Gene 525 (2013) 26–34

Contents lists available at SciVerse ScienceDirect

Gene

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

De novo transcriptome sequencing and comparative analysis of differentially expressed genes in *Gossypium aridum* under salt stress

Peng Xu [1], Zhangwei Liu [1], Xinqi Fan, Jin Gao, Xia Zhang, Xianggui Zhang, Xinlian Shen *

Key Laboratory of Cotton and Rapeseed (Nanjing), Ministry of Agriculture, PR China
The Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, PR China

ARTICLE INFO

Article history:
Accepted 22 April 2013
Available online 4 May 2013

Keywords:
RNA sequencing
Gene expression profiles
Salinity

ABSTRACT

Salinity stress is one of the most serious factors that impede the growth and development of various crops. Wild *Gossypium* species, which are remarkably tolerant to salt water immersion, are valuable resources for understanding salt tolerance mechanisms of *Gossypium* and improving salinity resistance in upland cotton. To generate a broad survey of genes with altered expression during various stages of salt stress, a mixed RNA sample was prepared from the roots and leaves of *Gossypium aridum* plants subjected to salt stress. The transcripts were sequenced using the Illumina sequencing platform. After cleaning and quality checks, approximately 41.5 million clean reads were obtained. Finally, these reads were eventually assembled into 98,989 unigenes with a mean size of 452 bp. All unigenes were compared to known cluster of orthologous groups (COG) sequences to predict and classify the possible functions of these genes, which were classified into at least 25 molecular families. Variations in gene expression were then examined after exposing the plants to 200 mM NaCl for 3, 12, 72 or 144 h. Sequencing depths of approximately six million raw tags were achieved for each of the five stages of salt stress. There were 2634 (1513 up-regulated/1121 down-regulated), 2449 (1586 up-regulated/863 down-regulated), 2271 (946 up-regulated/1325 down-regulated) and 3352 (933 up-regulated/2419 down-regulated) genes that were differentially expressed after exposure to NaCl for 3, 12, 72 and 144 h, respectively. Digital gene expression analysis indicated that pathways involved in "transport", "response to hormone stimulus" and "signaling" play important roles during salt stress, while genes involved in "protein kinase activity" and "transporter activity" undergo major changes in expression during early and later stages of salt stress, respectively.

1. Introduction

Salinity and drought are the main environmental stressors that severely affect plant growth and development. The mechanism of salt tolerance is one of the most important subjects in plant science, as salt stress decreases worldwide agricultural production. Over the long course of evolution, plants have developed several salt tolerance mechanisms, including ion homeostasis, osmotic homeostasis, stress damage control and repair, growth regulation and others (Zhu, 2002). The exposure of plants to salty environment results in many physiological, biochemical and molecular changes involving massive changes in gene expression. Transcript analysis allows the simultaneous detection of changes in the expression of numerous genes and enables the elucidation of salt tolerance mechanism with the ultimate goal of improving crop productivity in saline soil.

Cotton is an important cash crop worldwide and is a major source of fiber. Cotton is an excellent system for the study of the molecular basis of plant responses to water deficit and salinity, as it originates from wild perennial plants that are adapted to semiarid, subtropical environments that experience periodic drought and temperature extremes and grow in soils with high salt content (Kohel, 1974). The geographical distributions of the wild amphidiploids are located on or near the path of prevailing currents in both the Pacific and Atlantic oceans. They have floated there or been carried there intentionally or accidentally by humans. Long-distance dispersal has played an important role in the diversification of major evolutionary lines and in speciation within *Gossypium* lineages. Wild species seeds are tolerant to salt water, which has made long-distance dispersal possible (Wendel et al., 2010). However, modern cotton cultivars, which are often grown under unstressed conditions, are the result of intensive selection to produce

Abbreviations: COG, the cluster of orthologous groups; GO, Gene Ontology; KEGG, The Kyoto Encyclopedia of Genes and Genomes; DGE, digital gene expression; FDR, false discovery rates; LRPK, LRR receptor protein kinase; CPK32, calcium dependent protein kinase 32; SAPK, stress-activated protein kinase; LHT1, lysine histidine transporter 1; SSH, suppression subtractive hybridization; PRR, pseudo type-B response regulator; CML, CaM-like proteins; PCR, polymerase chain reaction; BLAST, basic local alignment search tool; TF, transcription factor.

* Corresponding author. Tel.: +86 25 84390291; fax: +86 25 84390366.
E-mail address: shenxinlian@yahoo.com.cn (X. Shen).

[1] These authors contributed equally to this work.

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.04.066

large amounts of specific types of fiber. Selection has unintentionally narrowed the genetic variability for salt tolerance. Liu et al. (1993) screened more than 3000 upland cotton lines for salt tolerance and identified only three lines that were tolerant to 0.4% NaCl stress condition during the seedling stage. Therefore, improving salt tolerance in upland cotton requires the use of valuable alleles from wild *Gossypium* species.

Gossypium aridum is a D-genome diploid species that grows in the Pacific coastal states of Mexico. Many wild *Gossypium* sp. are becoming extremely rare due to the disappearance of *in situ* areas. However, *G. aridum* appears not to be threatened according to a recent expedition carried out by some American scientists (Ulloa et al., 2006), probably because of the great diversity which is attributed by the ability of abiotic resistance including drought and salinity. Experiments on seed viability of coastal and inland forms of *Gossypium* showed that the seeds of most coastal ecotypes are remarkably tolerant to salt water immersion (Stephens, 1958). In fact, the germination rate of *G. aridum* seeds can reach 92% in 1.2% NaCl solution (data not shown). *G. aridum* is therefore a valuable wild species for understanding the mechanisms underlying salt tolerance in *Gossypium* sp. and improving salinity resistance in upland cotton. To date, however, only 55 nucleotide sequences or Expressed Sequence Tags (ESTs) from *G. aridum* have been deposited into the NCBI database. No information is available on the genome-wide response of *G. aridum* to salt stress.

In recent years, next-generation high-throughput DNA sequencing techniques have been valuable tools for life sciences research and have dramatically improved the efficiency and speed of gene discovery (Ansorge, 2009; Asmann et al., 2009; Gahlan et al., 2012; Hegedus et al., 2009; Hershkovitz et al., 2013; Mortazavi et al., 2008; Stockhammer et al., 2009; Sultan et al., 2008; Venturini et al., 2013; Zhang et al., 2010). In this study, we first utilized paired-end sequencing technology to characterize the root and leaf transcriptomes of *G. aridum* under salt stress, to lay a foundation for future experiments and to create a sequencing resource. We then compared the gene expression profiles of *G. aridum* during different stages of salt stress using a digital gene-expression system to obtain a list of candidate genes related to salt tolerance. The research will enhance understanding the mechanisms of salt tolerance mechanism in *Gossypium* sp. The functional genes identified in this study would be the important gene resource to improve salt tolerance in tetraploid cotton by transgenic technology.

2. Materials and methods

2.1. Plant materials

The wild *Gossypium* species *G. aridum* seeds were kindly supplied by the National Wild Cotton Plantation in Hainan Island, China. Seeds were pre-germinated in distilled deionized water in a plant growth chamber at 60% humidity and day/night temperature of 28/23 °C, photoperiod of 12 h light/12 h dark for 24 h. Then germinated seeds were planted into soil and cultured in the same chamber with the same conditions. The uniformly developed seedlings, approximately 20 cm in height with four true leaves were transferred into 1× Hoagland's nutrient solution and grew for additional 3 days in 100 ml test tubes. For transcriptome analysis, in order to create a sequence resource under high salt shock, uniform plants were treated for 12 h with a relatively high concentration of 300 mM NaCl. Leaf and root tissues were collected from treated and control plants. For DGE analysis, in order to employ a wide range of treatment times, *G. aridum* plants were treated with 200 mM NaCl for 3, 12, 72 and 144 h, and unstressed plants were used for the control. Root tissues were collected at every stage of stress. The treated plants were labeled G3 (3 h), G12 (12 h), G72 (72 h) and G144 (144 h), and the control was labeled G0. All tissues were immediately frozen in liquid nitrogen and stored at −70 °C.

2.2. cDNA library preparation and Illumina sequencing for transcriptome analysis

Total RNA was isolated from frozen tissues using the cold-acidic phenol method with a modified extraction buffer (10 mM Tris–HCl, pH 8.0, 25 mM EDTA, pH 8.0, 2% CTAB, 2% PVP). The RNA was then precipitated with ethanol, dissolved in DEPC-treated water and stored at −70 °C. To obtain complete gene expression information and to reduce sequencing costs, a pooled RNA sample obtained from roots and leaves from 12 h salt-stressed and unstressed plants was used to construct a reference library for *de novo* transcriptome sequencing. The individual RNA samples were quantified and examined spectrophotometrically for protein contamination (A_{260}/A_{280} ratio) and reagent contamination (A_{260}/A_{230} ratio) prior to library construction.

The cDNA library was then constructed following the Illumina manufacturer's instructions. The poly (A) + RNA was purified from 20 μg of pooled total RNA using Oligo(dT) magnetic beads and fragmented into short fragments. Using these short fragments as templates, random hexamer-primer was used to synthesize first-strand and second-strand cDNA. Short fragments were purified with a QiaQuick PCR Extraction Kit (QIAGEN, Germany) and ligated to sequencing adapters. The products were amplified by PCR to create a cDNA library.

The cDNA library was sequenced using the Illumina HiSeq™ 2000 system. The sequencing-received raw image data were transformed by base calling into raw reads. Transcriptome *de novo* assembly was carried out with a short reads assembling program—the SOAPdenovo (Li et al., 2010). The SOAPdenovo first combines reads with an identity value of 95% and a coverage length of 100 bp to form longer fragments without N, and these are called contigs. Then the reads are mapped back to contigs to detect contigs from the same transcript as well as the distances between these contigs. Next, the SOAPdenovo connects the contigs by using N to represent unknown sequences between each two contigs, and then forms Scaffolds. Paired-end reads are used again for gap filling of scaffolds to obtain sequences with least number of Ns and these cannot be extended on either end. Such sequences are defined as Unigenes. The generated unigene sequences were aligned by BLASTx to protein databases including the NCBI nr protein database, the Swiss-Prot database, the KEGG database and the COG database (E-value $\leq 10^{-5}$). The Blast2GO program (Altschul et al., 1997) was then used to obtain GO annotations of unigenes (http://www.blast2go.org).

2.3. DGE library preparation and sequencing

Total RNA was extracted from the roots of *G. aridum* plants treated for different periods of time with 200 mM NaCl (3, 12, 72 and 144 h) and the untreated control using the CTAB method. DGE libraries were prepared using the Illumina Gene Expression Sample Prep Kit (Illumina Inc., San Diego, CA, USA). The poly (A) + RNA was purified from 6 μg of total RNA using Oligo(dT) beads, and Oligo(dT) was used as a primer to synthesize the first- and second-strand cDNA. The bead-bound cDNA was subsequently digested with the restriction enzyme *Nla*III, which recognized and removed the CATG sites. The fragments (except for the 3′ cDNA fragments connected to the Oligo (dT) beads) were removed by washing, and Illumina adaptor 1 (sense: 5′ACACTCTTT CCCTACACGACGCTCTTCCGATC3′) was ligated to the sticky 5′ end of the digested bead-bound cDNA fragments. The junction of Illumina adaptor 1 and CATG site is the recognition site of *Mme*I, an endonuclease with separate recognition and digestion sites. *Mme*I was therefore used to digest the fragments 17 bp downstream of the CATG site, producing tags with adaptor 1. After removing the 3′ fragments using magnetic bead precipitation, Illumina adaptor 2 (sense: 5′GATCGGAAGAGCGG TTCAGCAGCAATGCCGAG3′) was ligated to the 3′ ends of the tags, producing tags with different adaptors at both ends to form a tag library.