Gene 660 (2018) 120–127

Contents lists available at ScienceDirect

Gene

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

Research paper

Comprehensive analysis of differentially expressed genes reveals the molecular response to elevated CO_2 levels in two sea buckthorn cultivars

Guoyun Zhang[a], Tong Zhang[a], Juanjuan Liu[a], Jianguo Zhang[a,b,*], Caiyun He[a,*]

[a] State Key Laboratory of Tree Genetics and Breeding & Key Laboratory of Tree Breeding and Cultivation, State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, China

[b] Collaborative Innovation Center of Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, China

ARTICLE INFO

Keywords:
Sea buckthorn
CO_2 concentration
Transcriptome
Differentially expressed genes
Molecular mechanism

ABSTRACT

Atmospheric carbon dioxide (CO_2) concentration increases every year. It is critical to understand the elevated CO_2 response molecular mechanisms of plants using genomic techniques. *Hippophae rhamnoides* L. is a high stress resistance plant species widely distributed in Europe and Asia. However, the molecular mechanism of elevated CO_2 response in *H. rhamnoides* has been limited. In this study, transcriptomic analysis of two sea buckthorn cultivars under different CO_2 concentrations was performed, based on the next-generation illumina sequencing platform and de novo assembly. We identified 4740 differentially expressed genes in sea buckthorn response to elevated CO_2 concentrations. According to the gene ontology (GO) results, photosystem I, photosynthesis and chloroplast thylakoid membrane were the main enriched terms in 'xiangyang' sea buckthorn. In 'zhongguo' sea buckthorn, photosynthesis was also the main significantly enriched term. However, the number of photosynthesis related differentially expressed genes were different between two sea buckthorn cultivars. Our GO and pathway analyses indicated that the expression levels of the transcription factors WRKY, MYB and NAC were significantly different between the two sea buckthorn cultivars. This study provides a reliable transcriptome sequence resource and is a valuable resource for genetic and genomic researches for plants under high CO_2 concentration in the future.

1. Introduction

Sea buckthorn (*Hippophae rhamnoides* L.), a thorny, nitrogen-fixing, deciduous shrub, has been widely cultivated for its nutritional and medicinal properties (Suryakumar and Gupta, 2011; Teleszko et al., 2015; Li et al., 2017). Sea buckthorn is environmentally important in coastal areas, as it fixes sand in place, preventing wind erosion (Ledwood and Shimwell, 1971; Singh, 2005). Transcriptomic studies have shown that the sea buckthorn tolerates cold and freeze stress well (Chaudhary and Sharma, 2015). Owing to its tolerance of a wide range of temperatures, the native habitat of sea buckthorn extends across Europe and Asia, including China, India, Mongolia, Russia, Sweden, Finland, and Norway (Lian and Chen, 1999). Recent studies of this plant have explored such diverse topics as the bioactive compounds it produces, and the mechanisms causing fruit ripening (Ghangal et al., 2012; Teleszko et al., 2015; Li et al., 2017). However, the molecular response mechanisms of sea buckthorn to elevated atmospheric carbon dioxide remain unknown.

Atmospheric CO_2 considered as the primary carbon source for plant photosynthesis and a major driver of climate change. Recently, studies of CO_2 become an important component of plant researches. Compared with the lowest concentrations, the average global atmospheric concentration of CO_2 was 394 $\mu mol \cdot mol^{-1}$ in 2012 (Meinshausen et al., 2011). This concentration are predicted to surpass 550 $\mu mol \cdot mol^{-1}$ by 2050 and reach 700 $\mu mol \cdot mol^{-1}$ by the end of 2100 (Leakey et al., 2009b). A large number of studies reveals that elevated CO_2 significantly affect the productivity and fitness of plant (Idso and Kimball, 1997; Ainsworth and Long, 2005; Springer and Ward, 2007; Becklin et al., 2017). For different plant species, the capacity to elevated atmospheric CO_2 relies on a series of physiological and biochemical processes. Hence, understanding the molecular mechanisms of tree growth can promote us to find certain key candidate genes, transcriptional factor and metabolism pathway. In past few decades, CO_2 studies found that several physiological characters related to plant growth were affected by elevated atmospheric CO_2 (Idso and Kimball, 1997; Ehlers et al., 2015; Becklin et al., 2017). Furthermore, a small number of

Abbreviations: DEGs, differentially expressed genes; RPKM, reads per kilobase per million reads; GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; ZG, zhongguo; XY, xiangyang; PPI, protein–protein interaction; RSEM, RNA-seq by Expectation Maximization; STRING, Search Tool for the Retrieval of Interacting Genes

* Corresponding authors.
E-mail addresses: zhangguoyun2014@gmail.com (G. Zhang), zhangjg@caf.ac.cn (J. Zhang), hecy@caf.ac.cn (C. He).

https://doi.org/10.1016/j.gene.2018.03.057
Received 24 December 2017; Received in revised form 5 March 2018; Accepted 16 March 2018
Available online 22 March 2018
0378-1119/

studies have provided insight into plant community dynamics (Ainsworth and Long, 2005; Leakey et al., 2009b; Norby et al., 2016) and large-scale transcriptomic studies have indicated that the expression levels of plant genes related to both photosynthesis (Kumar et al., 2017) and leaf development (Leakey et al., 2009a) are sensitive to elevated CO_2 levels. For example, transcriptomic analyses found that a series of photosynthesis regulatory genes were apparently affected by elevated CO_2 in *Jatropha curcas* L. (Kumar et al., 2017).

Previous studies on elevated CO_2 related molecular responses in plants concentrated on model species and few tree species and these studies mainly reveal secondary metabolism related genes during the plants development (Idso and Kimball, 1997; Tallis et al., 2010; Niu et al., 2016). Nowadays, with the high-speed development and widespread of next generation sequencing technologies, de novo RNA-sequencing and high throughput sequencing have promoted discovery and analysis of new genes in tree species. These approaches also provided insight for better understanding of transcriptional patterns and gene expression in plant development in response to diverse environments. Due to lack of genomic resources, few transcriptome studies for different tissues or development stages were performed for sea buckthorn.

However, there has been little research on gene differential expression for sea buckthorn under elevated CO2 concentration. In this study, we exposed two sea buckthorn cultivars to three different levels of atmospheric CO_2, including the current ambient concentration and two predicted future atmospheric CO_2 concentrations (550 and 700 $\mu mol \cdot mol^{-1}$). Then, we used RNA-seq techniques to identify the DEGs among two cultivars at the three levels of CO_2 exposure. A series of bioinformatics methods were used to determine the crucial differentially expressed genes affected by elevated CO_2 concentration and related metabolic pathways in sea buckthorn. This study first provided temporal expression pattern of elevated CO_2 response genes and comprehensive transcriptome sets for sea buckthorn.

2. Material and methods

2.1. Plant material and experimental treatments

On 11 March 2010, triennial “zhongguo” (ZG) and “xiangyang” (XY) sea buckthorn seedlings were planted in individual plastic pots (20 cm × 26 cm × 34 cm) containing a 5:3:2 mixture of clay soil, sand and peat. Eighteen randomly selected pots were moved into three controlled-environment chambers (AGC-2; Zhejiang University Electrical Equipment Factory, Hangzhou, China). Each chamber measured 3.5 m × 2.2 m × 3.2 m (length × width × height), with an average daytime active photosynthetic radiation of 800 ± 20 $\mu mol \cdot m^{-2} \cdot s^{-1}$, and a relative humidity of 65 ± 5%. Then, seedlings were exposed to three different CO_2 concentrations for three months (Table 1), starting 25 June 2010. The CO_2 concentrations in the chambers were continuously monitored by automatic CO_2 detection systems. The pots were rotated inside the chambers weekly to minimize the effects of microclimatic variation within the chambers.

2.2. RNA isolation and sequencing

The leaves, stems, and roots of all plants collected for biomass determination and RNA extraction. Total RNA from each sample was isolated using the RNeasy Plus Universal Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. The quality and quantity of total RNA was checked with an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Then, equal amounts of purification RNA from each separate tissue (leaf, stem, and root) of the plants under T0 treatment, and leaves of plants exposed to all three treatments (T0, T1, T2) were pooled to construct the cDNA libraries using the mRNA-seq sample preparation kit (Illumina, CA). Then, following the manufacture's recommendations, first-strand cDNA was synthesized from the generated short fragments using random hexamer primers. Then, second-strand cDNA synthesis was performed in a reaction mix. These cDNA fragments were purified using a QiaQuick PCR Purification Kit (Qiagen, Germany) and ligated to sequencing adapters. Suitable fragments were selected to create the final cDNA library by PCR amplification. The resulting 12 cDNA libraries were individually loaded into an illumina flow cell and sequenced on Illumina Hiseq2000 sequencing platform.

Table 1
The treatment of *Hippophae rhamnoides* sample.

	CO_2 concentration(μmol/mol)	Day temperature(°C)	Night temperature(°C)
T0	385 ± 20	25 ± 1	20 ± 1
T1	550 ± 20	28 ± 1	23 ± 1
T2	720 ± 20	31 ± 1	26 ± 1

2.3. De novo assembly and functional annotation

We filtered the raw data to generate clean reads by removing adaptor, unknown nucleotides and low-quality reads. All subsequent analyses were performed on these high-quality clean reads. To obtain more comprehensive transcripts, we mixed the clean reads from different tissues. The mixed clean short reads were assembled using Trinity (Grabherr et al., 2011).

Then, all assembled transcriptome sequences were compared with various sequence databases, including the NCBI non-redundant database (nr), NT, Swiss-Prot, and KOG. Based on these annotation, we annotated the assembled unigenes with the GO database using BLAST2GO with an E-value ≤ $1.0E^{-5}$ (Conesa et al., 2005). The KEGG pathway annotations were performed by KOBAS to the KEGG database.

2.4. Identification of DEGs

We generated six mRNA libraries representing the two sea buckthorn cultivars exposed to the three different CO_2 concentrations. The raw sequencing data were produced using Illumina sequencing. After the raw data were generated and the data processing was completed, the high-quality reads were mapped to assembled transcriptome reference sequences using Bowtie (Wang et al., 2011). We used RNA-seq by Expectation Maximization (RSEM) to count the number of mapped reads for each gene (Li and Dewey, 2011). The expression level of each unigene was calculated as reads per kilobase per million mapped reads (RPKM) (Ward et al., 2012). After calculating gene expression, we standardized the RPKM data using the trimmed mean of m-values (TMM) approach. We used DEGseq package to identify DEGs (Wang et al., 2010). The resulting *p*-values were adjusted using the Benjamini-Hochberg method of controlling the false discovery rate (FDR) (Shen et al., 2013). We defined DEGs in the same plant exposed to different levels of CO_2 as significantly affected if the adjusted *p*-value was < 0.001, the FDR was < 0.01, and the $|\log_2 ratio|$ was > 1.

Then, we performed a saturation analysis to test whether the number of detected genes increased proportional to the number of sequences. We used Pearson's correlation analysis to test for correlations between the parallel experiments (i.e. T0 vs T1, T0 vs T2, and T1 vs T2). A high correlation between parallel experiments would indicate the reproducibility of our results and show that our methods were stable. We further anatomized selected DEGs using integrated bioinformatics methods.

2.5. Differential expression analysis

Based on the annotations of the reference transcriptome sequences, we got annotation of selected DEGs. We used the cluster algorithm of