Gene 556 (2015) 153–162

Contents lists available at ScienceDirect

Gene

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

Transcriptome sequencing and comparative analysis reveal long-term flowing mechanisms in *Hevea brasiliensis* latex

Fang Wei [1], Shiqiao Luo [1], Qiankun Zheng [1], Jian Qiu [1], Wenfeng Yang [1], Ming Wu [1], Xianzhou Xiao *[,1]

Key Laboratory of Biology and Genetic Resources of Rubber Tree, Ministry of Agriculture, Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan 571737, China

ARTICLE INFO

Article history:
Received 12 September 2014
Received in revised form 18 November 2014
Accepted 21 November 2014
Available online 26 November 2014

Keywords:
Long-term flowing latex
Rubber tree
Gene expression
Transcriptome
Hevea brasiliensis

ABSTRACT

Background: The rubber tree, *Hevea brasiliensis*, is a major commercial source of natural rubber. Increasing the rubber yield of rubber trees is a very serious problem since the demands for high quality rubber materials are great. Establishment of a tapping system is based on an estimate of tapping intensity from the rubber tree. Latex flowing time is one of the most critical factors that determine the rubber yield. Long-term flow is a type of phenomenon of the rubber tree latex with longer flowing time than normal latex flow, and is always caused by intensive tapping. Thus, transcriptome and expression profiling data for long-term flowing latex (LFL) are needed as an important resource to identify genes and to better understand the biological mechanisms of latex flow in rubber trees.

Results: The transcripts were sequenced using the Illumina sequencing platform. After cleaning, quality checks and sequencing, 98,697 transcripts and 38,584 unigenes were assembled with the mean size of 1437.31 bp and 923.86 bp, respectively. In BLAST searches of our database against public databases, 65.17% (25,147) of the unigenes were annotated with gene descriptions, conserved protein domains, or gene ontology terms. Functional categorization further revealed 853 individual unigenes related to long-term flow. According to KEGG classification, the clusters for "cysteine and methionine metabolism", "energy", "oxidative phosphorylation", "terpenoid backbone biosynthesis", "plant hormone signal transduction" and "copper, potassium transporter" were significantly enriched metabolic pathways.

Conclusions: We conducted high-resolution transcriptome profiling related to LFL in *H. brasiliensis*. The research facilitates further studies on gene discovery and on the molecular mechanisms related to the estimation of tapping intensity and prolonging latex flowing time. We concluded that it was necessary to improve energy supplies for intensive tapping and the copper ion content of rubber tree latex could be considered as a standard to estimate tapping intensity.

1. Introduction

Natural rubber (cis-1,4-polyisoprene) is a type of high molecular weight polymeric substance. Currently, >90% of all natural rubber comes from one single tropical tree species — *Hevea brasiliensis* (para rubber tree) (Tang et al., 2010). Upon bark tapping, the laticifer vessels are opened and their latex is expelled due to the turgor pressure of the laticifer vessels (d'Auzac et al., 1989). The latex flow rate and duration of flow are the first intrinsic factors known to limit rubber yield: the faster and longer the latex flow, the higher the yield (d'Auzac et al., 1989). Increasing the rubber yield of rubber trees is very important to meet the demands for high quality rubber materials.

Long-term flow is a type of phenomenon typical of rubber tree latex with longer flowing time than normal flowing latex (NFL) whose flowing time is less than 6 h. The time of long-term flowing latex (LFL) usually ranges from 12–24 h but can be longer. On one hand, duration of flowing time increases latex yield per tapping; on the other hand, rubber trees expend metabolic energy when there is a too long flowing time and creates an added burden to regenerate latex flow. Generally speaking, intensive ethylene stimulation, low temperature and other factors can delay latex flow. Over the past few decades, studies of the mechanisms into what determines productivity have been

Abbreviations: LFL, Long-term flowing latex; BLAST, Basic local alignment search tool; COG, Clusters of orthologous groups; FDR, False discovery rate; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; Nr, Non-redundant; NCBI, National Center for Biotechnology Information; Q20 percentage, Percentage of bases whose quality was larger than 20 in clean reads; EST, Expressed Sequence Tags; qRT-PCR, Quantitative real-time reverse transcription polymerase chain reaction; RNA-Seq, RNA sequencing; TSC, Total solid content; Gb, Gigabases; ORF, Open Reading Frame; DEGs, Differentially expressed genes; IPP, Isopentenyl diphosphyate.

* Corresponding author.
E-mail addresses: wfang545@163.com (F. Wei), abc6712@163.com (S. Luo), universe101325@126.com (Q. Zheng), qiujian_@sina.com (J. Qiu), yangwenfeng117@163.com (W. Yang), dz.wm@163.com (M. Wu), xiaoxianzhou80@163.com (X. Xiao).
[1] Postal address: Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan 571737, China.

http://dx.doi.org/10.1016/j.gene.2014.11.048
0378-1119/

carried out using various methods and materials (Ko et al., 2003; She et al., 2013; Tang et al., 2010, 2013; Tian et al., 2013; Yeang, 2007). To date, chemical treatment with ethrel (2-chloroethylphosphonic acid), an ethylene releaser which acts on plant metabolism, is widely used to increase latex yields by up to 1.5–2.0 fold, and is now used as a routine stimulant in rubber tree plantations worldwide (Zhu and Zhang, 2009). Ethrel can improve latex yield mainly through prolonging latex flow (Tungngoen et al., 2009b). However, the negative aspect of ethrel stimulation cannot be neglected and may result in increased tapping panel dryness (TPD) with intensive tapping. LFL is therefore a good material to study how to delay latex flowing time at the molecular level with low concentration ethrel and low frequency tapping, or even without ethrel application. It is worth mentioning that this research will help understand the mechanisms of low-frequency and micro-cut tapping.

Sequencing technology has developed rapidly in recent years and transcriptome analysis has become increasingly powerful. RNA sequencing (RNA-seq) provides rapid and comprehensive analyses for plants and other organisms (Ozsolak and Milos, 2011; Van Verk et al., 2013; Wang et al., 2009). The depth of transcriptome sequencing provides a cost-effective means of qualitative and quantitative analyses of gene transcripts in many non-model plant species including the rubber tree. Compared to model plants, transcript resources in this economically important species have only been initiated relatively recently, with the analysis of Expressed Sequence Tags (EST) from latex (Chow et al., 2012, 2014; Duan et al., 2013; Li et al., 2012; Pootakham et al., 2011; Rahman et al., 2013; Triwitayakorn et al., 2011; Xia et al., 2011). In the present study, we generated a transcriptome dataset to provide genetic information to explore long-term flowing mechanisms of rubber tree latex using the Illumina HiSeq™ 2500 platform. In this research, a transcriptome study on the samples of NFL and LFL was performed to generate comprehensive gene expression analyses. We present a bioinformatic exploration, functional annotation, comparative analysis and the KEGG pathways of subset transcripts identified from significantly different expressions in the latex. We speculate that the assembled, annotated transcriptome sequences and transcript abundance patterns will provide a valuable genetic resource for further investigations of the molecular mechanisms of latex flowing and provide clues on how to ameliorate the damage caused by intensive tapping.

2. Materials and methods

2.1. Ethics statement

The study was approved by the Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan, China.

2.2. Plant material and determination of physiological parameters

The RRIM600 strain was planted at the experimental farm at the Chinese Academy of Tropical Agricultural Sciences in 1982. Two samples (each sample consisted of three biological replicates, each replicate consisted of three trees) comprising LFL and NFL were collected. An aliquot of 10 ml of latex per tree was collected 2 h after tapping in a 20-ml centrifuge tube placed in ice, and transported to the laboratory for immediate analysis. The parameters of total solid content (TSC), pH, thiols, inorganic phosphates, and sucrose contents were determined according to Eschbach et al. (1984).

Table 1
Physiological parameters for LFL and the control.

Physiological parameters	LFL	Control
Flowing time	25 h	4 h
Total solid content (%)	18.5 ± 1.3a	27.0 ± 0.5b
Sucrose content (mM)	7.38 ± 0.20a	14.33 ± 0.23b
Inorganic phosphorus (mM)	13.52 ± 0.58a	12.35 ± 0.40a
Thiols (mM)	1.504 ± 0.10a	0.823 ± 0.034b

Different letters in the same row indicate statistical significance at $p < 0.05$ while the same letter indicates no statistical significance.

2.3. RNA extraction and library construction for transcriptome analysis

Total RRIM600 RNA was extracted, including LFL (2 h after tapping) and NFL, two duplications every sample (each duplication consisted of three trees), according to the method of Tang et al. (2007). The quality and concentration of the extracted RNA were checked by agarose gel electrophoresis and measured by a spectrophotometer (Alphalmager 2200, USA). RNA was subjected to DNase I (Takara, China) digestion to remove genomic DNA contamination.

Both the quantity and quality of the RNA were verified using an Agilent 2100 Bioanalyzer (Agilent). A total of 20 μg of RNA was enriched equally from the samples for cDNA library preparation.

2.4. cDNA library construction and sequencing

cDNA library preparation and sequencing reactions were conducted using the Biomarker Technology Company, Beijing, China. Next, a Poly(A) mRNA Magnetic Isolation Module(NEB, E7490)was used to pool mRNA from RNA samples. The cDNA library was constructed following the manufacturer's instructions using a NEBNext mRNA Library Prep Master Mix Set for Illumina(NEB, E6110)and NEBNext Multiplex Oligos for Illumina (NEB, E7500). Short fragments were detected using 1.8% agarose gel electrophoresis and then were quantified by qPCR through a Library Quantification Kit-Illumina GA Universal system (Kapa, KK4824). Finally, the sequencing library was constructed by PCR amplification (15 cycles) and sequenced using the Illumina Genome Analyzer IIx sequencing platform. The Illumina cBOT (Illumina Inc., San Diego, CA) was then used for cluster generation which was sequenced on the Illumina HiSeqTM 2500 platform. Data analyses and base calling were performed using the Illumina instrument software.

2.5. Sequence analyses, assembly and annotation

All raw reads were cleaned by removing adapter sequences and low-quality sequences with ambiguous bases "N". Reads with more than 20% $Q < 30$ bases were also cleaned. De novo transcriptome assembly of these quality reads was performed using the Trinity (Grabherr et al., 2011; Rahman et al., 2013) software. The reads were mapped back to contigs with paired-end reads. This approach detected contigs from the same transcript as well as measuring the distances between these contigs. Next, Trinity connected the contigs using N to represent unknown sequences between each contig pair to form scaffolds. Finally, unigenes were generated with zero N values in the sequence that could not be extended on either end. Getorf software was used to predict ORF (http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html). In the final step, BLASTX alignments (e-value, 0.00001) were made between the recovered unigenes and protein databases such as the National Center for Biotechnology Information (NCBI) non-redundant protein (Nr), Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) and Cluster of Orthologous Groups of proteins (COG) and the best aligning results were used to determine the sequence direction of the unigenes.

2.6. Identification of differentially expressed unigenes

Unigene expression was calculated using the RPKM method (Trapnell et al., 2010). The RPKM method is able to eliminate the influence of

Table 2
Summary of Illumina transcriptome sequencing for rubber tree latex.

No. of reads	Data	GC(%)	Q20(%)	Q30(%)
80,992,220	20.37179	177.14	98.675	93.8475