Gene 527 (2013) 131–138

Contents lists available at SciVerse ScienceDirect

Gene

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

Analysis of the *Dendrobium officinale* transcriptome reveals putative alkaloid biosynthetic genes and genetic markers

Xu Guo [a], Ying Li [a], Chunfang Li [a], Hongmei Luo [a], Lizhi Wang [a], Jun Qian [a], Xiang Luo [a], Li Xiang [a], Jingyuan Song [a], Chao Sun [a], Haibin Xu [a], Hui Yao [a], Shilin Chen [a,b,*]

[a] *Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China*
[b] *Institute of Chinese Materia Medica, Academy of Chinese Medical Sciences, Beijing 100700, China*

ARTICLE INFO

Article history:
Accepted 27 May 2013
Available online 10 June 2013

Keywords:
Dendrobium officinale Kimura et Migo (*Orchidaceae*)
Transcriptome
454-ESTs
Alkaloid biosynthesis
Simple sequence repeats

ABSTRACT

Dendrobium officinale Kimura et Migo (*Orchidaceae*) is a traditional Chinese medicinal plant. The stem contains an alkaloid that is the primary bioactive component. However, the details of alkaloid biosynthesis have not been effectively explored because of the limited number of expressed sequence tags (ESTs) available in GenBank. In this study, we analyzed RNA isolated from the stem of *D. officinale* using a single half-run on the Roche 454 GS FLX Titanium platform to generate 553,084 ESTs with an average length of 417 bases. The ESTs were assembled into 36,407 unique putative transcripts. A total of 69.97% of the unique sequences were annotated, and a detailed view of alkaloid biosynthesis was obtained. Functional assignment based on Kyoto Encyclopedia of Genes and Genomes (KEGG) terms revealed 69 unique sequences representing 25 genes involved in alkaloid backbone biosynthesis. A series of qRT-PCR experiments confirmed that the expression levels of 5 key enzyme-encoding genes involved in alkaloid biosynthesis are greater in the leaves of *D. officinale* than in the stems. Cytochrome P450s, aminotransferases, methyltransferases, multidrug resistance protein (MDR) transporters and transcription factors were screened for possible involvement in alkaloid biosynthesis. Furthermore, a total of 1061 simple sequence repeat motifs (SSR) were detected from 36,407 unigenes. Dinucleotide repeats were the most abundant repeat type. Of these, 179 genes were associated with a metabolic pathway in KEGG. This study is the first to produce a large volume of transcriptome data from *D. officinale*. It extends the foundation to facilitate gene discovery in *D. officinale* and provides an important resource for the molecular genetic and functional genomic studies in this species.

1. Introduction

The stem of the plant *Dendrobium officinale* Kimura et Migo (*Orchidaceae*) is used in traditional Chinese medicine and has been used as an herbal medicine in many Asian countries for hundreds of years. The major medicinal components include alkaloids, polysaccharides, amino acids, and several trace mineral elements (Di, 2003; Li et al., 2011). Early work focused on characterizing alkaloid components; alkaloids have been successfully isolated, and their structures have been confirmed. The total alkaloid content of *D. officinale* has been measured at 0.02%. Although this is lower than the total alkaloid content of the related species *Dendrobium nobile*, *D. officinale* has been demonstrated to produce higher quality alkaloids (Chen et al., 2006). Currently, alkaloids are generated either by extraction (Morita et al., 2000) or by chemical synthesis (Kreis and Carreira, 2012), both processes that are inefficient and result in low yields. High market demand has led to the excessive harvesting and exploitation of *D. officinale*. Therefore, a biotechnology-based approach capable of producing stable alkaloids in large quantities is a promising strategy to meet market demand.

D. officinale alkaloids belong to the terpenoid indole alkaloid (TIA) class. The upstream biosynthetic pathways for all TIA products are conserved among alkaloid-producing plants and involve the generation of a strictosidine backbone (Supplementary Fig. 1). Although the various chemical and pharmacological properties of *D. officinale* alkaloids have been extensively studied, the biosynthetic pathways by which they are generated remain poorly understood. One method to improve our understanding of alkaloid biosynthetic pathways is

Abbreviations: BLAST, Basic Local Alignment Search Tool; CYP450, cytochrome P450-dependent monooxygenase; EC, enzyme commission; ESTs, expressed sequence tags; *G10H*, geraniol 10-hydroxylase; GO, Gene Ontology; HQ, high-quality; KEGG, Kyoto Encyclopedia of Genes and Genomes; MDR, multidrug resistance protein; MEP, methylerythritol 4-phosphate; MVA, mevalonic acid; NCBI, National Center for Biotechnology Information; NGS, next-generation sequencing; Nr, NCBI non-redundant protein; Nt, NCBI non-redundant nucleotide; qRT-PCR, quantitative real-time polymerase chain reaction; *SCS*, secologanin synthase; SSR, simple sequence repeat; *STR*, strictosidine synthase; TAIR, The Arabidopsis Information Resource; *TDC*, tryptophan decarboxylase; TIA, terpenoid indole alkaloid; TFs, transcription factors; *TSB*, tryptophan synthase.

* Corresponding author at: No.151, Malianwa North Road, HaiDian District, Beijing 100193, China. Tel.: +86 10 62899700; fax: +86 10 62899776.
E-mail address: slchen@implad.ac.cn (S. Chen).*URL:* http://www.implad.ac.cn (S. Chen).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.05.073

through the generation and analysis of expressed sequence tags (ESTs), a powerful tool in novel gene discovery and prediction (Li et al., 2010; Sun et al., 2010). Currently, only 800 *D. officinale* ESTs have been archived in the GenBank database. However, improvements in next-generation sequencing (NGS) technologies, such as 454 pyrosequencing, have made sequencing cheaper and faster and have made obtaining large EST datasets a viable option (Morozova et al., 2009; Simon et al., 2009). The 454-ESTs have been wildly used to refine predicted transcripts for genes, especially in non-model plants for which genomic sequence data are unavailable. For example, during chilling requirement fulfillment were analyzed by mining 454-ESTs from the flower buds of tree peony (Gai et al., 2011). In addition, 454 pyrosequencing allows the detection of simple sequence repeats (SSRs), also termed microsatellites, which are tandem repeats of two to six base-pair nucleotide motifs. It is likely that SSRs from ESTs (EST-SSRs) can provide a wealth of molecular markers for genetic breeding and can be used to establish physical genetic maps (Varshney et al., 2005).

Here, we present the first results of a study designed to characterize the transcriptome of the stems of *D. officinale* plants using NGS technology based on the 454 GS FLX Titanium platform. We provide an overview of the *D. officinale* stem transcriptome and discover candidate genes encoding enzymes involved in alkaloid biosynthesis. We additionally report the design of a set of SSRs that will facilitate the marker-assisted breeding of *D. officinale*.

2. Materials and methods

2.1. Plant material

D. officinale Kimura et Migo (*Orchidaceae*) plants were artificially cultivated and collected from the cultivation base of Jinhua, Zhejiang Province, China. The plant material was identified by Dr. Yulin Lin (Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, China), and *D. officinale* was obtained by vegetative propagation methods. Stem tissue was cut into small pieces and immediately frozen in liquid nitrogen and stored at −80 °C until required for further processing.

2.2. RNA preparation

Total RNA was extracted using the Plant RNA Isolation Mini Kit (BioTeke, Beijing, China). The RNA samples were treated with DNase I (TURBO DNase; Ambion, USA) to remove genomic DNA. The integrity of the total RNA was determined using agarose gel electrophoresis. Gels were stained with 1% ethidium bromide (EtBr-stained), and the RNA concentration was measured using a GeneQuant100 spectrophotometer (GE Healthcare, UK).

2.3. cDNA synthesis and 454 pyrosequencing

First-strand cDNA was produced from 2.1 μg of total RNA isolated from the stems of *D. officinale* according to the procedure provided with the Clontech SMART cDNA synthesis kit (Clontech, USA), with slight changes as described in our previous study (Sun et al., 2010). To remove long poly(A/T) tails from cDNA sequences, we designed a modified poly(T) primer (5′-AAG CAG TGG TAT CAA CGC AGT GCA GT(20)VN-3′) with a *Bsg*I site between the adapter and the poly(T) tail. Both the poly(T) primer and the Clontech SMART IV primer were used to synthesize first-strand cDNA. Double-stranded cDNA was synthesized using PCR Advantage II polymerase (Clontech, USA) with the following thermal profile: 1 min at 95 °C followed by 19 cycles of 95 °C for 15 s, 65 °C for 30 s, and 68 °C for 6 min. After amplification efficiency and quality were determined with a 1% agarose gel, the ds cDNA was purified using the PureLink TMPCR purification kit (Invitrogen, USA). The purified cDNA was treated overnight with *Bsg*I (NEB, USA) at 37 °C and recovered using the QIAquick PCR Purification Kit (Qiagen, USA). Finally, a total of 9 μg of ds cDNA was used for pyrosequencing with the GS FLX Titanium Kit.

2.4. 454 EST assembly and annotation

The resulting 454 raw sequence reads with weak signal or low quality were screened and trimmed by GS FLX pyrosequencing software to yield a final data set comprised of high-quality (HQ) (>99.5% accuracy of single-base reads) sequences. Prior to assembly, primer and adapter sequences were trimmed from the HQ dataset, and sequences shorter than 50 bp were removed. The remaining data were assembled into unique sequences (including contigs and singletons) using 454 GS *De Novo* Assembler software v2.6 (454 Life Sciences, Roche).

Similarity searches were carried out against the NCBI non-redundant nucleotide (Nt) database using the BLASTN algorithm with an *E*-value cut-off of 10^{-5} to find and remove ribosomal RNA sequences (Altschul et al., 1990). The remaining sequences were searched for/against public databases, including the *Arabidopsis* protein database at The Arabidopsis Information Resource (TAIR; http://www.arabidopsis.org) (version Tair10), the SwissProt protein database (http://www.expasy.ch/sprot; released on 03/04/2012), and the NCBI non-redundant protein (Nr) database (http://www.ncbi.nlm.nih.gov; released on 03/04/2012) using the BLASTX algorithm with an *E*-value cut-off of 10^{-5}. The functional categories of these unique sequences were further analyzed using the Gene Ontology (GO) database. The unique sequences were categorized based on AGI codes and TAIR GO slim provided by TAIR.

2.5. Pathway assignment with the KEGG database

Pathway assignments were made according to the Kyoto Encyclopedia of Genes and Genome (KEGG) mapping project (http://www.genome.ad.jp/kegg/kegg2.html) (version KEGG 58) (Ogata et al., 1999). Enzyme commission (EC) numbers were assigned to unique sequences that had BLASTX scores with an *E*-value cut-off of 10^{-5} as determined by searching the KEGG database. The unique sequences were allocated to specific biochemical pathways according to the corresponding EC distribution in the KEGG database.

2.6. Real-time PCR analysis

To assay the transcript levels of putative key enzyme genes in the roots, stems, leaves, and flowers of *D. officinale*, quantitative real-time PCR analysis was performed using an Applied Biosystems 7500 Real-Time PCR system with three replicates. Total RNA was treated with RNase-free DNase (TaKaRa). The reverse transcription reaction was performed using the PrimeScript™ 1st Strand cDNA Synthesis Kit (TaKaRa). Each reaction contained 10 μl 2 × SYBR premix Extaq™ (TaKaRa), 10 ng of cDNA and 200 nM gene-specific primers. The total volume was 20 μl. The cycling conditions were as follows: 95 °C for 30 s followed by 40 cycles of 95 °C for 5 s and then 60 °C for 34 s. A melting curve was performed from 60 to 95 °C to check the specificity of the amplified product. The mean value of three replicates was normalized using 18S as the internal control. Primer sequences for the real-time PCR assay are listed in Supplementary Table 1.

2.7. SSR analysis

The detection of SSRs was performed using the Sequence Repeat Identification Tool (http://www.grarnene.org/db/markers/ssrtool).