Genomics 105 (2015) 242–251

Contents lists available at ScienceDirect

Genomics

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

Profiling microRNA expression during multi-staged date palm (*Phoenix dactylifera* L.) fruit development

Chengqi Xin [a,b,c,1], Wanfei Liu [a,b,1], Qiang Lin [a,b,1], Xiaowei Zhang [a,b], Peng Cui [d], Fusen Li [a,b], Guangyu Zhang [a,b], Linlin Pan [a,b], Ali Al-Amer [a], Hailiang Mei [a,b,c], Ibrahim S. Al-Mssallem [a,e,*], Songnian Hu [a,b,**], Hasan Awad Al-Johi [a,***], Jun Yu [a,b,****]

[a] Joint Center for Genomics Research (JCGR), King Abdulaziz City for Science and Technology and Chinese Academy of Sciences, Prince Turki Road, Riyadh 11442, Saudi Arabia
[b] CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, No. 1 Beichen West Road, Chaoyang District, Beijing 100101, China
[c] University of Chinese Academy of Sciences, No. 19A Yuquan Road, Shijingshan District, Beijing 100049, China
[d] Biological and Environmental Sciences and Engineering Division, King Abdullah University of Sciences and Technology, Thuwal 23955-6900, Saudi Arabia
[e] Department of Biotechnology, College of Agriculture and Food Sciences, King Faisal University, Hofuf, Saudi Arabia

ARTICLE INFO

Article history:
Received 7 November 2014
Accepted 20 January 2015
Available online 29 January 2015

Keywords:
Phoenix dactylifera
MiRNA
Fruit development
MiRNA target
Gene regulation

ABSTRACT

MicroRNAs (miRNAs) play crucial roles in multiple stages of plant development and regulate gene expression at posttranscriptional and translational levels. In this study, we first identified 238 conserved miRNAs in date palm (*Phoenix dactylifera*) based on a high-quality genome assembly and defined 78 fruit-development-associated (FDA) miRNAs, whose expression profiles are variable at different fruit development stages. Using experimental data, we subsequently detected 276 novel *P. dactylifera*-specific FDA miRNAs and predicted their targets. We also revealed that FDA miRNAs function mainly in regulating genes involved in starch/sucrose metabolisms and other carbon metabolic pathways; among them, 221 FDA miRNAs exhibit negative correlation with their corresponding targets, which suggests their direct regulatory roles on mRNA targets. Our data define a comprehensive set of conserved and novel FDA miRNAs along with their expression profiles, which provide a basis for further experimentation in assigning discrete functions of these miRNAs in *P. dactylifera* fruit development.

1. Introduction

Date palm (*Phoenix dactylifera* L.), as a socio-cultural symbol for the Arabian Peninsula (Saudi Arabia and Gulf countries), has a long agriculture history [1] and is also an economically important food crop in tropical and subtropical regions. Its major edible part is its ripe fruit or the date that contains a large amount of monosaccharides and has high nutritional value [2]. The *P. dactylifera* date development and ripening is a complex process, including a long period from the initial pollination for fruit set to the terminal sugar-rich stage, during which its major metabolic mechanisms undergo a complex series of physiological events, such as carbon fixation, starch hydrolysis and synthesis, and glucose and fructose accumulation. The starch accumulated in the early stages eventually converts into easily-absorbed small sugar molecules, mostly monosaccharides, at the late stages, which include fructose, sucrose, and glucose [3]. It has been reported that several key enzyme-encoding genes, such as pyruvate kinase, glucosidase, synthase/phosphate synthase, and trehalose 6-phosphate synthase/phosphatase, play crucial roles in the *P. dactylifera* fruit development and ripening [3–5]. However, upstream regulators of these key genes, such as miRNAs, have yet to be studied in the species and in details.

MicroRNAs (miRNAs) are a class of endogenous small non-coding RNAs (20 to 24 nt in size), which are derived from single-stranded precursors that form a stem–loop secondary structure and common among animals, plants, and even viruses [6,7]. MiRNAs regulate gene expression by binding to their complementary sequences, leading to either cleavage-induced degradation or translational repression of their target transcripts [8]. MiRNAs have been studied in plants since 2002 and known to play important regulatory roles in plant hormone

* Correspondence to: I. Al-Mssallem, Joint Center for Genomics Research (JCGR), King Abdulaziz City for Science and Technology and Chinese Academy of Sciences, Prince Turki Road, Riyadh 11442, Saudi Arabia. Fax: +966 0135808411.
** Correspondence to: S. Hu, CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, No. 1 Beichen West Road, Chaoyang District, Beijing 100101, China. Fax: +86 10 84097458.
*** Correspondence to: H. Al-Johi, Joint Center for Genomics Research (JCGR), King Abdulaziz City for Science and Technology and Chinese Academy of Sciences, Prince Turki Road, Riyadh 11442, Saudi Arabia. Fax: +966 4813787.
**** Correspondence to: J. Yu, CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, No. 1 Beichen West Road, Chaoyang District, Beijing 100101, China. Fax: +86 10 84097540.
E-mail addresses: imssallem@gmail.com (I.S. Al-Mssallem), husn@big.ac.cn (S. Hu), haljohi@kacst.edu.sa (H.A. Al-Johi), junyu@big.ac.cn (J. Yu).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ygeno.2015.01.004

homeostasis [9], stress responses [10], and diverse developmental processes, including seed development, meristem and lateral organ development, root initiation, flowering and sex determination, fruit development, timing and phase transitions [11,12]. Since the first discovery of miRNAs from *Arabidopsis* in 2002 [13], plant miRNAs from various species have been intensely studied through both experimental and computational approaches. A recent update has pointed out that there have been 21,264 miRNAs discovered from 193 species and 5943 of them are plant in origin (miRBase, Release 19.0, August 2012) [14]. These plant-specific miRNAs include 713 from *Arabidopsis*, 721 from *Oryza*, 374 from *Populus*, and 719 from *Medicago truncatula*, and the rest are mostly from a dozen or so other plant species.

The *P. dactylifera* (*Pd*) genome has been sequenced [4,15], and its high-coverage sequences and the genome assemblies facilitate many aspects of genome-wide studies that include miRNA identification. Furthermore, the study of transcriptomic profiles for its fruit development and ripening allows direct comparison between miRNAs and their target genes expression [3,4]. In this study, we use a combined strategy—computational prediction and experiment identification—to study *P. dactylifera* or *Pd* miRNAs. In the computational approach, we use several filters based on known characteristics of plant miRNAs [13,16]. In the experimental approach, we sequence 6 small RNA libraries from different *P. dactylifera* fruit development stages, using the next-generation sequencing platforms. We correlate the miRNA expression profiles to their corresponding transcriptome from our previous studies of the *P. dactylifera* fruit development stages and find a clear negative correlation between most miRNAs and their target gene transcripts. Several important miRNAs are up-regulated during fruit development and ripening, and the mRNA targets of these fruit-development-associated (FDA) miRNAs are found involved in carbon metabolism, such as starch and sucrose metabolisms, and our results suggest that miRNA may play critical roles in regulating sugar accumulation and conversion, which are key transitional events to fruit ripening. Our genome-wide *Pd* miRNA characterization provides novel insights into the dynamics of miRNAs in regulating fruit development and evidence for understanding gene regulation in fruit development and ripening.

2. Results

2.1. Prediction of conserved miRNAs in P. dactylifera genome

In the plant kingdom, a substantial number of miRNAs are highly conserved in different lineages, ranging from mosses and gymnosperms to angiosperms. Such homologous miRNA families are also typically function-conserved and play essential regulatory roles across plant taxa [17]. Sequence and structure homologies are the central theory of computing-based approaches for miRNA prediction, and many computational methods and protocols have been developed to date [18,19]. We obtained 5943 known mature miRNA sequences from the miRBase (Release 19.0 August 2012) [14] and searched them against the *P. dactylifera* genome assembly put together by our group. The effort yielded 238 potential *Pd* miRNAs, which belong to 54 miRNA families (Supplementary Table S1 in Ref. [20]). Of the 54 miRNA families, the miR169 family is the largest with 24 members, similar to what has been observed in other plant species [21,22]. We have 6 other families, miR171, miR156, miR164, miR167, miR172 and miR529, whose members per family are all greater than 10 (Fig. 1A). Among the 54 miRNA families, 19 are highly conserved (more than 10 species including *Arabidopsis* and rice) among many plant species, of which 8 families (miR156, miR159, miR160, miR166, miR167, miR171, miR172, and miR396) are present in more than 30 monocots and dicots, whereas the other 35 show less evolutionary conservation (less than 10 species) (Fig. 2). Interestingly, several dicot-associated miRNAs (including miR391, miR479, miR828, miR845, miR856, miR1511, miR1918, miR2673, miR5225, miR5645, miR6034, and miR6150) are identified in *P. dactylifera* which is actually a monocotyledon plant but a rather primitive one according to plant systematics. Moreover, miR536, which is only identified in the lower plants, *Physcomitrella patens* (bryophyte) and *Selaginella moellendorffii* (lycophyte) (Fig. 2), is found for the first time in an angiosperm. These data suggest that the *Pd* miRNAs are both specific and complex.

The length of the 238 predicted conserved miRNAs ranges from 18 to 24 nt, where 21-nt miRNA takes the majority (43.28%), followed by the 19-nt (16.80%), 22-nt (15.96%), 20-nt (13.86%), 18-nt (6.30%), 23-nt

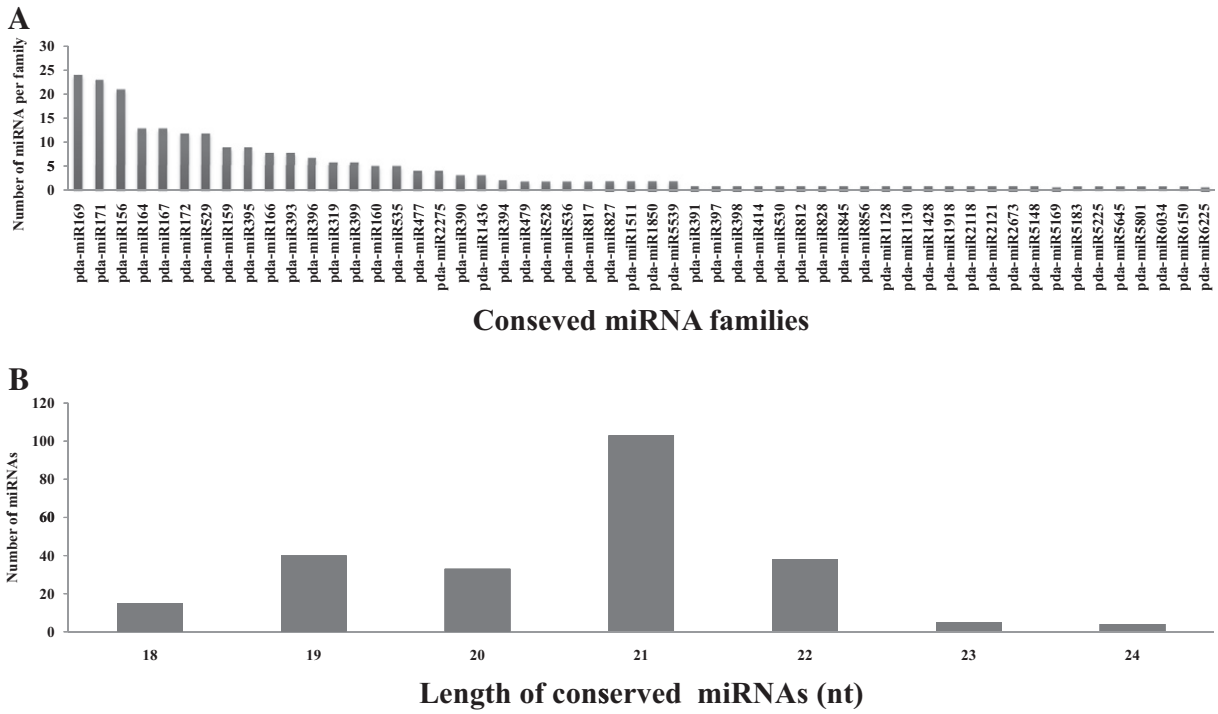

Fig. 1. The distribution of conserved miRNAs in *P. dactylifera*. A, MiRNA members in different miRNA families. B, The length distribution of conserved miRNAs.

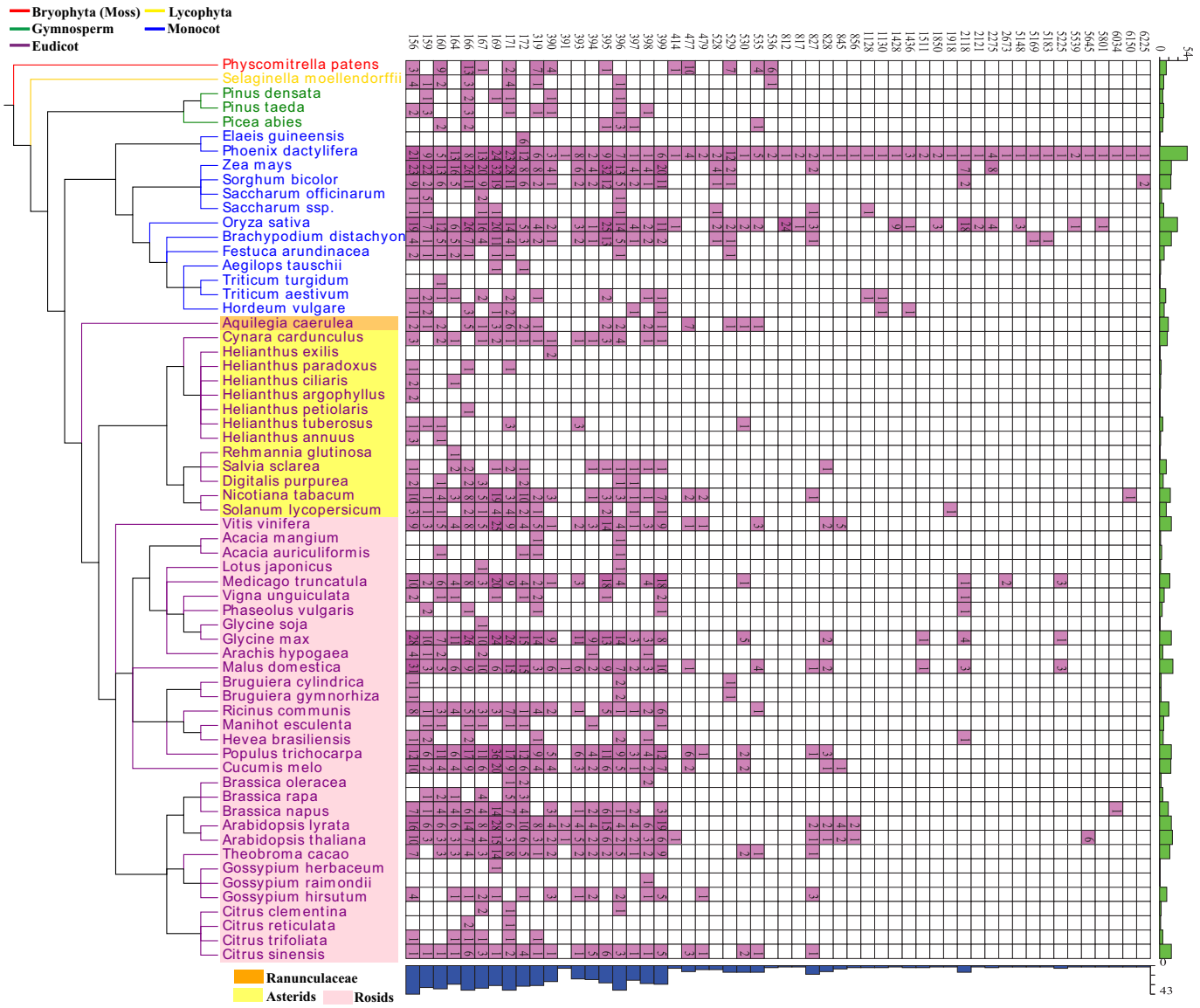

Fig. 2. Comparative analysis of conserved *P. dactylifera* miRNA families with other plant species. The left panel is a phylogenetic tree of 63 plant species built based on the NCBI common tree tool. The right panel shows the miRNAs of different families in each species. The right and the bottom bar diagrams display the miRNA families of each species and the species in each family, respectively. The top right numbers stand for miRNA families.

(2.10%), and 24-nt (1.68%) miRNAs (Fig. 1B). The predicted precursor sequences have a length range from 60 nt to 230 nt, and all are able to form the typical stem–loop structure with the mature miRNA on one of its secondary-structure-defined arms (Supplementary Fig. S1). We identified 128 (53.78%) and 110 (46.22%) miRNAs on the 3′ and 5′ arms, respectively, and those of the same family are not necessarily found on the same arm of the pre-miRNAs (Supplementary Table S1 in Ref. [20]). Further analysis showed that about half (109 miRNAs) of the predicted conserved mature *Pd* miRNA sequences began with uracil (U) (Supplementary Table S1 in Ref. [20]), and the specificity is consistent with previous results in other plants [21,23], due to the high-affinity of AGO proteins to U in the 5′ terminus of mature miRNA sequences.

2.2. Identification of novel Pd miRNAs involved in fruit development and ripening

To investigate the dynamics of miRNAs during *P. dactylifera* fruit development and ripening, we performed miRNA sequencing using samples from various developing and ripening fruits collected, which are divided into six stages: 0 DAF (days after fertilization), 15 DAF, 45 DAF, 75 DAF, 105 DAF, and 120 DAF. From the six independently-constructed small RNA libraries, we generated 67 to 87 million reads per library for the analysis (Supplementary Table S2 in Ref. [20]). Based on the sequencing data, we identified 78 conserved miRNAs that are expressed in at least one of the six libraries (Supplementary Table S3 in Ref. [20]); 87% (68) are shared by all six libraries and 91% (71) have both miRNA and miRNA* (the complementary strand to the mature miRNA) sequences.

One of the most important advantages of high-throughput sequencing is its usefulness in detecting novel miRNAs. Having set the criterion, where a miRNA candidate must have its miRNA* in at least one of the libraries or the mature sequence should be detected from more than two libraries, we identified 276 novel miRNAs associated with *P. dactylifera* fruit development and ripening (Supplementary Table S4 in Ref. [20]), of which 251 are expressed in all six libraries. The precursors of these novel miRNAs can be folded into the typical secondary structure by using RNAFold. Six of these novel miRNAs are selected for demonstrating their secondary structures (Fig. 3A), and we also confirmed their expressions using stem–loop RT-PCR. As a result, all

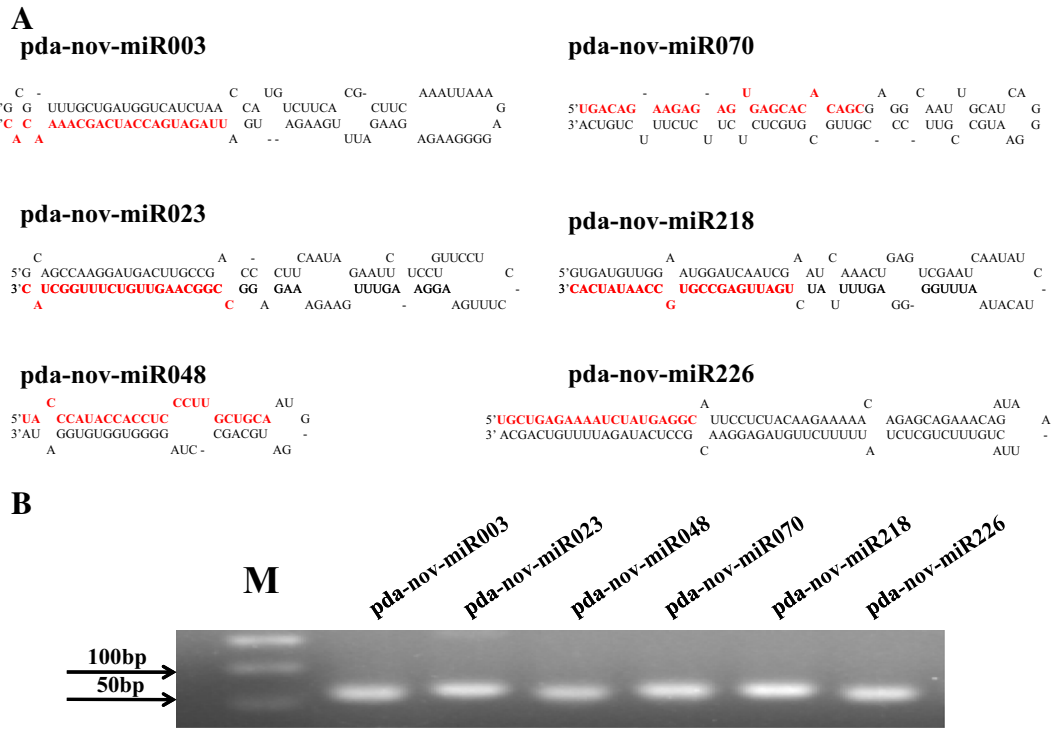

Fig. 3. Examples of novel *P. dactylifera* miRNAs identified in this study. A, Predicted stem–loop structures of novel miRNA precursors based on RNAFold. Mature miRNA sequences are highlighted in red. B, Stem–loop RT-PCR analysis of some novel miRNAs. The sizes of obtained PCR products are approximately ~60 bp. M indicates a 50-bp DNA ladder (TIANGEN).

the selected novel miRNAs are expressed in *P. dactylifera* fruits, and it suggests the effectiveness of our computational methods (Fig. 3B). The novel *Pd* miRNAs have a narrow length distribution of 20–25 nt, and we subsequently named them as pda-nov-miR + number (e.g., pda-nov-miR001) temporarily before obtaining their official accessions in the public databases.

2.3. Target prediction of conserved and novel miRNAs in P. dactylifera

The major challenge in revealing the function of a miRNA is to identify its regulatory target(s). Plant miRNA often shows perfect or near-perfect sequence complementarity to its targets [24,25], which allows effective target prediction based on computational approach. Using a web server, psRNATarget, and its default parameters, we identified 1135 potential targets for 167 conserved miRNAs (Supplementary Table S7 in Ref. [20]) and 795 potential targets for 199 novel miRNAs (Supplementary Table S8 in Ref. [20]) in *P. dactylifera*. KEGG analysis showed that these target genes participate in many biological pathways, such as starch and sucrose metabolism, glycolysis/gluconeogenesis, amino sugar and nucleotide sugar metabolism, citrate cycle (TCA cycle), fructose and mannose metabolism, and carbon metabolism (Supplementary Table S10 in Ref. [20]); many of these targets are conserved among various plant species, suggesting conserved regulatory roles. Among the pool of conserved targets, its majority are transcription factors [26], whereas others are associated with plant metabolism and response to environment stresses. These results are similar to those reported previously in other plant species, such as *Arabidopsis*, rice, corn, soybean, cotton, and tomato [11,24,27,28]. For examples, we found many transcription factors including SBP, ARF, NAC, MYB, HD-ZIP, GRAS, and AP2 as targets of miR156, miR160, miR164, miR159, miR166, miR171, and miR172 in *P. dactylifera*, respectively (Table 1 in Ref. [20]). In addition to the conserved transcription factors, several new targets related to sugar metabolism are also identified, including trehalose 6-phosphate synthase/phosphatase, pyruvate kinase, malate dehydrogenase, frutokinase, cysteine synthase A, serine O-acetyltransferase, and pyrophosphate-fructose-6-phosphate 1-phosphotransferase; all suggest regulatory roles in various metabolic pathways in *P. dactylifera* fruit development.

2.4. MiRNA expression profiles during fruit development and ripening

High-throughput sequencing data not only provide identity of miRNAs but also information of their expressions, especially for some with a low expression level. We identified 78 conserved and 276 novel miRNAs in *P. dactylifera*, which are expressed in the six libraries (Supplementary Table S3 in Ref. [20]). To determine which miRNAs are actually involved in fruit development and ripening as fruit-development-associated or FDA miRNAs, we identified 268 differentially expressed (DE) miRNAs (fold-change ≥ 2; $P \leq 0.001$) between every two adjacent libraries by comparing the normalized expression data of the 354 expressed miRNAs (including 78 conserved and 276 novel miRNAs; Table 1). We also clustered these DE miRNAs based on STEM (short time-series expression miner) [29], where 137 out of the 268 DE miRNAs were clustered into six expression patterns (Fig. 4; Supplementary Table S5 in Ref. [20]). There are many expression profiles identified to have different dynamics regarding to developmental timing (Fig. 4B). To validate some of the expression profiles experimentally, we

Table 1
Expression data of differentially expressed miRNAs between libraries.

Adjacent stages	Numbers of DE miRNAs		
	Total DE miRNAs	Up	Down
15/0 DAF	126	16	110
45/15 DAF	123	107	16
75/45 DAF	157	78	79
105/75 DAF	189	107	82
120/105 DAF	177	49	128

DE, differentially expressed (|log2fold-change| ≥ 1; p-value ≤ 0.001); Up, up regulation; Down, down regulation.

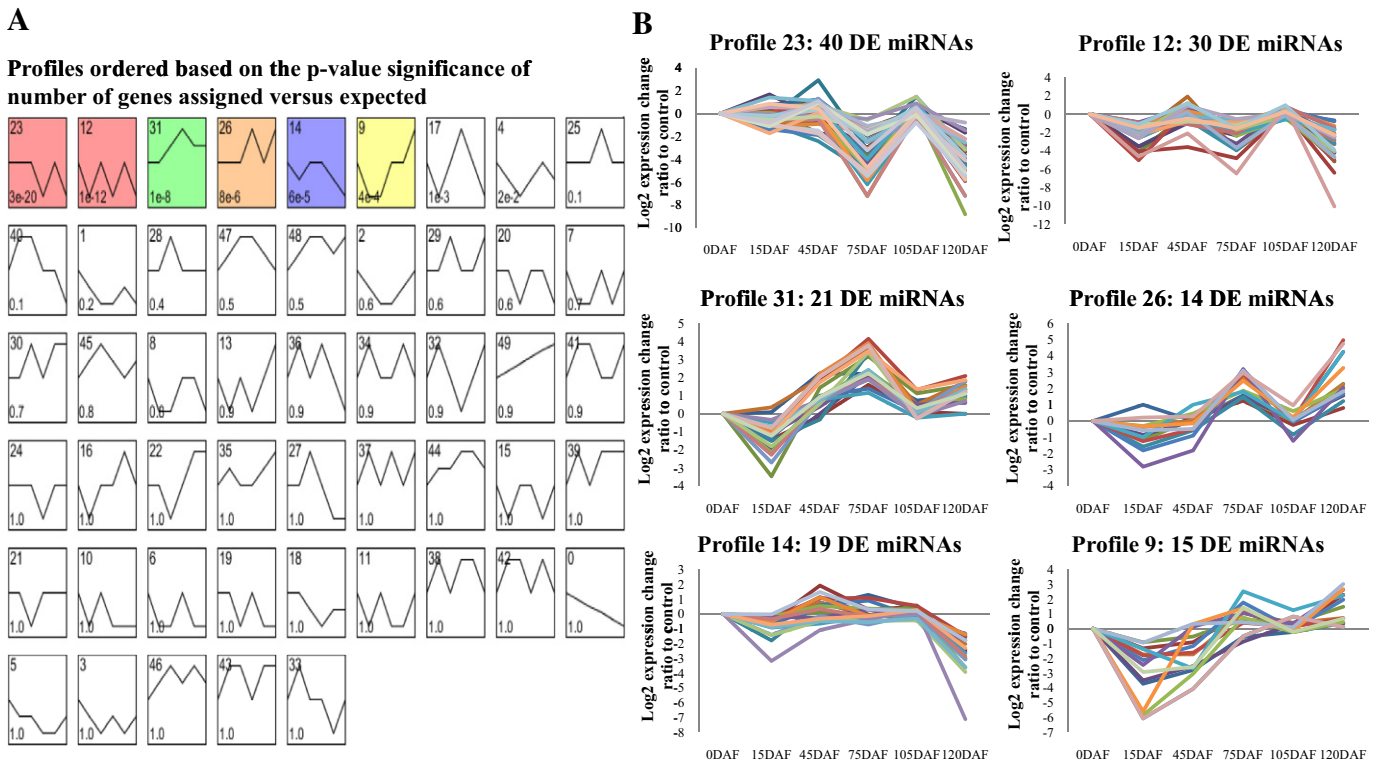

Fig. 4. MiRNA expression profiles analysis based on STEM clustering. A, Each box corresponds to a model expression profile in STEM results, and only colored profiles reach statistical significance in the analysis. The upper-left number in the box gives information about the order of profile and the bottom-left indicates p-values. B, Six significant clusters of miRNA profiles (colored in A) in *P. dactylifera* fruit development are displayed in time course (log2 miRNA expression ratios to control). DAF, days after fertilization.

performed stem–loop real-time quantitative RT-PCR. Most expression patterns in the tested miRNA set are overall confirmed, albeit with a trend discrepancy in one of the samples (constant vs. middle downward; Fig. 5); protocols based on different methodologies are in principle not at all equivalent in every aspect. Nevertheless, future focused studies are of essence in resolving the discrepancy.

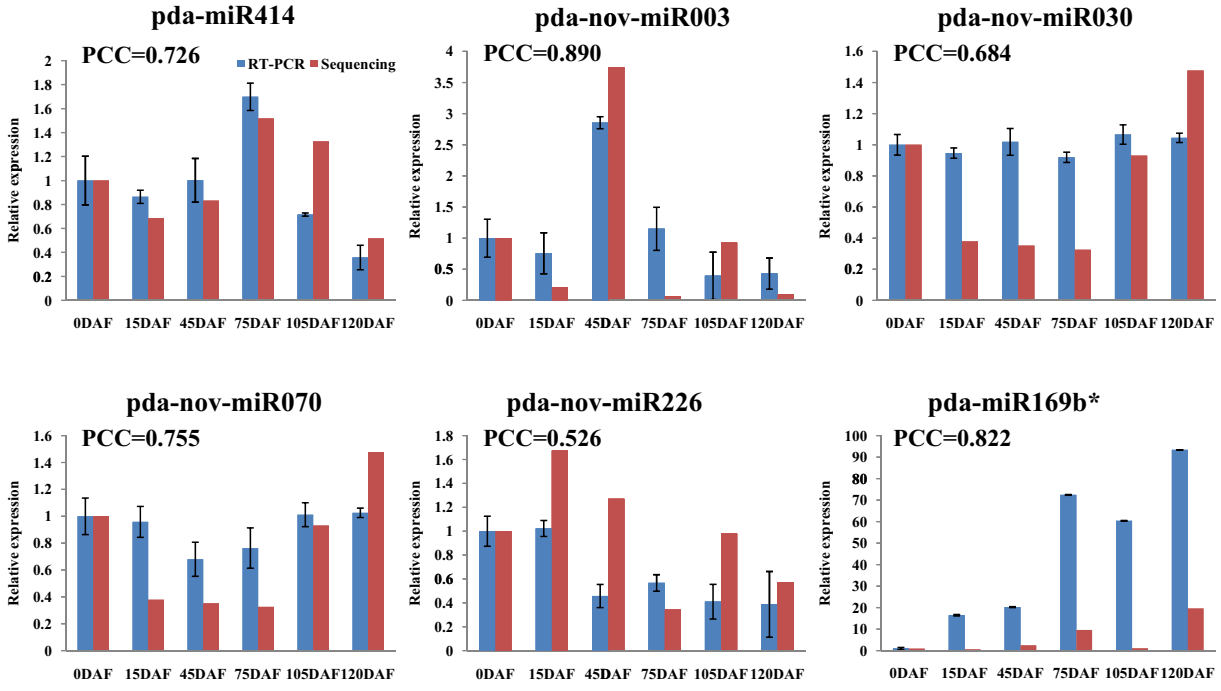

Fig. 5. Validation of miRNA expression based on quantitative RT-PCR. The expression levels of each miRNA are normalized by comparing to that of miRNA at 0 DAF, which is set to be 1.0. The experiments are repeated with three replicates and error bars represents standard errors. Correlation of the abundance between RT-PCR and sequencing data is calculated as Pearson correlation coefficient (PCC).

2.5. Expression correlation analysis between miRNAs and their corresponding targets

In general, expressions of most miRNAs are negatively correlated with their corresponding targets [30]. Information from our previous study of transcriptomic profiles for *P. dactylifera* fruit development and ripening allows the direct comparison between miRNAs and their targets' expression [3,4]. We correlated the expression of all expressed miRNAs, including 78 conserved miRNAs and 276 novel miRNAs, to that of their corresponding targets. Together, 579 out of 995 targets show negative correlation with their corresponding miRNAs (221 miRNAs) (Supplementary Fig. S2A; Supplementary Table S1), and such correlation suggests that a majority of the miRNAs are potentially involved in guiding the degradation of their target transcripts in the process of fruit development as FDA miRNAs. This result is very similar to that of rice grain filling reported recently [9].

We next applied gene ontology (GO) and KEGG pathway analyses to functionally categorize the miRNA targets. GO analysis showed that the genes annotated to the top scoring terms were mainly involved in developmental processes, cell differentiation and death, and regulation of cellular process (Table 2A, $p < 0.01$). Other terms include genes responsible for stimulus response, catalytic activity, and protein binding. The KEGG enrichment analysis revealed 5 pathways to be overrepresented (Table 2B, $p < 0.05$), one of which was related to starch and sucrose metabolism. These data clearly indicated that these targets were involved in fruit development and ripening, suggesting that miRNAs play important regulatory roles in the date formation and ripening.

In our data, the expression of pda-nov-miR110 is negatively correlated with its corresponding target—pyruvate kinase (XM_008804402 in NCBI), a key enzyme involved in glycolysis and contributes to the glucose and fructose accumulation in dates (Fig. 6A). In another case, the transcript of beta-glucosidase (XM_008777200), a glucosidase enzyme involved in starch and sucrose metabolism, is negatively correlated with the level of pda-nov-miR204 (Fig. 6B). Beta-glucosidase mainly catalyzes the hydrolysis of terminal non-reducing residues in beta-D-glucosides with the release of glucose [31], which contributes to the accumulation of glucose in the date. Another target involved in starch and sucrose metabolism, trehalose 6-phosphate synthase/phosphatase (XM_008794068), showed negative correlation with its corresponding miRNA pda-miR414 (Fig. 6C). Trehalose 6-phosphate synthase/phosphatase functions by catalyzing the transfer of glucose from UDP-glucose to glucose-6-phosphate and contributing to the glucose accumulation at late date developmental stages. A negative correlation is also observed between pda-nov-miR243 and its target serine O-acetyltransferase (XM_008791140), an enzyme involved in carbon and sulfur metabolisms (Fig. 6D).

In addition, we identified two well-known transcription factors, AP2 and SBP [11], which showed coordination with the expression of their corresponding miRNAs. AP2 plays a major role in the regulation of tomato fruit ripening, and its expression has been reported to be negatively regulated by corresponding miRNA [32,33]. In our data, two novel miRNAs, pda-nov-miR049 and pda-nov-miR111, all showed negative correlation with AP2 expression; the AP2 expression was fluctuated from 75DAF to 120DAF (Fig. 6E). SBP transcription factor has been shown to inhibit tomato fruit ripening [12,34]. In *P. dactylifera*, SBP expression is negatively correlated with all its corresponding miRNAs, and its expression exhibited a gradual decrease from 75DAF to 120DAF (Fig. 6F).

Although a miRNA and its target mRNAs are expected to show negative correlation, many studies have reported that this may not always be the case [9,35]. In our data, we observed examples of such cases. For instance, the expression of pda-nov-miR044, pda-nov-miR166, pda-nov-miR168, pda-nov-miR178, pda-nov-miR197, pda-nov-miR213, and pda-nov-miR228 is positively correlated with their targets (Supplementary Fig. S2B; Supplementary Table S1), which may be due to a similar feedback regulation between co-expressed miRNAs and their targets as previously reported in *Arabidopsis* and rice [9,35]. However, the concrete molecular mechanisms of these examples remain to be demonstrated experimentally in details.

3. Discussion

MiRNAs as a new class of regulatory factors have attracted much attention in recent studies, especially in plants, and many studies have been devoted to investigating the miRNA regulatory metabolisms underlying different plant developmental processes, such as grain development in rice and fruit development in tomato. There have been 5943 miRNAs belonging to 67 plant species reported in miRBase (release 19.0), and it has not accommodated a relatively thorough collection from *P. dactylifera*. In this study, we took a combined approach of bioinformatic prediction and high-throughput sequencing to characterize conserved and novel miRNAs and to categorize their expression pattern involved in *P. dactylifera* fruit development and ripening. The transcriptomic study of the same time-series multi-staged samples during fruit development allows a direct correlation analysis between miRNAs and their corresponding targets. Our study revealed miRNA dynamics for *P. dactylifera* fruit development and ripening, paving a way for further studies on gene regulation and control mechanisms.

3.1. Identification of conserved miRNAs in P. dactylifera

Conserved miRNA families can be computationally predicted due to their high conservation and widespread presence among plant species [17]. In our study, 238 conserved miRNAs belonging to 54 miRNA families were predicted based on computational methods, much more than what has been reported [36]. In the previous report, 21 conserved miRNA families from miRBase (release 19.0) supported by an early *P. dactylifera* genome assembly [15] were used to identify 81 individual miRNAs of 18 miRNA families. There have been some discrepancies between the two datasets. For instance, we did not identify miR168 and miR482 families in our results, and instead, members of the other 16 shared miRNA families were found much more in our results (Supplementary Table S2). The discrepancy may be due to several possible reasons. On one hand, the KACST genome reference we used in this study is 605.4 Mb in size [4], much larger than the Qatar genome assembly that is 382 Mb in size. Obviously, the former has

Table 2A
Functional analyses of miRNA target genes.

GO terms significantly enriched in the target genes						
Term	Function	P-value	Group count	Control count	Total count in group	Total count in control
GO: 0032502	Developmental process	0.000407	16	233	483	22,220
GO: 0030154	Cell differentiation and cell death	0.00083	16	213	483	22,220
GO: 0050794	Regulation of cellular process	0.000919	56	1497	483	22,220
GO: 0050896	Response to stimulus	0.00116	21	395	483	22,220
GO: 0004553	Catalytic activity	0.0023	6	32	483	22,220
GO: 0005515	Protein binding	0.00536	103	3401	483	22,220

Table 2B
Functional analyses of miRNA target genes.

KEGG pathways significantly enriched in the target genes						
Pathway	Function	P-value	Group count	Control count	Total count in group	Total count in control
2010	ABC transporters	0.000658	3	23	737	41,660
0920	Sulfur metabolism	0.001419	3	28	737	41,660
0500	Starch and sucrose metabolism	0.004244	7	144	737	41,660
3040	Spliceosome	0.011883	7	172	737	41,660
3015	mRNA surveillance pathway	0.01483	5	112	737	41,660

more material for an in-depth discovery. On the other hand, we checked the criteria applied to determine conserved miRNA candidates in the previous report and found that the maximum mismatch allowed between the predicted candidate miRNAs and the known miRNA reference sequences was 4 nucleotides. In contrast, a stricter criterion (the 3 nucleotide maximum mismatch) was used in our study, leading to the null result of miR168 and miR482 discovery.

Among the 54 miRNA families we identified, 12 miRNA families are first-identified in *P. dactylifera*, which had been known to be unique to dicots previously. More specially, miR477, which is only predicted in *P. patens* and another 7 dicots, and miR536, which is only identified in *P. patens* and *S. moellendorffii*, are now identified in *P. dactylifera*, which are the first-time reported miRNA belonging to these families in monocots. It is noteworthy that miR414, which has been found in 3 plant species, *Arabidopsis thaliana*, *Oryza sativa*, and *P. patens*, belonging to dicots, monocots, and mosses, respectively, is also found in *P. dactylifera*. These results suggest the complexity and specificity of the *Pd* miRNAs in the context of species evolution, where the gain-and-loss of particular miRNAs can be related to their conserved functions.

*3.2. The finding of miRNA*s with an abundance higher than their corresponding miRNAs*

Under normal cellular condition in animals or plants, a miRNA: miRNA* duplex disassembles after being transported out of the nucleus to the cytosol, and the mature miRNA is incorporated into the RISC (RNA-induced silencing complexes) for target recognition, whereas the miRNA* strands are thought to be degraded rapidly [7]. Therefore, the miRNA levels are usually higher than their corresponding miRNA*. However, a recent study in plants proved that both miRNA and miRNA* can silence different targets [37,38], and ample studies have shown that miRNA*s can have a higher expression level than their miRNAs in plants [9,35,39]. Here, we also encountered similar situations; in our datasets, six miRNA*s (miR160a*, miR164b*, miR169b*, miR395f*, miR396a*, and miR5225*) are expressed at a higher level than their corresponding miRNAs (Supplementary Fig. S3; Supplementary Table S3), suggesting that these miRNA*s may be genuine products of their pre-miRNAs or that both the miRNAs and their miRNA*s may play important regulatory roles in fruit development and ripening. To analyze the function of these highly expressed

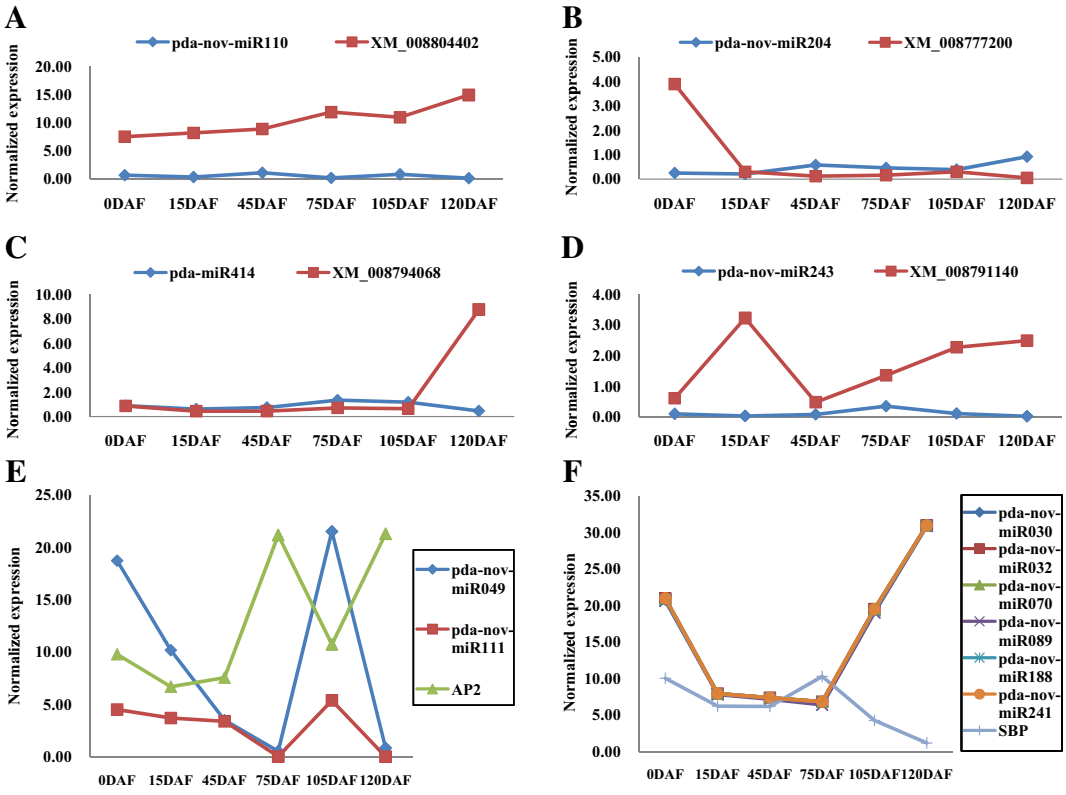

Fig. 6. Expression profiling of four sugar metabolism-related target genes and two transcription factors with their corresponding miRNAs in *P. dactylifera* fruit development. A–D, The expression of four sugar metabolism-related targets and their corresponding miRNAs. E, The expression of AP2 and its corresponding miRNAs. F, The expression of SBP and its corresponding miRNAs. The y-axis indicates normalized expression: TPM for miRNA and RPKM for their corresponding target genes.

miRNA*s, we used psRNATarget with default parameters to predict targets of these miRNA*s, and obtained 31 targets of 4 miRNA*s (Supplementary Table S9 in Ref. [20]). The majority of potential targets predicted for miR160a* and miR5225* are transposon element-encoded transcripts, which have been also predicted as potential miRNA targets in rice [35]; whereas miR395f* was predicted to target some conserved transcription factors including GRAS, MYB, HLH, and F-box; these results suggest conserved functions of some miRNA*s among plant lineages and species.

3.3. Conserved and non-conserved targets of miRNAs in P. dactylifera

The identification of miRNA target genes is essential for functional analysis. To identify the function of *Pd* miRNAs, we used psRNATarget, a web server, for genome-wide analysis. Our results confirmed, as have been shown in other studies, that many miRNAs and their targets are conserved among plant species from mosses to flowering plants. For example, each of miR156, 159, 160, 164, 166, 171, and 172 is predicted to target one exclusive transcription factor family in *Arabidopsis*, rice, grapevine, soybean, tomato, and mosses, and the functional conservation suggests that the regulation of target gene expressions by their corresponding miRNAs arose early in plant evolution. We also predicted some new targets in this study; for example, for the conserved miR167, we predicted TPS:trehalose 6-phosphate synthase/phosphatase as its new target, which has not yet been reported in other plant species.

In addition, 874 targets for 199 novel miRNAs are also predicted in our study. Comparing the targets of conserved miRNAs, we found that the novel *Pd* miRNA targets exhibit more diverse functions, not limited to TFs, although some of them still do. For example, the novel miRNAs 049, 111, 200, and 212, similar to the conserved miR172, target AP2; the novel miRNAs 052, 057, 150, 218, and 219, similar to the conserved miR171, target GRAS; the novel miRNAs 030, 032, 070, 089, 188, and 241, similar to the conserved miR156, target SBP; the novel miRNAs 005 and 160, similar to the conserved miR164, target NAC; and the novel miRNAs 065 and 258, similar to the conserved miR159, target MYB. These results suggest that some novel miRNAs, or a fraction of species-specific miRNAs in general, together with their conserved counterparts, may play regulatory roles in a synergetic way to control the expression of the conserved targets.

3.4. Roles of miRNAs in P. dactylifera fruit development and ripening

The *P. dactylifera* fruit development and ripening is a complex process, including a series of highly coordinated molecular and cellular events associated with unique transcriptomic profiles, and studies have shown gene regulations in complicated metabolic activities, such as the synthesis of starch and other related sugars [3,4]. In our data, many of the predicted target genes are associated with function-conserved transcription factors involved in fruit development.

Phytohormones are important for plant growth and development [40]. Studies on tomato fruit development have shown that auxin is a key regulator of tomato fruit set and development [11]. In our data, miR160 has the potential to target five auxin response factors (ARFs), and albeit expressed at a particularly low level during *P. dactylifera* fruit development; it is closely related to the high transcriptional activity of ARFs [41,42].

Date development and its ripening are highly relevant to sugar accumulation and polysaccharide-to-monosaccharide conversions, including a series of events, such as carbon fixation, starch hydrolysis–synthesis, and glucose–fructose accumulation. Our KEGG analysis defined miRNAs that play roles in carbon metabolism, starch–sucrose metabolism, fructose–mannose metabolism, glycolysis, and citrate cycle. For example, a novel *Pd* miRNA, pda-nov-miR110, is predicted to target sucrose–phosphate synthase (SPS), a key enzyme in sucrose synthesis, showing a higher expression level at the late stages, and the result agrees with a previous study in *P. dactylifera* [4]. Three genes encoding β-amylases are predicted to be targeted by eight novel *Pd* miRNAs in our study, which are a group of key enzymes involved in starch–sucrose metabolism. One target of pda-nov-miR215 is predicted to be pyrophosphate:fructose-6-phosphate 1-phosphotransferase (PFP), which catalyzes the reversible reaction of fructose 6-phosphate (Fru 6-P) and pyrophosphate (PPi) to fructose 1,6-bisphosphate (Fru 1,6-P_2) and inorganic phosphate (Pi) [43]; it is gradually being up-regulated in fruit development.

4. Conclusion

Our combined approach based on computational and experimental methodology has worked well in the genome-wide identification of miRNAs in *P. dactylifera*, revealing dynamic regulation of miRNAs between individual developmental stages of *P. dactylifera* fruit development and ripening. We contribute basic information and expression patterns of 514 miRNAs to the *Pd* miRNA repertoire, especially the FDA miRNAs, for further study on fruit development and ripening. Our miRNA–target correlation analysis is also affirmative, where confirmation of the common mechanisms and suggestive novel mechanisms are both assured. Our datasets and initial analysis provide essential data and information for deciphering the complex regulatory network governing *P. dactylifera* fruit development and ripening.

5. Materials and methods

5.1. Plant materials

Samples were collected from *P. dactylifera* trees grown in Al-Kharj (24°08′54″N, 47°18′18″E), Saudi Arabia. Fruits of *Khalas* cultivar at different developmental stages, including 0 DAF (days after fertilization), 15 DAF, 45 DAF, 75 DAF, 105 DAF, and 120 DAF, were collected, washed with distilled water, frozen in liquid nitrogen, and stored at −80 °C for further use.

5.2. Computational prediction of conserved miRNAs

In Fig. 1 in Ref. [20], we summarized the procedure for searching conserved *Pd* miRNA homologues in miRBase (release 19.0). Briefly, all the known mature plant miRNA sequences were subjected to a BLAST search with *P. dactylifera* genome sequences using BLASTn [44] with the following parameter settings: expected values were set at 1000; the default word-match size between query and database sequences was set at 7; and the numbers of descriptions and alignments were raised to 1000. Sequences with no more than 3 mismatches to known miRNAs and a length between 18 and 25 nt were kept for further analysis. After removing repeat and protein-encoding sequences, the remaining sequences were subjected to do secondary structure folding. The candidate sequences along with 100-nt upstream and downstream flanking sequences were ran through RNAFold [45]. MFE (minimal folding free energy), AMFE (adjusted MFE), MFEI (minimal folding free energy index), length of sequence, percent of nucleotide (A, U, G, and C), A + U content, G + C content, and number of base pairs were calculated as previously reported. The following criteria were used to determine the candidates of potential miRNA or pre-miRNA: (1) the predicted mature miRNA was allowed to have 0–3 nucleotide mismatches in sequence with the best matched known plant mature miRNA and sequence length was between 18 and 25 nucleotides; (2) the RNA sequence can fold into an appropriate stem–loop hairpin secondary structure; (3) the predicted mature miRNA is located in one arm of the hairpin structure; (4) the predicted mature miRNA sequence had no more than 6 mismatches with its opposite miRNA* sequence in the other arm; (5) the predicted mature miRNA sequence had at least 16 base pairs involved in Watson–Crick or G/U base pairing with the miRNA* sequence in the stem portion of the hairpin; (6) no loop or break in the miRNA or miRNA* sequence; (7) the MFE allowed for a

candidate pre-miRNA was −25 kcal/mol; (8) the A + U content of the candidate pre-miRNA was within 30%–70%; (9) the predicted secondary structure had a higher MFEI over 0.85.

5.3. Library construction, sequencing, and data analysis

Total RNAs were isolated from fruits using the CTAB (2%) method [46], then RNase-free DNaseI was used to remove the DNA contamination for 15 min at 37 °C. The RNA was quantified using NanoDrop™ Spectrophotomer ND-8000 and Qubit, and then checked on 1% agarose gel. Small RNA fractions (less than 40 nt) were isolated (FlashPAGE™ Fractionator system, Ambion) and cloned (SOLiD™ Small RNA Expression Kit, Applied Biosystems) according to the manufacturer's protocols. Sequencing data were acquired with SOLiD™ System 4.0.

The low quality reads were processed (RNA2MAP, RNA_pipeline_0.4.0) as follows. First, sequences that mapped to a dataset containing *P. dactylifera* rRNA, tRNA, snoRNA, mRNA and Rfam RNA families (excluding known miRNA sequences) were filtered. Second, sequences that could be perfectly mapped onto miRNA precursors and mature miRNAs of our previously identified conserved miRNAs were identified as known miRNAs. Third, the remaining unmapped reads were aligned to the genome sequence in attempting to identify novel miRNAs involved in fruit development and ripening in *P. dactylifera*. The reads that mapped to the genome sequence were considered as novel *P. dactylifera* miRNA candidates.

The candidate sequences along with 100-nt upstream and downstream flanking sequences were analyzed by running them through RNAFold. The basic criteria applied to predict the novel miRNAs are the same as previously used. Furthermore, only the candidates that satisfied the following criteria were considered as novel *Pd* miRNAs. First, to decrease background noise, small RNAs with a read number of less than 20 are filtered out. Second, small RNAs mapped to more than 10 loci in the *P. dactylifera* genome are excluded. Third, the candidates possessing their corresponding miRNA* sequences in at least one of the libraries or the mature sequence are detected from more than two libraries [9].

5.4. Differential expression and quantitative real time RT-PCR analysis of miRNAs

To investigate the expression patterns of miRNAs among the six different stages of fruit development and ripening, the read count of each expressed miRNA was normalized to transcripts per million (TPM) using the following formula: normalized expression = actual read count / total read count × 1,000,000 [39]. After normalization, the expression was set to 0.01 for miRNAs that are not expressed. The differentially expressed miRNAs between two adjacent developmental stages were identified based on the following two criteria: $p \leq 0.001$ and fold change ≥ 2 using "DEGseq" [47], an R package. STEM (short time-series expression miner) was used for clustering and visualizing possible profiles of differentially expressed (DE) miRNAs [29]. The maximum unit change in model profiles between time points was adjusted to 1 and the maximum number of model profiles to 50. STEM was run by using the "log normalized data" option with all other settings set to the default.

Total RNA was extracted from fruit at each developmental stage. RNase-free DNase I (NEB, USA) was used to remove DNA contamination. Complementary DNA was synthesized from total RNA using miRNA-specific stem–loop primers obtained from a commercial service (Sangon, Shanghai) in a 15-μl reaction volume with the SuperScript™ II reverse transcriptase (Invitrogen). The reaction was incubated for 30 min at 16 °C, followed by 60 cycles of pulsed RT at 30 °C for 30 s, 42 °C for 30 s and 50 °C for 1 s to increase the sensitivity of miRNA detection [48], then followed by 85 °C for 5 min to inactivate the reaction. A 25-μl PCR reaction containing SYBR green mix (TIANGEN) was incubated in a 96-well optical plate at 95 °C for 15 min, followed by 40 cycles of 95 °C for 15 s, 60 °C for 15 s and 72 °C for 32 s, in a ABI 7500 Sequence Detection system. Melting curve analysis was carried out for each PCR product to avoid non-specific amplification. The comparative Ct method was used to calculate the fold change. All the stem–loop RT primers were designed according to Chen et al. [49] and listed in Supplementary Table S6 in Ref. [20].

5.5. MiRNA target gene prediction and function analysis

The target genes of miRNAs were predicted according to their perfect or nearly perfect complementarity between miRNAs and the target genes through a homology algorithm. The potential targets of *Pd* miRNAs were predicted by using the psRNATarget web server with the default parameters [50]. Gene ontology annotation and KEGG pathways were applied to analyze miRNA targets based on their web tools.

5.6. Data availability

The dataset was deposited to the NCBI database and the BioProject ID is PRJNA266266.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2015.01.004.

Author contributions

CX, WL, QL, ISM, SH and JY designed the experiments. CX, XZ and FL generated the sequencing data. CX and AA performed the RT-PCR experiments. WL, QL, CX, GZ, PL and HM analyzed data. HAJ, SH, JY supplied the reagents and materials. CX, WL, PC and JY wrote the manuscript. All authors read and accepted the final manuscript.

Acknowledgment

We acknowledge my colleagues Saeed Mohammed Al-Qahtani, Majed Abdullah Majrashi, and Abdullah Ganim Al-Salem, who kindly assisted us to collect the samples, and Dr. Yimei Cai for her help in the experiment. The work is supported by KACST grant AT428-29 from the Joint Center for Genomics Research (JCGR), King Abdulaziz City for Science and Technology (KACST), Kingdom of Saudi Arabia. Technical supports were provided by the CAS Key Laboratory of Genome Science and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, the People's Republic of China.

References

[1] J.F. Morton, Fruits of warm climates, Purdue University Center for New Crops and Plants, 1987. 5–11 (p. Date).
[2] W. Al-Shahib, R.J. Marshall, The fruit of the date palm: its possible use as the best food for the future? Int. J. Food Sci. Nutr. 54 (4) (2003) 247–259.
[3] Y. Yin, et al., High-throughput sequencing-based gene profiling on multi-staged fruit development of date palm (*Phoenix dactylifera*, L.), Plant Mol. Biol. 78 (6) (2012) 617–626.
[4] I.S. Al-Mssallem, et al., Genome sequence of the date palm *Phoenix dactylifera* L. Nat. Commun. 4 (2013) 2274.
[5] F. Bourgis, et al., Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning, Proc. Natl. Acad. Sci. 108 (30) (2011) 12527–12532.
[6] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2) (2004) 281–297.
[7] O. Voinnet, Origin, biogenesis, and activity of plant microRNAs, Cell 136 (4) (2009) 669–687.
[8] R.W. Carthew, E.J. Sontheimer, Origins and mechanisms of miRNAs and siRNAs, Cell 136 (4) (2009) 642–655.
[9] T. Peng, et al., Characterization and expression patterns of microRNAs involved in rice grain filling, PLoS ONE 8 (1) (2013).
[10] R. Sunkar, Y.F. Li, G. Jagadeeswaran, Functions of microRNAs in plant stress responses, Trends Plant Sci. 17 (4) (2012) 196–203.
[11] R. Karlova, et al., Identification of microRNA targets in tomato fruit development using high-throughput sequencing and degradome analysis, J. Exp. Bot. 64 (7) (2013) 1863–1878.

[12] S. Moxon, et al., Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening, Genome Res. 18 (10) (2008) 1602–1609.
[13] B.J. Reinhart, et al., MicroRNAs in plants, Genes Dev. 16 (13) (2002) 1616–1626.
[14] A. Kozomara, S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data, Nucleic Acids Res. 39 (Database issue) (2011) D152–D157.
[15] E.K. Al-Dous, et al., De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*), Nat. Biotechnol. 29 (6) (2011) 521–527.
[16] B.H. Zhang, et al., Evidence that miRNAs are different from other RNAs, Cell. Mol. Life Sci. 63 (2) (2006) 246–254.
[17] M.W. Jones-Rhoades, Conservation and divergence in plant microRNAs, Plant Mol. Biol. 80 (1) (2012) 3–16.
[18] F. Thiebaut, et al., Computational identification and analysis of novel sugarcane microRNAs, BMC Genomics 13 (2012) 290.
[19] A. Katiyar, et al., Identification of miRNAs in sorghum by using bioinformatics approach, Plant Signal. Behav. 7 (2) (2012) 246–259.
[20] C. Xin, et al., Genome-wide identification and expression profiles of MicroRNA during multi-staged date palm (*Phoenix dactylifera* L.) fruit development, *Data in brief* (2015) (submitted for public).
[21] O. Patanun, et al., Computational identification of microRNAs and their targets in cassava (*Manihot esculenta* Crantz), Mol. Biotechnol. 53 (3) (2013) 257–269.
[22] T.P. Frazier, et al., Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum*), Planta 232 (6) (2010) 1289–1308.
[23] V. Dhandapani, et al., Identification of potential microRNAs and their targets in *Brassica rapa* L. Mol. Cells 32 (1) (2011) 21–37.
[24] M.W. Rhoades, et al., Prediction of plant microRNA targets, Cell 110 (4) (2002) 513–520.
[25] L. Zhou, et al., Genome-wide identification and analysis of drought-responsive microRNAs in *Oryza sativa*, J. Exp. Bot. 61 (15) (2010) 4157–4168.
[26] L.S. Liu, M.J. White, T.H. MacRae, Transcription factors and their genes in higher plants — functional domains, evolution and regulation, Eur. J. Biochem. 262 (2) (1999) 247–257.
[27] E. Bonnet, et al., Detection of 91 potential in plant conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes, Proc. Natl. Acad. Sci. U. S. A. 101 (31) (2004) 11511–11516.
[28] B.H. Zhang, X.P. Pan, E.J. Stellwag, Identification of soybean microRNAs and their targets, Planta 229 (1) (2008) 161–182.
[29] J. Ernst, Z. Bar-Joseph, STEM: a tool for the analysis of short time series gene expression data, BMC Bioinf. 7 (2006) 191.
[30] L.J. Xue, J.J. Zhang, H.W. Xue, Characterization and expression profiles of miRNAs in rice seeds, Nucleic Acids Res. 37 (3) (2009) 916–930.
[31] V. Veena, et al., Isolation and characterization of β-glucosidase producing bacteria from different sources, Afr. J. Biotechnol. 10 (66) (2013) 14891–14906.
[32] M.Y. Chung, et al., A tomato (*Solanum lycopersicum*) APETALA2/ERF gene, SlAP2a, is a negative regulator of fruit ripening, Plant J. 64 (6) (2010) 936–947.
[33] R. Karlova, et al., Transcriptome and metabolite profiling show that APETALA2a is a major regulator of tomato fruit ripening, Plant Cell Online 23 (3) (2011) 923–941.
[34] K. Manning, et al., A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening, Nat. Genet. 38 (8) (2006) 948–952.
[35] R. Yi, et al., Identification and expression analysis of microRNAs at the grain filling stage in rice (*Oryza sativa* L.) via deep sequencing, PLoS ONE 8 (3) (2013).
[36] Y. Xiao, et al., Characterization and evolution of conserved microRNA through duplication events in date palm (*Phoenix dactylifera*), PLoS ONE 8 (8) (2013).
[37] E.A. Devers, et al., Stars and symbiosis: microRNA- and microRNA*-mediated transcript cleavage involved in arbuscular mycorrhizal symbiosis, Plant Physiol. 156 (4) (2011) 1990–2010.
[38] X. Zhang, et al., *Arabidopsis* Argonaute 2 regulates innate immunity via miRNA393 (*)-mediated silencing of a Golgi-localized SNARE gene, MEMB12, Mol. Cell 42 (3) (2011) 356–366.
[39] T. Peng, et al., Differential expression of the microRNAs in superior and inferior spikelets in rice (*Oryza sativa*), J. Exp. Bot. 62 (14) (2011) 4943–4954.
[40] A. Santner, M. Estelle, Recent advances and emerging trends in plant hormone signalling, Nature 459 (7250) (2009) 1071–1078.
[41] S.B. Tiwari, G. Hagen, T. Guilfoyle, The roles of auxin response factor domains in auxin-responsive transcription, Plant Cell 15 (2) (2003) 533–543.
[42] T.J. Guilfoyle, G. Hagen, Auxin response factors, Curr. Opin. Plant Biol. 10 (5) (2007) 453–460.
[43] J.H. Groenewald, F.C. Botha, Down-regulation of pyrophosphate: fructose 6-phosphate 1-phosphotransferase (PFP) activity in sugarcane enhances sucrose accumulation in immature internodes, Transgenic Res. 17 (1) (2008) 85–92.
[44] S.F. Altschul, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (17) (1997) 3389–3402.
[45] R.B. Denman, Using RNAfold to predict the activity of small catalytic RNAs, Biotechniques 15 (6) (1993) (p. 1090–&).
[46] S. Chang, J. Puryear, J. Cairney, A simple and efficient method for isolating RNA from pine trees, Plant Mol. Biol. Report. 11 (2) (1993) 113–116.
[47] L. Wang, et al., DEGseq: an R package for identifying differentially expressed genes from RNA-seq data, Bioinformatics 26 (1) (2010) 136–138.
[48] E. Varkonyi-Gasic, et al., Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs, Plant Methods 3 (2007) 12.
[49] C. Chen, et al., Real-time quantification of microRNAs by stem–loop RT-PCR, Nucleic Acids Res. 33 (20) (2005) e179.
[50] X. Dai, P.X. Zhao, psRNATarget: a plant small RNA target analysis server, Nucleic Acids Res. 39 (Web Server issue) (2011) W155–W159.