Gene xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Gene

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

Research paper

Genome-wide analysis of long non-coding RNAs at the mature stage of sea buckthorn (*Hippophae rhamnoides* Linn) fruit

Guoyun Zhang [a], Aiguo Duan [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

Article history:
Received 18 August 2016
Received in revised form 3 October 2016
Accepted 13 October 2016
Available online xxxx

Keywords:
Sea buckthorn
Fruit
lncRNAs
Transcriptome

ABSTRACT

Long non-coding RNAs (lncRNAs), which are >200 nt longer transcripts, potentially play important roles in almost all biological processes in plants and mammals. However, the functions and profiles of lncRNAs in fruit is less understood. Therefore, it is urgent and necessary to identify and analyze the functions of lncRNAs in sea buckthorns. Using RNA-sequencing, we synthetically identified lncRNAs in mature fruit from the red and yellow sea buckthorn. We obtained 567,778,938 clean reads from six samples and identified 3428 lncRNAs in mature fruit, including 2498 intergenic lncRNAs, 593 anti-sense lncRNAs, and 337 intronic lncRNAs. We also identified 3819 and 2295 circular RNAs in red and yellow sea buckthorn Fruit. In the aspects of gene architecture and expression, our results showed significant differences among the three lncRNA subtypes. We also investigated the effect of lncRNAs on its cis and trans target genes. Based on target genes analysis, we obtained 61 different expression lncRNAs (DE-lncRNAs) between these two sea buckthorns, including 23 special expression lncRNAs in red fruit and 22 special expression lncRNAs in yellow fruit. Importantly, we found a few DE-lncRNAs play cis and trans roles for genes in the Carotenoid biosynthesis, ascorbate and aldarate metabolism and fatty acid metabolism pathways. Our study provides a resource for lncRNA studies in mature fruit. It probably encourages researchers to deeply study fruit-coloring. It expands our knowledge about lncRNA biology and the annotation of the sea buckthorn genome.

1. Introduction

Sea buckthorn (*Hippophae rhamnoides* Linn) belongs to the genus *Hippophae* Elaeagnaceae family, which is distributed throughout Asia, Europe, and several other parts of the world (Li and Schroeder, 1996). In addition, it is well known for its important economic and medicinal values (Rousi, 1971; Rongsen, 1992; Li and Small, 2003). In these parts of world, sea buckthorn fruit is widely used in the production of juice, wine and jam (Bal et al., 2011). In addition, because of the high levels of isoprenylated flavonoids, the sea buckthorn fruit has been used for preventing liver fibrosis, improving eyesight, treating fever, lowering cholesterol, facilitating urination, and lowering blood pressure (Gao et al., 2003; Suomela et al., 2006; Suryakumar and Gupta, 2011). Owing to the many different colors of fruits, sea buckthorn is an ideal research model organism used to study fruit pigments. Recently, the genome sequencing of sea buckthorn has been completed. However, a lot of important information has not been exploited completely, especially in terms of lncRNAs. Therefore, it is urgent and necessary to identify lncRNAs and understand the functions of lncRNAs in sea buckthorn.

Following the advance of DNA sequencing technology, recent research has challenged the traditional central dogma of biology. According to past studies, a set of increasing evidence shows that only 1–2% of eukaryotic genomes have protein-coding capacity, but >90% are transcribed, which are mainly transcribed as noncoding RNAs (ncRNAs) (Carninci et al., 2005; Cheng et al., 2005; Qi et al., 2013; Song et al., 2016). A lot of studies have indicated that ncRNAs play important roles in many respects, including regulation of transcriptional, post-transcriptional, and epigenetic levels in eukaryotes (Wapinski and Chang, 2011; Kim and Sung, 2012; Hangauer et al., 2013). Since the first report of lncRNAs in humans (Lukiw et al., 1992), many lncRNAs have been identified in a few plant species, including *Arabidopsis*, *Zea mays*, rice, cotton and kiwifruit (Boerner and McGinnis, 2012; Liu et al., 2012; Y.-C Zhang et al., 2014; Hao et al., 2015; Tang et al., 2016; Zou et al., 2016). For example, sexual reproduction in rice is influenced by the lncRNAs XLOC_057324, which acts as a miR164 target mimic and plays a role in panicle development and fertility in rice and tomatoes

Abbreviations: lncRNAs, long non-coding RNAs; DE-lncRNAs, different expression lncRNAs; ncRNAs, noncoding RNAs; CPC, Coding Potential Calculator; FPKM, fragment per kilobase of exon model per Million mapped reads; CIRI, CircRNA identifier.

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

http://dx.doi.org/10.1016/j.gene.2016.10.017
0378-1119/

(Rymarquis et al., 2008; Y.-C Zhang et al., 2014). Previous researchers also identified 664 lncRNAs responsive to drought stress (W. Zhang et al., 2014a).

In our study, we firstly identified 3428 genome-wide lncRNAs, using RNA-seq data from sea buckthorn mature fruits. In addition, the structural characteristics of these lncRNAs and mRNAs were compared and analyzed. Furthermore, based on genomic information, we predicted the functions of the lncRNAs, which contribute to further illuminating the roles of the lncRNAs in the mature fruits.

2. Material and methods

2.1. Plants

In this study, two sea buckthorn groups with diverse phenotypes of fruit color were investigated. The red fruit and yellow fruit sea buckthorn, which were the first filial generation of *H. rhamnoides* L. subsp. *mongolica* Rousi × *H. rhamnoides* L. subsp. chinensis Rousi, is characterized by fruit color. The plants were cultivated under natural conditions in the wild. Healthy and fresh fruits were excised in the mature stage. Three biological replicates were collected for each samples. The materials were reserved in liquid nitrogen for RNA sequencing.

2.2. RNA isolation, library preparation, and sequencing

According to the manufacturer's instructions, total RNA was isolated using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). RNA concentration and integrity was respectively measured using Qubit RNA Assay Kit and the RNA Nano 6000 Assay Kit (Thermo Scientific, Waltham, USA). In addition, 3 μg RNA for each sample was prepared to construct the library. First, ribosomal RNA was removed by using an EpicentreRibo-zero rRNA Removal Kit (Thermo Scientific, Waltham, USA). Subsequently, the high strand-specificity of the libraries was generated following the manufacturer's recommendations. Finally, the libraries were sequenced on an Illumina Hiseq 2500 platform and 125-bp paired-end reads were generated (SRP076880).

2.3. Transcriptome assembly and lncRNA identification

Firstly, clean data were obtained by removing reads containing adapters, reads containing over 10% of ploy-N, and low-quality reads (>50% of bases whose Phred scores were <5%) in the raw data. We applied the clean data to all subsequent analyses. Then, reference genome and annotation files of sea buckthorn were obtained from our research directly. Using Bowtie2 (version 2.2.9) (Langmead and Salzberg, 2012; Ye et al., 2015) and TopHat2 (version 2.0.9) (Trapnell et al., 2012; Kim et al., 2013), we built an index of the reference genome and compared the clean data to the reference genome. The mapped reads of each sample were assembled using both Scripture (Guttman et al., 2010) and Cufflinks (version 2.2.1) (Trapnell et al., 2010). Then, all transcriptomes from six samples were merged to reconstruct a comprehensive transcriptome using Cuffmerge (version 2.2.1). After the final transcriptome was generated, Cuffdiff (version 2.2.1) was used to estimate the expression levels of all transcripts. We used a pipeline to select the lncRNAs as follows: (1) we reserved transcripts assembled by both two tools or appearing in two samples;(2) we discarded transcripts that overlapped with known mRNAs, transcripts with FPKM scores <0.5 in all samples and transcripts shorter than 200 bp; 3) we used the CPC (Kong et al., 2007) to predict transcripts with coding potential. All transcripts with CPC scores >0 were removed.

Finally, to eliminate transcripts that contained known protein domains involved in the Pfam database (Bateman et al., 2002), we subjected these remaining transcripts to HMMER (Eddy, 2009) analysis. The remaining transcripts were considered reliably expressed lncRNAs. According to lncRNAs' location, lncRNAs were classified to long intergenic noncoding RNAs (lincRNAs), intronic lncRNAs and antisense lncRNAs (lncNATs) categories by cufflinks (Liu et al., 2012).

2.4. Identification of circular RNAs and AS events

Circular RNAs are a class of long noncoding RNA that were recently rediscovered as diverse, highly abundant, conserved and naturally occurring RNAs in eukaryotes. In this study, we identified circular RNAs using CIRI tool with default parameters (Gao et al., 2015). Because of the prevalence of alternative splicing (AS) events within circular RNAs, we also identified the AS events of circular RNAs using CIRI-AStool (Gao et al., 2016).

2.5. Conservation analysis of lncRNAs

The six representative plant genomes, including rice, maize, *Arabidopsis*, populus, tomato and watermelon, were downloaded from online databases. Using the BLASTN tool (version 2.2.24), the lncRNA sequences of sea buckthorn were aligned against these six plant genomes with a cutoff threshold E-value <1e−5 and coverage of >20% of matched regions.

2.6. Target gene prediction and functional analysis of lncRNAs

To explore the function of lncRNAs, we first predicted the cis- and trans- target genes of lncRNAs. According to past research (Y.-C Zhang et al., 2014b), lncRNAs may play a cis role acting on neighboring target genes. In this study, we searched coding genes 10/100 k upstream and downstream of lncRNA using perl script. In this study, the trans target genes were predicted based on the expression level of lncRNA and target genes (Li et al., 2014a). So, we calculated the Pearson's correlation coefficients (R) between lncRNAs and mRNAs using R tool. LncRNAs trans target genes were identified with $|R| \geq 0.95$. Then, we showed functional analysis of the target genes for lncRNAs by using the BLAST2GO software (Conesa et al., 2005; Du et al., 2010). Significance was expressed as a p-value < 0.05.

2.7. Different expression analysis of lncRNAs

According to present study, Cufflinks (version 2.2.1) was used to perform expression levels for lncRNAs and coding genes in each sample by calculating FPKM (Trapnell et al., 2010). We described the FPKM of HR divided FPKM of HY as fold change (FC) lncRNA expression level. The differentially expressed lncRNAs were selected with a $\log_2$FC ≥ 1 or ≤−1 and with statistical significance (False Discovery Rate q-value < 0.05). Then, based on the functional enrichment analysis of the target genes, we performed analysis for the target genes of DE-lncRNAs.

3. Results

3.1. Identification of lncRNAs in sea buckthorn fruit

We obtained raw data generated from the Illumina HiSeq 2500 platform. After discarding adaptor and low-quality sequences, we obtained >42,632,472 clean reads. In addition, the percentage of clean reads ranged from 97 to 99% (Table 1). Then, we mapped the clean reads to the sea buckthorn reference genome. Using both Scripture and Cufflinks (version 2.2.1) software, we reconstructed the comprehensive transcripts. After basic filtering, we obtained 5148 tentative lncRNAs. Then, we performed protein coding potential analysis to further confirm these lncRNAs using the CPC and Pfamscan. After these rigorous selections, 2498 lincRNAs, 337 intronic lncRNAs, and 593 lncNATs from the mature fruit were identified and used for further analysis.

Please cite this article as: Zhang, G., et al., Genome-wide analysis of long non-coding RNAs at the mature stage of sea buckthorn (*Hippophae rhamnoides* Linn) fruit, Gene (2016), http://dx.doi.org/10.1016/j.gene.2016.10.017