South African Journal of Botany 116 (2018) 34–41

Contents lists available at ScienceDirect

South African Journal of Botany

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

Transcriptome analysis of *Abeliophyllum distichum* NAKAI reveals potential molecular markers and candidate genes involved in anthocyanin biosynthesis pathway

J.H. Choi [a], H. Kim [b], T.K. Hyun [a,*]

[a] Department of Industrial Plant Science and Technology, College of Agricultural, Life and Environmental Sciences, Chungbuk National University, Cheongju 28644, Republic of Korea
[b] Gangwon Forest Science Institute, Chuncheon 24207, Republic of Korea

ARTICLE INFO

Article history:
Received 4 November 2017
Received in revised form 9 February 2018
Accepted 23 February 2018
Available online xxxx

Editor: Kalina Ananieva

Keywords:
Abeliophyllum distichum
RNA-seq
De novo assembly
Anthocyanin
Antioxidant

ABSTRACT

De novo transcriptome assembly from RNA-sequencing (RNA-seq) data has been successfully used in biological studies worldwide for gene characterization, functional genomic studies, understanding biological processes and developing molecular markers in non-model plants. This study employed *de novo* assembly to characterize the transcriptome of *Abeliophyllum distichum*, which is a monotypic genus and plant endemic to Korea. After *de novo* assembly and aligning the sequence to the public databases, 28,842 (59.32%) of the 48,623 unigenes were annotated. A total of 1724 potential simple sequence repeats (SSRs) were identified in 1469 unigenes. SSRs with a tri-nucleotide repeat motif were the most abundant in the *A. distichum* transcriptome. In addition, the flower extract of *A. distichum* possessed the highest level of anthocyanins and displayed the highest antioxidant activity, although the leaf extract contained a higher total flavonoid content. Furthermore, we identified 55 unigenes from the *A. distichum* transcriptome that encode putative enzymes involved in anthocyanin biosynthesis. We analyzed the organ-specific expression pattern of anthocyanin biosynthesis-related genes using real-time PCR analysis and determined that the organ-specific transcription of functional genes affects organ-specific anthocyanin accumulation. Taken together, our transcriptome analysis is the first step toward functional genomics and will assist future studies of the molecular genetics associated with molecular markers in *A. distichum*, as well as other members of the Oleaceae family.

1. Introduction

As a result of the advent and rapid development of next-generation sequencing, the field of biological research has rapidly evolved to high-throughput, information-intensive investigations that produce massive amounts of data, such as genomics, transcriptomics, metagenomics, proteogenomics, non-coding RNA discovery and single nucleotide polymorphism (SNP) detection. In addition, RNA-sequencing (RNA-seq) for transcriptome analysis has allowed the identification of the metabolic pathways of bioactive compounds and the analysis of the related gene expression in various samples and tissues. For instance, this approach has been applied to investigate saponin biosynthesis in Siberian ginseng (Hwang et al., 2015), American ginseng (Sun et al., 2010) and Korean ginseng (Jayakodi et al., 2014) and *Oplopanax elatus* (Eom et al., 2017); flavonoid and anthocyanin biosynthesis in Korean black raspberry (Hyun et al., 2014), tea plant (Guo et al., 2017), *Epimedium pseudowushanense* (Pan et al., 2017) and litchi (Zhang et al., 2016); taxane biosynthesis in *Taxus mairei* (Hao et al., 2011); carotenoid biosynthesis in ori'ental melon (Shin et al., 2017), watermelon (Zhu et al., 2017), as well as gac (Hyun et al., 2012), and fatty acid biosynthesis in *Perilla frutescens* (Kim et al., 2016) and oil palm (Jin et al., 2017). These indicate that *de novo* transcriptome assembly from RNA-seq data is a valuable approach in the study of non-model plant species lacking sequenced genomes.

Abeliophyllum distichum Nakai, commonly called white forsythia, belongs to the family Oleaceae and is a monotypic genus endemic to Korea (Kim et al., 2016). Although *A. distichum* is used as an ornamental plant in some countries, due to its horticultural value (Kim et al., 2016), a recent report showed that *A. distichum* extract attenuates inflammatory responses via reactive oxygen species (ROS) and MEK/ERK signaling pathways in lipopolysaccharide (LPS)-stimulated RAW 264.7 cells (Choi et al., 2017). In addition, pharmaceutical studies have revealed further pharmaceutical applications of *A. distichum* based on its antihypertensive (Oh et al., 2003), antidiabetic (Li et al., 2013), and anticancer (Park et al., 2014) activities. Furthermore, the presence of biologically active compounds, including acteoside, eutigoside B, isoacteoside, cornoside, rutin, chlorogenic acid, caffeic acid, gentisic acid, ferulic acid, and quercetin indicates the potential of *A. distichum* as a crude drug and dietary health supplement (Li et al., 2013; Choi et al., 2017).

* Corresponding author.
E-mail address: taekyung7708@chungbuk.ac.kr (T.K. Hyun).

https://doi.org/10.1016/j.sajb.2018.02.401
0254-6299/

In spite of its growing economic importance, with ornamental, nutritional and medicinal values, to our knowledge, only rudimentary sequence information is available in the public database (National Center for Biotechnology Information [NCBI], https://www.ncbi.nlm.nih.gov/) comprising of 18 nucleotide sequences, 133 genes (collected information about gene loci), 6 PopSet (sequence sets from phylogenetic and population studies) and 179 proteins.

Therefore, this study attempted to analyze and annotate the transcriptome of *A. distichum*, by assembling short paired-end Illumina reads. Overall, we generated 155.8 M sequence reads assembled into 48,623 unigenes with an average length of 631 bp. Based on BLASTX against the NCBI database, we found that 59.32% of all unigenes matched to known genes in various plants. In addition, comprehensive analysis suggested that flowers and leaves are main sources of flavonoid biosynthesis in *A. distichum*, and differences in gene expression affect the accumulation of secondary metabolites including flavonoids and anthocyanins. Moreover, we identified 1724 cDNA-derived simple sequence repeats (SSRs), which may facilitate genetic characterization and be useful for analyzing genetic diversity within the Oleaceae family.

2. Materials and methods

2.1. Plant materials and RNA isolation

A. distichum were grown in the fields of the Nature Environment Research Park of Gangwon Province, South Korea, under normal cultivation conditions. Various tissues, including flowers, leaves, and stems of *A. distichum* were collected. Total RNA was isolated from each sample using an Easy-spin™ llp Plant RNA Extraction Kit (Intron Biotechnology, Seongnam, South Korea), according to the manufacturer's instructions, and quantified using an Optizen Nano Q spectrophotometer (Mecasys, Deajeon, South Korea). The total RNA of each sample was then pooled in equivalent quantities for cDNA preparation.

2.2. Illumina sequencing and assembly

For transcriptome analysis, mRNA from total RNA was isolated with oligo(dT) magnetic beads and fragmented to short pieces. Then, the cDNA library was synthesized as described by Hyun et al. (2012) and sequenced on an Illumina HiSeq™ 2500 sequencing platform. The raw reads were cleaned with the removing adapter sequences, empty reads, low-quality reads (with ambiguous sequence, *N*), and reads with more than 10% Q < 20 bases (i.e., with a base quality of less than 20), using Trimmomatic v.0.33. After obtaining the clean data (157,580,490 clean paired-end reads), transcriptome assembly was accomplished by using Trinity software, with a default k-mer parameter (K = 25). Trinity contigs of high similarity were clustered into groups with CD-HIT-EST (v.4.6.1–2012–08-27), and gene open reading frames were predicted using Transdecoder v.2.0.1 (http://transdecoder.sourceforge.net/).

For functional annotation, the assembled unigenes were identified by sequence comparisons with the NCBI non-redundant protein (NR) database (http://www.ncbi.nlm.nig.gov), the Cluster of Orthologous Groups (COG) database (http://www.ncbi.nlm.nih.gov/COG/), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (http://www. Genome.jp/kegg), by applying BLAST searches with a cut-off E-value of 10^{-5}. With the NR annotation, the Blast2GO program was used to gain gene ontology (GO) annotation of the unigenes.

2.3. Mining for SSRs

The MISA (Microsatellite) Perl script (http://pgrc.ipk-gatersleben.de/misa) was used to identify microsatellites in the unigenes of *A. distichum*. The minimum SSR length criteria were defined as five iterations for di-nucleotide repeats, four iterations for tri-nucleotide repeats, and three iterations for the other repeat units.

2.4. Gene expression analysis by quantitative real-time PCR (qRT-PCR)

Total RNA isolated from the flowers, leaves, and stems was treated with gDNA remover for genomic DNA elimination, and reverse-transcribed into cDNA using the ReverTra Ace® qPCR RT Master Mix with qDNA Remover (Toyobo, Co., Ltd., Osaka, Japan), according to the manufacturer's recommendations. qRT-PCR was performed using the SYBR® Green Real-time PCR Master Mix (Toyobo, Co., Ltd., Osaka, Japan) in a CFX96™ Real-time system (Bio-Rad), with default parameters. The expression levels of each gene were normalized to the constitutive expression level of elongation factor 1-alpha (EF-1α). The specific primer pairs used in qRT-PCR are listed in Supplementary Table 1.

2.5. Determination of reducing power and oxygen radical antioxidant capacity (ORAC)

The reducing power assay was determined according to the method described by Lee et al. (2013). Ten milligrams of the freeze-dried materials were soaked in MeOH (1 ml) for 24 h. After centrifugation, the supernatant was used for determination of reducing power and ORAC. For the total reducing power of each MeOH extract, 30 μl of each extract were mixed with 0.2 ml of 200 mM sodium phosphate buffer (pH 6.6) and 0.2 ml of 1% potassium ferricyanide. After incubation at 50 °C for 20 min, the mixture was combined with 1 ml of 10% trichloroacetic acid, and then centrifuged at 6500 rpm for 10 min. Next, a 500 μl aliquot of the supernatant was mixed with 500 μl of deionized water and 100 μl of 0.1% ferric chloride, and the absorbance was measured at 750 nm using an iMark™ microplate reader (Bio-Rad). The reducing power of each sample was expressed as μg of BHT (butylated hydroxytoluene) equivalents per mg of dry weight.

The ORAC assay was conducted according to Choi et al. (2017). A total of 150 μl of 0.08 μM fluorescein diluted in phosphate buffer (75 mM, pH 7.0) was combined with 25 μl of phosphate buffer (blank), Trolox standard (6.25–50 μM), or each MeOH extract in separate wells of a microplate. After incubation at 37 °C for 10 min in dark, 25 μl of fresh 2,2′-azobis(isobutyramidine) dihydrochloride (0.12 g/ml) was added. The fluorescence intensity was monitored every minute for 90 min, using 485 and 525 nm excitation and emission wavelengths, respectively, in a SpectraMax Gemini EM microplate reader (Molecular Devices, CA, USA). Results were expressed as μmol of Trolox equivalents per g dry weight.

2.6. Analysis of total flavonoid and anthocyanin contents

The total flavonoid content was analyzed with the colorimetric method described by Hyun et al. (2013). An aliquot (50 μl) of each MeOH extract was mixed with 10 μl of 10% aluminum nitrate, 10 μl of 1 M potassium acetate, and 430 μl of 80% ethanol. The mixture was incubated at room temperature for 40 min, and then the absorbance was determined at 415 nm. The total flavonoid concentration was calculated in mg of quercetin equivalents (QE) per mg dry weight of sample.

The anthocyanins were quantified as described by Zhou et al. (2012). The absorbance was determined at 530 and 657 nm, respectively, using an Epoch microplate spectrophotometer (BioTek Instruments, Inc., VT, USA). Anthocyanin concentration was calculated using the following formula:

$$Q_{\text{Anthocyanin}} = (A_{530} - 0.25A_{657}) \times M^{-}1$$

where $Q_{\text{Anthocyanin}}$ is the relative concentration of anthocyanin, A_{530} and A_{657} are the absorptions at 530 and 657 nm, respectively, and M is the dry weight (g) of the plant materials used for extraction.