In silico identification and characterization of conserved plant microRNAs in barley

Wu, Xiaoming; Hornyik, Csaba; Bayer, Micha; Marshall, David; Waugh, Robbie; Zhang, Runxuan

Published in:
Central European Journal of Biology

DOI:
10.2478/s11535-014-0308-z

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Cent. Eur. J. Biol. • 9(9) • 2014 • 841-852
DOI: 10.2478/s11535-014-0308-z

VERSITA

Central European Journal of **Biology**

In silico identification and characterization of conserved plant microRNAs in barley

Research Article

Xiaoming Wu[1], Csaba Hornyik[2], Micha Bayer[2], David Marshall[2], Robbie Waugh[2,3], Runxuan Zhang[2]*

[1]*The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, 710049 Xi'an, China*

[2]*The James Hutton Institute, Invergowrie, Dundee, United Kingdom*

[3]*Division of Plant Sciences, College of Life Sciences, University of Dundee, Dundee, UK*

Received 11 November 2013; Accepted 20 January 2014

Abstract: Background /Methodology: Taking advantage of a newly available barley (*Hordeum vulgare*) genome sequence assembly and increasing number of miRNAs identified in other plant species, we carried out computational analyses to identify and characterize miRNAs conserved in barley. We investigate the locations of miRNAs on the barley genome assembly and provide annotation of the functions of their predicted target genes. We compare our results to previous miRNA studies and publicly available barley small RNA libraries. Results: 116 mature miRNA sequences from 60 miRNA families have been found in the barley genome assembly by our miRNA identification pipeline. Closely related cereal crops contain most of the miRNA families that we found in the barley genome assembly. Most miRNA genes were located in intergenic regions or introns. Among the 116 mature miRNAs predicted, 80 have been reported in previous barley miRNA studies. Eight mature miRNA sequences have never reported in the previous barley miRNA studies. Conclusions: This *in silico* study has provided updated information in characterizing plant miRNAs in barley. The identified miRNA and precursor sequences, their genomic locations as well as predicted target transcripts will serve as valuable resources for future studies.

Keywords: *microRNAs • Barley • In silico • Genome sequence • Small RNAs • Crop*

1. Introduction

Micro RNAs (miRNAs) are a class of non-coding small RNAs which play key roles in plant biological processes such as development, signal transduction and environmental stress response [1–3]. In plants, miRNAs originate from transcripts with strong secondary structures transcribed by RNA polymerase II from coding or intergenic regions that are referred to as primary miRNA (pri-miRNA) transcripts. These pri-miRNAs are processed by the RNase III enzyme DICER-LIKE 1 (DCL1) into shorter, stem-loop RNAs called pre-miRNAs or miRNA precursors. The pre-miRNAs are further cleaved by DCL1 generating a miRNA/miRNA* duplex with a 2-nucleotide 3' overhang. The miRNA/miRNA* duplex is exported from the nucleus into the cytoplasm where the 20-24nt miRNAs (mature miRNA) are incorporated into an RNA Induced Silencing Complex (RISC) containing one of the Argonaute proteins (Ago) and the miRNA* (star sequence) is usually degraded. The miRNAs guide the effector complexes to their target transcripts at their complementary site directing cleavage of the transcripts or inhibiting translation [4,5].

In higher eukaryotes, such as Drosophila and *Caenorhabditis elegans*, miRNA sequences account for as much as *ca*. 1% of the genome [6] and they are estimated to control the activity of 50% of all protein-coding genes in mammals [7]. miRNAs have been known to play important roles in plants, such as shoot morphogenesis, vegetative to reproductive phase

* *E-mail: Runxuan.zhang@hutton.ac.uk*

transition, floral differentiation and development, root initiation, vascular development as well as hormone signalling and homeostasis [1,2]. They also regulate gene expression in response to environmental stresses, such as pathogen attack, oxidative stress, dehydration and phosphate and sulphate limitation [3,8]. To date, 5,940 plant mature miRNAs have been found and deposited at miRBase Release 19 [9]. Many plant mature miRNAs have been found to be widely conserved in both structure and sequence across species [6,10].

Compared to experimental techniques for the detection of miRNAs, computational methods offer a fast, accurate and less expensive alternative that can more successfully identify of low copy number miRNAs [10]. The principles of these computational methods are based on: 1) the major characteristics of hairpin stem-loop secondary structures, 2) a high degree of evolutionary conservation between species, and 3) measurements of minimal folding free energy [11]. Homology based methods, in combination with secondary structure conservation, have been widely used in identification of new miRNAs in other related species [12–15].

Barley (*Hordeum vulgare*) is one of the world's most important crops and is ranked fourth in both area and tonnage produced (http://faostat.fao.org). So far there have been limited studies carried out on the barley microRNA transcriptome. Due to the lack of a barley genome sequence, all of these studies searched for miRNA precursor sequences in ESTs and other limited sequence resources [4,10,16–18]. Recently a 4.98 Gb physical map of the barley genome was published [19] along with a whole genome shotgun sequence assembly totalling 1.9 Gbp, with 410 Mbp containing the majority of barley genes anchored directly to the genetic map. The availability of this genome assembly provides a unique opportunity for a deeper and more comprehensive characterization of the barley microRNA transcriptome using computational-based methods.

Here we have carried out analyses to identify and characterize conserved miRNAs in barley. Their location on the barley genome was investigated and we have predicted their targets, providing annotation of their target genes. We compared our results to those from previous studies confirming the miRNAs identified in this work.

2. Analysis Procedures

2.1 Reference miRNAs

miRBase [9,20], the main data repository for storing microRNA information, has served for many computational miRNA discovery studies in different species [12–15,21]. A total of 5,940 mature micro RNA sequences from plants were extracted from miRBase Release 19, including 67 barley miRNAs. Duplicated sequences were removed and finally a dataset containing 3228 non-redundant mature miRNA sequences was obtained and used in the subsequent analysis.

2.2 Barley *cv*. Morex Genome

The International Barley Genome Sequencing Consortium (IBSC) has recently published a draft genome assembly of the barley cultivar *Morex* [19]. The associated physical map consists of 9,265 BAC contigs and is represented by a minimum titling path of ~ 67,000 BACs. Sequences currently anchored to the physical map include 6,300 fully sequenced BACs and ~304,000 BAC end pairs (1.136 Gbp) and WGS sequence contigs (1.9 Gb). Datasets originating from this project and referenced here were obtained from the public ftp server at the Munich Information Center for Protein Sequence, MIPS (ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data/).

2.3 *In silico* Analysis

2.3.1 Identification of miRNAs families in barley

All 5,940 plant mature miRNAs from miRBase release 19, which includes 69 barley miRNA sequences from 57 families were selected. We reduced this to a list of 3228 non-redundant miRNA sequences. These non-redundant miRNA sequences were then mapped to all possible positions with perfect matches on the *cv*. *Morex* assembly using Bowtie (version 0.12.7) [22]. 276 non-redundant miRNA sequences matched 2787 positions in the assembly. Mapped sequences were inspected visually using the Tablet assembly viewer [23]. The sequences that mapped to multiple loci in the genome assembly were filtered using 30 as a threshold. To recover miRNAs with relatively large family sizes, we filtered the miRNA list down to 255 non-redundant miRNA sequences, which mapped to 738 positions in the barley genome. 1167 flanking regions of these filtered miRNAs were then extracted using tools (excise_candidate.pl) in mirDeep-p [24] from the barley genome using a window size of 250 nt, where up to 2 potential precursor sequences were excised from one mapped location corresponding to the short sequences processed from the right or left arm of the potential precursor sequence. Then, the secondary structure and minimum folding energy of all the candidate precursor sequences were calculated using RNAfold [25]. Finally, MIRCheck [26] was used to examine the match patterns of the hairpin structure. 272 miRNAs from 267 precursors passed

MIRCheck, which corresponds to 116 mature miRNA sequences from 60 miRNA families.

2.3.2 miRNA target predictions

The small RNA target analysis tool psRNATarget (http://plantgrn.noble.org/psRNATarget/) was employed to predict the targets of the putative 116 mature miRNAs in the nucleotide sequences of all predicted transcript sequences from the barley genome project. psRNATarget predicts small RNA targets by reverse complementary matching between small RNA and target transcripts and evaluating the target site accessibility by calculating unpaired energy required to open secondary structure around the small RNA target site [27]. It also predicts the translational inhibition or cleavage degradation by presence/absence of a mismatch in the central complementary region of the small RNA sequence. The threshold for the maximum expectation value was set to 2.0. Hsp size (length of complimentary scoring) was set to 19. Target accessibility was set to 25 and the rest of the parameters were kept as defaults. 17 bp upstream and 13 bp downstream of the target site were used for target accessibility analysis. The range of the central mismatch leading to translation inhibition was between nine and 11 nt. The result is shown in Table S2. The top hits to rice annotations and HarvEST 35 (http://harvest.ucr.edu/), a barley resource with functional annotation, for the target genes are also shown in Table S2.

2.3.3 Target Annotations

Both the high-confidence (HC) and low-confidence (LC) predicted gene sequences from the IBSC barley assembly [19] were searched using BLAST in batch against peptides sequences from rice (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.1/all.dir/all.pep.gz), and barley EST sequences from the Harvest35 assembly (http://www.harvest-web.org/hweb/bin/wc.dll?hwebProcess~hmain~&versid=5). Protein databases were queried using NCBI standalone blastx version 2.25+ (ftp://ftp.ncbi.nih.gov/blast/executables/blast+/2.2.25/) using e-value cutoffs of 1 and limiting output to a single top hit only. The BLAST against the Harvest35 EST sequences was done in a similar fashion but using blastn instead.

2.3.4 miRNA verification using small RNA libraries

sRNA sequencing data from Schreiber *et al*. (2011) [4] with accession numbers SRX029116, SRX029117, SRX029118, SRX029119, SRX029120, and data from Curaba *et al*. (2012) [18] with accession numbers SRX154723, SRX154724, SRX154725, were downloaded from NCBI GEO database [28]. SRA Toolkit (version 2.1.16) [29] was used to extract the raw data. Adaptor sequences were removed and reads were trimmed based on quality values using a threshold of 20 by *cutadapt (*http://code.google.com/p/cutadapt/*)*. *Blastn (2.2.26)* [30] was used to compare miRNA mature sequences with the read sequence in the libraries. The parameters for the sequence search were set as e_value equals 0.1 and *word_size* equals 7. A match was defined as <= 2 mismatches between compared sequences. Finally, the miRNAs with less than 5 matched reads were removed from the list of verified mature miRNA sequences.

3. Results and Discussion

3.1 Conserved barley miRNAs

3.1.1 In silico identification and characterization of barley miRNAs

Aligning all known plant miRNA sequences to the barley *cv. Morex* sequence assembly and checking the secondary structures of the extracted sequences at these loci, allowed us to identify many conserved miRNAs. The analysis workflow we adopted is shown in Figure 1. 272 miRNAs from 267 precursors were predicted by our analysis pipeline, where 5 precursors host two mature miRNA sequences each. This corresponds to 116 unique mature miRNA sequences coming from 60 miRNA families. Here we define a miRNA by its mature miRNA sequence together with its precursor sequence, genome location and strand information. A unique combination of these features defined a miRNA. MicroRNA families are defined here as paralogous MIRNA loci producing identical or nearly identical mature miRNAs. The number of mismatches between mature miRNAs in the same family typically ranges from zero to two, but up to four being acceptable under certain conditions [31]. The list of predicted miRNAs, their precursors and their locations on the genome assembly are shown in Table S1.

Despite predicting 116 miRNAs, we were only able to identify 20 of the 57 barley miRNA families deposited in miRBase release 19. There could be several reasons for this poor recovery: First, the barley genome assembly is incomplete. Although a substantial proportion of the barley genome has been sequenced and assembled [19], the missing segments might have a higher representation of non-coding sequences. Second, sequence errors in the genome assembly in the miRNA locations will affect the number of miRNA retrieved. Third, we found that 45 of 57 miRNA families in miRBase 19 have been uniquely found in barley. Without the conservations, these miRNA families are

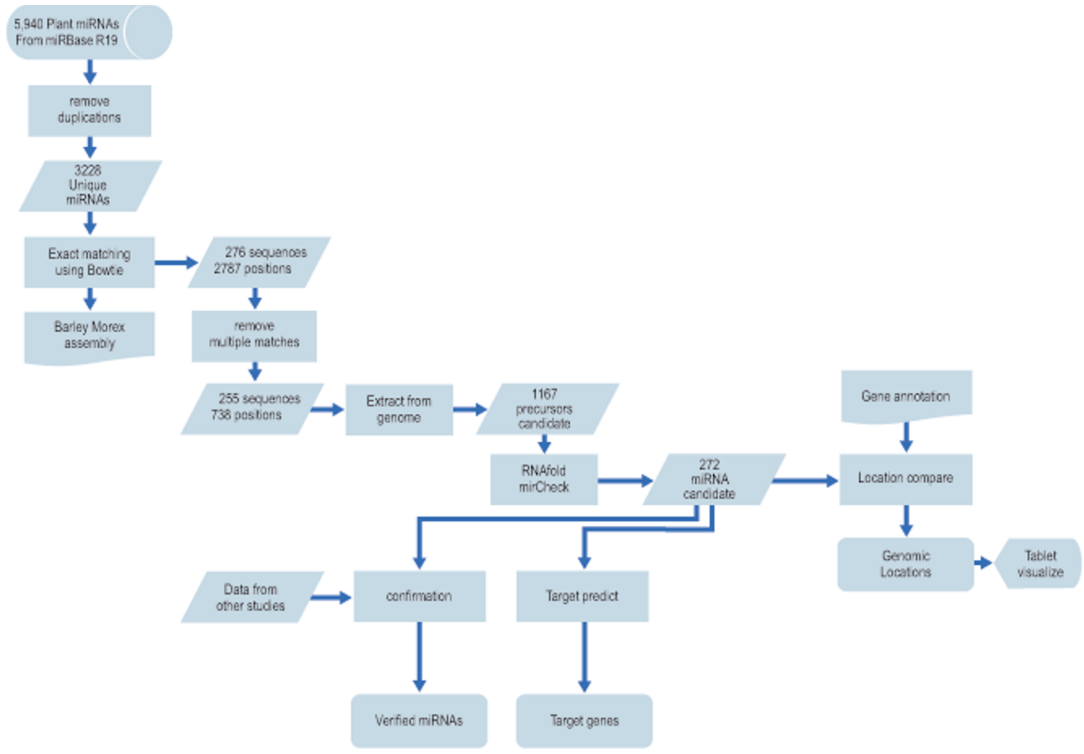

Figure 1. Flowchart for in silico identification and characterization of conserved barley miRNAs.

less confident to be true miRNAs rather than other small RNA sequences from double stranded RNAs [6,32]. Last, the miRNAs in miRBase are derived from multiple barley varieties. Different varieties may have differences in their miRNA sequences, which are not reflected in the genome assembly. In fact, we have also observed very low level of overlap among the 4 published studies on identified barley miRNAs, different experiment conditions, biological materials, analysis approaches, genomic resources, may all contribute to the discrepany observed. Therefore, more accurate and comprehensive genome resource and small RNA sequences of more tissue types should be available, computational tools needs to be assessed and/or improved and different analysis approaches needs to be taken to have a comprehensive, reproducible and consensus set of miRNAs in barley.

The length distributions of the predicted miRNAs and their precursors are shown in Figures 2a and 2b. The most abundant groups of mature miRNAs are 21nt in length, followed by those of 22nt and 20nt in length. The length of predicted barley miRNA precursors has a wide distribution of between 60 to 230nt, in agreement with previous findings in the literature [33]. We ranked the identified miRNA families by size (Figure 3). Different members of the same miRNA family come from different loci in the genome and have different pri-miRNA sequences. Similar to other studies in monocots we found that miR395 and miR169 families are amongst the largest in barley [34]. Regulation of family members may be different. Larger families may be indicative of more complex dynamics among the miRNAs, their targets and their regulators.

3.1.2 Conservation between species

The histogram in Figure 4a shows conservation of the identified miRNA families given the limited information in miRBase 19. The highest similarity to barley was found in the closely related crops rice, *Brachypodium distachyon*, maize and wheat. The related cereal species *Sorghum bicolor* ranked the 7th. The extensive conservation of miRNA families observed in cereals and closely related species suggest that miRNA regulation is probably widely conserved among these species. Of the 60 miRNA families conserved in barley, 19 are conserved and occurred in at least 2 species in miRBase (Figure 4b). These highly conserved miRNA families may reflect their fundamental role in regulating orthologous processes in different species [5,35].

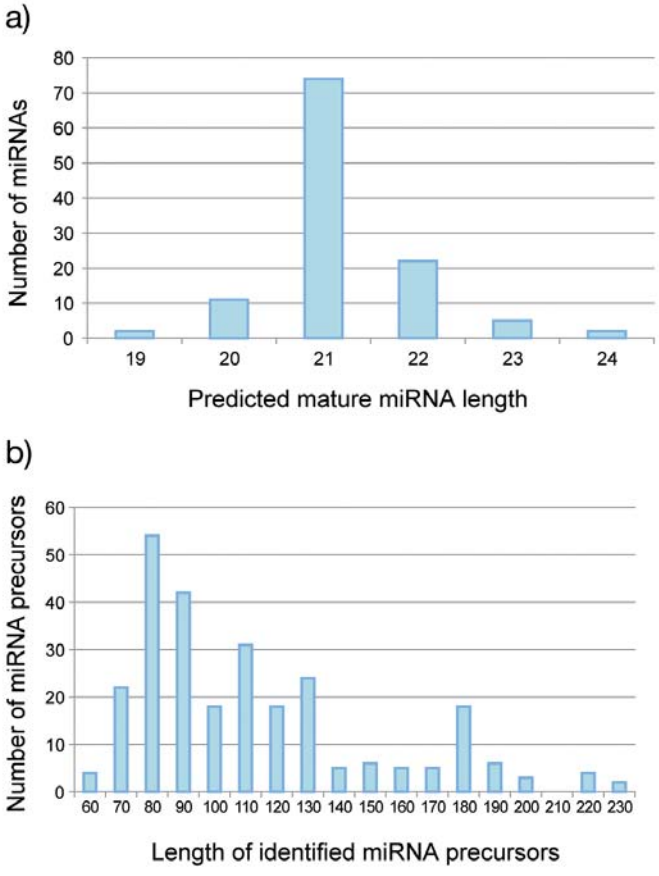

Figure 2. Length distribution of identified miRNAs and their precursors. (a) most miRNAs in barley are of 21nt in length (b) Identified barley miRNA precursors have a wide length distribution between 60 to 230nts.

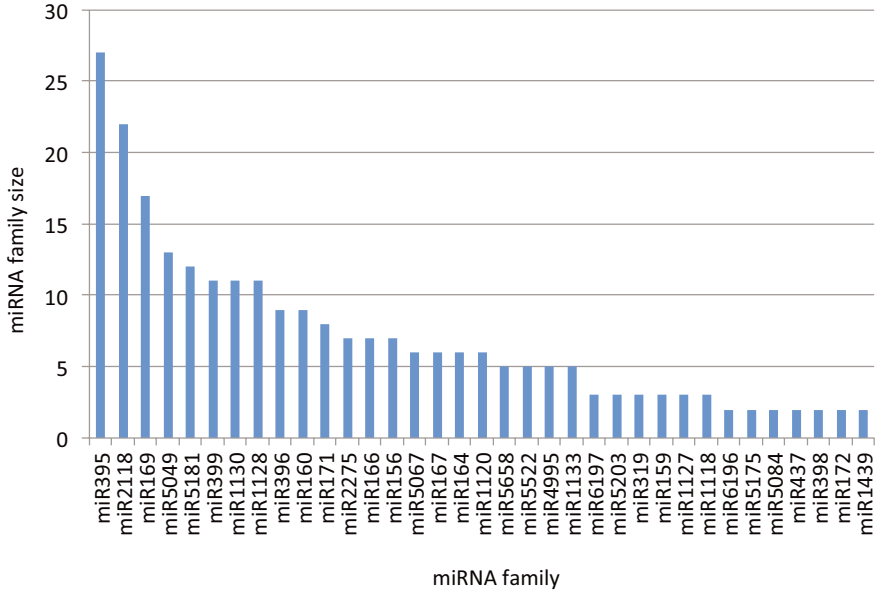

Figure 3. Size of top 24 miRNA families. Different members in a miRNA family originate from different loci from the barley genome and may have different precursor sequences.

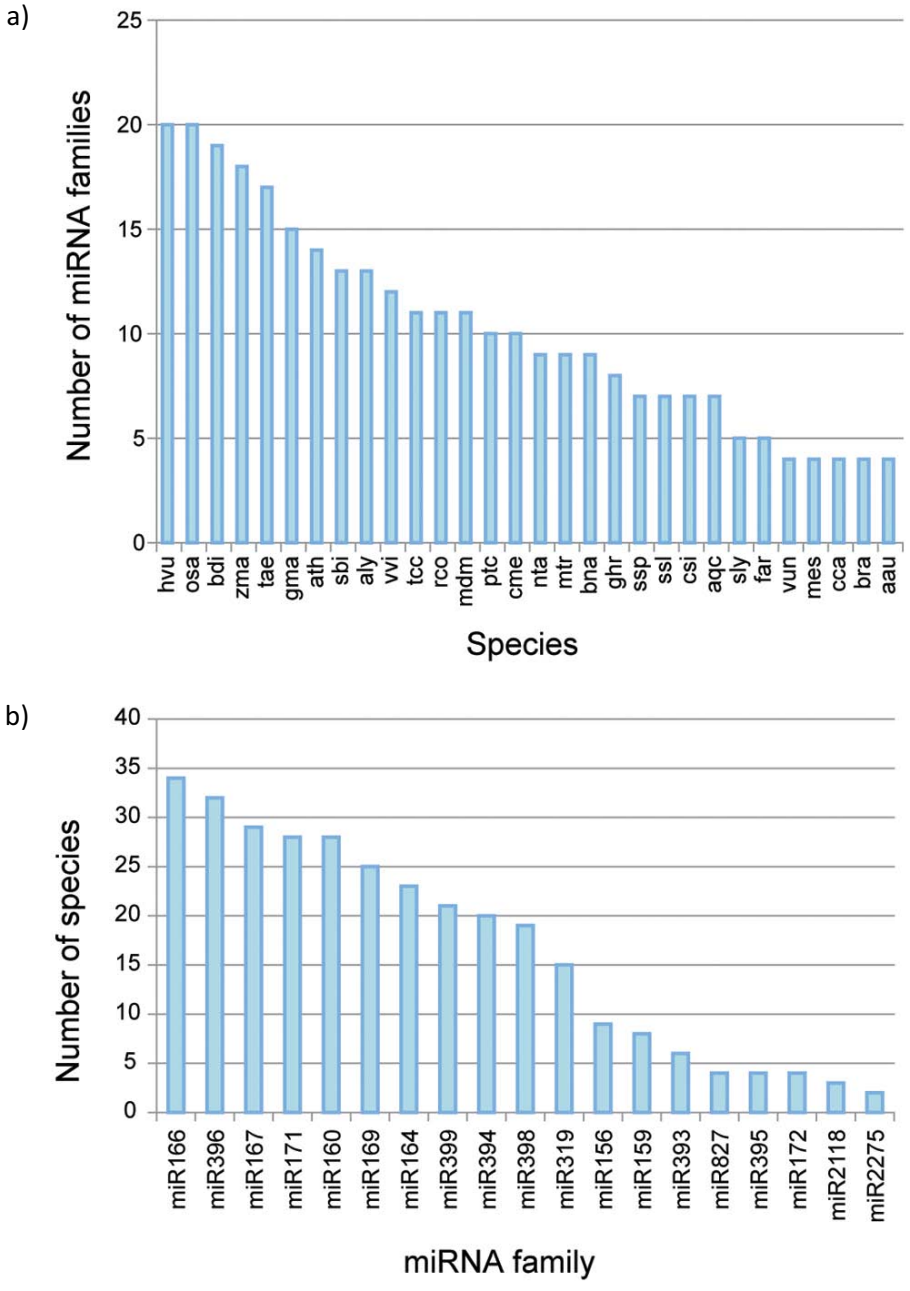

Figure 4. (a)Conservation of barley miRNA families in other plant species. Cereal and closely related crops, *Oryza sativa* (osa), *Brachypodium distachyon*(bdi), *Zea mays* (zma), *Triticum aestivum* (tae), are the top four plant species that contained conserved miRNAs. The rest of the species and corresponding abbreviations are: *Glycine max* (gma), *Arabidopsis thaliana* (ath), *Sorghum bicolor* (sbi), *Arabidopsis lyrata* (aly), *Vitis vinifera* (vvi), *Theobroma cacao* (tcc), *Ricinus communis* (rco), *Malus domestica* (mdm), *Populus trichocarpa* (ptc), *Cucumis melo* (cme), *Nicotiana tabacum* (nta), *Medicago truncatula* (mtr), *Brassica napus* (bna), *Gossypium hirsutum* (ghr), *Saccharum ssp.* (ssp), *Salvia sclarea* (ssl), *Citrus sinensis* (csi), *Aquilegia caerulea* (aqc), *Solanum lycopersicum* (sly), *Festuca arundinacea* (far), *Vigna unguiculata* (vun), *Manihot esculenta* (mes), *Cynara cardunculus* (cca), *Brassica rapa* (bra), and *Acacia auriculiformis* (aau).(b) Most conserved miRNA families. miRNA families conserved across many species may reflect their fundamental roles in regulating orthologous processes in different species.

We found that the predicted large miRNA families show a clear separation of groups in terms of conservation. The 10 largest families divide into two groups; four classified as more conserved and six as less conserved miRNA families (Table 1). The four more conserved families (miR169, miR399, miR396

miR family	number of conserved species	family size
miR395	4	27
miR2118	3	22
miR169	25	17
miR5049	1	13
miR5181	1	12
miR399	21	11
miR1130	1	11
miR1128	1	11
miR396	32	9
miR160	28	9

Table 1. Family size and conservation between species of 10 largest miRNA families

and miR160) are conserved in over 20 plant species. These represent miRNAs that play important roles in the plant and may have been constrained during evolution. The miR169 family has been found in rice to be induced by high salinity, transiently inhibiting the CCAAT-box binding transcription factor NF-YA (Nuclear Factor Y, subunit A) genes [36]. Expression of this miRNA in Arabidopsis was strongly down-regulated by nitrogen starvation [37] and drought stress [38] with its target, NF-YA, being up-regulated. The miR399 family plays key roles in inorganic phosphate (Pi) homeostasis by targeting an ubiquitin-conjugating E2 enzyme. It is up-regulated upon Pi starvation to increase the uptake of Pi in Arabidopsis. Overexpression of miR399 causes Pi accumulation in shoots that display Pi toxicity symptoms [39–41]. miR396 regulates cell proliferation in developing leaves through the repression of *GRF* activity and a decrease in the expression of cell cycle genes [42]. Overexpression of miR396 results in narrow-leaf phenotypes due to reduction in cell number [43]. It also plays a role in syncytium formation during cyst nematode infection in Arabidopsis [44]. Finally, miR160 negatively regulates three genes that encode AUXIN RESPONSE FACTORs (ARF10, -16, and -17) that are transcription factors involved in auxin signal transduction during many stages of plant growth and development. miR160 also plays important roles in seed germination and post-germination [45], and disrupting miR160 mediated regulation causes developmental defects in leaves, flowers and roots [46,47].

A further four miRNA families have been found in only one species; miR5181 in Brachypodium, miR1128 in wheat, miR5049 and miR1130 in barley. The miR2118 family is conserved in maize, rice and sorghum and the miR395 family is conserved in four species - barley, wheat, maize and Brachypodium distachyon. All six less conserved miRNA families are found only in cereal crops and closely related grass species.

3.1.3 Origins of Predicted miRNAs

The origins and genomic locations of miRNAs have significant implications for both their function and regulation. Intragenic (intronic and exonic) miRNAs can play important roles in the transcriptional regulation and RNA processing of their host genes. Evidence has shown that exonic miRNA biogenesis can incur the possibility of destabilizing the corresponding protein leading to a reduction in protein synthesis [48]. Intronic miRNAs can be subject to inefficient splicing or alternative splicing which can affect the miRNA production. Expression of intronic miR400 is down-regulated by heat treatment due to an alternative splicing event which causes the miRNA to be retained in the host gene [49]. To investigate miRNA origins, we first added introns to the barley annotation file from the International Barley Sequencing Consortium (ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data/genes/barley_HighConf_genes_MIPS_21Aug12_Transcript_and_CDS_structure.gff) using Genome Tools (version 1.4.1) and converted the .gff file into .bed format. We obtained the annotation of miRNAs by intersecting the miRNA precursor genomic locations with barley high confidence predicted gene annotations using BedTools [50]. As shown in Figure 5, we found that 52 of 272 the miRNAs that passed MIRCheck were located in coding regions and were therefore transcribed by RNA polymerase II. These could be subject to temporal or spatial regulation. 21 of these 52 miRNAs were exclusively located in introns and 15 were located in exon regions. 16 miRNAs could be found in both intron and exon regions, which indicate that some miRNAs are produced only in certain spliced isoforms of their co-existing genes.

3.2 Targets of Predicted miRNAs in barley

Figure 6a shows the relationships between predicted miRNA families and the number of target genes. It shows that the number of genes targeted by different miRNA families varies greatly, from just one to as many as 79 genes. The majority of miRNA families target less than 17 genes, however, only six miRNA families target a single gene and 90% of the miRNA families have multiple target genes. The miRNA families with the most target transcripts are shown in Figure 6b and include miR2118, miR1128, miR5658, miR5021, miR1439, miR1171, miR5169 and miR6197. As these appear to be conserved across the grasses, they likely represent evolutionary grass specific miRNAs. We found that 18

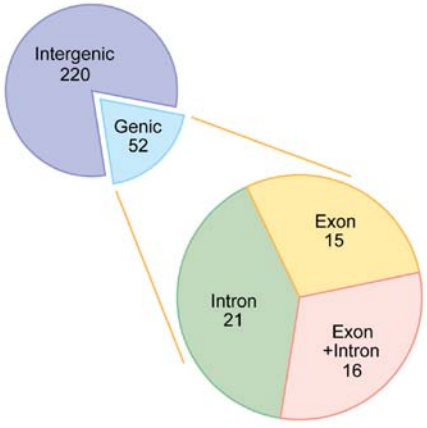

Figure 5. Origin of miRNA precursors using barley gene annotation of 26 159 high confidence genes with homology support from other plant genomes. The majority of miRNA precursors are transcribed from intergenic regions.

of the 610 target genes were targeted by two miRNA families indicating that while miRNA families target multiple genes, each gene is rarely subject to regulation by more than one miRNA family.

The predicted target transcripts of identified miRNAs are associated with a variety of functions, such as flowering, vegetative growth and responses upon environment stress. Known targets for miRNAs has been correctly predicted, such as miR172 target AP2, miR160 target CCAAT-box binding transcription factor Y, miR399 target ubiquitin-conjugating enzyme, miR396 target growth-regulating factor as well as miR160 target auxin response factors. We also identified additional target for miR160, disease resistance RPP13-like protein 1 gene, which may be indicative of miRNA functions relating disease resistance.

We also investigated the putative targets of the less conserved large families. MiR1128 putatively targets Spc97 / Spc98 family genes, which are spindle pole body (SBP) components that form a complex with gamma-tubulin. MiR1130 targets metallo-beta-lactamase family genes. MiR5049 targets oxidoreductase, short chain dehydrogenase/reductase family and UDP-glucoronosyl and UDP-glucosyl transferase domain proteins. MiR5181 targets include WD-40 repeat family genes and peroxisome assembly proteins. Mir2118 targets NBS-LRR disease resistance genes and lipid transfer proteins, which participate in cutin formation, embryogenesis, defence reactions against pathogens, symbiosis and the adaptation of plants to different environment conditions [51]. MiR395 targets a sulfate transporter, which has been reported to mediate sulphate accumulation and allocation in *Arabidopsis thaliana* [52]. Although false discoveries exist, these predictions may be helpful for choosing targets to further investigate functions of miRNA families, where their roles are generally unknown.

3.3 Identification of novel miRNAs in barley

A number of studies on miRNAs in barley have been carried out to date based on sequence homology or using high throughput sequencing of short RNA reads. A computational-based update performed by Colaiacovo *et al.* (2010) by searching for barley miRNA homologues in EST database identified 156 miRNAs belonging to 50 miRNA families [16]. Kantar *et al.* (2010) also carried out computer-based survey by looking for homologues of known miRNAs in barley EST datasets, identifying 28 miRNAs belonging to 18 miRNA families [10]. Lv *et al*. (2012) identified 126 conserved miRNAs from 58 families and 133 barley specific miRNAs from 50 families from a small RNA library of four mixed tissues and from barley EST and genome sequences screened for qualified secondary structures [17]. Schreiber *et al.* (2011) identified 100 miRNAs in barley by high-throughput sequencing of small RNAs, 44 of which were classified as barley specific. Barley BAC sequences and other related genomes have also been used to search for potential miRNA precursors [4]. Curaba *et al*. (2012) used small RNA libraries from barley seed to detect known and new miRNAs based on the presence of their precursors in a cDNA database [18]. 84 known miRNAs and 7 new miRNAs were identified in their analysis.

We compared our identified miRNAs with the results in the above studies according to sequence similarity between mature sequences. The mature sequences identified in this study were searched using BLAST [30] against reported mature miRNA sequences from all the above studies allowing a maximum mismatch of two nucleotides. A detailed comparison of the results is

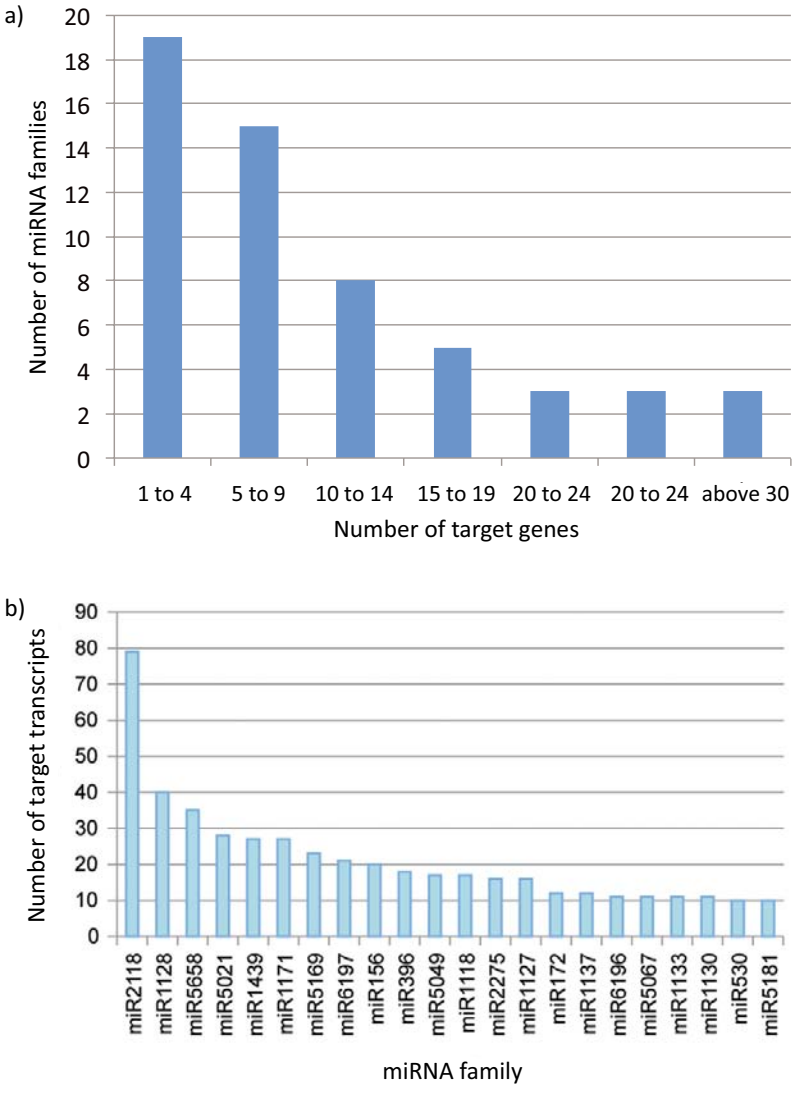

Figure 6. (a) Target gene number distributions for miRNA families. The number of transcripts targeted by one miRNA family varies greatly from one to 79. The majority of the miRNA families target 17 genes or less and only six miRNA families target one single gene. (b) Top miRNA families with most target transcripts. miRNA families with most target transcripts appear to be conserved in the grasses. It is therefore possible that they represent evolutionary grass specific miRNAs.

data source	No. of mature miRNA discovered	No. of miRNA matching our result	Total number matching all studies
Lv [17]	133 novel 126 conserved	13 49 62 (total)	80
Schreiber [4]	44	5	
Kantar [10]	28	12	
Curaba [18]	84 conserved	29	
Colaiacovo [16]	7 novel 156 conserved	0 36	

Table 2. The number of predicted miRNAs in barley compared with previously reported mature miRNA sequences

presented in Table 2. In total, we found that 80 mature miRNAs in our study were also reported in one or more of previous studies. Importantly, we also identified 36 novel miRNAs in barley using our *in silico* approach.

We then matched the mature miRNAs identified here to two sets of publicly available small RNA sequence libraries. As a result, we confirmed 61 and 43 mature miRNAs from the libraries reported by Schreiber *et al.* (2011) [4] and Curaba *et al.* (2012) [18]. Of these, 5 and 4 respectively, has not been reported in previous studies. After removing duplicates, 88 mature miRNA sequences in our study were verified by their presence in sRNA sequence libraries and/or reported by previous studies, which accounts for 75.9% of all the miRNAs identified. Among these, eight mature miRNA sequences have never been reported in any of the previous studies (Table 3) and the 28 predicted miRNA sequences without enough support evidence from previous literature or small RNA sequencing libraries were shown in Table S3.

In summary, we have provided an updated characterization of the conserved plant miRNAs in barley by taking advantage of the recently available genome sequence assembly. The identified miRNA and precursor sequences, genomic locations as well as predicted target transcripts will serve as valuable resources for further functional studies of barley miRNAs. 116 mature miRNA sequences coming from 60 miRNA families have been predicted in the barley genome assembly by our miRNA identification pipeline. Closely related cereal crops *Oryza sativa* (rice), *Brachypodium distachyon*, *Zea mays* (maize), *Triticum aestivum* (wheat) and *Sorghum bicolor* contain most of the miRNA families found in the barley, likely indicating a conservation of miRNA regulation mechanisms in cereal crops. Most of the miRNA genes were located in intergenic or intronic regions according to high confidence annotation. miRNA target prediction results showed that the number of genes targeted by different miRNA families varies greatly. While many miRNA families target multiple genes, each gene is rarely subject to regulation by more than one miRNA family. Among the identified 116 mature miRNAs, 36 have never been reported in previous barley miRNA studies. A blast search against available barley sRNA libraries showed that 88 mature miRNAs identified here have five or more reads in these libraries. We considered this observation as confirmatory evidence of their existence. Eight predicted miRNAs, confirmed by sRNA library analyses, are described for the first time in this paper.

Mir ID	Strand	Sequence
PlantMir10097_x1	+	AAGAATTTAGGAACGGAGGGA
PlantMir10346_x1	+	ACTTATTATGGACCGGAGGGA
PlantMir10500_x1	+	AGGCAGTGGCTTGGTTAAGGG
PlantMir11799_x1	-	CATGGGCAGTCTCCTTGGCTA
PlantMir12599_x59	-	TGCACGTGCCCTGCTTCTCCA
PlantMir12647_x1	+	TGGCATACAGGGAGCCAGGCA
PlantMir12709_x1	+	TGTAGATACTCTCTAAGGCTT
PlantMir12894_x37	-	CAGTTCAAGAAAGCTGTGGAA

Table 3. Eight newly discovered miRNA sequences

However, due to the limitation of the barley genome assembly, and possible mis-annotations, the miRNAs identified using this *in silico* approach only accounts for a part of the barley miRNAs deposited in miRBase. Nevertheless, the miRNAs identified in this study along with the associated analyses could still serve as a reference and be of value for further detailed investigation of miRNA families.

Acknowledgments

The authors would like to acknowledge the support given by the Scottish Government Rural and Environment Science and Analytical Services Division (RESAS) Research Programme (WP 5.2) and the Chinese Scholarship Council that funded the visit of Xiaoming Wu at the James Hutton Institute. We would also like to thank John Brown and Sarah Mckim for their helpful suggestions to improve the manuscript, Linda Milne for helpful discussions on data analysis, Iain Milne and Gordon Stephen for the ongoing development of the Tablet assembly viewer software, which was instrumental in the visual validation of our results. We would also like to thank Cavan Convery, John D Brown and Pat Carnegie for the art work.

Electronic supplementary material

The online version of this article (doi:10.2478/s11535-014-0308-z) contains supplementary material, which is available to authorized users.

References

[1] Mishra N. S., Mukherjee S. K. A Peep into the Plant miRNA World. The Open Plant Science Journal. 2007, 1, 1-9.

[2] Zhang B., Pan X., Cobb G. P., Anderson T. A. Plant microRNA: a small regulatory molecule with big impact. Dev. Biol. 2006, 289, 3-16.

[3] Schwach F., Moxon S., Moulton V., Dalmay T. Deciphering the diversity of small RNAs in plants: the long and short of it. Brief Funct Genomic Proteomic. 2009, 8, 472-481.

[4] Schreiber A., Shi B.-J., Huang C.-Y., Langridge P., Baumann U. Discovery of barley miRNAs through deep sequencing of short reads. BMC Genomics. 2011, 12, 129.

[5] Chen X. MicroRNA biogenesis and function in plants. FEBS Lett. 2005, 579, 5923-5931.

[6] Bartel D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004, 116, 281-297.

[7] Krol J., Loedige I., Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. Nat. Rev. Genet. 2010, 11, 597-610.

[8] Khraiwesh B., Zhu J. K., Zhu J. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochim. Biophys. Acta. 2012, 1819, 137-148.

[9] Griffiths-Jones S. The microRNA Registry. Nucleic Acids Res. 2004, 32, D109-111.

[10] Kantar M., Unver T., Budak H. Regulation of barley miRNAs upon dehydration stress correlated with target gene expression. Funct. Integr. Genomics. 2010, 10, 493-507.

[11] Zhang B., Pan X., Wang Q., Cobb G. P., Anderson T. A. Computational identification of microRNAs and their targets. Comput. Biol. Chem. 2006, 30, 395-407.

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

[13] Zhang W., Luo Y., Gong X., Zeng W., Li S. Computational identification of 48 potato microRNAs and their targets. Comput. Biol. Chem. 2009, 33, 84-93.

[14] Martinez G., Forment J., Llave C., Pallas V., Gomez G. High-throughput sequencing, characterization and detection of new and conserved cucumber miRNAs. PLoS One. 2011, 6, e19523.

[15] Moxon S., Jing R., Szittya G., Schwach F., Rusholme Pilcher R. L., Moulton V., et al. Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. Genome Res. 2008, 18, 1602-1609.

[16] 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. 2010, 11, 595.

[17] Lv S., Nie X., Wang L., Du X., Biradar S. S., Jia X., et al. Identification and Characterization of MicroRNAs from Barley (Hordeum vulgare L.) by High-Throughput Sequencing. Int. J. Mol. Sci. 2012, 13, 2973-2984.

[18] Curaba J., Spriggs A., Taylor J., Li Z., Helliwell C. miRNA regulation in the early development of barley seed. BMC Plant Biol. 2012, 12, 120.

[19] International Barley Genome Sequencing C., Mayer K. F., Waugh R., Brown J. W., Schulman A., Langridge P., et al. A physical, genetic and functional sequence assembly of the barley genome. Nature. 2012, 491, 711-716.

[20] Griffiths-Jones S., Grocock R. J., van Dongen S., Bateman A., Enright A. J. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006, 34, D140-144.

[21] Kim B., Yu H. J., Park S. G., Shin J. Y., Oh M., Kim N., et al. Identification and profiling of novel microRNAs in the Brassica rapa genome based on small RNA deep sequencing. BMC Plant Biol. 2012, 12, 218.

[22] Langmead B., Trapnell C., Pop M., Salzberg S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009, 10, R25.

[23] Milne I., Bayer M., Cardle L., Shaw P., Stephen G., Wright F., et al. Tablet--next generation sequence assembly visualization. Bioinformatics. 2010, 26, 401-402.

[24] Yang X., Li L. miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics. 2011, 27, 2614-2615.

[25] Hofacker I. L., Stadler P. F. Memory efficient folding algorithms for circular RNA secondary structures. Bioinformatics. 2006, 22, 1172-1176.

[26] Jones-Rhoades M. W., Bartel D. P. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol. Cell. 2004, 14, 787-799.

[27] Dai X., Zhao P. X. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 2011, 39, W155-159.

[28] Edgar R., Domrachev M., Lash A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002, 30, 207-210.

[29] Wheeler D. L., Barrett T., Benson D. A., Bryant S. H., Canese K., Chetvernin V., et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2008, 36, D13-21.
[30] Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. Basic local alignment search tool. J. Mol. Biol. 1990, 215, 403-410.
[31] Meyers B. C., Axtell M. J., Bartel B., Bartel D. P., Baulcombe D., Bowman J. L., et al. Criteria for annotation of plant MicroRNAs. Plant Cell. 2008, 20, 3186-3190.
[32] Jones-Rhoades M. W., Bartel D. P., Bartel B. MicroRNAs AND THEIR REGULATORY ROLES IN PLANTS. Annual Review of Plant Biology. 2006, 57, 19-53.
[33] Thakur V., Wanchana S., Xu M., Bruskiewich R., Quick W. P., Mosig A., et al. Characterization of statistical features for plant microRNA prediction. BMC Genomics. 2011, 12, 108.
[34] Jiao Y., Song W., Zhang M., Lai J. Identification of novel maize miRNAs by measuring the precision of precursor processing. BMC Plant Biol. 2011, 11, 141.
[35] Lelandais-Briere C., Sorin C., Declerck M., Benslimane A., Crespi M., Hartmann C. Small RNA diversity in plants and its impact in development. Curr. Genomics. 2010, 11, 14-23.
[36] Zhao B., Ge L., Liang R., Li W., Ruan K., Lin H., et al. Members of miR-169 family are induced by high salinity and transiently inhibit the NF-YA transcription factor. BMC Mol. Biol. 2009, 10, 29.
[37] Zhao M., Ding H., Zhu J. K., Zhang F., Li W. X. Involvement of miR169 in the nitrogen-starvation responses in Arabidopsis. New Phytol. 2011, 190, 906-915.
[38] Li W. X., Oono Y., Zhu J., He X. J., Wu J. M., Iida K., et al. The Arabidopsis NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. Plant Cell. 2008, 20, 2238-2251.
[39] Chiou T. J., Aung K., Lin S. I., Wu C. C., Chiang S. F., Su C. L. Regulation of phosphate homeostasis by MicroRNA in Arabidopsis. Plant Cell. 2006, 18, 412-421.
[40] Lin S. I., Chiang S. F., Lin W. Y., Chen J. W., Tseng C. Y., Wu P. C., et al. Regulatory network of microRNA399 and PHO2 by systemic signaling. Plant Physiol. 2008, 147, 732-746.
[41] Branscheid A., Sieh D., Pant B. D., May P., Devers E. A., Elkrog A., et al. Expression pattern suggests a role of MiR399 in the regulation of the cellular response to local Pi increase during arbuscular mycorrhizal symbiosis. Mol. Plant Microbe Interact. 2010, 23, 915-926.
[42] Rodriguez R. E., Mecchia M. A., Debernardi J. M., Schommer C., Weigel D., Palatnik J. F. Control of cell proliferation in Arabidopsis thaliana by microRNA miR396. Development. 2010, 137, 103-112.
[43] Liu D., Song Y., Chen Z., Yu D. Ectopic expression of miR396 suppresses GRF target gene expression and alters leaf growth in Arabidopsis. Physiol Plant. 2009, 136, 223-236.
[44] Hewezi T., Maier T. R., Nettleton D., Baum T. J. The Arabidopsis microRNA396-GRF1/GRF3 regulatory module acts as a developmental regulator in the reprogramming of root cells during cyst nematode infection. Plant Physiol. 2012, 159, 321-335.
[45] Liu P. P., Montgomery T. A., Fahlgren N., Kasschau K. D., Nonogaki H., Carrington J. C. Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. Plant J. 2007, 52, 133-146.
[46] Mallory A. C., Bartel D. P., Bartel B. MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. Plant Cell. 2005, 17, 1360-1375.
[47] Liu X., Huang J., Wang Y., Khanna K., Xie Z., Owen H. A., et al. The role of floral organs in carpels, an Arabidopsis loss-of-function mutation in MicroRNA160a, in organogenesis and the mechanism regulating its expression. Plant J. 2010, 62, 416-428.
[48] Colaiacovo M., Lamontanara A., Bernardo L., Alberici R., Crosatti C., Giusti L., et al. On the complexity of miRNA-mediated regulation in plants: novel insights into the genomic organization of plant miRNAs. Biol. Direct. 2012, 7, 15.
[49] Yan K., Liu P., Wu C. A., Yang G. D., Xu R., Guo Q. H., et al. Stress-induced alternative splicing provides a mechanism for the regulation of microRNA processing in Arabidopsis thaliana. Mol. Cell. 2012, 48, 521-531.
[50] Quinlan A. R., Hall I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010, 26, 841-842.
[51] Kader J. C. Lipid-Transfer Proteins in Plants. Annu Rev Plant Physiol Plant Mol. Biol. 1996, 47, 627-654.
[52] Liang G., Yang F., Yu D. MicroRNA395 mediates regulation of sulfate accumulation and allocation in Arabidopsis thaliana. Plant J. 2010, 62, 1046-1057.