Journal of Plant Physiology 170 (2013) 80–92

Contents lists available at SciVerse ScienceDirect

Journal of Plant Physiology

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

Bioinformatics prediction of miRNAs in the *Prunus persica* genome with validation of their precise sequences by miR-RACE

Yanping Zhang[a,1], Youhuang Bai[b,1], Jian Han[a], Ming Chen[b], Emrul Kayesh[a], Weibing Jiang[a], Jinggui Fang[a,*]

[a] College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China
[b] College of Plant Protection, Zhejiang University, Hangzhou 310095, China

ARTICLE INFO

Article history:
Received 27 April 2012
Received in revised form 17 August 2012
Accepted 21 August 2012

Keywords:
Prunus persica
microRNAs
Computational prediction
qRT-PCR
miR-RACE

ABSTRACT

We predicted 262 potential MicroRNAs (miRNAs) belonging to 70 miRNA families from the peach (*Prunus persica*) genome and two specific 5′ and 3′ miRNA rapid amplification of cDNA ends (miR-RACE) PCR reactions and sequence-directed cloning were employed to accurately validate 61 unique *P. persica* miRNAs (Ppe-miRNAs) sequences belonging to 61 families comprising 97 Ppe-miRNAs. Validation of the termini nucleotides in particular can define the real sequences of the Ppe-miRNAs on peach genome. Comparison between predicted and validated Ppe-miRNAs through alignment revealed that 43 unique orthologous sequences were identical, while the remaining 18 exhibited some divergences at their termini nucleotides. Quantitative real-time polymerase chain reaction (qRT-PCR) was further employed to analyze the expression of all the 61 miRNAs and 10 putative targets of 8 randomly selected Ppe-miRNAs in peach leaves, flowers and fruits at different stages of development, where both the miRNAs and the putative target genes showed tissue-specific expression.

Introduction

MicroRNAs are small endogenous non-coding 20-22 nucleotides (nt) long RNA molecules derived from genome-encoded stem-loop precursors. These small RNAs play very important roles in post-transcriptional gene regulation through degradation of target mRNAs or by repression of targeted gene translation in organisms by base pairing with their target genes (Carrington and Ambros, 2003; Bartel, 2004; Mallory and Vaucheret, 2004; Navarro et al., 2006; Chekulaeva and Filipowicz, 2009; Krol et al., 2010). MiRNAs play vital roles in a variety of biological and metabolic process in plants, such as leaf morphogenesis and polarity, floral organ identity and differentiation, organ development, signaling pathways and in responses to biotic and abiotic stresses (Aukerman and Sakai, 2003; Juarez et al., 2004; Guo et al., 2005; Jones-Rhoades et al., 2006; Xu et al., 2007; Nag et al., 2009; Inui et al., 2010). Studies on miRNAs have grown to be a key research area with the identification and characterization of miRNAs in various organisms being fundamentally important.

Hundreds of miRNAs from diverse plants have recently been discovered, with many being highly evolutionarily conserved between species (Floyd and Bowman, 2004; Zhang et al., 2006b; Wang et al., 2011b). The conservation of miRNA sequences within the plant kingdom has made it possible to predict and validate conserved miRNAs and their target genes from all plant species through bioinformatics and experimental methods (Axtell and Bartel, 2005; Fahlgren et al., 2007). With the advances in bioinformatics, many computational methods have been employed for the prediction and identification of miRNAs (Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Zhang et al., 2006c; Yin et al., 2008) leading to discovery of miRNAs in dozens of plants (Zhang et al., 2006a, 2007, 2012; Qiu et al., 2007; Xie et al., 2007; Gleave et al., 2008; Yin et al., 2008; Carra et al., 2009; Song et al., 2009, 2010; Yu et al., 2010; Zeng et al., 2010). These discoveries further proved that the computational or bioinformatics-based method of predicting novel miRNAs is very useful, as this option has the capacity to reveal miRNAs which are usually not easily detected by direct cloning. Although computational prediction of miRNA can theoretically discover all the potential miRNAs based on the criteria set for characterization of miRNAs, there is an inherent chance that some false predictions will occur, probably due to the evolution of miRNAs and the non-availability of miRNA end characteristics that exactly define the start and stop codes of mature miRNAs. This is the main disadvantage of computational prediction compared to direct cloning of miRNAs. Initial studies on miRNAs were mainly focused on determining their expression using Northern blotting and/or qRT-PCR techniques which, despite being robust, can only confirm the existence and size, but not the entire precise sequences of miRNAs targeted for identification. The newly reported miR-RACE

* Corresponding author.
E-mail address: fanggg@njau.edu.cn (J. Fang).
[1] First Co-authors.

0176-1617/$ – see front matter © 2012 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.jplph.2012.08.021

(Song et al., 2009) can overcome the shortcomings of Northern blotting and qRT-PCR techniques as it is preferentially suitable in the validation of precise sequences, especially both ends, of computationally predicted miRNAs in an organism. This efficiency of miR-RACE is exemplified by successful utilization in earlier studies on the identification of precise miRNA sequences in several fruit crops (Song et al., 2009; Yu et al., 2010; Wang et al., 2011a; Zhang et al., 2012). Validation of the precise sequences of miRNAs is as critical due to its importance in downstream research, such as determining the mechanisms by which miRNAs act on their targets, as well as in studies on miRNA evolution.

Prunus persica is also among the most economically important fruit crops worldwide, having played an important role in human life for a long time. It is cultivated both as an ornamental and for fruit production. With the rapid development in plant genomics, the entire genomes of many plants have been sequenced in succession, with the peach genome sequence being completed and released publically on April 1, 2010 (http://www.rosaceae.org/peach/genome). The availability of this important resource makes it possible to initiate research on peach functional genomics, including the prediction and characterization of Ppe-miRNAs. Although several Ppe-miRNAs have been predicted and validated based on the peach expressed sequence tag (EST) database (Zhang et al., 2012), they compare poorly with the number of miRNAs reported in *Arabidopsis* and other plants, thus calling for identification and characterization of many more Ppe-miRNAs. To the best of our knowledge, there are no reports on large-scale prediction and determination of precise sequences of peach miRNAs. This study was therefore mooted to address this information gap, whereby we used all previously identified miRNAs from plant species deposited in the miRNA registry database (Release 17.0, April 2011) to search for homologs of these miRNAs in the peach genome database (http://www.rosaceae.org/peach/genome). From this exercise, we identified a total of 70 Ppe-miRNA families comprised of 262 potential Ppe-miRNAs, and we subsequently predicted candidate target genes for all these Ppe-miRNAs. Spatiotemporal expression of 61 Ppe-miRNAs and 10 putative targets genes of 8 randomly selected Ppe-miRNAs were also studied in peach leaves, flowers, and fruits at different stages of development. The findings of this study provide a very important foundation for further studies needed to check the role of Ppe-miRNAs on the control of peach development and also to assess putative functions for miRNAs in plants.

Materials and methods

Plant material

The samples were collected from peach 'liangji' grown at the Peach Resource Garden of The Jiangsu Academy of Agricultural Sciences, China. Young and old leaves (1 month old and 6 months old), flower buds, small and big size flowers (semi-open and fully open), and fruits at different stages of development (1 cm, 3 cm and 4.5 cm diameter) were collected. After collection, all the samples were immediately frozen in liquid nitrogen and stored at −80 °C.

Prediction of potential Ppe-miRNA and their putative targets

All known plant miRNAs were downloaded from the miRNA registry database (Release 17.0, April 2011) (Griffiths-Jones et al., 2008), and non-redundant 2060 miRNAs were used as query for mining conserved miRNAs in *Prunus persica*. The peach genome sequence and peach predicted CDS were downloaded from the Peach Genome v1.0 (Release 1.0, April 2011; http://www.rosaceae.org/peach/genome; http://www.rosaceae.org/node/355). A flowchart illustrating the search for potential miRNAs is shown in Fig. 1. The Patscan program was used to identify matches to known miRNAs in the target genomic sequence with 0–2 substitutions and no insertions/deletions. The miRcheck algorithm, developed by Jones-Rhoades (2010), was used to evaluate the local secondary structure of the miRNA to identify homology miRNA candidates. These resulting candidates were retained as conserved miRNA candidates of *P. persica* only if they met the following criteria: (i) no hit against the NCBI non-redundant database using Blastn 2.2.21 release. One potential source of false positives is the mis-annotation of miRNA binding sites in target genes as miRNA hairpins because most plant miRNAs are highly complementary to target miRNAs (Jones-Rhoades, 2010). (ii) The location of putative miRNA within its hairpin precursor in *P. persica* is consistent with the location of the query miRNA within its hairpin precursor in other plants. Each miRNA family has a characteristic location in which the mature miRNA is always found (Jones-Rhoades, 2010). We defined the candidate miRNA as Ppe-miRNA and these were divided into different families, just as their homologous miRNAs in other plants who, as queried, the members of each family were then given the number a-. . .z, according to the order they were predicted. The secondary structures of stem-loop hairpin sequences were predicted using RNAfold (Hofacker and Stadler, 2006). For further prediction of the miRNA targets, all *P. persica* transcripts sequences were used and all the miRNA targets were predicted through the psRNA target web server with the default parameters (Dai and Zhao, 2011).

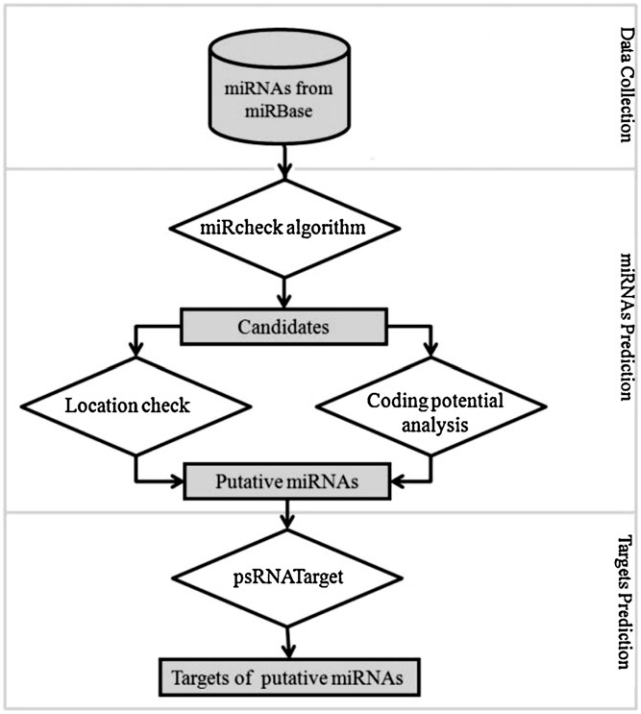

Fig. 1. Schematic representation of the search procedure used to identify homology of peach miRNA and their genes with known plant miRNAs and target genes.

Low molecular RNA extraction and construction of small RNA cDNA libraries

Total RNA isolated from 100 mg of previously collected tissue using TRIZOL reagent (Invitrogen, Life Technologies, Carlsbad, CA) (Zhang et al., 2012). Low and high molecular weight RNAs (LMW and HMW-RNAs) were isolated with 4 M LiCl (Adai et al., 2005; Song et al., 2010). The procedure of construction cDNA libraries of small RNAs was followed as previously reported to generate the miRNA-enriched library (Song et al., 2009). The HMW-RNA samples