African Journal of Biotechnology Vol. 10(74), pp. 16769-16773, 23 November, 2011
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.235
ISSN 1684–5315 © 2011 Academic Journals

Full Length Research Paper

Efficient extraction of RNA from various *Camellia* species rich in secondary metabolites for deep transcriptome sequencing and gene expression analysis

Cong Jiang[1#], Qiang Wen[1,2#], Ying Chen[1], Li-an Xu[1]* and Min-ren Huang[1]

[1]The Key Laboratory of Forest Genetics and Biotechnology (Nanjing Forestry University), Ministry of Education, Nanjing 210037, China.
[2]Plant Bio-tech Key Laboratory for Jiangxi Province, Nanchang 330032, China.

Accepted 31 October, 2011

Camellia species, an important economic plants widely distributed in Asia, are recalcitrant to RNA extraction. Here, we developed a method for high quality RNA isolation. Based on the RNA isolated from flower buds, deep transcriptome sequencing of *Camellia oleifera*, *Camellia chekiangoleosa* and *Camellia brevistyla* were successfully carried out. About 600,000 readings produced in a single 454 sequencing run were assembled into 49,909 contigs and 72,877 singlets, 41.06% of which were annotated. Subsequently, chalcone synthase (*Chs*) gene of *C. chekiangoleosa* was first cloned for gene expression analysis. All these works could lay the foundation for future molecular studies of Camellias.

Key words: Camellias, RNA extraction, transcriptome sequencing, gene expression analysis.

INTRODUCTION

Camellias are very important economic plants producing a lot of oil in China (Shanan and Ying, 1982). Among the 238 identified species of Camellia, half can be cultivated for oil production (Zhang, 1981), of which *Camellia oleifera* is widely distributed in the south of China and is known as eastern "olive" universally (Zhang et al., 2008). In addition, the famous woody flowers, such as *Camellia chekiangoleosa* with a lovely red flower, and lots of cultivars were developed from wild Camellia species by selection and breeding in China and Japan (Shanan and Ying, 1982). During the first half of the Eighteenth Century, Camellias have become fashionable ornamentals in the milder zones of the world. In view of an important breeding resource, the development of Camellia germplasm is of great urgency. Now, a molecular biology research has been developed to investigate the genetic resources of Camellia, in order to improve the exploitation and preservation of these species.

In recent years, transcriptome sequencing has become more encompassing, including mRNA transcript-expression analysis (full-length mRNA, expressed sequence tags (ESTs) and ditags), but to select the appropriate input sample (total RNA or mRNA), the quality and quantity is one of the current challenges for researchers (Thomas and Timothy, 2008); so, RNA extraction from specific plant organs and tissues is crucial. The Camellia tissues that present high levels of polysaccharides, polyphenolics and secondary metabolites (Yoshida et al., 1994; Vijayan et al., 2009) are recalcitrant plants for isolation of RNA. So far, there are no reports about the comparison and evaluation of RNA extraction from several species of Camellia and even its use for high-throughput transcript analysis. This study aims to obtain high quality RNA from Camellias for deep transcriptome sequencing, although, it initially attempts to clone the

*Corresponding author. E-mail: laxu@njfu.edu.cn. Fax: +86-25-8542-7882.

Abbreviations: CTAB, Cetyltrimethyammonium bromide; **DEPC,** diethyl pyrocarbonate; **PVP,** polyvinylpyrrolidone; **RACE,** rapid amplification of cDNA ends; **RT-PCR,** reverse transcription-polymerase chain reaction.

#These authors contributed equally to this works.

valuable genes, and even made a foundation for conservation of germplasm resource of Camellias.

MATERIALS AND METHODS

Camellia species used for RNA extraction were grown in Camellia Germplasm Bank of Jiangxi Academy of Forestry, including two clones of *C. oleifera*, *C. chekiangoleosa*, *Camellia brevistyla* and *Camellia cuspidata*. The tabacco, Chinese tulip tree, Chinese red pine and the tissue-cultured poplar were obtained in the experimental fields of Nanjing Forestry University in November and the tissues, which were healthy, were completely differentiated. Samples were snap-frozen in liquid nitrogen and stored at -80°C.

RNA extraction protocol

Fresh plant tissue (0.3 g) was ground to a very fine powder in liquid nitrogen, using a precooled (with liquid nitrogen) mortar and pestle. The powder was transferred to 500 ul extraction buffer [100 mM Tris-HCl (PH 8.0), 2.0 M NaCl, 25 mM EDTA, 2% (w/v) CTAB, 2% (w/v) PVP, and addition of 3% (v/v) β-mercaptoethanol just prior to use], incubated at 65°C for 30 min, and was vigorously shaken every 8 to 10 min. After an equal volume of chloroform was extracted, the upper aqueous phase was transferred to a new tube. Ethanol was added slowly to the final concentration of 20% (v/v) and precipitated for 30 min, after which the supernatant was transferred to another tube. One-third of the volume of 10 M LiCl was added and precipitated for 4 h at 4°C, while the tube was centrifuged at 14000 g for 20 min at 4°C. Then, the supernatant was discarded and the spare pellets were resuspended in 250 µl of SSTE buffer [1 M NaCl, 0.5% (w/v) SDS, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA] prewarmed. The suspension was transferred to a new tube and extracted with PCI and chloroform, and then the supernatant was transferred to 2 volumes of precooled absolute ethanol, while the RNA was precipitated at -20°C for 2 h. After the tube was centrifuged at 14000 g and 4°C for 20 min, the pellets were washed twice with 1 ml of precooled 70% ethanol, by centrifuging at 10000 g and 4°C for 2 min. Dried at room temperature, the pellets were dissolved in 50 ul DEPC-treated water. Then, the rough RNA was treated with DNase(Takara Bio Inc., Japan). RNA yield was measured with the ND 1000 UV-Vis Spectrophotometer (NanoDrop Technologies, New Zealand) and was checked by electrophoresis on 1% agarose TAE gels.

Deep transcriptome sequencing

First-strand cDNA was produced using the RNA extracted from flower buds by Clontech's SMART cDNA synthesis kit. Then, Advantage 2 PCR Kits (Clontech) were used for ds cDNA amplification. In this step, the optimal number of PCR cycles was determined using the optimization tube. PureLink™ PCR Purification Kit (Invitrogen) was used for cDNA purification. Deep transcriptome sequencing of *C. oleifera*, *C. chekiangoleosa* and *C. brevistyla* was carried out by 454 GS FLX. For gene name annotation, the assembled sequences were compared against nr protein sequence databases, using blastx with a significant threshold of E ≤ 10-5 (Hahn et al., 2009).

Full length cDNA cloning and gene expression

Full-length cDNA, encoding Chalcone synthase (*Chs*) gene, was cloned by RACE using 5'-Full RACE Kit (Takara) and 3'-Full RACE Core Set Ver.2.0 (Takara). For semi-quantitative RT-PCR, primers were first designed based on the results of transcriptome sequencing and RACE, in order to amplify the three commonly-used genes, including the housekeeping genes: *Actin* and the two specific expression genes: Fatty acid desaturase 2 (*Fad2*) and *Chs*. Then, 1 µg of the total RNAs, extracted from the flower buds of two clones of *C. oleifera* (*C. oleifera*1 and *C. oleifera*2), *C. chekiangoleosa, C. brevistyla* and *C. cuspidate*, was reversely transcribed with reverse-transcriptase (promega) at 25°C for 5 min initially, and then at 42°C for 60 min, before it was finally transcribed at 70°C for 15 min. Furthermore, the housekeeping gene, *Actin*, was selected as the internal control (Thellin et al., 2003).

RESULTS AND DISCUSSION

RNA isolation

RNA, with high quality, is essential for transcriptome sequencing. Due to the features of Camellia materials as a recalcitrant plant, we switched over to CTAB methods and tested a lot of modification. According to the feature of phenolic compounds, we tried to use 25 mM EDTA, 3% mercaptoethano and 2% polyvinyl pyrrolidone that contributed to the prevention of polyphenolic oxidation and a removal of polyphenolic (Wang and Vodkin, 1994). To solve the problem of polysaccharides, several measures were also taken. The higher concentration of Na^+ of the extraction buffer would increase the solubility of polysaccharides (Chang et al., 1993), thus reducing their coprecipitation with RNA in later steps of the protocol. However, low concentration of ethanol was used to precipitate the polysaccharides, while Li^+ was used to precipitate RNA differentially from polysaccharides and DNA. SDS extraction was added after LiCl precipitation, which was necessary for the removal of contaminants, organic molecules and residual Li^+. By means of that method, a trial was made to extract the RNA from different organs of *C. oleifera*, several species of Camellia and other plant tissues including the tissue culture of poplar, tobacco, Chinese tulip tree and Chinese red pine in just one day, and was used to obtain good results. Due to the differences in the effective material, the amount of RNA extraction was slightly different. In sum, the average yield was about 200 ug RNA g^{-1} fresh tissues of Camellias. The A260/A280 absorbance ratios were 1.96 to 2.12, and the A260/A230 was always above 2.0 (Table 1). The integrities of RNA, extracted from the flower buds of four Camellias, were tested on 1.0% agarose gel electrophoresis, which showed clear, discrete rRNA with no apparent degradation (Figure 1). The results showed that the study's protocol was adaptive and efficient.

Transcriptome sequencing

Nowadays, the sequencing technology is widely used, due to the fact that the large-scale sequencing is a new method for gene expression, identification and cloning of novel genes. Before transcriptome sequencing, cDNA

Table 1. Absorbance ratio and yield of RNA extraction from several organs of four species of Camellia, as well as other plant species.

S/N	Species	Organ	Absorbance ratio		Yield (μg/g)
			A260/A230	A260/A280	
1	*C. oleifera*	Flower buds	2.02±0.08	2.06±0.04	174±37
2	*C. oleifera*	Leave buds	2.20±0.2	2.03±0.05	320±45
3	*C. oleifera*	Stems	1.92±0.3	2.00±0.03	102±25
4	*C. oleifera*	Roots	1.55±0.3	1.96±0.04	98±15
5	*C. chekiangoleosa*	Flower buds	2.51±0.04	2.04±0.01	392±35
6	*C. brevistyla*	Flower buds	2.71±0.08	2.04±0.02	254±30
7	*C. cuspidata*	Flower buds	2.22±0.02	2.07±0.01	142±18
8	Chinese Tulip Tree	Roots	2.04±0.2	1.98±0.04	180±20
9	Chinese Tulip Tree	Leaves	2.26±0.3	2.00±0.06	265±40
10	Tobacco	Leaves	2.43±0.1	2.05±0.1	424±45
11	Tobacco	flower buds	2.48±0.1	2.07±0.03	377±72
12	Chinese red pine	Leaves	2.34±0.1	2.12±0.05	220±35
13	Tissue-cultured Poplar	Leaves	2.56±0.1	2.02±0.03	89±20

Each value is indicated by the mean of 5 (n = 5), replicates recorded ±SE.

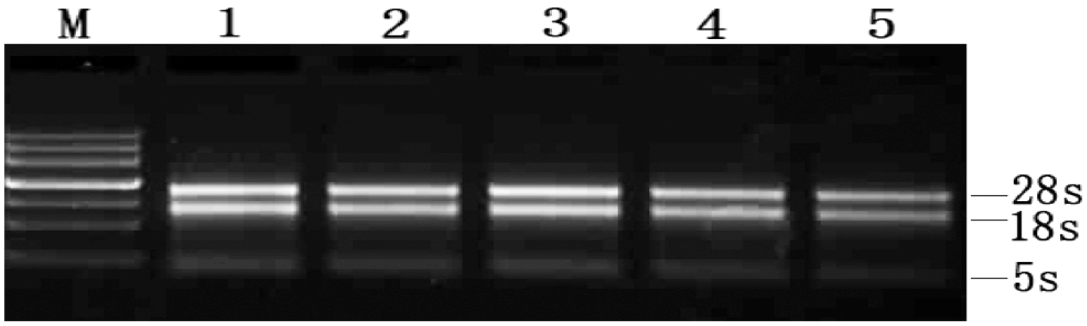

Figure 1. Agarose gel electrophoresis of RNA extracted from the flower buds of four Camellia species. M: maker (tiangen); 1 and 2: the clones of *C. oleifera* (*C. oleifera1* and *C. Oleifera2*); 3: *C. chekiangoleosa*; 4: *C. brevistyla*; 5: *C. cuspidata.*

should be synthesized from the total RNA; then, deep transcriptome sequencing of *C. oleifera*, *C. chekiangoleosa* and *C. brevistyla* could be completed using 454 GS FLX. However, 217,996, 219,906 and 154,597 readings of *C. oleifera, C. chekiangoleosa* and *C. brevistyla*, respectively were generated in just a run. After removal of the adaptor sequence and low complexity sequence, 182,766, 190,545 and 132,147 readings of *C. oleifera*, *C. chekiangoleosa* and *C. brevistyla*, respectively, with high quality, remained. These readings were assembled into 49,909 contigs [*C. oleifera* (15,733), *C. chekiangoleosa* (19,397) and *C. brevistyla* (14,779)] and 72,877 singlets [*C. oleifera* (20,606), *C. chekiangoleosa* (26,882) and *C. brevistyla* (25,389)]. The N50 of the three species were 487, 504 and 542 bp, which showed little difference. Above all, 26,005 contigs [*C. oleifera* (7,786), *C. chekiangoleosa* (9,732) and *C. brevistyla* (8,487)] and 24,411 singlets [*C. oleifera* (6,306), *C. chekiangoleosa* (8,041) and *C. brevistyla* (9,064)] were annotated using BLAST searches (E-value ≤ 1e-5) against the Nr database. A comparison of this study with recently published results, which employed 454 sequencing by Cheung et al. (2006) and Vera et al. (2008), showed that in this study, the proportion for high-quality readings was about 90%, which was superior to their reports. As such, the readings per sequencing run produced were also 2 to 5 times more than those results. Based on the information, we could estimate gene expression in different species, develop genetic markers, and even detect and validate SNP.

Gene expression analysis

In the process of research on an economic plant, it is meaningful to find out the molecular mechanisms of biological features, especially the expression of some crucial genes in different tissues and different

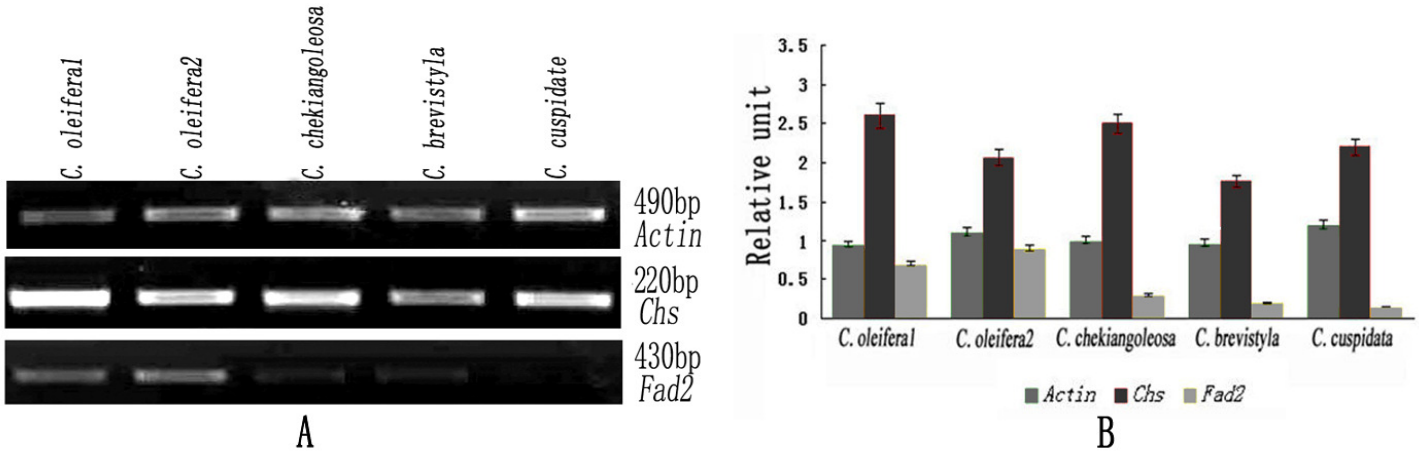

Figure 2. Expression analysis of *Chs* and *Fad2* in flower buds of four Camellia species by semi-quantitative PCR. A: the RT-PCR results of *C. oleifera1, C. oleifera2, C. chekiangoleosa, C. brevistyla, C. cuspidate*, followed by Actin, *Chs,* Fad2-PCR amplification; B: Their relative units in different results of RT-PCR were shown.

developmental stages. Flavonoids play important roles in flower color formation, pollen development, auxin transportation, UV damage prevention, disease resistance and stress response in plant. The *Chs* gene catalyzed the first committed step in the biosynthesis pathway of it (Hans et al., 1986). The *Chs* gene is relatively conservative, and it could be crucial for it to hold functions during the process of evolution. Based on the results of transcriptome sequencing, the *Chs* gene of *C. chekiangoleosa* was cloned by RACE. The full length of the gene was 1469bp (GenBank: HQ704701.1). The full length of the gene was 1469 bp, and it predicted an open reading frame of 1167 bp corresponding to 389 amino acids, flanked by 125 bp of 5'-untranslated sequence and 3'-untranslated sequence of 117 bp. This was the first report that described the isolation of a gene from *C. chekiangoleosa*. A comparison of sequencing results and *Chs* sequences of *Camellia sinensis*, embodied by GenBank, showed that the homology was more than 90%. It preliminarily confirmed that the *Chs* gene was a functional gene, whose sequen-ces were conserved among Camellias.

As another important gene, fatty acid desaturase 2 (*Fad2*) is involved in the conversion of oleic acid into linoleic acid in plant and is associated with the quality of the oil (Tao et al., 2006). We analyzed the expression of the *Chs* and *Fad2*. In the results (Figure 2), using the *Actin* as a control, the expression of *Chs* was higher than *Actin*, while the expression of *Fad2* was relatively lower. Meanwhile, the expression of the same gene in different species was not always the same. According to the oil content of those species, the expression of the *Fad2* gene of *C. oleifera*1 and *C. oleifera*2 was higher than the other three species of Camellia, which indicated that *Fad2* may participate in the process of oil synthesis, while *Chs* gene expression in different species were more or less uniform, even if the flower color of these species were not the same (the flower color of *C. chekiangoleosa* was red, while the others were white). We could not find the obvious connection between the *Chs* gene cloned by us and the flower color, even if the *Chs* gene is essential to the flavonoids synthesis.

ACKNOWLEDEGMENTS

This research was supported by the National Department Public Benefit Research Foundation of China (No. 201204413), and the Jiangxi Provincial Natural Science Foundation of China (No.2009GQN0047). And, we thank Dr. Bodenes (INRA France), who provided valuable opinions and revised the manuscript.

REFERENCES

Chang S, Puryear J, Caimey J (1993). A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Rep. 11: 113–116.
Cheung F, Haas BJ, Goldberg SM, May GD, Xiao Y, Town CD (2006). Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology. BMC Genomics, 7: p. 272.
Hahn DA, Ragland GJ, Shoemaker DD, Denlinger DL (2009). Gene discovery using massively parallel pyrosequencing to develop ESTs for the flesh fly *Sarcophaga crassipalpis*. BMC Genomics, 10: p. 234.
Hans S, Heinz S (1986). Structure of the chalcone synthase gene of Antirrhinum majus. Mol. Gen. Genet. 202: 429-434.
Shanan H, Ying G (1982). The comprehensive utilization of camellia fruits. Am. Camellia Yearbk. 37: 104–107.
Tao F, Zhu SW, Fan J, Cheng BJ (2006). Cloning and Sequence Analysis of Maize FAD2 Gene. J. Plant Physiol. Mol. Biol. 32(6): 649-656.
Thellin O, Zorzi W, Lakayec B, Bormanc D, Coumans B, Hennen G, Grisar T, Igout A, Heinen E (2003). Housekeeping genes as internal standards: use and limits. Plant Mol. Biol. 52: 1025-1036.
Thomas J, Timothy H (2008). Transcriptome sequencing with the Genome Sequencer FLX system. Nat. Methods, 5: 6-8.
Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH (2008). Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol. Ecol. 17(7):

1636-1647.
Vijayan K, Zhang WJ, Tsou CH (2009). Molecular taxonomy of Camellia (Theaceae) inferred from nrITS sequences. Am. J. Bot. 96: 1348-1360.
Wang CS, Vodkin LO (1994). Extraction of RNA from tissues containing high levels of procyanidins that bind RNA. Plant Mol. Biol. Rep. 12: 132-145.
Yoshida T, Nakazawa T, Hatano T, Yang RC, Yang LL, Yen KY, Okuda T (1994). A dimeric hydrolysable tannin from *Camellia oleifera*. Phytochemist. 37: 241-244.
Zhang HD (1981). A taxonomy of the genus Camellia. Sun Yatsen University Forum, 1: 52-160.
Zhang DL, Stack L, Zhang RQ, Yu JF, Xie BX, Chen YZ, Ruter JM (2008). Teaoil camellia - eastern "olive" for the world. Acta. Hortic. 769: 43-48.