

ANNOTATION AND ANALYSIS OF THE TRANSCRIPTOME OF BEARING BRANCH

ZHU, B.¹ – ZHANG, L.^{1*} – XU, X. B.¹ – ZOU, B. Y.² – GUAN, Q. M.² – YANG, J. J.² – YU, Z.²

¹*Forestry College, Jiangxi Agricultural University, Nanchang 330045, China*

²*Jiangxi BoJun Agriculture Holding Limited, Nancheng 344700, China*

**Corresponding author*

e-mail: LLPPMM222@163.com

(Received 4th Jun 2019; accepted 16th Oct 2019)

Abstract. The analysis of the Transcriptome of different jujube, and the mining of their functional genes was performed. And the genes were annotated and analyzed, and the difference gene was screened out to provide the theoretical basis for the fresh jujube. The sequencing of the cDNA library was performed using synthesis technology and the Illumina HiSeq platform. The sequencing results of splicing and functional annotation. A total of 29402 Unigenes were annotated, while in the Nr, Swiss-Prot, GO, COG, KEGG and Pfam databases 29330, 24358, 21771, 11028, 6962 and 28940 Unigenes, respectively. At the same time, 355 new genes were annotated. The 50 functional groups in the Go database gene function annotation, 14608 metabolic processes in biological processes, accounting for 67.1%; reference KEGG database, 110 metabolic pathways, the number of annotations, gene function ribosomes, amino acid biosynthesis. Human embryonic stem cells, carbon metabolism, and plant hormone signaling are more abundant; it is noted that COD in 11028 single-gene distribution families has general function prediction, transcription, replication, recombination and repair, and secondary metabolites in 25 genes. Biosynthesis, transport and catabolism. This study is the first time to use RNA-seq high-throughput sequencing technology to perform transcriptome sequencing and functional analysis of transcriptome results in fresh-eating jujube leaves, to explore genes for biological processes, cell components and molecular functions, and to study different functional genes, effective biosynthetic pathways and regulatory mechanisms.

Keywords: *fresh jujube, Bearing branch, RNA-seq, differential gene expression, gene ontology, transcriptomics*

Introduction

Ziziphus jujuba Mill is a plant belonging to the Rhamnaceae family. It's the main cultivated species in China. Cultivation and distribution all over the country. The species has recently been introduced to Asia, Europe, America, Africa and Oceania, but is generally not cultivated economically (Qu et al., 1993). *Jujube* are different from other fruits. Its branches can be divided into once branch of jujube, jujube stock and bearing branch. The Bearing branch is the basic unit of the result of the jujube tree. It is also known as the result branch, the exfoliative branch, the falling branch, and the two type

branch (Tang et al., 2012). Bearing branch can be divided into lignification jujube hoists and shedding jujube cranes (Yang et al., 2007, 2014). Recently, many studies have been made on the lignification of jujube cranes. Scholars believe that lignification jujube is better in result performance. Study on the cultivation of fresh Chinese jujube in the South. Study on the performance, photosynthetic efficiency, nutrient transport and nutrient accumulation of lignification jujube hoist (Yan et al., 2013; Wang et al., 2014). The majority of jujube stock is formed by the accessory bud. In practice, callus or pruning of jujube head is easy to form jujube crane. The jujube cranes can also be lifted on the base of the jujube head and the two branches of the year. Most jujube cranes are able to blossom, and the small part of the jujube is suspended from the top of the jujube head (Shi et al., 1999; Tang et al., 2012; Wang et al., 2014). In fact, jujube crane can be divided into fruit jujube crane and non-fruiting jujube crane from its form, bear fruit Jujube cranes are divided into lignification jujube cranes and non-lignification jujube cranes. The study of jujube crane is directly related to the value of the breeding and breeding of jujube trees. It is also related to the economic and production value of jujube trees. Study the development rule of jujube crane to provide basic theory for production management (Sun et al., 1996; Chen et al., 2015).

RNA-Seq can respond to the expression of genes at the transcriptional level (Reich et al., 2018). High energy sequencing can be used to better study the expression of different genes in different jujube cranes. Thus, the molecular formation mechanism of the flower bud differentiation and results of jujube crane is described better.

Materials and methods

Experimental materials

Sample from Jiangxi Xianlv modern agricultural demonstration garden Magu jujube No. 1, Take 3 different jujube cranes: first is Lignification jujube hanging leaf (A1, A2, A3), Non lignification jujube hanging leaf (B1, B2, B3). Non lignification jujube leaves without flower bud differentiation (C1, C2, C3). Lignification leaves of jujube leaves on 3 jujube trees (A1, A2, A3), B1 and C1 are the same trees, B2 and C2 are the same trees, B3 and C3 are the same trees. A total of 9 samples (3 Biological duplicates). Rapid freezing of liquid nitrogen after sampling. Put in the -80°C refrigerator and spare.

Construction of cDNA library and sequencing of Transcriptome

RNA sequencing is commissioned by Hangzhou Jing Jie Biological Technology Co., Ltd. After testing the sample. Magnetic beads with Oligo (dT). Complementary pairing through A-T and use The combination of mRNA's ployA tail enriching the mRNA of eukaryotes. Then add fragmentation buffer to break mRNA into short segments. Take mRNA as a template. Synthesis of a chain cDNA with six base random primers. And then add the buffer, *dNTPs* and *DNA* polymerase I. Synthesis of two chain cDNA. Then using

AMPure XP beads to purify double stranded cDNA. Purification of double stranded cDNA for terminal repair. Adding A tail and connecting sequenced joint. Then AMPure XP beads is used to select the size of the fragment. Finally, the final PCR library is enriched and the final cDNA library is obtained (Qi et al., 2011; Ding et al., 2017).

Statistics and comparison of sequencing

The original data reads of the machine. Clean Data was filtered after removing the reads containing the joint, repeated and low quality of the sequencing. Sequence alignment with the specified reference genome. The obtained Mapped Data. Quality evaluation of insert fragment length test, randomness test and so on. This project uses the specified genome as a reference for sequence alignment and subsequent analysis. Using *TopHat2* (Kim et al., 2013) Alignment of Clean Reads with the reference genome. Obtaining location information on the reference genome or gene. And sequence characteristic information of sequence samples. *TopHat2* is based on the comparison of software *Bowtie2* (Langmead et al., 2012). Comparison of the transcriptome sequence Reads to the genome, Recognition of splice points between exons by analyzing comparison results (Splicing Junction).

Discovery of new genes

Based on the selected reference genome sequence. Using Cufflinks software to splice Mapped Reads. And compare with the original genome annotation information. Looking for the original unnoted transcriptional area. Discovery of the new transcript and new genes of the species. So as to supplement and improve the original genome annotation information. Filtering the encoded peptide chain is too short or contains only a single exon sequence (less than 50 amino acid residues). Using BLAST software to compare the new genes that are excavated with the *NR*, *Swiss-Prot*, *GO*, *COG*, *KOG*, *Pfam*, *KEGG* database, KEGG Orthology results of new genes were obtained by using KOBAS2.0, Using HMMER software compared with Pfam database after predicting the amino acid sequence of the new gene. Annotated information of new genes (Zhu et al., 2018).

Functional annotation and analysis of the Transcriptome

Filtering and assembly of Unigenes. Based on the selected reference genome sequence. Using Cufflinks software to splice Mapped Reads. And compare with the original genome annotation information. Looking for the original unnoted transcriptional area. Discovery of the new transcript and new genes of the species. So as to supplement and improve the original genome annotation information. Filtering the encoded peptide chain is too short or contains only a single exon sequence (less than 50 amino acid residues). Using *BLAST* software to compare the new genes that are excavated with the *NR*, *Swiss-Prot*, *GO*, *COG*, *KOG*, *Pfam*, *KEGG* database. KEGG Orthology results of new genes were obtained by using KOBAS2.0. Using HMMER software compared with Pfam database after predicting

the amino acid sequence of the new gene, Obtain annotated information of the measured genes (Liu et al., 2017; Niu et al., 2018; Zhang et al., 2018; Chen et al., 2018).

Results and analysis

Sequencing data output statistics

Repeat sequencing of 3 material samples with high energy sequencing technology. The total number of pair-end Reads of clean Data in the ReadSum of A2 is maximum. Up to 28708241. The total base number of Clean Data in BaseSum is 8612472300. The minimum of A3's ReadSum in all the samples is 21709049. The BaseSum 6512714700 of A3. The percentage of two bases of G and C in Clean DataGC is more than 44%. The percentage of Clean Data with a mass greater than or equal to 30 of a base is greater than 88%. The percentage of Clean Data with a mass greater than or equal to 20 of a base is greater than 94%. The above analysis indicates that the quality of sequencing is good, and the data can be used for subsequent All-Unigenes database establishment, gene function annotation, classification and same-sex analysis. The output statistics of the sample data in this project are shown in *Table 1*.

Table 1. *Sequencing data statistics table*

SampleID	ReadSum	BaseSum	GC(%)	Q20(%)	Q30(%)
A1	24, 974, 459	7, 492, 337, 700	45. 18%	95. 07%	88. 88%
A2	28, 708, 241	8, 612, 472, 300	44. 78%	95. 25%	89. 25%
A3	21, 709, 049	6, 512, 714, 700	44. 34%	94. 81%	88. 42%
B1	26, 961, 553	8, 088, 465, 900	44. 44%	95. 25%	89. 25%
B2	26, 875, 609	8, 062, 682, 700	45. 00%	95. 27%	89. 26%
B3	27, 843, 858	8, 353, 157, 400	44. 96%	95. 28%	89. 30%
C1	22, 799, 545	6, 839, 863, 500	44. 79%	95. 25%	89. 24%
C2	24, 859, 991	7, 457, 997, 300	45. 08%	95. 34%	89. 41%
C3	24, 962, 466	7, 488, 739, 800	45. 00%	95. 32%	89. 35%

Ratio of transcriptional data to reference genome sequence

This not only provides a data base for variable splicing analysis, but also enables more Reads to be compared to the reference genome, and improves the utilization of sequencing data.

Comparison efficiency refers to the percentage of Mapped Reads that accounts for Clean Reads, which is the most direct manifestation of the data utilization of the transcriptional group. The efficiency of comparison is not only affected by the quality of data sequencing, but also related to the advantages and disadvantages of the reference genome assembly, and the biological classification relationship between reference genome and sequencing samples. By comparing efficiency, it is possible to assess whether the selected reference genome assembly can meet the needs of information analysis (*Table 2*).

Table 2. A statistical table of sequence alignment of the sample sequence and the selected reference genome

sampleID	total_read	reads_mapped (%)	multi_mapped (%)	uniq_mapped (%)	reads_plus (%)	reads_minus (%)
A1	49948918	34289534 (68. 6)	486538 (1. 42)	33802996 (98. 58)	37956271 (62. 85)	22435941 (37. 15)
A2	57416482	39460995 (68. 73)	484144 (1. 23)	38976851 (98. 77)	42123971 (63. 49)	24220188 (36. 51)
A3	43418098	29243769 (67. 35)	362679 (1. 24)	28881090 (98. 76)	32184159 (64. 11)	18019123 (35. 89)
B1	53923106	36479830 (67. 65)	477714 (1. 31)	36002116 (98. 69)	39978099 (64. 07)	22423068 (35. 93)
B2	53751218	36611439 (68. 11)	487630 (1. 33)	36123809 (98. 67)	40024889 (63. 69)	22814098 (36. 31)
B3	55687716	38058734 (68. 34)	524265 (1. 38)	37534469 (98. 62)	41600309 (63. 49)	23922692 (36. 51)
C1	45599090	31073411 (68. 14)	426693 (1. 37)	30646718 (98. 63)	34299648 (63. 54)	19678520 (36. 46)
C2	49719982	33687509 (67. 75)	499050 (1. 48)	33188459 (98. 52)	37803787 (63. 55)	21683112 (36. 45)
C3	49924932	34395909 (68. 90)	478825 (1. 39)	33917084 (98. 61)	37594222 (63. 12)	21969233 (36. 88)

More than 67% samples in total percentage of Reads compared to the reference genome Reads number in Clean Reads accounted for the percentage of the number of Reads, compared to the reference genome at various locations in the Clean Reads accounted for less than 2%, the number of Reads percentage compared to the reference genome location in Clean Reads only accounted for more than 98%, the percentage ratio the number of Reads positive reference genome chains in Clean Reads accounted for more than 62% (including more than reads), compared to the number of Reads and percentage of negative strand genome reference in Clean Reads accounted for less than 38% (including more than reads).

Discovery of new genes

The total number of newly discovered genes obtained from corresponding database annotation information was 355, of which 196 were 300-1000bp in length and 155 in length 1000bp. COG notes to 37, of which there are 19 300-1000bp in length, length of more than 1000bp 18; GO notes to 205, of which there are 123 300-1000bp in length, length of more than 1000bp 79; KEGG notes to 65, of which there are 38 300-1000bp in length, length of more than 1000bp there are 27; KOG notes to 155, of which there are 77 300-1000bp in length, length of more than 1000bp 77; Pfame notes to 138, of which there are 68 300-1000bp in length, length of more than 1000bp 68; Swissprot notes to 196, of which there are 100 300-1000bp in length a length of more than 1000bp, 95; NR notes to 344, of which there are 188 300-1000bp in length, length of more than 1000bp 152. The specific distribution of the genes for each data annotation is shown in *Table 3* and *Figure 1*.

Table 3. New discovery of gene functional annotation results

Anno_Database	Annotated_Number	300<=length<1000	length>=1000
COG	37	19	18
GO	205	123	79
KEGG	65	38	27
KOG	155	77	77
Pfam	138	68	68
Swissprot	196	100	95
nr	344	188	152
All	355	196	155

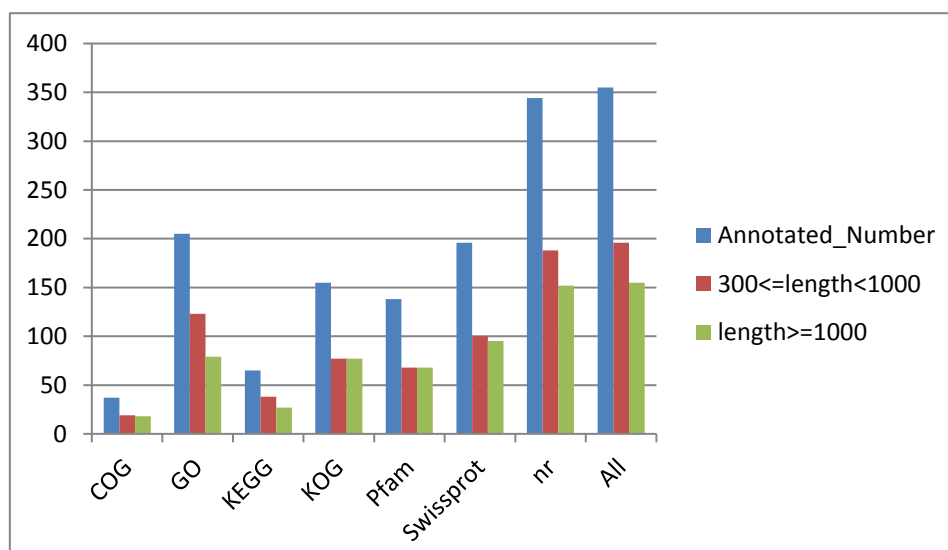


Figure 1. New gene annotation distribution

Annotation and functional classification of genes

The total number of genes in the annotation was 29402, of which 12539 were 300-1000bp in length and 16307 in length greater than that of 1000bp. COG notes to 11028, of which there are 3757 300-1000bp in length, length of more than 1000bp 7172; GO notes to 21771, of which there are 8900 300-1000bp in length, length of more than 1000bp 12413; KEGG notes to 6962, of which there are 2683 300-1000bp in length, length of more than 1000bp there are 4154; KOG notes to 16821, of which there are 6294 300-1000bp in length, length of more than 1000bp 10278; Pfame notes to 28940, of which there are 12273 300-1000bp in length, length of more than 1000bp 16115; Swissprot note to 24358, which is 300-1000bp in length 9435, length of more than 1000bp 14589; NR notes to 29330, of which there are 12491 300-1000bp in length, length of more than 1000bp 16286.

The number of functional genes that are ultimately annotated by each database is shown in Table 4.

Table 4. Statistical analysis of functional annotation of gene

Anno_Database	Annotated_Number	300<=length<1000	length>=1000
COG	11028	3757	7172
GO	21771	8900	12413
KEGG	6962	2683	4154
KOG	16821	6294	10278
Swissprot	24358	9435	14589
Pfam	28940	12273	16115
nr	29330	12491	16286
All	29402	12539	16307

Gene GO classification

GO database is a structured standard biological annotation system built by GO Organization (Gene Ontology Consortium) in 2000, aiming at establishing standard vocabulary system of genes and their product knowledge, which is suitable for all kinds of species. The GO annotation system is a directed acyclic graph, which contains three main branches, namely Biological Process, Molecular Function and Cellular Component. In this study, the Unigene obtained from three samples of fresh jujube was compared in the GO functional database. A total of 21771 annotations were selected in 50 functional groups of GO biology. The metabolic processes involved in biological processes (14608, 67.1%), cell processes (12146, 55.79%), and single biological processes (10464, 48.06%). The main components related to annotation of cell components were cells (9818, 45.09%), membrane (8930, 41.02%), cell components (9760, 44.83%), and membrane (6609, 30.36%). The main parts of the molecular functional annotations are catalytic activity (11882, 54.57%), and binding (9688, 44.5%). This result also shows the main distribution of the annotation function. See *Table 5*.

Classification of KEGG metabolic pathways

To KEGG database as a reference, there are 6962 notes to the KEGG database, can be classified into 110 groups of gene transcription in the metabolic pathways (*Table 6*), which notes to the biosynthesis of ribosomes, amino acid biosynthesis, carbon metabolism, plant hormone signaling, plant pathogen interaction, splicing, RNA transport, and starch sucrose metabolism, endoplasmic reticulum protein processing, phenylpropanoid is relatively high. Glucosinolate biosynthesis and degradation of aromatic compounds, sulfur relay systems, non-homologous termination, glycosphingolipid biosynthesis - ganglion series, beet red pigment biosynthesis, lipoic acid metabolism, other types of O-glycan biosynthesis and Vancomycin resistance, caffeine metabolic pathways is to compare notes low.

Table 5. GO annotation classification of Unigene in jujube leaf transcriptional group

#GO_classify1	GO_classify2	Trans gene
cellular component	extracellular region	594
cellular component	cell	9818
cellular component	nucleoid	17
cellular component	membrane	8930
cellular component	virion	81
cellular component	cell junction	431
cellular component	extracellular matrix	11
cellular component	membrane-enclosed lumen	413
cellular component	macromolecular complex	2218
cellular component	organelle	6339
cellular component	extracellular region part	48
cellular component	organelle part	2962
cellular component	virion part	81
cellular component	membrane part	6609
cellular component	cell part	9760
molecular function	protein binding transcription factor activity	108
molecular function	nucleic acid binding transcription factor activity	502
molecular function	catalytic activity	11882
molecular function	receptor activity	153
molecular function	guanyl-nucleotide exchange factor activity	30
molecular function	structural molecule activity	512
molecular function	transporter activity	1435
molecular function	binding	9688
molecular function	electron carrier activity	154
molecular function	antioxidant activity	207
molecular function	channel regulator activity	1
molecular function	metallochaperone activity	4
molecular function	enzyme regulator activity	228
molecular function	protein tag	6
molecular function	translation regulator activity	1
molecular function	nutrient reservoir activity	44
molecular function	molecular transducer activity	303
biological process	reproduction	337
biological process	cell killing	3
biological process	immune system process	234
biological process	metabolic process	14608
biological process	cellular process	12146
biological process	reproductive process	827
biological process	biological adhesion	49
biological process	signaling	1087
biological process	multicellular organismal process	1343
biological process	developmental process	1785
biological process	growth	281
biological process	locomotion	9
biological process	single-organism process	10464
biological process	rhythmic process	26
biological process	response to stimulus	3905
biological process	localization	3404
biological process	multi-organism process	525
biological process	biological regulation	4495
biological process	cellular component organization or biogenesis	2970

Table 6. KEGG metabolic pathway annotation table

#pathway	pathway_id	Gene_number
Glycolysis / Gluconeogenesis	ko00010	102
Citrate cycle (TCA cycle)	ko00020	39
Pentose phosphate pathway	ko00030	45
Pentose and glucuronate interconversions	ko00040	66
Fructose and mannose metabolism	ko00051	54
Galactose metabolism	ko00052	57
Ascorbate and aldarate metabolism	ko00053	40
Fatty acid biosynthesis	ko00061	44
Fatty acid elongation	ko00062	35
Fatty acid degradation	ko00071	38
Synthesis and degradation of ketone bodies	ko00072	8
Cutin, suberine and wax biosynthesis	ko00073	21
Steroid biosynthesis	ko00100	19
Ubiquinone and other terpenoid-quinone biosynthesis	ko00130	18
Oxidative phosphorylation	ko00190	89
Photosynthesis	ko00195	36
Photosynthesis - antenna proteins	ko00196	12
Purine metabolism	ko00230	128
Caffeine metabolism	ko00232	1
Pyrimidine metabolism	ko00240	101
Alanine, aspartate and glutamate metabolism	ko00250	44
Glycine, serine and threonine metabolism	ko00260	60
Cysteine and methionine metabolism	ko00270	65
Valine, leucine and isoleucine degradation	ko00280	59
Valine, leucine and isoleucine biosynthesis	ko00290	25
Lysine biosynthesis	ko00300	12
Lysine degradation	ko00310	20
Arginine and proline metabolism	ko00330	72
Histidine metabolism	ko00340	16
Tyrosine metabolism	ko00350	42
Phenylalanine metabolism	ko00360	92
Tryptophan metabolism	ko00380	27
Phenylalanine, tyrosine and tryptophan biosynthesis	ko00400	38
beta-Alanine metabolism	ko00410	46
Taurine and hypotaurine metabolism	ko00430	9
Selenocompound metabolism	ko00450	25
Cyanoamino acid metabolism	ko00460	85
Glutathione metabolism	ko00480	94
Starch and sucrose metabolism	ko00500	166
N-Glycan biosynthesis	ko00510	39
Other glycan degradation	ko00511	16
Other types of O-glycan biosynthesis	ko00514	2
Amino sugar and nucleotide sugar metabolism	ko00520	106
Glycosaminoglycan degradation	ko00531	17
Glycerolipid metabolism	ko00561	47
Inositol phosphate metabolism	ko00562	58
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	ko00563	19
Glycerophospholipid metabolism	ko00564	74
Ether lipid metabolism	ko00565	28
Arachidonic acid metabolism	ko00590	10
Linoleic acid metabolism	ko00591	11
alpha-Linolenic acid metabolism	ko00592	37
Sphingolipid metabolism	ko00600	20
Glycosphingolipid biosynthesis - globo series	ko00603	14
Glycosphingolipid biosynthesis - ganglio series	ko00604	5

#pathway	pathway_id	Gene_number
Pyruvate metabolism	ko00620	78
Glyoxylate and dicarboxylate metabolism	ko00630	51
Propanoate metabolism	ko00640	26
Butanoate metabolism	ko00650	20
C5-Branched dibasic acid metabolism	ko00660	11
One carbon pool by folate	ko00670	16
Carbon fixation in photosynthetic organisms	ko00710	52
Thiamine metabolism	ko00730	9
Riboflavin metabolism	ko00740	9
Vitamin B6 metabolism	ko00750	9
Nicotinate and nicotinamide metabolism	ko00760	14
Pantothenate and CoA biosynthesis	ko00770	30
Biotin metabolism	ko00780	17
Lipoic acid metabolism	ko00785	3
Folate biosynthesis	ko00790	11
Porphyrin and chlorophyll metabolism	ko00860	45
Terpenoid backbone biosynthesis	ko00900	46
Monoterpenoid biosynthesis	ko00902	11
Limonene and pinene degradation	ko00903	8
Diterpenoid biosynthesis	ko00904	21
Brassinosteroid biosynthesis	ko00905	9
Carotenoid biosynthesis	ko00906	33
Zeatin biosynthesis	ko00908	28
Sesquiterpenoid and triterpenoid biosynthesis	ko00909	24
Nitrogen metabolism	ko00910	29
Sulfur metabolism	ko00920	35
Phenylpropanoid biosynthesis	ko00940	143
Flavonoid biosynthesis	ko00941	23
Anthocyanin biosynthesis	ko00942	13
Flavone and flavonol biosynthesis	ko00944	9
Stilbenoid, diarylheptanoid and gingerol biosynthesis	ko00945	13
Isoquinoline alkaloid biosynthesis	ko00950	23
Tropane, piperidine and pyridine alkaloid biosynthesis	ko00960	23
Betalain biosynthesis	ko00965	5
Glucosinolate biosynthesis	ko00966	8
Aminoacyl-tRNA biosynthesis	ko00970	54
Biosynthesis of unsaturated fatty acids	ko01040	33
Carbon metabolism	ko01200	199
2-Oxocarboxylic acid metabolism	ko01210	55
Fatty acid metabolism	ko01212	68
Degradation of aromatic compounds	ko01220	8
Biosynthesis of amino acids	ko01230	206
Vancomycin resistance	ko01502	2
ABC transporters	ko02010	15
Ribosome biogenesis in eukaryotes	ko03008	83
Ribosome	ko03010	230
RNA transport	ko03013	167
mRNA surveillance pathway	ko03015	109
RNA degradation	ko03018	103
RNA polymerase	ko03020	41
Basal transcription factors	ko03022	40
DNA replication	ko03030	45
Spliceosome	ko03040	171
Proteasome	ko03050	45
Protein export	ko03060	41
Base excision repair	ko03410	43
Nucleotide excision repair	ko03420	49

#pathway	pathway_id	Gene_number
Mismatch repair	ko03430	33
Homologous recombination	ko03440	49
Non-homologous end-joining	ko03450	7
Phosphatidylinositol signaling system	ko04070	56
Plant hormone signal transduction	ko04075	188
Ubiquitin mediated proteolysis	ko04120	96
Sulfur relay system	ko04122	8
SNARE interactions in vesicular transport	ko04130	27
Regulation of autophagy	ko04140	21
Protein processing in endoplasmic reticulum	ko04141	157
Endocytosis	ko04144	135
Phagosome	ko04145	67
Peroxisome	ko04146	74
Plant-pathogen interaction	ko04626	180
Circadian rhythm - plant	ko04712	46

COG functional classification

In this study, 11028 Unigenes, annotated into COG, were distributed in 25 gene families (Table 7).

Table 7. The function table annotated in the COG Library

ID	Class_Name	Numbers
R	General function prediction only	3500
K	Transcription	1767
L	Replication, recombination and repair	1749
T	Signal transduction mechanisms	1669
O	Posttranslational modification, protein turnover, chaperones	960
G	Carbohydrate transport and metabolism	821
J	Translation, ribosomal structure and biogenesis	797
Q	Secondary metabolites biosynthesis, transport and catabolism	749
E	Amino acid transport and metabolism	740
S	Function unknown	615
C	Energy production and conversion	562
P	Inorganic ion transport and metabolism	530
I	Lipid transport and metabolism	457
M	Cell wall/membrane/envelope biogenesis	417
D	Cell cycle control, cell division, chromosome partitioning	310
V	Defense mechanisms	244
H	Coenzyme transport and metabolism	235
A	RNA processing and modification	185
Z	Cytoskeleton	182
U	Intracellular trafficking, secretion, and vesicular transport	180
F	Nucleotide transport and metabolism	147
B	Chromatin structure and dynamics	87
N	Cell motility	25
Y	Nuclear structure	1
W	Extracellular structures	0

Such as general function prediction, transcription, replication, recombination and repair, signal transduction, signal transduction, signal transduction, post-translational modification, protein turnover, mate, sugar transport and metabolism, translation,

ribosome structure and biosynthesis of secondary metabolites biosynthesis, transport and catabolism and other functions (Allen et al., 2018; Astaneh et al., 2018; Li et al., 2018; Nisavic et al., 2018; Wang et al., 2018; Wu et al., 2018).

Conclusion

The sequence of cDNA library using Using Illumina HiSeq platform, the sequencing results were linked and functional annotations were analyzed to reveal the overall functional expression pattern of transcripts from different leaves of jujube. In recent years, there have been many studies based on high-energy sequencing, including *Daphne odora Thunb* (Yang et al., 2017), *Michelia macclurei Dandy* (Li et al., 2017), *Rhododendron calvenscens Balf.f.* (Li et al., 2017), *Carthamus tinctorius L.* (Dong et al., 2017), *Penthorum chinense Pursh* (Yuan et al., 2017) and other plants. Transcriptome genes are classified into 110 metabolic pathways, including ribosome, amino acid biosynthesis, carbon metabolism, plant hormone signal transduction, plant pathogen interaction, splicer, RNA transport, starch and sucrose metabolism, endoplasmic reticulum protein processing, phenylpropanoid biosynthesis and so on. This study is the first time to use RNA-seq high-throughput sequencing technology to perform transcriptome sequencing and functional analysis of transcriptome results in fresh-eating jujube leaves, to explore genes for biological processes, cell components and molecular functions, and to study different functional genes, effective biosynthetic pathways and regulatory mechanisms.

Analysis of lignin-related genes in lignified jujube suspension

The results showed that the expression of lignin-related genes was more active in comparison with non-lignified jujube suspension, all of which were up-regulated genes. There were 11 up-regulated genes: *gene 29690*, *gene 4518*, *gene 30089*, *gene 1272*, *gene 21988*, *gene 16475*, *gene 16475*, *gene 14268*, *gene 14269*, *gene 10082* and *gene 29495*.

The main functional genes in *gene 29690* family are monooxygenase activity and lignin metabolism. In *gene 14269* gene family, it includes the functions to be expressed in almost other gene families. Meristem growth regulation, chitin decomposition, glycolysis, cytokinin response, abscisic acid response, lignin biosynthesis, polar transport of auxin, cell tip growth.

The up-regulation of gene expression related to Lignin Metabolism in lignified jujube suspension also confirms that the cause of lignified jujube suspension is related to the synergy of gene expression mentioned above. Jiang et al. (2018) found that 16 lignin genes, including 2 *PAL*, 5 *CCR*, 3 *4CL*, 2 *CADH2* and 4 *LAC*, were up-regulated with shoot development, suggesting that they may be related to developmental lignin accumulation. Lin et al. (2018) studied the characteristics and rules of Lignin Metabolism in okra fruits and found that the significant increase of lignin and cellulose content was

related to phenylalanine ammonia lyase (*PAL*), cinnamyl alcohol 4-hydroxylase (*C4H*), 4 coumarin acyl-CoA ligase, protopectin and flavonoid enzymes.

PRX52 is a major lignin and secondary cell wall biosynthetic gene, especially in xylem vessel. The mutant of *Arabidopsis thaliana* showed a 70-80% decrease in lignin content, which was syringyl lignin (Fernandez-Perez et al., 2015). Previous studies have investigated its possible role in ABA-mediated plant defense against bacterial and fungal responses (Mohr et al., 2007). *TCP3* and *MYB12* were found to be involved in the regulation of *PRX52* (Arro et al., 2017) in grape tendril transcriptome studies. During lignification, *PRX52* interacts with *TCP3* more extensively than it responds to adversity. *PRX52* also shares the phenylpropane metabolic pathway in lignin biosynthesis with *MYB12* (response switch gene for flavonol metabolism) (Czemmel et al., 2012). Phenylalanine aminolytic enzyme (*PAL*) is a rate-limiting enzyme in the metabolic pathway of lignin phenylpropane, which catalyzes the conversion of phenylalanine to cinnamic acid and coenzyme lipids, and is one of the most important enzymes in lignin synthesis (Matus et al., 2008). In this experiment, homologous genes were expressed in the transcripts of lignified jujube suspension.

Differential gene expression analysis of lignified jujube suspension related to photosynthesis

The results showed that the genes annotated in GO were up-regulated in *gene 4228*, *gene 10305* and *gene 42934*, down-regulated in *gene 6204*, *gene 29074*, *gene 24178*, *gene 15775*, *gene 2674*, *gene 4496* and *gene 296156*. In the annotation of metabolic pathway in KEGG database, three genes participated in photosynthesis expression, of which *gene 4496*, *gene 24178* were down-regulated and *gene 4293* was up-regulated.

In the *gene 4496* gene family, there are reactive oxygen species, photosynthesis and redox processes involved in biological processes. Molecular function is involved in iron oxidase activity and iron binding. Chloroplasts are involved in cell components. In *gene 24178* expression, serine cysteine biosynthesis, photosynthetic adaptation, regulation of hydrogen peroxide metabolism, regulation of superoxide metabolism, photosynthesis and so on. The molecular functions include transferase activity, pyridoxal phosphate binding activity and cysteine synthetase activity. The thylakoid cavity is involved in the up-regulation of cell composition. *Gene 4293* family genes are up-regulated, including sulfur amino acid metabolism, glycine metabolism, vitamin metabolism, cell amino acid biosynthesis, aromatic amino acid metabolism, lipid metabolism, coenzyme biosynthesis, cold reaction, biological stimulation detection, blue light response. High light intensity response, red light response, far red light response, PSII related light capture complex II catabolism process, regulation of hydrogen peroxide metabolism process, chlorophyll biosynthesis process, starch biosynthesis process. Only the chloroplast envelope is involved in the expression.

Liu (2008) studied the temporal and spatial expression of *cab-PhE3*, *cab-PhE1*, *cab-PhE5*, *cab-PhE8*, *cab-PhE10* and *cab-PhE11* genes in *Phyllostachys heterocycla* (Carr.)

Mitford cv. Pubescens under photosynthesis (Zhang et al., 2009). Through proteomics, bioinformatics and *VIGS* analysis, it was found that *TYLCCNV*/*TYLCCNB* invasion was involved in stress and defense, energy production, photosynthesis, protein homeostasis, metabolism, cell structure, signal transduction, transcription, transport and cell growth/division. Effective methylation of N-methyl was promoted through N-adenosine-L-methionine cycle II pathway (Zhang, 2014). Molecular mechanism of carbon flux regulation is fixed carbon expression in photosynthesis in different wild-type algae and *Arabidopsis thaliana*. *Arabidopsis thaliana* is used as a model to study the transformation of genetic regulation from vegetative growth to reproductive growth (Khan et al., 2014; Carrieri et al., 2018). The study also proved that both lignified and non-lignified jujube suspension had gene expression, and more up-regulated genes were expressed in lignified jujube suspension.

Summary

By comparing lignified jujube suspension with non-lignified jujube suspension, using GO and KEGG database annotations, the differential functional genes were analyzed from three database systems.

For lignified jujube suspension analysis, 11 genes were up-regulated with lignin gene expression, and there was no up-regulated with non-lignified jujube suspension. The results showed that lignified jujube suspension was related to lignin synthesis. There are 42 genes involved in lignin biosynthesis. They have synergistic effects on Chitinase activity, polysaccharide binding, cell wall tissue, root hair elongation, regulation of carbohydrate biosynthesis, cellulose biosynthesis and meristem growth. The formation of lignified jujube suspension is related to the co-regulation of photosynthesis and endogenous hormones.

REFERENCES

- [1] Allen, D. T., Carrier, D. J., Gong, J. L., Hwang, B. J., Licence, P., Moores, A., Pradeep, T., Sels, B., Subramaniam, B., Tam, M. K. C., Zhang, L., Williams, R. M. (2018): Advancing the Use of Sustainability Metrics in ACS Sustainable Chemistry & Engineering. – ACS Publications 6(1): 1.
- [2] Arro, J., Cuenca, J., Yang, Y. Z., Liang, Z. C., Cousins, P., Zhong, G. Y. (2017): A transcriptome analysis of two grapevine populations segregating for tendril phyllotaxy. – Horticulture Research 4: 17032.
- [3] Astaneh, R. K., Bolandnazar, S., Nahandi, F. Z., Oustan, S. (2018): The Effects of Selenium On some Physiological Traits and K, Na Concentration of Garlic (*Allium Sativum* L.) Under NaCl Stress. – Information Processing in Agriculture 5(1): 156-161.
- [4] Carrieri, D., Lombardi, T., Paddock, T., Cano, M., Goodney, G. A., Nag, A., Old, W., Maness, P.-C., Seibert, M., Ghirardi, Yu, J. P. (2017): Transcriptome and proteome

- analysis of nitrogen starvation responses in *Synechocystis* 6803 Δ glgC, a mutant incapable of glycogen storage. – *Algal Research* 21: 64-75.
- [5] Chen, Z. L., Liu, S. P., Liu, C. H. (2015): Discussion on growth and development of jujube and jujube crane. – *Chinese agronomy bulletin* 31(28): 104-111.
- [6] Chen, H., Wu, Y. G., Yang, Y. Z. (2018): The data assembly of the young leaves and mature leaves of patchouli and the gene function annotation. – *Molecular Plant Breeding* 16(7): 2139-2154.
- [7] Czempl, S., Heppel, S. C., Bogs, J. (2012): R2R3 MYB transcription factors: key regulators of the flavonoid biosynthetic pathway in grapevine. – *Protoplasma* 249: 109-118.
- [8] Ding, Y. H. (2017): High-throughput sequencing of small RNA and functional verification of microRNA in cotton male sterility induced by high temperature stress. – *Huazhong Agricultural University*.
- [9] Dong, Y. Y., Liu, X. M., Yao, N., Li, H. Y. (2017): Transcription factor analysis based on transcriptome data of different tissues of Safflower. – *Chinese herbal medicine* 48(21): 4515-4522.
- [10] Fernandez-Perez, F., Pomar, F., Pedreno, M. A., Novo-Uzal, E. (2015): The suppression of AtPrx52 affects fibers but not xylem lignification in *Arabidopsis* by altering the proportion of syringyl units. – *Physiol Plant* 154: 395-406.
- [11] Jiang, H. G., Wang, S. G., Liu, Y. (2018): Screening and expression analysis of lignin genes in shoots of Cizhu bamboo based on transcriptome sequencing. – *Plant studies* 38(03): 415-421.
- [12] Khan, M. R., Ai, X. Y., Zhang, J. Z. (2014): Genetic regulation of flowering time in annual and perennial plants. – *Wiley Interdiscip. Rev.* 5: 347-359.
- [13] Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S. L. (2013): TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. – *Genome Biology* 14(4): R36.
- [14] Langmead, B., Salzberg, S. L. (2012): Fast gapped-read alignment with bowtie. – *Nature Methods* 9(4): 357-359.
- [15] Li, Q. Y., Zhong, C. L., Jiang, Q. B. (2017): *Molecular Plant Breeding* 15(11): 4396-4404.
- [16] Li, T. Q., Liu, X. F., Wanyou, M. (2017): Transcriptome analysis of wild plant *Rhododendron longstalk* based on high-throughput sequencing. – *Plant studies* 37(06): 825-834.
- [17] Li, Z., Han, C. G. (2018): Economics of Biomass Gasification: A Review of the Current Status. – *Energy Sources Part B Economics Planning & Policy* 13(2): 137-140.
- [18] Lin, X., Pan, X., Kang, X. (2018): Effects of picking time on Lignin Metabolism and related enzyme activities of Chuanqiu okra. – *Zhejiang Agricultural Journal* 09: 1519-1525.
- [19] Liu, Y. L. (2008): Isolation and expression of light-harvesting pigment binding protein gene in *Phyllostachys pubescens* photosystem. – *Chinese Academy of Forestry Sciences*.
- [20] Liu, J. T., Zhu, H. S., Wen, Q. F. (2017): Separation and browning of WRKY transcription factor gene of silk gourd. – *Journal of Agricultural Biotechnology* 25(12): 1950-1960.

- [21] Matus, J. T., Aquea, F., Arce-Johnson, P. (2008): Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes. – *BMC Plant Biol.* 8: 83.
- [22] Mohr, P. G., Cahill, D. M. (2007): Suppression by ABA of salicylic acid and lignin accumulation and the expression of multiple genes, in *Arabidopsis* infected with *Pseudomonas syringae* PV. Tomato. – *Funct Integr Genomics* 7: 181-191.
- [23] Nisavic, M., Stoiljkovic, M., Crnolatac, I., Milosevic, M., Rilak, A., Masnikosa, R. (2018): Highly Water-Soluble Ruthenium(II) Terpyridine Coordination Compounds Form Stable Adducts with Blood-Borne Metal Transporting Proteins. – *Arabian Journal of Chemistry* 11(3): 291-304.
- [24] Niu, J. Q., Su, J. M., Wang, X. M. (2018): Data analysis of *Dendrobium nobile* leaves transcriptome based on high throughput sequencing. – *Molecular Plant Breeding* 16(03): 747-756.
- [25] Qi, Y. X., Liu, Y. B., Rong, W. H. (2011): rRNA-Seq and its applications: a new technology for transcriptomics. – *Yichuan(Hereditas)* 33(11): 1191-1202.
- [26] Qu, Z. Z., Wang, Y. H. (1993): Fruit tree of China. Jujube volume. – China Forestry Publishing House.
- [27] Reich, I., Ijaz, U. Z., Gormally, M., Smith, C. J. (2018): 16s rRNA sequencing reveals likely beneficial core microbes within faecal samples of the eu protected slug *geomalacus maculosus*. – *Scientific Reports* 8(1): 10402.
- [28] Sun, Q. H., Cheng, Z. Y., Qi, H. R. (1996): Observation and report on the growth and fruiting habit of Jinzao jujube crane. – *Shaanxi forestry science and technology* 3: 10-11.
- [29] Tang, Z. J., He, M., Gao, J. P. (2012): Influence of pinching and sprouting on the formation of lignification jujube crane. – *North horticulture* 22: 26-28.
- [30] Wang, S., Yan, C. (2014): Comparison of accumulation capacity of Photosynthate from 2 types of jujube crane in southern China. – *Forestry Science* 50(06): 90-97.
- [31] Wang, S., Yan, C., Deng, B. L. (2014): Comparison of nutrient transport and accumulation ability of two types of jujube crane in southern China. – *Journal of Central South University of Forestry and Technology* 34(12): 35-39.
- [32] Wang, M., Zhang, D. Q., Su, J., Dong, J. W., Tan, S. K. (2018): Assessing Hydrological Effects and Performance of Low Impact Development Practices Based On Future Scenarios Modeling. – *Journal of Cleaner Production* 179: 12-23.
- [33] Wu, X., Song, M., Qiu, P., Li, F., Wang, M., Zheng, J., Wang, Q., Xu, F., Xiao, H. (2018): A Metabolite of Nobiletin, 4 '-Demethylnobiletin and Atorvastatin Synergistically Inhibits Human Colon Cancer Cell Growth by Inducing G0/G1 Cell Cycle Arrest and Apoptosis. – *Food & Function* 9(1): 87-95.
- [34] Yan, C., Wang, S., Shao, F. X. (2013): Comparison of photosynthetic efficiency between leaves of lignification and non lignification of jujube leaves in southern jujube. – *Economic Forest Research* 31(02): 113-117.
- [35] Yang, Y. R., Zhao, J., Liu, M. J. (2007): Research progress of jujube crane. – *Acta agronomica Sinica* S2: 53-57.

- [36] Yang, L., Zhang, W., Xu, Y. T. (2014): Effects of dormant pruning on Germination of ash dates, Jun jujube and date jujube crane. – Xinjiang Agricultural Sciences 51(11): 1984-1989.
- [37] Yang, Y., Liu, H., Qiu, D. (2017): Transcriptome data analysis of *Stellera chamaejasme* based on high-throughput sequencing. – Chinese herbal medicine 48(22): 4740-4747.
- [38] Yuan, C., Fang, P., Zhong, W. J. (2017): Transcriptome sequencing and analysis of *Phellodendron chinensis*. – Chinese herbal medicine 48(21): 4507-4514.
- [39] Zhang, D., Li, W., Li, D. (2009): The relationship between isoflavone content in soybean leaves and relative expression of PAL gene. – Soybean Science 28(4): 670-673.
- [40] Zhang, Y., Zhou, Y., Chen, Q. (2014): Molecular basis of flowering time regulation in *Arabidopsis*. – China. Bull. Bot. 49: 469-482.
- [41] Zhang, S. P., Zhang, S. H., Qiu, S. L. (2018): Based on the transcriptional sequence analysis of the related genes of sunflower green anthocyanin in purple back. – Acta Agriculturae Nucleatae Sinica 32(04): 639-645.
- [42] Zhu, L. S., Liu, G., He, S. E. (2018): Analysis of the transcriptional characteristics of *Eucalyptus grandis* based on Illumina HiSeq 2000 sequencing technology. – Molecular Plant Breeding 13: 1-13.