



OPEN

# Identification and the potential involvement of miRNAs in the regulation of artemisinin biosynthesis in *A. annua*

Shazia Khan, Athar Ali, Monica Saifi, Parul Saxena, Seema Ahlawat &amp; Malik Zainul Abdin✉

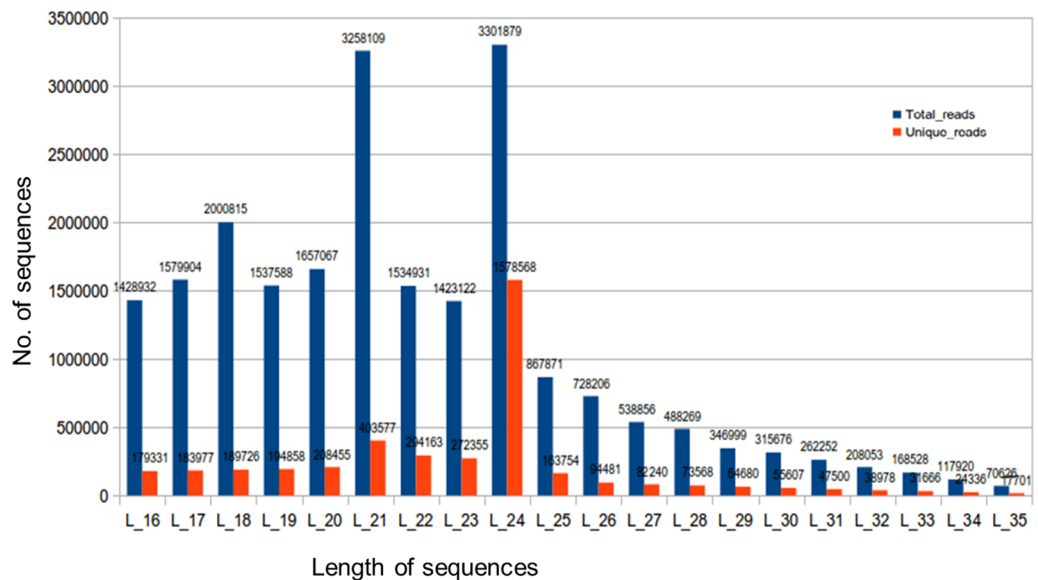
Micro RNAs (miRNAs) play crucial regulatory roles in multiple biological processes. Recently they have garnered the attention for their strong influence on the secondary metabolite production in plants. Their role in the regulation of artemisinin (ART) biosynthesis is, however, not fully elucidated. ART is a potent anti-malarial compound recommended by WHO for the treatment of drug-resistant malaria. It is produced by *Artemisia annua* (*A. annua*). The lower *in planta* content of ART necessitates a deep understanding of regulatory mechanisms involved in the biosynthesis of this metabolite. In this study, using modern high throughput small RNA-sequencing by Illumina Nextseq 500 platform for identification and stem-loop RT PCR for validation, miRNAs were identified in the leaf sample of *A. annua* plant. Here, we report a total of 121 miRNAs from *A. annua* that target several important genes and transcription factors involved in the biosynthesis of ART. This study revealed the presence of some important conserved miRNA families, miR396, miR319, miR399, miR858, miR5083 and miR6111 not identified so far in *A. annua*. The expression patterns and correlation between miRNAs and their corresponding targets at different developmental stages of the plant using real-time PCR indicate that they may influence ART accumulation. These findings thus, open new possibilities for the rational engineering of the secondary metabolite pathways in general and ART biosynthesis in particular.

## Abbreviations

ART	Artemisinin
<i>A. annua</i>	<i>Artemisia annua</i>
miRNA	MicroRNA
TF	Transcription factor

*Artemisia annua* (*A. annua*), an important medicinal herb is the only natural source of artemisinin (ART). ART, a sesquiterpene lactone is the most reliable and widely used drug to quell Malaria, which is one of the most devastating diseases and a leading cause of deaths worldwide. According to WHO, there were a staggering 228 million cases and 405,000 deaths reported in 2018 alone, despite the disease being preventable as well as curable<sup>1</sup>. WHO has recommended the use of ART only in combination of other drugs, to slow down the development of resistance in *Plasmodium* sps. against it and to maintain its efficacy. The combination treatment is widely known as artemisinin-based combination therapies (ACTs). In wild type *A. annua*, ART is present in the aerial parts, in extremely low quantities (0.02–1.07%)<sup>2</sup>. Numerous efforts were made to generate ART synthetically or semi-synthetically, but does not meet industrial practices due to high production cost. The insufficient supply of ART produced by plants or synthetic methods, resulted in the lack of ACTs available to the patients and causes approximately half a million mortalities every year<sup>3</sup>. Despite various efforts to produce semi-synthetic ART at low cost, the natural source of ART remains the main provider. Thus, researchers are investigating effective strategies to enhance *in planta* content of ART. The Ian Graham lab and Bill and Malinda gates foundation invested huge fund in production of high ART yielding hybrids and succeeded in development of HYB8001R that could produce 1.4% ART of dry weight<sup>4</sup>. Reconstitution of the pathway and increasing the area under cultivation could be a promising solution to increase ART production. Ding et al., 2020 suggested the potential fields

Department of Biotechnology, School of Chemical and Life Sciences, Centre for Transgenic Plant Development, Jamia Hamdard, New Delhi 110062, India. ✉email: mzabdin@jamiahamdard.ac.in



**Figure 1.** Length distribution graph of small RNA sequences in leaf sample library of *A. annua*.

suited best for the cultivation of *A. annua* plant worldwide. They found that contrary to the higher demands in African region the potential fields are minimal. The best suited land resources were predicted to be distributed around the south eastern coast of Brazil and Australia<sup>5</sup>. Thus, the strong international cooperation is needed to resolve this spatial mismatch.

The reconstitution of the pathway, requires the complete elucidation of the biosynthetic pathway genes and regulatory elements. In plants, the secondary metabolites biosynthesis and accumulation is highly regulated. ART biosynthesis has been extensively studied and it was found that two metabolic pathways; cytoplasmic mevalonic acid (MVA) and plastidial methylerythritol phosphate (MEP) are involved in providing the basic carbon precursors of isoprene units for the biosynthesis of ART. To date, genes encoding enzymes of MVA and MEP pathways such as HMGR, IDI, DXS, DXR, HDR and FPS as well as genes for all five downstream enzymes of ART biosynthesis pathway namely ADS, CYP71AV1/CPR, ADH1, ALDH1, and DBR2 are over-expressed in *A. annua* plants<sup>6–13</sup>. Shen et al. (2018) showed up to 3.2% enhancement in the production of ART in *A. annua* transgenic lines that were simultaneously overexpressing the FPS, HMGR and DBR2 genes<sup>14</sup>. Role of TF in the regulation of complex pathways is well known and cannot be neglected, when a lot of reports are advocating their regulatory role in secondary metabolite biosynthesis. Consequently, TFs were cloned and their effect on ART content was studied. It was shown that AaMYB1 significantly enhance the transcription of key genes of the ART biosynthetic pathway, especially ADS and CYP71AV1, that resulted in an increased ART content in transgenic plants<sup>15</sup>. Similar results were obtained when other TFs such as AaERF1, AabZIP, bHLH, AaWRKY, AaMYC2, AaGSW1, AaORA and AaTAR1 were overexpressed<sup>16–19</sup>. Due to the fine tuning of pathways and probably due to the presence of other key players, these efforts were however, unable to effectively boost up the entire metabolic flux towards ART biosynthesis. In a study, it was however, shown that the overexpression or loss of miR163 alters entire secondary metabolite profile in mir163 mutant of *Arabidopsis*<sup>20</sup>. Also, another study revealed that miR156-targeted SPL9 directly binds terpene synthase promoter and regulate  $\beta$ -caryophyllene synthesis in *Arabidopsis*<sup>21</sup>. Thus, it seems reasonable to assume that miRNAs can also be used as effective regulators for modulating ART biosynthesis based on their targets in ART biosynthetic pathway. Here, we are reporting, miRNAs from *A. annua* and their possible involvement in ART biosynthesis.

## Results

**Sequencing of small RNAs and data analysis.** A high-quality RNA (RIN number 8.4), isolated from the leaves of *A. annua* at pre-flowering stage was subjected to next generation sequencing (NGS), thereby resulted in generation of 36,392,552 raw reads. As a result of further analysis, i.e. selection of size, trimming of adapters and low-quality reads, 34,741,778 clean reads (unique reads 5,228,750) were remained. Further removing the other non-coding RNA sequences such as snRNA, siRNA, tRNA, rRNA, snoRNAs resulted in 21,835,603 clean reads (unique reads 4,199,521). Among the clean reads the maximum reads were of 24nt followed by 21nt and then 20, 22 and 23 nt reads. The length distribution graph is given in Fig. 1. The 23 and 24nt sequences possess 5'-adenosine at the first position, generally considered as endogenous siRNAs<sup>22</sup>. These are one of the most abundant classes of small RNAs in plants, while most plant miRNAs tend to be 21 and 22nt in length<sup>23</sup>.

**Identification of *A. annua* conserved and novel miRNAs.** The clean reads obtained were subsequently analysed to predict conserved and novel miRNAs in *A. annua*. As a result, 121 miRNAs were identified, 80 of which were conserved across plant species whereas, 41 candidates were predicted as *A. annua* specific novel miRNAs (supplementary file S1). The conserved miRNAs belong to 23 different families of miRNAs. Varying

S. No	MicroRNA	Abundance (reads per million)
1	miR396f	261,298
2	miR159a	73,683
3	miR166a-3p	47,883
4	miR166m	19,512
5	miR166n	19,240
6	miR396g-5p	11,558
7	miR396b-5p	8,444
8	miR166g-3p	6,714
9	miR162a-3p	5,741
10	miR159a.1	4,930

**Table 1.** Ten most abundant conserved miRNA families in *A. annua* leaf sample library.

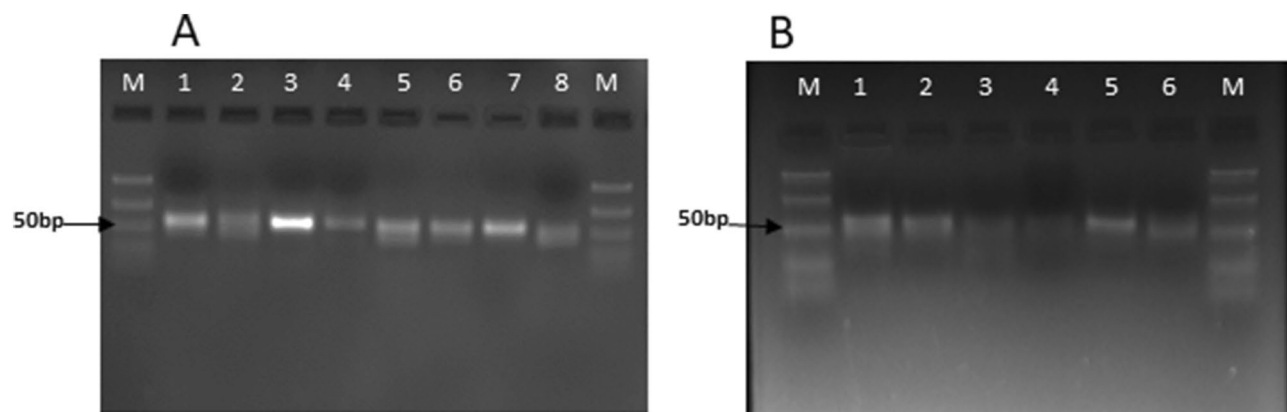
S. No.	miRNA family	No. of members
1	miR166	16
2	miR396	12
3	miR159	9
4	miR162	5
5	miR156	4
6	miR167	4
7	miR319	4
8	miR398	4
9	miR160	2
10	miR164	2
11	miR165	2
12	miR168	2
13	miR172	2
14	miR399	2
15	miR403	2
16	miR170	1
17	miR171	1
18	miR393	1
19	miR394	1
20	miR858	1
21	miR5083	1
22	miR5368	1
23	miR6111	1

**Table 2.** No. of members in conserved miRNA families in *A. annua*.

level of miRNA expression was found in the library. Out of all, the maximum expression was observed for miRNA396f, followed by miRNA159a with 261,298 and 73,683 reads, respectively. The ten most abundant miRNAs in the sample are shown in Table 1.

A total of 248 sequences of 18–22 nt were also identified that were either 1-2nt shorter or longer than miRNA in miRbase-21 (supplementary file S1). In our data, the highest number of miRNAs belongs to miR166 and miR396 families having 16 members and 12 members respectively. While, families miR170, miR171, miR393, miR394, miR858, miR5083, miR5368, and miR6111 were found to have only one member. Number of members in each family are summarized in Table 2.

Furthermore, a total of 41 sequences were successfully identified as novel miRNA candidates in *A. annua*. These sequences have a length ranging from 18 to 24 bp, but the maximum among them were 20–22 bp long. The minimum fold energy (MFE) observed was between  $-18.3 \text{ kcal}\cdot\text{mol}^{-1}$  to  $-84.4 \text{ kcal}\cdot\text{mol}^{-1}$ . This showed good agreement with previous reports on novel miRNA predictions in other species (MFE of *Catharanthus roseus*,  $-43.2 \text{ kcal}\cdot\text{mol}^{-1}$ ; Opium poppy (*Papaver somniferum*),  $-34.09 \text{ kcal}\cdot\text{mol}^{-1}$ ; pigeonpea (*Cajanus cajan*),  $-34.8 \text{ kcal}\cdot\text{mol}^{-1}$ ; *Arabidopsis thaliana*, rice (*Oryza sativa*), soybean (*Glycine max*), *Medicago truncatula*, *Saccharum officinarum*, sorghum (*Sorghum bicolor*) and maize (*Zea mays*) from  $-40$  to  $100 \text{ kcal}\cdot\text{mol}^{-1}$ )<sup>24–27</sup>. The minimum fold energy index (MFEI) is an important criterion to characterize miRNAs. When the MFEI is more



**Figure 2.** Agarose gel based detection of miRNAs (A) conserved miRNAs M:50 bp ladder, 1: miR5083, 2: miR858b, 3: miR159a, 4: miR396, 5: miR172a-3p, 6: miR166i, 7: miR166g, 8: miR162 (B) novel miRNAs M: 50 bp ladder, 1:miRn-7, 2: miRn-10, 3: miRn-11, 4: miRn-18, 5:miRn-29, 6: miRn-26.

than 0.85, the sequence is most likely to be miRNA. It was observed that the MFEIs of stem loop structures of the miRNAs were ranged from 0.87 to 2.1. All the information along with precursor sequences is given in supplementary file S1.

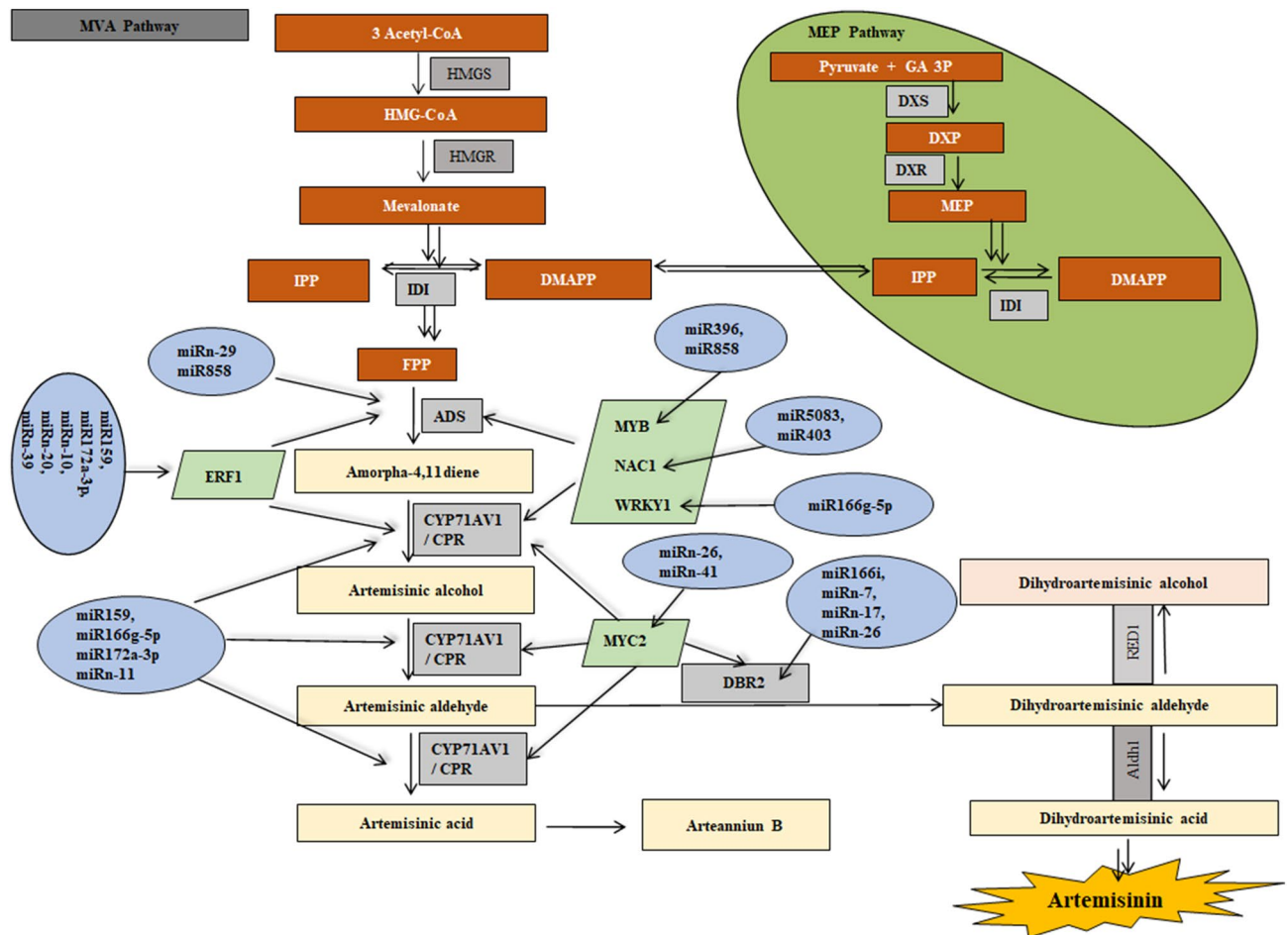
**Validation of conserved and novel miRNAs.** The miRNAs obtained by NGS were validated using stem-loop PCR according to published protocol<sup>28</sup>. Of the 15 selected miRNAs (8 conserved and 7 novel miRNAs), we confirmed the expression of 14 miRNAs (Fig. 2).

To further strengthen our results, all these miRNAs were also checked at different developmental stages of the plant. All these miRNAs were successfully amplified in all the stages. Thus, the results obtained in this experiment and the data obtained by NGS present enough support for the existence of these miRNAs in *A. annua*.

**Target prediction.** As a result of target prediction, 54 unique targets of 52 conserved miRNAs, and more than 60 targets of 30 novel miRNAs were successfully identified. Results are given in supplementary file S1. Target interpretation was mainly focused on their putative roles in ART biosynthesis. Multiple miRNAs were found to target single mRNA. For instance, three families of miRNAs, namely miR159, miR166, miR172, and four novel miRNAs namely miRn-8, miRn-20, miRn-36 and miRn-37 putatively targets cytochrome p450 reductase (CPR) mRNA. CPR is a redox partner of CYP71AV1, an important enzyme in ART biosynthesis. Another important miRNA family is miR396 family. The three members of this family, viz. miR396, miR396f and miR396g-5p targets DXR1, a rate limiting enzyme of plastidial MEP pathway. Similarly, other enzymes of MVA and MEP pathways such as HMGR, IDI, DXS, HDR and FPP are targeted by several miRNAs. The details are given in supplementary file S1. These enzymes play an important role in producing basic isoprene units for production of all terpenoids. It was observed that miR858 putatively target transcript for the enzyme ADS. Other significant miRNAs were miRn-11, miRn-17, miRn-18, that targets CYP71AV1 mRNA. Both of the enzymes are key players in ART biosynthesis and thus, these miRNAs could play crucial roles in the ART biosynthesis regulation. In addition to above mentioned miRNAs, we have identified miRNAs that targets TFs. We observed that miR858 targets transcription factor family MYB. In *A. annua*, MYB1 is involved in ART biosynthesis regulation<sup>15</sup>. It was found that in *Arabidopsis*, miRNA858 targets several members of MYB family that regulate the synthesis of flavonoids<sup>29</sup>. Consistent with these reports, miR858 in *A. annua* targets various members of MYB. In addition, miR858 was also predicted to target ADS transcript. ADS is one of the key enzymes in ART biosynthesis regulation. Another important TF, NAC1 is a putative target of miRNA families 403 and 5,083. In literature, NAC1 is reported to regulate two key genes ADS and CYP71AV1 of ART biosynthesis pathway. WRKY1 the third TF putatively targeted by miR166g is also reported to drag ART biosynthesis pathway in positive direction and help in accumulation of ART by regulating pathway genes. Similarly, MYC2 is the putative target of miRn-26 and miRn-41. Apart from targeting transcripts of gene, few miRNAs targets promoter region of the genes. This includes miR159, miR172a-3p, miR166i, miRn-7, miRn-10, miRn-24, and miRn-29. Of which miR166i and miRn-7 target DBR2 gene promoter; miR159, miR172a-3p and miRn-10 target ERF1 gene promoter while miRn-24 and miRn-29 target promoters of WRKY1 and ADS genes, respectively.

Additionally, we have found that the mRNAs of  $\beta$ -amyrin synthase,  $\beta$ -farnesene synthase, and linalool synthase, which are involved in the formation of other terpene molecules are targeted by some conserved and *A. annua* specific novel miRNAs listed in supplementary file S1. Many researchers showed that blocking the pathways involved in biosynthesis of other terpene molecules, redirect the carbon flux towards ART biosynthesis in *A. annua*. Some important miRNAs involve in ART biosynthesis were given in Fig. 3. The detailed target prediction data is available in supplementary file S1.

**Differential expression and correlation of miRNAs and their putative targets.** The expressions of six miRNAs, viz. miR858, miR5083, miR159, miRn-7, miRn-10, miRn-29 and their putative target genes evaluated at different developmental stages viz., vegetative, pre-flowering and flowering stages of the plant using

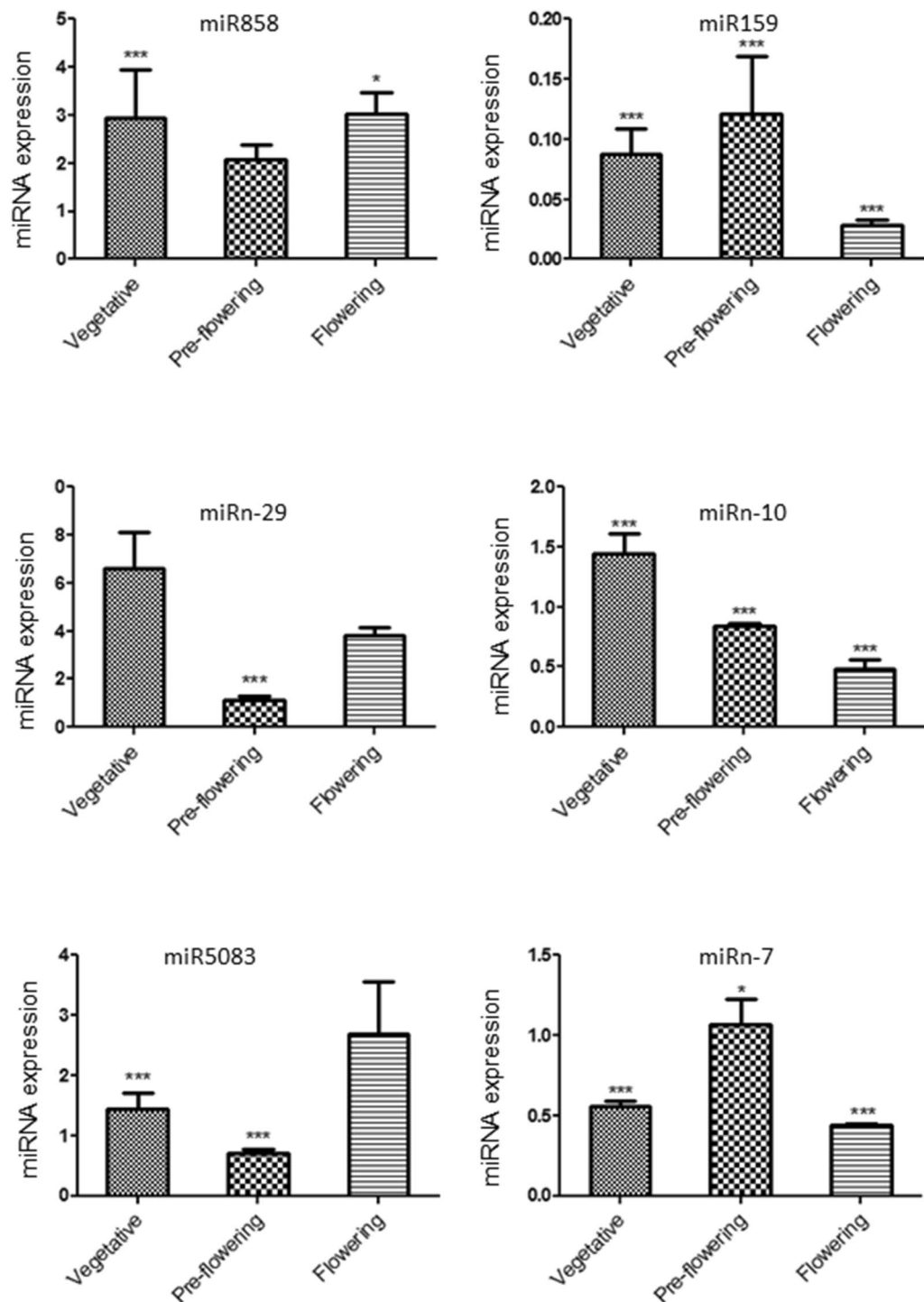


**Figure 3.** ART biosynthesis and putative involvement of miRNAs. The MVA pathway and MEP pathway products are shown in orange boxes. *HMG-CoA* hydroxymethylglutaryl CoA, *IPP* isopentenyl diphosphate, *DMAPP* dimethylallyl diphosphate, *FPP* farnesyl diphosphate, *DXP* 1-deoxy-D-xylulose 5-phosphate, *MEP* methyl erythritol phosphate, enzymes are indicated in grey boxes. *HMGs* HMG-CoA synthase, *HMGR* HMG-CoA reductase, *IDI* IPP isomerase, *DXS* 1-deoxy-D-xylulose 5-phosphate synthase, *DXR* DXP reductoisomerase, *ADS* amorpha-4, 11-diene synthase, *CYP71AV1 / CPR* cytochrome P450 mono-oxygenase, *CPR* cytochrome P450 reductase, *DBR2* artemisinic aldehyde delta-11(13)-double bond reductase, *ALDH1* aldehyde dehydrogenase, *RED1* dihydroartemisinic aldehyde reductase, transcription factors are indicated in green boxes and miRNAs are indicated in blue circles.

qRT-PCR are shown in Figs. 4 and 5. Results showed that miR858 was negatively correlated with its targets MYB1 and ADS (Pearson coefficient  $-0.52$ ,  $-0.50$ ). Similarly, miR159 was negatively correlated with its target CPR (Pearson coefficient  $-0.68$ ). Interestingly, the expression of miR5083 was positively correlated with its target NAC1 (Pearson coefficient  $+0.4$ ). Of the four miRNAs viz. miR159, miRn-7, miRn-10, and miRn-29 that target promoter regions of ERF1, DBR2 and ADS genes, the expression levels of miRn-7 had a positive correlation with its target gene DBR2 expression (Pearson coefficient  $+0.39$ ). The expressions of miRn-10, miR159 and miRn-29 were not found to be correlated with the expression of their target genes.

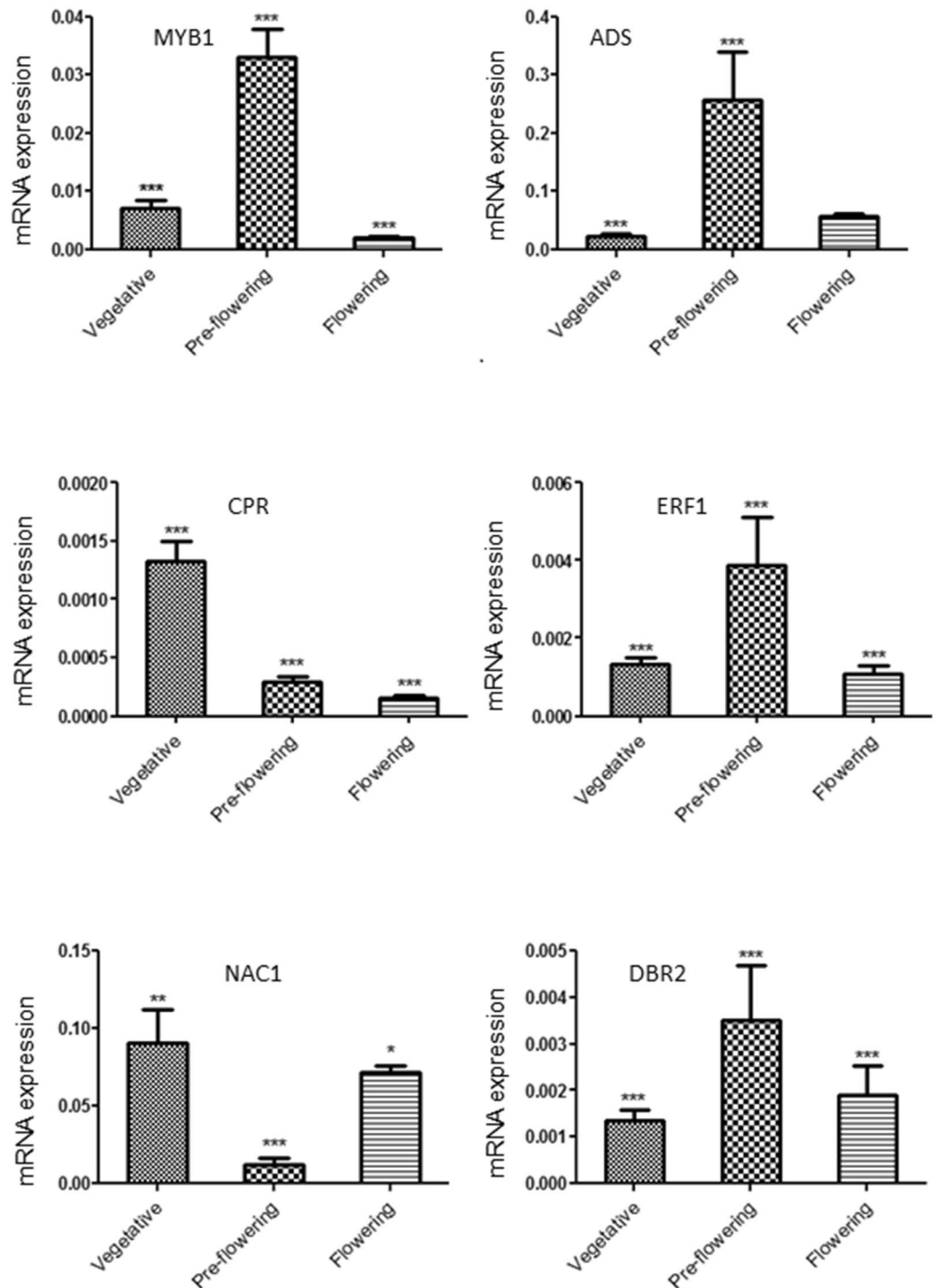
## Discussion

Micro RNAs are small non coding RNAs that exert extensive impact on the regulation of gene expression in both plants and animals. In recent years, they have garnered attention due to their involvement in a number of complex biological processes including the secondary metabolite production in plants<sup>30</sup>. Their possible involvement in ART biosynthesis is however, yet to be elucidated. Previous three reports on miRNAs identification and target prediction in *A. annua* were based on computational identification and predicted only six, thirteen and eleven miRNAs, respectively that may represent only a fraction of its miRNAome<sup>31-33</sup>. Here, we sought to identify and validate miRNAs present in *A. annua* plant using NGS along with qPCR. Modern deep sequencing/NGS has many advantages over computational prediction specially in species that do not have full genome sequences. A total of 121 miRNAs were identified that include some important conserved miRNAs like miR396, miR319, miR399, miR858, miR5083 and miR6111, which were not identified so far in *A. annua*. It was recently shown that miR319 regulate TCP4 proteins, involved in suppression of trichome branching in *Arabidopsis thaliana*<sup>34</sup>. Trichomes are factories for synthesis and accumulation of most secondary metabolites including ART<sup>35,36</sup>.



**Figure 4.** Relative expression of selected miRNAs levels quantified by real-time PCR at vegetative, pre-flowering and flowering stages of the *A. annua* plant. The expression levels of miRNAs were normalized to the expression level of U6 RNA. Relative expression was calculated using  $2^{\Delta\Delta Ct}$  equation, where  $Ct = (Ct_{miRNA} - Ct_{U6})$ . ANOVA followed by Dunnett's test was applied to all miRNAs compared to U6. The error bar shows the standard error. \* $P < 0.05$ , \*\*\* $P < 0.001$ .

Besides, reports on miR858 family mediated regulation of secondary metabolites are accumulating in model as well as non-model plants<sup>37,38</sup>. Thus, identification of these new classes of miRNAs may provide new directions to unveil the regulatory mechanisms controlling secondary metabolites biosynthesis and accumulation, and for rational engineering of ART biosynthetic pathway. NGS also allowed us to predict 41 novel miRNAs, based on



**Figure 5.** Relative expression of selected mRNAs levels at vegetative, pre-flowering and flowering stages of the *A. annua* plant. The expression levels of mRNAs were normalized to the expression level of  $\beta$ -actin. Relative expression was calculated using  $2^{\Delta\Delta Ct}$  equation, where  $Ct = (Ct_{\text{gene}} - Ct_{\beta\text{-actin}})$ . ANOVA followed by Dunnett's test was applied to all mRNAs compared to  $\beta$ -actin. The error bar shows the standard error. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

their hairpin structures. Expression of *A. annua* specific novel miRNAs were lower as compared to conserved miRNAs, which is often the case of species-specific miRNAs<sup>22,39</sup>. Validation of fifteen miRNAs by stem-loop RT-PCR at different developmental stages of the plant further strengthen our NGS data.

Interestingly, some of the miRNAs target multiple genes, for instance, miR396e had eight targets while others like miR159a, miR159b, miR159c and miR159f were found to target the mRNA of the same gene. These results exhibit the complex miRNA-target network in *A. annua*.

This study was focused on identification of miRNAs involved in the regulation of ART biosynthesis. Results showed that three families of conserved miRNAs, viz. miR159, miR172 and miR166 were targeting a *cytochrome P450 reductase (CPR)* gene. CPR, along with CYP71AV1 catalyzes three steps of ART biosynthesis, oxidizing amorpho 4,11 diene to artemisinic acid<sup>40,41</sup>. qPCR data suggested that higher expression of miR159 is negatively correlated with target CPR gene expression (Pearson coefficient -0.68) might play a role in the repression of CPR gene at the pre-flowering stage. Another significant miRNA, miR858 targets ADS mRNA. ADS is an important enzyme in ART biosynthesis regulation. Expression of miR858 at different developmental stages is negatively correlated with ADS gene expression (Pearson coefficient - 0.50). miR858 was also targeted MYB TF as discussed later in this section. Besides, we also identified the nine *A. annua* specific miRNAs targeting important genes in ART biosynthesis.

Some of the TFs and miRNAs, while independently regulating their targets, collaborate with each other to regulate gene expression. Literature showed the involvement of at least seven families of TFs in ART biosynthesis that include AP2/ERF, bZip, bHLH, NAC, WRKY, MYC and MYB families<sup>15,16,42,43</sup>. In our results, it was found that miR858 targets several MYB TFs. The qPCR results revealed that the miR858 accumulation is negatively correlated with that of MYB1 (Pearson coefficient - 0.52), when checked at different developmental stages of the plant. Hernández et al.<sup>15</sup> predicted the binding sites of R2R3 type of AaMYB1 in ADS, CYP71AV1, DBR2 promoters. They showed that the content of ART doubles in transgenic lines when AaMYB1 was cloned and overexpressed. Evidences on regulation of MYBs by miR858 are accumulating in plants such as *Arabidopsis*, *Actinidia arguta* (kiwifruit), *Vitis vinifera* (grapes), and *Ipomoea batatas* (sweetpotato), where they play key roles in the synthesis of secondary metabolites like anthocyanin and flavonoids<sup>37,38,44,45</sup>. It was also found that miR858 is involved in fungal susceptibility, and inhibiting miR858 by miRNA mimic technique trigger fungal resistance in *Arabidopsis thaliana*<sup>46</sup>. We speculate that miR858 may play similar roles in *A. annua*. NAC1, another important TF involved in ART biosynthesis was found to be the target of two families of conserved miRNAs, miR403 and miR5083. Most of the plant miRNAs exert negative regulation on their target genes. There are findings where their expressions are however, positively correlated. For example, Kawashima et al.<sup>47</sup> revealed that the expressions of miR395 and its target gene, sulphate transporter (SULTR2;1) were positively correlated in the roots of *Arabidopsis* plants. Thus, it was interesting to discover a positive correlation in the expression of miR5083 with its target NAC1 gene (Pearson coefficient + 0.4). Similar kind of correlations were also reported in other living systems, suggesting that this is not restricted only to plants<sup>48</sup>. miR166g targets AaWRKY1 that regulates ART biosynthesis by binding to the W box in the ADS promoter<sup>49</sup>. It was shown that transgenic lines overexpressing AaWRKY1 lead to accumulate 1.9-fold higher ART content<sup>50</sup>. Similarly, TF AaERF1 was the putative target of miR159, miR172a-3p, miRn-10, miRn-20, and miRn-39, while two miRNAs, miRn-26 and miRn-41 could bind MYC2. We speculate that it will be more rational to modulate the TFs that regulate the multiple genes of ART pathway, by targeting miRNAs than to overexpress individual gene to enhance ART content.

Target prediction has led us to identify few miRNAs that target promoter regions of the important genes involved in ART biosynthesis. One of such miRNAs is miRn-7 that targets DBR2 gene promoter region. Induction of DBR2 gene leads to higher accumulation of ART<sup>10</sup>. Similarly, miRNA159, miR172a-3p, miR166i, miRn-24 and miRn-29 target the promoter regions of ERF1, DBR2 WRKY1, and ADS genes. It was observed that miRNAs that target the promoters of genes, activate their expression at the transcriptional level. For instance, miR5658 directly activates AT3G25290 expression by targeting its promoter in *Arabidopsis thaliana*<sup>51</sup>. Similarly, Place et al.<sup>52</sup> reported that miR-373 induces the expression of two genes, cold-shock domain-containing protein C2 (CSDC2) and E-cadherin by binding the promoters in PC-3 prostate cancer cells<sup>52</sup>. We observed that the expression of miRn-7 is positively correlated with the expression of its target gene DBR2 (Pearson coefficient + 0.39) at different developmental stages of *A. annua* plant. We presume that it could play important regulatory role in ART biosynthesis. The expression of miRn-29, miRn-10 and miR159 were not correlated with the expression of their target genes (Pearson coefficient < 0.3).

Apart from miRNAs targeting ART biosynthesis, it was interesting to discover microRNAs that target genes involved in biosynthetic pathways of other secondary metabolites. These pathways interact with ART pathway while competing for the common precursors generated by MVA pathway. We found that one conserved miRNA, miR162 was putatively targeting chalcone synthase mRNA while five novel miRNAs namely, miRn-1, miRn-4, miRn-18, miRn-20, miRn-26 were targeting chrysanthemyl synthase, cinnamyl alcohol dehydrogenase,  $\beta$ -amyrin synthase, linalool synthase and  $\beta$ -farnesene synthase mRNAs, respectively. Researchers showed that by inhibiting the pathways competing for the common precursors lead to divert the carbon flux towards ART biosynthesis and resulted in accumulation of higher ART in transgenic *A. annua* plants<sup>53,54</sup>. Further studies would however, be required for assessing the impact of these miRNAs on secondary metabolite profile. Thus, this study provides a new insight into the possible miRNA-directed molecular processes that regulate ART biosynthesis in *A. annua*.

## Methods

To identify miRNAs in the wild type *A. annua*, 'cim-arogya' variety was grown in the herbal garden of Jamia Hamdard, New Delhi, India. The seeds of these plants were obtained from Ipca Laboratories Pvt Ltd, M.P., India. It was a high artemisinin producing variety and contains 0.9 to 1.1% of artemisinin content<sup>55</sup>. Plant leaf samples at pre flowering stage were collected for RNA isolation.

**RNA extraction.** For RNA extraction, leaf samples were grounded with mortar and pestle in liquid nitrogen. RNA was isolated using TRIzol reagent (Invitrogen, USA) as per manufacturer's instructions. Total RNA integrity was checked on 1% formaldehyde gel and concentration of RNA was determined using nanodrop electrophotometer (Thermo Fischer Scientific, USA) while the RNA quality was checked on the BioAnalyzer (Agilent, USA).



**Library preparation and sequencing.** RNA sample were sent for high throughput sequencing to M/S Sandor Lifesciences Pvt Ltd, Hyderabad, India. Small RNA library was generated using Illumina NextSeq 500 platform. NEB adaptor having sequence 'AGATCGGAAGAGCACACGTCT' was added for sequencing. Redundant reads were counted and reads per million (RPM) values were calculated. Low quality reads were removed before processing data. Raw reads were then processed, and adapter sequences were trimmed using Cutadapt tool. Remaining reads were then searched against Rfam database to eliminate other non-coding RNAs like rRNA, tRNA, snRNA and snoRNAs.

**Identification of *A. annua* miRNAs.** The clean reads thus obtained were used to predict conserved and novel miRNAs. For conserved miRNA prediction clean reads were aligned to miRBase-21 (<https://www.mirbase.org/>) allowing up to 2 mismatches. The unaligned sequences left were used to predict novel miRNAs. At the time of sequencing the sRNAs, the genome of *A. annua* was not available. Although, it has been assembled in 2018, but the unavailability of complete annotation limits its application as good reference genome. Thus, for novel miRNA prediction, unique sRNA sequences were mapped to the EST sequences of *A. annua* and transcriptome of *A. annua* submitted to the National Center for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov/>). Mature as well as stem-loop structures were predicted for novel miRNAs using mireap\_0.2 software. The criteria used were as follows: (i) The putative mature miRNAs were required to have length between 18 and 24 bases and at least have reads per million (RPM) > 2; (ii) There were no more than five mismatches and four bulges between mature miRNA candidate and the opposite hairpin arm; and (iii) the upper limit of minimum fold energy (MFE) of miRNAs was set to  $-18 \text{ kcal}\cdot\text{mol}^{-1}$ . Hairpin loops of all novel miRNAs were constructed by mfold web server using RNA folding form.

**Target search of *A. annua* miRNAs.** The functional annotation of miRNAs was obtained using psRNATarget<sup>56</sup>, one of the most reliable and widely used tool to predict the targets gene of the miRNAs in plants with default parameters. *A. annua* nucleotide sequences from NCBI database (<https://www.ncbi.nlm.nih.gov/unigene>) were taken as data set for predicting the targets.

**Stem-loop RT-PCR.** Results obtained from the next generation sequencing (NGS) were validated using stemloop RT-PCR, one of the most commonly used and reliable method for determination of miRNAs. Fifteen miRNAs were selected for validation by stemloop RT-PCR according to the published protocol<sup>28</sup>. The basis of selection of miRNAs was their targets in the artemisinin biosynthesis pathway. Amplification was obtained using standard PCR. Stemloop RT primers were designed for each of the fifteen miRNAs for reverse transcription and cDNA was prepared using RevertAid cDNA synthesis kit (Thermo Fischer scientific, USA). Reverse transcription reaction mixture was kept in thermal cycler (Thermo fischer scientific) under following condition: 1 cycle of 16 °C for 30 min; 60 cycles of 30 °C for 30 s, 42 °C for 30 s, 50 °C for 1 s, and 1 cycle of 70 °C for 15 min as described by Gasic et al.<sup>28,57</sup>.

**Relative expression profiling of miRNAs and their putative target genes in ART biosynthesis pathway.** To analyse the possible association of miRNAs and their putative target genes in ART biosynthesis pathway, expression of six selected miRNAs as well as their target genes were analysed by RT-qPCR. The basis of selection of miRNAs was their putative targets in ART biosynthesis pathway as well as their specificity. The miRNAs having only one or two targets were selected for expression analysis.

The leaf tissues from different developmental stages were taken. The stages of growth were considered from the days after transplanting (DAT). Leaves were collected at vegetative (90 days from DAT), pre-flowering (180 days from DAT-budding stage), and flowering stages (200 days from DAT) for RNA isolation. Three biological replicates were taken from each stage. Primers for miRNAs were designed as described by Gassic et al., and the primers for the genes were designed using IDT Primer quest tool. Primer sequences used in the study are given in supplementary file S1. Total RNA was isolated using TRIzol reagent (Invitrogen, USA). 1 µg of high-quality RNA (RIN 0.8–0.95) was taken for reverse transcription reactions in 20 µl reaction mixture using RevertAid cDNA synthesis kit (Thermo Scientific, USA).

After RT reaction cDNAs were diluted 10 times. The reactions for qPCR were performed in Roche light cycler 480 (Roche) with SYBR Green PCR Master Mix (Roche) in 10 µl tubes on 96 well plates. All reactions were replicated three times. To normalize the data U6 and β-actin were taken as internal controls for miRNAs and genes respectively. To finally calculate the relative expression,  $2^{-\Delta\text{Ct}}$  equation was used, where  $\text{Ct} = (\text{Ct}_{\text{miRNA}} - \text{Ct}_{\text{U6}})$  for miRNAs and  $\text{Ct} = (\text{Ct}_{\text{gene}} - \text{Ct}_{\beta\text{-actin}})$  for genes. The specificity of amplification was checked using melt curve analysis; no primer dimers were observed.

**Statistical analysis.** The significant differences of miRNAs and their target genes with their reference genes at different developmental stages of the plant were calculated. Analysis was performed using one-way ANOVA followed by Dunnett's post hoc test. The data are presented as means ± S.E. The correlation between each miRNA and its target mRNA was investigated using the Pearson correlation test. A perfect negative correlation was defined as  $r = -1$  and a perfect positive correlation as  $r = +1$ . Correlation values between -0.3 to +0.3 were taken as no correlation between miRNA and their targets<sup>58</sup>. Results were considered statistically significant, when  $p < 0.05$ . All statistical analysis were performed using the GraphPad Prism 5.00, GraphPad Software, Inc., La Jolla, CA, USA<sup>59</sup>.

Received: 21 February 2020; Accepted: 17 June 2020

Published online: 12 August 2020

## References

- Organization, W. H. *World Malaria Report 2019*. (2019).
- Organization, W. H. *The Use of Non-Pharmaceutical Forms of Artemisia*. (World Health Organization, 2019).
- Xie, D.-Y., Ma, D.-M., Judd, R. & Jones, A. L. J. P. R. Artemisinin biosynthesis in *Artemisia annua* and metabolic engineering: Questions, challenges, and perspectives. **15**, 1093–1114 (2016).
- Suberu, J., Gromski, P. S., Nordon, A., Lapkin, A. J. J. O. P. & Analysis, B. Multivariate data analysis and metabolic profiling of artemisinin and related compounds in high yielding varieties of *Artemisia annua* field-grown in Madagascar. **117**, 522–531 (2016).
- Ding, F. *et al.* Mapping worldwide environmental suitability for *Artemisia annua* L. *Sustainability* **12**, 1309 (2020).
- Teoh, K. H., Polichuk, D. R., Reed, D. W. & Covello, P. S. Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua*. *Botany* **87**, 635–642. <https://doi.org/10.1139/B09-032> (2009).
- Tang, K., Shen, Q., Yan, T. & Fu, X. Transgenic approach to increase artemisinin content in *Artemisia annua* L. *Plant Cell Rep.* **33**, 605–615. <https://doi.org/10.1007/s00299-014-1566-y> (2014).
- Han, J. L. *et al.* Effects of overexpression of the endogenous farnesyl diphosphate synthase on the artemisinin content in *Artemisia annua* L. *J. Integr. Plant Biol.* **48**, 482–487. <https://doi.org/10.1111/j.1744-7909.2006.00208.x> (2006).
- Alam, P. & Abdin, M. Over-expression of HMG-CoA reductase and amorpha-4, 11-diene synthase genes in *Artemisia annua* L. and its influence on artemisinin content. *Plant Cell Rep.* **30**, 1919–1928. <https://doi.org/10.1007/s00299-011-1099-6> (2011).
- Yuan, Y. *et al.* Overexpression of artemisinic aldehyde  $\Delta 11$  (13) reductase gene-enhanced artemisinin and its relative metabolite biosynthesis in transgenic *Artemisia annua* L. *Biotechnol. Appl. Biochem.* **62**, 17–23 (2015).
- Liu, M. *et al.* Characterization of a trichome-specific promoter of the aldehyde dehydrogenase 1 (ALDH1) gene in *Artemisia annua*. *Plant Cell Tissue Organ Cult.* **126**, 469–480. <https://doi.org/10.1007/s11240-016-1015-4> (2016).
- Ma, D. *et al.* Overexpression of a type-I isopentenyl pyrophosphate isomerase of *Artemisia annua* in the cytosol leads to high artemisinin B production and artemisinin increase. *Plant J.* **91**, 466–479 (2017).
- 13Ma, D., Li, G., Zhu, Y. & Xie, D.-Y. J. F. I. P. S. Overexpression and suppression of *Artemisia annua* 4-hydroxy-3-methylbut-2-enyl diphosphate reductase 1 gene (AaHDR1) differentially regulate artemisinin and terpenoid biosynthesis. **8**, 77 (2017).
- Shen, Q. *et al.* The genome of *Artemisia annua* provides insight into the evolution of Asteraceae family and artemisinin biosynthesis. *Mol. Plant* **11**, 776–788. <https://doi.org/10.1016/j.molp.2018.03.015> (2018).
- Matias-Hernandez, L. *et al.* AaMYB1 and its orthologue AtMYB61 affect terpene metabolism and trichome development in *Artemisia annua* and *Arabidopsis thaliana*. *Plant J.* **90**, 520–534. <https://doi.org/10.1111/tpj.13509> (2017).
- Shen, Q., Yan, T., Fu, X. & Tang, K. Transcriptional regulation of artemisinin biosynthesis in *Artemisia annua* L. *Sci. Bull.* **61**, 18–25. <https://doi.org/10.1007/s11434-015-0983-9> (2016).
- Yu, Z.-X. *et al.* The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. *Mol. Plant* **5**, 353–365. <https://doi.org/10.1093/mp/ssr087> (2012).
- Chen, M. *et al.* GLANDULAR TRICHOME-SPECIFIC WRKY 1 promotes artemisinin biosynthesis in *Artemisia annua*. **214**, 304–316 (2017).
- Lu, X. *et al.* A a ORA, a trichome-specific AP 2/ERF transcription factor of *Artemisia annua*, is a positive regulator in the artemisinin biosynthetic pathway and in disease resistance to *Botrytis cinerea*. **198**, 1191–1202 (2013).
- Ng, D. W. *et al.* cis- and trans-Regulation of miR163 and target genes confers natural variation of secondary metabolites in two *Arabidopsis* species and their allopolyploids. *Plant Cell* **23**, 1729–1740. <https://doi.org/10.1105/tpc.111.083915> (2011).
- Yu, Z. X. *et al.* Progressive regulation of sesquiterpene biosynthesis in *Arabidopsis* and *Patchouli* (*Pogostemon cablin*) by the miR156-targeted SPL transcription factors. *Mol. Plant* **8**, 98–110. <https://doi.org/10.1016/j.molp.2014.11.002> (2015).
- Axtell, M. J. Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.* **64**, 137–159. <https://doi.org/10.1146/annurev-arplant-050312-120043> (2013).
- Axtell, M. J. & Meyers, B. C. Revisiting criteria for plant microRNA annotation in the era of big data. *Plant Cell* **30**, 272–284. <https://doi.org/10.1105/tpc.17.00851> (2018).
- Zhang, B. H., Pan, X. P., Cox, S. B., Cobb, G. P. & Anderson, T. A. Evidence that miRNAs are different from other RNAs. *Cell Mol Life Sci* **63**, 246–254. <https://doi.org/10.1007/s00018-005-5467-7> (2006).
- Shanmugavadeivel, P., Soren, K., Konda, A. K., Chaturvedi, S. & Singh, N. Identification of potential stress responsive microRNAs and their targets in *Cajanus* spp. *Agri Gene* **1**, 33–37. <https://doi.org/10.1016/j.aggene.2016.06.001> (2016).
- Boke, H. *et al.* Regulation of the alkaloid biosynthesis by mi RNA in opium poppy. *Plant Biotechnol. J.* **13**, 409–420. <https://doi.org/10.1111/pbi.12346> (2015).
- Shen, E. M. *et al.* The miRNAome of *Catharanthus roseus*: Identification, expression analysis, and potential roles of microRNAs in regulation of terpenoid indole alkaloid biosynthesis. *Sci. Rep.* **7**, 43027. <https://doi.org/10.1038/srep43027> (2017).
- Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E. F. & Hellens, R. P. Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods* **3**, 12. <https://doi.org/10.1186/1746-4811-3-12> (2007).
- Dubos, C. *et al.* MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* **15**, 573–581. <https://doi.org/10.1016/j.tplan.2010.06.005> (2010).
- Gupta, O. P., Karkute, S. G., Banerjee, S., Meena, N. L. & Dahuja, A. Contemporary understanding of miRNA-based regulation of secondary metabolites biosynthesis in plants. *Front. Plant Sci.* **8**, 374. <https://doi.org/10.3389/fpls.2017.00374> (2017).
- Barozai, M. Y. K. Identification of microRNAs and their targets in *Artemisia annua* L. *Pak. J. Bot.* **45**, 461–465 (2013).
- Pani, A., Mahapatra, R. K., Behera, N. & Naik, P. K. J. G. Proteomics. Computational identification of sweet wormwood (*Artemisia annua*) microRNA and their mRNA targets. *Genomics Proteomics Bioinform.* **9**, 200–210. [https://doi.org/10.1016/S1672-0229\(11\)60023-5](https://doi.org/10.1016/S1672-0229(11)60023-5) (2011).
- Perez-Quintero, A. L. *et al.* Mining of miRNAs and potential targets from gene oriented clusters of transcripts sequences of the anti-malarial plant, *Artemisia annua*. *Biotechnol. Lett.* **34**, 737–745. <https://doi.org/10.1007/s10529-011-0808-0> (2012).
- Vadde, B. V. L., Challa, K. R. & Nath, U. The TCP 4 transcription factor regulates trichome cell differentiation by directly activating GLABROUS INFLORESCENCE STEMS in *Arabidopsis thaliana*. *Plant J.* **93**, 259–269. <https://doi.org/10.1111/tpj.13772> (2018).
- Judd, R. *et al.* Artemisinin biosynthesis in non-glandular trichome cells of *Artemisia annua*. *Mol. Plant* **12**, 704–714. <https://doi.org/10.1016/j.molp.2019.02.011> (2019).
- Olsson, M. E. *et al.* Localization of enzymes of artemisinin biosynthesis to the apical cells of glandular secretory trichomes of *Artemisia annua* L. *Phytochemistry* **70**, 1123–1128. <https://doi.org/10.1016/j.phytochem.2009.07.009> (2009).
- 37X., G. *et al.* miR828 and miR858 regulate homoeologous MYB2 gene functions in *Arabidopsis* trichome and cotton fibre development. *Nat. Commun.* **5**, 3050, <https://doi.org/10.1038/ncomms4050> (2014).
- Li, Y. *et al.* MicroRNA858-mediated regulation of anthocyanin biosynthesis in kiwifruit (*Actinidia arguta*) based on small RNA sequencing. *PLoS ONE* **14**, e0217480. <https://doi.org/10.1371/journal.pone.0217480> (2019).
- Fahlgren, N. & Carrington, J. C. *Plant MicroRNAs 51–57* (Springer, New York, 2010).

40. Shen, Q. *et al.* Overexpression of the cytochrome P450 monooxygenase (cyp71av1) and cytochrome P450 reductase (cpr) genes increased artemisinin content in *Artemisia annua* (Asteraceae). *Genet. Mol. Res.* **11**, 3298–3309. <https://doi.org/10.4238/2012> (2012).
41. Xiang, L. *et al.* Enhancement of artemisinin biosynthesis by overexpressing dxr, cyp71av1 and cpr in the plants of *Artemisia annua* L. *Plant Omics* **5**, 503 (2012).
42. Shen, Q. *et al.* The jasmonate-responsive AaMYC2 transcription factor positively regulates artemisinin biosynthesis in *Artemisia annua*. *New Phytol.* **210**, 1269–1281. <https://doi.org/10.1111/nph.13874> (2016).
43. Lv, Z. *et al.* Overexpression of a novel NAC domain-containing transcription factor gene (AaNAC1) enhances the content of artemisinin and increases tolerance to drought and *Botrytis cinerea* in *Artemisia annua*. *Plant Cell Physiology* **57**, 1961–1971. <https://doi.org/10.1093/pcp/pcw118> (2016).
44. He, L. *et al.* Uncovering anthocyanin biosynthesis related microRNAs and their target genes by small RNA and degradome sequencing in tuberous roots of sweetpotato. *BMC Plant Biol.* **19**, 232. <https://doi.org/10.1186/s12870-019-1790-2> (2019).
45. Tirumalai, V., Swetha, C., Nair, A., Pandit, A. & Shivaprasad, P. miR828 and miR858 regulate VvMYB114 to promote anthocyanin and flavonol accumulation in grapes. *J. Exp. Bot.* <https://doi.org/10.1093/jxb/erz264> (2019).
46. Camargo-Ramirez, R., Val-Torregrosa, B., & San Segundo, B. MiR858-mediated regulation of flavonoid-specific MYB transcription factor genes controls resistance to pathogen infection in *Arabidopsis*. *Plant Cell Physiol.* **59**, 190–204. <https://doi.org/10.1093/pcp/pcx175> (2018).
47. Kawashima, C. G. *et al.* Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. *Plant J.* **57**, 313–321. <https://doi.org/10.1111/j.1365-313X.2008.03690.x> (2009).
48. Nunez, Y. O. *et al.* Positively correlated miRNA-mRNA regulatory networks in mouse frontal cortex during early stages of alcohol dependence. *BMC genomics* **14**, 725. <https://doi.org/10.1186/1471-2164-14-725> (2013).
49. Ma, D. *et al.* Isolation and characterization of AaWRKY1, an *Artemisia annua* transcription factor that regulates the amorpha-4, 11-diene synthase gene, a key gene of artemisinin biosynthesis. *Plant cell physiology* **50**, 2146–2161. <https://doi.org/10.1093/pcp/pcp149> (2009).
50. Han, J., Wang, H., Lundgren, A. & Brodelius, P. E. Effects of overexpression of AaWRKY1 on artemisinin biosynthesis in transgenic *Artemisia annua* plants. *Phytochemistry* **102**, 89–96. <https://doi.org/10.1016/j.phytochem.2014.02.011> (2014).
51. Yang, G. *et al.* MicroRNAs transcriptionally regulate promoter activity in *Arabidopsis thaliana*. *J. Integr. Plant Biol.* <https://doi.org/10.1111/jipb.12775> (2019).
52. Place, R. F., Li, L.-C., Pookot, D., Noonan, E. J. & Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci.* **105**, 1608–1613. <https://doi.org/10.1073/pnas.0707594105> (2008).
53. Ro, D.-K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nat. Genet.* **440**, 940. <https://doi.org/10.1038/nature04640> (2006).
54. Ali, A. *et al.* RNAi-mediated modulation of squalene synthase gene expression in *Artemisia annua* L. and its impact on artemisinin biosynthesis. *Rendiconti Lincei* **28**, 731–741. <https://doi.org/10.1007/s12210-017-0647-6> (2017).
55. 55Khanuja, S. P. S. *et al.* (Google Patents, 2008).
56. Dai, X., Zhuang, Z. & Zhao, P. X. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **46**, W49–W54. <https://doi.org/10.1093/nar/gky316> (2018).
57. Varkonyi-Gasic, E. *Plant Epigenetics 163–175* (Springer, New York, 2017).
58. Akoglu, H. J. T. J. O. E. M. User's guide to correlation coefficients. **18**, 91–93 (2018).
59. Motulsky, H. J. G. S. Prism 5 statistics guide, 2007. **31**, 39–42 (2007).

## Acknowledgements

The research facilities developed from UGC-SAP (DRS-I and DRSII) grant sanctioned to the Department of Biotechnology and the senior research fellowship to SK under UGC-BSR scheme by University Grants Commission, Government of India are gratefully acknowledged.

## Author contributions

S.K., A.A., and M.Z. conceived the study and designed the experiments. S.K., A.A., M.S., S.A., and P.S. conducted the experiments. S.K. wrote the manuscript. M.Z. interpreted and discussed the results and edited the manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-69707-3>.

**Correspondence** and requests for materials should be addressed to M.Z.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020