Journal of Biotechnology 189 (2014) 48–57

Contents lists available at ScienceDirect

Journal of Biotechnology

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

Transcriptome characterization and differential expression analysis of cold-responsive genes in young spikes of common wheat

Shujuan Zhang[a,b,c,1], Guoqi Song[a,b,c,1], Jie Gao[a,b,c], Yulian Li[a,b,c], Dong Guo[a,b,c], Qingqi Fan[a,b,c], Xinxia Sui[a,b,c], Xiusheng Chu[a,b,c], Chengyan Huang[a,b,c], Jianjun Liu[a,b,c], Genying Li[a,b,c,*]

[a] Crop Research Institute, Shandong Academy of Agricultural Sciences, China
[b] Key Laboratory of Wheat Biology & Genetic Improvement On North Yellow & Huai River Valley, Ministry of Agriculture, China
[c] National Engineering Laboratory For Wheat & Maize, Jinan 250100, Shandong, China

ARTICLE INFO

Article history:
Received 26 May 2014
Received in revised form 21 August 2014
Accepted 25 August 2014
Available online 21 September 2014

Keywords:
Triticum aestivum L.
Reproductive stage
Transcriptome
Digital gene expression profile
Frost responsive gene

ABSTRACT

With the frequent occurrence of climatic anomalies, spring frost has become a significant limiting factor on wheat production, especially during the reproductive growth stage. A high-throughput sequencing technology was applied and a total of 54 million clean reads that corresponded to 7.44 Gb of total nucleotides were generated. These reads were then *de novo* assembled into 120,715 unigenes with an average length of 627 bp. Functional annotations were then obtained by aligning all unigenes with public protein databases. In total, 9657 potential EST-SSRs were identified, and 6310 primer pairs for 1329 SSRs were obtained. Meanwhile, a comparison of four tag-based digital gene expression libraries, which was built from the control and cold-treated young spikes were performed. Overall, 526 up-regulated and 489 down-regulated genes were identified, and GO and KEGG pathway analyses of those genes were further conducted. Based on these results, a series of candidate genes involved in cold response pathways were identified, and 12 of them were confirmed by qRT-PCR. The combination of RNA-Seq and digital gene expression analysis in this study provides a powerful approach for investigating the transcriptional changes and obtained a large number of unigenes annotated to public databases.

1. Introduction

Common wheat (*Triticum aestivum* L.) is one of the most important food crops in the world. With the frequent occurrence of climatic anomalies, spring frost, has become a significant limiting factor on wheat production. Spring frost injury frequently occurs when low temperatures coincide with sensitive plant growth stages (Slafer and Rawson, 1995; Mahfoozi et al., 2001), from pistil and stamen differentiation stage to anther differentiation stage (Fowler and Limin, 2004; Single, 1974; Single and Marcellos, 1974). To date, very few studies on the underneath mechanism involved in frost response at the reproductive stages of common wheat were reported (Steponkus, 1984; Campbell and Close, 1997; Rinalducci et al., 2011), while most of them were focused on frost tolerance at vegetative developmental stages, and *VRN*, *Fr* and *CBF* transcription factors were the major genetic loci related to frost resistance at this period (Galiba et al., 1995; Galiba et al., 2009; Toth et al., 2003). To understand the molecular mechanism and signaling processes response to frost during wheat reproductive stage, transcriptome analysis of young spikes of common wheat Jimai22 at pistil and stamen differentiation stage, were carried out in the present study. Jimai22 was an excellent commercial cultivar in China, with the multiple resistances to abiotic stresses and the highest yield in winter wheat.

Common wheat has a large genome (17 Gb) with a high proportion of repetitive sequences (>80%) and complex polyploid genome makes genetic and functional analyses extremely challenging. Recently, several wheat genome sequencing efforts (Brenchley et al., 2012; Ling et al., 2013; Jia et al., 2013) produced incomplete results lacking in physical maps, location of genes and assembled scaffolds that make map-based cloning very difficult. The database of putative full-length cDNAs for common wheat, TriFLDB, has released approximately 16,000 full-length cDNAs (http://trifldb.psc.riken.jp/index.pl) (Mochida et al., 2009). Although this dataset is a useful reference for transcript mapping,

* Corresponding author at: Crop Research Institute, Shandong Academy of Agricultural Sciences, 202 Gongyebei Road, Jinan 250100, Shandong, China. Tel.: +86 531 83178122.
E-mail address: lgy111@126.com (G. Li).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jbiotec.2014.08.032
0168-1656/

it is incomplete and not enough for sequence annotation and characterization (Choulet et al., 2010).

Transcriptome sequencing has been proven to be an efficient method for gene discovery. With advances in sequencing technology, RNA-Seq, which is cost-efficient and provides sufficient information, has not only been used in model plants, but also in non-model plants whose genomes have never been sequenced (Zhang et al., 2010; Kakumanu et al., 2012; Wei et al., 2011; Wang et al., 2010; Bräutigam and Gowik, 2010). Cantu et al. (2011) assembled 1.4 million 454 reads into 30,497 contigs to study grain protein content related genes in common wheat. Li et al. (2011) performed an mRNA tag analysis of common wheat seedlings to gain an overview of responses to H_2O_2 treatments in common wheat. Pont et al. (2011) assembled 934,928 reads generated from 454 sequences to understand the polyploidization events of common wheat. Pellny et al. (2012) utilized rice sequences and 1.5 million public ESTs as references to study the transcriptome of the developing starchy endosperm of common wheat. Oono et al. (2013) carried out *de novo* assembly of 73.8 million high-quality reads by RNA-Seq to generate transcriptome profiles of the common wheat cultivar 'Chinese Spring' responding to 10 days of Pi starvation. Duan et al. (2012) optimized *de novo* assembly of the hexaploid wheat transcriptome using 16.2 Gb short reads, documented every detail of the assembly process and discussed the effects of each assembling step to gain insight into the transcriptome assembly.

Another method of great value for expression analysis is DGE (digital gene expression), which uses 17–21 bp short fragments from the whole transcriptome as gene-specific tags and calculates the expression level of a gene from the frequency of its tag. The clean tags are collected and mapped to the genome. The number of tags is proportional to the abundance of cognate transcripts in the specific tissue. This method provides much more qualitative and quantitative description of gene expression than previous microarray-based assays. In contrast with high-throughput RNA-seq, both methods provide similar assessments of relative transcript abundance, but DGE better detects the expression differences for poorly expressed genes and does not exhibit transcript length bias (Lewis et al., 2011).

The combination of RNA-Seq and DGE allows us to easily perform transcriptome analysis without the need for an already-assembled reference genome. In the present study, we sampled the pooled transcriptomes of cold-treated and control (untreated) common wheat young spikes using Illumina paired-end sequencing technology to generate a large-scale EST database and develop EST-SSRs. The assembled and annotated gene expression profiles provide a valuable resource for identification of differentially expressed genes during cold response, and enable us to understand the underlying molecular mechanism of frost tolerance at the reproductive stage of common wheat. The produced EST datasets together with the new transcript data will also serve as good resource for novel gene discovery and marker-assisted selection in wheat breeding.

2. Materials and methods

2.1. Plant materials and RNA extraction

Seeds of common wheat (*T. aestivum* L.) cultivar Jimai22 were germinated and grown in a 4 °C chamber for vernalization. After 40 days, the seedlings were transplanted to greenhouse and grown at 22 °C/16 °C day/night. Forty days after transplanting, most of the young spikes developed into pistil and stamen differentiation stage, then the seedlings were divided into two groups. One was used as the control sample at 22 °C/16 °C in green house, while the other was moved to a chamber and cultured for frost treatment at 0 °C for 48 h. Young spikes were pooled simultaneously after 48 h cold treatment and stored in liquid nitrogen for RNA extraction. Total RNA was isolated using the TRIzol reagent according to the manufacturer's instructions (Invitrogen, Carlsbad, California, USA). RNA quality was tested using a 2100 Bioanalyzer RNA Nanochip (Agilent, Santa Clara, CA, USA). RNA concentration was quantified using a NanoDrop ND-1000 Spectrophotometer (Nano-Drop, Wilmington, DE, USA).

2.2. Preparation of cDNA library and Illumina sequencing for transcriptome analysis

Solexa sequencing was performed as a commercial service in Beijing Genomics Institute (BGI; Shenzhen, China). Transcriptome library was prepared and sequenced on Illumina HiSeq™ 2000 using paired-end technology. The sequencing data were deposited in the NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/Traces/sra) with accession number SRX375489.

2.3. Data filtering and de novo assembly

The raw reads were cleaned by removing the adapter sequences or low-quality sequences. *De novo* assembly of the clean reads was performed using Trinity (http://trinityrnaseq.sourceforge.net/) with default K-mers = 25. Trinity combines three independent software modules: Inchworm, Chrysalis, and Butterfly, applied sequentially to process large volumes of reads. Inchworm assembles the data into the unique sequences of transcripts. Chrysalis clusters the Inchworm contigs into clusters and constructs complete de brujin graphs for each other. Butterfly then processes each graph independently to extract full-length splicing isoforms and to tease apart transcripts derived from paralogous genes. We used Trinity to connect the contigs and obtained sequences that cannot be extended on either end known as unigenes. Finally, the TIGR Gene Indices Clustering tools (TGICL) (Pertea et al., 2003) is used to get rid of redundant unigene and further assemble all the unigenes to form a single set of non-redundant unigenes.

2.4. Gene annotation

Unigenes were aligned with the Nr, Swiss-Prot, the KEGG and COG database using BLASTx with an *E*-value of less than 10^{-5}. The best aligning results were used to determine the sequence direction of unigenes. If the results of different databases conflicted with each other, a priority order of Nr, Swiss-Prot, KEGG and COG was followed. For unigenes that could not align to any of the above databases were analyzed using ESTScan to determine its sequence direction.

After Nr annotation, the Blast2GO program (Conesa et al., 2005) was used to get GO annotation. The unigene sequences were aligned to the COG database to predict and classify possible functions. To investigate the metabolic pathway annotation of unigenes, we aligned the unigene sequences to the KEGG database.

2.5. EST-derived SSR detection and primer design

Potential SSR markers were detected among all unigenes using the MicroSAtellite software (MISA; http://pgrc.ipk-gatersleben.de/misa/). The parameters were adjusted for identification of perfect mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of 12, 6, 5, 5, 4 and 4 repeats, respectively. Primer pairs were designed using BatchPrimer3 (Kortt et al., 1991). The major parameters for primers design were set as follows: primer length of 18–23 bases, PCR product size of 100–200 bp, GC content of 40–55%, and annealing temperatures of 50–60 °C.