

# PmiRKB: a plant microRNA knowledge base

Yijun Meng<sup>1,2,3</sup>, Lingfeng Gou<sup>1</sup>, Dijun Chen<sup>1</sup>, Chuanzao Mao<sup>2</sup>, Yongfeng Jin<sup>3</sup>, Ping Wu<sup>2,\*</sup> and Ming Chen<sup>1,2,\*</sup>

<sup>1</sup>Department of Bioinformatics, <sup>2</sup>State Key Laboratory of Plant Physiology and Biochemistry and <sup>3</sup>Institute of Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou 310058, P. R. China

Received June 4, 2010; Revised July 21, 2010; Accepted July 29, 2010

## ABSTRACT

MicroRNAs (miRNAs), one type of small RNAs (sRNAs) in plants, play an essential role in gene regulation. Several miRNA databases were established; however, successively generated new datasets need to be collected, organized and analyzed. To this end, we have constructed a plant miRNA knowledge base (PmiRKB) that provides four major functional modules. In the ‘SNP’ module, single nucleotide polymorphism (SNP) data of seven *Arabidopsis* (*Arabidopsis thaliana*) accessions and 21 rice (*Oryza sativa*) subspecies were collected to inspect the SNPs within pre-miRNAs (precursor microRNAs) and miRNA–target RNA duplexes. Depending on their locations, SNPs can affect the secondary structures of pre-miRNAs, or interactions between miRNAs and their targets. A second module, ‘Pri-miR’, can be used to investigate the tissue-specific, transcriptional contexts of pre- and pri-miRNAs (primary microRNAs), based on massively parallel signature sequencing data. The third module, ‘MiR–Tar’, was designed to validate thousands of miRNA–target pairs by using parallel analysis of RNA end (PARE) data. Correspondingly, the fourth module, ‘Self-reg’, also used PARE data to investigate the metabolism of miRNA precursors, including precursor processing and miRNA- or miRNA\*-mediated self-regulation effects on their host precursors. PmiRKB can be freely accessed at <http://bis.zju.edu.cn/pmirkb/>.

## INTRODUCTION

The biological significance of miRNAs was widely recognized at the beginning of this century. Since then, miRNAs have been extensively studied in both plants

and animals (1,2). The miRNAs, ~21 nt in length, can recognize their targets based on complementary recognition sites, normally resulting in transcript cleavages in plants or translational repression in animals (1,2). The functionalities of miRNA targets vary widely, although a sizable portion were transcription factors (2,3), suggesting extraordinarily complex regulatory networks that miRNAs involved in. To date, many miRNA families, either highly conserved or species-specific, have been cloned and functionally characterized (2,3). With the advent of next-generation sequencing, significant advances have been achieved in the miRNA research area, and valuable sequencing data were generated (4,5). Thus, standardized databases are required for data deposition, organization, parsing and analysis, while also allowing for user queries. Currently, there are several established miRNA databases, such as miRBase (6) and the plant microRNA database (PMRD) (7). These two databases catalog various organisms, and numerous plant species, respectively. Although extremely comprehensive, they only provide general information for miRNAs, such as sequences, secondary structures, experimental evidences, and references. Hence, by using such kind of databases, only basic queries of specific miRNAs are available. Apparently, a need exists for databases providing more specific tools and in-depth information regarding the miRNAs.

Here, we focus on the miRNAs of two model plants, the eudicot *Arabidopsis* and the monocot rice. PmiRKB (<http://bis.zju.edu.cn/pmirkb/>) contains four major functional modules which are titled: ‘SNP’, ‘Pri-miR’, ‘MiR–Tar’ and ‘Self-reg’ (Figure 1). The ‘SNP’ module contains SNPs within the pre-miRNAs among different *Arabidopsis* or rice subspecies based on currently released SNP data. To investigate the potential effects of SNPs on the secondary structure transformation of pre-miRNAs, the stem-loop structured pre-miRNAs were predicted. The ‘SNP’ module also includes SNPs present in the mature miRNAs and their target sites,

\*To whom correspondence should be addressed. Tel: +86 571 88206612; Fax: +86 571 88206612/13; Email: mchen@zju.edu.cn  
Correspondence may also be addressed to Ping Wu. Email: clspwu@zju.edu.cn

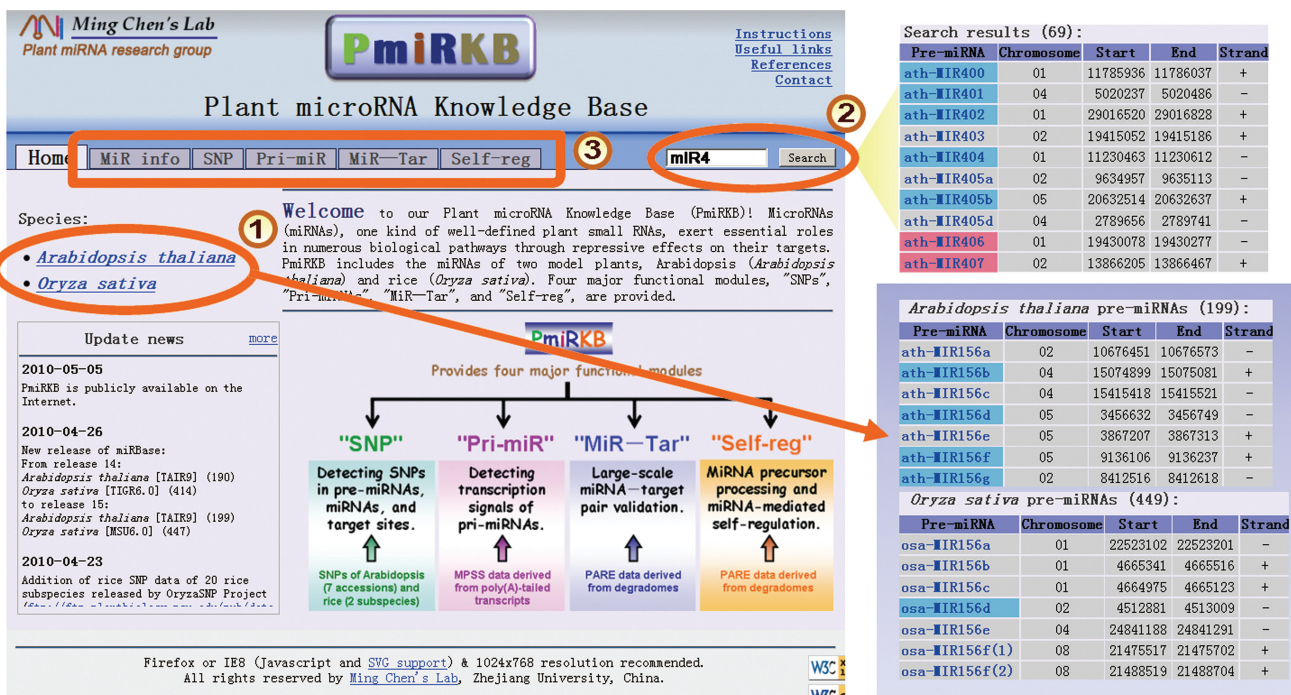
The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

which may affect miRNA–target interactions. The second module, ‘Pri-miR’, provides data regarding the transcriptional contexts of pre- and pri-miRNAs, which include tissue-specific considerations. To our knowledge, this is the first large-scale attempt to elucidate the transcriptional ranges of pri-miRNAs *in planta*. The third module, ‘MiR–Tar’, can identify cleavage signals present within predicted miRNA–target recognition sites based on PARE data derived from plant degradomes. As a result, this module can serve as the reference for *in vivo* miRNA–target pair validation. Lastly, the ‘Self-reg’ module provides data regarding miRNA precursor processing, and miRNA- or miRNA\*-mediated cleavage effects on their host precursors, based on cleavage signals detected by PARE. Together, our online service of PmiRKB implemented in PostgreSQL + Apache + hypertext preprocessor (PHP) + scalable vector graphics (SVG) provides an unprecedented resource for plant miRNA research, ensuring its value for biologists.

**DATA COLLECTION**

Sequence information for the miRNAs included in PmiRKB was retrieved from miRBase (<http://www.mirbase.org/index.shtml>; release 15) (6). Genomic information for Arabidopsis and rice were downloaded from

the Arabidopsis information resource (TAIR release 9, <ftp://ftp.arabidopsis.org/home/tair/Sequences/>) (8), and the rice genome annotation project established by the institute for genome research (currently named the J. Craig Venter Institute) (TIGR rice genome release 6.1, [ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic\\_Projects/o\\_sativa/annotation\\_dbs/](ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/)) (9), respectively. The SNPs of seven Arabidopsis accessions (Col-0, Bur-0, Tsu-1, Cvi-0, Ler-1, Bay-0, and Sha; the genome of Col-0 served as the reference) reported by Weigel’s group (10) were retrieved from TAIR ([ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR9\\_genome\\_release/TAIR9\\_gff3/Variation\\_GFF/](ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR9_genome_release/TAIR9_gff3/Variation_GFF/), release 9) (8). The SNPs between two rice subspecies, Nipponbare and 93–11 (the genome of Nipponbare served as the reference), were a gift from the Paterson group (11), while the SNPs belonging to 20 rice subspecies (Nipponbare, Tainung 67, Li-Jiang-Xin-Tuan-Hei-Gu, M 202, Azucena, Moroberekan, Cypress, Dom-Sufid, N 22, Dular, FR13 A, Aswina, Rayada, IR64-21, Shan-Huang Zhan-2, Pokkali, Swarna, Sadu-Cho, Minghui 63 and Zhenshan 97B; the genome of Nipponbare served as the reference) were obtained from the OryzaSNP Project ([ftp://ftp.plantbiology.msu.edu/pub/data/Oryza\\_SNP/](ftp://ftp.plantbiology.msu.edu/pub/data/Oryza_SNP/)) (12,13). All of the SNPs, and their genomic positions, were manually checked based on the current



**Figure 1.** Overview of the accesses to the functional modules of PmiRKB. There are three options to perform queries of the data maintained in PmiRKB. In this figure they are circled or boxed and are numbered 1 to 3. The first is an option to search *Arabidopsis thaliana* or *Oryza sativa* miRNAs. An example of each species’ output format is shown in the purple shaded box on the lower right of the figure as indicated by the arrow. Second, there is an option to perform a search of the entire knowledge base. An example of the output for the search query is shown on the upper right of the figure to the right of the yellow triangle. Third, there are five functional modules: ‘MiR info’, ‘SNP’, ‘Pri-miR’, ‘MiR–Tar’ and ‘Self-reg’, which can be clicked on to access the various datasets described in the illustrated boxes below. Note that in both the miRNA lists and the search results (shown on the right panels), the pre-miRNAs with SNPs in their mature miRNAs are shaded in red, and those with SNPs in other regions of the pre-miRNAs are shaded in blue. Additional resources are also available that include (clockwise from the top right on the PmiRKB homepage): ‘Instructions’, ‘Useful links’, ‘References’, ‘Contact’, an illustrated overview of the four major modules of PmiRKB, as well as a panel that provides ‘Update News’.

genomic information available from TAIR (release 9) (8) and TIGR (release 6.1) (9). Detected inconsistencies were removed.

A majority of the MPSS and the PARE data of Arabidopsis and rice were retrieved from the plant MPSS databases ([http://mpss.udel.edu/at/mpss\\_index.php](http://mpss.udel.edu/at/mpss_index.php) and [http://mpss.udel.edu/at\\_pare/](http://mpss.udel.edu/at_pare/) for Arabidopsis, and [http://mpss.udel.edu/rice/mpss\\_index.php](http://mpss.udel.edu/rice/mpss_index.php) and [http://mpss.udel.edu/rice\\_pare/](http://mpss.udel.edu/rice_pare/) for rice, respectively) (14–16). In addition, a portion of PARE data of Arabidopsis and short-read sequences derived from poly(A)-tailed transcripts and degradomes in rice were collected from recent reports (17–19). For users, more detailed descriptions of the features and the usages of these high-throughput sequencing (HTS) data can be obtained from the link ‘Instructions’ of PmiRKB.

All the data retrieved for this study were summarized in Table 1.

### MicroRNA TARGET PREDICTION

MiRNA targets were predicted *in silico* using the algorithm employed by miRU with default settings (20), and the messenger RNA (mRNA) sequences of Arabidopsis and rice were retrieved from TAIR (release 9) and TIGR (release 6.1), respectively. A random subset of the predicted results was manually checked with the results generated by miRU to ensure the accuracy of our prediction program written in C language.

### DATABASE CONSTRUCTION AND CONTENT

SNPs present in pre-miRNAs and mature miRNAs in Arabidopsis and rice were extracted from the whole-genome SNP datasets, and downstream targets were predicted for each miRNA using the algorithm employed by miRU (20). These results were integrated into PmiRKB, along with the SNPs present in predicted miRNA target sites and the SNPs with potential to alter the secondary structures of pre-miRNAs according to RNAfold prediction (21) (see Table 2 for summary). The resulting ‘SNP’ module can identify different secondary conformations of

a specific pre-miRNA among various Arabidopsis or rice subspecies, which result from SNP-induced thermal stability alteration (Figure 2A). Moreover, SNPs present in the miRNA–target binding regions, which have the potential to directly disturb miRNA–target interactions, were taken into account (Figure 2A). Accordingly, the output from the ‘SNP’ module can direct the investigation of phenotypic or physiological divergences among different Arabidopsis or rice subspecies, which may be attributed to SNP-involved, miRNA-mediated regulatory processes.

HTS data, largely represented by MPSS data (22), were retrieved from public resources (see details in ‘Data Collection’ section). These short reads were derived from sequencing of the poly(A)-tailed transcripts in various tissues (14). Although many poly(A)-tailed transcripts are represented by mRNAs, a majority of pri-miRNAs transcribed by RNA polymerase II also possess poly(A) tails (2,23). Thus, these HTS datasets can be used to detect the transcriptional signals of pri-miRNAs in a tissue-specific manner. For this purpose, all the HTS signatures were mapped to the corresponding genomes, and only the signatures with unique genomic loci were reserved for further investigation. The short reads located within ~5 kb on either side of a pre-miRNA were used to reveal the transcriptional context of the corresponding pri-miRNA. Moreover, expression levels of the short reads associated with each library prepared from distinct tissues were normalized in RPM (reads per million) to allow cross-library comparisons. Thus, the transcriptional signals of certain pri-miRNA can be queried in a tissue-specific manner by using the module ‘Pri-miR’ (Figure 2D–F).

In plants, most miRNAs cleave their downstream targets depending on the highly complementary recognition sites (2). The resulting 3′ cleavage remnants are usually utilized for miRNA–target pair validation and slicing site mapping by using modified 5′ rapid amplification of cDNA ends (RACE) (3). However, this classical method would be time-consuming, tedious, and costly once applied to large-scale validation of miRNA–target pairs. To overcome these limitations, the modified 5′ RACE has been combined with the newly available HTS

**Table 1.** Data sources used for the construction of PmiRKB

| Data type            | Species     | Data source   | Reference |
|----------------------|-------------|---|-----------|
| MicroRNA information | Arabidopsis | miRBase, release 15: ( <a href="http://www.mirbase.org/cgi-bin/mirna_summary.pl?org=ath">http://www.mirbase.org/cgi-bin/mirna_summary.pl?org=ath</a> )  | (6)       |
|                      | Rice        | miRBase, release 15: ( <a href="http://www.mirbase.org/cgi-bin/mirna_summary.pl?org=osa">http://www.mirbase.org/cgi-bin/mirna_summary.pl?org=osa</a> )  |           |
| Genomic information  | Arabidopsis | TAIR, release 9: ( <a href="ftp://ftp.arabidopsis.org/home/tair/Sequences/">ftp://ftp.arabidopsis.org/home/tair/Sequences/</a> )  | (8)       |
|                      | Rice        | TIGR release 6.1: ( <a href="ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/">ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/</a> )   | (9)       |
| SNP data             | Arabidopsis | TAIR, release 9: ( <a href="ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR9_genome_release/TAIR9_gff3/Variation_GFF/">ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR9_genome_release/TAIR9_gff3/Variation_GFF/</a> )  | (8,10)    |
|                      | Rice        | <a href="http://chibba.agtec.uga.edu/snp/index.php?table_name=-injaSnps">http://chibba.agtec.uga.edu/snp/index.php?table_name=-injaSnps</a><br>OryzaSNP Project: ( <a href="ftp://ftp.plantbiology.msu.edu/pub/data/Oryza_SNP/">ftp://ftp.plantbiology.msu.edu/pub/data/Oryza_SNP/</a> )  | (11–13)   |
| HTS data             | Arabidopsis | Plant MPSS databases: ( <a href="http://mpss.udel.edu/at/mpss_index.php">http://mpss.udel.edu/at/mpss_index.php</a> for MPSS data, and <a href="http://mpss.udel.edu/at_pare/">http://mpss.udel.edu/at_pare/</a> for PARE data)<br>GEO (Gene Expression Omnibus): ( <a href="http://www.ncbi.nlm.nih.gov/geo/">http://www.ncbi.nlm.nih.gov/geo/</a> ) | (14–19)   |
|                      | Rice        | Plant MPSS databases: ( <a href="http://mpss.udel.edu/rice/mpss_index.php">http://mpss.udel.edu/rice/mpss_index.php</a> for MPSS data, and <a href="http://mpss.udel.edu/rice_pare/">http://mpss.udel.edu/rice_pare/</a> for PARE data)<br>GEO: ( <a href="http://www.ncbi.nlm.nih.gov/geo/">http://www.ncbi.nlm.nih.gov/geo/</a> )                   |           |



**Table 2.** Statistical result of SNPs used for the construction of PmiRKB

| Species          | Subspecies               | Pre-miRNAs <sup>a</sup> | MiRNAs <sup>b</sup> | Target sites <sup>c</sup> | Others <sup>d</sup> | Sum <sup>e</sup> |
|------------------|--------------------------|-------------------------|---------------------|---------------------------|---------------------|------------------|
| Arabidopsis      | Bur-0                    | 148                     | 11                  | 239                       | 569 414             | 569 795          |
|                  | Tsu-1                    | 147                     | 11                  | 218                       | 500 987             | 501 349          |
|                  | Cvi-0                    | 213                     | 23                  | 343                       | 807 540             | 808 085          |
|                  | Ler-1                    | 66                      | 3                   | 145                       | 304 370             | 304 580          |
|                  | Bay-0                    | 79                      | 12                  | 119                       | 316 805             | 316 999          |
|                  | Sha                      | 105                     | 12                  | 156                       | 402 326             | 402 580          |
| Sum <sup>f</sup> |                          | 758                     | 72                  | 1220                      | 2 901 442           | 2 903 388        |
| Rice             | Tainung 67               | 0                       | 0                   | 0                         | 228                 | 228              |
|                  | Minghui 63               | 9                       | 0                   | 17                        | 60 681              | 60 707           |
|                  | Zhenshan 97B             | 7                       | 0                   | 12                        | 38 653              | 38 672           |
|                  | Sadu-Cho                 | 10                      | 0                   | 15                        | 60 634              | 60 659           |
|                  | N 22                     | 4                       | 0                   | 25                        | 58 266              | 58 295           |
|                  | FR13 A                   | 6                       | 1                   | 24                        | 66 304              | 66 334           |
|                  | Moroberekan              | 0                       | 0                   | 8                         | 22 098              | 22 106           |
|                  | Dom-Sufid                | 0                       | 0                   | 1                         | 2185                | 2186             |
|                  | Aswina                   | 6                       | 0                   | 18                        | 66 834              | 66 858           |
|                  | Dular                    | 9                       | 1                   | 30                        | 63 690              | 63 729           |
|                  | Rayada                   | 9                       | 1                   | 29                        | 66 053              | 66 091           |
|                  | Pokkali                  | 8                       | 0                   | 18                        | 55 738              | 55 764           |
|                  | IR64-21                  | 6                       | 0                   | 15                        | 58 244              | 58 265           |
|                  | Azucena                  | 2                       | 0                   | 4                         | 15 348              | 15 354           |
|                  | Shan-Huang Zhan-2        | 7                       | 0                   | 13                        | 48 100              | 48 120           |
|                  | Li-Jiang-Xin-Tuan-Hei-Gu | 0                       | 0                   | 1                         | 3206                | 3207             |
|                  | Swarna                   | 11                      | 1                   | 16                        | 54 020              | 54 047           |
|                  | Cypress                  | 2                       | 1                   | 7                         | 23 959              | 23 968           |
|                  | M 202                    | 0                       | 0                   | 0                         | 5493                | 5493             |
|                  | 93-11                    | 48                      | 6                   | 89                        | 382 279             | 382 416          |
| Sum <sup>f</sup> |                          | 144                     | 11                  | 342                       | 1 152 013           | 1 152 499        |

<sup>a</sup>Number of SNPs that reside in pre-miRNAs (precursor microRNAs) in each subspecies.

<sup>b</sup>Number of SNPs that reside in mature microRNAs in each subspecies.

<sup>c</sup>Number of SNPs that reside in the microRNA target recognition sites in each subspecies.

<sup>d</sup>Number of SNPs with other genomic locations in each subspecies.

<sup>e</sup>Total number of SNPs in each subspecies.

<sup>f</sup>Sum of the SNPs in all the subspecies belonging to each categories.

Note that, the SNPs within pre-miRNAs, mature microRNAs and microRNA target recognition sites may overlap with each other, resulting in the inconsistencies between 'Sum<sup>e</sup>' and the sum of all the categories ('Pre-miRNAs<sup>a</sup>', 'MiRNAs<sup>b</sup>', 'Target sites<sup>c</sup>' and 'Others<sup>d</sup>') in each subspecies.

to develop high-throughput methods, such as PARE (15-18). In this study, a large-scale investigation of *in vivo* miRNA-target pairs was performed based on the public PARE datasets. For this purpose, the PARE short reads were mapped to the mRNAs. The signatures containing overlapping regions with the predicted target sites were considered to be potential cleavage signals that support the specific miRNA-target regulatory relationships. Besides, the sample origin and the normalized expression level (in RPM) of each short read are provided for reference. Thus, the functional module, 'MiR-Tar', was established (Figure 2B).

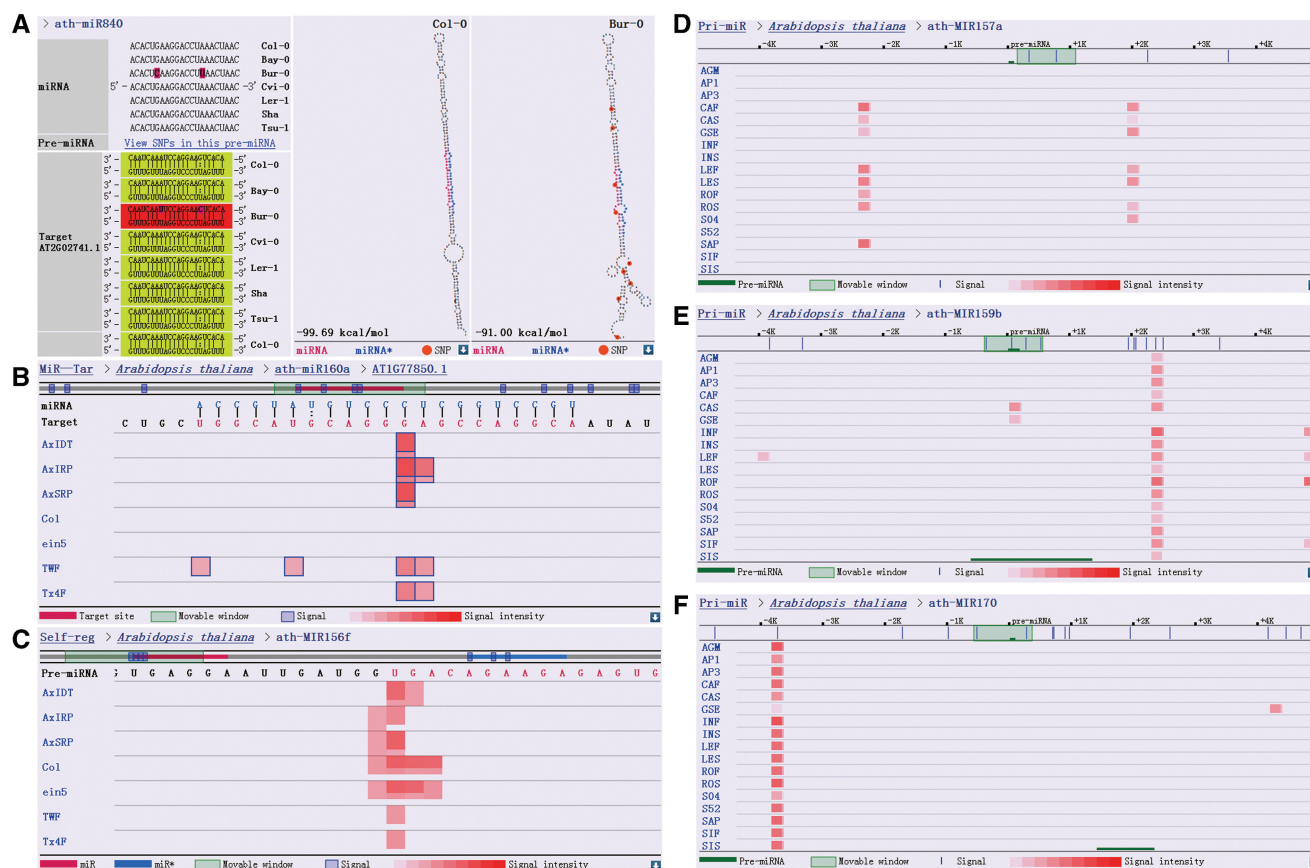
Theoretically, 3' remnants generated after the pri-miRNA processing by Dicer-like 1 (DCL1) can be included in plant degradome libraries and detected by PARE sequencing (17,24). Thus, the DCL1-mediated first-step cleavages on both strands of the stem region of a specific pri-miRNA, which result in two poly(A)-tailed remnants, can be validated by PARE data. On the other hand, mature miRNA, and occasionally its miRNA\*, can recognize their host precursor as a target based on the complementary sequence present in the stem region. This type of miRNA- or miRNA\*-mediated self-regulation can

be also represented by plant degradome PARE sequencing data (15,24). Therefore, the module, 'Self-reg', was created in PmiRKB to provide a comprehensive view of miRNA precursor processing and miRNA- or miRNA\*-mediated self-cleavages on their host precursors (Figure 2C). Similar to the 'MiR-Tar' module, the sample origin and the normalized expression level (in RPM) of each short read are provided for reference.

PmiRKB additionally includes the module, 'MiR info', which provides sequences of mature miRNAs and pre-miRNAs in FASTA format, clustered pre-miRNAs (within 50 kb), experimental evidences, references and external links to miRBase (6) and PMRD (7). Several useful web links for accessing miRNA-related data and additional analytical tools are provided by 'Useful links'. Beginners may also appreciate the resources provided in 'Instructions' and 'References' (Figure 1).

## DATABASE IMPLEMENTATION

PmiRKB, with web pages written in PHP scripting language, is maintained on an Apache2 HTTP server containing a PHP5 module. The connection between



**Figure 2.** Output data from the four major modules of PmiRKB. (A) An example of SNPs identified in the ‘SNP’ module (top left panel, highlighted in dark pink), which have the potential to influence the stability of miRNA–target RNA duplexes (lower left panel, indicated by different background colors). The effect of SNPs (indicated by orange circles) on the secondary structure of a specific pre-miRNA (precursor miRNA), ath-MIR840 for example, is illustrated on the right two panels. (B) Output data from the ‘MiR–Tar’ module that indicates ATIG77850.1 is a genuine target of ath-miR160a, based on the intense PARE signals detected in the middle of the miRNA–target duplex. The 100-nt sequence (gray horizontal line) surrounding the target site (red horizontal line) is shown. The blue outlines of the PARE signals denote the signatures that possess unique loci in the corresponding transcriptome. (C) The result given by ‘Self-reg’ indicates that the Dicer-like 1 (DCL1)-mediated first-step cleavages on the stem region of a specific pri-miRNA (primary miRNA) can be reflected by PARE data. In most cases, cleavage signals generated from the left arms of the pri-miRNAs were much easier to be detected than those from the right arms, likely due to the longer remnants generated from the left ones after DCL1-mediated cleavages. Cleavage signals within the miRNA\* region of the pre-miRNA, ath-MIR156f, may represent miRNA-mediated self-regulation. For both (B and C), the signal intensity at a specific position is enhanced by the supporting of distinct signatures (represented by the red overlapping signals in different lengths in vertical orientation). (D, E and F) are output data from the ‘Pri-miR’ module. (D) represents tissue-specific transcription signals associated with the 3’-downstream region of ath-MIR157a, which could be hardly observed in the reproductive organs including inflorescences and siliques (the libraries lack of related signals, i.e. AGM, API, AP3, INF, INS, SIF, and SIS, were prepared from the reproductive organs of Arabidopsis). The transcription signals with distinct tissue-origin patterns were observed surrounding certain pre-miRNAs, such as ath-MIR159b (E) and ath-MIR170 (F), indicating that independent transcription units may exist. For (D, E and F), the signal intensity at a specific position is enhanced by the supporting of distinct signatures (represented by the red overlapping signals in different lengths in horizontal orientation).

PmiRKB and the PostgreSQL 8.4 database, containing all of the analytical results, was accomplished by the PostgreSQL client extension of PHP5. Dynamic graphics provided in the four major modules were implemented by SVG, while a few Javascript codes were employed for the dynamic effects experienced on the web pages, such as the movable windows.

## FEATURES

Although huge quantities of plant SNP data are currently available, few analyses have been applied for a large-scale investigation of their biological significances for specific types of genes, such as miRNAs. Based on the analytical

results accessible from our PmiRKB, SNPs with the potential to alter the secondary structures of pre-miRNAs can be identified. Furthermore, the SNPs within the predicted miRNA–target duplexes, which may interrupt miRNA–target regulation, were also included in the ‘SNP’ module of PmiRKB.

Technological advances in sequencing have greatly accelerated the studies on plant transcriptomes (5). However, analyses of these newly generated HTS data remain incomplete. Here, the ingenious use of HTS data derived from poly(A)-tailed transcripts enabled us to uncover the transcriptional contexts of pri-miRNAs. Considering the lack of information regarding pri-miRNAs in the current miRNA databases, the ‘Pri-miR’

module is a valuable resource for the study of plant miRNA transcription. Specifically, it can facilitate primer or probe designing for pri-miRNA detection. Besides, the level of a specific pri-miRNA will not only indicate the extent of miRNA transcription, but also the stability of the pri-miRNA *in vivo*. Thus, 'Pri-miR' will also provide valuable insights into the efficiency of pri-miRNA processing in a specific tissue.

Another valuable resource provided by PmiRKB is the whole-transcriptome identification of miRNA–target pairs *in planta*, based on the PARE data (15–17). For researchers who intend to validate miRNA–target pairs of interest, the 'MiR–Tar' module can provide the information of cleavage signal intensities. Thus, the priority of each miRNA–target pair to be validated could be determined before modified 5' RACE assays. A sizable portion of 3' cleavage remnants, especially those at trace levels, are difficult to be detected by the traditional RACE method. Instead, they could be reflected by PARE data, considering the in-depth feature of HTS technology. Together, the 'MiR–Tar' module provides a comprehensive set of miRNA–target pair candidates, including novel pairs with subtle regulatory relationships. Finally, the 'Self-reg' module provides novel insights into miRNA precursor processing and miRNA- or miRNA\*-mediated self-regulation of miRNA genes.

## FUTURE DEVELOPMENT

As the interest in the miRNA world grows, more and more datasets are being generated. Correspondingly, there is an urgent need to organize, parse, and analyze the data produced. This can be partially accomplished through the development of both comprehensive and specialized databases. Here, the creation of a plant miRNA knowledge base, PmiRKB, has been described. The potential for its four functional modules to provide valuable insights into the transcription, processing and regulation of miRNA genes has been demonstrated.

PmiRKB has been updated based on the current release of miRBase (from release 14 to 15) (6). In the near future, more plant species will be included, and PmiRKB will be timely updated following the new release of miRBase (6). In addition, as more HTS data will become publicly available, the results provided by the four modules of PmiRKB could be further refined.

## ACKNOWLEDGEMENTS

The authors thank Dr Ramanjulu Sunkar for his kindness to provide a free access to the valuable rice degradome sequencing data during the construction of PmiRKB. The authors thank Dr Christian Klukas for his kind discussions. The authors also thank Dr Michael Galperin and the two anonymous referees for their constructive and helpful suggestions.

## FUNDING

Funding for open access charge: National High Technology Research and Development Program of

China ("863" Program) (2008AA10Z125); National Natural Sciences Foundation of China (30771326, 30971743, 31050110121); Program for New Century Excellent Talents in University of China (NCET-07-0740).

*Conflict of interest statement.* None declared.

## REFERENCES

- Kim, V.N., Han, J. and Siomi, M.C. (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell. Biol.*, **10**, 126–139.
- Voynet, O. (2009) Origin, biogenesis, and activity of plant microRNAs. *Cell*, **136**, 669–687.
- Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B. (2006) MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.*, **57**, 19–53.
- Morozova, O. and Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics*, **92**, 255–264.
- Simon, S.A., Zhai, J., Nandety, R.S., McCormick, K.P., Zeng, J., Mejia, D. and Meyers, B.C. (2009) Short-read sequencing technologies for transcriptional analyses. *Annu. Rev. Plant Biol.*, **60**, 305–333.
- Griffiths-Jones, S., Saini, H.K., van Dongen, S. and Enright, A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
- Zhang, Z., Yu, J., Li, D., Zhang, Z., Liu, F., Zhou, X., Wang, T., Ling, Y. and Su, Z. (2010) PMRD: plant microRNA database. *Nucleic Acids Res.*, **38**, D806–D813.
- Huala, E., Dickerman, A.W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M., Huang, W. *et al.* (2001) The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.*, **29**, 102–105.
- Yuan, Q., Ouyang, S., Liu, J., Suh, B., Cheung, F., Sultana, R., Lee, D., Quackenbush, J. and Buell, C.R. (2003) The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. *Nucleic Acids Res.*, **31**, 229–233.
- Clark, R.M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T.T., Fu, G., Hinds, D.A. *et al.* (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science*, **317**, 338–342.
- Feltus, F.A., Wan, J., Schulze, S.R., Estill, J.C., Jiang, N. and Paterson, A.H. (2004) An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. *Genome Res.*, **14**, 1812–1819.
- McNally, K.L., Childs, K.L., Bohnert, R., Davidson, R.M., Zhao, K., Ulat, V.J., Zeller, G., Clark, R.M., Hoen, D.R., Bureau, T.E. *et al.* (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl Acad. Sci. USA*, **106**, 12273–12278.
- McNally, K.L., Bruskiwich, R., Mackill, D., Buell, C.R., Leach, J.E. and Leung, H. (2006) Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. *Plant Physiol.*, **141**, 26–31.
- Nakano, M., Nobuta, K., Vemaraju, K., Tej, S.S., Skogen, J.W. and Meyers, B.C. (2006) Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. *Nucleic Acids Res.*, **34**, D731–D735.
- German, M.A., Pillay, M., Jeong, D.H., Hetawal, A., Luo, S., Janardhanan, P., Kannan, V., Rymarquis, L.A., Nobuta, K., German, R. *et al.* (2008) Global identification of microRNA–target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.*, **26**, 941–946.
- Zhou, M., Gu, L., Li, P., Song, X., Wei, L., Chen, Z. and Cao, X. (2010) Degradome sequencing reveals endogenous small RNA targets in rice (*Oryza sativa* L. ssp. *indica*). *Front. Biol.*, **5**, 67–90.
- Li, Y.F., Zheng, Y., Addo-Quaye, C., Zhang, L., Saini, A., Jagadeeswaran, G., Axtell, M.J., Zhang, W. and Sunkar, R. (2010)

- Transcriptome-wide identification of microRNA targets in rice. *Plant J.*, **62**, 742–759.
18. Addo-Quaye,C., Eshoo,T.W., Bartel,D.P. and Axtell,M.J. (2008) Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Curr. Biol.*, **18**, 758–762.
  19. Zhang,G., Guo,G., Hu,X., Zhang,Y., Li,Q., Li,R., Zhuang,R., Lu,Z., He,Z., Fang,X. *et al.* (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.*, **20**, 646–654.
  20. Zhang,Y. (2005) miRU: an automated plant miRNA target prediction server. *Nucleic Acids Res.*, **33**, W701–W704.
  21. Hofacker,I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.
  22. Brenner,S., Johnson,M., Bridgham,J., Golda,G., Lloyd,D.H., Johnson,D., Luo,S., McCurdy,S., Foy,M., Ewan,M. *et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.*, **18**, 630–634.
  23. Lee,Y., Kim,M., Han,J., Yeom,K.H., Lee,S., Baek,S.H. and Kim,V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.*, **23**, 4051–4060.
  24. Meng,Y., Gou,L., Chen,D., Wu,P. and Chen,M. (2010) High-throughput degradome sequencing can be used to gain insights into microRNA precursor metabolism. *J. Exp. Bot.*, [Epub ahead of print, July 19, 2010; doi: 10.1093/jxb/erq209].