Scientia Horticulturae 239 (2018) 171–180

Contents lists available at ScienceDirect

Scientia Horticulturae

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

Transcriptome analysis reveals the hawthorn response to the infection of apple chlorotic leaf spot virus

Wei Guo, Nan Guo, Weijia Li, Hongyan Dai*

College of Horticulture, Shenyang Agricultural University, Shenyang, 110866, China

ARTICLE INFO

Keywords:
Hawthorn
ACLSV
High-throughput sequencing
Differentially expressed genes
Plant-pathogen interaction

ABSTRACT

Viral diseases have become an important disease in the growth and development of fruit trees. Apple chlorotic leaf spot virus (ACLSV) is one of the most important viruses that harm fruit trees. We performed a high-throughput sequencing study to identify the differentially expressed genes (DEGs) induced by ACLSV in hawthorn using uninfected plants as a control. The results indicated that we have identified 125,469 unigenes, of which 9910 were expressed differentially. GO and KEGG enrichment analysis of the DEGs revealed that most DEGs were in the "stimulus response" GO term of "biological processes" and the "plant-pathogen interaction" KEGG pathway. Quantitative RT-PCR analysis about some defense-related genes showed a high consistency with the transcriptome data. This is the first study characterizing changes in plants when challenged with ACLSV through the transcriptome.

1. Introduction

Hawthorn (*Crataegus pinnatifida*) belongs to the subfamily Maloideae of the Rosaceae, which is distributed in the northern temperate zone worldwide (Rigelsky and Sweet, 2002). At present, there are a thousand varieties of hawthorn in the world, and the utilization of its edible and medicinal value in China has lasted for 2500 years (Dai et al., 2007). Hawthorn fruit contains 18 amino acids and various mineral nutrients that are essential to the human body, and its vitamin C content is much higher than that in pears and apples. All the fruits and leaves of hawthorn contain medicinal ingredients, which lower the blood pressure, cholesterol, and blood lipids and protect blood vessels (Rigelsky and Sweet, 2002; Kao et al., 2005). In addition to the high economic and medicinal value, hawthorn trees play a role in environmental protection and virescence and are the pioneer tree species of soil and water conservation.

Viruses are an important pathogen that endangers perennial fruit trees. After infection, the physiological functions of fruit trees are severely damaged, the growth is weakened, the yield is reduced, and the quality is deteriorated (Pusey, 1991). Virus disease can cause long-term damage and even death to fruit trees, which is an enormous barrier to fruit tree production. Due to the severity, long-term effects and the difficulty of prevention, virus diseases of fruit trees have been of concern and value worldwide (Desvignes et al., 1999). Apple chlorotic leaf spot virus (ACLSV) is one of the severely damaging orchard latent viruses. ACLSV has a wide host range and is distributed in fruit-producing areas around the world (Marini et al., 2008; Martelli et al., 1994). ACLSV can induce serious viral symptoms on stone fruit trees, such as the dark green hollow spot and the deformation of the leaves and fruits of peach, the bark split of plum, and the graft incompatibility of apricot (Dunez et al., 1972; Desvignes and Boyé, 1989; Hounsa et al., 1995). Because ACLSV usually does not show obvious symptoms in pome fruit trees, such as apples, pears, and hawthorns, it is difficult to detect the virus in the early stages of infection (Rana et al., 2010). However, with the development of the tree, the virus accumulates year by year. The disease can weaken the growth of trees, decrease the yield and quality, and cause a serious loss in fruit production. Sometimes, the orchards have no harvest because the trees die early (Gella, 1989).

At present, transcriptional experimental techniques mainly include gene chips based on hybridized technology and RNA-Seq based on high-throughput sequencing. RNA-Seq, which is widely considered to be superior to microarray-based methods, has advantages in terms of test repeatability, wide detection range and quantitative accuracy (Zhao and Feng, 1996). RNA-Seq is not only applicable to species with known genome sequences but also to species with unknown genome sequences. In this way, plant genes are gradually excavated, and the genetic networks of plant anabolism is more and more understood. This provides a rich database for comprehensively studying the plant immune mechanism and the influence of pathogens on plants. Wang et al. (2017) found that many genes involved in cellulose biosynthesis, protein amino acid phosphorylation, protein biosynthesis, protein glycosylation, glycolysis and cellular macromolecular complex assembly were

* Correspondence author.
E-mail address: daihy1970@163.com (H. Dai).

https://doi.org/10.1016/j.scienta.2018.05.017
Received 9 February 2018; Received in revised form 4 April 2018; Accepted 8 May 2018

up-regulated in *Brachypodium distachyon* infected with barley stripe mosaic virus compared with the controls, which may be related to the replication and movement of the viruses. In previous report, 3442 differentially expressed genes (DEGs) were screened from disease-resistant cotton variety vaccinated with the *Verticillium* wilt fungus V991 strain and the unvaccinated same variety plant, which revealed the lignin and phenylalanine pathway plays an important role in plant defense responses against *Verticillium* wilt bacteria (Xu et al., 2011). Comparative transcriptome analysis of leaf spot susceptible and resistant buffalograss lines challenged with *Curvularia inaequalis* identified twenty-one transcripts that regulate plant immunity and three transcripts encoding pathogenesis-related (PR) proteins downstream of systemic acquired resistance (Amaradasa and Amundsen, 2016).

Due to the sequencing depth and the huge amounts of high-throughput sequencing data, the tested tissues may be infected with pathogens, and the sequencing data may contain sequence information of viruses, viroids, or other pathogens. Therefore, high-throughput sequencing can also be applied to the discovery and assembly of plant virus genomes. So far, high-throughput sequencing has become a common tool for finding known or unknown viruses, and reports of discovering known and unknown viruses by high-throughput sequencing are commonplace. In a recent study, the full-length sequence of apple stem grooving virus was obtained from apple and pear transcriptome data (Jo et al., 2016). Previous experiments in our laboratory also assembled three ACLSV isolates of hawthorn using transcriptome data (Guo et al., 2016).

However, the current research on ACLSV mainly focuses on the identification of isolates, virus detection and detoxification techniques, but little has been reported about the pathogenic mechanism of ACLSV. To understand the pathogenesis of ACLSV and explore the effects of ACLSV on the hawthorn host, we selected the same variety of hawthorn plants infected with ACLSV and uninfected for transcriptome sequencing. It turned out that DEGs existed in many biological pathways, and "plant-pathogen interaction" pathway enriched maximum amount different genes. These results help us understand and control the mechanism of hawthorn response to ACLSV.

2. Materials and methods

2.1. Sample preparation

The trees of hawthorn (*Crataegus pinnatifida*) variety 'Kaiyuanruanzi' used in this research were maintained in the National Hawthorn Germplasm Repository in Shenyang Agricultural University. Samples collected from the ACLSV infected and uninfected mature leaves of hawthorn trees at 35 days after bloom were used for transcriptome sequencing and quantitative RT-PCR (qRT-PCR) testing. The infected hawthorn was sequenced previously, and only one virus, ACLSV, was annotated (Wang and Dai, 2014). We also detected the two hawthorn materials with high sensitivity TaqMan probe qRT-PCR, one for the negative and one for the positive (Zheng et al., 2016). So we called the ACLSV-infected samples "P_leaf" and the healthy samples "N_leaf". Therefore, the material was suitable for subsequent experiments and data analysis.

2.2. cDNA library preparation and illumina sequencing

Total RNA samples of ACLSV-infected and healthy hawthorn leaves in three biological replicates were submitted to Novogene Bioinformation Technology Co., Ltd, Beijing, China, for cDNA library preparation and high-throughput sequencing. The cDNA library was sequenced on an Illumina sequencing platform (HiSeq™ 4000).

2.3. De novo assembly

Raw data (raw reads) were filtered by removing the reads containing an adaptor, including the reads whose N content was more than 0.1% (N is the base unable to determine) and low-quality reads whose quality score Q was no more than 20, accounting for more than 50% of the whole reads. The remaining data were termed the clean data (clean reads). The clean data was de novo assembled using Trinity software (Grabherr et al., 2011) to construct the full-length transcript, the longest transcript of each cluster was taken as a unigene. The Illumina data set has been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under accession number SRP131934.

2.4. Functional annotation

Unigenes were annotated using BLASTx against seven databases. These databases include the NCBI non-redundant (NR) protein and nucleotide (NT) databases (http://www.ncbi.nlm.nih.gov), the Swiss-Prot protein database (http://www.ebi.ac.uk/uniprot/), the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (http://www.genome.jp/kegg), the Cluster of Orthologous Groups (COG) database (http://www.ncbi.nlm.nih.gov/COG), and the Protein family (Pfam) database (http://pfam.sanger.ac.uk/). The alignments from the NR database were used in blast2GO (https://www.blast2go.com/) to obtain Gene Ontology (GO) annotation (http://www.geneontology.org/) (Ashburner et al., 2000). GO functional classification of all the unigenes was used to view the distribution of the gene functions.

2.5. Differentially expressed genes analysis

The clean reads of each sample were mapped to the reference sequence assembled by Trinity. Gene expression levels were measured in RNA-Seq analysis as reads per kilobase of transcript per million mapped reads (RPKM) (Mortazavi et al., 2008). DESeq Software (Anders and Huber, 2010) was used to identify DEGs with the screening threshold for padj < 0.01 (padj is the adjusted P value, also called q value) and two-fold change (log2 ratios ≥ 1). These DEGs were used for subsequent function clustering analysis.

2.6. Quantitative RT-PCR analysis

A total of 1 μg of RNA was used to synthesize the cDNA with the PrimeScript® RT reagent kit (TaKaRa, Dalian, China) according to the manufacturer's instructions. The cDNA was diluted five times as the template for the reaction. qPCR was conduct in an ABI 7500 Real Time PCR Detection System (Applied Biosystems, US). The primers for DEGs validation were synthesized by GENEWIZ, Inc. (Suzhou, China) and are given in Supplementary Table 1. The qRT-PCR reaction was carried out in a final volume of 20 μL containing 10 μL Real Master Mix SYBR Green (CWBIO, China), 1 μL each forward and reverse primers (10 μM), and 1 μL diluted cDNA. The PCR reaction conditions were denaturation at 95 °C for 5 min followed by 40 cycles of 95 °C for 10 s, annealing at the appropriate temperature (from 58 to 60 °C) for 30 s, extension at 72 °C for 30 s, following melting curve range from 60 to 95 °C. All reactions were done in triplicate. An endogenous *ACTIN* gene was used for the normalization (Dai et al., 2013). The relative fold changes (P_leaf/N_leaf) in genes expression were calculated using the comparative Ct ($2^{-\Delta\Delta Ct}$) method.

3. Results

3.1. RNA-Seq and de novo transcriptome assembly

Six samples were chosen for transcriptome sequencing, with three duplicates for each group of infected and uninfected virus samples. We obtained 42.03 Gb of clean data after filtering the redundant data, and each sample encompassed approximately 7 Gb of clean data. Of the clean reads, the GC content is approximately 47%, and more than