



# Computational screening of miRNAs and their targets in saffron (*Crocus sativus* L.) by transcriptome mining

Ayat Taheri-Dehkordi<sup>1</sup> · Roohangiz Naderi<sup>1</sup> · Federico Martinelli<sup>2</sup> · Seyed Alireza Salami<sup>1</sup>

Received: 6 September 2021 / Accepted: 18 October 2021 / Published online: 9 November 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

**Main conclusion** A robust workflow for the identification of miRNAs and their targets in saffron was developed. MicroRNA-mediated gene regulation in saffron is potentially involved in several biological processes, including the biosynthesis of highly valuable apocarotenoids.

**Abstract** Saffron (*Crocus sativus* L.) is the most expensive spice in the world and a major source of apocarotenoids. Even though miRNAs (20–24 nt non-coding small RNAs) are important regulators of gene expression at transcriptional and post-transcriptional levels, their role in saffron has not been thoroughly investigated. As a result, a workflow for computational identification of miRNAs and their targets can be useful to uncover the regulatory networks underlying biological processes in this valuable plant. The efficiency of several assembly tools such as Trans-ABYSS, Trinity, Bridger, rnaSPAdes, and EvidentialGene was evaluated based on both reference-based and reference-free metrics using transcriptome data. A reliable workflow for computational identification of miRNAs and their targets in saffron was described. The EvidentialGene was found to be the most efficient de novo transcriptome assembler for saffron as a complex triploid model, followed by the Trinity. In total, 66 miRNAs from 19 different families that target 2880 genes, including several transcription factors involved in the flowering transition, were identified. Three of the identified targets were involved in the terpenoids backbone biosynthesis. CsCCD and CsUGT genes involved in the apocarotenoids biosynthetic pathway were targeted by csa-miR156g and csa-miR156b-3p, revealing a unique post-transcriptional regulation dynamic in saffron. The identified miRNAs and their targets add to our understanding of the many biological roles of miRNAs in saffron and shed new light on the control of the apocarotenoid biosynthetic pathway in this valuable plant.

**Keywords** Apocarotenoids · Bioinformatics analysis · Flowering transition · Metabolic engineering · Transcriptome assembly

## Introduction

*Crocus sativus* L., also known as saffron, is the most precious spice in the world (Chib et al. 2020). This species spreads from central Europe to western China (Harpke et al. 2013). Iran is known as the world's largest producer of

saffron, and in recent years, more attention has been dedicated to this valuable plant due to its advantages over other crops, such as higher price and low water needs (Vahedi et al. 2018; Taheri-Dehkordi et al. 2020). The color, flavor, and aroma of saffron are attributed to particular apocarotenoids named crocin, picrocrocin, and safranal (Tarantilis et al. 1995). Increasing demand for saffron's active ingredients, on the one hand, and its sterile nature ( $2n = 3x = 24$ ), which renders traditional breeding methods ineffective for increasing the amounts of its active ingredients, on the other, have prompted researchers to turn their attention to the use of molecular breeding methods (Taheri-Dehkordi et al. 2020). Metabolic engineering is an approach for increasing apocarotenoids in saffron. Using this approach necessitates a

Communicated by Anastasios Melis.

✉ Seyed Alireza Salami  
asalami@ut.ac.ir

<sup>1</sup> Department of Horticultural Science, Faculty of Agricultural Science and Engineering, University of Tehran, Karaj, Iran

<sup>2</sup> Department of Biology, University of Florence, Sesto Fiorentino, Italy

thorough grasp of the regulatory mechanisms governing the apocarotenoid biosynthetic pathway in this plant.

MicroRNAs (miRNAs) are endogenous small non-coding RNAs (20–24 nucleotides in length) that are thought to be the most important post-transcriptional regulators of gene expression (Bartel 2018). These molecules are a class of important cellular organic molecules that finely modulate gene expression in relation to environmental signals although their role in plant species such as saffron has been scarcely investigated. From their discovery, studies have focused on the understanding of their function at the post-transcriptional regulation stage in response to developmental and environmental conditions. Mature miRNAs are obtained from longer non-coding pre-miRNAs, through a process of multiple cleavage steps, where a complex system of enzymes is involved (Kurihara and Watanabe 2010). miRNAs are key effectors in gene silencing via hybridization with their mRNA targets, having a complementary sequence (Jeong et al. 2010). They are widely known in all plant genomes and divided into different families with 1–32 loci. Typically, members of the same family, miRNAs are characterized by identical or very similar mature sequences. Approximately 20 families are known in the plant kingdom. Among them, some miRNAs are in primitive land plants while others have been generated through recent evolutionary events (Sunkar et al. 2008).

In plants, transcription of miRNA genes (MIR) by RNA Polymerase II (Pol II) results in primary miRNA (pri-miRNA). Pri-miRNAs then fold back to create hairpin structures that are recognized by a family of enzymes known as Dicer-like (DCL). DCL enzymes have different members in different plant species, for example, *Arabidopsis thaliana* and grape have four DCL members (Liu et al. 2009). DCL1 is the member of the family responsible for the cleavage of pri-miRNA to precursor miRNA (pre-miRNA). Further cleavage of pre-miRNA by DCL1 leads to the release of a miRNA/miRNA\* duplex (Axtell et al. 2011). The addition of a methyl group to the 3' end of the miRNA/miRNA\* duplex by Hua Enhancer 1 (HEN1) methyltransferase results in its stability (Yu 2005). After the miRNA/miRNA\* duplex is separated in the cytoplasm, the guide miRNA binds to Argonaute (AGO) proteins and integrates into the RNA-induced silencing complex (RISC) (Budak and Akpinar 2015). Finally, the fully functional RISC complex can regulate gene expression via either target cleavage or translational inhibition (Wu et al. 2009).

miRNAs have various roles in plant growth and development such as regulation of terpenoid (Chen et al. 2020) and alkaloids (Verma et al. 2020) biosynthesis, shoot apical meristem (SAM), leaf development, and the flowering transition (D'Ario et al. 2017). The number of newly characterized miRNA is enhancing (Fileccia et al. 2017, 2019), most of them in the model plant *A. thaliana* (Kozomara

and Griffiths-Jones 2014). However, their role is still far to be elucidated although many of their families have been shown to have a role in plant environmental stress responses (Martinelli et al. 2018) such as miR169, a miRNA highly conserved across the plant kingdom, linked to plant abiotic stress responses (Li et al. 2008). Several plant miRNAs have been studied but the complex networks of regulation mechanisms driven by miRNA are far to be fully elucidated (Budak and Akpinar 2015). Recently a miRNAome of durum wheat has been characterized for different plant organs (Fileccia et al. 2017). The majority of miRNA targets are transcription factors (Chow et al. 2016), and they work together to form a regulatory network that regulates gene expression in a variety of biological processes (Samad et al. 2017). Previous works dealing with identifying miRNAs and their targets were performed in different species (Xin et al. 2010; Tang et al. 2012). Tissue-specific analysis of miRNAs was conducted in wheat (Liu et al. 2015; Agharbaoui et al. 2015). Few studies have been performed dealing with the annotation and chromosome mapping of key functional miRNAs. A strategy for minimizing the annotation of misidentifying TE sequences as miRNAs was proposed (Lucas and Budak 2012). Two new scripts, BSumirPredictor<sup>^</sup> and BSumirLocator<sup>^</sup>, have been generated for enhancing miRNA prediction and identification, for genomes characterized by high repetitive genomic sequences (Alptekin et al. 2017). These methods allow us to clarify how miRNA precursors are distributed in the plant genome and how they are present in the transcriptome in relation to the presence of transposon elements. Identification of miRNAs in plants is mostly accomplished through two methods: small RNA sequencing and computational (in silico) analysis. With the advent of Next-Generation Sequencing (NGS) technology, a massive quantity of genomic and transcriptomic data has been generated, which has greatly aided in the computational identification of miRNAs in plants, particularly those lacking a sequenced genome. Computational techniques were successfully identified potential miRNAs in a variety of plants, including *Hypericum* spp. (Petijová et al. 2020), *Citrullus lanatus* (Zakeel et al. 2019), and *Arundo donax* L. (Jike et al. 2018). Furthermore, the availability of databases for computationally identified miRNAs, such as miRBase (Kozomara et al. 2019), has contributed to the identification of miRNAs in plants. Furthermore, several reliable criteria for distinguishing miRNAs from other small RNAs and avoiding false-positive outcomes through miRNA identification have been described (Kurihara and Watanabe 2010).

Despite the importance of miRNAs in regulating several processes in plants, the significance of these molecules in saffron has yet to be fully explored. Thus, the primary goal of this study was to provide a computational strategy for identifying miRNAs and their targets using saffron transcriptome data. The efficiency of various de novo transcriptome

assembly tools was tested using both reference-based and reference-free criteria. To decrease the possibility of false-positive results, a precise and reliable workflow for identifying miRNAs, their targets, and their functional annotation was developed. These findings will help us better understand the potential regulatory mechanism (s) involved in apocarotenoids biosynthesis as well as the flowering transition in saffron.

## Materials and methods

### Data acquisition and preprocessing

The meta-analysis was carried out using raw RNA-seq data from different saffron floral parts (i.e. stigma, anther, tepal, and stamen) (Table 1). Initially, the SRA files were converted to FASTQ format using the SRA Toolkit's (2.10.7) `fasterq-dump` command (Leinonen et al. 2010). FastQC (version 0.11.9) was used to check the data quality (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). After initial quality assessment, barcode sequences, short (less than 50 nucleotides in length), and low-quality ( $Q < 20$ ) sequences

and adapters were trimmed using Trimmomatic (Version 0.39) (Bolger et al. 2014).

### De novo transcriptome assembly

Four distinct assemblers with varying  $k$ -mer indexes, as well as EvidentialGene (18 May 07; <http://arthropods.eugenesis.org/EvidentialGene>), a tool for merging multiple assemblies, were used to create the best assembly. Trans-ABYSS (V2.0.1) (Robertson et al. 2010) and Trinity (v2.10.0) (Grabherr et al. 2011) assemblers are based on *de Bruijn* graphs and are designed to deal with RNA-seq data. Bridger (r2014-12-01) (Chang et al. 2015), on the other hand, relies on splicing graphs. This tool provides a novel framework for de novo transcriptome assembly and bridges the gap between the techniques used in the Cufflinks and Trinity pipelines to overcome Trinity restrictions. rnaSPAdes (3.14.1) (Bushmanova et al. 2019), a genome-based assembler based on *de Bruijn* graphs, was also investigated. The purpose was to evaluate the performance of a single-cell genome assembler to that of existing assemblers tuned for transcriptome data. It should be mentioned that RNA-Seq mode, which uses two  $k$ -mers, is available starting with version 3.9.0 of this software.

**Table 1** The four *Crocus sativus* RNA-Seq data sets used for transcriptome assembly

No.	Study	Run	Tissue	Protocol	Reads		Source
					No.	Length	
1	PRJNA562470	SRR10028148	Tepal	PE	8.6	250	
		SRR10028149			8.7	250	
		SRR10028150	Stigma		14.4	250	
		SRR10028151			14.5	250	
		SRR10028156	Anther		15.3	250	
		SRR10028157			15.6	250	
2	PRJNA400472	SRR5985553	Stigma	PE	23.0	125	(Tan et al. 2019)
		SRR5985554			19.6	125	
		SRR5985555			16.8	125	
		SRR5985556			25.0	125	
		SRR5985557			23.0	125	
		SRR5985558			15.3	125	
		SRR5985559			17.8	125	
		SRR5985560			20.3	125	
3	PRJNA273002	SRR1767300	Tepal	PE	24.1	101	(Jain et al. 2016)
		SRR1767301	Stamen		19.3	101	
		SRR1767302	Stigma		22.9	101	
4	PRJNA277895	SRR1909702	Whole flower without stigma	PE	37.7	72	(Baba et al. 2015)
		SRR1910564	Whole flower without stigma		37.7	72	
		SRR1909704	Stigma		29.5	72	
		SRR1910567	Stigma		29.5	72	

Read numbers are given in millions

## Evaluation of different assemblies

Various metrics were analyzed to compare different assemblies and their efficiency. These metrics are divided into two basic categories: 1- Reference-based metrics: the number of reads mapped back to the transcriptome (RMBT), Ex90N50, and the average percentage of open reading frames (ORF) 2- Reference-free metrics: transcripts larger than 1000 bp, BUSCO, and full-length protein-coding transcript reconstruction.

### Reads mapped back to the transcriptome (RMBT)

The Bowtie2 (Langmead and Salzberg 2012) was used to calculate the number of reads that could be mapped back to the assembly. This index can provide a basic comprehension of the quality of a transcriptome assembly; nevertheless, additional criteria are required to evaluate the performance of each assembler.

### Ex90N50

The Ex90N50 was used instead of the Nx statistical index using the Trinity toolkit (Grabherr et al. 2011). This modified index includes information on transcript expression. Because short, poorly expressed transcripts can dominate a transcriptome assembly, this so-called expression-aware statistic makes up for it. Ex90N50 computes the N50 statistic in the usual way, but it is confined to highly expressed transcripts, which account for 90% of all normalized data. Ex90N50 values were computed using Salmon (v0.11.3) (Patro et al. 2017).

### The average percentage of open reading frames (ORF)

TransRate (Smith-Unna et al. 2016) is a tool for assessing the quality of a de novo transcriptome assembly in the absence of a reference. This tool quickly identifies common transcriptome assembly issues using only sequenced reads and an assembly as input. The average percentage of open reading frames (ORFs) was employed in this study.

### Transcripts longer than 1000 bp

One of the indications of assembly quality is the length of the transcripts. In this study, the number of transcripts longer than 1000 bp was calculated using rnaQUAST (v2.0.1) (Bushmanova et al. 2016).

## BUSCO

To find universal single-copy orthologs, BUSCO v4.0.6 (Seppey et al. 2019) was employed. This tool detects candidate ortholog genes in assemblies and investigates the presence and frequency of single-copy orthologs as an evaluation criterion. BUSCO gives a quantitative assessment of the completeness of an assembly in terms of expected gene content. The results are divided into four categories: (1) complete and single-copy, (2) complete and duplicated, (3) fragmented, and (4) missing BUSCOs. To account for the different isoforms reconstructed by the assembly tool, the values of complete single-copy and complete duplicated BUSCOs were summed and the missing BUSCO values were also taken into account. The embryophyta odb10.2019-11-20 database, which comprises 1,375 BUSCO groups, was used for the evaluations.

### Reconstruction of full-length protein-coding transcripts

To estimate the number of full-length protein-coding transcripts, Blastx was run against the UniProtKB/Swiss-Prot database, followed by the Trinity tool's scripts. For each assembly, the number of proteins that had more than 90% of their protein length covered by assembled transcripts was reported. It should be noted that Blastx was executed with the e-value set to 1e-20 and the `-max_target_seqs` parameter set to 1. Setting the maximum number of target sequences to one significantly reduced execution time, but only the first result that exceeded the e-value criterion was reported (Hölzer and Marz 2019).

### Calculation of the normalized evaluation score

The normalized evaluation score for each assembly was calculated according to Hölzer and Marz (2019). The performance of each assembler  $a_k \in \{a_1, \dots, a_7\}$  on the RNA-Seq dataset  $di \in \{d_1\}$  was evaluated using seven selected metrics  $mj \in \{m_1, \dots, m_7\}$ . For each combination of the data set  $di$  and the metric  $mj$ , a vector  $v^{i,j}$  of raw data  $r_k^{r,j}$  was assigned to each assembler  $a_k$  (Eq. 1).

$$v^{i,j} = (r_1^{i,j}, \dots, r_7^{i,j}). \quad (1)$$

Equation 2 was used to normalize the data in the range of (0, 1):

$$\text{normalized}(v_k^{i,j}) = \frac{v_k^{i,j} - \min v^{i,j}}{\max v^{i,j} - \min v^{i,j}} = n_k^{i,j}. \quad (2)$$

Finally, an index known as the overall metric score (OMS) was considered for each assembly, which is the algebraic sum of the normalized values from the preceding phase. The best assembly was the one with the highest OMS (the highest possible score = 7).

### Computational identification of miRNAs from the saffron transcriptome

An in silico pipeline was created to identify conserved miRNAs from the saffron transcriptome. The procedure and bioinformatics tools used for each step are summarized in Fig. 1. To begin, blastx was run against the nr database to eliminate protein-coding sequences. Then blastn was run against the Rfam 14.3 database (September 2020) to remove sequences from other non-coding RNAs. Rfam is a collection of structural RNA families that contains both non-coding RNA genes and *cis*-regulatory elements. The sequences of all Viridiplantae mature miRNAs were then obtained from the miRBase database (Release 22.1). CD-HIT was used to remove duplicate miRNA sequences from the miRBase database. The total number of miRNAs was initially 10,051, but after removing all sequences that were 100% identical, the total number of sequences was decreased to 5,645.

Known miRNAs were detected according to Alptekin et al. (2017). Two Perl scripts, SUMirFind and SUMirFold (Lucas and Budak 2012), as well as a Python script, SUMirPredictor v2, were used in this protocol. SUMirFind employs blastn with optimized parameters for short sequences and assigns equal scores to long mismatched sequences as well as small non-mismatched sequences (-task blastn-short -ungapped -penalty -1 -reward 1). After aligning the sequences against known miRNAs, the results are limited to candidate miRNAs that have a maximum of two mismatches. SUMirFold then sends a 700-nucleotide sequence from the candidate sequence containing the mature miRNA to UNAFold to generate and evaluate potential pre-miRNA secondary structures. SUMirFold selects the structures with the minimum free energy (MFE) and discards the remaining sequences. If the pri-miRNA structure meets all of the criteria, 20 nucleotides are isolated before and after the mature miRNA sequence, and the secondary structure of the miRNA is re-examined. The scripts SUMirFind and SUMirFold essentially provide evidence for the presence of a suitable secondary structure, the precursor miRNA (pre-miRNA), which contains a mature miRNA sequence. Finally, SUMirPredictor evaluates potentially eligible precursor sequences based on previously reported miRNA core properties (Kurihara and Watanabe 2010), i.e. (i) Potential precursors or hairpin structures cannot have multi-loop structures above the mature miRNA position, (ii) mature miRNA and miRNA\* sequences cannot extend to the head

of the hairpin structure, and (iii) mismatches in the DICER-LIKE enzyme cleavage regions of mature miRNA and miRNA\* sequences are not permitted.

Minimal free energy (MFE) is another important characteristic for the evaluation of miRNA secondary structure. Lower MFE indicates that the secondary structure is thermodynamically more stable. The pre-miRNA sequences should have a highly negative MFE and a minimal free folding energy index (MFEI) (Zhang et al. 2006a). The MFE and the adjusted minimum folding energy (AMFE) which is the MFE of a 100 bp sequence of the predicted pri-miRNAs were calculated using Eqs. 3 and 4, respectively.

$$\text{MFEI} = \frac{\text{AMFE}}{(G + C)\%}, \quad (3)$$

$$\text{AMFE} = \frac{\text{MFE}}{\text{Sequence length}(bp)} \times 100. \quad (4)$$

### Phylogenetic and conservation analysis of the identified miRNAs

To investigate the phylogenetic relationships of several identified miRNAs (miR156, miR171, miR319), mature miRNA sequences of experimentally validated miRNAs from dicot (*A. thaliana*) and monocot (*Zea mays* L.) species, were obtained from the miRBase database. Multiple sequence alignment and subsequent generation of the phylogenetic tree were performed using the ClustalW tool of MEGA X (10.2.5) software (Kumar et al. 2018). For generating a phylogenetic tree, the maximum likelihood statistical method based on the Tamura-Nei model (Tamura and Nei 1993) with 1000 boot-strapped replicates was used. WebLogo (Crooks 2004) was used to conduct a conservation analysis on the identified miRNAs.

### Identification of miRNA targets

Three distinct tools that use various algorithms to identify miRNA targets were used to search for the targets of predicted miRNAs and minimize false-positive results. These tools include psRNATarget (Dai et al. 2019) using the pre-defined V2 scoring scheme (2017 release) with default settings, TAPIR (precise mode using the RNAhybrid search engine) (Bonnet et al. 2010), and psRobot (Wu et al. 2012). psRNATarget is regarded as one of the most reliable and accurate tools for determining miRNA targets (Srivastava et al. 2014). Along with evaluating multiple target sites in an mRNA molecule, this tool considers complementarity and accessibility to the target site; thus, a miRNA sequence may have more than one identified target (Dai and Zhao 2011). TAPIR performs alignment based on a penalty score

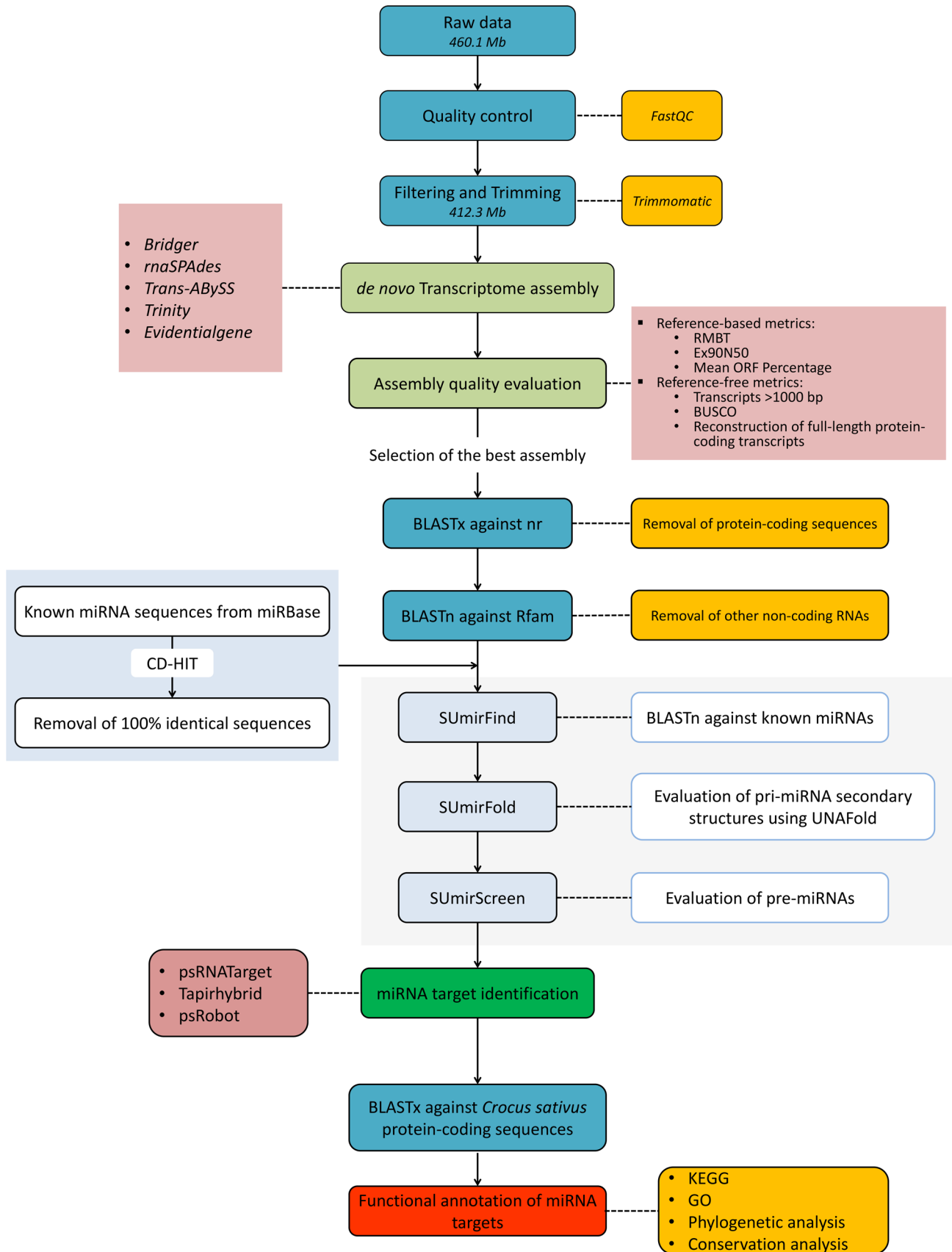


Fig. 1 Workflow of in silico miRNA identification in the transcriptome of *Crocus sativus*

that assesses the number of mismatches, gaps (created by bulges and loop structures), and the number of G:U pairs in the duplex. psRobot predicts the target using a modified Smith–Waterman algorithm with a unique scoring system.

miRNA sequences were used as queries to search against *C. sativus* protein-coding sequences. To reduce false-positive results while preserving sensitivity in finding targets, miRNA candidates discovered by at least two tools were deemed miRNA targets. Following that, blastx (evalue = 5e-10) was used to search the target sequences against the UniRef100 database.

### Functional annotation, biochemical pathway, and network analysis of miRNA targets

The GO terms were retrieved based on UniRef100 identifiers using the UniProt database Retrieve/ID mapping tool (<https://www.uniprot.org>) and classified into three subgroups included cellular components, molecular function, and biological processes using WEGO (Ye et al. 2006). The KEGG Automatic Annotation Server (KAAS) (Moriya et al. 2007) was used to investigate biochemical pathways enriched in miRNA targets. The interaction network between the identified miRNAs and their targets was constructed using Cytoscape 3.8.2 (Shannon 2003). The online version of iTAK was used to identify transcription factors among the identified miRNA targets (Zheng et al. 2016).

## Results

### Data quality control and trimming

A total of ~460.1 million paired-end reads were downloaded from 22 RNA-seq samples of different saffron flower parts, with a mean GC content of 45.59 percent. Trimming produced ~412.3 Mb of high-quality (Q < 20) reads for performing de novo assemblies. In other words, data volume and GC content were both reduced by 10.4% and 0.32 percent, respectively. Table S1 summarizes the detailed results of the trimming process.

### Performance of de novo transcriptome assembly tools

Four different assemblers (Trinity, Trans-ABYSS, rnaSPAdes, and Bridger) with different *k-mer* sizes and a tool for merging different assemblies (i.e. EvidentialGene) were used to identify the best transcriptome assembly tool for saffron. In total, seven de novo transcriptome assemblies were generated. To compare assembly tools, calculated metrics for each assembly were normalized in a (0, 1) range. The normalized scores for each assembly tool are summarized in Table 2.

With the exception of rnaSPAdes, the percentage of RMBT ranged from 97.94% (Trinity) to 92.08% (Trans-ABYSS; *k-mer*=25). rnaSPAdes had the lowest performance

**Table 2** Results for all selected evaluation metrics

<i>k-mer</i>		Trinity	Trans-ABYSS		rnaSPAdes	EvidentialGene	Bridger	
		default	25	32	default		25	32
Evaluation metrics								
Bowtie 2								
1	RMBT	<b>1</b> <sub>97.94</sub>	<b>0.73</b> <sub>92.08</sub>	<b>0.88</b> <sub>95.25</sub>	<b>0</b> <sub>76.15</sub>	<b>0.95</b> <sub>96.77</sub>	<b>0.79</b> <sub>93.41</sub>	<b>0.89</b> <sub>95.48</sub>
Trinity								
2	Ex90N50	<b>0.38</b> <sub>1185</sub>	<b>0.06</b> <sub>701</sub>	<b>0</b> <sub>602</sub>	<b>1</b> <sub>2150</sub>	<b>0.90</b> <sub>1997</sub>	<b>0.96</b> <sub>2092</sub>	<b>0.95</b> <sub>2069</sub>
3	No. of full length transcripts	<b>0.69</b> <sub>6633</sub>	<b>0</b> <sub>4888</sub>	<b>0.24</b> <sub>5499</sub>	<b>0.59</b> <sub>6387</sub>	<b>1</b> <sub>7421</sub>	<b>0.30</b> <sub>5639</sub>	<b>0.44</b> <sub>6005</sub>
rnaQUAST								
4	Transcripts ≥ 1,000 bp	<b>0.14</b> <sub>51405</sub>	<b>0.72</b> <sub>128970</sub>	<b>0</b> <sub>32891</sub>	<b>0.95</b> <sub>159789</sub>	<b>1</b> <sub>165818</sub>	<b>0.90</b> <sub>152252</sub>	<b>0.94</b> <sub>158359</sub>
TransRate								
5	The average ORF percentage	<b>0.7</b> <sub>46.16</sub>	<b>0.94</b> <sub>49.82</sub>	<b>0.85</b> <sub>48.55</sub>	<b>0.08</b> <sub>36.83</sub>	<b>1</b> <sub>50.77</sub>	<b>0</b> <sub>35.59</sub>	<b>0</b> <sub>35.56</sub>
BUSCO								
6	Complete BUSCOs	<b>0.77</b> <sub>1288</sub>	<b>0.06</b> <sub>1222</sub>	<b>0.34</b> <sub>1248</sub>	<b>0.40</b> <sub>1253</sub>	<b>1</b> <sub>1309</sub>	<b>0</b> <sub>1216</sub>	<b>0.32</b> <sub>1246</sub>
7	Missing BUSCOs	<b>0.93</b> <sub>50</sub>	<b>0</b> <sub>78</sub>	<b>0.43</b> <sub>65</sub>	<b>0.17</b> <sub>73</sub>	<b>1</b> <sub>48</sub>	<b>0.10</b> <sub>75</sub>	<b>0.27</b> <sub>70</sub>
	Summarized metric (0,1)-score	<b>4.61</b>	<b>2.52</b>	<b>2.75</b>	<b>3.19</b>	<b>6.85</b>	<b>3.05</b>	<b>3.81</b>

The results for all selected metrics (rows) based on Bowtie 2 (Langmead and Salzberg 2012), Trinity toolkit (Grabherr et al. 2011), rnaQUAST (Bushmanova et al. 2016), Transrate (Smith-Unna et al. 2016), and BUSCO (Simão et al. 2015; Waterhouse et al. 2018) utilities for all assembly tools (columns) are shown. For each metric, normalized scores (in bold) are displayed in the range between 0 and 1. Next to the normalized values, the raw values are given in subscript. The MS of (0,1)-normalized scores is summarized in the last row. The most efficient assembly is the one with the highest summarized metric score. Ex90N50 values are calculated in the same way as N50 values, but only for the top 90% most highly expressed transcripts, which account for 90% of the total normalized expression data. Complete BUSCOs are calculated as the sum of single-copy and duplicated Benchmarking Universal Single-Copy Orthologs

in this metric (76.15 percent). Assemblies with the longest to the shortest Ex90N50 values were as follow: rnaSPAdes (2,150 bp), Bridger ( $k\text{-mer}=25$ , 2,092 bp), Bridger ( $k\text{-mer}=32$ , 2069 bp), EvidentialGene (1997 bp), Trinity (1185 bp), Trans-ABYSS ( $k\text{-mer}=25$ , 701 bp), and Trans-ABYSS ( $k\text{-mer}=32$ , 602 bp). The number of full-length protein-coding transcripts differed significantly between assemblies. The EvidentialGene (7421) identified the greatest number of full-length transcripts, followed by Trinity (6633). By reconstructing 4888 transcripts, Trans-ABYSS  $k\text{-mer}=25$  performed poorly in this index. EvidentialGene (165,818) and Trans-ABYSS ( $k\text{-mer}=32$ ; 32,891) had the highest and lowest number of transcripts larger than 1,000 bp, respectively. In terms of average ORF percentage, EvidentialGene performed the best (50.77%); however, Bridger performed the worst ( $k\text{-mer}=25$ ; 35.59% and  $k\text{-mer}=32$ ; 35.56%).

The number of complete and missing BUSCOs was also taken into account to further evaluate the completeness and accuracy of the assemblies (Fig. S1). EvidentialGene outperformed the other assemblers with 1319 complete BUSCOs out of 1375, and Trinity took second place with 1288 complete BUSCOs. Bridger had a poor performance in this metric ( $k\text{-mer}=25$ ; 1216). The lowest and the highest number of missing BUSCOs were observed in EvidentialGene (48) and Trans-ABYSS ( $k\text{-mer}=25$ ; 78), respectively.

Although none of the assembly tools outperformed the others in all metrics, the EvidentialGene tool, with 6.85 out of a possible 7, performed the best, followed by the Trinity tool with 4.61 points and the Bridger ( $k\text{-mer}=32$ ; 3.81). Trans-ABYSS ( $k\text{-mer}=25$ ) had the lowest performance, with a score of 2.52.

### Identification of miRNAs in the saffron transcriptome

SUmiRFind was used to search non-protein-coding transcripts of saffron (474,280 transcripts) against 5,645 known mature plant miRNAs. In total, 8,030 transcripts with two or fewer mismatches were found. The use of SUmirFold to examine secondary structure features of candidate miRNAs resulted in 175 sequences (2.17%) that were able to form hairpin structure and meet the basic criteria for miRNA identification (including 20 suspicious stem-loop structures) (Fig. 2a); Suspicious sequences are miRNA-miRNA\* duplexes with no mismatch; such sequences may be related to inverted repeats or siRNA sequences (Lucas and Budak 2012).

The resulting sequences were then evaluated by the SUmirPredictor script, and 125 miRNAs were entered in the next step. Following manual examination of the candidate miRNAs to diminish false-positive results and removal of duplicates, 66 miRNAs belonging to 19 families were identified

based on secondary structure prediction and miRNA annotation criteria (Table 3). The highest frequency of miRNAs was found in the miR156 family, followed by miR164, miR319, and mir396 (Fig. 2b). To validate the identified miRNAs, the characteristics of mature and pre-miRNA sequences, as well as pre-miRNA hairpin structures, were examined. The length of the identified mature miRNAs was between 20 and 23 nt with an average of about 21 nt (Fig. 2c; Table 4).

Despite the observed similarity in mature miRNA length, pre-miRNA length and sequence were very different. Pre-miRNAs ranged in length from 99 nucleotides (miR536) to 211 nucleotides (miR319b), with an average length of 134.5 nucleotides (median = 126; SD = 29.47) (Fig. 2d). The identified miRNAs were located at both 5' and 3' strands, with the majority (54.55%) found on the 5' strand. MFE and MFEI are two other important criteria for determining miRNA-related secondary structures. In the current study, the MFE of pre-miRNAs ranged from - 22.7 to - 89.5 kcal/mol, with an average of - 55.76 kcal/mol (median = - 54.25; SD = 12.96), and MFEI ranged from - 0.68 to - 1.32 kcal/mol (median = - 0.91; SD = 0.11) (Fig. 2e).

### Conservation and phylogenetic results

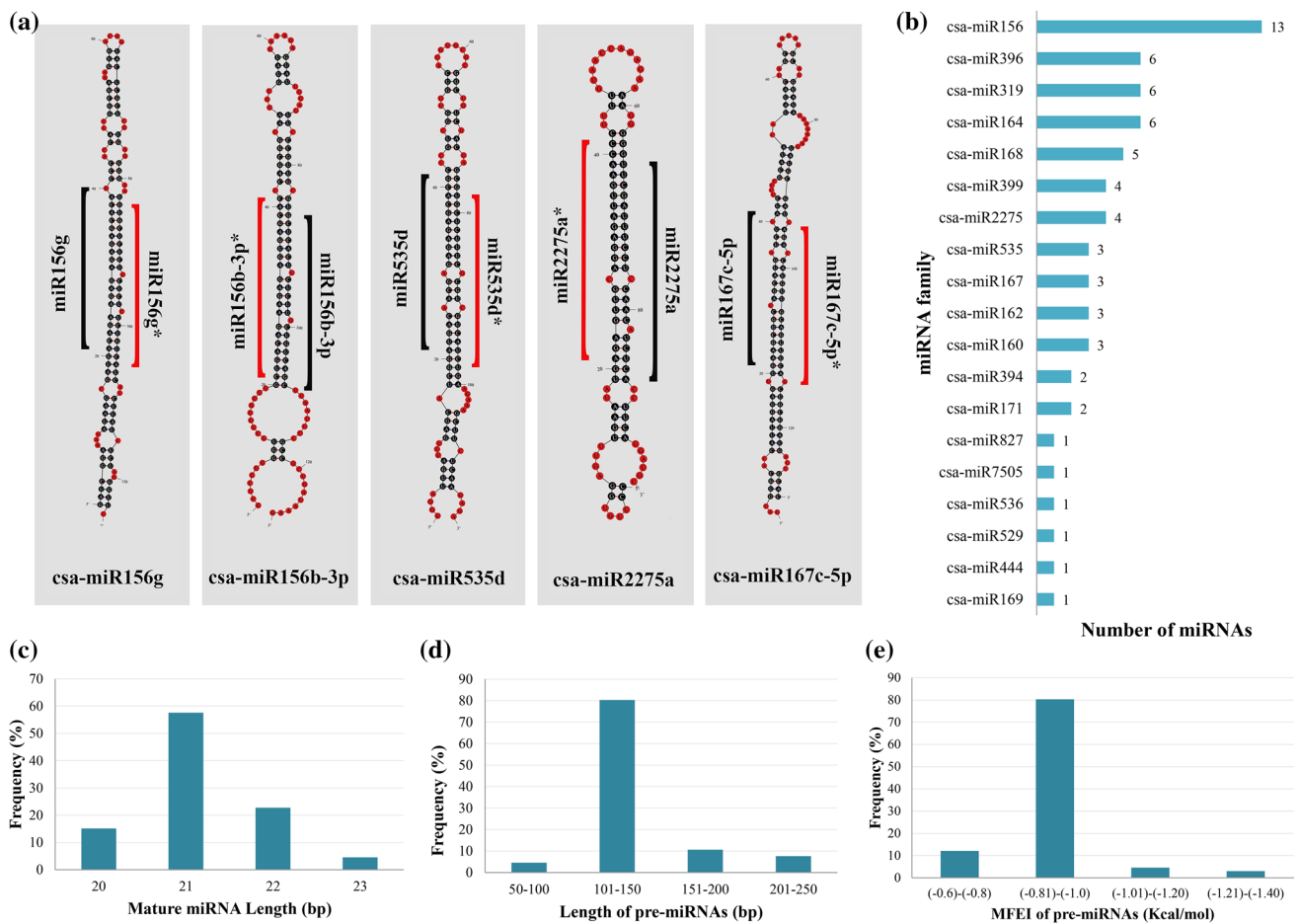
In the present study, phylogenetic and conservation analyses of several identified miRNAs along with previously reported miRNAs were performed. The alignment and conservation analyses of miR156, miR171, and miR319 families were performed along with experimentally validated miRNAs reported for *A. thaliana* and *Z. mays* L. using ClustalW and WebLogo (Fig. 3). The results showed the conserved nature of these miRNAs in different species. Phylogenetic analysis showed that these miRNAs are closely related to *Z. mays* L. and *A. thaliana*. Previous reports have also confirmed the conserved nature of pre-miRNAs as well as mature miRNAs (Zhang et al. 2006b).

### Identification and functional analysis of miRNA targets

As shown in Fig. 4, based on the results of psRNATarget, TAPIRHybrid, and psRobot, 9537, 2972, and 3377 hypothetical targets were identified for all predicted miRNAs, respectively; Of these, 2880 targets were identified by at least two tools and were considered as targets of miRNAs identified in this study.

Based on the gene ontology (GO) results of the predicted targets of miRNAs, 548 unigenes were associated with at least one GO term (Fig. 5). In total, 1,056 GO terms were classified into 40 functional groups, which are members of three basic GO groups called molecular function, cellular components, and biological processes. 438 (41.5%) GO





**Fig. 2** **a** Stem–loop structures of a number of predicted mi-RNAs. Mature miRNA and miRNA\* are identified by brackets, **b** number of identified miRNAs of each family, **c** frequency of mature miRNA

length, **d** frequency of pre-miRNA length, **e** frequency of MFEI range of pre-miRNAs

terms were related to molecular function, 335 (31.7%) GO terms were related to cellular components, and 283 (26.8%) GO terms were related to biological processes.

The major molecular functions associated with the predicted targets were "binding," "catalytic activity," and "transporter activity," which corresponded to miRNAs' major role in the control of transcription cascades by targeting transcription factors (Nazarov et al. 2013). The main targets in the category of cellular components were "cell section," "cell," and "organelle," while the main targets in the category of biological processes were "cellular process", "metabolic process", and "biological regulation".

**miRNAs involved in the biosynthesis of terpenoids**

The analysis of biochemical pathways using the KAAS showed that 770 miRNA targets are involved in 266 metabolic pathways. Three miRNA targets were found in one of the major pathways involved in the metabolism of apocarotenoids, known as “terpenoid backbone biosynthesis”

(ko00900). Four identified miRNAs (csa-miR444b, csa-miR2275-3p, csa-miR2275c, csa-miR169o-3p) could control the biosynthesis of terpenoids by regulating the expression of enzymes involved in this pathway.

**miRNA targets involved in the biosynthesis of apocarotenoids**

Using strict criteria for miRNA target identification, two potential miRNA targets in the apocarotenoids biosynthetic pathway were discovered. According to the results, csa-miR156g and csa-miR156b-3p could regulate the expression of *Carotenoid cleavage dioxygenase (CCD)* and *UDP-glucosyltransferase (UGT)*, respectively (Fig. 6).

**miRNAs involved in flowering transition through regulation of transcription factors**

miRNAs have been found to target 227 transcription factors (TFs) from 27 different families. The results indicated

**Table 3** Identified pri-miRNAs in the saffron transcriptome

Predicted primary miRNA	Mature miRNA sequence	Mature miRNA length (bp)	Number of mismatches with conserved miRNA	miRbase hit
csa-miR156	CUGACAGAAGAGAGUGAGCAC	21	0	ama-miR156
csa-miR156a	UGACAGAAGAGAGUGAGCACC	21	1	bna-miR156a
csa-miR156a-3p	UGCUCACUUCUCUUGCUGUCAG	22	1	vca-miR156a-3p
csa-miR156b-3p	GCUCACUUCUCUUGCUGUCAGCU	23	2	csi-miR156b-3p
csa-miR156c	AUGACAGAAGAGAGAGAGCAC	21	2	smo-miR156c
csa-miR156e-3p	GCUCUCUGCUCUCUCUGUCAUC	22	1	zma-miR156e-3p
csa-miR156f	AUGACAGAAGAGAGAGAGCACA	22	1	gma-miR156f
csa-miR156f-5p	AUGACAGAAGAGAGAGAGCAC	21	1	csi-miR156f-5p
csa-miR156g	ACAGAAGAGAGAGAGCAG	20	1	gma-miR156g
csa-miR156i	CAGAAGAGAGAGAGCACAGC	20	2	cas-miR156i
csa-miR156j	GCUGACAGAAGAGAGUGAGCAC	22	1	cme-miR156j
csa-miR156j-3p	UGCUCUCUGCUCUCUCUGUCAUC	23	1	zma-miR156j-3p
csa-miR156k-3p	GCUCACUUCUCUUGCUGUCAGC	22	2	zma-miR156k-3p
csa-miR160e	UGCCUGGCUCCUGGAUGCCAUC	23	1	mes-miR160e
csa-miR160e	CUGCCUGGCUCCUGGAUGCC	21	1	pab-miR160e
csa-miR160g	UGCCUGGCUCCUGGAUGCCA	21	0	lus-miR160g
csa-miR162a	UAGAUAAACCUCUGCAUCCAG	21	1	hpe-miR162a
csa-miR162b	UCGAUAAACCGCUGCAUCCAG	21	2	osa-miR162b
csa-miR162b-5p	UGGAUGCAGCGUUUAUCGAUC	22	2	ath-miR162b-5p
csa-miR164a-3p	CAUGUGCCUAUCUUCUCCACC	21	2	bdi-miR164a-3p
csa-miR164b-3p	AUGUGCCUAUCUUCUCCACC	20	1	zma-miR164b-3p
csa-miR164c	UGGAGAAGCAGGGCACGUGCA	21	1	osa-miR164c
csa-miR164c-5p	UGGAGAAGCAGGGCACGUGCA	21	1	ath-miR164c-5p
csa-miR164e	UGGAGAAGCAGGGCACGUGUA	21	2	osa-miR164e
csa-miR164h-5p	UGGAGAAGCAGGGCACGUGUA	21	1	zma-miR164h-5p
csa-miR167b-3p	AGACCAUGUUGCAGUUUCAUC	21	1	vca-miR167b-3p
csa-miR167c-5p	GAAGCUGCCAGCAUGAUCUUA	21	2	ath-miR167c-5p
csa-miR167j	UGAAGCUGCCAGCAUGAUCUUA	22	0	mdm-miR167j
csa-miR168-5p	UCGCUUGGUGCAGGUCGGGAAC	22	0	lja-miR168-5p
csa-miR168b	UCGCUUGGUGCAGGUCGGGAA	21	2	pab-miR168b
csa-miR168b-3p	CCCGCCUUGCAUCAAACUGAAU	21	1	vca-miR168b-3p
csa-miR168b-3p	CCCGCCUUGCAUCAAACUGAA	20	1	zma-miR168b-3p
csa-miR168c-5p	UCGCUUGGUGCAGGUCGGGAA	21	1	bra-miR168c-5p
csa-miR169o-3p	GGGAGGUCUUCUUGGCUCGC	20	2	zma-miR169o-3p
csa-miR171a	UGAGCCGCGCCAAUAUCACGC	21	2	gma-miR171a
csa-miR171b-3p	UUGAGCCGCGCCAAUAUCACG	21	1	ath-miR171b-3p
csa-miR2275-3p	UUUGGUUUCUCCAAUAUCUCG	22	0	tae-miR2275-3p
csa-miR2275a	UUCGAUUUCUCCAAUAUCUCA	22	1	aof-miR2275a
csa-miR2275b-3p	UUCAUUUUCUCCAAUAUCUCA	22	1	zma-miR2275b-3p
csa-miR2275c	UUUGAUUUCUCCAAUAUCUCA	22	0	aof-miR2275c
csa-miR319	GUUGGACUGAAGGGAGCUCC	20	1	cpa-miR319
csa-miR319b	UUGGACUGAAGGGAGCUCCAA	21	2	ath-miR319b
csa-miR319c	GUUGGACUGAAGGGAGCUCCA	21	2	cas-miR319c
csa-miR319c-3p	UUGGACUUAAGGGAGCUCCAA	21	1	mtr-miR319c-3p
csa-miR319d	UUGGACUGAAGGGAGCUCCA	20	2	pab-miR319d
csa-miR319p	GGUUGGACUGAAGGGAGCUCC	21	2	gma-miR319p
csa-miR394b	UUGGCAUUCUGUCCACCUCC	20	1	pab-miR394b
csa-miR394c	UUGGCAUUCUGUCCACCUCCA	22	0	mes-miR394c
csa-miR396	UCCACGGCUUUCUUGAACU	20	1	smo-miR396

**Table 3** (continued)

Predicted primary miRNA	Mature miRNA sequence	Mature miRNA length (bp)	Number of mismatches with conserved miRNA	miRbase hit
csa-miR396a	UUCCACAGCUUUCUUGAACUG	21	1	mdm-miR396a
csa-miR396b-5p	UUCCACAGCUUUCUUGAACUG	21	1	ath-miR396b-5p
csa-miR396e	UUCCACGGCUUUCUUGAACUG	21	0	cme-miR396e
csa-miR396e	UUCCACAGCUUUCUUGAACUGG	22	1	gma-miR396e
csa-miR396f	UCCCACGUCUGUCUUGAACUU	21	2	pab-miR396f
csa-miR399b	UGCCAAAGGAGAGUUGCCCUA	21	0	cme-miR399b
csa-miR399b-5p	AGGGCUUCUCUCCUUUGGCAG	21	0	csi-miR399b-5p
csa-miR399f	GGCAGCUCUCCUUUGGCAAG	20	1	pab-miR399f
csa-miR399g	AGGGCUUCUCUCCUUUGGCAGG	22	1	cme-miR399g
csa-miR444b	UGUAGUUGCUGCCUCAAGCUU	21	1	bdi-miR444b
csa-miR529d	AGAAGAGAGAGAGCACAGCCC	21	0	ppt-miR529d
csa-miR535-3p	GUGCUCUCUCUCGUUGUCAUC	21	1	vca-miR535-3p
csa-miR535a	UGACAACGAGAGAGAGCACGC	21	1	mes-miR535a
csa-miR535d	AUGACAACGAGAGAGAGCACG	21	2	mes-miR535d
csa-miR536	UCGUGCCACGCUGUGUGCGUU	21	1	aof-miR536
csa-miR7505b	ACAGUUUUAGAAACCAUCCCU	21	2	gra-miR7505b
csa-miR827	UUAGAUGAUCAUCAACAAACA	21	1	nta-miR827

that the majority of TFs belonged to SBP, MYB, GRAS, NAC, and Trihelix families, respectively (Fig. S2a). Seventeen miRNA families were found to be capable of targeting TFs, with miR156 being the most abundant (24.23%), followed by miR319 (12.78%) and miR169 (10.13%). MiR827 and miR168 had the lowest number of members (Fig. S2b).

Several miRNAs are identified to be involved in flowering transition through the regulation of TFs. For instance, SBP (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE; SPL or SBP) domain proteins were identified as one of the most abundant targets of the miR156 family. Also, TCP transcription factors were identified as one of the miR319 targets. GRAS TFs were identified as one of the targets of miR171. Another group of TFs was bZIP which was identified as a target of miR396 and miR2275.

### Construction of miRNA-mRNA network

The miRNA-mRNA interaction network was constructed in Cytoscape 3.8.2 using 66 identified miRNAs and 1187 mRNA targets (Fig. 7). The results showed that, among all miRNAs, miR156 and miR169 had the most connections with targets, with 344 and 135 connections, respectively. There were 14 targets co-regulated by 16 different miRNA families. The most common families were miR156 and miR169, which shared four targets with other miRNA families.

### Discussion

Recent advances in experimental and computational technologies have led to a breakthrough in discovering the role of miRNAs. In particular, the advent of NGS has revolutionized miRNAs discovery, and as a result, many new miRNAs have been discovered recently. miRNAs are regarded as the major regulators of gene expression at the transcriptional or post-transcriptional level in different organisms (Bartel 2018). Moreover, a miRNA can participate in the regulation of multiple genes or mRNAs in a wide variety of plant growth and developmental stages (Dehury et al. 2013) that makes them a versatile tool for metabolic engineering, crop improvement, etc. (Sabzehzari and Naghavi 2019).

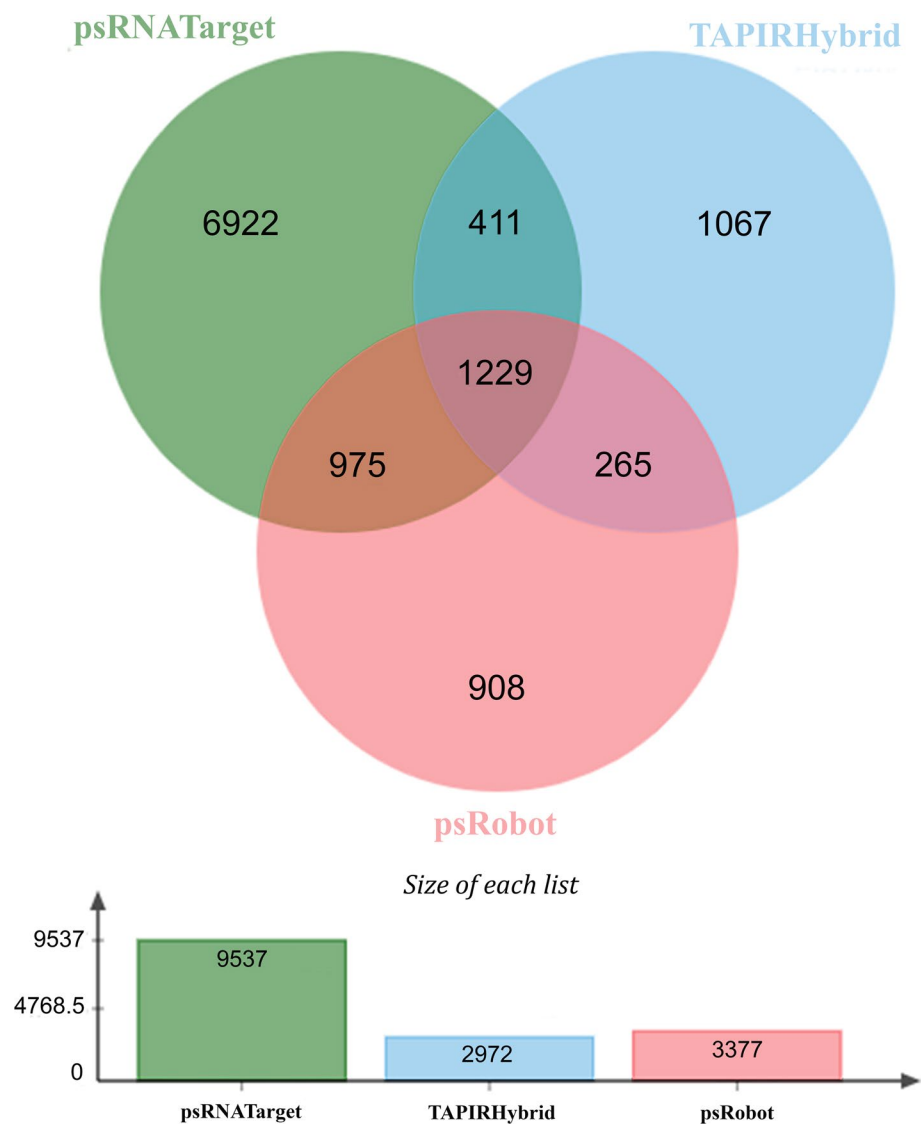
In the present study, an *in silico* approach was adopted for the identification of miRNAs in the saffron transcriptome. Because there was no saffron reference genome available at the time of the analyses, a *de novo* transcriptome assembly that could be a fair representation of all samples had to be created first. So, for this purpose, seven transcriptome assemblies were created using four different assembly tools and a tool for merging different assemblies. For determining the quality of a genome or transcriptome assembly, evaluation criteria are essential; however, opinions differ on which evaluation criteria are best for *de novo* assembling (Hölzer and Marz 2019). Our study included 22 RNA-seq samples of saffron flower and for evaluation of the assemblies both biological/reference-based (RMBT,

**Table 4** Characteristics of pre-miRNAs identified in the Saffron transcriptome

Identified miRNAs	pre-miRNA length	Position of hair-pin structure	Pre-miRNA MFE (Kcal/mol)	Pre-miRNA GC%	Pre-miRNA MFEI (Kcal/mol)
csa-miR156	124	5'	- 50	46.77	- 0.86
csa-miR156a	126	5'	- 50	45.24	- 0.88
csa-miR156a-3p	124	3'	- 50	45.97	- 0.88
csa-miR156b-3p	128	3'	- 50	46.09	- 0.85
csa-miR156c	127	5'	- 55.7	49.61	- 0.88
csa-miR156e-3p	129	3'	- 55.8	49.61	- 0.87
csa-miR156f	127	5'	- 55.7	49.61	- 0.88
csa-miR156f-5p	127	5'	- 55.7	49.61	- 0.88
csa-miR156g	125	5'	- 54.8	48.80	- 0.90
csa-miR156i	123	5'	- 54.6	48.78	- 0.91
csa-miR156j	126	5'	- 50	46.03	- 0.86
csa-miR156j-3p	129	3'	- 55.8	49.61	- 0.87
csa-miR156k-3p	126	3'	- 50	45.24	- 0.88
csa-miR160e	124	5'	- 57.6	49.19	- 0.94
csa-miR160e	122	3'	- 57.6	47.54	- 0.99
csa-miR160g	124	5'	- 57.6	49.19	- 0.94
csa-miR162a	137	5'	- 43.6	42.34	- 0.75
csa-miR162b	137	3'	- 46.9	43.07	- 0.79
csa-miR162b-5p	135	3'	- 43.6	43.70	- 0.74
csa-miR164a-3p	180	5'	- 66.7	40.56	- 0.91
csa-miR164b-3p	180	3'	- 66.7	40.56	- 0.91
csa-miR164c	180	3'	- 66.7	40.56	- 0.91
csa-miR164c-5p	180	5'	- 66.7	40.56	- 0.91
csa-miR164e	133	5'	- 50.9	46.62	- 0.82
csa-miR164h-5p	133	5'	- 52.3	46.62	- 0.84
csa-miR167b-3p	136	5'	- 50.8	41.18	- 0.91
csa-miR167c-5p	130	3'	- 48.4	40.77	- 0.91
csa-miR167j	136	5'	- 50.8	41.18	- 0.91
csa-miR168-5p	110	5'	- 50.7	55.45	- 0.83
csa-miR168b	110	5'	- 50.7	55.45	- 0.83
csa-miR168b-3p	110	5'	- 50.7	55.45	- 0.83
csa-miR168b-3p	108	3'	- 50.7	56.48	- 0.83
csa-miR168c-5p	110	5'	- 50.7	44.45	- 0.83
csa-miR169o-3p	125	5'	- 45.2	49.60	- 0.73
csa-miR171a	125	3'	- 55.1	56	- 0.79
csa-miR171b-3p	123	3'	- 54.3	55.28	- 0.80
csa-miR2275-3p	110	3'	- 34.8	33.64	- 0.94
csa-miR2275a	100	3'	- 39	44	- 0.89
csa-miR2275b-3p	100	3'	- 22.7	33	- 0.69
csa-miR2275c	109	3'	- 32	31.19	- 0.94
csa-miR319	207	3'	- 88.6	46.38	- 0.92
csa-miR319b	211	3'	- 89.5	46.46	- 0.91
csa-miR319c	209	3'	- 88.6	46.41	- 0.91
csa-miR319c-3p	109	3'	- 39.7	36.70	- 0.99
csa-miR319d	209	3'	- 88.6	46.41	- 0.91
csa-miR319p	207	3'	- 88.6	46.38	- 0.92
csa-miR394b	106	5'	- 51.2	50.49	- 0.98
csa-miR394c	103	5'	- 51.2	50.49	- 0.98
csa-miR396	180	5'	- 70	35.56	- 1.09



**Fig. 4** **a** Venn diagram depicting the number of miRNA targets identified by various tools, **b** gene ontology chart of miRNA targets predicted in the saffron transcriptome

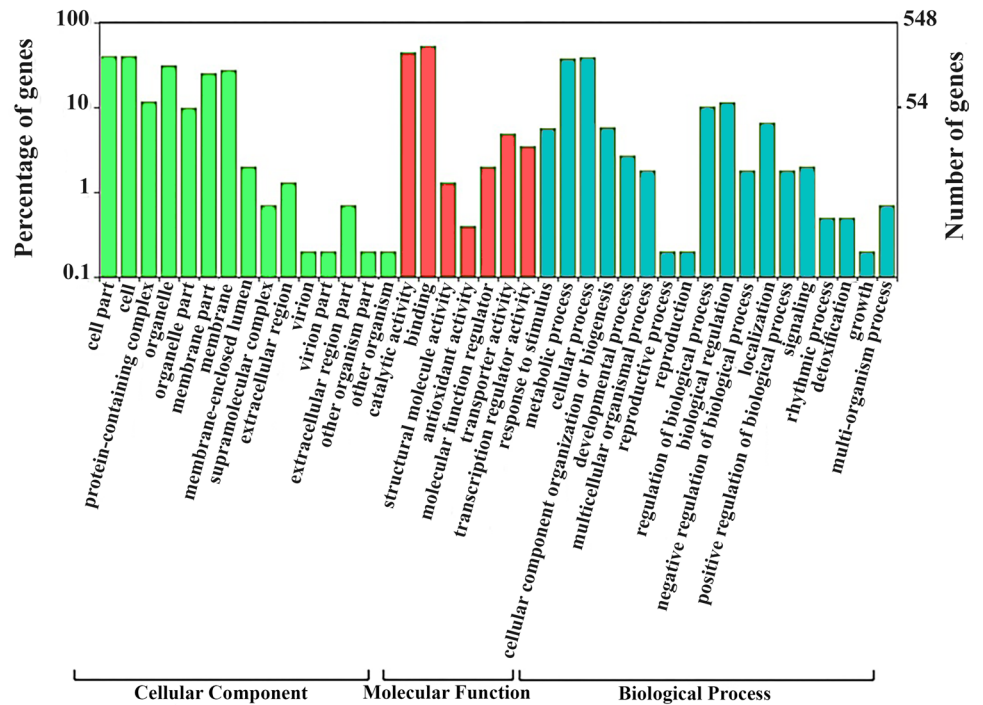


Ex90N50, and the average percentage of ORFs) and statistical criteria without the need for reference (transcripts longer than 1000 bp, BUSCO, and reconstruction of full-length protein-coding transcripts) were used. Overall, a workflow for miRNA identification in saffron was developed, which can be applied to other non-model species as well.

Comparison of the findings with those of other studies confirms that different assembly tools generate significantly different assemblies (Li et al. 2008). Thus, it is crucial to compare those assemblies using different evaluation criteria and select the most accurate and complete assembly for downstream analyses. Initially, the values obtained for each index were normalized, and the sum of the normalized values of different indices for each assembly (OMS) was used as a criterion for comparing the efficiency of assemblies. Even though none of the assembly tools had the highest scores across all metrics, EvidentialGene

outperformed the others. Trinity and Bridger ( $k\text{-mer} = 32$ ) were ranked second and third, respectively, with TransABySS ( $k\text{-mer} = 25$ ) performing the worst. In general, if a large percentage of the transcripts can be mapped back to the assembly, it can be considered accurate and complete; thus, RMBT is a measure of assembly completeness for de novo assembled transcripts (Moreton et al. 2014). A higher percentage of RMBT indicates a more efficient assembly. Except for rnaSPAdes, the RMBT percentage for all assemblies was greater than 92 percent. In this study, the percentage of RMBT was higher than in previous studies. For example, the RMBT percentage in almond (*Prunus dulcis* Mill.) and pistachio (*Pistacia vera* L.) was reported to be 83.68% and 89.93%, respectively (Mousavi et al. 2014; Moazzam Jazi et al. 2017). Ex90N50 is another commonly used metric for comparing the performance of different assemblies. It indicates that 90% of the normalized expression data is assembled on contigs

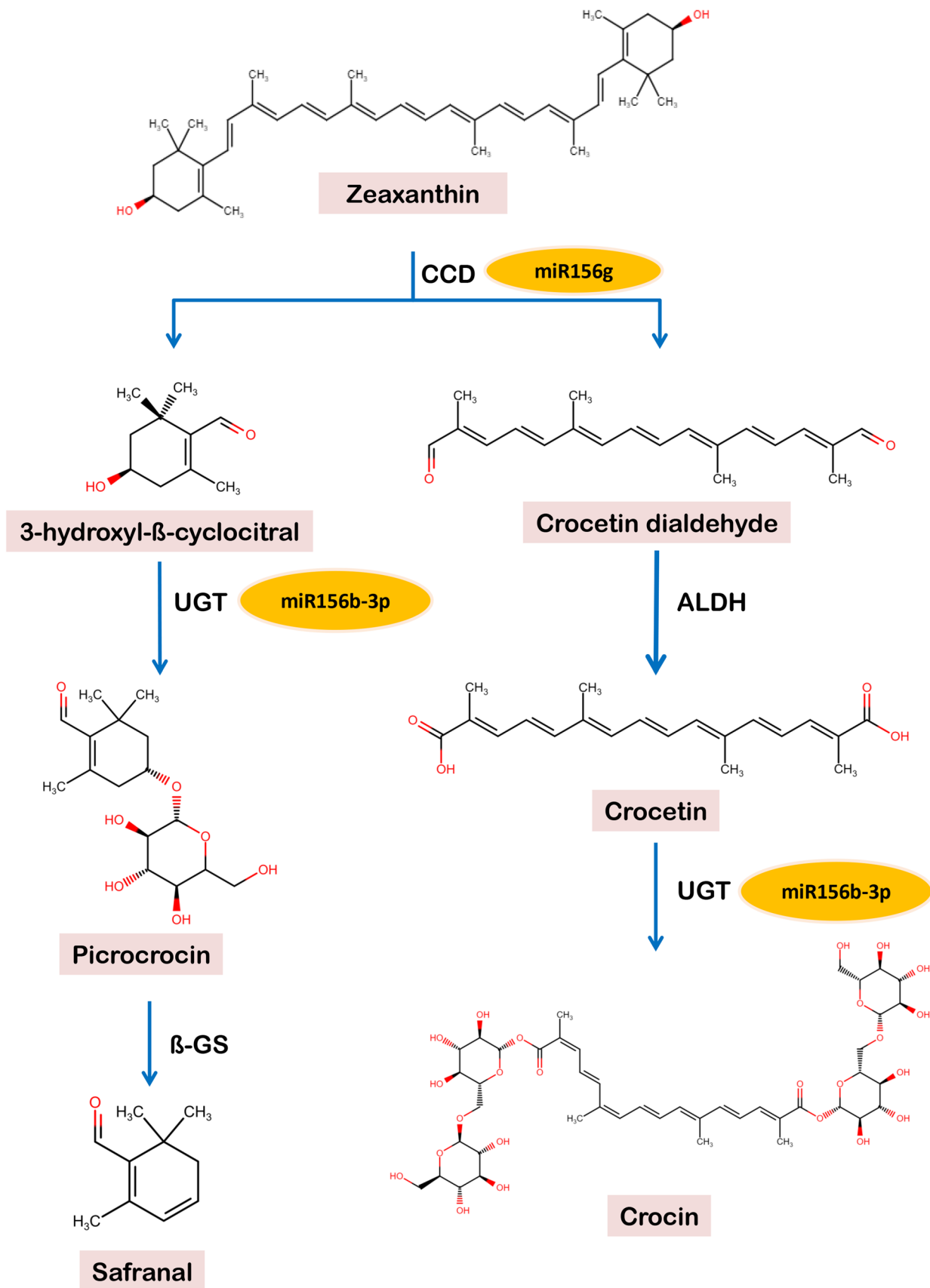
**Fig. 5** Gene ontology chart of miRNA targets predicted in the saffron transcriptome



of length N50. There was a significant difference in performance between the best (2,150 bp for rnaSPAdes) and worst (602 bp for Trans-ABYSS,  $k\text{-mer}=32$ ) assemblies. EvidentialGene performed admirably in the reconstruction of full-length transcripts, reconstructing 7,421 transcripts. Previous studies have also reported on the use of this evaluation metric (Nakashima et al. 2014; Chopra et al. 2014). EvidentialGene outperformed other assemblies in the case of Transcripts  $\geq 1,000$  bp. EvidentialGene and Trans-ABYSS ( $k\text{-mer}=25$ ) were very similar in that they both contained roughly half of the ORFs. BUSCO is another metric that can be used as a reference for determining the completeness of an assembly. It will scan a database of single-copy ortholog genes with a high degree of conservation. It examines the transcriptome assembly to see whether each BUSCO group is complete, duplicated, fragmented, or missing. The most complete BUSCOs (1,309) were found in EvidentialGene, which takes advantage of merging different assemblies. This finding is consistent with that of Mamrot et al. (2017) who also reported better performance of merging assembly tools. Furthermore, EvidentialGene had the lowest number of missing BUSCOs, indicating its efficiency in assembling transcriptomes. In summary, each assembler uses a different algorithm and can only identify a limited number of distinct transcripts (He et al. 2015). As a result, it may be useful to merge different assemblies produced by different assembly tools in order to reconstruct a more comprehensive transcriptome

as a reference; however, the downside to this approach is that it is time-consuming and complicated.

De novo transcriptome assembly has previously been performed in the *Crocus* species. Several assemblers are used in some studies. Jain et al. (2016), for example, investigated the transcriptome of various saffron organs such as stigma, petals, leaves, and corms using various assembly tools such as Trinity, Oases, Velvet, ABYSS, SOAPdenovo, and CLC Genomics Workbench; after examining indices such as average contig length, N50, and the number of created contigs, Oases was finally introduced as the best assembler. However, in most studies, only one assembler is used; for example, Hu et al. (2020) used the Trinity tool to investigate the transcriptome changes of saffron stigmas during the flowering transition. In another study, Ahrazem et al. (2019) used Trinity as the only assembly tool for stigma RNA-Seq to investigate the pathway of crocin biosynthesis in three *Crocus* species: *C. ancyrensis*, *C. cartwrightianus*, and *C. sativus*. More information on which metrics best predict de novo transcriptome assembly quality will undoubtedly aid in the development of protocols for producing high-quality assemblies. Taken together, EvidentialGene demonstrated the highest efficiency for de novo transcriptome assembly in saffron; thus, it appears that using the tools for merging different assemblies will help in achieving a more complete assembly, which will aid in increasing the accuracy of downstream analysis. However, if resources are limited and only one assembly



**Fig. 6** A diagram of the apocarotenoids biosynthesis pathway



tool is available, the Trinity is recommended as it was the second best performing assembly.

An in silico workflow was developed in this study to identify miRNAs based on saffron flower transcriptome. Following the removal of protein-coding and non-coding RNAs (other than miRNAs) from the generated transcriptome, primary miRNA sequences were identified using a homology-based search against known plant miRNAs. To reduce the possibility of false-positive results, the homology search was made as strict as possible. Finally, 66 miRNAs from 19 families were identified using secondary structure prediction and miRNA annotation criteria. All of the identified miRNAs were found on the stem of the stem-loop structure. The length of pre-miRNAs varied greatly, with an average of 134.5 nt. These findings are consistent with those of Barozai et al. (2012), who discovered variation in the length of pre-miRNAs. The majority (57.58%) of mature miRNAs had a length of 21 nt followed by 22 nt (22.73%), 20 nt (15.15%), and 23 nt (4.55%). These findings were also consistent with previous research because many mature plant miRNAs have 19–24 nt and are skewed toward 21 nt (Kurihara and Watanabe 2010; Kurtoglu et al. 2014). The MFE values were highly negative (mean =  $-55.76$  kcal/mol), owing to the stable secondary structure of pre-miRNAs in comparison to other noncoding RNAs (Bonnet et al. 2004). Furthermore, pre-miRNAs had an average MFEI of  $-0.9$  kcal/mol, compared to 0.64 for tRNAs, 0.59 for rRNAs, and 0.65 for mRNAs (Zhang et al. 2006c). The MFE values were extremely negative (mean =  $-55.76$  kcal/mol), owing to the stable secondary structure of pre-miRNAs in comparison to other noncoding RNAs (Bonnet et al. 2004). Furthermore, pre-miRNAs had an average MFEI of  $-0.9$  kcal/mol, compared to 0.64 for tRNAs, 0.59 for rRNAs, and 0.65 for mRNAs (Zhang et al. 2006a). As a result, the identified miRNAs are almost certainly true miRNAs.

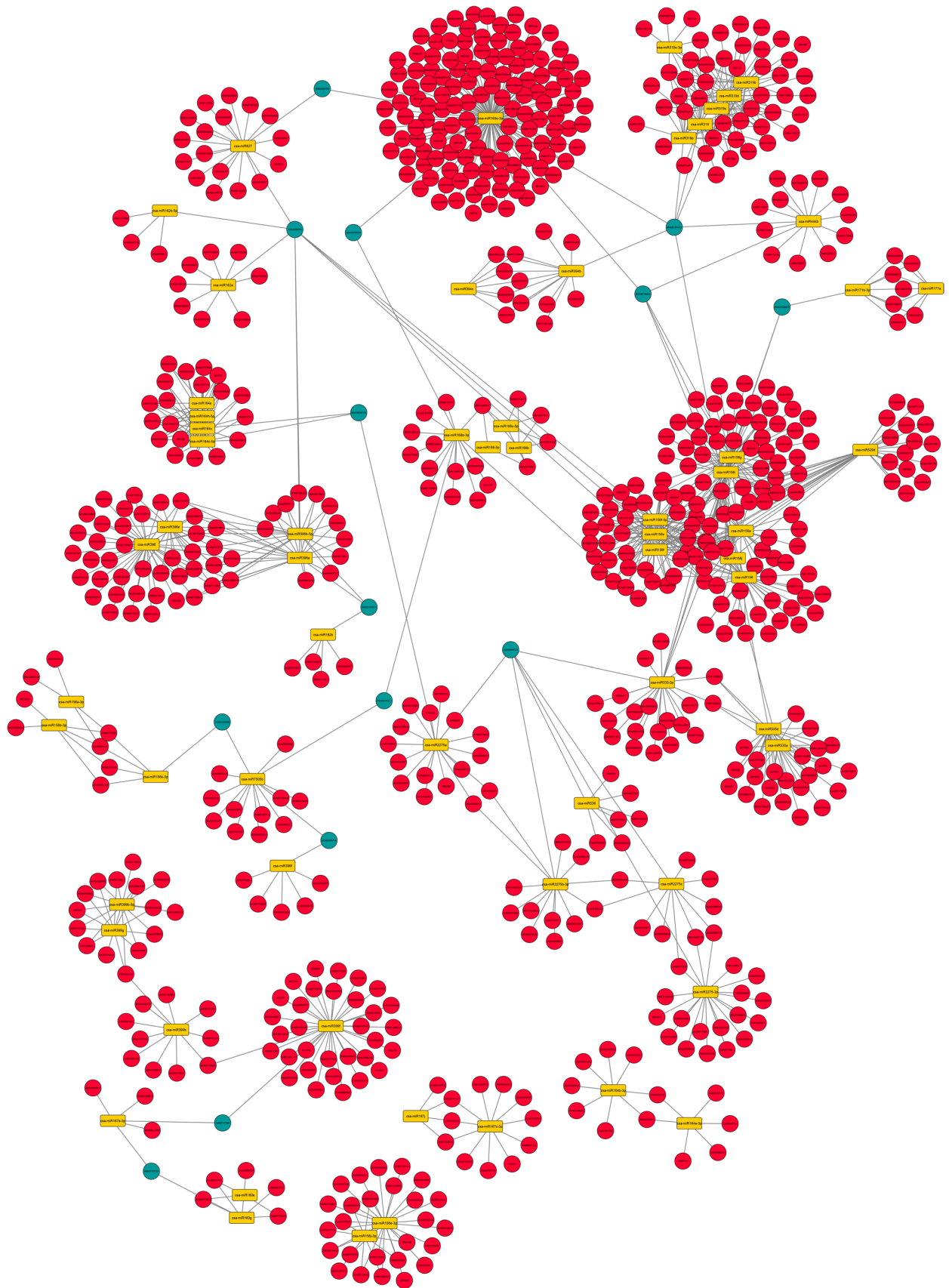
Plant miRNAs share a high degree of sequence similarity and are conserved in both precursor and mature sequences across distantly related taxa (Zhang et al. 2006a). A homology-based search was performed to determine the phylogenetic relationship of the identified *csa*-miRNAs to those of experimentally validated miRNAs from other species. The observed variation in miRNA sequence may allow them to target different miRNAs (Roy et al. 2020). The results of phylogenetic and conservation analysis support previous researchers' findings that miRNAs are conserved across species (Zhang et al. 2006c; Barozai et al. 2012).

miRNAs are thought to be the most important regulators of gene expression at both the transcriptional and post-transcriptional levels (Catalanotto et al. 2016). These molecules bind to complementary sites on target mRNAs and suppress expression by inhibiting mRNA translation or cleavage (Fahlgren and Carrington 2010). Thus, identifying miRNA targets using in silico analysis is a straightforward

method. In the current study, three different tools were used to identify the target of miRNAs with high reliability, reduce false-positive results, and maintain sensitivity, as previously described. The targets identified by at least two tools were considered as possible targets of the identified miRNAs. The identified targets were found in a wide range of biological and metabolic processes. Four miRNAs (*csa*-miR444b, *csa*-miR2275-3p, *csa*-miR2275c, *csa*-miR169o-3p) have been identified as being directly involved in the biosynthesis of terpenoids. Terpenoids are a large class of plant chemicals that contain at least 40,000 different compounds that play various roles in plant growth and development (Tholl 2015). All terpenoids are made up of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) (Kubeczka 2020). IPP and DMAPP are produced in plants via two distinct pathways known as the mevalonic acid (MVA) and methylerythritol phosphate (MEP) pathways. The five-carbon products of the MEP pathway, in particular, are preferred for the biosynthesis of a variety of products, including carotenoids and their derivatives (e.g. apocarotenoids) (Tholl 2015). Since previous studies found that miRNAs play a role in the regulation of the terpenoid backbone biosynthesis pathway in *Persicaria minor* (Samad et al. 2019), *Panax notoginseng* (Wei et al. 2015), and *Ginkgo biloba* (Ye et al. 2020), it appears that miRNAs may also play a role in the regulation of the MEP pathway in saffron.

Despite their importance, the main source of apocarotenoids is the extraction from the saffron plant, which is expensive. So researchers are attempting to produce these compounds in various organisms (e.g., bacteria, yeast, etc.) via metabolic engineering. As a result, their biosynthesis pathway has received much attention, and the enzymes involved in this pathway have been extensively studied. However, there has been little research into its regulatory mechanisms (Liu et al. 2020). To our knowledge, only one study on the regulators of the apocarotenoids pathway has been published, and the role of a transcription factor (*CsULT1*) has been reported (Ashraf et al. 2015), but no reports on the role of miRNAs in the regulation of this pathway have been published to date.

One of the most important findings of this study was the identification of two miRNAs (*csa*-miR156g and *csa*-miR156b-3p) that may control the biosynthesis of apocarotenoids, as CCD and UGT enzymes were identified as miRNA targets. These enzymes are required for the biosynthesis of crocin, picrocrocin, and safranal, the three main compounds found in saffron. The first step in the biosynthesis of apocarotenoids is the cleavage of carotenoids by CCD enzymes. CCD catalyzes the conversion of zeaxanthin to crocetin dialdehyde and 3-hydroxyl- $\beta$ -cyclocitral, whereas UGT converts picrocrocin to crocin and crocetin to picrocrocin (Tan et al. 2019). Glycosylation of carotenoids by the UGT enzyme results in stable and water-soluble apocarotenoids (Liu et al.



**Fig. 7** miRNA–mRNA interaction network in the transcriptome of saffron. miRNAs (yellow rectangles), miRNA targets (red circles), shared miRNA target (blue circles)

2020). As a result, the current findings of the potential roles of miRNAs would undoubtedly help to further understand the regulatory mechanisms of apocarotenoids biosynthesis in the context of metabolic engineering. These miRNAs contribute to our understanding of potential regulators of this pathway, which could be used to increase the amount of effective components of saffron via metabolic engineering.

Due to the importance of higher flower production in saffron and the processes underlying flowering transition, attention has recently been focused on the flowering control mechanisms in this valuable plant, because the saffron industry's sustainable development is dependent on the quality and quantity of stigmas (Hu et al. 2020). Thus, studying the molecular mechanisms of flowering transition in saffron is critical for resolving these issues. Based on previous research on the role of miRNAs in plant flowering control (Schwab et al. 2005; Jung et al. 2011), and the lack of such research in saffron, an attempt was made to investigate several identified miRNAs that may be related to the flowering transition in saffron. Hu et al. (2020) proposed a model for flowering transition in saffron that included several TFs involved in this process. Surprisingly, the transcription factors SPL, TCP, and bZIP were identified as miRNA targets in the current study. SPL transcription factors, also known as floral activators (Gandikota et al. 2007), were identified as *csa*-miR156 targets. It has also been reported that miR156 regulates SPL3 expression in *A. thaliana* via transcript cleavage and translational inhibition (Gandikota et al. 2007). SPL genes are found in plants as part of a multigene family. *A. thaliana*, for example, has 16 SPL genes. Many SPL genes play critical roles in plant growth processes such as vegetative phase change, somatic embryogenesis, anthocyanin biosynthesis, gibberellin (GA) biosynthesis and signaling, and plant stress response (Wang et al. 2018). Overexpression of miR156 in transgenic *Populus x canadensis* reduced the expression of miR156-targeted SPL genes, resulting in a prolongation of the juvenile phase (Wang et al. 2011). This finding suggests that the role of miR156/SPL in the vegetative growth stages of annual herbaceous plants and perennial trees is well protected (Li et al. 2018). TCP transcription factors were also found to be targeted by *csa*-miR319. This TF family has a conserved basic helix–loop–helix domain that is important for DNA binding or protein–protein interactions (Cubas et al. 1999). During plant development, gibberellin interacts with the miR319/TCP module in addition to the miR156/SPL module. According to reports, the miR319/TCP module regulates flowering time. Excess miR319 accumulation in *Arabidopsis* negatively regulates TCP class II genes, including TCP4, resulting in a late

flowering phenotype. Plants with miR319-resistant TCP variants or overexpressed TCPs, on the other hand, flower earlier (Sarvepalli and Nath 2011). TCP4 targets of miR319 have been shown to bind to the CONSTANS (CO) promoter, inducing expression and positively regulating the flowering period (Liu et al. 2017). However, miR319-regulated TCP genes can also suppress target genes at specific developmental stages (Burko et al. 2013).

## Conclusions

In the current study, an in silico workflow was developed for identifying miRNAs in saffron using transcriptome data. The efficiency of various de novo transcriptome assembly tools was evaluated and EvidentialGene was introduced as the most efficient tool for transcriptome assembly in this plant. A total of 66 miRNAs were discovered, representing 19 distinct miRNA families. The targets of saffron miRNAs were found to be involved in several processes. Four miRNAs have been identified as being involved in the terpenoids backbone biosynthesis. The apocarotenoids biosynthetic pathway enzymes CCD and UGT were also recognized as potential miRNA targets. A number of TFs were also discovered to be miRNA targets involved in the flowering transition. Because there is no evidence of species-specific identification of miRNAs in saffron, the number of miRNAs and their targets may be greater than the number estimated in this study. The current findings highlight the importance of miRNAs and their targets in regulating various biological processes in saffron and, in particular, improve our understanding of the control mechanisms of the apocarotenoids biosynthesis pathway in this precious plant.

## Author contribution statement

ATD conducted the computational analyses, processed the results, and wrote the paper. RN and FM conceived and designed the experiments and revised the manuscript. SAS conceived and designed the experiments, contributed analysis tools or data, and revised the manuscript. All authors read and approved the manuscript.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00425-021-03761-7>.

**Funding** The authors received no financial support for the research.

**Data availability** The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Ethical approval** No animals or harmful substances have been used for this research.

## References

- Agharbaoui Z, Leclercq M, Remita MA et al (2015) An integrative approach to identify hexaploid wheat miRNAsome associated with development and tolerance to abiotic stress. *BMC Genomics* 16:339. <https://doi.org/10.1186/s12864-015-1490-8>
- Ahrazem O, Argandoña J, Fiore A et al (2019) Multi-species transcriptome analyses for the regulation of crocins biosynthesis in *Crocus*. *BMC Genomics* 20:320. <https://doi.org/10.1186/s12864-019-5666-5>
- Alptekin B, Akpınar BA, Budak H (2017) A comprehensive prescription for plant miRNA identification. *Front Plant Sci* 7:1–28. <https://doi.org/10.3389/fpls.2016.02058>
- Ashraf N, Jain D, Vishwakarma RA (2015) Identification, cloning and characterization of an ultrapetala transcription factor CsULT1 from *Crocus*: a novel regulator of apocarotenoid biosynthesis. *BMC Plant Biol* 15:25. <https://doi.org/10.1186/s12870-015-0423-7>
- Axtell MJ, Westholm JO, Lai EC (2011) Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol* 12:221. <https://doi.org/10.1186/gb-2011-12-4-221>
- Baba SA, Mohiuddin T, Basu S et al (2015) Comprehensive transcriptome analysis of *Crocus sativus* for discovery and expression of genes involved in apocarotenoid biosynthesis. *BMC Genomics* 16:698. <https://doi.org/10.1186/s12864-015-1894-5>
- Barozai MYK, Baloch IA, Din M (2012) Identification of microRNAs and their targets in *Helianthus*. *Mol Biol Rep* 39:2523–2532. <https://doi.org/10.1007/s11033-011-1004-y>
- Bartel DP (2018) Metazoan microRNAs. *Cell* 173:20–51. <https://doi.org/10.1016/j.cell.2018.03.006>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Bonnet E, Wuyts J, Rouze P, Van de Peer Y (2004) Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* 20:2911–2917. <https://doi.org/10.1093/bioinformatics/bth374>
- Bonnet E, He Y, Billiau K, Van de Peer Y (2010) TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics* 26:1566–1568. <https://doi.org/10.1093/bioinformatics/btq233>
- Budak H, Akpınar BA (2015) Plant miRNAs: biogenesis, organization and origins. *Funct Integr Genomics* 15:523–531. <https://doi.org/10.1007/s10142-015-0451-2>
- Burko Y, Shleizer-Burko S, Yanai O et al (2013) A role for APETALA1/FRUITFULL transcription factors in tomato leaf development. *Plant Cell* 25:2070–2083. <https://doi.org/10.1105/tpc.113.113035>
- Bushmanova E, Antipov D, Lapidus A et al (2016) rnaQUAST: a quality assessment tool for de novo transcriptome assemblies. *Bioinformatics* 32:2210–2212
- Bushmanova E, Antipov D, Lapidus A, Prjibelski AD (2019) rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience*. <https://doi.org/10.1093/gigascience/giz100>
- Catalanotto C, Cogoni C, Zardo G (2016) MicroRNA in control of gene expression: an overview of nuclear functions. *Int J Mol Sci* 17:1712. <https://doi.org/10.3390/ijms17101712>
- Chang Z, Li G, Liu J et al (2015) Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biol* 16:30. <https://doi.org/10.1186/s13059-015-0596-2>
- Chen C, Zhong Y, Yu F, Xu M (2020) Deep sequencing identifies miRNAs and their target genes involved in the biosynthesis of terpenoids in *Cinnamomum camphora*. *Ind Crops Prod* 145:111853. <https://doi.org/10.1016/j.indcrop.2019.111853>
- Chib S, Thangaraj A, Kaul S et al (2020) Development of a system for efficient callus production, somatic embryogenesis and gene editing using CRISPR/Cas9 in saffron (*Crocus sativus* L.). *Plant Methods* 16:47. <https://doi.org/10.1186/s13007-020-00589-2>
- Chopra R, Burow G, Farmer A et al (2014) Comparisons of de novo transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp) RNA-seq data. *PLoS ONE* 9:e115055. <https://doi.org/10.1371/journal.pone.0115055>
- Chow C-N, Zheng H-Q, Wu N-Y et al (2016) PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res* 44:D1154–D1160. <https://doi.org/10.1093/nar/gkv1035>
- Crooks GE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190. <https://doi.org/10.1101/gr.849004>
- Cubas P, Lauter N, Doebley J, Coen E (1999) The TCP domain: a motif found in proteins regulating plant growth and development. *Plant J* 18:215–222. <https://doi.org/10.1046/j.1365-313X.1999.00444.x>
- D’Ario M, Griffiths-Jones S, Kim M (2017) Small RNAs: big impact on plant development. *Trends Plant Sci* 22:1056–1068. <https://doi.org/10.1016/j.tplants.2017.09.009>
- Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res* 39:W155–W159. <https://doi.org/10.1093/nar/gkr319>
- Dai X, Zhuang Z, Zhao PX (2019) psRNATarget V2: a high-performance plant small rna target analysis server. In: plant and animal genome XXVII conference (January 12–16, 2019). PAG
- Dehury B, Panda D, Sahu J et al (2013) In silico identification and characterization of conserved miRNAs and their target genes in sweet potato (*Ipomoea batatas* L.) expressed sequence tags (ESTs). *Plant Signal Behav* 8:e26543. <https://doi.org/10.4161/psb.26543>
- Fahlgrén N, Carrington JC (2010) miRNA target prediction in plants. *Plant microRNAs*. Springer, pp 51–57
- Fileccia V, Bertolini E, Ruisi P et al (2017) Identification and characterization of durum wheat microRNAs in leaf and root tissues. *Funct Integr Genomics* 17:583–598. <https://doi.org/10.1007/s10142-017-0551-2>
- Fileccia V, Ingraffia R, Amato G et al (2019) Identification of microRNAs differentially regulated by water deficit in relation to mycorrhizal treatment in wheat. *Mol Biol Rep* 46:5163–5174. <https://doi.org/10.1007/s11033-019-04974-6>
- Gandikota M, Birkenbihl RP, Höhmann S et al (2007) The miRNA156/157 recognition element in the 3’ UTR of the *Arabidopsis* SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. *Plant J* 49:683–693. <https://doi.org/10.1111/j.1365-313X.2006.02983.x>
- Grabherr MG, Haas BJ, Yassour M et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652. <https://doi.org/10.1038/nbt.1883>
- Harpke D, Meng S, Rutten T et al (2013) Phylogeny of *Crocus* (Iridaceae) based on one chloroplast and two nuclear loci: ancient hybridization and chromosome number evolution. *Mol Phylogenet Evol* 66:617–627. <https://doi.org/10.1016/j.ympev.2012.10.007>

- He B, Zhao S, Chen Y et al (2015) Optimal assembly strategies of transcriptome related to ploidies of eukaryotic organisms. *BMC Genomics* 16:65. <https://doi.org/10.1186/s12864-014-1192-7>
- Hölzer M, Marz M (2019) De novo transcriptome assembly: a comprehensive cross-species comparison of short-read RNA-Seq assemblers. *Gigascience* 8:1–16. <https://doi.org/10.1093/gigascience/giz039>
- Hu J, Liu Y, Tang X et al (2020) Transcriptome profiling of the flowering transition in saffron (*Crocus sativus* L.). *Sci Rep* 10:9680. <https://doi.org/10.1038/s41598-020-66675-6>
- Jain M, Srivastava PL, Verma M et al (2016) De novo transcriptome assembly and comprehensive expression profiling in *Crocus sativus* to gain insights into apocarotenoid biosynthesis. *Sci Rep* 6:22456. <https://doi.org/10.1038/srep22456>
- Jeong D-H, German MA, Rymarquis LA et al (2010) Abiotic stress-associated miRNAs: detection and functional analysis. *Plant microRNAs*. Springer, pp 203–230
- Jike W, Sablok G, Bertorelle G et al (2018) In silico identification and characterization of a diverse subset of conserved microRNAs in bioenergy crop *Arundo donax* L. *Sci Rep* 8:1–13. <https://doi.org/10.1038/s41598-018-34982-8>
- Jung J-H, Seo PJ, Kang SK, Park C-M (2011) miR172 signals are incorporated into the miR156 signaling pathway at the SPL3/4/5 genes in *Arabidopsis* developmental transitions. *Plant Mol Biol* 76:35–45. <https://doi.org/10.1007/s11103-011-9759-z>
- Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 42:D68–D73. <https://doi.org/10.1093/nar/gkt1181>
- Kozomara A, Birgaoanu M, Griffiths-Jones S (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res* 47:D155–D162. <https://doi.org/10.1093/nar/gky1141>
- Kubeczka K-H (2020) History and sources of essential oil research. *Handbook of essential oils*. CRC Press, pp 3–39
- Kumar S, Stecher G, Li M et al (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Kurihara Y, Watanabe Y (2010) Processing of miRNA precursors. *Plant microRNAs*. Springer, pp 231–241
- Kurtoglu KY, Kantar M, Budak H (2014) New wheat microRNA using whole-genome sequence. *Funct Integr Genomics* 14:363–379. <https://doi.org/10.1007/s10142-013-0357-9>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Leinonen R, Sugawara H, Shumway M, Collaboration INSD (2010) The sequence read archive. *Nucleic Acids Res* 39:D19–D21
- Li W-X, Oono Y, Zhu J-KJ et al (2008) The *Arabidopsis* NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. *Plant Cell* 20:2238–2251. <https://doi.org/10.1105/tpc.108.059444>
- Li X-Y, Lin E-P, Huang H-H et al (2018) Molecular characterization of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) gene family in *Betula luminifera*. *Front Plant Sci* 9:608. <https://doi.org/10.3389/fpls.2018.00608>
- Liu Q, Feng Y, Zhu Z (2009) Dicer-like (DCL) proteins in plants. *Funct Integr Genomics* 9:277–286. <https://doi.org/10.1007/s10142-009-0111-5>
- Liu H, Searle IR, Watson-Haigh NS et al (2015) Genome-wide identification of microRNAs in leaves and the developing head of four durum genotypes during water deficit stress. *PLoS ONE* 10:e0142799. <https://doi.org/10.1371/journal.pone.0142799>
- Liu J, Cheng X, Liu P et al (2017) MicroRNA319-regulated TCPs interact with FBHs and PFT1 to activate CO transcription and control flowering time in *Arabidopsis*. *PLoS Genet* 13:e1006833. <https://doi.org/10.1371/journal.pgen.1006833>
- Liu T, Yu S, Xu Z et al (2020) Prospects and progress on crocin biosynthetic pathway and metabolic engineering. *Comput Struct Biotechnol J* 18:3278–3286. <https://doi.org/10.1016/j.csbj.2020.10.019>
- Lucas SJ, Budak H (2012) Sorting the wheat from the chaff: identifying miRNAs in genomic survey sequences of *Triticum aestivum* chromosome 1AL. *PLoS ONE* 7:e40859. <https://doi.org/10.1371/journal.pone.0040859>
- Mamrot J, Legaie R, Ellery SJ et al (2017) De novo transcriptome assembly for the spiny mouse (*Acomys cahirinus*). *Sci Rep* 7:8996. <https://doi.org/10.1038/s41598-017-09334-7>
- Martinelli F, Cannarozzi G, Balan B et al (2018) Identification of miRNAs linked with the drought response of tef [*Eragrostis tef* (Zucc.) Trotter]. *J Plant Physiol* 224–225:163–172. <https://doi.org/10.1016/j.jplph.2018.02.011>
- Moazzam Jazi M, Seyedi SM, Ebrahimie E et al (2017) A genome-wide transcriptome map of pistachio (*Pistacia vera* L.) provides novel insights into salinity-related genes and marker discovery. *BMC Genomics* 18:627. <https://doi.org/10.1186/s12864-017-3989-7>
- Moreton J, Dunham SP, Emes RD (2014) A consensus approach to vertebrate de novo transcriptome assembly from RNA-seq data: assembly of the duck (*Anas platyrhynchos*) transcriptome. *Front Genet* 5:190. <https://doi.org/10.3389/fgene.2014.00190>
- Moriya Y, Itoh M, Okuda S et al (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35:W182–W185. <https://doi.org/10.1093/nar/gkm321>
- Mousavi S, Alisoltani A, Shiran B et al (2014) De novo transcriptome assembly and comparative analysis of differentially expressed genes in *Prunus dulcis* Mill. in response to freezing stress. *PLoS ONE* 9:e104541. <https://doi.org/10.1371/journal.pone.0104541>
- Nakashima K, Jan A, Todaka D et al (2014) Comparative functional analysis of six drought-responsive promoters in transgenic rice. *Planta* 239:47–60. <https://doi.org/10.1007/s00425-013-1960-7>
- Nazarov PV, Reinsbach SE, Muller A et al (2013) Interplay of microRNAs, transcription factors and target genes: linking dynamic expression changes to function. *Nucleic Acids Res* 41:2817–2831. <https://doi.org/10.1093/nar/gks1471>
- Patro R, Duggal G, Love MI et al (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14:417–419. <https://doi.org/10.1038/nmeth.4197>
- Petijová L, Jurčáková Z, Čellárová E (2020) Computational screening of miRNAs and their targets in leaves of *Hypericum* spp. by transcriptome-mining: a pilot study. *Planta* 251:49. <https://doi.org/10.1007/s00425-020-03342-0>
- Robertson G, Schein J, Chiu R et al (2010) De novo assembly and analysis of RNA-seq data. *Nat Methods* 7:909–912. <https://doi.org/10.1038/nmeth.1517>
- Roy S, Nath D, Paul P, Chakraborty S (2020) Computational identification of conserved microRNAs and functional annotation of their target genes in Citrus limon. *S Afr J Bot* 130:109–116. <https://doi.org/10.1016/j.sajb.2019.12.009>
- Sabzehzari M, Naghavi MR (2019) Phyto-miRNAs-based regulation of metabolites biosynthesis in medicinal plants. *Gene* 682:13–24. <https://doi.org/10.1016/j.gene.2018.09.049>
- Samad AFA, Sajad M, Nazaruddin N et al (2017) MicroRNA and transcription factor: key players in plant regulatory network. *Front Plant Sci* 8:565. <https://doi.org/10.3389/fpls.2017.00565>
- Samad AFA, Rahnamaie-Tajadod R, Sajad M et al (2019) Regulation of terpenoid biosynthesis by miRNA in *Piscaria minor* induced by *Fusarium oxysporum*. *BMC Genomics* 20:1–22
- Sarvepalli K, Nath U (2011) Hyper-activation of the TCP4 transcription factor in *Arabidopsis thaliana* accelerates multiple aspects of plant maturation. *Plant J* 67:595–607. <https://doi.org/10.1111/j.1365-3113.2011.04616.x>

- Schwab R, Palatnik JF, Riester M et al (2005) Specific effects of microRNAs on the plant transcriptome. *Dev Cell* 8:517–527. <https://doi.org/10.1016/j.devcel.2005.01.018>
- Seppy M, Manni M, Zdobnov EM (2019) BUSCO: assessing genome assembly and annotation completeness. *Gene prediction*. Springer, pp 227–245
- Shannon P (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>
- Simão FA, Waterhouse RM, Ioannidis P et al (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Smith-Unna R, Bournsnel C, Patro R et al (2016) TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res* 26:1134–1144. <https://doi.org/10.1101/gr.196469.115>
- Srivastava PK, Moturu T, Pandey P et al (2014) A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. *BMC Genomics* 15:348. <https://doi.org/10.1186/1471-2164-15-348>
- Sunkar R, Zhou X, Zheng Y et al (2008) Identification of novel and candidate miRNAs in rice by high throughput sequencing. *BMC Plant Biol* 8:25. <https://doi.org/10.1186/1471-2229-8-25>
- Taheri-Dehkordi A, Naderi R, Martinelli F, Salami SA (2020) A robust workflow for indirect somatic embryogenesis and cormllet production in saffron (*Crocus sativus* L.) and its wild allies *C. caspius* and *C. speciosus*. *Heliyon* 6:e05841. <https://doi.org/10.1016/j.heliyon.2020.e05841>
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>
- Tan H, Chen X, Liang N et al (2019) Transcriptome analysis reveals novel enzymes for apo-carotenoid biosynthesis in saffron and allows construction of a pathway for crocetin synthesis in yeast. *J Exp Bot* 70:4819–4834. <https://doi.org/10.1093/jxb/erz211>
- Tang R, Li L, Zhu D et al (2012) Mouse miRNA-709 directly regulates miRNA-15a/16-1 biogenesis at the posttranscriptional level in the nucleus: evidence for a microRNA hierarchy system. *Cell Res* 22:504–515. <https://doi.org/10.1038/cr.2011.137>
- Tarantilis PA, Tsoupras G, Polissiou M (1995) Determination of saffron (*Crocus sativus* L.) components in crude plant extract using high-performance liquid chromatography-UV-visible photodiode-array detection-mass spectrometry. *J Chromatogr A* 699:107–118. [https://doi.org/10.1016/0021-9673\(95\)00044-N](https://doi.org/10.1016/0021-9673(95)00044-N)
- Tholl D (2015) Biosynthesis and biological functions of terpenoids in plants. *Biotechnology of isoprenoids*. Springer, pp 63–106
- Vahedi M, Kabiri M, Salami SA et al (2018) Quantitative HPLC-based metabolomics of some Iranian saffron (*Crocus sativus* L.) accessions. *Ind Crops Prod* 118:26–29. <https://doi.org/10.1016/j.indcrop.2018.03.024>
- Verma P, Singh N, Khan SA et al (2020) TIAs pathway genes and associated miRNA identification in *Vinca* minor: supporting aspido-perma and eburnamine alkaloids linkage via transcriptomic analysis. *Physiol Mol Biol Plants* 26:1695–1711. <https://doi.org/10.1007/s12298-020-00842-x>
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63. <https://doi.org/10.1038/nrg2484>
- Wang J-W, Park MY, Wang L-J et al (2011) MiRNA control of vegetative phase change in trees. *PLoS Genet* 7:e1002012. <https://doi.org/10.1371/journal.pgen.1002012>
- Wang L, Liu N, Wang T et al (2018) The GhmiR157a–GhSPL10 regulatory module controls initial cellular dedifferentiation and callus proliferation in cotton by modulating ethylene-mediated flavonoid biosynthesis. *J Exp Bot* 69:1081–1093. <https://doi.org/10.1093/jxb/erx475>
- Waterhouse RM, Seppy M, Simão FA et al (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* 35:543–548
- Wei R, Qiu D, Wilson IW et al (2015) Identification of novel and conserved microRNAs in *Panax notoginseng* roots by high-throughput sequencing. *BMC Genomics* 16:835. <https://doi.org/10.1186/s12864-015-2010-6>
- Wu L, Zhang Q, Zhou H et al (2009) Rice microRNA effector complexes and targets. *Plant Cell* 21:3421–3435. <https://doi.org/10.1105/tpc.109.070938>
- Wu H-J, Ma Y-K, Chen T et al (2012) PsRobot: a web-based plant small RNA meta-analysis toolbox. *Nucleic Acids Res* 40:W22–W28. <https://doi.org/10.1093/nar/gks554>
- Xin M, Wang Y, Yao Y et al (2010) Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (*Triticum aestivum* L.). *BMC Plant Biol* 10:123. <https://doi.org/10.1186/1471-2229-10-123>
- Ye J, Fang L, Zheng H et al (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res* 34:W293–W297. <https://doi.org/10.1093/nar/gkl031>
- Ye J, Zhang X, Tan J et al (2020) Global identification of *Ginkgo biloba* microRNAs and insight into their role in metabolism regulatory network of terpene trilactones by high-throughput sequencing and degradome analysis. *Ind Crops Prod* 148:112289. <https://doi.org/10.1016/j.indcrop.2020.112289>
- Yu B (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science* 307:932–935. <https://doi.org/10.1126/science.1107130>
- Zakeel MCM, Safeena MIS, Komathy T (2019) In silico identification of microRNAs and their target genes in watermelon (*Citrullus lanatus*). *Sci Hortic (amsterdam)* 252:55–60. <https://doi.org/10.1016/j.scienta.2019.02.012>
- Zhang B, Pan X, Cannon CH et al (2006a) Conservation and divergence of plant microRNA genes. *Plant J* 46:243–259. <https://doi.org/10.1111/j.1365-3113X.2006.02697.x>
- Zhang B, Pan X, Cobb GP, Anderson TA (2006b) Plant microRNA: a small regulatory molecule with big impact. *Dev Biol* 289:3–16. <https://doi.org/10.1016/j.ydbio.2005.10.036>
- Zhang BH, Pan XP, Cox SB et al (2006c) Evidence that miRNAs are different from other RNAs. *Cell Mol Life Sci C* 63:246–254. <https://doi.org/10.1007/s11033-011-1004-y>
- Zheng Y, Jiao C, Sun H et al (2016) iTAK: a Program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol Plant* 9:1667–1670. <https://doi.org/10.1016/j.molp.2016.09.014>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.