Gene 559 (2015) 155–163

Contents lists available at ScienceDirect

Gene

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

Gene set by *de novo* assembly of *Perilla* species and expression profiling between *P. frutescens* (L.) var. *frutescens* and var. *crispa*

Wei Tong [a,1], Soon-Jae Kwon [b,1], Jeongsoo Lee [c], Ik-Young Choi [c], Yong-Jin Park [a], Seung Hun Choi [d], Kyu Jin Sa [d], Byeong Wan Kim [d], Ju Kyong Lee [d,*]

[a] Department of Plant Resources, College of Industrial Science, Kongju National University, Yesan 340-702, Republic of Korea
[b] Advanced Radiation Technology Institute, Korea Atomic Energy Research Institute, Jeongeup 580-185, Republic of Korea
[c] National Instrumentation Center for Environmental Management, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-742, Republic of Korea
[d] Department of Applied Plant Sciences, Oriental Bio-herb Research Institute, Kangwon National University, Chuncheon 200-701, Republic of Korea

ARTICLE INFO

Article history:
Received 27 January 2014
Received in revised form 13 January 2015
Accepted 14 January 2015
Available online 15 January 2015

Keywords:
Perilla
Transcriptome
DEG profiling
Leaf color
Weedy and cultivar

ABSTRACT

Perilla frutescens (L.) Britt. is a self-pollinating annual species and is widely cultivated in China, Korea and Japan as an economic crop and a source of medicine and spices. In this study, we sequenced one cultivar variety (PF98095) of *P. frutescens* (L.) var. *frutescens* Britt., which was assembled as reference and other three varieties (PF11109, weedy of var. *frutescens*, PF06336 and PF06353, cultivars of *varieties crispa*) in order to carry out comparative expression profiling within cultivar and weedy in *varieties frutescens* and between *varieties frutescens* and *varieties crispa* of cultivar type in *P. frutescens*. Assembly of PF98095, annotation mapping, DEG (differentially expressed gene) profiling, and comparative analysis were performed. We found that more than 65% of the reads were mapped to the reference of *P. frutescens* gene set. Moreover, we detected 22,962 DEGs in the weedy variety compared to the cultivar, and also, 22,138 and 23,845 DEGs were identified in two cultivars according to the reference, respectively. The DEGs and functional classification were developed to analyze the differences between weedy and cultivar and between *varieties frutescens* and *varieties crispa* of *Perilla*. Furthermore, candidate genes for the different color and seed size of *Perilla* were identified that could be further investigated in future study. The herein results may play a significant role, and contribute in functional transcriptome studies of *Perilla*.

1. Introduction

Perilla frutescens (L.) Britt. is a self-fertilizing crop which is widely cultivated in East Asia, i.e. China, Korea and Japan. The species includes two cultivated types on the basis of their morphology and dual uses. The one is *P. frutescens* var. *frutescens* Britt., an oil crop (Ren in Chinese, Dlggae in Korean and Egoma in Japanese). The other one is *P. frutescens* var. *crispa* Decaisne, a Chinese medicine or vegetable crop (Zisu in Chinese, Chajoki in Korean and Shiso in Japanese) (Lee and Ohnishi, 2003). In Korea, var. *frutescens* is an important oil and vegetable crop, but the cultivation of var. *crispa* is very little due to the decreased use of its seeds in Chinese medicine. It is occasionally found in a relict form in Korea (Lee and Ohnishi, 2001; Lee and Kim, 2007). However, in Japan, var. *crispa* is extensively cultivated as a spicy vegetable crop, while the cultivation of var. *frutescens* has declined to small-scale operations in mountain areas (Nitta et al., 2003; Nitta and Ohnishi, 1999). Despite the geographic proximity between Korea and Japan, the cultivation and utilization of the two cultivated types of *Perilla* crop are significantly different (Lee and Ohnishi, 2001; Nitta et al., 2003). The seeds of var. *frutescens* are used like sesame seeds as a seasoning in Korea, China and Japan. These two cultivated types can be distinguished from each other based on their typical fragrance and the size and degree of hardness of their seeds (Nitta et al., 2003). In spite of distinct morphological characters between these two crops, they have being cross-fertile with each other by artificial pollination (Honda et al., 1990, 1994). Because there have been very few efforts to improve *Perilla* through breeding programs, several genotypes of *Perilla* crop still occur as cultivars in farmer's fields in several areas of South Korea (Park et al., 2005).

For maximizing the usage of *Perilla* germplasm resources and developing more efficient breeding programs, the morphological variations among cultivated and weedy types of *Perilla* crop need to be characterized. The variability of some *Perilla* genotypes has been described in terms of their morphology, cultivation, usage, and diversification (Lee and Ohnishi, 2001; Nitta et al., 2003, 2005), as well as their genetic diversity and relationships with weedy forms using random amplified

Abbreviations: DEG, differentially expressed gene; RAPD, random amplified polymorphic DNA; AFLP, amplified fragment length polymorphism; SSR, simple sequence repeat; NGS, next-generation sequencing; GO, gene ontology; RPKM, reads per kilo base of exon model per million mapped reads; KEGG, Kyoto Encyclopedia of Genes and Genomes; SAGE, serial analysis of gene expression; CAGE, cap analysis of gene expression; SNP, single nucleotide polymorphism.

* Corresponding author.
E-mail address: jukyonglee@kangwon.ac.kr (J.K. Lee).
[1] Equally contributed to this research.

http://dx.doi.org/10.1016/j.gene.2015.01.028
0378-1119/

polymorphic DNA (RAPD, Nitta and Ohnishi, 1999), amplified fragment length polymorphism (AFLP, Lee et al., 2002; Lee and Ohnishi, 2003) markers and sequence tagged microsatellite (STMS, Verma et al., 2010) markers. Kwon et al. (2005) isolated and characterized 13 polymorphic microsatellite primers from *P. frutescens* var. *frutescens* Britt. by using a modified method, which provided valuable resources for different genetic studies in *Perilla* genome research program. Jung et al. (2009) and Kim et al. (2011) reported that cultivated and weedy types of *Perilla* showed differentiable and variable germination. Morphological, RAPD, AFLP and SSR (Simple Sequence Repeats) analysis demonstrated that these weedy plants can be also grouped into two types: one belonging to the group of var. *frutescens* and another belonging to the group of var. *crispa* (Lee and Kim, 2007). However, these analyses are limited because of the small numbers of *Perilla* accessions used and loci analyzed. Distribution pattern of diversity and survey, exploration and germplasm collection of cultivated and weedy type *P. frutescens* (L.) var. *frutescens* Britt. and its uses in India Himalaya were developed, which also addressed the brief information on botanical characterization on the basis of systematic study of cultivated and weedy types and thrust on areas for germplasm collection from diverse habitats (Pandey et al., 2008). Sa et al. (2012) evaluated the morphological characteristics of 54 accessions of cultivated and the weedy types of *Perilla* species, and reported that the cultivated type of var. *frutescens* might be regarded as a more domesticated type than the cultivated type of var. *crispa*. Gong et al. (1997) illustrated that the structural genes that encode the enzymes of the biosynthetic pathway for anthocyanin are expressed in the red forma of *Perilla*, but not in the green forma. Two genes involved in vacuole transport of anthocyanin, *PfGST1*, encoding glutathione S-transferase (GST), and *PfCHI1*, encoding chalcone isomerase (CHI), have been identified through a PCR-select subtraction, which illustrated that the expression level of *PfGST1* and *PfCHI1* was higher in red *Perilla* (Yamazaki et al., 2008).

Recent advances in the next-generation sequencing (NGS) provided a fast, cost-effective, and reliable approach to generate large expression datasets for functional genomic analysis. RNA-Seq, a revolutionary advance in genome-scale sequencing, is a more comprehensive and efficient way to measure transcriptome composition, obtain RNA expression patterns, and discover new exons and genes (Mortazavi et al., 2008; Simon et al., 2009; Morozova et al., 2009; Wold and Myers, 2008; Wang et al., 2009; Yang et al., 2011). Since RNA-Seq is not limited to detecting transcripts that correspond to existing genomic sequences, it is particularly attractive for non-model organisms with genomic sequences that are yet to be determined (Vera et al., 2008; Kristiansson et al., 2009; Meyer et al., 2009; Wang et al., 2010). In addition, this new approach is very sensitive, allowing detections of low abundant transcripts.

In the present study, we sequenced and assembled one cultivar variety (PF98095) of *P. frutescens* (L.) var. *frutescens* Britt. Three additional varieties were sequenced (PF11109, weedy of *P. frutescens* (L.) var. *frutescens*, PF06336 and PF06353, cultivars of *P. frutescens* (L.) vars. *crispa*) in order to perform the comparative expression profiling between cultivar and weedy in *varieties frutescens* and between *varieties frutescens* and *varieties crispa* of cultivar type in *P. frutescens*. The comparative DEG profiling and functional classification based on GO and KEGG between weedy and cultivar as well as the DEGs between cultivars of *varieties frutescens* and *varieties crispa* were predicted. Moreover, contigs and DEGs indicating different leaf colors and seed sizes of *Perilla* were identified, putting in evidence some candidate genes or relevant proteins that could be further investigated. Our data and observations may provide an important resource for future studies upon *P. frutescens* var. *frutescens* Britt. and *P. frutescens* var. *crispa* Decaisne.

2. Materials and methods

2.1. Plant material and sequencing

Two plants of *P. frutescens* var. *frutescens* Britt. (PF98095 and PF11109) and two plants of *P. frutescens* var. *crispa* Decaisne (PF06336 and PF06353) were collected from Korea and Japan as shown in Table 1. Of the four samples, PF11109 is weedy-type, while the other three are cultivars. The leaves of PF98095 and PF11109 are green, and PF06336 and PF06353 are green and red, respectively. After fostered and harvesting, seeds of four plants were stored at 4 °C and then germinated at 25 °C under illumination at a 16 h light period in a greenhouse. After about 5 weeks, when the leaves had grown to a length of about 3 cm, the younger leaves were used for isolation of RNA. RNA isolation of the samples was performed using TRIzol reagent according to the instructions of the manufacturer. The extracted RNA were checked using RNA Pico Chip on Agilent 2100 Bioanalyzer (Agilent technologies, Santa Clara, CA, US) to estimate the relative purity and concentration.

RNA-Seq data of the four samples were generated using Illumina HiSeq 2000 sequencer platform (Illumina CA, US) according to the manufacture's library and sequencing protocol. Briefly, a total of 10 μg total RNA were initiated and cDNA fragments were isolated and enriched by PCR to create the final cDNA library. Pools of four samples were sequenced on a HiSeq 2000 for 100 bp paired-end cycles (2 × 101 bp) following the manufacturer's instructions (Illumina, San Diego, CA). The processing of fluorescent images into sequences, base-calling, and quality value calculations were performed using the Illumina data processing pipeline (version 1.8).

2.2. Transcriptomic reference construction and annotation

For the construction of the transcriptome reference set, the *de novo* assembly of the PF98095 RNA-Seq data was performed using Trinity software (http://TrinityRNASeq.sourceforge.net). The raw reads from NGS sequencing were filtered with at least Phred quality score 20 and with at least read length 50 bp of HiSeq 2000 data before assembly.

Contigs were annotated using BlastX which is based on sequence similarity to known proteins of the non-redundant (NR) database in NCBI (http://www.ncbi.nlm.nih.gov/). Afterwards, the genes were classified in base of their gene ontology (GO) terms, as obtained using the Blast2GO software (ver. 2.7) with mysql DBMS (ver. 5.0.77).

2.3. Comparative DEG profiling and functional classification

Normalization is used to remove non-biological influences on biological data, and to make data comparable between experiments, runs, and lanes. It enables accurate comparisons of the expression levels between and within samples. The expression levels of transcripts from RNA-Seq were normalized by RPKM method (reads per kilo base of exon model per million mapped reads), which is a formula to quantify gene expression by Mortazavi et al. (2008). The cutoff value for determining gene transcriptional activity was determined based on a 95% confidence interval for all RPKM values. To deduce the putative gene function, transcripts were subjected to BlastX analysis against a UniProt database. The putative functions of query transcripts were defined by the first subject hits.

After RPKM data normalization, DEGseq package (Wang and Wang, 2013) was applied to estimate the differential gene expression. In this study, differentially expressed genes (DEGs) were assessed by fold-change (FC) and $\log_2$ values of FC. Genes that exhibited an estimated absolute $\log_2$ ratio ≥ 1 (more than 2-fold change) were considered to be differentially expressed genes. PF11109, PF06336 and PF06353 were compared to the reference gene set derived from PF98095 and the DEGs were characterized according to the $\log_2$ ratio 1 to 7, respectively. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed to classify the DEGs and seek significantly enriched pathway terms among DEGs. Candidate genes (contigs) and DEGs related to different leaf colors and seed sizes