Journal of Integrative Agriculture 2017, 16(3): 591–604

Available online at www.sciencedirect.com

ScienceDirect

RESEARCH ARTICLE

High-throughput sequencing of highbush blueberry transcriptome and analysis of basic helix-loop-helix transcription factors

SONG Yang[1], LIU Hong-di[1], ZHOU Qiang[2], ZHANG Hong-jun[1], ZHANG Zhi-dong[2], LI Ya-dong[2], WANG Hai-bo[1], LIU Feng-zhi[1]

[1] *Research Institute of Pomology, Chinese Academy of Agricultural Sciences/Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (Germplasm Resources Utilization), Ministry of Agriculture, Xingcheng 125100, P.R.China*

[2] *College of Horticulture, Jilin Agricultural University, Changchun 130118, P.R.China*

Abstract

The highbush blueberry (*Vaccinium corymbosum*), Duke, was used to construct a *de novo* transcriptome sequence library and to perform data statistical analysis. Mega 4, CLC Sequence Viewer 6 software, and quantitative PCR were employed for bioinformatics and expression analyses of the basic helix-loop-helix (BHLH) transcription factors of the sequencing library. The results showed that 28.38 gigabytes of valid data were obtained from transcriptome sequencing and were assembled into 108 033 unigenes. Functional annotation showed that 32 244 unigenes were annotated into Clusters of Orthologous Groups (COG) and Gene Ontology (GO) databases, whereas the rest of the 75 789 unigenes had no matching information. By using COG and GO classification tools, sequences with annotation information were divided into 25 and 52 categories, respectively, which involved transport and metabolism, transcriptional regulation, and signal transduction. Analysis of the transcriptome library identified a total of 59 *BHLH* genes. Sequence analysis revealed that 55 genes of that contained a complete BHLH domain. Furthermore, phylogenetic analysis showed that *BHLH* genes of blueberry (Duke) could be divided into 13 sub-groups. PCR results showed that 45 genes were expressed at various developmental stages of buds, stems, leaves, flowers, and fruits, suggesting that the function of BHLH was associated with the development of different tissues and organs of blueberry, Duke. The present study would provided a foundation for further investigations on the classification and functions of the blueberry BHLH family.

Keywords: blueberry, bioinformatics, transcriptome sequencing, basic helix-loop-helix, transcription factor

1. Introduction

The highbush blueberry, a new fruit in China, contains multiple beneficial substances that play important roles in improving the immune system of humans and in scavenging free radicals (Basu *et al.* 2010; Rendeiro *et al.* 2012). Molecular biology methods are extensively being used to improve the quality of highbush blueberry. However, our understanding of the blueberry genome is currently limited, thus the study on the function of various genes is limitd.

Received 7 April, 2016 Accepted 1 August, 2016
SONG Yang, Mobile: +86-13504299435, E-mail: songyang1225@163.com; Correspondence LIU Feng-zhi, Mobile: +86-13904295109, E-mail: liufengzhi6699@126.com

doi: 10.1016/S2095-3119(16)61461-2

In recent years, the use of high-throughput sequencing, combined with techniques in bioinformatics, has provided new avenues and approaches to study the transcriptomes of various species. In particular, this technology has been proven to be suitable for species without genomic data. It has been widely applied for the study of crop quality and mechanisms of stress tolerance in different species, such as peaches (Zhou *et al.* 2014), grapes (Wang *et al.* 2014), oranges (Guo *et al.* 2015), bananas (Yang *et al.* 2015), and rice (Huang *et al.* 2014).

Several recent studies have explored transcriptome sequencing of highbush blueberry. Rowland *et al.* (2012) analyzed the relationship of transcriptomic data of leaves, fruits, and flower buds with cold acclimation and flower bud differentiation in blueberry, Bluecrop. Li *et al.* (2012) used the transcriptome sequencing method to identify differences between mature fruit skin and fruit flesh in the blueberry, Northland. Several significant differences in 22 different transcription factors were identified, which included 52 basic helix-loop-helix (BHLH) transcription factors.

BHLHs are a class of transcription factors that are involved in plant growth and development. And BHLHs consist of a highly conserved domain that recognizes and binds to the E-box (CANNTG) of target genes to control their expression (Feller *et al.* 2011). The *BHLH* genes of *Arabidopsis* are divided into 12 large groups, of which some genes in Group III, VII, and XII are better understood for their functions and mechanisms of action (Heim *et al.* 2003). Some of the *BHLH* genes in these three groups play important roles in low-temperature acclimation (Chinnusamy *et al.* 2003), gene expression induced by abscisic acid and jasmonic acid (Abe *et al.* 1997; Qi *et al.* 2011), regulation of flavonoid biosynthesis (Nesi *et al.* 2000; Zhang *et al.* 2003; Xu *et al.* 2013; Nemie-Feyissa *et al.* 2015), pistil development (Heisler *et al.* 2001; Rajani and Sundaresan 2001), phytochrome signal transduction (Martınez-Garcıa *et al.* 2000; Qiu *et al.* 2015), and brassinosteroid signal transduction (Friedrichsen *et al.* 2002). In *Arabidopsis*, Group III consists of *Inducer of CBF Expression 1* (*ICE1*), *MYC2*, *Transparent Testa 8* (*TT8*), *GLABRA 3* (*GL3*), and *Enhancer of Glabra 3* (*EGL3*). Group VII contains *ALCATRAZ* (*ALC*), *SPATULA* (*SPT*), *Phytochrome Interacting Factor 3* (*PIF3*), *PIF4*, *Phytochrome Interacting Factor 3-Like* 1 (*PIL1*), and *PIL2*. Group XII comprises *BR Enhanced Expression*.

Although blueberry transcriptome analysis has been previously reported, these have not been involved in the systematic classification of the BHLH family of transcription factors. This study investigated the transcriptomic data of different organs and BHLH family transcription factors in the highbush blueberry variety, Duke, to generate basic information on the regulatory mechanism underlying growth and development.

2. Materials and methods

2.1. Materials

From April 2015 to January 2016, tests were performed at the Horticultural Crop Germplasm Resources Laboratory (Xingcheng, Liaoning) of the Agriculture Section of the Fruit Tree Research Institute, Chinese Academy of Agricultural Sciences. The source materials were seven-year-old seedlings of blueberry (*Vaccinium corymbosum*), Duke, which was collected from a berry garden of the Fruit Tree Research Institute, Chinese Academy of Agricultural Sciences and a berry garden at Jilin Agricultural University, China. Top flower buds of one-year-old branches, one-year-old branches (top section), leaves, flowers (flowering), fruits at 40 d after flowering (green fruit), fruits at 65 d after flowering (pink fruits), and fruits at 80 d after flowering (blue fruits) were obtained. To maintain consistency of the source material, only fruits at the top of inflorescences were collected and immediately frozen in liquid nitrogen and then stored at –70°C.

2.2. Library construction and sequencing

Following the method of Cheng *et al.* (1993), total RNA was extracted from different tissues and organs. The integrity, purity, and quality of RNA were confirmed before library construction using 1.2% agarose gel electrophoresis, an Agilent 2100 analyzer (Aglient, USA) and NanoDrop spectrophotometer (Thermo Scientific, USA). Reverse transcription and quantitative PCR kits were purchased from Fermentas Co., USA. *Escherichia coli* DH5α cells were purchased from Beijing Tiangen Co., China. Primer synthesis and gene sequencing were performed by Shanghai Sangon Co., China.

Library construction and sequencing were performed by Beijing Biomarker Company (http://www.biomarker.com.cn). The procedures are described below. After total RNA was digested using DNase I, oligo(dT) beads were used to enrich mRNAs. Then mRNAs were cut into short segments, which were converted to single-stranded and double-stranded cDNAs. The resulting cDNAs were purified, their sticky ends repaired, and their 3′ ends ligated with “A” oligobases and then ligated to connectors. Finally, fragment sizes were selected, and the final library was constructed by PCR amplification. Qubit 2.0 and Agilent 2100 were used to respectively measure the concentrations and the inserted fragment sizes of the constructed library. Quantitative PCR was performed to accurately quantify the effective concentration of the library. After meeting the requirements for quality, an Illumina HiSeq 2500 sequencing machine was used to perform sequencing.