase pathway is one of the oxidation pathways of soybean fatty acids [24].

To investigate the *cis*-acting function of lncRNAs, the potential functions of all 1,805 differentially expressed lncRNAs and protein-coding mRNA genes were analyzed. GO and KEGG analysis showed that these genes were mainly related to oxidation-reduction process, integral component of membrane and metal ion binding. These results indicate that many lncRNAs regulate gene expression through *cis*-action, which is consistent with previous studies [25,26]. Here, a network of lncRNAs and mRNAs was constructed, which showed that a total of 604 lncRNAs and 1,484 mRNAs had regulatory relationships. The mRNAs related to fatty acid synthesis were analyzed, and 115 target genes of 77 lncRNAs were found to be directly or indirectly involved in this biosynthesis. In most cases, a single lncRNA regulates the expression of multiple mRNAs. For example, the lncRNA MSTRG. 5008.1 regulated two genes,

NM_001251413.1 and XM_006588443.1, which may be two alleles in the FAD family. In addition, a single mRNA is regulated by multiple lncRNAs. For example, NM_001248245.2 was co-regulated by lncRNA MSTRG. 25520.1 and lncRNA MSTRG. 25534.2, and it may be involved in lipid metabolism.

Compare that to previous reports, Chen et al. identified only a few lncRNAs involved in the regulation of the fatty acid synthesis pathway in several *Arabidopsis* strains under different drought treatments [21]. For example, At4G13180 is regulated by seven lncRNAs, but five of them are up-regulated and thus promote the fatty acids synthesis. Probably because *Arabidopsis thaliana* is a model plant, limited results were obtained for lncRNAs associated with fatty acid synthesis. Another study found that 39 mRNAs in the seeds of two different peony varieties may be related to lipid metabolism [15]. The results of the present study are larger than the data obtained in previous studies, probably because soybean is one of the most important oil crops with high oil content. In addition, this study focused on the fatty acid synthesis pathway in the lipid synthesis pathway. In the functional analysis, several processes involved in fatty acid synthesis were taken into consideration, including nearly 20 metabolic pathways, such as fatty acid degradation, linolenic acid metabolism, and unsaturated fatty acid biosynthesis. This is the largest study on lncRNAs associated with fatty acid synthesis in soybean. This study expanded the known database of soybean lncRNAs and provided important information for future studies on the function of lncRNAs in soybean pods. Understanding the role of lncRNAs in regulating of fatty acid synthesis in soybean seeds is of great value for breeding soybean varieties with high oleic acid content.

Materials And Methods

Plant materials

In this study, the soybean pods of the fourth-generation low-linolenic acid 'mutant 72' (MT72) at the filling stage (50 d after flowering) were used as the experimental material and those of 'JiNong 18'(JN18) were used as the control (Fig. 12). The cultivar JN18 (Ji Shen bean 2006) was bred at the biotechnology center of Jilin Agricultural University, Changchun, China. In 2015, the low-linolenic acid mutant 'MT72' was obtained after soybean was exposed to 0.5% ethylmethanesul-fonate for 6 h. Gas chromatographic results showed that 50 d after flowering, the level of unsaturated fatty acid in 'JN18' were 0.175% linolenic acid, 43.5% oleic acid, and 29.4% linoleic acid (Additional file 9: Table S9), whereas those of saturated fatty acids were 15.8% palmitic acid and 7.33% stearic acid (GC-7890, Agilent, Santa Clara, CA, USA) (Fig. 13). The samples were collected from the experimental and control groups 50 d after soybean flowering. The pods were wrapped in aluminum foil, quickly frozen using liquid nitrogen and then stored at -80 °C.

LncRNAs library construction and RNA sequencing

Total RNA was isolated and purified using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's procedure. RNA integrity was assessed by Agilent 2100 with RIN number >7.0.

Approximately 5 µg of the total RNA was used to deplete ribosomal RNA using Ribo-Zero Gold Kits (Illumina, San Diego, USA). After the removal of ribosomal RNAs, the remaining RNAs were fragmented into small pieces using divalent cations under high temperature. Subsequently, the cleaved RNA fragments were reverse-transcribed to create cDNA. The average insert size for the final cDNA library was 300 bp (±50 bp). Finally, 150 bp paired-end sequencing was performed using an Illumina Hiseq 4000 (LC Bio, China) following the manufacturer's recommended protocols.

Identification of lncRNAs and differential expression analysis

First, reads that contained adaptor contamination and low quality and undetermined bases were removed using Cutadapt [27]. Then, sequence quality was verified using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) [28]. The sequences were then mapped to the reference genomes of two soybean varieties using HISAT2 [29]. Soybean reference genome and gene annotation were obtained from the National Center for Biological Information web resource (<https://www.ncbi.nlm.nih.gov/genome/?term=Glycine%20max>). After final transcriptome generation, StringTie was used to estimate the expression levels of all the transcripts [30]. After removing the known mRNAs and transcripts measuring <200 bp, the lncRNAs were predicted for the remaining transcripts using CPC and CNCI [31]. The differentially expressed lncRNAs that had a |log₂ (fold change)| value of > 1 and a statistically significant p value of ≤ 0.05 were selected using the edgeR package in R [32].

Functional enrichment analysis

To explore the functions of lncRNAs in soybean development, their *cis*-target genes were predicted. The enrichment of the predicted genes from cellular components, molecular functions, and biological processes was analyzed using the Gene Ontology (GO) database (<http://www.geneontology.org>). Pathway analysis of potentially differently expressed genes was conducted using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg/mapper.html>) [33].

Quantitative real-time PCR validation

To verify the sequencing results, qRT-PCR was performed to detect the expression levels of lncRNAs and target genes in soybean pods at 50 d after flowering. The total RNAs of young soybean pods were extracted using the abovementioned methods and were reverse-transcribed using the UEIris II RT-PCR System with dsDNase (Code R2028, US Everbright Inc., Suzhou, China). Fast Super EvaGreen qPCR Master Mix (Code S2008; US Everbright Inc.) was used for the qRT-PCR reactions, which were performed under the following conditions: initial 3-min denaturation at 95 °C, followed by 40 cycles of 30 s denaturation at 95 °C, 30 s at 60 °C annealing temperature and a 30-s extension at 72 °C, followed by a

10 min final extension at 72 °C. All reactions were performed in triplicates. The relative gene expression values were calculated using the $2^{-\Delta\Delta CT}$ method [34].

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Competing interests

The authors declare that there are no competing interests in the reported research.

Availability of data

All data pertaining to the present study has been included in the Figures/supplementary files of the manuscript. The RNA-seq datasets are available in the GEO database (Accession ID: GSE161260, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE161260>).

Funding

This research was partially supported by Jilin Provincial Science and Technology Development Project (NO. 20200201027JC).

Author information

Affiliations

Crop germplasm Resources Innovation Laboratory, College of Life Sciences, Jilin Agricultural University, Changchun 130118, China

Dan Yao, Bohan Ma, Yue Li, Meng Teng, Aijing Zhang & Qiuzhu Zhao

Plant Biotechnology Centre, College of Agronomy, Jilin Agricultural University, Jilin, Changchun 130118, China

Author contributions

BHM collected the samples, performed the experiments and data analyses, made the Figures and Tables, and wrote the manuscript. DY and JZ joined the samples collection and edited the manuscript. MASK, YL, and MT conceived and designed the study. QZZ, AJZ and AL supervised the experiments and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank Peiwu Wang for his help in editing this paper. We thank HaoYu Shen (Hangzhou Lianchuan Bio Technologies Co) for providing help in the experiments.

References

1. Saito Y, Itakura K, Kuramoto M, Kaho T, Ohtake N, Hasegawa H, Suzuki T, Kondo N. Prediction of protein and oil contents in soybeans using fluorescence excitation emission matrix. *Food Chem.* 2021;365(5):130403.
2. Jing G, Tang DP, Yao Y, Su YK, Shen Y, Bai Y, Jing W, Zhang Q, Lin F, Guo DQ et al. Seed specifically over-expressing DGAT2A enhances oil and linoleic acid contents in soybean seeds. *Biochem Biophys Res Commun.* 2021;568:143-150.
3. Liu DG, Mewalal R, Hu RB, Tuskan GA, Yang XH. New technologies accelerate the exploration of non-coding RNAs in horticultural plants. *Hortic Res.* 2017;4:1-8.
4. Liu J, Wang H, Chua NH. Long noncoding RNA transcriptome of plants. *Plant Biotechnol J.* 2015;13(3):319-328.
5. Kang X, Cai JJ, Chen YX, Yan YC, Yang ST, He RQ, Wang D, Zhu YL. Pod-shattering characteristics differences between two groups of soybeans are associated with specific changes in gene expression. *Funct Integr Genomic.* 2020;20(2):201-210.
6. Bardou F, Ariel F, Simpson CG, Romero-Barrios N, Laporte P, Balzergue S, Brown JWS, Crespi M. Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. *Dev Cell.* 2014;30(2):166-176.
7. Cao P, Fan WJ, Li PJ, Hu YX. Genome-wide profiling of long noncoding RNAs involved in wheat spike development. *BMC Genomics.* 2021;22(1):493.
8. He H, Zhou Y-F, Yang Y-W, Zhang Z, Lei M-Q, Feng Y-Z, Zhang Y-C, Chen Y-Q, Lian J-P, Yu Y. Genome-wide analysis identified a set of conserved lncRNAs associated with domestication-related traits in rice. *Int J Mol Sci.* 2021;22(9):4742.

9. Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Chettoor AM, Givan SA, Cole RA, Fowler JE et al. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* 2014;15(2):R40.
10. Jiang N, Cui J, Shi YS, Yang GL, Zhou XX, Hou XX, Meng J, Luan YS. Tomato lncRNA23468 functions as a competing endogenous RNA to modulate NBS-LRR genes by decoying miR482b in the tomato-*Phytophthora infestans* interaction. *Hortic Res.* 2019;6:28.
11. Liu X, Hao L, Li D, Zhu LH, Hu SN. Long non-coding RNAs and their biological roles in plants. *Genom Proteom Bioinf.* 2015;13(3):137-147.
12. Baruah PM, Kashyap P, Krishnatreya DB, Bordoloi KS, Gill SS, Agarwala N. Identification and functional analysis of drought responsive lncRNAs in tea plant. *Plant Gene.* 2021;27(17):100311.
13. Zhao XB, Gan LM, Yan CX, Li CJ, Sun QX, Wang J, Yuan CL, Zhang H, Shan SH, Liu JN. Genome-wide identification and characterization of long non-coding RNAs in peanut. *Genes.* 2019;10(7):536.
14. Bhatia G, Singh A, Verma D, Sharma S, Singh K. Genome-wide investigation of regulatory roles of lncRNAs in response to heat and drought stress in *Brassica juncea* (Indian mustard). *Environ Exp Bot.* 2020;171:103922.
15. Yin DD, Li SS, Shu QY, Gu ZY, Wu Q, Feng CY, Xu WZ, Wang LS. Identification of microRNAs and long non-coding RNAs involved in fatty acid biosynthesis in tree peony seeds. *Gene.* 2018;666:72-82.
16. Ma BH, Zhang AJ, Zhao QJ, Li ZY, Lamboro A, He HB, Li Y, Jiao SQ, Guan SY, Liu SY. Genome-wide identification and analysis of long non-coding RNAs involved in fatty acid biosynthesis in young soybean pods. *Sci Rep.* 2021;11(1):7603.
17. Zhang JY, Xu HQ, Yang YM, Zhang XQ, Huang ZW, Zhang D. Genome-wide analysis of long non-coding RNAs (lncRNAs) in two contrasting soybean genotypes subjected to phosphate starvation. *BMC Genomics.* 2021;22(1):433-433.
18. Ma XL, Zhang XQ, Traore SM, Xin ZY, Ning LL, Li K, Zhao KK, Li ZF, He GH, Yin DM. Genome-wide identification and analysis of long noncoding RNAs (lncRNAs) during seed development in peanut (*Arachis hypogaea* L.). *BMC Plant Biol.* 2020;20(1):192.
19. Pang JL, Zhang X, Ma XH, Zhao J. Spatio-temporal transcriptional dynamics of maize long non-coding RNAs responsive to drought stress. *Genes.* 2019;10(2):138.
20. Shen EH, Zhu XT, Hua SJ, Chen HY, Ye CY, Zhou LH, Liu Q, Zhu QH, Fan LJ, Chen X. Genome-wide identification of oil biosynthesis-related long non-coding RNAs in allopolyploid *Brassica napus*. *BMC Genomics.* 2018;19:745.
21. Chen K, Huang Y, Liu CN, Liang Y, Li MT. Transcriptome profile analysis of Arabidopsis reveals the drought stress-induced long non-coding RNAs associated with photosynthesis, chlorophyll synthesis, fatty acid synthesis and degradation. *Front Plant Sci.* 2021;12:643182.
22. He M, Qin CX, Wang X, Ding NZ. Plant unsaturated fatty acids: biosynthesis and regulation. *Front Plant Sci.* 2020;11:390.
23. Liu JY, Li P, Zhang YW, Zuo JF, Li G, Han X, Dunwell JM, Zhang YM. Three-dimensional genetic networks among seed oil-related traits, metabolites and genes reveal the genetic foundations of oil

- synthesis in soybean. *Plant J.* 2020;103(3):1103-1124.
24. Maschietto V, Marocco A, Malachova A, Lanubile A. Resistance to *Fusarium verticillioides* and fumonisin accumulation in maize inbred lines involves an earlier and enhanced expression of lipoxygenase (LOX) genes. *J Plant Physiol.* 2015;188:9-18.
 25. Fan TQ, Zhang QX, Hu YY, Wang ZJ, Huang YJ. Genome-wide identification of lncRNAs during hickory (*Carya cathayensis*) flowering. *Funct Integr Genomic.* 2020;20(4):591-607.
 26. Yan Q, Wu F, Yan ZZ, Li J, Ma TT, Zhang YF, Zhao YF, Wang YR, Zhang JY. Differential co-expression networks of long non-coding RNAs and mRNAs in *Cleistogenes songorica* under water stress and during recovery. *BMC Plant Biol.* 2019;19(1):23.
 27. Martin M. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *Embnet Journal.* 2011;17(1).
 28. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357-U354.
 29. Kim D, Landmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015;12(4):357-U121.
 30. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33(3):290-+.
 31. Gao Y, Wang JF, Zhao FQ. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.* 2015;16:4.
 32. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* 2016;11(9):1650-1667.
 33. Sun L, Luo HT, Bu DC, Zhao GG, Yu KT, Zhang CH, Liu YN, Chen RS, Zhao Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 2013;41(17).
 34. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods.* 2001;25(4):402-408.

Figures

Figure 1

Identification of lncRNAs in soybean. (a) Venn diagram of the lncRNAs detected in each sample (MT_50d 'MT72', JN_50d 'JN18'); (b) Chart showing the different types of lncRNAs; (c) Numbers of exons in lncRNAs and protein-coding mRNAs; (d) Comparison of the lengths of lncRNAs and protein-coding mRNAs; (e) Expression levels of lncRNAs along 20 soybean chromosomes. The outermost layer represents all soybean chromosomes (Chr01–20), whereas the inner layer of the figure represents data

from three samples per group, which are JN_50 and MT_50 from the outside to the inside; (f) Distribution of different types of lncRNAs. The antisense, intronic, and sense lncRNAs—as well as long intervening/intergenic noncoding RNAs—are represented by different concentric rings based on the lncRNA loci in the genome.

Figure 2

Expression patterns of differentially expressed lncRNAs. (a) Venn diagram of differentially expressed lncRNAs in five different comparison groups; (b) Regulation of differentially expressed lncRNAs in five different comparison groups.

Figure 3

Volcano map of differentially expressed lncRNAs (red and blue dots represent significantly up-regulated and downregulated lncRNAs, respectively). (a) JN_50d versus JN_30d; (b) JN_50d versus JN_40d; (c) MT_50d versus JN_50d; (d) MT_50d versus MT_30d; (e) MT_50d versus MT_40d.

Figure 4

Clustered heat map of differentially expressed lncRNAs in five different comparison groups. (a) JN_50d versus JN_30d; (b) JN_50d versus JN_40d; (c) MT_50d versus JN_50d; (d) MT_50d versus MT_30d; (e) MT_50d versus MT_40d.

Figure 5

Expression patterns of differentially expressed mRNAs. (a) Venn diagram of differentially ex-pressed mRNAs in five different comparison groups; (b) Regulation of differentially expressed mRNAs in five different comparison groups.

Figure 6

Clustered heat map of differentially expressed target mRNAs in five different comparison groups. (a) JN_50d versus JN_30d; (b) JN_50d versus JN_40d; (c) MT_50d versus JN_50d; (d) MT_50d versus MT_30d; (e) MT_50d versus MT_40d.

Figure 7

Gene ontology analysis of target protein-coding genes of differentially expressed lncRNAs in five different comparison groups. (a) JN_50d versus JN_30d; (b) JN_50d versus JN_40d; (c) MT_50d versus JN_50d; (d) MT_50d versus MT_30d; (e) MT_50d versus MT_40d.

Figure 8

KEGG analysis of the target protein-coding genes of differentially expressed lncRNAs in five different comparison groups. (a) JN_50d versus JN_30d; (b) JN_50d versus JN_40d; (c) MT_50d versus JN_50d; (d) MT_50d versus MT_30d; (e) MT_50d versus MT_40d.

Figure 9

KEGG metabolic pathways. (a) fatty acid elongation; (b) alpha-linolenic acid; (c) linoleic acid; (d) biosynthesis of unsaturated fatty acids; (e) fatty acid degradation.

Figure 10

Validation of the expression pattern of highly expressed lncRNAs and protein-coding genes by qRT-PCR. * Represents $p < 0.05$, there is statistical difference, ** represents $p < 0.01$ with statistically significant difference. (a) lncRNAs; (b) target mRNAs.

Figure 11

lncRNAs-mRNAs network. The yellow and purple represent mRNAs and lncRNAs, respectively. This network diagram shows partial results of the interactions between lncRNAs and mRNAs.

Figure 12

Soybean [*Glycine max* (L.) Merr] 50 d after flowering (samples were collected seven times between 20 and 80 d after flowering).

Figure 13

Gas chromatogram of soybean pods at different stages. (a) Gas chromatographic peak of fatty acid content in soybean at different stages (the starting points from left to right are palmitic acid, stearic acid, oleic acid, linoleic acid, arachidonic acid, linolenic acid and behenic acid); (b) Relative content of fatty acids in soybean in different periods.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryfiles.zip](#)