Scientia Horticulturae 235 (2018) 116–123

Contents lists available at ScienceDirect

Scientia Horticulturae

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

Transcriptome analysis of carotenoid biosynthesis in the *Brassica campestris* L. subsp. *chinensis* var. *rosularis Tsen*

Hongmei Tian, Ling Fang*, Qi'an Zhang, Mingxia Wang, Yan Wang, Li Jia

Institute of Horticulture, Anhui Academy of Agricultural Sciences He'fei, Anhui, 230031, PR China

ARTICLE INFO

Keywords:
Transcriptome
RNA-sequencing
Brassica campestris L. subsp. *chinensis* var. *rosularis Tsen*
Carotenoid
Differentially expressed genes
Expression analysis

ABSTRACT

Brassica campestris L. subsp. *chinensis* var. *rosularis Tsen* is one of important vegetables grown in Yangtze river basin in winter. "Heixinwu" and "Huangxinwu" were two major varieties, and the carotenoid content of "Huangxinwu" was much richer than "Heixinwu". In this study, we used transcriptome sequencing to identify the repertoire of genes expressed between "Heixinwu" and "Huangxinwu", aiming to delineate the molecular mechanisms of carotenoid biosynthesis. A total of 19.02 gigabase pairs (Gbp) of data with 18,920,370 high-quality reads were obtained, and 55,076 unigenes with average length of 775.96 base pairs were identified by *de novo* assembly. Of these, 39,454 unigenes (71.64%) were further annotated by comparison to public protein databases. A total of 13,018 (33.00%) unigenes were mapped into 122 pathways by searching against the Kyoto Encyclopedia of Genes and Genomes Pathway database (KEGG). Differentially expressed genes (DEGs) analysis identified 1184 DEGs from the comparison of "Heixinwu" and "Huangxinwu", including 610 up-regulated and 574 down-regulated genes. Additionally, 14 carotenoid synthase genes were identified from the transcriptome. The expression patterns of the 14 genes related to carotenoid biosynthesis were analyzed by qRT-PCR to explore their putative functions. This transcriptome dataset will aid in understanding and carrying out future studies on the molecular basis of carotenoid formation and contribute to future artificial production and applications.

1. Introduction

Brassica campestris L. subsp. *chinensis* var. *rosularis Tsen* (20 = 2x = 20) belongs to cruciferae and is native to China. It is mutated from non-heading Chinese cabbage. Owing to the strong resistance of cold and unique flavor, it is widely cultivated all over the country. "Heixinwu" and "Huangxinwu" are the mainly varieties, and "Huangxinwu" rich in carotenoids, which are good for human health. β-carotene and α-carotene serve as precursors for vitamin A, and lycopene serve as precursors for antioxidants (Fiedor and Burda, 2014; Fraser and Bramley, 2004). Carotenoids are a group of isoprenoid molecules generally regarded to as pigments because of their characteristic colour in the yellow to red range. These colours reflect the carotene composition and content. Wisutiamonkul reported that the dark yellow pulp in ripe 'Chanee' durian correlates with higher levels of β-carotene and α-carotene, in comparison to light yellow pulp in 'Monthone' duiran (Wisutiamonkul et al., 2015). They also analyzed the carotenoid accumulation and gene expression during durian fruit growth and ripening (Wisutiamonkul et al., 2017a).

In most photosynthetic organisms, carotenoids play a vital role in protecting the photosynthetic machinery from photo-oxidation sensitized by excited molecules of chlorophyll (Cunningham and Gantt, 2007), and accumulate as secondary metabolites in chromoplasts of flower, fruits, seeds or roots (Devitt et al., 2010). At present, the carotenoid biosynthetic pathway has been well studied. Carotenoid biosynthesis has been proposed via pathways through isopentenyl pyrophosphate (IPP) synthesis by two independent pathways, 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway in plastids and mevalonate (MVA) pathway in cytosol (Chaudhary et al., 2010; DellaPenna and Pogson, 2006; Devitt et al., 2010). The carotenoid biosynthesis include a series of reactions, such as condensation, dehydrogenation and cyclization, hydroxylation, epoxidation, and so on. The first carotenoid compound is formed by the condensation of two molecules of geranylgeranyl diphosphate (C_{20}) into phytoene (C_{40}) catalyzed by phytoene synthase (*PSY*). Lycopene is transformed from phytoene via ζ-carotene by phytoene desaturase (*PDS*) and τ-carotene desaturase (*ZDS*). The lycopene cyclization catalyzed by lycopene β- cyclase (*LCYB*) or lycopene ε-cyclase (*LCYE*) is an important branching point in carotenoid pathway, which result in the production of β-carotene and α-carotene, respectively. β-carotene and α-carotene are catalyzed to lutein and zeaxanthin by β-hydroxylase (*HYB*), respectively. Finally, zeaxanthin catalyzed by zeaxanthin epoxidase (*ZEP*) to form

* Corresponding author.
E-mail address: 13955100418@139.com (L. Fang).

https://doi.org/10.1016/j.scienta.2018.02.040
Received 27 July 2017; Received in revised form 12 February 2018; Accepted 15 February 2018

violaxanthin (Pogson et al., 1996).

Carotenoid biosynthesis and its regulation have been studied in many plants species, such as arabidopsis (Pogson et al., 1996), tomato (Fraser et al., 1994; Giuliano et al., 1993), citrus (Kato et al., 2004), turnip (Li et al., 2015), and so on. Genes involved in carotenoid biosynthesis were induced by various environmental stimuli. PSY was induced by light in tomato seedlings (Giuliano et al., 1993). High light and salt stress elicited the gene expression of PSY and carotenoid hydroxylase, resulting in a rapid accumulation of astaxanthin (Steinbrenner and Linden, 2001). “Huangxinwu” accumulated carotenoid sharply under the temperature of 5 °C–−10 °C, while the leaves changed from green to yellow in production practice.

Recently, RNA sequencing (RNA-seq) has become a powerful technology to profile transcriptomes due to its high-throughput, accuracy, and reproducibility (Vijay et al., 2013). RNA-seq technology was used to identify the mechanism of carotenoid synthesis in *brassica campestris* L. subsp. *chinensis* var. *rosularis Tsen*. Additionally, The expression of carotenoid biosynthetic genes were investigated in the leaves of “Heixinwu” and “Huangxinwu”. The data will promote future genetic and genomics studies on the molecular mechanisms of carotenoid formation, and contribute to future applications.

2. Materials and method

2.1. Plant material and RNA isolation

Near isogenic lines (NILs) from “Heixinwu” (Black green blades) and “Huangxinwu” (Yellow blades) had been constructed. The seeds were sown in plug tray (50 holes) containing 50% grass peat and 50% vermiculite under natural irradiance. Irrigation was applied as necessary, and nutrient solution was sprayed at regular intervals. When the leaves expansion, these plants were transformed to plastic plot (17 × 17 × 20 cm). Leaves of 2-month-old plants were collected and immediately frozen in liquid nitrogen and stored at −80 °C for later use. “Heixinwu” and “Huangxinwu” chose two samples as repeats, respectively. T05 and T06 samples were taken on “Heixinwu”. T03 and T04 samples were taken on “Huangxinwu”.

2.2. RNA extraction, library construction and RNA-seq

Total RNA from each sample was isolated using a total RNA kit (TRIzol Regent, Tiangen, Beijing, China). The RNA concentration was quantified with a Nanodrop ND 1000 spectrophotometer (Nanodrop Technologies Inc., Rockland, DE, USA), and RNA integrity was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies,Santa Clara, CA, USA).

The mRNA-seq library was constructed using Illumina’s TruSeq RNA Sample Preparation Kit (Illumina Inc, San Diego, CA, USA). The isolation of mRNA, fragment interruption, cDNA synthesis, addition of adapters, PCR amplification and RNA-Seq were performed by staff at Beijing BioMarker Technology (Beijing, China). Poly-A mRNA was isolated using poly-T oligo-attached magnetic beads, and then broken into small pieces using divalent cations under an elevated temperature. The cleaved RNA fragments were then primed with random hexamers and submitted to the synthesis of the first-strand and second-strand cDNAs. The cDNAs fragments

processed for end repair and finally ligated to paired-end adaptors. Then, the cDNA fragments with 200 bp (± 25 bp) size were separated by agarose gel electrophoresisand enriched by PCR amplification. Finally, the mRNA-seq library was constructed for sequencing on the Illumina HiSeq™ 2500 sequencing platform.

2.3. Sequence data analysis and assembly

To obtain high-quality clean read data for *de novo* assembly, the raw reads from mRNA-seq were filtered by discarding the reads with adaptor contamination, masking low-quality reads with ambiguous ‘N’ bases and removing the reads in which more than 10% bases had a Q-value < 30 (Wang et al., 2010; Xie et al., 2012). The clean reads were assembled into contigs using the Trinity method (http://trinityrnaseq.sourceforge.net/), which is efficient in reconstructing full-length transcripts across a broad range of expression levels and sequencing depths (Grabherr et al., 2011). We used the Trinity method with an optimized k-mer length of 25 for *de novo* assembly. Subsequently, the contigs were linked into transcripts according to the paired-end information of the sequences. Then the transcripts were clustered based on nucleotide sequence identity. The longest transcripts in the cluster units were regarded as unigenes to eliminate redundant sequences, and then the unigenes were combined to produce the final assembly used for annotation.

2.4. Functional annotation

Functional annotations were carried out using sequence comparison with public databases. All unigenes were compared with the NCBI non-redundant nucleic acid database (NT), the NCBI non-redundant protein database (NR http://www.ncbi.nlm.nih.gov/), the Swiss-Prot database (http://www.expasy.ch/sprot), Gene Ontology (GO) (http://www.geneontology.org/), and the Clusters of Orthologous Groups database (http://www.ncbi.nlm.nih.gov/COG/) using BLAST. Pathway assignments were carried out based on the KEGG database (http://www.genome.jp/kegg).

2.5. Unigene differential expression analysis

Unigene expression levels were calculated with the formula:

FPKM (A) = cDNA Fragments/[Mapped Fragments (millions) × Transcript Length (kb)]. In the formula, we assigns FPKM (A) to be the expression of gene A, cDNA Fragments to be number of fragments compared to some transcript, Mapped Fragments (millions) to be the total number of mapped reads. The FPKM method is able to eliminate the influence of different gene lengths and sequencing discrepancy within the calculation of gene expression.

2.6. Detection of candidate SSR markers

The assembled sequences longer than 1 kb were used for the detection of SSR markers. Potential SSR markers were detected using MISA software (http://pgrc.ipk-gatersleben.de/misa/). The parameters were set for the identification of perfect dinucleotide motifs with a minimum of six repeats, and tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of five repeats (Wei et al., 2011; Zeng et al., 2010).

2.7. Real-time quantitative PCR analysis (qRT-PCR)

The *Brassica campestris* L. subsp. *chinensis* var. *rosularis Tsen* transcriptome database was mined for genes involved in carotenoid biosynthetic pathways. Total RNA was isolated from leaves using TaKaRa RNAiso Reagent (TaKaRa Code: D312) according to the manufacturer’s instructions. First-strand cDNA was synthesized using PrimeScript™ RT Reagents Kit (TaKaRa Code: DRR037A) following the manufacturers’ instructions. The related gene primers were designed using the Primer 5.0 software (PE Applied Biosystems). For quantitative real-time PCR analysis, the β-actin gene was used as an internal constitutively expressed control (house-keeping gene). These primer sequences were shown in Table 1. The size of PCR products ranged from 90 to 120 bp. Each real-time PCR was performed in triplicate with a 25-μL final volume on a Bio-Rad iQ™ thermocycler according to the manufacturer’s instructions. Reactions performed without template yielded no product.