DOI: 10.1007/s10535-014-0441-6 BIOLOGIA PLANTARUM 58 (4): 627-640, 2014

Natural genetic variation in *MIR172* isolated from *Brassica* species

S.M. SHIVARAJ, P. DHAKATE, P. MAYEE, M.S. NEGI, and A. SINGH*

Department of Biotechnology, TERI University, 10 Institutional Area, Vasant Kunj, Delhi -110070, India

Abstract

The present study reports a natural variation in *microRNA172* (*MIR172*) family members isolated from six species of genus *Brassica*. The analysis of nucleotide polymorphism across 44 *Brassica MIR172* homologs revealed a higher conservation in the predicted precursors relative to flanking regions. Single nucleotide polymorphisms (SNPs) were detected in miRNA and miRNA*. The 21-nt miRNA sequence was conserved in all *MIR172* members except MIR172a. However, the miRNA* sequence was conserved only in *MIR172a* compared to *A. thaliana*. Non-canonical *Brassica* variants of precursor miR172a were detected wherein SNP at 5' terminal in mature miR172a resulted in a sequence identical to mature miR172e. SNPs and indels in precursors resulted in varied stem-loop structures of differing stabilities (ΔG) implying a differential efficiency of miRNA biogenesis. A sequence based phylogram revealed ortholog specific groupings of *MIR172* irrespective of genetic background. A Northern analysis in *Brassica juncea* displayed the cumulative expression of miR172 isoforms in all tissues representing different developmental stages with levels gradually increasing from vegetative to reproductive stages. Detection of high content of miR172 in roots indicates the possibility of additional roles of *Brassica* miR172 in root development.

Additional key words: *Arabidopsis thaliana*, evolution, flowering, orthologs, polyploidy, single nucleotide polymorphism.

Introduction

MicroRNAs (miRNAs*)* are a class of small non-coding RNAs (~21-nt) known to regulate target gene expression by directing transcript cleavage, translation repression (Rhoades *et al*. 2006, Rogers and Chen 2013), methylation (Chellappan *et al*. 2010, Wu *et al*. 2010), and histone modification (Chuang and Jones 2007, Singh and Campbell 2013). Although most miRNA loci are transcribed by RNA polymerase II (Lee *et al*. 2004, Xie *et al*. 2005), a few mammalian and viral miRNAs are transcribed by RNA polymerase III (Borchert *et al*. 2006, Diebel *et al*. 2010). In plants, the primary transcript forms a stem-loop structure which is processed in a step-wise manner by DICER-LIKE 1 (DCL1) to release a precursor of miRNA followed by a miRNA:miRNA* duplex (Reinhart *et al*. 2002, Chen 2005). The mechanistic aspects of miRNA biogenesis have been described in detail by Rogers and Chen (2013). Deep sequencing and expressed sequence tag (EST) based predictions have led to characterization of majority of plant miRNAs (Zhang *et al*. 2006, Fahlgren *et al*. 2007, Subramanian *et al*. 2008, Sunkar *et al*. 2008, Colaiacovo *et al*. 2010). Comparative genomics approaches have also been employed to isolate orthologs of miRNA families (Kusumanjali *et al*. 2011, Kumari *et al*. 2013). Characterization of miRNA gene sequences from complex plant genomes has provided insights into evolution of miRNA homologs. Whereas paralogs are considerably diverse except in the region mapping to mature miRNA (Rhoades *et al*. 2006), orthologs display lower sequence divergence even in the regions flanking mature miRNA (Zhang *et al*. 2006).

Plant miRNAs are implicated in regulation of key developmental phases (Khraiwesh *et al*. 2012, Liu *et al*. 2013). The miR172 regulates processes, such as flowering, tuberization, and nodulation (Aukerman and Sakai 2003, Martin *et al*. 2009, Yan *et al*. 2013). In

Submitted 16 October 2013, *last revision* 11 February 2014, *accepted* 25 February 2014.
Abbreviations: DAS - days after sowing; pre-miRNA - precursor microRNA; SNP - single nucleotide polymorphism.
Acknowledgements: This work was supported by grants (BT/PR10071/AGR/36/31/2007 and BT/PR628/AGR/36/674/2011) received from the Department of Biotechnology, Govt. of India. The authors thank Dr. S.B. Tripathi and Dr. A. Agnihotri for offering valuable suggestions and sharing germplasm resources. Karuna Kusumanjali and Gunjan Kumari are acknowledged for technical assistance. The contribution of Mr. Hari Ram Gupta in field-related activities is duly recognized. Financial assistance as SRF to S.M. Shivaraj and Pratiksha Mayee from the Department of Biotechnology, Govt. of India, and to Priyanka Dhakate from the Council of Scientific and Industrial Research, Govt. of India is gratefully acknowledged. Infrastructural support from TERI and the TERI University are duly acknowledged.
* Corresponding author: fax: (+91) 11 26122874, e-mail: asingh@teri.res.in

A. thaliana, miR172 determines floral organ identity and flowering time by regulating *APETALA2* (*AP2*) and *AP2*-like genes, such as *TARGET OF EAT1* (*TOE1*), *TARGET OF EAT2* (*TOE2*), and *TARGET OF EAT3* (*TOE3*), respectively (Aukerman and Sakai 2003). The miR172 mediated regulation of *AP2* occurs both by transcript cleavage and by translation repression (Chen 2004, Schwab *et al*. 2005, Zhu and Helliwell 2011). Five functionally redundant paralogs (*MIR172a*-*e*) are known to be controlled by distinct upstream regulators (Yumul *et al*. 2013) and give rise to three mature isoforms of miR172 differing at 5' or 3' terminal nucleotides (Zhu and Helliwell 2011).

MIR172 orthologs have been isolated and characterized from *A. thaliana*, rice, barley, tomato, and potato (Martin *et al*. 2009, Nair *et al*. 2010, Karlova *et al*. 2011, Lee and An 2012). In spite of variation observed in flowering time in diploid and allo-tetraploid *Brassica* species (Schranz and Osborn 2000, Lukens 2004), the sequence divergence and evolutionary fate of *MIR172* orthologs remains unknown. Comparative genomics permits identification of *Brassica* homologs based on *A. thaliana* sequence information owing to their underlying evolutionary relatedness in spite of their split from a common ancestor about 20 to 24 MYA (Lysak 2005). After divergence, whole genome duplications, natural inter-specific hybridizations, and genomic re-arrangements in *Brassica* species have resulted in expansion of gene copy number relative to *A. thaliana* (Ziolkowski *et al*. 2006, Mun *et al*. 2009).

In the present study, we employed comparative genomics to identify, isolate, and characterize sequence variation in paralogs and orthologs of *MIR172* from six *Brassica* species. We describe the extent and distribution of variations in miRNA, miRNA*, precursor, and sequences flanking the precursor of *MIR172* genes. We also report the spatio-temporal expression pattern of miR172 in *B. juncea* by quantifying the cumulative levels of miRNA isoforms.

Materials and methods

Brassica rapa (L.) cv. Ragini, *Brassica nigra* (L.) W.D.J. Koch cv. IC257, *Brassica oleracea* (L.) var. *capitata*, *Brassica carinata* A. Braun cv. Pusa swarnim, *Brassica juncea* (L.) Czern cv. Varuna, and *Brassica napus* (L.) cv. GSC5 were grown in the experimental field of *TERI-Gram*, Haryana, India (October 2010 - April 2011)**.** From 30-d-old plants, young leaves were harvested and flash frozen in liquid nitrogen. DNA was isolated using the method described by Doyle and Doyle (1990).

Precursor sequences of *A. thaliana MIR172a* (AT2G28056; 102 nt), *MIR172b* (AT5G04275; 95 nt), *MIR172c* (AT3G11435; 133 nt), *MIR172d* (AT3G55512; 124 nt), and *MIR172e* (AT5G59505; 125 nt) were employed as query to perform homology searches at *BRAD* (http://brassicadb.org). *BLASTN* was employed to screen the databases *B. rapa Chromosome v1.1* and *BAC* sequences of *B. rapa* (KBr) keeping an expected threshold at 1E-10. Genomic sequences identified as high scoring pairs (HSPs) were retrieved. For isolation of *MIR172*, forward and reverse primers were designed 500 bp upstream and 100 bp downstream, respectively, of the predicted 21-mer miRNA. *MIR172a, MIR172b, MIR172d*, and *MIR172e* were amplified using primer pairs *MIR172a* Fwd: 5'-CATCAGGTCTTCTCTG CTTAGT-3' and *MIR172a* Rev: 5'-ACCCACCTGAAG AAGATCTGG-3'; *MIR172b* Fwd: 5'-TGTAGACA CGTCAGCCCTTG-3' and *MIR172b* Rev: 5'-AAG GCCCTAATCCGTCAAAG-3'; *MIR172d* Fwd: 5'-GGC TACGCAACACATCTCTTC-3' and *MIR172d* Rev: 5'-TGTTGGTCTGGATTTCTGGA-3'; *MIR172e* Fwd: 5'-GTCCTCTATTCCTGACAGG-3' and *MIR172e* Rev: 5'-CAAGAGTAGGAAACAAGAAC-3', respectively. Oligonucleotides were synthesized by *MWG Biotech* (Munich, Germany). PCR was carried out using 50 ng of genomic DNA, 2.5 mM $MgSO_4$, 0.2 mM dNTPs, 0.5 μM each primer, 1× reaction buffer, and 1 U *Pfu* DNA polymerase (*Fermentas*, Vilnius, Lithuania). Thermo-cycling parameters for amplification of *MIR172* included initial denaturation at 95 °C for 3 min, 35 cycles at 95 °C for 60 s, at 53 °C for 60 s, and at 72 °C for 90 s, and final extension at 72 °C for 5 min. PCR products were cloned in *pGEMT Easy* (*Promega*, Madison, WI, USA), transformed into *Escherichia coli* DH5α and finally sequenced (*Macrogen*, Seoul, South Korea). At least three clones were sequenced to form a consensus sequence before submitting to GenBank for assignment of accession numbers.

B. rapa chromosomal sequences (~100 to 130 kb) or *BACs* containing predicted *MIR172* orthologs were retrieved from *BRAD*. These were used for *ab initio* gene prediction (*FGENESH*, http://linux1.softberry.com/berry.phtml). The coding sequences of predicted genes were employed as query to launch homology searches (*BLASTN*) using an expected threshold of 1E-4 in *TAIR v. 10.0* (http://www.arabidopsis.org). Coordinates of *A. thaliana* orthologs identified as high scoring pairs (HSPs) were employed for synteny analysis.

Sequences were aligned using *Clustal X v. 2.0* (Larkin *et al*. 2007). Nucleotide diversity (π) was examined through sliding window analysis (*DNASP v. 5.10*, Librado and Rozas 2009) keeping window and step size at 25 and 2 bp, respectively. Pairwise alignment scores were calculated using *BioEdit v. 7.0.5.3* (Hall 1999). Phylogenetic relationships were analysed using the Bayesian statistical framework implemented in *BEAST* (Drummond and Rambaut 2007). The alignment file in the nex format was used to generate an xml file through *BEAUti v. 1.6.2* with default parameters (the Hasegawa, Kishino, and Yano model of DNA substitution; strict clock with rate = 1.0; tree prior = coalescent tree;

constant size; Markov chain Monte Carlo chain length = 10 000 000). The xml file was used as input file in *BEAST v. 1.6.2* to permute trees which were analyzed through *Tree Annotator v. 1.6.2* (Drummond and Rambaut 2007) and visualized using *FigTree v. 1.3.1* (http://tree.bio.ed.ac.uk/software/figtree).

Precursor, miRNA and miRNA* regions in *Brassica MIR172* sequences were mapped based on *A. thaliana* sequence information. Secondary structures of predicted precursors were generated using *MFOLD v. 3.5* (Zuker 2003, http://mfold.rna.albany.edu/?q=mfold) keeping a folding temperature at 37 °C and ionic conditions at 1 M NaCl. The nomenclature suggested by Song *et al*. (2010) has been followed throughout the manuscript to specify miRNA related sequences *viz*. gene or locus, precursor, and mature miRNA sequences.

For expression analysis, root, leaf, apical meristem, closed bud, open flower, and pod samples were harvested from *B. juncea* cv. Varuna at 30, 75, and 110 d after sowing (DAS) and flash frozen in liquid nitrogen. Total RNA was extracted using a *Trizol* reagent (*Invitrogen*, Carlsbad, CA, USA) as per manufacturer's instructions. Small RNA blotting was carried out according to the protocol described by Kumari *et al*. (2013). Briefly, 50 µg of total RNA was electrophoresed on a 17 % (m/v) denaturing polyacrylamide gel at 400 V in 1× Tris-borate-EDTA (TBE) buffer. For positive control and sizing, 23 and 21 base sized oligonucleotides (5'-TAG AATCTTGATGATGCTGCAGA-3' and 5'-AGAATC TTGATGATGCTGCAG-3') were employed. The upper half of the gel was excised and stained using ethidium bromide for normalization, whereas the lower half of the gel was electro-blotted onto a *Hybond-N*$^+$ membrane (*GE Healthcare*, *Amersham Pharmacia Biotech*, Buckingham-shire, UK) using a semi-dry blotting apparatus (*Apelex*, Massy, France) at 350 mA for 2 h. The transferred RNA was UV cross-linked and the N^+ membrane was pre-hybridized in a phosphate buffer [7 % (m/v) sodium dodecyl sulphate, (SDS), 1 mM EDTA, 0.25 M NaH_2PO_4, pH 7.2)] at 37 °C for 2 h. One micromolar antisense oligonucleotide (5'-CTGCAGCATCATCAA GATTCT-3'), end labelled with P^{32}-γATP using T4 polynucleotide kinase (*Fermentas*), was used as probe. The probe was purified by gel filtration (*Sephadex G-25*) and hybridization was carried out at 37 °C overnight. The N^+ membrane was washed twice in a washing buffer (2× SSC, 0.1 % SDS) at 37 °C for 10 min. The blots were exposed to X-ray film using an intensifying screen (*Kodak X-Omatic Regular*) at -80 °C for two days and subsequently developed.

Results

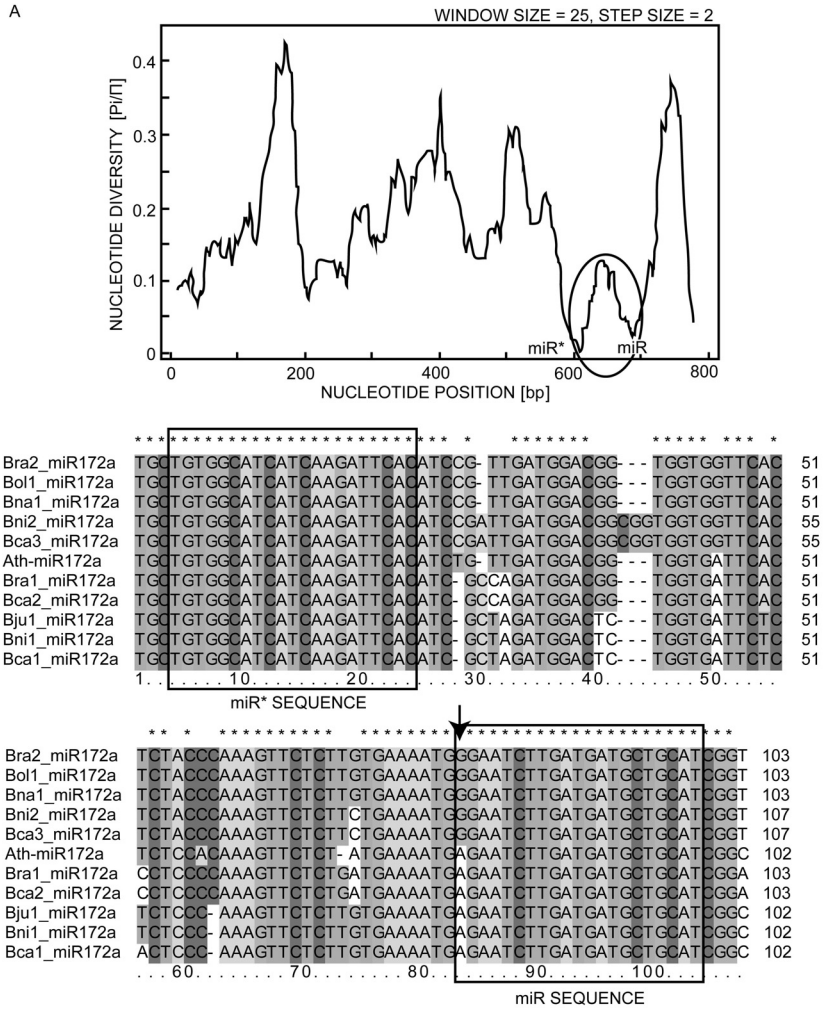
Homology search for *Brassica MIR172* led to the identification of *B. rapa* genomic regions located on chromosome A03 (coordinates: 11440436 to 11440537) and on chromosome A10 (coordinates: 11965680 to

Table 1. Size variation [bp] in *Brassica MIR172* orthologs.

Plant species	*MIR172a*	*MIR172b*	*MIR172d*	*MIR172e*
B. rapa	*Bra1_MIR172a* -715 *Bra2_MIR172a* -706	*Bra1_MIR172b* -677 *Bra2_MIR172b* -677	*Bra1_MIR172d* -575	*Bra1_MIR172e* -651 *Bra2_MIR172e* -643 *Bra3_MIR172e* -648
B. nigra	*Bni1_MIR172a* -697 *Bni2_MIR172a* -699	*Bni1_MIR172b* -653 *Bni2_MIR172b* -655 *Bni3_MIR172b* -658 *Bni4_MIR172b* -626	*Bni1_MIR172d* -587	not cloned
B. oleracea	*Bol1_MIR172a* -702	*Bol1_MIR172b* -621 *Bol2_MIR172b* -679 *Bol3_MIR172b* -642	*Bol1_MIR172d* -581 *Bol2_MIR172d* -580	*Bol1_MIR172e* -616 *Bol2_MIR172e* -596
B. juncea	*Bju1_MIR172a* -697	*Bju1_MIR172b* -658 *Bju2_MIR172b* -677	*Bju1_MIR172d* -586	*Bju1_MIR172e* -607 *Bju2_MIR172e* -607
B. napus	*Bna1_MIR172a* -704	*Bna1_MIR172b* -678	*Bna1_MIR172d* -574 *Bna2_MIR172d* -582	*Bna1_MIR172e* -596 *Bna2_MIR172e* -613
B. carinata	*Bca1_MIR172a* -697 *Bca2_MIR172a* -714 *Bca3_MIR172a* -699	*Bca1_MIR172b* -670 *Bca2_MIR172b* -664	*Bca1_MIR172d* -585 *Bca2_MIR172d* -575	*Bca1_MIR172e* -602 *Bca2_MIR172e* -604
A. thaliana	*At_MIR172a* -757	*At_MIR172b* -644	*At_MIR172d* -564	*At_MIR172e* -631

11965816) corresponding to precursors of miR172a and miR172c, respectively. Similarly, *BAC* accessions AC189388, AC189487, and AC189648, corresponding to precursors of miR172b, miR172d, and miR172e, respectively, were retrieved as *HSPs*. Orthology between *MIR172* loci of *A. thaliana* and *B. rapa* was established by analysing the conservation in the identities of flanking genes. A chromosome position (11385162 to 11511975) in A03 mapped to chromosome 2 of *A. thaliana* spanning a region between AT2G27430 to AT2G28480 that includes *MIR172a* (AT2G28056). AC189388 mapped to chromosome 5 of *A. thaliana* spanning the region between AT5G04110 to AT5G04940 that includes *MIR172b* (AT5G04275). A chromosome position (11941366 to 11991366) in A10 mapped to a region between AT3G11410 and AT3G11520 that contains *MIR172c* (AT3G11435). Similarly, AC189487 mapped to chromosome 3 of *A. thaliana,* spanning a region between AT3G55250 to AT3G55530 that contains *MIR172d* (AT3G55512). Likewise, AC189648 mapped to a region on chromosome 5 in *A. thaliana* from AT5G58820 to AT5G59540 that contains *MIR172e* (AT5G59505) lying within this stretch. Orthology was further validated using the AVID alignment tool (Bray *et al*. 2003, data not shown). Two paralogous blocks of *B. rapa* defined by chromosomal coordinates A10 (16831535 to 16831634) and A03 (693081 to 693180) were detected for *MIR172b*. Chromosomal coordinates A02 (6174070 to 6174194) and A10 (8871470 to 8871571) were detected for *MIR172e*. Similarly, additional paralogous blocks A07 (10771602 to 10771703) and A04 (2690050 to 2690173) were detected for *MIR172a* and *MIR172d,* respectively, in *B. rapa* (data not shown).

The underlying sequence similarity between genomes of *A. thaliana* and *Brassica* species permitted the isolation of homologs of *MIR172a-e* from the Indian cultivars of *Brassica* species that were assigned a nomenclature (Table 1) and submitted to GenBank (Table 1 Suppl.). A total of 14, 11, 10, and 9 variants of *MIR172b, MIR172e, MIR172a*, and *MIR172d*, respectively, were isolated (Table 1). Furthermore, multiple sequence variants were recovered from single genetic backgrounds. For instance, 2 to 3 variants of *MIRNA172a*, *MIRNA172b*, *MIRNA172d*, and *MIRNA172e* were identified from *B. carinata* cv. Pusa



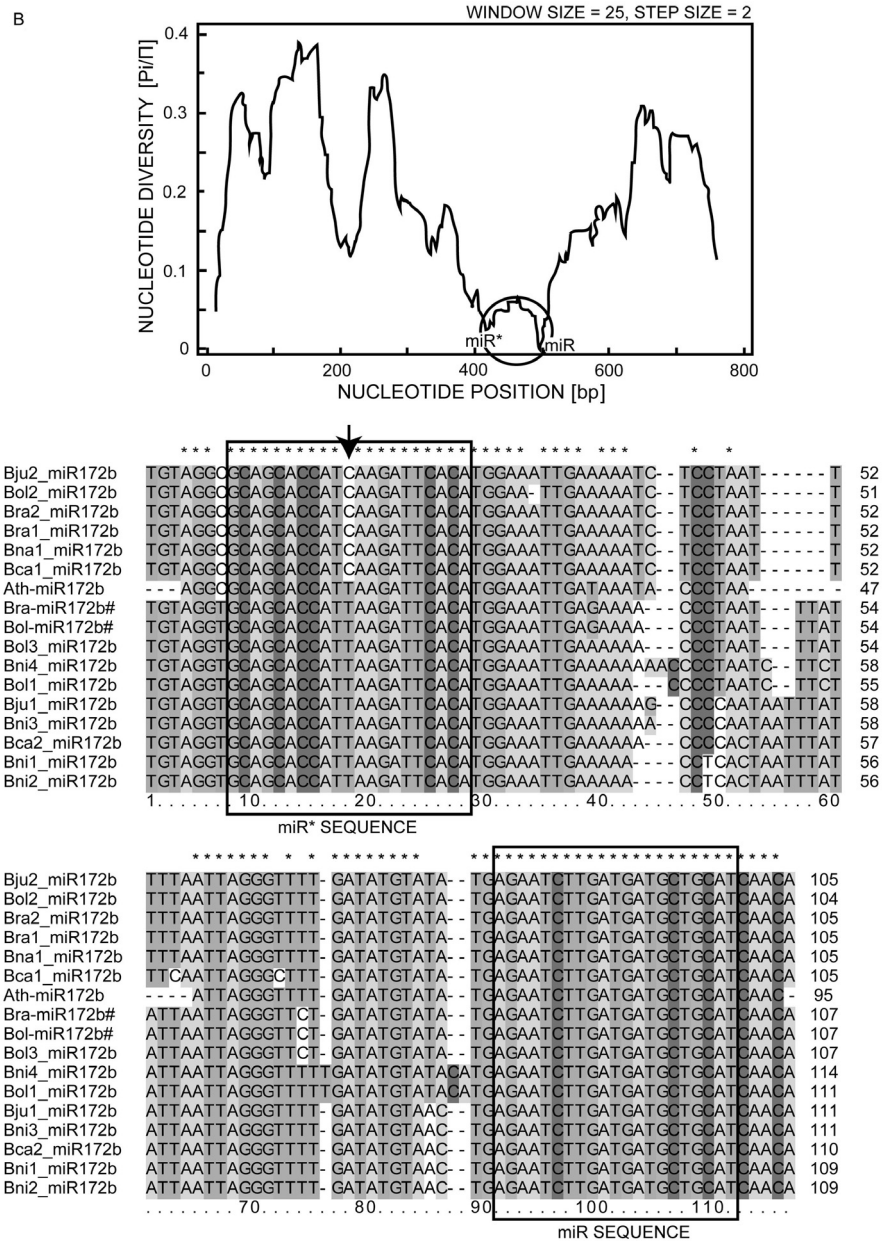


Fig. 1. Average nucleotide diversity across the complete length of aligned sequences of *MIR172a* (*A*) and *MIR172b* (*B*) orthologs of *Brassica* species and *A. thaliana*. The region mapping to the predicted precursor is encircled in the *sliding window*, whereas the multiple sequence alignment of the predicted precursor is indicated by miRNA and miRNA* (the *boxed regions*) mapping to locations of the least polymorphism. SNP within miRNA and miRNA* is marked with the *arrow*. Bra_miR172b# and Bol_miR172b# represents sequence accessions MI0010662 and MI0010666, respectively, from miRBase.

swarnim (Table 1). Interestingly, sequence variants of identical sequence lengths were also recovered from a single genetic background *Bra1_MIR172b* and *Bra2_MIR172b* (677 bp each), *Bju1_MIR172e* and *Bju2_MIR172e* (607 bp each). *MIR172b* represented the most heterogeneous class with an average pairwise similarity of 0.73 followed by *MIR172a*, *MIR172e*, and *MIR172d* with average pairwise similarity values of 0.80, 0.93, and 0.95, respectively (Table 2 Suppl.). Heterogeneity in *MIR172* sequences was also contributed by precursor length polymorphism with pre-miR172b (a size range 104 to 114 nt) and pre-miR172a (a size range 102 to 107 nt) representing seven and three distinct classes, respectively. No size polymorphism was identified in the case of pre-miR172d and pre-miR172e (Table 3 Suppl.).

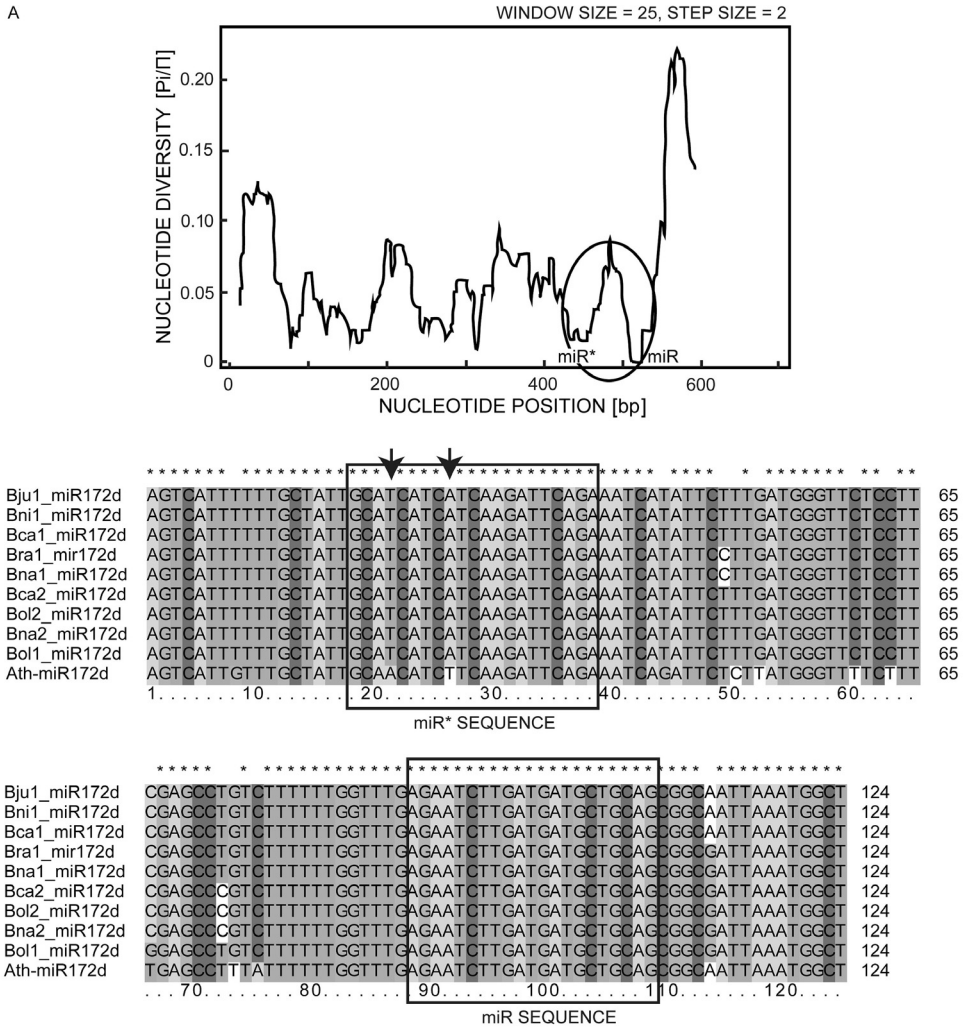
The sliding window analysis of all *MIR172* family members (Fig. 1 and 2) displayed a low polymorphism in precursors compared to flanking sequences. A negligible polymorphism was observed in miRNA and miRNA*. *MIR172a* displayed the polymorphism peaking in regions outside the predicted precursor ($\pi \approx 0.4$), whereas the miRNA and miRNA* sequences were completely

conserved ($\pi \approx 0$). In *MIR172b* orthologs, nucleotide diversity was the least ($\pi < 0.05$) in the regions defining miRNA and miRNA*. Outside the predicted precursor, however, the π value increased to 0.4. The nucleotide diversity for *MIR172d* was the highest ($\pi > 0.2$) in the region immediately downstream to the miRNA sequence, whereas the miRNA and miRNA* regions were completely conserved ($\pi \approx 0$). Similarly, *MIR172e* displayed the highest polymorphism ($\pi = 0.1$) in the regions flanking the predicted precursor, whereas the miRNA and miRNA* *per se* were invariant ($\pi \approx 0$). The analysis of multiple sequence alignment revealed several interesting polymorphisms located in the precursors and flanking region of *MIR172* family members (Figs. 1, 2, 3, and 4 Suppl.).

In *MIR172a*, a bi-allelic SNP (A/G) was detected in terminal 5' nucleotide of mature miRNA. Consequently, the precursor backbones Bra1_miR172a, Bca1_miR172a, Bca2_miR172a, Bju1_miR172a, and Bni1_miR172a bearing 5' - "A" were predicted to generate the canonical miR172a (5'-AGAAUCUUGAUGAUG CUGCAU-3'). However, the miR172a precursors of Bra2_miR172a, Bol1_miR172a, Bna1_miR172a, Bni2_miR172a, and Bca3_miR172a bearing 5' - "G" were predicted to give rise to a miR172a variant which is identical to miR172e (5'-GGAAUCUUGAUG AUGCUGCAU-3').

In contrast to *MIR172a*, the mature miRNA sequences of *MIR172b*, *MIR172d,* and *MIR172e* were conserved, though SNPs were detected in their miRNA* sequence. In *MIR172b* a SNP (C/T) was found in miRNA* within *Brassica* genus as well as between *Brassica* species and *A. thaliana*. In *MIR172d* two SNPs (T/A and A/T) in miRNA* were identified with respect to *A. thaliana*. In *MIR172e*, two prominent variations were detected within a precursor, *viz*. a lone SNP (T/C) in miRNA* and 29 bp deletion relative to *A. thaliana*.

The sequence alignment of *MIR172* family members from *A. thaliana* and *Brassica* species was carried out to analyze a genetic relationship in a phylogram (Fig. 3). *MIR172a*, *MIR172d*, and *MIR172c* were found to be genetically distinct from *MIR172b* and *MIR172e*. Further, *MIR172* members clustered in an ortholog specific manner with sub-clusters IA, IB.1, IB.2, IIA, and IIB corresponding to *MIR172a*, *MIR172c*, *MIR172d*,



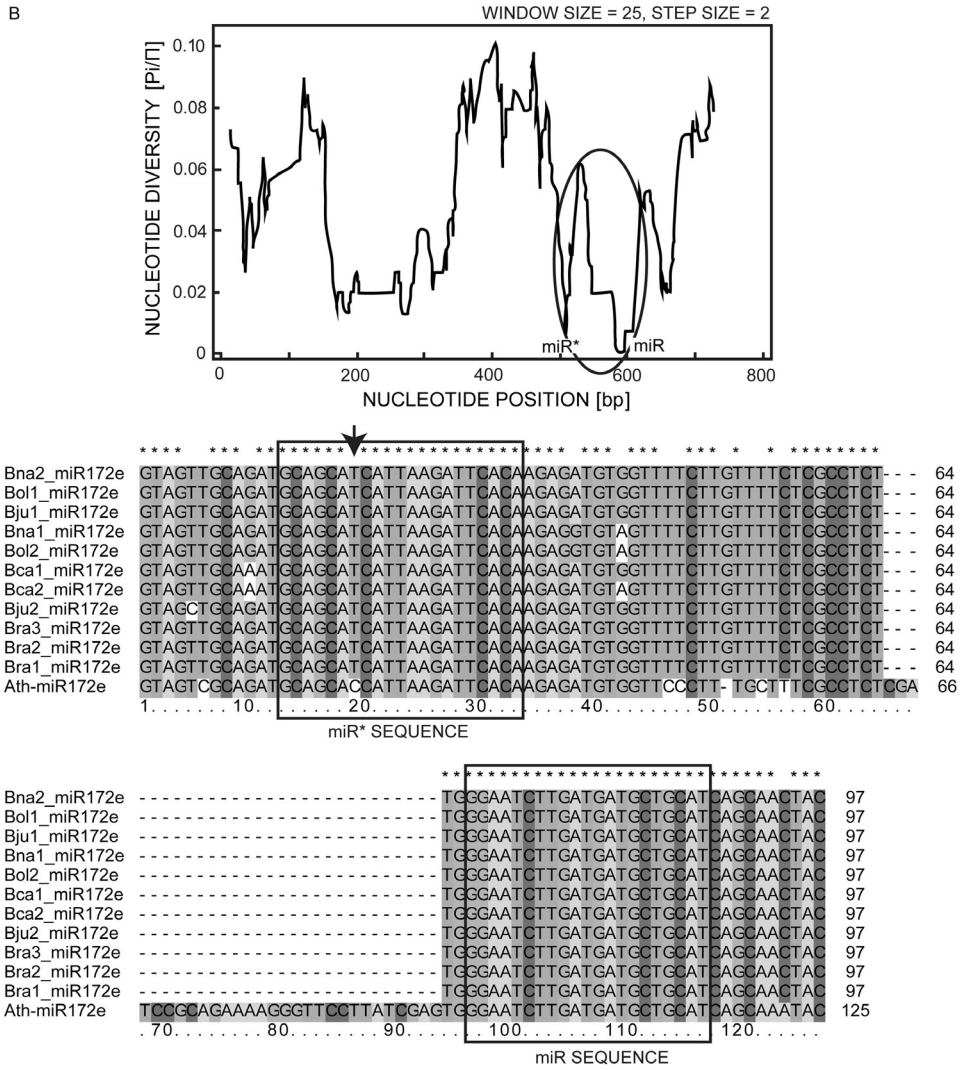


Fig. 2. Average nucleotide diversity across the complete length of aligned sequences of *MIR172d* (*A*) and *MIR172e* (*B*) orthologs of *Brassica* species and *A. thaliana*. The region mapping to the predicted precursor is encircled in the *sliding window*, whereas the multiple sequence alignment of the predicted precursor is indicated by miRNA and miRNA* (the *boxed regions*) mapping to locations of the least polymorphism. SNPs within miRNA* are marked with the *arrows*. The miR172e precursor sequence alignment revealed 29 bp deletion (5'-CGATCCGCAGAAAAGGGTTCCTTATCGAG-3') relative to *A. thaliana*.

MIR172b, and *MIR172e*, respectively. Interestingly, the members of *Brassica* subgroups IA.1 and IA.2 were distinguished by 5'-A and -G residues in the mature miRNA, respectively. Furthermore, *A. thaliana MIR172a-e* were found to branch out from *Brassica* orthologs. Within the *Brassica* orthologs, genome specific groupings were observed with sequences isolated from "A" genomes clustering separately from "B" genomes.

A total of 23 structural variants were predicted across the *MIR172* isolated from 6 different *Brassica* species. The secondary structure, ΔG, type and position of variation of pre-miR172a, pre-miR172b, pre-miR172d, and pre-miR172e are provided in Figs. 4 and 5, and 5 and 6 Suppl., respectively. Representative SNPs have been sequentially numbered in the 5' to 3' direction suffixed by a specific SNP allele; indels are marked with the *asterisks* in the fold-back structures. Consistent with the sequence variation, a wide array of structural variants were generated for both *Brassica* miR172a and miR172b precursors (Figs. 4, 5). Even though *Brassica MIR172b* orthologs represented the highest sequence diversity (Fig. 2 Suppl.), *Brassica MIR172a* represented the largest variation in topology of stem-loop structures (Fig. 4). Structural diversity was not observed in miR172d and miR172e precursors (Fig. 5 and 6 Suppl.).

Fourteen distinct polymorphic loci have been mapped on four pre-miR172a fold-back structures derived from *Brassica* variants (Fig. 4). Among these, a SNP (A/G) at position 13 mapped to the 5' terminal nucleotide of mature miRNA. ΔG of *Brassica* pre-miR172a ranged from -163.34 to -193.30 kJ mol^{-1}. Precursors of miR172b represented the highest structural diversity and generated three types of fold-back structures (Fig. 5). These differed

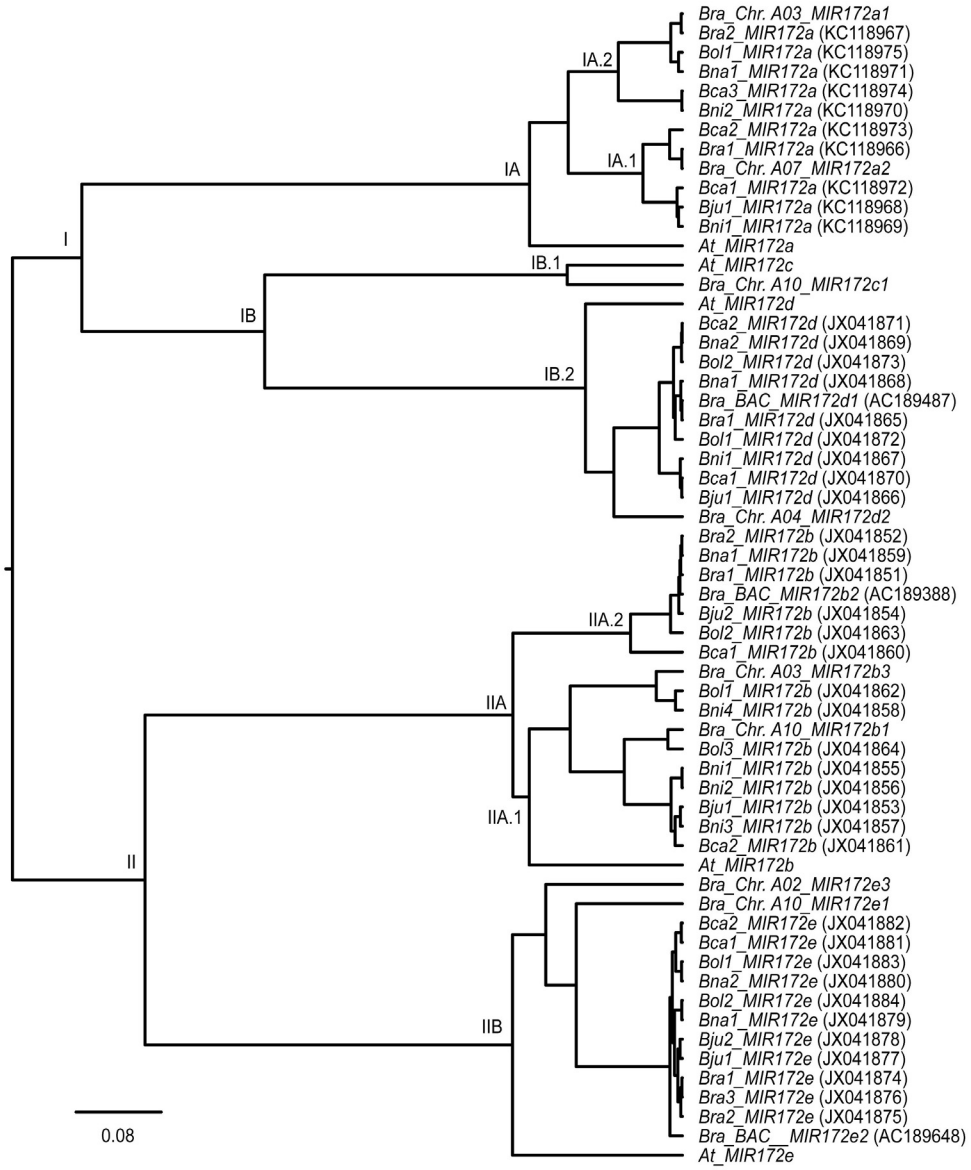


Fig. 3. A phylogram generated from the complete cloned length of *MIR172* sequences depicts genetic relationships between *MIR172* family members derived from *Brassica* species and *A. thaliana*. *MIR172* sequences formed two major clusters with *MIR172a*, *MIR172d*, and *MIR172c* sequences in one cluster and *MIR172b* and *MIR172e* sequences in another cluster. Within *Brassica* orthologs, genome specific groupings were observed with sequences from "A" and "B" genomes forming separate clusters. *Arabidopsis MIR172a-e* formed a sister clade to *Brassica* sequences.

with respect to at least 13 poly-morphic sites. Among all the polymorphic sites noted, a single SNP (6A/C) was found specific to group I (Fig 5*A*), three SNPs (1U/C, 2U/C, 5U/C) were specific to group II, and two SNPs (8U/A, 9A/C) were specific to group III (Fig. 5*B*). The groupings based on structure were found to recapitulate clustering observed in the phylogram. However, in group II, the secondary structure of Bol2_miR172b was found to be distinct from Bra1_miR172b and Bca1_miR172b (Fig. 5*B*). In spite of structural differences, Bol2_miR172b shared a high pairwise similarity of 0.99 with Bra1_miR172b and 0.96 with Bca1_miR172b, whereas Bra1_miR172b and Bca1_miR172b shared a similarity value of 0.98. Hence, the observed sequence similarities did not translate into structural similarities in this group. The distinguishing indel, 3* in Bol2_miR172b, was absent in remaining members that have an adenine residue in the corresponding position. Following an *in silico* insertion of adenine residue in Bol2_miR172b at 3*, the fold-back structure was found to conform to the topology of Bra1_miR172b (Fig. S7). Remarkably, in group III, position 10 appeared to be a hotspot of variation in Bni4_miR172b and Bol1_miR172b (Fig. 5*B*) displaying a transversion (U/C), whereas Bol3_miR172b harboured a deletion. The fold-back structures of miR172b with additional 20-nt

flanking sequences at either ends revealed well defined terminal loop with a considerable structural variation in a distal stem region owing to SNPs. On the other hand, the ~15 bp proximal stem region displayed a 3-nt unpaired bulge in a few miR172b variants (Bni3_miR172b, Bju1_miR172b, Bca2_miR172b, Bni1_miR172b, and Bol3_miR172b). Perfect linear stems devoid of any bulges were observed in other variants Bca1_miR172b, Bra1_miR172b, Bol2_miR172b, Bni4_miR172b, and Bol1_miR172b (data not shown). In general, *Brassica* miR172d precursors were found to form a structurally homogenous class with lower ΔG compared to *A. thaliana* (Fig. 5 Suppl.). The miR172e precursors from *Brassica* species displayed ΔG comparable to the *A. thaliana* counterpart in spite of a 29 bp deletion.

The spatio-temporal expression analysis revealed a gradual increase in miR172 accumulation during the transition from the vegetative to reproductive phase in all samples (Fig. 6). Tissue samples from the pre-flowering stage (30 DAS) showed the least accumulation of miR172, whereas corresponding tissue samples harvested after flowering, including roots (75 and 110 DAS), represented the maximum accumulation. Buds at 75 DAS had a markedly lower content of miRNA compared to buds at 110 DAS. Open flowers (110 DAS) and pods (110 DAS) accumulated miR172 to the same extent as bud and root samples at 110 DAS.

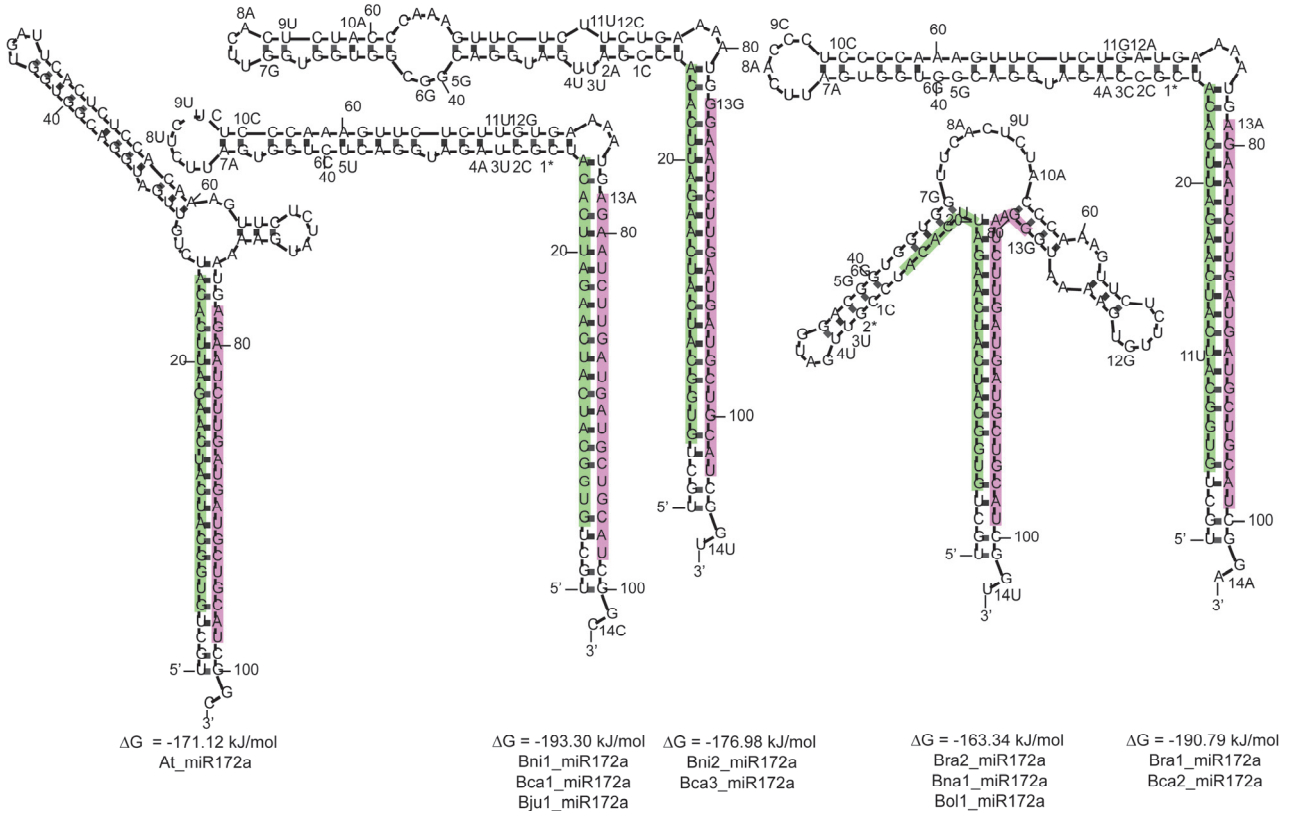


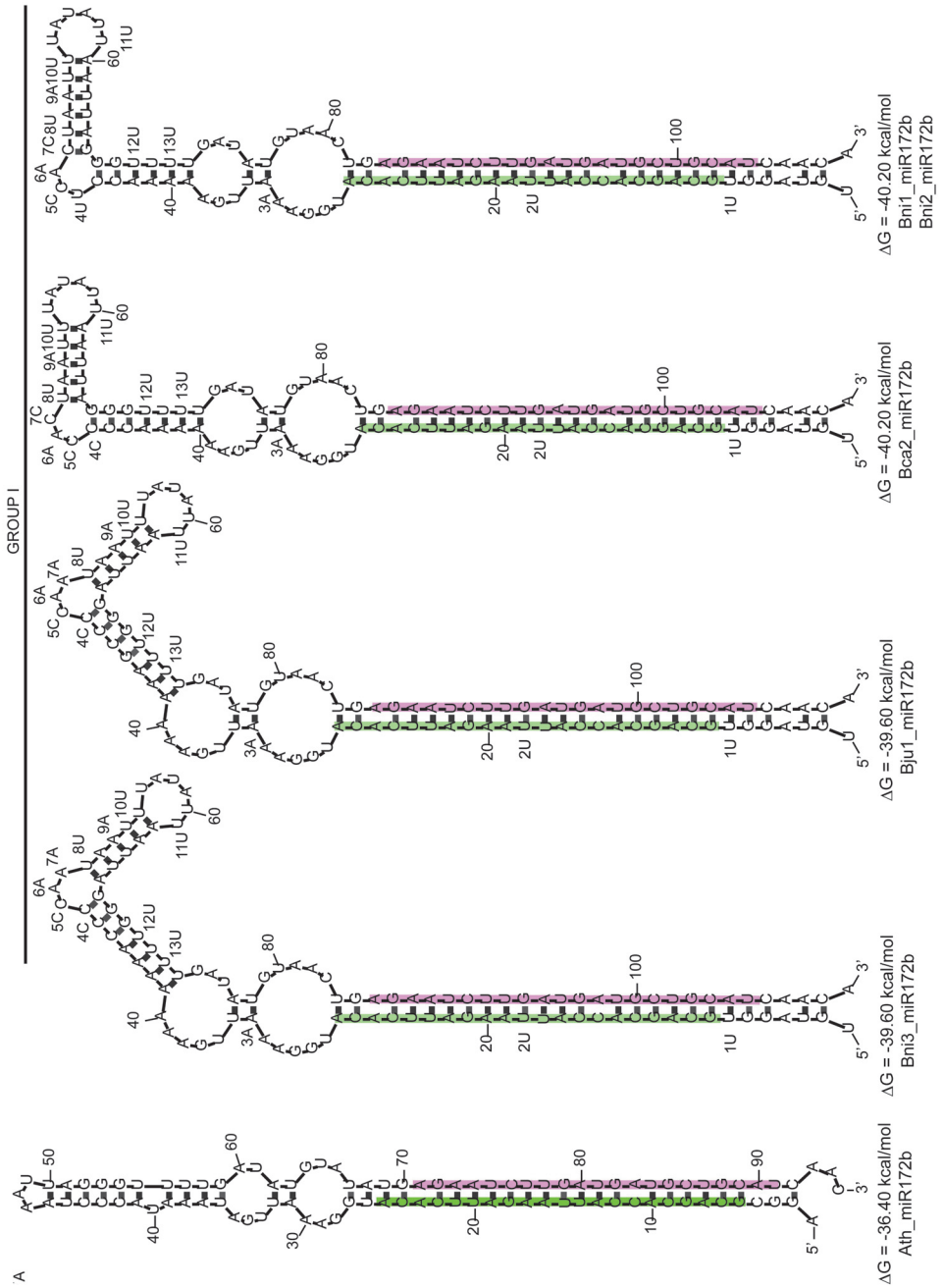
Fig. 4. The predicted fold-back structures of miR172a precursors; corresponding ΔG values are depicted. Polymorphisms within precursors of *Brassica* species are marked numerically in the order of their appearance in direction from 5' to 3', and indels are indicated by *. The miRNA sequence is highlighted in *pink*, whereas miRNA* in *green*.

Discussion

This is the first study describing the isolation and sequence characterization of *MIR172* family members from six *Brassica* species. *A. thaliana* miR172 precursor sequences were deployed to identify orthologs in the sequenced genome of *B. rapa*. The analysis of synteny between *B. rapa* and *A. thaliana* validated the conservation of gene content and order in genomic context of *MIR172*. In general, *Brassica MIR172* orthologs revealed sequence and size variations characterized as SNPs and indels. The discovery of multiple *MIR172* sequence variants from individual genetic backgrounds confirms the proposed model of *Brassica* evolution describing the diversification of paralogs after whole genome duplication events (Lysak 2005).

More specifically, our analysis revealed a higher sequence conservation in precursors relative to flanking sequences in all the *Brassica MIR172* family members. A

greater positive selection pressure was found to be imposed on sequences defining the miR172 stem-loop structure since its stability is likely to influence processing mature miRNA. Similar observations have been made in miRNA family members of *Arabidopsis* (Ehrenreich and Purugganan 2008, Warthmann *et al*. 2008) and rice (Wang *et al*. 2010). Within the predicted precursors, the sequence conservation was found to be particularly high in regions mapping to 21-nt miRNA followed by miRNA* owing to purifying selection. The base complementarity between miRNA and miRNA* is vital and must remain conserved as the base stacking of the stem enhances stability thereby facilitating efficient processing miRNA. It has already been reported that plant miRNAs do not tolerate more than four mismatches with miRNA* and a critical minimum complementarity is imperative for effective miRNA biogenesis (Meyers *et al*. 2008). While the polymorphism in the predicted mature miR172 is likely to alter ΔG of hybridization of miR172 to the target binding sites, the polymorphism in miR172* is speculated to impact biogenesis of mature miRNA. Hence, the polymorphism lying in both these sequences indirectly alters the extent of miR172 mediated down-regulation of *AP2*, *TOE1*, and *TOE2*.



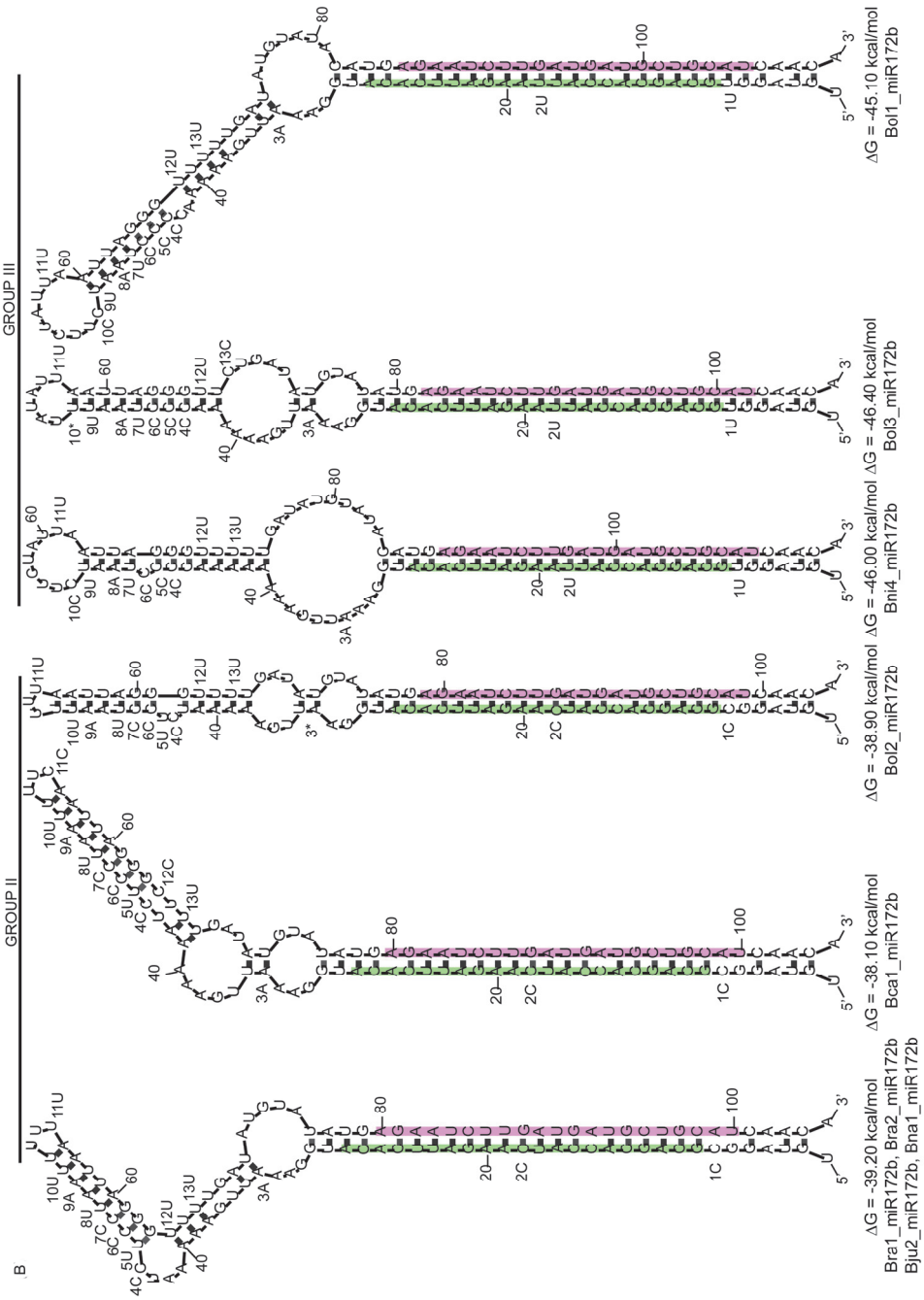


Fig. 5. The predicted fold-back structures of miR172b precursors are classified into three groups based on structural similarity; group I (*A*), group II, and group III (*B*). Polymorphism within precursors of *Brassica* species have been marked numerically in the order of appearance in direction from 5' to 3'; indels are indicated by *. The miRNA sequence is highlighted in *pink*, whereas miRNA* in *green*.

A. thaliana miR172a and miR172e have been reported to differ only with respect to the 5' terminal nucleotide (A or G) respectively, though primary and precursor sequences are distinct (miRBase v. 19; Kozomara and Griffiths-Jones 2011). Remarkably, in *Brassica* species, the 5' SNP (A to G) in miR172a resulted in a 21-nt sequence identical to miR172e. This presents a dramatic evolutionary consequence wherein miR172e is predicted to accumulate at higher amounts compared to miR172a. Liu *et al*. (2008) have discussed that miRNA isoforms differing in 5' terminal nucleotides expand the functional diversity by altering target spectra modestly *via* changed 3′ pairing. Hence, a modification of mature miR172a sequence is expected to result in down-regulation of miR172e specific targets. The possibility of miRNA variants down regulating target genes with

varying efficiency has also been illustrated (Debernardi *et al*. 2012). The miR396 variants were shown to regulate *GRF2* transcript with differential efficiency. There are no studies yet in *A*. *thaliana* that describe the specificity of interaction between paralogs of *MIR172* with diverse targets (*AP2*, *TOE1*, *TOE2*, *TOE3*, *SCHLAFMUTZE*, *SCHNARCHZAPFEN*). Nevertheless, the polyploidy induced expansion of *MIR172* and target genes in *Brassica* species are expected to present complex combinatorial interfaces.

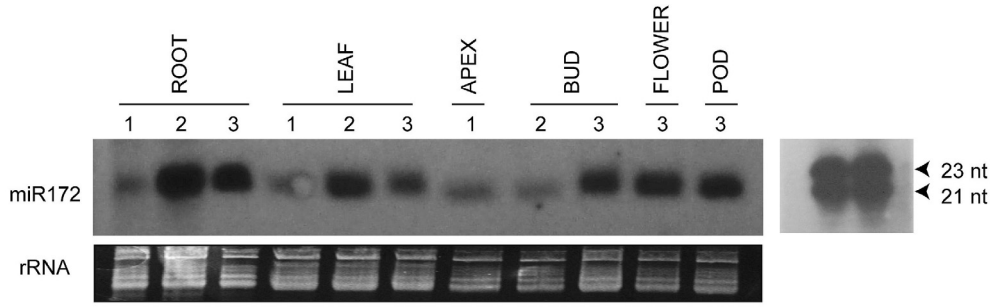


Fig. 6. Northern blot analysis indicating the miR172 expression in *Brassica juncea* cv. Varuna tissue samples harvested at 30 DAS (1), 75 DAS (2), and 110 DAS (3). Ethidium bromide stained rRNA was used as loading control (the *lower panel*). An increased accumulation of miR172 was observed in post-flowering samples (the *upper panel*). Sense oligonucleotides (21 and 23 bases) were used as markers (the *arrows*).

In the phylogram, orthologs of a specific *MIR172* family clustered together irrespective of the genetic background implying that gene duplication preceded speciation. Our findings are in accordance with Mica *et al*. (2006) wherein similar ortholog specific groupings were observed. The ancestral split between *Brassica* and *Arabidopsis* (Lysak *et al*. 2005, Ziolkowski *et al*. 2006) was evident by *A. thaliana* forming a separate clade. Within the *MIR172* family specific clades, species dependent groupings were observed. A similar pattern of clustering is reported in phylograms of *Brassica MIR164C* and *MIR165a* orthologs (Kusumanjali *et al*. 2011, Kumari *et al*. 2013).

To understand if the sequence variation impacted the stability of precursor, *in-silico* fold-back structures were generated. Surprisingly, the similarity in secondary structures of *Brassica* precursors was found to correlate well with the extent of genetic relatedness uncovered within the phylogram of *MIR172* except for pre-miR172b members of group II (Fig. 5*B*). Interestingly, the structural difference of Bol2_miR172b relative to Bra1_miR172b and Bca1_miR172b was ascribed to a point mutation at a critical position (3*, Fig. 5*B*). This unique observation implies that the topology of fold-back structure is influenced by the position of SNP.

To analyze if the observed polymorphism mapped to regions known to influence biogenesis of miRNA, fold-back structures of miR172b variants were generated after the inclusion of additional nucleotide bases. Surprisingly, in spite of the polymorphisms, fold-back structures of miR172b variants displayed perfect linear stems in the proximal region known to be crucial for biogenesis of miRNA:miRNA* duplex, thereby implying efficient processing (Werner *et al*. 2010). SNPs were observed even in a distal region in few variants resulting in minor bulges. However, such variations are unlikely to impair the biogenesis of miRNA:miRNA* duplex. Bologna *et al.* (2013) have discussed that a terminal loop in a structured proximal region is sufficient for processing. It may, therefore, be inferred that pre-miR172b structural variants are processed efficiently during biogenesis. The identification of structural variants of miR172 precursors from a single genetic background indicates a possibility of varying efficiency in biogenesis of 21-mer.

In order to understand the functional outcome of vast sequence diversity unravelled in *Brassica MIR172* loci, the possibility of expansion of *MIR172* expression domains was analyzed in amphidiploid *B. juncea*. Spatio-temporal expression studies of miR172 revealed its gradual accumulation from pre-flowering to post-flowering stages. This was true for all tissue samples tested. It may be noted that the signal intensity in the Northern blots was likely to be contributed by all miR172 isoforms since the oligonucleotide probe was perfectly complementary to miR172a and miR172b and a nearly complementary to miR172c, miR172d, and miR172e. A strong expression of miR172 was observed in root samples of *B. juncea*, whereas the least accumulations reported in *A. thaliana* (Jung *et al*. 2007) and rice (Zhu *et al*. 2009) indicates a possibility of its novel functions in *B. juncea*.

In conclusion, genome complexity in *Brassica* species has played a major role in the sequence divergence of *MIR172* homologs, but a strong evolutionary constraint has been imposed on the mature miR172 sequence so as to maintain its interaction with target genes. Expansion in the expression domain of *MIR172* into roots indicates a possibility of a diversified function acquired during the course of evolution which is corroborated by identification of novel *MIR172* variants.

References

Aukerman, M.J., Sakai, H.: Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. - Plant Cell **15**: 2730-2741, 2003.

Bologna, N.G., Schapire, A.L., Palatnik, J.F.: Processing of plant microRNA precursors. - Briefings Funct. Genomics **12**: 37-45, 2013.

Borchert, G.M., Lanier, W., Davidson, B.L.: RNA polymerase III transcribes human microRNAs. - Nat. struct. mol. Biol. **13**: 1097-1101, 2006.

Bray, N., Dubchak, I., Pachter, L.: AVID: a global alignment program. - Genome Res. **13**: 97-102, 2003.

Chellappan, P., Xia, J., Zhou, X., Gao, S., Zhang, X., Coutino, G., Vazquez, F., Zhang, W., Jin, H.: siRNAs from miRNA sites mediate DNA methylation of target genes. - Nucl. Acids Res. **38**: 6883-6894, 2010.

Chen, X.: A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. - Science **303**: 2022-2025, 2004.

Chen, X.: MicroRNA biogenesis and function in plants. - FEBS Lett. **579**: 5923-5931, 2005.

Chuang, J.C., Jones, P.A.: Epigenetics and microRNAs. - Pediatric Res. **61**: 24R-29R, 2007.

Colaiacovo, M., Subacchi, A., Bagnaresi, P., Lamontanara, A., Cattivelli, L., Faccioli, P.: A computational-based update on microRNAs and their targets in barley (*Hordeum vulgare* L.). - BMC Genomics **11**: 595, 2010.

Debernardi, J.M., Rodriguez, R.E., Mecchia, M.A., Palatnik, J.F.: Functional specialization of the plant miR396 regulatory network through distinct microRNA-target interactions. - PLoS Genet. **8**: e1002419, 2012.

Diebel, K.W., Smith, A.L., Van Dyk, L.F.: Mature and functional viral miRNAs transcribed from novel RNA polymerase III promoters. - RNA **16**: 170-185, 2010.

Doyle, J.J., Doyle, J.L.: Isolation of plant DNA from fresh tissues. - Focus **12**: 13-15, 1990.

Drummond, A.J., Rambaut, A.: BEAST: bayesian evolutionary analysis by sampling trees. - BMC Evol. Biol. **7**: 214, 2007.

Ehrenreich, I.M., Purugganan, M.D.: Sequence variation of microRNAs and their binding sites in *Arabidopsis*. - Plant Physiol. **146**: 1974-1982, 2008.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dang, J.L., Carrington, J.C.: High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of *MIRNA* genes. - PLoS ONE **2**: e219, 2007.

Hall, T.A.: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. - Nucl. Acids Symp. Ser. **41**: 95-98, 1999.

Jung, J.H., Seo, Y.H., Seo, P.J., Reyes, J.L., Yun, J., Chua, N.H., Park, C.M.: The *GIGANTEA*-regulated microRNA172 mediates photoperiodic flowering independent of *CONSTANS* in *Arabidopsis*. - Plant Cell **19**: 2736-2748, 2007.

Karlova, R., Rosin, F.M., Busscher-Lange, J., Parapunova, V., Do, P.T., Fernie, A.R., Fraser, P.D., Baxter, C., Angenent, G.C., De Maagdet, R.D.: Transcriptome and metabolite profiling show that APETELA2a is a major regulator of tomato fruit ripening. - Plant Cell **23**: 923-941, 2011.

Khraiwesh, B., Zhu, J.K., Zhu, J.H.: Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. – Biochim. Biophys. Acta **1819**: 137-148, 2012.

Kozomara, A., Griffiths-Jones, S.: miRBase: integrating microRNA annotation and deep-sequencing data. - Nucl. Acids Res. **39** (Suppl.): D152-D157, 2011.

Kumari, G., Kusumanjali, K., Srivastava, P.S., Das, S.: Isolation and expression analysis of *miR165a* and *REVOLUTA* from *Brassica* species. - Acta Physiol. Plant. **35**: 399-410, 2013.

Kusumanjali, K., Kumari, G., Srivastava, P.S., Das, S.: Sequence conservation and divergence in *miR164C1* and its target, *CUC1*, in *Brassica* species. - Plant Biotechnol. Rep. **6**: 149-163, 2011.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G.: Clustal W and Clustal X version 2.0. - Bioinformatics **23**: 2947-2948, 2007.

Lee, D.Y., An, G.: Two AP2 family genes, *SUPERNUMERARY BRACT* (*SNB*) and *OsINDETERMINATE SPIKELET1* (*OsIDS1*), synergistically control inflorescence architecture and floral meristem establishment in rice. - Plant J. **69**: 445-461, 2012.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., Kim, V.N.: MicroRNA genes are transcribed by RNA polymerase II. - EMBO J. **23**: 4051-4060, 2004.

Librado, P., Rozas, J.: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. - Bioinformatics **25**: 1451-1452, 2009.

Liu, H., Shen, D., Jia, S., Li, W., Li, J., Liu, J., Han, S., Wang, Y.: Microarray-based screening of the microRNAs associated with caryopsis development in *Oryza sativa*. - Biol. Plant. **57**: 255-261, 2013.

Liu, N., Okamura, K., Tyler, D.M., Phillips M.D., Chung, W.J., Lai, E.C.: The evolution and functional diversification of animal microRNA genes. - Cell Res. **18**: 985-996, 2008.

Lukens, L.N., Quijada, P.A., Udall, J., Pires, J.C., Schranz, M.E., Osborn, T.C.: Genome redundancy and plasticity within ancient and recent *Brassica* crop species. - Biol. J. Linn. Soc. **82**: 665-674, 2004.

Lysak, M.A., Koch, M.A., Pecinka, A., Schubert, I.: Chromosome triplication found across the tribe *Brassiceae*. - Genome Res. **15**: 516-525, 2005.

Martin, A., Adam, H., Diaz-Mendoza, M., Zurczak, M., Gonzalez-Schain, N.D., Suarez- Lopez, P.: Graft-transmissible induction of potato tuberization by the microRNA *miR172*. - Development **136**: 2873-2881, 2009.

Meyers, B.C., Axtell, M. J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J., Griffiths-Jones, S., Jacobsen, S.E., Mallory, A.C., Martienssen, R.A., Poethig, R.S., Qi, Y., Vaucheret, H., Vazquez, F., Voinnet, O., Watanabe, Y., Weigel, D. Zhu, J.K.: Criteria for annotation of plant microRNAs. - Plant Cell **20**: 3186-3190, 2008.

Mica, E., Gianfranceschi, L., Enrico Pè, M.: Characterization of five microRNA families in maize. - J. exp. Bot. **57**: 2601-2612, 2006.

Mun, J.H., Kwon, S.J., Yang, T.J., Seol, Y.J., Jin, M., Kim, J.A., Lim, M.H., Kim, J.S., Baek, S., Choi, B.S., Yu, H.J., Kim, D.S., Kim, N., Lim, K.B., Lee, S.I., Hahn, J.H., Lim, Y.P., Bancroft, I., Park, B.S.: Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. - Genome Biol. **10**: R111, 2009.

Nair, S.K., Wang, N., Turuspekov, Y., Pourkheirandish, M., Sinsuwongwat, S., Chen, G., Sameri, M., Tagiri, A., Honda, I., Watanabe, Y., Kanamori, H., Wicker, T., Stein, N.,

Nagamura, Y., Matsumoto, T., Komatsuda, T.: Cleistogamous flowering in barley arises from the suppression of microRNA-guided *HvAP2* mRNA cleavage. - Proc. nat. Acad. Sci. USA **107**: 490-495, 2010.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., Bartel, D.P.: Micro-RNAs in plants. - Genes Dev. **16**: 1616-1626, 2002.

Rhoades, M.W., Bartel, D.P., Bartel, B.: MicroRNAs and their regulatory roles in plants. - Annu. Rev. Plant Biol. **57**: 19-53, 2006.

Rogers, K., Chen, X.: Biogenesis, turnover, and mode of action of plant microRNAs. - Plant Cell **25**: 2383-2399, 2013.

Schranz, M.E., Osborn, T.C.: Novel flowering time variation in the resynthesized polyploid *Brassica napus*. - J. Heredity **91**: 242-246, 2000.

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M., Weigel, D.: Specific effects of microRNAs on the plant transcriptome. - Dev. Cell **8**: 517-527, 2005.

Singh, P.K., Campbell, M.J.: The interactions of microRNA and epigenetic modifications in prostate cancer. - Cancer **5**: 998-1019, 2013.

Song, L., Axtell, M.J., Fedoroff, N.V.: RNA secondary structural determinants of miRNA precursor processing in *Arabidopsis*. - Curr. Biol. **20**: 37-41, 2010.

Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W.B., Zhu, J.K., Yu, O.: Novel and nodulation-regulated microRNAs in soybean roots. - BMC Genomics **9**: 160, 2008.

Sunkar, R., Zhou, X., Zheng, Y., Zhang, W., Zhu, J.K.: Identification of novel and candidate miRNAs in rice by high throughput sequencing. - BMC Plant Biol. **8**: 25, 2008.

Wang, Y., Shen, D., Bo, S., Chen, H., Zheng, J., Zhu, Q. H., Cai, D., Helliwell, C., Fan, L.: Sequence variation and selection of small RNAs in domesticated rice. - BMC Evol. Biol. **10**: 119, 2010.

Warthmann, N., Das, S., Lanz, C., Weigel, D.: Comparative analysis of the *MIR319a* microRNA locus in *Arabidopsis* and related Brassicaceae. - Mol. Biol. Evol. **25**: 892-902, 2008.

Werner, S., Wollmann, H., Schneeberger, K., Weigel, D.: Structure determinants for accurate processing of miR172a in *Arabidopsis thaliana*. - Curr. Biol. **20**: 42-48, 2010.

Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C., Qi, Y.: DNA methylation mediated by a microRNA pathway. - Mol. Cell **38**: 465-475, 2010.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., Carrington, J.C.: Expression of *Arabidopsis MIRNA* Genes. - Plant Physiol. **138**: 2145-2154, 2005.

Yan, Z., Hossain, M.S., Wang, J., Valdés-López, O., Liang, Y., Libault, M., Qiu, L., Stacey, G.: miR172 Regulates Soybean Nodulation. – Mol. Plant-Microbe Interact. **26**: 1371-1377, 2013.

Yumul, R.E., Kim, Y.J., Liu, X., Wang, R., Ding, J., Xiao, L., Chen, X.: *POWERDRESS* and diversified expression of the *MIR172* gene family bolster the floral stem cell network. - PLoS Genet. **9**: e1003218, 2013.

Zhang, B., Pan, X., Cannon, C.H., Cobb, G.P., Anderson, T.A.: Conservation and divergence of plant microRNA genes. - Plant J. **46**: 243-259, 2006.

Zhu, Q.H., Upadhyaya, N.M., Gubler, F., Helliwell, C.A.: Over-expression of miR172 causes loss of spikelet determinacy and floral organ abnormalities in rice (*Oryza sativa*). - BMC Plant Biol. **9**:149, 2009.

Zhu, Q.H., Helliwell, C.: Regulation of flowering time and floral patterning by miR172. - J. exp. Bot. **62**: 487-495, 2011.

Ziolkowski, P.A., Kaczmarek, M., Babula, D., Sadowski, J.: Genome evolution in *Arabidopsis/Brassica*: conservation and divergence of ancient rearranged segments and their breakpoints. - Plant J. **47**: 63-74, 2006.

Zuker, M.: Mfold web server for nucleic acid folding and hybridization prediction. - Nucl. Acids Res. **31**: 3406-3415, 2003.