Genome-wide identification of circRNAs involved in tomato fruit coloration

Rong Zhou [a, b], Liping Xu [c], Liping Zhao [a, b], Yinlei Wang [a, b], Tongmin Zhao [a, b, *]

[a] Institute of Vegetable Crop, Jiangsu Province Academy of Agricultural Sciences, Jiangsu, Nanjing, China
[b] Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Jiangsu, Nanjing, China
[c] Institute of Vegetable Research in Xining, Qinghai, Xining, China

ARTICLE INFO

ABSTRACT

Circular RNAs (circRNAs) are an emerging class of non-coding RNAs in plants. Our aim is to identify the circRNAs with different expression levels between tomato fruits from two cultivars ('Jinling Fenyu' and 'Jinling Moyu') at two different developmental stages. Following high-throughput sequencing, 3796 circRNAs were identified, and 243 circRNAs were shared in the four samples. As compared with the fruit at mature green stage, the expression levels of 273 and 89 circRNAs were significantly altered in the fruit at turning stage from 'Jinling Fenyu' and 'Jinling Moyu', respectively. Moreover, the parental genes of the circRNAs with significantly different expression level were mainly involved in metabolic, cellular and single-organism process and played roles in catalytic activity and binding based on GO (Gene Ontology) analysis. The results suggested that circRNAs were widespread in tomato and were generated from different chromosomes and diverse genomic regions. Some circRNAs were specific in tomato fruits at different developmental stages, which enriches the number of circRNAs in plants involved in fruit coloration and ripening. This study provides the first genome-wide profile of circRNAs involved in tomato fruit coloration and lays a foundation for studying the potential biological functions of circRNAs involved in fruit ripening.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Diverse classes of non-coding RNAs were found in cells, such as microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and circular RNAs (circRNAs) [1—3]. CircRNAs generate exons or introns in the nucleus, transfer to the cytoplasm through the nuclear pore complex and function in the cytoplasm [1,4,5]. CircRNAs were abundant with stable expression, which played crucial roles in regulating the processes of growth and development in animals [6—8]. In plants, circRNAs with potential biological functions have become the latest research focus [2,3,9].

With the development of high-throughput sequencing, functional genomics and novel bioinformatic approaches, circRNAs have been identified in plants, such as Arabidopsis, tomato and rice [2,3,10]. Ye et al. (2015) found 1,2307 and 6012 circRNAs in rice and Arabidopsis, respectively [2]. In total, 2354 circRNAs and 854 circRNAs were identified in rice and tomato, respectively [3,10]. As compared with animals, the studies concerning plants circRNAs were scarce [1,2,7]. A comprehensive detection and genome-wide identification of circRNAs is an important step [11].

Tomato (*Solanum lycopersicum* L.) is a model plant for scientific study, especially for fruit development [12—14]. Fruit ripening is a complicated developmental process that involves dramatic changes in color, texture and flavour of the fruits [15]. Color change is one of the most obvious characteristics that accompanies fruit ripening, which is coordinated with the expression of a large number of ripening-related genes [15,16]. The ripening stage of tomato fruits can be characterized by the color changing (mature green, breaker, turning and ripening), which is carried out through chlorophyll degradation and lycopene biosynthesis [15]. CircRNAs are prevalent as crucial non-coding regulators in plants [3,10,17]. However, there have been no reports concerning circRNA identification in tomatoes involved in fruit coloration.

Therefore, we firstly collected the tomato fruits from two cultivars at two different developmental stages with three replications. Then, 12 libraries from the tomato fruit samples were constructed for genome-wide circRNAs identification. The

* Corresponding author. Institute of Vegetable Crop, Jiangsu Province Academy of Agricultural Sciences, Jiangsu, China.
E-mail address: tmzhaomail@163.com (T. Zhao).

expression levels of circRNAs in two tomato cultivars at different stages were compared. Moreover, functional analysis of parental genes for the circRNAs was performed. Our hypothesis is that (1) the number and type of circRNAs differed in two cultivars and two stages, and (2) circRNAs might play important roles in tomatoes during the process of fruit coloration. Our aim is to (1) genome-widely identify the circRNAs in tomato fruits, and (2) acquire the circRNAs with different expression levels between tomato fruits at two cultivars and two developmental stages. This study will uncover the key circRNAs involved in tomato coloration during fruit ripening. The outcome will have potential implications for understanding the function of this new-emerging non-coding RNA in fruit development of plant.

2. Materials and method

2.1. Fruit sample preparation

Two tomato varieties, 'Jinling Fenyu' with pink ripening fruit and 'Jinling Moyu' with red and green stripes ripening fruit, were used as plant materials. The seeds of two varieties were sown in a plug tray (54 cm length, 27 cm width, 50 holes) with sphagnum substrate (Pindstrup, Denmark). After 30 days, the uniform seedlings were transferred to plastic pots (36-cm diameter, 28 cm height) containing the same substrate. The plants were grown in the greenhouse at $16-26\,^{\circ}C$ air temperature, $50-70\%$ relative humidity (RH) and ambient CO_2 concentration throughout the experiment. The plants were irrigated by flooding the nutrient solution according to the Japanese Garden test formula two times per week. In order to collect the fruit samples at different developmental stages, the fully open flowers were tagged from the 90-days plants at full-bloom stage. The fruit at the stage of mature green (MG) and turning (T) were harvested at 43 and 50 DPA (days post anthesis), respectively. The fruit samples of 'Jinling Fenyu' at 43 and 50 DPA (F43 and F50) and 'Jinling Moyu' at 43 and 50 DPA (M43 and M50) were taken from 5 individual plants with three replications. The samples were mixed and immediately frozen in liquid nitrogen and then stored at $-80\,^{\circ}C$ before extraction.

2.2. Library construction

The total RNAs of the 12 samples extracted using RNAprep Pure Plant Kit (DP441, Tiangen biotech, Co., Ltd, Beijing, China). The rRNA was removed from 2 μg qualified RNA per sample using NEB Next® rRNA Depletion Kit (NEB, USA). The linear RNA was removed from the rRNA-depleted RNA by incubating at $37\,^{\circ}C$ for 1 h using 20U RNase R (Epicentre, Madison, WI, USA). The 12 libraries were constructed using NEB Next® Ultra™ small RNA Sample Library Prep Kit for Illumina® (NEB, USA).

2.3. CircRNAs sequencing and identification

CircRNAs fragmentation was conducted using fragmentation buffer. The first strand cDNA was synthesized using random hexamer primer and the second strand cDNA was synthesized by adding dUTPs, RNase H, DNA Polymerase I and buffer. Cohesive ends were converted into blunt ends using T4 DNA and Klenow DNA polymerase. After the adenylation of 3′ ends, the NEBNext Adaptor with hairpin loop structure were ligated. The library fragments were purified by adding AMPure XP Beads (Beckman Coulter, Beverly, USA). The second strand cDNA with U was degraded using USER enzyme (NEB, USA). Then, the PCR was conducted to obtain the circRNA library. The library insert size and the accurate quantitative analysis of the library effective concentration was detected using Agilent 2100 and Q-PCR, respectively. The

library was qualified when the insert size was ~220bp and the library effective concentration>2 nM. The 12 qualified libraries were sequenced using Illumina Hiseq X-Ten platform (Biomarker Technologies, China).

The clean reads with high quality were obtained by removing the junk reads with adaptor, ploy-N and low quality from raw reads. Then, the clean reads were blasted with reference genome sequence of *Solanum Lycopersicum* (ftp://ftp.ensemblgenomes.org/pub/release32/plants/fasta/solanum_lycopersicum/dna). The CircRNAs were taken into account using Find-circ software when (1) GU/AG was on the sides of splice site; (2) clear breakpoint could be detected; (3) two mismatches; (4) breakpoint appeared in the position within 2 nucleotide (nt); (5) at least two reads supported junction; (6) the score of blasting to the right position of short sequence should be ≥ 35 higher than blasting to other positions.

2.4. Expression level comparison of circRNAs

The circRNAs with low expression levels were deleted before comparing the expression levels. In detail, the counts per million (cpm) of circRNAs were calculated. The circRNAs with cpm>1 in three replicated samples at the same time were kept. The expression level of the remained circRNAs with high quality were compared using the DESeq R package (1.10.1). To control false discovery rate (FDR), the *P* values were adjusted using the Benjamini-Hochberg correction method. The circRNA was considered to be up- or down-regulated when $| \log_2 \text{(fold change)} |>1$ and *P* value < 0.05. The expression level of circRNAs identified in three replications was averaged as the final value.

2.5. Functional analysis of circRNA parental genes

The GO (Gene Ontology) enrichment analysis for the parental genes of the differently expressed circRNAs (F43 *vs* F50, M43 *vs* M50, F43 *vs* M43 and F50 *vs* M50) was performed.

3. Results and discussion

3.1. Overview of circRNA sequencing data and circRNA identification in tomato fruits

The GC content and Q30 of the clean reads were $\geq 42.86\%$ and 90.92% in each library, indicating that clean reads obtained in this study were high-quality. High-throughput sequencing identified 45,032,805, 45,865,305, 50,176,215 and 50,177,347 clean reads in the library of F43, F50, M43 and M50, respectively. The mapped ratios of the four samples were 99.11%, 99.37%, 99.73% and 99.62%, respectively. Moreover, the unique mapped ratios of the four samples were 94.95%, 95.35%, 96.36% and 96.17%, respectively. These indicated that the data were qualified for further circRNA identification.

In total, 3796 unique circRNAs were identified in tomato fruits, which uncovered the widespread occurrence of circRNAs in tomatoes and enriched the circRNAs number in plants. Of the circRNAs, 553 unique circRNAs were generated from Chr1, which accounted for the most (14.57%) (Fig. 1A). Moreover, 3083 (81.22%) circRNAs were produced from exons of protein-coding genes; 527 (13.88%) circRNAs were from intergenic regions; 186 (4.90%) circRNAs were produced from introns (Fig. 1B). These suggested that different chromosomes produced different numbers of circRNA and exonic circRNAs accounted for the most, while intronic circRNAs accounted for the least, which was in accordance with Zuo et al. (2016) [3].

To obtain the primary circRNAs in tomatoes that affected fruit coloration, tomatoes from two cultivars at the stage of mature