1 Research article

Transcriptome profiling and digital gene expression analysis of genes associated with salinity resistance in peanut

Q3 Q2 Jiongming Sui [1], Pingping Jiang [1], Guilong Qin [1], Shupeng Gai, Dan Zhu, Lixian Qiao, Jingshan Wang *

College of Life Science, Qingdao Agricultural University, Key Laboratory of Qingdao Major Crop Germplasm Resource Innovation and Application, Key Lab of Plant Biotechnology in Universities of Shandong, Qingdao 266109, China

A B S T R A C T

Background: Soil salinity can significantly reduce crop production, but the molecular mechanism of salinity tolerance in peanut is poorly understood. A mutant (S1) with higher salinity resistance than its mutagenic parent HY22 (S3) was obtained. Transcriptome sequencing and digital gene expression (DGE) analysis were performed with leaves of S1 and S3 before and after plants were irrigated with 250 mM NaCl.

Results: A total of 107,725 comprehensive transcripts were assembled into 67,738 unigenes using TIGR Gene Indices clustering tools (TGICL). All unigenes were searched against the euKaryotic Ortholog Groups (KOG), gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, and these unigenes were assigned to 26 functional KOG categories, 56 GO terms, 32 KEGG groups, respectively. In total 112 differentially expressed genes (DEGs) between S1 and S3 after salinity stress were screened, among them, 86 were responsive to salinity stress in S1 and/or S3. These 86 DEGs included genes that encoded the following kinds of proteins that are known to be involved in resistance to salinity stress: late embryogenesis abundant proteins (LEAs), major intrinsic proteins (MIPs) or aquaporins, metallothioneins (MTs), lipid transfer protein (LTP), calcineurin B-like protein-interacting protein kinases (CIPKs), 9-cis-epoxycarotenoid dioxygenase (NCED) and oleosins, etc. Of these 86 DEGs, 18 could not be matched with known proteins.

Conclusion: The results from this study will be useful for further research on the mechanism of salinity resistance and will provide a useful gene resource for the variety breeding of salinity resistance in peanut.

1. Introduction

Over the past few decades, there has been a dramatic increase in the salinization of arable land. As the land available for conventional agriculture becomes increasingly limited, plants grown on marginal soils will be exposed to higher levels of soil salinity. Soil salinity is a major abiotic stress responsible for reduced growth and yield of many crops [1]. Consequently, a better understanding of salt tolerance in crops is needed.

Peanut (Arachis hypogaea L.), which is an important oil-crop and protein production in the tropics and subtropics [2], is likely to face increased drought and salinity stresses in the near future [3]. Hence, genes responsible for resistance to drought and salinity stress in peanut need to be identified and studied. Unfortunately, little progress has been made in the study of salinity tolerance in peanut, in part because of the lack of peanut germplasm with high resistance to salinity stress. In our previous studies, we conducted in vitro mutagenesis (with pingyangmycin as the mutagen) and directed screening with a medium containing NaCl to generate mutants with salt tolerance [4]. One mutant (designated S1) with enhanced salinity tolerance was obtained. This mutant had a much higher germination rate than its mutagenic parent Huayu 22 (designated S3) in a 0.7% NaCl solution, and its self-pollinated offspring grew better than S3 in a saline-alkali field in Dongying City, China. Little is known, however, about the molecular mechanisms resulting in salt tolerance in peanut.

High-throughput RNA-sequencing (RNA-Seq) is a recent and effective technology for the analysis of gene expression, the discovery of novel transcripts, and the identification of differentially expressed genes (DEGs). This powerful technology makes it possible to study non-model organisms [5,6,7].

To investigate the molecular basis for the salinity-tolerance in peanut, we compared the transcriptome and digital gene expression (DGE) profiles in the leaves of S1 and its salinity-sensitive parent, S3, before and after the application of a salinity-stress treatment. We identified the specific transcripts related to salinity-stress resistance in peanut, and we discussed the possible roles of the DEGs.

* Corresponding author.
E-mail addresses: lxqiao73@163.com (L. Qiao), jswang319@126.com (J. Wang).
[1] These authors contribute equally to this study.
Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

2. Materials and methods

2.1. Plant growth and stress treatments

The seeds of S1 (the mutant with enhanced salinity tolerance) and S3 (Huayu 22, the control) were grown in a growth chamber with a dark/light cycle of 8/16 h at 28°C for six weeks. Then, the seedlings of each genotype were irrigated with 250 mM NaCl for salinity stress under culture-room conditions according to our previous report [8]. At 0, 6, 12, 24, and 48 h after the seedlings were subjected to the NaCl solution, the leaves of the S1 and S3 seedlings were removed and placed in liquid nitrogen.

2.2. Library construction and transcriptome sequencing

A total amount of 1.5-μg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) f and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3′ ends of DNA fragments, NEBNext Adaptor with hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 150–200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then, 3-μl USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, The RNA from each combination of seedling type (S1 or S3) and time after salt treatment was pooled and then analyzed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Sequencing of the RNA was carried out by Novogene Bioinformatics Technology Co., Ltd. (Beijing, China) on an Illumina HiSeq2000 sequencer. For no reference genome, after the acquisition of clean reads, the clean reads needs to be spliced to obtain the reference sequence for subsequent analysis. All clean sequence read data were deposited in the NCBI SRA database (accession number SRR3114511), and then they were assembled into comprehensive unigenes using Trinity and TGICL [9].

2.3. Transcriptome functional annotation

The assignment of sequence orientations and functional annotations of the all-unigenes were determined by BLASTx against the following databases: the NCBI non-redundant (NR) protein database, the Swiss-Prot protein database with an E-value cut-off of 10^{-5}, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database with an E-value of 10^{-10}, and the euKaryotic Ortholog Groups of proteins (KOG) database with an E value of 10^{-3}. The all-unigenes were assigned to Gene Ontology (GO) categories with an E-value cut-off of 10^{-6}. In addition, unigenes were aligned with the NCBI nucleotide (NT) databases using BLASTn with an E value of 10^{-5}.

2.4. Digital gene expression (DGE) sequencing and mapping

The RNA samples from S1 and S3 were labeled with the sampling times (0, 6, 12, 24, and 48 h after NaCl treatment) as follows: S1_0, S1_6, S1_12, S1_24, and S1_48 for S1, and S3_0, S3_6, S3_12, S3_24, and S3_48 for S3. Each combination of genotype and sampling time after salt treatment was represented by two replicate RNA samples. DGE sequencing was carried out with a single 50-bp end read for each reaction; all clean sequence read data were deposited in the NCBI SRA database (accession number SRR3204213 and SRR3204348). Then, all reads of each library were separately mapped onto the unigenes using the default parameters in SOAP, and the uniquely mapped reads were extracted for abundance quantification. Finally, unigene expression was normalized using the value of RPKM (reads per kilobase per million reads). Multiple comparisons were carried between the data sets of different samples.

Expression was compared both within each genotype and between the two genotypes. The comparison between S1 and S3 samples resulted in D series data sets, which represented the DEGs between S1 and S3 samples in response to salinity stress treatment; they were denoted as D_0, D_6, D_12, D_24 and D_48 with the sampling time point of 0, 6, 12, 24 and 48 h. Between the genotypes, expression was compared at 0, 6, 12, 24, and 48 h. If the level of expression was significantly different (the adjusted P-value < 0.05) in a comparison, the gene was considered to be differentially expressed. Within S1 and S3, expression was compared between each two sampling time of 0, 6, 12, 24, and 48 h; if the level of expression was significantly up- or down-regulated (the adjusted P-value < 0.05) in a comparison, this gene was proposed to be responsive to salinity stress. Pathways that were statistically significant (FDR ≤ 0.05) were enriched with KEGG.   Q4

2.5. Real-time PCR analysis

To determine whether the expression analyses were correct, we performed real-time PCR analysis on selected DEGs. Reverse transcription were performed using an Invitrogen SuperScript Reagent Kit. For real-time PCR, the SYBR® Premix Ex Taq™ (TAKARA) was used on a Bio-Rad CFX96 real-time PCR detection system (Bio-Rad, Hercules, CA). Gene expression was analyzed for S1 and S3 samples at 0, 6, 12, 24, and 48 h after application of the salinity-stress treatment. All reactions for each gene were performed in three biological replications with a 20-μl reaction volume. The relative expression level of each gene among samples was calculated using the $2^{-\triangle\triangle Ct}$ method with normalization to the internal reference *actin* gene from peanut. The parameters of the thermal cycle were 95°C for 30 s, followed by 40 cycles of 95°C for 10 s and 50–56°C for 25 s.

3. Results

3.1. Transcriptomic sequencing and de novo assembly

The transcriptomic analysis of pooled samples resulted in a total of 62,887,762 clean reads and 7.86 G clean base pairs. The comprehensive reads were assembled into transcripts using paired-end reads, resulting in 107,725 comprehensive transcripts. With the criteria of more than 50-bp overlap and 90% identity, the transcripts were further assembled into 67,738 unigenes using TGICL. The size of unigenes ranged from 201 to 18,360 bp with an average length of 766 bp; the N50 value was 1362 bp (Fig. S1).

3.2. Function annotation and classification

Predicted functions of these unigenes were obtained by searching against several protein databases. The number and percentage of the 67,738 unigenes that were annotated in the NR, NT, SwissProt, and PFAM databases are indicated in Table S1.

All unigenes were searched against the euKaryotic Ortholog Groups (KOG) database to divide ortholog clusters by phylogenetical relations. A total of 10,571 (15.61%) of the 67,738 unigenes were assigned to the 26 function categories (Fig. 1, Table S2). The top five categories were "General function prediction only" (1983, 18.76%), "Posttranslational modification, protein turnover, chaperones" (1369, 12.95%), "Signal