Ecotoxicology and Environmental Safety 158 (2018) 18–27

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

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

Comprehensive transcriptome profiling of soybean leaves in response to simulated acid rain

Long Yang[1], Yongchao Xu[1], Rui Zhang, Xiaotong Wang, Chao Yang*

State Key Laboratory of Crop Biology, Shandong Agricultural University, Tai'an, China

ARTICLE INFO

Keywords:
Simulated acid rain (SAR)
RNA-Seq
Soybean
Transcriptomic study
Gene network

ABSTRACT

As a source of edible oil and protein, soybean is a major globally important economic crop; Improving its production has been an important objective of soybean breeding. Acid rain has been shown to influence soybean growth and productivity, with consequent adverse impacts on its production for use by human populations. In this study, RNA sequencing technology was utilized to examine changes in gene expression when soybean was exposed to simulated acid rain (SAR). We sampled soybean leaves at five time intervals (0, 6, 30, 54, 78, and 102 h), and built the cDNA library. In total, 54,175 expression genes were found, including 2016 genes with differential expression. A total of 416 genes were considered, as they were closely related to the response to SAR. Genes related to the regulation of sulfur and nitrogen metabolism, carbohydrate metabolism, photosynthesis, and reactive oxygen species were among those differentially expressed in response to SAR. In this study, we examined the response mechanisms of soybean under SAR exposure. Our findings will improve our understanding of the molecular mechanisms employed by soybean in responding to abiotic stress, and therefore provides important information in developing soybean breeding to improve tolerance to these stresses.

1. Introduction

In recent decades, acid rain, ozone depletion, and global warming (the greenhouse effect) have emerged as three major global environmental problems (Abbasi et al., 2013). Public attention has increasingly focused on acid rain, due to the increasing degradation it causes and low solubility. The term 'acid rain' was first used by Smith in 1872, in his book Air and Rain: The Beginnings of Chemical Climatology, published in that year (Smith, 1872). With the development of global industry, the occurrence of acid rain has been aggravated by overconsumption of fossil fuel. After Europe and North America, China is considered to be the third largest area affected by acid rain in the world (Hogan, 1998; Larssen et al., 1999). The damage caused by acid rain has wide-ranging and serious impacts on the natural world, including various ecosystems and human populations, as well as on buildings (BÄCk et al., 1994; Grantz et al., 2003; Liu and Liu, 2011). Plant respond both directly and indirectly to acid rain. There have been numerous academic studies which have found the effects of acid rain-related damage on plants to be multiple and complex, including inhibiting seed germination, inducing leaf necrosis, altering respiration and photosynthesis, and decreasing crop yield (Dise and Verry, 2001; Dolatabadian et al., 2013; Evans, 1982; Gabara et al., 2003; Jagels et al., 2002; Neves et al., 2009; Velikova et al., 1998). However, previous studies exploring the effect of acid rain on plants have to date mainly focused on the physiological and biochemical levels, while gene expression patterns under acid rain stress have merely been reported for Arabidopsis (Lee et al., 2006; Liu et al., 2012). It is therefore little understanding of the molecular mechanisms involved in the effects of acid rain on crops.

Multiple technologies have been developed in the past few years that are effective in unveiling the molecular basis of many biological processes in living organisms. These include microarrays, cDNA or expressed sequence tag (EST) sequencing, and serial analysis of gene expression (SAGE), approaches which have provided huge gene datasets at the transcriptome level (Ward et al., 2012). These sequencing technologies exhibit some common drawbacks, such as low throughput, high cost, low sensitivity, high background signal, and cloning bias (Bellin et al., 2009; Wang et al., 2016). However, the introduction of next-generation high-throughput RNA sequencing technology (RNA-Seq) has remedied these disadvantages; its sequencing depth and sufficient sensitivity makes it more suitable and powerful in discovering new gene expression profiles than previously developed technologies (Li et al., 2014; Wang et al., 2009). RNA-Seq has been making a significant contribution to the understanding of molecular mechanisms in

* Corresponding author.
E-mail address: ileman@sdau.edu.cn (C. Yang).
[1] Equal contributors.

https://doi.org/10.1016/j.ecoenv.2018.04.015
Received 31 January 2018; Received in revised form 4 April 2018; Accepted 6 April 2018

many plant species (Duhoux et al., 2015; Xu et al., 2013; Yuan et al., 2016).

As a source of edible oil and protein, soybean is a globally important economic crop; improving its production has been an important objective of soybean breeding (Yang et al., 2011). Acid rain has been shown to influence soybean growth and productivity, with consequent adverse impacts on its production for use by human populations. A large number of studies on soybean cultivars were undertaken in the 1980s, especially in the USA and Japan, which provide a comprehensive analysis of soybean growth parameters, agronomic traits, yield components, and seed quality after exposure to acid rain (Evans et al., 1984, 1983, 1985, 1986; Kohno and Kobayashi, 1989; Porter et al., 1989; Smith et al., 1991). However, the response to acid rain stress in soybean is still poorly understood at the transcriptome level. In the present study, we used RNA-seq to illuminate the transcriptomic responses to acid rain in soybean. In the transcriptomes of leaf samples from plants treated with simulated acid rain (SAR) or (untreated) control plants, a total of 385 unigenes were found to be highly relevant to SAR response and tolerance. Moreover, the findings were further explicated, and a putative model was proposed for the regulatory network under conditions of SAR stress. These results could expand the horizons of our understanding of the molecular mechanisms that operate under conditions of SAR stress, and provide information useful to enhancing stress tolerance in soybean seeding strategies.

2. Materials and methods

2.1. Plant materials and SAR treatment

Soybean (Glycine max cv. Nannong 1138-2) plants were grown at Shandong Agricultural University under normal soybean cultivation conditions. Four-week-old soybean seedlings were treated with SAR. In our pre-experiments, the growth of soybean seedlings was obviously inhibited by SAR with a pH of 2.5. Therefore, the SAR solutions used in the experiment were adjusted to pH 2.5 using a mixture of HNO_3 and H_2SO_4 in the ratio of 5:1, and prepared at the following concentrations (mg L-1): $CaCl_2$, 3.1; NH_4Cl, 2.8; NaCl, 0.91; KCl, 0.75. Once they had grown for 25 days (the third true leaf was developed), soybean plants were sprayed evenly for six days (at 9:00 a.m. each day) with 10 ml SAR per plant. The leaf samples were collected at stages of 0, 6, 30, 54, 78, and 102 h, and immediately frozen in liquid nitrogen for RNA extraction.

2.2. RNA preparation, library construction, and sequencing

Total RNA for transcriptome sequencing was isolated from each sample using TRIzol (Invitrogen, CA, USA) according to the manufacturer's instructions. The purity of the total RNA was checked using a NanoPhotometer® spectrophotometer (Implen, CA, USA), and RNA integrity and concentration were evaluated using an RNA Nano 6000 Assay Kit, which is part of the Bioanalyzer 2100 System (Agilent Technologies, Palo Alto, CA, USA). Briefly, 3 μg RNA per sample was used to prepare the RNA samples.

The RNA samples for sequencing library construction were generated using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs (NEB), MA, USA) according to the manufacturer's recommendations, and index codes were added in order to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo attached magnetic beads, and then broken into small fragments using divalent cations under conditions of elevated temperature in a NEB Next First Strand Synthesis reaction buffer (5 ×) (NEB). First-strand cDNA was synthesized using a random hexamer primer and M-MuLV Reverse Transcriptase (NEB), followed by second-strand cDNA synthesis using buffer, deoxyribonucleotide triphosphates (dNTPs), DNA polymerase I, and Ribonuclease H (RNase H). Suitable fragments (150–200 bp in length) were purified with the AMPure XP system (Beckman Coulter, Beverly, USA), and library quality was assessed on the Agilent Bioanalyzer 2100 System. Cluster amplification was conducted with a cBot Cluster Generation System using a TruSeq PE Cluster Kit v3-cBot-HS (Illumina2500, CA, USA), according to the manufacturer's instructions. Finally, 12 cDNA libraries of six samples (two replicates) at the different stages of SAR treatment were sequenced on the Illumina Hiseq. 2500 platform, and 125 bp/150 bp paired-end reads were generated.

2.3. Sequencing data analysis

2.3.1. Quality control

Raw data in fastq format were firstly processed. In this step, clean data were obtained from raw data by removing low-quality data (reads with ambiguous bases (N) and fragments of less than 40 bp in length), and rRNA reads using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). At the same time, the guanine-cytosine (GC) content and quality scores (Q20, Q30) of the clean data were calculated. All the downstream analyses were based on clean data with high quality.

2.3.2. Differential expression analysis

Paired-end clean reads were aligned to the reference soybean genome (Schmutz et al., 2010b) using TopHat v2.0.12, and differential gene expression was calculated with the software program Cufflinks (Trapnell et al., 2012) using RPKM (reads per kilo base of the exon model per million mapped reads). Filter criteria of fold change ≥ 2.0 and a false discovery rate (FDR) of ≤ 0.001 were adopted to determine the significance of the differential gene expression. All unigenes were analyzed via sequence alignment with the NR (NCBI Non-redundant Protein), GO (Gene Ontology), COG (Clusters of Orthologous Groups), and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases, using BLAST2GO (Conesa et al., 2005).

2.4. Quantitative real-time PCR (qRT-PCR) analysis

Total RNA was isolated from different soybean samples using the RNeasy Plus Mini Kit (Qiagen, Valencia, CA, USA). The PrimeScript RT reagent kit with gDNA Eraser (Takara, Dalian, China) was used to synthesize the first-strand cDNA for qRT-PCR analysis. In order to verify the expression profiles, qRT-PCR analysis of 11 cDNA samples was performed using SYBR Premix Ex Taq (Takara). The qRT-PCR mixture consisted of 10 μL Realtime PCR Master Mix, 0.8 μL cDNA sample, 1 μL gene-specific forward and reverse primers (Supplementary Tables S1), and 8.2 μL nuclease-free water. The tubulin housekeeping gene was used as the internal control (Supplementary Table S1). The thermal cycling protocol comprised an initial heating step at 95 °C for 10 min, followed by 39 cycles of three-step reactions (95 °C for 20 s, 56 °C for 50 s, and 72 °C for 30 s), and a final extension at 72 °C for 10 min. Three replicates of each reaction were carried out, and the relative expression levels were calculated from cycle threshold values using the $2^{-\Delta\Delta CT}$ method.

3. Results and discussion

3.1. RNA-sequencing and analysis of six developmental stages under SAR

We sequenced 12 cDNA libraries of six samples with two repeats under SAR treatment at different stages; a total of 32.66 Gb bases were sequenced using the Hiseq. 2500 platform (Illumina). Approximately 29.44 G reads with 87.38% Q30 bases and a GC content of 48.00% were obtained after stringent quality checking and data cleansing, based on 109 million paired-end raw reads (Table 1). The saturation sequencing was deemed sufficient for quantitative analysis of gene expression. On average, 77.31% clean reads were aligned to the soybean reference genome (Schmutz et al., 2010a), and matched either unique or multiple genomic locations. A total of 54,175 gene expression were mapped to