Plant Gene 12 (2017) 33–42

Contents lists available at ScienceDirect

Plant Gene

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

De novo transcriptome sequencing and analysis of *Euphorbia pekinensis* Rupr. and identification of genes involved in diterpenoid biosynthesis

Xiaoying Cao, Fangzhou Zhang, Bo Yuan, Laisheng Meng, Xuqin Yang, Jihong Jiang*

The Key Laboratory of Biotechnology for Medicinal Plant of Jiangsu Province, Jiangsu Normal University, Xuzhou, China

ARTICLE INFO

Keywords:
E. pekinensis
Expression profiling
Diterpene synthase
Cytochrome P450

ABSTRACT

Euphorbia pekinensis Radixz (EPR) is a commonly used Chinese herbal medicine prepared from the dried roots of *E. pekinensis* Rupr. It has been revealed in chemical analysis that diterpenoids are the primary component of EPR bestowing toxicity and bioactivity; however, little is known about these bioactive and toxic diterpenoid biosynthesis pathways and diterpene diversity in *E. Pekinensis*. In this study, we sequenced root, stem, and leaf transcriptomes of *E. pekinensis*, and conducted *de novo* assembly using 429 million clean reads with Trinity software. Then, 157,491 unigenes were generated whose N50 contained 1826 bp. All of the unigenes were interpreted using Nt, Nr, COGs, SwissProt, GO, and KEGG databases. The expression profiling based on FPKM showed 665 genes up-regulated in roots were enrichment pathways for CYP450s and transcript factors. There are extremely diverse specialised diterpene pathways therein, and most specialised diterpenes are limited to several plant genera, species, or special tissues according to their taxonomy. Using homology, we identified 26 diterpenoid synthase candidates which comprised 25 class I proteins, and only one class II protein, in TPS-a and TPS-c sub-families, among which five class I proteins were highly expressed in root. Additionally, 23 CYP450 enzymes from sub-families of CYP71D and CYP726 were identified, and three were highly expressed in root, for which expression patterns were similar to the aforementioned five class I diTPSs expressed in root and may be involved in macroditerpenoid biosynthesis in EPR. Our results provided a beneficial genomic resource which is able to be used for future functional genomics research on *E. pekinensis*.

1. Introduction

Euphorbiae pekinensis Radix (EPR), which is recorded in the Chinese Pharmacopoeia and prepared from dried roots of *E. pekinensis* Rupr, is a commonly used Chinese herbal medicine in the treatment of gonorrhea, oedema, warts, and migraine. Pharmacological studies show that EPR has bioactivities capable of resisting tumours, viruses, and bacteria, and enhancing immunity as well as certain skin-irritant, tumour promoting, and pro-inflammatory properties (Cao et al., 2010; Xue et al., 2007). Its bioactivity is largely due to the secondary metabolites of chemically diverse diterpenoids. To this day approximately 27 diterpenoids have been isolated from EPR with casbane, cembrane, lathyrane, abietane, ent-atisane, isopimarane, and ent-kaurane skeletons (Tian et al., 2016; Wang et al., 2016; Wang et al., 2014; Tao et al., 2013; Hou et al., 2013; Shao et al., 2011; Liang et al., 2009; Kong et al., 2002). Most of these diterpenoids conferred an anti-tumour effect, in which euphpekinensin, a casbane dierpenoid, has the highest cytotoxic activity (IC_{50} = 0.06 μg/mL) against human KB cells *in vitro* (Kong et al., 2002). Some of these diterpenoids are toxic such as (−)-(1*S*)-15-hydroxy-18-carboxycembrene, a cembrane-type diterpenoid, which is a nephrotoxic substance (Zhang et al., 2016).

Diterpenes are among the most diverse types of compounds and are significant plant metabolites. These compounds are all 20-carbon compounds coming from common precursor geranylgeranyl pyrophosphates. Converting geranylgeranyl diphosphates into casbene is the first critical process for biosynthesising macrocyclic diterpenoids seen

Abbreviations: EPR, *Euphorbia pekinensis* Radixz; GO, gene ontology; COGs, Clusters of Orthologous Groups; KEGG, Kyoto Encyclopedia of Genes and Genomes; NCBI, National Center for Biotechnology Information; qRT-PCR, quantitative real-time reverse transcription PCR; Nt, non-redundant nucleotide database; NR, non-redundant protein database; FPKM, fragments per kb per million fragments; FDR, false discovery rate; DEG, differentially expressed gene; SRA, Sequence Read Archive; RNA-seq, RNA-sequencing; NGS, next generation sequencing; CYP450, cytochrome P450; diTPS, diterpene synthase; UTRs, untranslated regions; MVA, mevalonate pathway; MEP, 2-methyl-D-erythritol-4-phosphate pathway; AACT, acetoacetyl-CoA thiolase; HMGS, HMG-CoA synthase; HMGR, HMG-CoA reductase; MVK, mevalonate kinase; PMK, phosphomevalonate kinase; PMD, diphospho-mevalonate decarboxylase; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; MCT, 2-*C*-methyl-derythritol 4-phosphate cytidylyltransferase; CMK, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol kinase; MDS, 2-*C*-methyl-derythritol 2,4-cyclodiphosphate synthase; HDS, 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase; HDR, 1-hydroxy-2methyl-2-(*E*)-butenyl 4-diphosphate reductase

* Corresponding author.
E-mail address: jhjiang@jsnu.edu.cn (J. Jiang).

http://dx.doi.org/10.1016/j.plgene.2017.07.001
Received 13 March 2017; Received in revised form 30 June 2017; Accepted 5 July 2017
Available online 12 July 2017
2352-4073/ © 2017 Elsevier B.V. All rights reserved.

in secondary metabolism (Zerbe and Bohlmann, 2015); however, from the time when the gene encoding this first critical process was identified, only a biosynthetic step was clarified in the process of transforming casbene to jolkinol C, a lathyrane diterpenoid, and no other steps for biosynthesising diterpenoids derived from casbene have been identified in the pathways (Luo et al., 2016). Synthesising macrocyclic diterpenoids using biological methods includes a variety of ring closures, oxidations, and, for esters, the addition of acyl groups (King et al., 2014). Most oxidation reactions are attributed to cytochrome P450s (Osbourn, 2010; Mizutani, 2012). Cytochrome P450s and alcohol dehydrogenases are probably involved in ring closures and play a catalytic role in the generation of an intramolecular bond: acyltransferases are the strongest candidates involved in diterpenoid ester biosynthesis (Luo et al., 2016; Muchiri and Walker, 2012). Identifying biosynthetic enzymes involved in the formation of diterpenoids and clarifying the way in which these diterpenoids are generated can shed light on how this plant can produce such chemical diversity among its derivative compounds. In addition, they can inspire the development of approaches for increasing the supply of particular diterpenoids used in trials or therapy.

Although diterpenoids have significant chemical diversity, people fail to sufficiently and sustainably access these compounds as a result of their low yields and overharvesting of plant materials in the wild. Some are probably limited to certain tissues, organs, or cell types in space or restricted to certain growth and development stages of plants in time (Zerbe et al., 2013). Recent successes in approaches from metabolic engineering and synthetic biology have attracted broad attention as economically competitive approaches for producing diterpenes of high value (Ajikumar et al., 2010; Zhou et al., 2012); however, these approaches need a knowledge of specific diterpene pathways, enzymes, and genes.

RNAseq, or deep sequencing of mRNA, has become a useful tool for discovering new enzymes and genes and for exploring functional spaces of specific diterpene metabolism, especially in non-model organisms not containing genomic sequences (Boutanaev et al., 2015; Xu et al., 2015; Guzman et al., 2014). To date, people still cannot obtain genomic information about *E. pekinensis*, and there are only 21 nucleotide sequences deposited in the NCBI GenBank database.

In existing studies, an attempt has been made to reveal the characteristics of transcriptome profiles of roots, stems, and leaves of *E. pekinensis* by using Illumina HiSeq platforms and explore genes involved in diterpenoid biosynthesis. We obtained 157,491 unigenes with an N50 of 1826 bp and an average length of 1150 bp. Annotations were made for all of these unigenes with available nucleotide databases and proteins. In addition, we selected candidate genes which involved in the synthesis of diterpene metabolism using biological methods and investigated the differential expression patterns related to these genes. In summary, the annotation and transcriptome assembly of *E. pekinensis* tissues can supply significant genome information which will facilitate future research into the functional genome and metabolism regulation of medicinal compositions of *E. pekinensis*.

2. Materials and methods

2.1. Plant and RNA extraction

Live *E. pekinensis* was obtained from Xuzhou, Jiangsu Province, China during May, 2015. The plants were then grown in the greenhouse at Jiangsu Normal University. The roots, stems, and leaves of *E. pekinensis* were harvested and cleaned using tap water, and then the roots were chopped into small pieces. All the above tissues were harvested with three independent duplicates. All of the samples were frozen immediately in liquid nitrogen and preserved at $-80\ ^{\circ}C$ before further processing.

The RNeasy Plant Mini Kit (Qiagen) was adopted to extract total RNAs from the roots, leaves, and stems following the instructions provided by the manufacturer. Then, the RNase-free DNase I (Takara, Japan) was used to treat the RNAs to eliminate genomic DNA which was contaminated. Moreover, the Agilent Bioanalyzer 2100 system (Agilent Technologies Inc.) and the RNA Nano 6000 Assay Kit was used to examine RNA integrity.

2.2. Preparation and sequencing of cDNA

By using a Paired End Sample Prep Kit (Illumina), cDNA libraries were built based on the instructions supplied by the manufacturer. In brief, the authors purified and extracted mRNA from each sample using Magnetic Oligo (dT) beads. Then, before cDNA synthesis, the RNA subjected to purification was cleaved into those with a mean fragment size of 330 nt. Short fragments then underwent purification and were linked to sequencing adapters. Suitable fragments were then purified by agarose gel electrophoresis to be used as templates in PCR amplification, and the final PCR products were sequenced using the Illumina HiSeq platform.

2.3. De novo transcriptome assembly and annotation

The raw reads obtained *via* the cDNA library sequencing were initially processed to produce clean reads by filtering adapter sequences, empty adapter, low-quality, or unknown readouts. The clean reads were stored in the NCBI Sequence Read Archive (SRA) Sequence Database under accession number SRP097008. The clean reads were assembled using the Trinity software to generate unigenes without redundancy (Grabherr et al., 2011). All of the assembled unigenes to be further annotated showed a length of at least 200 bp. The unigene sequences were then associated with BLASTX to NCBI non-redundant protein (NR), SwissProt database, Kyoto Encyclopedia of Genes and Genomes (KEGG), Clusters of Orthologous Groups (COG), and with BLASTN to nucleotide (Nt) databases resulting in functional annotations (E-values $< 10^{-5}$) of the unigene sequences (Apweiler et al., 2004; Kanehisa et al., 2004; Ashburner et al., 2000; Tatusov et al., 2000). The Blast2GO software was used to find the GO term distribution of each transcript (Conesa et al., 2005).

2.4. Identification of DEGs

The fragments per kilobase per million fragments mapped (FPKM) approach was used for the measurement of gene expression levels in root, stem, and leaf tissues. The expression values were calculated by only retaining both pairs of paired-end reads with unique locations. By using the edgeR program, the DEGs with a log-fold expression change were judged against the criterion: (log FC) > 2 or < -2 based on a high significance value ($P < 0.005$) and the threshold of false discovery rates (FDR < 0.001) (Robinson et al., 2010).

2.5. Identification and analysis of genes related to diterpenoids

For identifying cytochrome P450 (P450) candidates and diterpene synthase (diTPS), custom databases were built according to protein sequences which are available publicly and represent the least populous sequence sets without redundancy. In this way, the functional ranges resembled those used for diTPS and P450 genes. The authors then used tBLASTx with a minimum read length of 150 amino acids and an E-value threshold of 1.0×10^{-50} to perform gene mining of important candidates from produced assemblies. Sequence alignments of proteins were performed by using the CLUSTAL W program. Subsequently, the maximum likelihood method was adopted to build a phylogenetic tree in MEGA 6 using 1000 bootstrap values but otherwise with standard settings.

Download English Version:

https://daneshyari.com/en/article/5590961

Download Persian Version:

https://daneshyari.com/article/5590961

Daneshyari.com