Plant Physiology and Biochemistry 105 (2016) 224–232

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

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

Research article

Transcriptome *de novo* assembly and analysis of differentially expressed genes related to cytoplasmic male sterility in cabbage

Shuai Wang, Chao Wang*, Xiao-Xuan Zhang, Xin Chen, Jiao-Jiao Liu, Xue-Fang Jia, Si-Qi Jia

College of Horticulture, Northeast Agricultural University, Harbin, 150030, China

ARTICLE INFO

Article history:
Received 28 December 2015
Received in revised form
14 April 2016
Accepted 15 April 2016
Available online 16 April 2016

Keywords:
Cabbage (*Brassica oleracea* L. var. *capitata*)
Cytoplasmic male sterility
Transcriptome analysis
Differentially expressed genes
Anther development

ABSTRACT

Cytoplasmic male sterility (CMS) is a maternally inherited trait producing abnormal pollen during anther development. To identify the critical genes and pathways that are involved in the sterility and to better understand the underlying mechanisms, cabbage anthers at different developmental stages were cytologically examined and the transcriptomes were analyzed in CMS line and its maintainer line using the next-generation sequencing (NGS) technology. Microscopy showed that anther development in the CMS line was abnormal in the tetrad stage and failed to produce fertile pollen. We obtained 55,663,594 and 54,801,384 raw transcriptome reads from the sterile and maintainer lines, respectively, and assembled these reads into 68,851 unigenes with an average size of 1028 bp. By using the fragments assigned per kilobase of target per million mapped reads (FPKM) method, 5592 differentially expressed genes were identified, consisting of 3403 up- and 2089 down-regulated genes. Furthermore, there were 1011 and 45 genes specifically expressed in the maintainer or sterile line, respectively. Gene Ontology (GO) functional annotation and enrichment analysis of metabolic pathways were performed to map and analyze the candidate genes that may be involved in male sterility. Expression of eighteen genes was examined using qRT-PCR and their expression patterns were found to be same as the sequencing data. A clear cytological difference exists between the sterile and maintainer lines. The differentially expressed genes are associated with carbohydrate and energy metabolisms, or encode transcription factors, heat shock proteins and other stress proteins. Identification of these candidate genes provides a comprehensive understanding of the mechanism underlying CMS in cabbage.

1. Introduction

Cabbage (*Brassica oleracea* L. var. *capitata*), one of the most important species in the *Brassicaceae* family, is widely grown and economically important in the world. The chromosomal relationships among the Brassica species are triangle as described by UN (1935). There are three diploid progenitors, *Brassica rapa* (AA, 2n = 20; turnip rape, turnip and Chinese cabbage), *Brassica nigra* (BB, 2n = 16; black mustard) and *B. oleracea* (CC, 2n = 18; cabbage, cauliflower, broccoli, kale, kohlrabi and brussels sprout). In addition, there are three different amphidiploids, *Brassica napus* (AACC, 2n = 38; oilseed rape and swede), *Brassica juncea* (AABB, 2n = 36; Indian or brown mustard) and *Brassica carinata* (BBCC, 2n = 34; Abyssinian or Ethiopian mustard). These species are derived from spontaneous chromosome doubling via meiotic nondisjunction after hybridizations of three diploid progenitors (Chiang et al., 1993). Cabbage is an important source of human nutrition and has significant levels of vitamins (A, C, B6, and K), folic acid and minerals such as iron or calcium. A large number of cabbage cultivars are F_1 hybrids produced by crossing male-sterile or self-incompatible lines with pollination lines for heterosis (Fang et al., 2003). However, self-incompatible lines have some limitations in the production of hybrid seed, such as poor seed purity, high time-consuming and labor costs for parental seed propagation. It also often leads to germplasm degeneration after long-term selfing. On the other hand, production of hybrid seeds using male-sterile lines is more reliable and economical for efficient utilization of heterosis in cabbage (Prakash and Verma, 2004).

Male sterility is a phenomenon whereby the plant is unable to generate functional pollen or anther, microspore or male gamete. Genetically, male sterility is generally divided into genetic male sterility (GMS) and cytoplasmic male sterility (CMS) (Yi et al., 2006), which is maternally inherited. CMS has been identified

* Corresponding author. College of Horticulture, Northeast Agricultural University, No.59 Mucai Street, Xiangfang District, Harbin, 150030, China.
E-mail address: wangchao504@126.com (C. Wang).

http://dx.doi.org/10.1016/j.plaphy.2016.04.027
0981-9428/

and characterized in about 153 angiosperm taxa. The pistil of CMS plant is viable and can be fertilized to produce seed, however, the male floral organ is abnormal, and unable to produce viable pollen. As such, CMS is widely used in commercial hybrid seed production in the species such as maize (Liu et al., 2002), rice (Virmani et al., 1982) and radish (Lee et al., 2008).

CMS is generally believed to originate from the interactions between mitochondrial and nuclear genes. A large number of analyses for association of mitochondrial DNA fragments with CMS show that the abnormal chimeric genes are formed by frequent intramolecular or intermolecular genomic recombination that results in new gene expression patterns and even new genes. Due to the common structure, most of the chimeric genes contain a subunit of the ATP synthase genes or cytochrome oxidase genes and open reading frame (*orf*), which could disrupt mitochondrial function at the critical moment of anther development, leading to CMS. In addition, dramatic changes in gene expression profiles during pollen development in CMS lines have been observed. Identification of these genes contributes to the elucidation of the molecular mechanisms of male sterility, as well as the metabolic pathways involved. RNA-seq, also known as whole transcriptome shotgun sequencing (WTSS), employs the NGS technologies to sequence cDNA. It facilitates the discovery of differentially expressed genes (DEGs) induced by developmental and environmental conditions and helps delineate the underlying mechanisms of interaction, as well as associated signaling pathways. NGS technology has therefore been applied to study the molecular mechanisms of CMS in many species, such as oilseed rape (Yan et al., 2013), chili pepper (Liu et al., 2013), and kenaf (Chen et al., 2014).

Ogura cytoplasmic male sterility (Ogu CMS), originally identified in an unknown male sterile variety of Japanese radish, is the earliest and one of the most extensively analyzed CMS in cruciferous vegetables. It has been reported that a chimeric gene containing *orf138* and *orfB* in mitochondria is involved in Ogu CMS. *B. oleracea* possessing Ogu CMS was originally produced in 1974 (Bannerot et al., 1974). Although it is stable and easy transfer, the Ogura CMS plants are sensitive to low temperature with poor flower development, which limits its utilization in *Brassica*. These drawbacks were addressed *via* protoplast fusion (Pelletier et al., 1983). The Ogura male-sterile cytoplasm was transferred into cauliflower in 1992 without any of its original defects (Walters et al., 1992), and was later transferred into cabbage and other subspecies (Sigareva and Earle, 1997). Fang et al. developed Ogu CMS cabbage line with favorable traits such as appropriate flowering time, higher seed setting and combining ability (Fang et al., 2001). Our group bred an improved Ogu CMS line and it is designated PM. Subsequently, PM was used as the female parent to cross with high-quality inbred lines including male parents in a multiple-generation backcrossing program. As a result, stable sterile lines and corresponding maintainer line PF were obtained. The sterile line and maintainer line are similar in both flowering and economic traits. Recurrent backcrossing and selection for 10 generations resulted in similar nuclear genetic backgrounds between PM and PF. This pair of lines are highly valuable for studying the interactions between plant mitochondrial and nuclear genes.

Most reports of cabbage sequence analysis using RNA-Seq primarily focus on the development of molecular markers (Izzah et al., 2014). To understand the mechanisms of CMS and identify genes that are differentially expressed during the pollen development in PM and its maintainer PF, we conducted cytological observations and transcriptome analyses. Comparison of the transcriptome data from the CMS and maintainer line yielded large number of differentially expressed unigenes. Some of them were further verified by real-time quantitative PCR (qRT-PCR). The results would facilitate the elucidation of the cytological and molecular mechanisms of CMS in cabbage.

2. Materials and methods

2.1. Plant materials

Cabbage CMS line PM and its maintainer line PF were planted at the experimental station of Northeast Agricultural University (Harbin, Heilongjiang, China) under suitable conditions with 16 h of light (sunlight + halogen lamps), at 28 ± 3 °C, followed by 8 h of darkness at 20 ± 4 °C during the spring in 2013. Under greenhouse conditions, PF line produced normal floral organs with four long and two weak and short stamens in each flower (Fig. 1A). In PM plants, the flowers were different with significant degradation of stamens (Fig. 1B). The long stamens in PF were healthy, and clustered together; the pistils were shorter than the stamens. The two short stamens were located contralaterally. The surfaces of all stamens were covered with yellow pollen grains (Fig. 1C). The stamens in the PM were thin with short filaments. Mature stamens shriveled and dried into triangles, leaving the pistils slightly over the stamens. In addition, the PM stamens were pale-yellow and did not release pollens (Fig. 1D).

We collected floral buds that contain pollens at various developmental stages for microscopic study (Fig. 1E). A total of 2 g floral buds (<3 mm in size) were pooled and flash-frozen in liquid nitrogen and stored at −80 °C for RNA-Seq analysis.

2.2. Microscopy

The floral buds were fixed at room temperature in an ethanol and acetic acid (3:1) mixture for 24 h, and dehydrated through an ethanol gradient. The dehydrated buds were embedded in paraffin, sectioned into 8-μm thick slices, mounted on slides, stained with hematoxylin and Fast Green, and viewed and pictured under a light microscope (Nikon-YS100) (Lee et al., 2008).

2.3. mRNA sequencing, de novo assembly, and annotation

Total RNA was extracted using the TRIzol reagent (Invitrogen, USA), digested with DNase I (TransGen, Beijing, China) to remove residual DNA and purified using an EasyPure RNA Purification Kit (TransGen, Beijing, China) following the manufacturer's instructions. RNA purity was verified using a Bioanalyzer 2100 (Agilent) with 2.2 ≥ OD260/280 ≥ 1.8, 28 S/18 S ≥ 1.7 and RIN ≥ 9.0. Then, RNA from three samples from each line was mixed in equal quantity (Liu et al., 2013). 6 μg mixed RNA was used to create normalized cDNAs. The cDNAs were amplified according to the Illumina RNA-Seq protocols and sequenced using the Illumina HiSeq 2000 system at BGI, Shenzhen (Yan et al., 2013). The length of reads was 90 bp and paired ends were obtained. Then, we removed the 3'- adaptor fragments, low-quality sequences (<20% of the bases with a quality score Q ≤ 10), and impurities (N>5%) using an in-house PERL (http://www.perl.org) scripts. The clean reads were *de novo* assembled with Trinity which combines three independent software modules such as Inchworm, Chrysalis, and Butterfly. Trinity partitions the sequence data into many individual *de Bruijn* graphs, each representing the transcriptional complexity at a given gene or locus, and then processes each graph independently to extract full-length splicing isoforms and to tease apart transcripts derived from paralogous genes to avoid impact on differential splicing. The process was as follows: First, reads with certain lengths of overlap were combined to generate contigs, and mapped back to identify them from the same transcripts. Using paired-end reads, we detected contigs from the same transcript as well as the distances between these contigs. The contigs were then connected