Gene 547 (2014) 126–135

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Transcriptome profiling and digital gene expression analysis of *Fallopia multiflora* to discover putative genes involved in the biosynthesis of 2,3,5,4′-tetrahydroxy stilbene-2-O-β-D-glucoside

Wei Zhao [a,b], Wanxia Xia [a,b], Jiewen Li [b], Shujing Sheng [b], Lei Lei [a,b], Shujing Zhao [a,b,*]

[a] School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510640, People's Republic of China
[b] Department of Pharmacy, General Hospital of Guangzhou Military Command, Guangzhou 510010, People's Republic of China

ARTICLE INFO

Article history:
Received 26 April 2014
Received in revised form 19 June 2014
Accepted 20 June 2014
Available online 24 June 2014

Keywords:
Biosynthesis
Digital gene expression
Fallopia multiflora
Transcriptome profiling
2,3,5,4′-Tetrahydroxy stilbene-2-O-β-D-glucoside

ABSTRACT

The compound 2,3,5,4′-tetrahydroxy stilbene-2-O-ß-D-glucoside (THSG) synthesized by *Fallopia multiflora* (*F. multiflora*) exhibits pharmacological potency. However, the mechanistic details of its biosynthesis pathway are still vague. To clear this ambiguity, we performed de novo transcriptome assembly and digital gene expression (DGE) profiling analyses of *F. multiflora* using the Illumina RNA-seq system. RNA-seq generated approximately 70 million high-quality reads that were assembled into 65,653 unigenes (mean length = 750 bp), including 26,670 clusters and 38,983 singletons. A total of 48,173 (73.4%) unigenes were annotated using public protein databases with a cut-off e-value above 10^{-5}. Furthermore, we investigated the transcriptome difference of four different *F. multiflora* tissues using DGE profiling. Variations in gene expression were identified based on comparisons of transcriptomes from various parts of a high-level THSG- and a low-level THSG-producing *F. multiflora* plant. Clusters with similar differential expression patterns and enriched metabolic pathways with regard to the differentially expressed genes putatively involved in THSG biosynthesis were revealed for the first time. Our data provides the most comprehensive sequence resource regarding *F. multiflora* so far. Taken together, the results of this study considerably extend the knowledge on THSG production.

1. Introduction

Fallopia multiflora (*F. multiflora*) has been used in traditional Chinese medicine for thousands of years. Recent research showed that *F. multiflora* (i) has potent antioxidative and cytoprotective effects (Steele et al., 2013), (ii) has enhanced purgative properties and promotes diuresis as well as choleretic effects (Xie et al., 2012), and (iii) shows a neuroprotective effect against glutamate-induced neurotoxicity (Jang et al., 2013). The main active ingredients of *F. multiflora* include anthraquinone, phospholipids, and stilbenes, among other compounds. Regarding stilbenes, 2,3,5,4′-tetrahydroxy stilbene-2-O-ß-D-glucoside (THSG) is an important active component in *F. multiflora* which is used as a chemical marker for the identification of *F. multiflora* quality (Committee, 2010).

Pharmacological research revealed that THSG plays a vital role on maintaining human health by exhibiting antihyperlipidemic (Wang et al., 2012), anti-oxidative, anti-inflammatory and endothelial-protective activities (Lv et al., 2011). Furthermore, THSG can protect osteoblastic MC3T3-E1 cells by inhibiting the release of bone-resorbing mediators and oxidative damage of these cells (Zhang et al., 2012), and suppress atherosclerosis by altering the expression of key proteins that may be novel molecular targets responsible for this pathological condition (Yao et al., 2013). Sun et al. reported that THSG might provide a potentially new strategy for preventing and treating neurodegenerative disorders such as Parkinson's disease and also demonstrated that THSG may protect neurons against 1-methyl-4-phenylpyridinium (MPP+)-induced cell death through the improvement of mitochondrial function, decrease of oxidative stress and inhibition of apoptosis (Sun et al., 2011).

Thus far, little is known about the molecular mechanism of the biosynthesis of THSG in *F. multiflora*. Our previous study showed that precursor feeding of methyl jasmonate and salicylic acid in suspension cultures of *F. multiflora* could increase THSG production (Shao et al., 2012), and the stilbene synthase gene *FmPKS* was isolated from the rhizomes of *F. multiflora*. The expression level of *FmPKS* depends on the THSG content in different tissues, but transgenic expression in *Arabidopsis thaliana* was not detectable (Sheng, 2010). We also constructed suppression subtractive hybridization (SSH) libraries to screen for possible genes involved in THSG biosynthesis, but only twelve non-redundant differentially expressed sequence tags were obtained (Zhao

Abbreviations: DGE, digital gene expression tag profiling; ESTs, expressed sequence tags; *F. multiflora*, *Fallopia multiflora*; SSH, suppression subtractive hybridization; THSG, 2,3,5,4′-tetrahydroxy stilbene-2-O-ß-D-glucoside.

* Corresponding author at: Department of Pharmacy, General Hospital of Guangzhou Military Command, No. 111, Liuhua Road, Yuexiu District, Guangzhou 510010, People's Republic of China.
E-mail addresses: zhaoweih2@163.com (W. Zhao), gzzsjzhs@163.com (S. Zhao).

http://dx.doi.org/10.1016/j.gene.2014.06.041
0378-1119/

et al., 2014). Therefore, this attempt did not clarify the mechanisms of the THSG biosynthesis pathway.

In recent years, next generation sequencing based genome-wide gene expression profiling techniques, such as RNA-seq and Digital Gene Expression (DGE), have dramatically improved the efficiency and speed of gene discovery. As transcriptome analysis is essential to interpret the functional elements of a given genome in a particular period, transcriptome comparison from the same sample of different periods or states can reveal differential genes that cause the sample phenotypic differences comprehensively. Based on genome-wide expression profiles by sequencing, DGE is a powerful tool to identify and quantify gene expression on the whole genome level.

Pursuing the objective to identify novel genes involved in THSG biosynthesis in our study, we performed the first global analysis of *F. multiflora* transcriptomes using Illumina HiSeq™ 2000. Almost six billion clean nucleotides were generated, including 123,762 contigs and 65,653 unigenes. Furthermore, we compared the gene expression profiles of four different *F. multiflora* tissues (root, stem and leaf) using the DGE approach. The assembled, annotated transcriptome sequences and gene expression profiles provide an invaluable resource for the identification of *F. multiflora* genes involved in THSG biosynthesis and tissue specific functions.

2. Materials and methods

2.1. Plant materials and preparation

Deqing City (Guangdong Province) has been regarded as the authentic origin of *F. multiflora*. It is characterized by an average concentration of 2% (weight/weight) of THSG in its roots. An investigation on how the geographic origin of *F. multiflora* specimen correlates with their THSG content in roots revealed that isolates from Chongqing contain the lowest amount, generally below 0.1‰ (our preliminary data; the article is under review.). We selected *F. multiflora* isolated from these two areas as research material for our study. *F. multiflora* roots from the two origins were used as a basis for transcriptome sequencing (TrCD), and Deqing *F. multiflora* roots (Dr), stems (Ds), leaves (Dl) and Chongqing *F. multiflora* roots (Cr) were used for DGE analysis.

F. multiflora plants were collected in March 2010 and were subsequently maintained in the medicinal plant garden of the Department of Pharmacy, Guangzhou Liuhuaqiao Hospital, Guangzhou, China. As the samples were gathered in wild field, it was not possible to assess the age of the plants. THSG content was determined by HPLC using a Dikma Diamonsil C18 column (250 mm × 4.6 mm, tablet path 5 μm), according to Sheng's method (Sheng et al., 2010).

2.2. RNA extraction

Total RNA was extracted from *F. multiflora* tissues using the Plant Total RNA Isolation kit (BioTeke, P. R. China) and treated with DNase I (TaKaRa, Dalian, P. R. China). Ethidium bromide (EtBr) staining, agarose gel electrophoresis and spectrophotometric (NanoDrop 2000, USA) analysis were performed to examine the quality and concentration of total RNA.

2.3. cDNA library preparation and transcriptome sequencing

Total RNA collected from Deqing *F. multiflora* roots and Chongqing *F. multiflora* roots were mixed equally for RNA-seq. mRNA was purified from total RNA (60 μg) using oligo (dT) magnetic beads. Fragmentation buffer was added for cleavage of mRNA into short fragments. Taking these short fragments as templates, a random hexamer-primer solution was used to synthesize the first-strand cDNA. The second-strand cDNA was synthesized using an appropriate buffer, dNTPs, RNaseH and DNA polymerase I, respectively. Short fragments were purified with QiaQuick PCR extraction kit and resolved with EB buffer for end extension and tailing A. Subsequently, the short cDNA fragments were ligated to sequencing adapters. After separation by agarose gel electrophoresis the suitable fragments were selected as templates for the PCR amplification. Finally, the complete library was sequenced using Illumina HiSeq™ 2000.

2.4. Analysis of transcriptome sequencing results

Image data output from the sequencing machine was transformed by base calling into sequence data, which is termed 'raw data' or 'raw reads'. Image deconvolution and quality value calculations were performed using the Illumina GA pipeline v1.6. Subsequently, the raw reads were cleaned by removing reads with adaptors, reads with unknown nucleotides higher than 5% and low quality reads. Transcriptome de novo assembly was carried out by applying a program managing the assembly of short reads, Trinity (Grabherr et al., 2011). First, Trinity combines reads with certain minimum length of overlap to form longer fragments, which are termed contigs. Then the reads are mapped back to contigs; with paired-end reads the program is able to detect contigs from the same transcript as well as the distances between these contigs. Finally, Trinity connects the contigs resulting in sequences that cannot be further extended on either end. Such sequences are defined as 'unigenes'. By family clustering, the unigenes will be divided into two classes. One class contains clusters that are denoted with the prefix 'CL' followed by the cluster identifier. In one cluster, there are several unigenes which display similarity values higher than 70%. The other class contains singletons that are denoted with the prefix 'unigene'. In the final step, BLASTx alignment analysis (e-value < 0.00001) between unigenes and protein database searches using nr, Swiss-Prot, KEGG and COG is performed. The most significant alignment results are used to decide the sequence direction of unigenes. If results of different databases conflict with each other, a priority order of nr, Swiss-Prot, KEGG and COG should be followed when deciding the sequence direction of unigenes. When a unigene happened to be aligned to none of the databases mentioned above, ESTScan (Iseli et al., 1999) was applied to decide its sequence direction.

Unigene annotation provides valuable information of its expression and functional annotation. This information consists of protein function annotation, COG functional annotation and Gene Ontology (GO) functional annotation of unigenes. Unigene sequences are first aligned by BLASTx to protein databases like nr, Swiss-Prot, KEGG and COG (e-value < 0.00001) and aligned by BLASTn to the nucleotide databases nt (e-value < 0.00001), thereby retrieving proteins with the highest sequence similarity.

2.5. DGE library construction and sequencing

Tag library construction for the four *F. multiflora* samples (Dr, Ds, Dl, Cr) was performed in parallel using the Illumina Gene Expression Sample Prep Kit. In short, mRNA was isolated from total RNA using magnetic oligo(dT) beads. Then the first and second-strand cDNA were synthesized using oligo(dT) as primer. Bead-bound cDNA was subsequently digested with the restriction enzyme *Nla*III, which recognizes and cleaves the CATG sites. The cDNA fragments with the sticky 3′ ends were then purified with the help of magnetic beads. The Illumina adapter 1 was added to their 5′ ends. The junction between Illumina adaptor 1 and the CATG site is a recognition site for *Mme*I, which is an endonuclease with separated recognition and digestion sites. This enzyme cuts 17 bp downstream of the CATG site, producing tags containing adaptor sequence 1. After removing the 3′ fragments by magnetic bead precipitation, Illumina adaptor 2 is ligated to the 3′ ends of the tags, generating tags with different adaptors at both ends to form a tag library. After 15 cycles of linear PCR amplification, 105 bp fragments are purified by using 6% TBE polyacrylamide gel electrophoresis. After denaturation, the single-chain molecules are fixed onto the Illumina sequencing chip. Each molecule is extended into a single-molecule cluster