Genomics 99 (2012) 299–307

Contents lists available at SciVerse ScienceDirect

Genomics

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

Transcriptome analysis of rosette and folding leaves in Chinese cabbage using high-throughput RNA sequencing

Fengde Wang [a], Libin Li [a], Huayin Li [a], Lifeng Liu [a], Yihui Zhang [a], Jianwei Gao [a,*], Xiaowu Wang [b,**]

[a] Institute of Vegetables, Shandong Academy of Agricultural Sciences/Shandong Key Laboratory of Greenhouse Vegetable Biology/Shandong Branch of National Vegetable Improvement Center, Jinan 250100, China
[b] Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China

ARTICLE INFO

Article history:
Received 7 December 2011
Accepted 15 February 2012
Available online 23 February 2012

Keywords:
Chinese cabbage
Rosette leaves
Folding leaves
Transcriptome
High-throughput RNA sequencing

ABSTRACT

In this study, we report the first use of RNA-sequencing to gain insight into the wide range of transcriptional events that are associated with leafy head development in Chinese cabbage. We generated 53.5 million sequence reads (90 bp in length) from the rosette and heading leaves. The sequence reads were aligned to the recently sequenced Chiifu genome and were analyzed to measure the gene expression levels, to detect alternative splicing events and novel transcripts, to determine the expression of single nucleotide polymorphisms, and to refine the annotated gene structures. The analysis of the global gene expression pattern suggests two important concepts, which govern leafy head formation. Firstly, some stimuli, such as carbohydrate levels, light intensity and endogenous hormones might play a critical role in regulating the leafy head formation. Secondly, the regulation of transcription factors, protein kinases and calcium may also be involved in this developmental process.

1. Introduction

Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*) is a widely cultivated and economically important vegetable crop in Asia. Chinese cabbage originated in China, and it has now become increasingly popular in other countries. Chinese cabbage has a tight leafy head, which is the storage organ for the cabbage and is the part that is commonly eaten. The leafy head is composed of a large number of heading leaves, which usually initiate after the rosette stage. The rosette leaves differentiate at the rosette stage and serve as the photosynthetic organs during the vegetative growth, whereas the heading leaves surrounding the shoot apexes are compact enough to form a tight head or heart. The production of Chinese cabbage is usually hampered by the poor heading of the leaves [1]. The understanding of the leafy head development at a molecular level will greatly facilitate the genetic improvement of the Chinese cabbage yield, its nutritional value and the quality of its appearance. The initiation and developmental process of the leafy head may be influenced by many factors, including the uneven distribution of auxin levels in the leaves, temperature, weak light, short days and the carbohydrate nutrition level [2]. However, because the results of many physiological experiments have not been explained or verified completely, how the plant senses environmental signals and subsequently turns on downstream gene regulation networks is largely unknown [1].

During cultivation, the vegetative growth of Chinese cabbage is divided into five stages, including the germination, seedling, rosette, heading and dormancy stages [3]. At the end of rosette stage, the young leaves, named folding leaves (FL), begin to fold inward, and the heading leaf differentiation is initiated. It has been reported that spraying auxin onto the dorsal side of the FL induces them to bend inward, whereas the juvenile leaves (JL) and rosette leaves (RL) do not react in the same way [2]. In addition, transcripts of the *BcpLH* gene have been detected in the FL but not in the JL or in the RL, indicating that this gene may play an important role in regulating the leafy head development in Chinese cabbage [1]. Compared with the other leaf types, the FL are unique because they accept external environmental stimuli and transmit these signals into morphological responses. The initiation of the leafy head is characterized by the bending inward of the leaves and is marked by the differentiation of the FL. Therefore, the development of the FL is an excellent model system to investigate the regulation, differentiation and development of the leafy head.

Recently, the mRNA differential display technique [4], the analysis of expressed sequence tags [5] and the differential hybridization method [1] have been used to examine the expression patterns of the RL and FL genes and have provided the first illustration of transcriptome dynamics during leafy head development. However, these data are far from being complete due to the limitations of these approaches.

RNA-sequencing (RNA-Seq) is a novel, high-throughput, deep-sequencing technology that is widely used for genomics research

*Corresponding author. Fax: +86 531 83179060.
**Corresponding author. Fax: +86 10 62146163.
E-mail addresses: jianweigao3@yahoo.com (J. Gao), wangxw@mail.caas.net.cn (X. Wang).

0888-7543/$ – see front matter © 2012 Published by Elsevier Inc.
doi:10.1016/j.ygeno.2012.02.005

and provides new strategies to analyze the functional complexity of transcriptomes. In particular, the Solexa/Illumina sequencing technology has many advantages as a revolutionary tool for transcriptome analysis, such as high coverage at a relatively low cost [6,7], and it has been used to investigate transcriptomes in plants, such as *Arabidopsis*, rice and berry [8–10].

To understand further the complexity of the transcriptome during leafy head development at the level of the whole genome, we performed the first global analysis of the transcriptomes from RL and FL in Chinese cabbage using the Solexa/Illumina RNA-Seq platform. This comprehensive analysis of the transcriptome dynamics serves as a blueprint for the gene expression profile of leafy head development. The data may substantially improve the global view of the Chinese cabbage transcriptome during leafy head development and pave the way for its further analysis and application to breeding practices. In addition, our analysis revealed the expression of numerous novel transcriptional units (TU), single nucleotide polymorphisms (SNPs) and alternative splicing (AS) events. We also refined the gene structures and found that compared to the existing gene annotations, a large number of the genes could be extended at the 5′ end, the 3′ end or at both ends.

2. Results

2.1. RNA-Seq and mapping of the sequence reads

To obtain a global view of transcriptome relevant to leafy head development in Chinese cabbage inbred line Fushanbaotou, the high-throughput RNA-Seq analysis on poly(A)-enriched RNAs from RL and FL libraries, respectively, was performed using the Solexa/Illumina platform. After filtering out the low-complexity reads, the low-quality reads and the repetitive reads, 27,088,098 usable reads for the RL and 26,386,316 usable reads for the FL were obtained (Table 1). To identify the gene expression patterns in the RL and the FL of Chinese cabbage, we mapped the reads against the sequenced Chinese cabbage genome (http://brassicadb.org/brad/) using the SOAP2 software [11], which was set to allow two base mismatches. Of the total reads, 70.6% of the reads matched to unique (67.7%) or multiple (2.9%) genome locations (Table 1). In addition, the annotated exons from the Chinese cabbage genome were used as reference genes to assign each read to a specific gene. As shown in Table 1, approximately 60% of the reads were mapped to unique genes, and 2.83% of the reads were mapped to multiple reference genes.

2.2. Global gene expression pattern analysis

The analysis of differentially expressed genes (DEGs) between the FL and RL libraries should aid our understanding of the molecular events involved in leafy head development. To confirm that the differences in the gene expression patterns observed among the developmental stages are statistically significant, we compared the RPKM-derived read count using Audic's method [12] with some modifications. A threshold value of FDR $\leq$ 0.001 and an absolute value of $\log_2$Ratio $\geq$ 1 were used to judge the significance of the differences in gene expression. Based on these criteria, 2255 genes were differentially expressed between the RL and FL including 953 genes that were up-regulated and 1302 genes that were down-regulated in the FL tissues (Supplementary Table S1). To understand their biological function, these DEGs were subjected to an on-line (http://brassicadb.org/brad/searchAll.php) BlastX search against the *Arabidopsis* genome.

The Gene Ontology (GO) database includes information on biological processes, molecular functions and cellular components and is an international standardized gene functional classification system, which offers a dynamically updated controlled vocabulary and a strictly defined concept that comprehensively describes the properties of genes and their products in any organism. To facilitate the global analysis of gene expression, a GO analysis was performed by mapping each differentially expressed gene into the records of the GO database (http://www.geneontology.org/). Under the biological process category, 20 GO categories were significantly enriched (corrected p-value $\leq$ 0.05) in the DEGs (Fig. 1A; Supplementary Table S2). Of these, a large number of gene responses to various stimuli, including carbohydrate, light intensity, endogenous and other biotic and abiotic stimuli, were prominently represented, suggesting that these stimuli are needed for leafy head development. Under the category of molecular function (Fig. 1B; Supplementary Table S2), the main functional groups of the DEGs (151 genes) were genes with nucleic acid binding transcription factor (TF) activity, including the MYB protein family, the zinc finger protein family, the AUX/IAA protein family, the WRKY protein family, and the bHLH protein family, which are involved in various plant developmental processes and stimuli responses. For the cellular component category (Fig. 1C; Supplementary Table S2), most of the DEGs were associated with the external encapsulating structure, the non-membrane-bounded organelle, the intracellular non-membrane-bound organelle and the cell periphery, whereas only a few DEGs were assigned to the chromosomal components, the chromatin, and the chromosome.

Because leafy head formation is affected by the auxin distribution in the leaves, a lower temperature, a greater day–night temperature difference, weak light, a short day length and sufficient carbohydrate nutrition [2], four subgroups were characterized from the group of DEGs that were related to plant development and the stimuli response (Table 2). The first subgroup of the DEGs was shown to be involved in transcriptional regulation. A number of the transcription factors were identified, such as *MYB*, *LBD*, *HD-ZIPIII*, *TCP*, *MADS*, *zf-HD*, Zinc finger protein, *NAM*, *WRKY*, *bZIP*, *bHLH*, *ANT*, *GRF*, *HB*, *AP2-EREBP* and *ERF*. The second subgroup was shown to be composed of 38 genes with homologies to genes that encode protein kinases, such as *CDKs*, *MAPK*, *MAPKK*, *MAPKKK*, *RLK* and *CDPK*. It is noteworthy that the expression of 21 members of the *CDKs* was down-regulated, whereas the expression of 8 members of the *RLKs* was up-regulated. The third subgroup contained 34 genes, including 15 calcium-binding proteins, 10 calcium-binding EF hand family proteins, 6 calmodulin-binding proteins and 3 calcium:cation antiporters. The final subgroup that was analyzed was related to auxin synthesis (*YUCCA*), transport (*PIN*), signaling (*AUX/IAA* and *ARF*), inactivation (*GH3*) and response (*SAUR*).

Table 1
Summary of read number.

Reads category	Map to genome		Map to annotation gene	
	RL	FL	RL	FL
Total reads	27,088,098	26,386,316	27,088,098	26,386,316
Total base pairs	2,437,928,820	2,374,768,440	2,437,928,820	2,374,768,440
Total mapped reads	19,158,357	18,610,411	16,869,340	16,735,761
Perfect match	13,569,026	13,163,404	11,401,649	11,347,040
≤2 bp mismatch	5,589,331	5,447,007	5,467,691	5,388,721
Unique match	18,398,107	17,826,502	16,143,576	15,953,181
Multi-position match	760,250	783,909	725,764	782,580
Total unmapped reads	7,929,741	7,775,905	10,218,758	9,650,555

2.3. Identification of splice variants and the discovery of novel transcripts

Alternative splicing is a mechanism that brings remarkable diversity to proteins and allows a gene to generate different mRNA transcripts that are translated into distinguishable proteins [13,14]. To assess the genome-wide AS events during leafy head development,