PsRobot: a web-based plant small RNA meta-analysis toolbox

Hua-Jun Wu[1,2], Ying-Ke Ma[1,2], Tong Chen[1,2], Meng Wang[1,*] and Xiu-Jie Wang[1,*]

[1]The State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China and [2]Graduate University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

Small RNAs (smRNAs) in plants, mainly microRNAs and small interfering RNAs, play important roles in both transcriptional and post-transcriptional gene regulation. The broad application of high-throughput sequencing technology has made routinely generation of bulk smRNA sequences in laboratories possible, thus has significantly increased the need for batch analysis tools. PsRobot is a web-based easy-to-use tool dedicated to the identification of smRNAs with stem-loop shaped precursors (such as microRNAs and short hairpin RNAs) and their target genes/transcripts. It performs fast analysis to identify smRNAs with stem-loop shaped precursors among batch input data and predicts their targets using a modified Smith–Waterman algorithm. PsRobot integrates the expression data of smRNAs in major plant smRNA biogenesis gene mutants and smRNA-associated protein complexes to give clues to the smRNA generation and functional processes. Besides improved specificity, the reliability of smRNA target prediction results can also be evaluated by mRNA cleavage (degradome) data. The cross species conservation statuses and the multiplicity of smRNA target sites are also provided. PsRobot is freely accessible at http://omicslab.genetics.ac.cn/psRobot/.

INTRODUCTION

MicroRNAs (miRNAs) and small interfering RNAs (siRNAs) are two major classes of endogenous regulatory small RNAs (smRNAs) in plants. They are usually 21–24 nucleotides (nt) long, and both function by pairing to targets via sequence complementarity (1). miRNAs are usually generated from limited genomic loci and mainly work post-transcriptionally to down-regulate target mRNAs, whereas siRNAs have much broader origins and can function at both transcriptional and post-transcriptional levels (2–6). Both miRNAs and siRNAs are typically identified by cloning and sequencing of small size RNAs (7). The development and application of the high-throughput sequencing technology have significantly advanced the studies on smRNAs, but also imposed increasing numbers of laboratories facing the tasks of data analysis.

Plant miRNAs are usually ~21 nt long and processed from the pairing stem region of longer precursors with hairpin-shaped secondary structures (8). The presence of stem-loop precursor as the lowest energy folding form has been considered as one of the key criteria for the identification of new miRNAs (9,10). However, as many genomic loci giving rise to siRNAs can also be folded into hairpin-shaped structures (11,12), searching for miRNAs by their expression evidence and the presence of stem-loop precursors may yield many false-positive candidates. According to the community-agreed plant miRNA annotation criteria (9,13), the presence of dominantly expressed candidate sequence and the detection of the pairing sequence (miRNA*) are required for miRNAs (13). These constraints improved the prediction specificity, but the results are still far from ideal.

It has been shown that plant miRNAs mainly pair with their targets via nearly perfect sequence complementarity, very similar to the manner of siRNAs (9,13). This has made the predictions for plant miRNA targets relatively straightforward, yielding limited numbers of targets per miRNA rather than hundreds of ones in animals (9,13). Yet, it still requires large amount of work to identify real miRNA targets from the predicted candidate list. In

*To whom correspondence should be addressed. Tel: +86 10 64806590; Fax: +86 10 64873428; Email: xjwang@genetics.ac.cn
Correspondence may also be addressed to Meng Wang. Tel: +86 10 64806591; Fax: +86 10 64873428; Email: mengwang@genetics.ac.cn

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

addition, increasing lines of evidence have shown that gaps are also tolerable in the pairing between plant miRNAs and their targets (14), which will enlarge the candidate target list and produce more difficulties for experimental validation (15).

Most of the available smRNA target prediction software was designed for animal miRNAs and not ideal for plant data because of the common presence of large bulges in animal miRNA and target alignments as well as the different contributions of miRNA seed sequences or central sequences to the stability of miRNA: target pairs of animals and plants. Several recently developed plant miRNA target prediction tools, such as targetFinder (16,17), psRNATarget (18) and CleaveLand (19), have provided great help to researches. However, they are also limited by the requirement of local installation, lack of degradome data supports or dependence on the target prediction results of third-party software (Supplementary Table S1).

PsRobot is designed to partially solve the aforementioned problems in plant miRNA and target prediction. It incorporates the commonly agreed criteria to identify smRNAs with stem-loop shaped precursors from user-uploaded sequences and predict their targets. The multiple user adjustable parameters made the software able to meet different needs from users. To facilitate better classification and functional analysis of input sequences, psRobot integrates the expression information of input sequences in reported plant smRNA binding protein pull-down assay or mutants of major smRNA biogenesis pathway genes. For example, strong association with ARGONAUTE1 (AGO1) protein and expression depletion in *dcl1* mutant will strengthen the confidence of a smRNA with stem-loop precursor as an miRNA (1). It also incorporates the available mRNA degradome data for users to evaluate the reliability of miRNA target prediction results. The multiplicity of miRNA binding sites on a single target as well as the cross species conservation status of the target sites are also analyzed and provided. PsRobot can be either used online or downloaded and installed locally. The local version offers a larger capacity for input data size and has the function to incorporate user-uploaded degradome data.

METHODS AND RESULTS

The Stem-loop smRNA Prediction Function

Input, algorithm and parameters
The stem-loop smRNA prediction function takes input smRNA sequences in FASTA or plain text format. For each query smRNA, the software finds its perfectly matched genomic origins, and extracts various lengths of upstream and downstream sequences as precursors, assuming that the smRNA may originate from either the 5′ or 3′ end of precursors, with 10 nt extension at one end of the precursor each time, till reaching the user defined precursor length. The secondary structures of the extracted precursor sequences are then evaluated by the MFOLD program (20). Precursor sequences with stem-loop structure as the minimal free energy folding

form and the corresponding query smRNA will be selected and reported in the result pages (Fig. 1a), either in html or text format. If the conservation analysis function of smRNA sequences is selected, the cross species conservation status of smRNAs will be analyzed by aligning the query small sequences to eight selected plant genomes using the BLAST program (allowing up to two mismatches in the smRNA sequences), and the ClustalW (21) alignment of the identified smRNA homologous sequences will be included in the outputs (Fig 1b). The repetitive sequence regions of the 26 preloaded genomes were identified by the RepeatMasker program (http://repeatmasker.org) and stored in the background database. Every genomic locus of the query sequences will be searched against the database to identify repeat sequence originated smRNAs. Parameters for users to adjust include smRNA conservation analysis, the minimal and maximal numbers of mismatched nucleotides within the query smRNA sequences in the obtained precursor structures, the maximal lengths of extracted precursors, and the permission of large loop sequence in the qualified precursors. Although the precursor structures of most canonical plant miRNAs are very short and well paired, there are still some with large bulge or hairpin loops, such as ath-MIR393a and ath-MIR167d (22,23). Enable the ''Retain large loop small RNA'' function will include precursors with large loops in the prediction results.

Data collection
A preloaded species needs to be selected to define the origins of smRNAs. Up to 26 completed plant genomes are currently supported by psRobot, and future finished genomes will also be incorporated on their release. The *Arabidopsis thaliana* genome was downloaded from the Arabidopsis Information Resource (TAIR) (24) and the rice genome was from Rice Genome Annotation Project (RGAP) (25) and the Rice Annotation Project Database (RAP-DB) (26) databases. Other plant genomes were downloaded from the Phytozome genome database (27). SmRNA deep sequencing data in smRNA biogenesis mutants and AGO-associated libraries were collected from the NCBI Gene Expression Omnibus database with datasets GSE11094, GSE14695, GSE16959, GSE13605, GSE10036, GSE6682, GSE5343 and GSE6682 for *Arabidopsis thaliana* (28–35) and datasets GSE20748 and GSE18250 for rice (36,37).

smRNA biogenesis and functional data
PsRobot incorporates published smRNA sequencing data from major plant smRNA biogenesis and function associated protein complexes or gene mutants and returns this information together with the stem-loop prediction results. If a query sequence presents in any of the preloaded database, its sequence reads in each database will be listed (Fig. 1c). As it has been shown that smRNAs of different origins are processed by different Dicer-like (DCL) family proteins and associate with different AGO protein complexes, such information will facilitate users to evaluate the types and functions of the query smRNAs. For example, decreased expression in *dcl1* mutant together
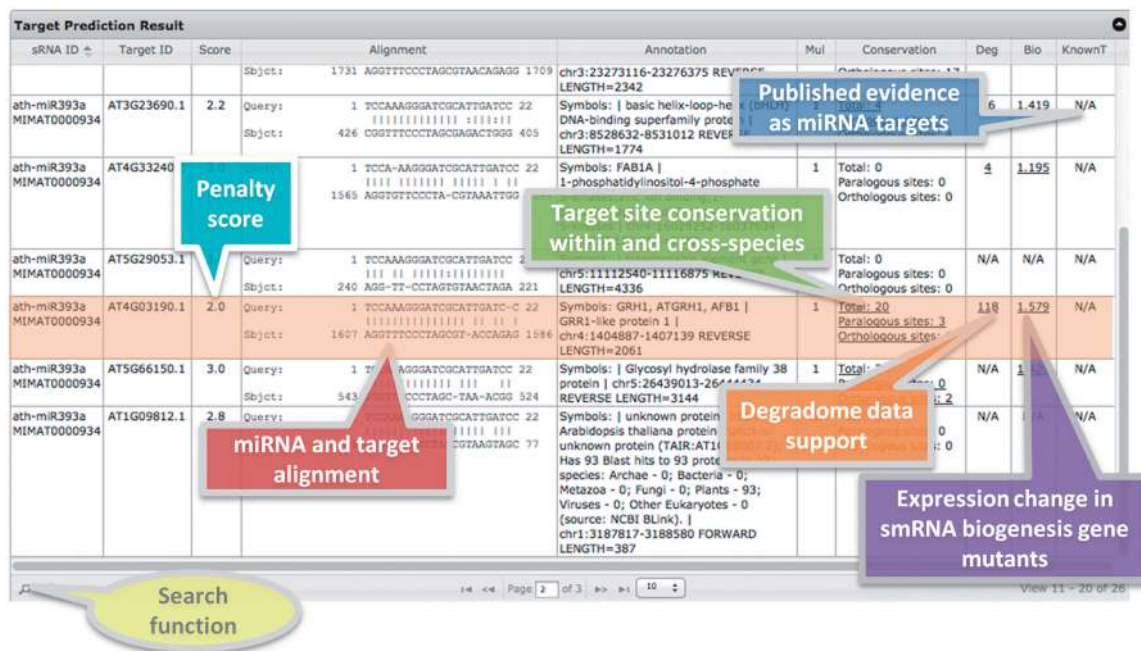
Figure 1. Output of stem-loop smRNA prediction. (**A**) The genomic mapping and stem-loop precursor prediction of the query smRNAs, detailed information of genomic location, precursor sequence, secondary structure of each predicted stem-loop smRNA loci and their folding energies are included. Sequence in red and capitalized letters is the query smRNA sequence; (**B**) shows the conservation status of the smRNA in 8 plant species; (**C**) shows the normalized reads of a given smRNA in smRNA biogenesis mutants and AGO-associated libraries.

with enrichment in the AGO1 association libraries provide strong evidence for authentic miRNAs (1), whereas decreased expression in *dcl3* mutant together with enrichment in the AGO4 association libraries indicate the smRNAs as siRNAs (38).

Performance
To evaluate the accuracy of stem-loop smRNA prediction function, all miRBase (release v17) (39) recorded *Arabidopsis thaliana* miRNAs were collected and used as the test dataset. Among the 213 nonredundant sequences, 202 (~94%) miRNAs were successfully identified as stem-loop smRNAs.

The smRNA Target Prediction Function

Input and parameters
To use the smRNA target prediction function, users can either select the known miRNAs from a plant species or submit their own sequences. The software will search for target sites among the pre-loaded genes/transcripts of the corresponding genome or the user-uploaded target library. The uploaded query smRNA sequences and target library should be either in FASTA or plain text format. The parameters for users to adjust include the following: (1) penalty score for the alignment between smRNAs and targets, which is defined by the formulas below; (2) the boundaries of essential sequence region, within which

mismatches or gaps will receive double penalty scores than other regions; (3) the threshold for the total number of gaps within the smRNA and target alignment region; and (4) the region within which gaps are permitted. Degradome sequences mapped within the target sites will be analyzed and presented. Only preloaded degradome data are available for the online version of psRobot, yet users can incorporate their own degradome data via the psRobot_deg program once psRobot is installed locally.

Algorithms
As the pairings between smRNAs and target mRNAs involve global sequence alignment of smRNAs and local sequence alignment of targets, we applied a modified Smith–Waterman algorithm (Formula 1) (40) with the defined scoring system (Formula 2) to calculate the alignment scores between the query smRNAs and targets. The penalty score of each candidate alignment is obtained by subtracting the actual alignment score from the ideal perfect global pairing score (Formula 3). Alignments meet the penalty score cutoff will be backtraced and reported in the result page (Fig. 2). Exhaustive search will be performed on each mRNA to search for the potential presence of multiple target sites (target multiplicity). Parallel computing method is used to accelerate the speed. A standalone local version of this function is also available for download to facilitate analysis on large datasets.

Figure 2. Result summary table of smRNA target prediction function.

Formula 1:

$$S(0,j) = 0, 0 \leq j \leq n; S(i,0) = 0, 0 \leq i \leq m;$$

$$S(i,j) = \max \begin{cases} S[i-1][j-1] + \omega(q_i, r_j) \\ S[i-1][j] + \omega(q_i, -) \\ S[i][j-1] - \omega(match) + \omega(-, r_j) \end{cases} \left| \begin{array}{c} Match/Mismatch/(G:Upair) \\ Insertion \\ Deletion \end{array} \right\}, 0 \leq i \leq m, 0 \leq j \leq n$$

notes: q represents smRNA sequences; r represents target sequences; m = length(q); n = length(r).

Formula 2:

$$\omega(match) = 3, \omega(mismatch) = -1, \omega(G:Upair)$$
$$= 1, \omega(deletion) = \omega(insertion) = -1.1$$

notes: deletion represents gaps on query smRNAs; insertion represents gaps on candidate targets.

Formula 3:

$$P(m,j) = m \times \omega(match) - S(m,j)$$

Output of the smRNA target prediction function

The primary output of the smRNA target prediction function is summarized in a sortable and searchable table (Fig. 2), with the query sequence, target alignment, alignment penalty score, target annotation, multiplicity of target sites and other supporting information. The contents of the expandable links in the result table are summarized in the next sections.

Target site conservation

Homolog gene groups of eight plant species (*Arabidopsis thaliana, Brachypodium distachyon, Carica papaya, Oryza sativa, Populus trichocarpa, Sorghum bicolor, Vitis vinifera,* *Zea mays*) are generated using OrthoMCL (41) with the default parameters, and serve as the source data for the conservation analysis of target sites. Predicted targets are searched against this source data for both paralogous and orthologous sequences with conserved target sites (Fig. 2). The multiple alignments of the conserved target sites and the alignments between smRNA and targets for homolog genes can be viewed via the hyperlink in the "Conservation" column (Fig. 3b and c).

Degradome data

It has been shown that the miRNA cleavage products of targets can be cloned and detected by high-throughput sequencing technology, generating mRNA degradome data among which the 5′ ends of sequences mark the cleavage sites of miRNAs (28,37,42–44). PsRobot integrates well-produced datasets of *Arabidopsis thaliana* (GSE11094) and rice (GSE18248) in the target prediction results (Fig. 2) (28,37). The abundance of degradome sequences (after normalized to reads per million, RPM) is marked at the starting genomic loci of each sequence (Fig. 3d). Candidate target sites with abundant degradome sequences are more likely to be authentic miRNA targets. The position with the most abundant degradome sequences should represent the miRNA cleavage site.

Figure 3. Illustration of expandable links in the target prediction result table. (**A**) A sample line of the target prediction result table; (**B**) target site conservation analysis results. Homolog genes with the same smRNA target site, both within and cross-species, are shown in the ClustalW multiple alignment format; (**C**) the alignments between miRNA and targets provoked by mousing over the hyperlinks of gene IDs listed in (B); (**D**) Degradome data of the predicted smRNA target. Positions listed in the table are orders of nucleotides in the smRNA sequences, plus upstream and downstream 5 nt each; 'Loci number' represents the number of perfectly mapped loci of each degradome sequence, marked at the start nucleotide position of each sequence. Normalized reads starting at each nucleotide position in the degradome data are listed in following rows. The histogram shows the distribution of the degradome reads. In the example, the target should be cleaved after the 12th nucleotide according to the degradome data; (**E**) Expression change of target gene in smRNA biogenesis pathway gene mutants compared to wild-type plants. Fold changes of the normalized expression values between mutants and wildtype are shown in the table and the Y axis of the plot. Data from different datasets are distinguished by colors in the plot.

Target expression in smRNA biogenesis mutants

As the production of smRNAs will be significantly impaired in the mutants of genes involved in plant smRNA biogenesis pathways, such as the *dcl, hyl1, hen1* and *rdr* family genes (38), expression increment of genes in these mutants may indicate an inhibitory target effect by smRNAs. To facilitate inspections from this perspective, psRobot collected the published microarray data and integrated them in the target prediction results (Fig. 3e). Currently, this function is only available for *Arabidopsis*

thaliana (GSE2473, GSE3011, GSE24887) (45,46) and will be expanded to other species on the availability of required data.

Performance

To test the reliability of the smRNA target prediction results, we selected 75 *Arabidopsis thaliana* miRNAs with at least one reported target in the ASRP database (Supplementary Table S2) (47). A total of 995 genes were predicted as the targets of the 75 miRNAs using the

default parameters, of which 306 (31%) targets were reported by the ASRP database, representing 89% of the 344 validated *Arabidopsis* miRNA targets (Supplementary Table S1). Significant improvement of the prediction results was achieved by combining the information of target site conservation, detectability in degradome data and expression change in smRNA biogenesis mutants as filters, as demonstrated by the reduction of total predicted targets to 542, of which 292 (54%) were confirmed targets from ASRP (85% of validated targets) (Supplementary Table S1).

CONCLUSIONS

Computational prediction of miRNAs and their targets have suffered high false-positive rate because of limited constraints to apply. To generate more specific prediction results, psRobot integrated the biogenesis and protein association information of plant smRNAs, as well as conservation, cleavage and smRNA dependency information for mRNAs. These information can facilitate users to quickly identify *bona fide* miRNAs or other functional stem-loop smRNAs and their candidate targets. The ability to handle both single or batch sequence input and the availability of online and local version of the software renders it high flexibility in application.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2.

ACKNOWLEDGEMENTS

The authors thank Dr. Guan-Zheng Luo for the discussion of web structure. The authors also thank Guihai Feng, Yong-Xin Liu and Xin-Min Zhang for testing the local software.

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. Vazquez,F., Vaucheret,H., Rajagopalan,R., Lepers,C., Gasciolli,V., Mallory,A.C., Hilbert,J.L., Bartel,D.P. and Crete,P. (2004) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol. Cell.*, **16**, 69–79.
3. Borsani,O., Zhu,J., Verslues,P.E., Sunkar,R. and Zhu,J.K. (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell*, **123**, 1279–1291.
4. Wang,X.J., Gaasterland,T. and Chua,N.H. (2005) Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana. Genome biology*, **6**, R30.
5. Zhang,Y., Liu,X.S., Liu,Q.R. and Wei,L. (2006) Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Res.*, **34**, 3465–3475.
6. Wang,H., Chua,N.H. and Wang,X.J. (2006) Prediction of trans-antisense transcripts in Arabidopsis thaliana. *Genome Biol.*, **7**, R92.
7. Reinhart,B.J., Weinstein,E.G., Rhoades,M.W., Bartel,B. and Bartel,D.P. (2002) MicroRNAs in plants. *Genes Dev.*, **16**, 1616–1626.
8. Mallory,A.C. and Vaucheret,H. (2006) Functions of microRNAs and related small RNAs in plants. *Nat. Genet.*, **38 Suppl**, S31–S36.
9. Rhoades,M.W., Reinhart,B.J., Lim,L.P., Burge,C.B., Bartel,B. and Bartel,D.P. (2002) Prediction of plant microRNA targets. *Cell*, **110**, 513–520.
10. Meyers,B.C., Simon,S.A. and Zhai,J. (2010) MicroRNA processing: battle of the bulge. *Curr. Biol.: CB*, **20**, R68–R70.
11. Zilberman,D., Cao,X., Johansen,L.K., Xie,Z., Carrington,J.C. and Jacobsen,S.E. (2004) Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.*, **14**, 1214–1220.
12. Dunoyer,P., Brosnan,C.A., Schott,G., Wang,Y., Jay,F., Alioua,A., Himber,C. and Voinnet,O. (2010) An endogenous, systemic RNAi pathway in plants. *EMBO J.*, **29**, 1699–1712.
13. 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. *et al.* (2008) Criteria for annotation of plant MicroRNAs. *Plant Cell*, **20**, 3186–3190.
14. Jones-Rhoades,M.W. and Bartel,D.P. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell*, **14**, 787–799.
15. Dai,X., Zhuang,Z. and Zhao,P.X. (2010) Computational analysis of miRNA targets in plants: current status and challenges. *Brief. Bioinform.*, **12**, 115–121.
16. Allen,E., Xie,Z., Gustafson,A.M. and Carrington,J.C. (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, **121**, 207–221.
17. 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., Dangl,J.L. *et al.* (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PloS one*, **2**, e219.
18. Dai,X. and Zhao,P.X. (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.*, **39**, W155–W159.
19. Addo-Quaye,C., Miller,W. and Axtell,M.J. (2009) CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics*, **25**, 130–131.
20. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
21. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
22. Wang,X.J., Reyes,J.L., Chua,N.H. and Gaasterland,T. (2004) Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. *Genome Biol*, **5**, R65.
23. Sunkar,R. and Zhu,J.K. (2004) Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *Plant Cell*, **16**, 2001–2019.
24. Lamesch,P., Berardini,T.Z., Li,D., Swarbreck,D., Wilks,C., Sasidharan,R., Muller,R., Dreher,K., Alexander,D.L., Garcia-Hernandez,M. *et al.* (2011) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.*, **40**, D1202–D1210.
25. Ouyang,S., Zhu,W., Hamilton,J., Lin,H., Campbell,M., Childs,K., Thibaud-Nissen,F., Malek,R.L., Lee,Y., Zheng,L. *et al.* (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.*, **35**, D883–D887.

26. Tanaka,T., Antonio,B.A., Kikuchi,S., Matsumoto,T., Nagamura,Y., Numa,H., Sakai,H., Wu,J., Itoh,T., Sasaki,T. *et al.* (2008) The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Res.*, **36**, D1028–D1033.

27. Goodstein,D.M., Shu,S., Howson,R., Neupane,R., Hayes,R.D., Fazo,J., Mitros,T., Dirks,W., Hellsten,U., Putnam,N. *et al.* (2011) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.*, **40**, D1178–D1186.

28. 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.

29. Fahlgren,N., Sullivan,C.M., Kasschau,K.D., Chapman,E.J., Cumbie,J.S., Montgomery,T.A., Gilbert,S.D., Dasenko,M., Backman,T.W., Givan,S.A. *et al.* (2009) Computational and analytical framework for small RNA profiling by high-throughput sequencing. *RNA*, **15**, 992–1002.

30. Hardcastle,T.J. and Kelly,K.A. (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*, **11**, 422.

31. Montgomery,T.A., Yoo,S.J., Fahlgren,N., Gilbert,S.D., Howell,M.D., Sullivan,C.M., Alexander,A., Nguyen,G., Allen,E., Ahn,J.H. *et al.* (2008) AGO1-miR173 complex initiates phased siRNA formation in plants. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 20055–20062.

32. Mi,S., Cai,T., Hu,Y., Chen,Y., Hodges,E., Ni,F., Wu,L., Li,S., Zhou,H., Long,C. *et al.* (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell*, **133**, 116–127.

33. Kasschau,K.D., Fahlgren,N., Chapman,E.J., Sullivan,C.M., Cumbie,J.S., Givan,S.A. and Carrington,J.C. (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS. Biol.*, **5**, e57.

34. Lu,C., Kulkarni,K., Souret,F.F., MuthuValliappan,R., Tej,S.S., Poethig,R.S., Henderson,I.R., Jacobsen,S.E., Wang,W., Green,P.J. *et al.* (2006) MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. *Genome Res.*, **16**, 1276–1288.

35. Howell,M.D., Fahlgren,N., Chapman,E.J., Cumbie,J.S., Sullivan,C.M., Givan,S.A., Kasschau,K.D. and Carrington,J.C. (2007) Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in Arabidopsis reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell*, **19**, 926–942.

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

37. Wu,L., Zhang,Q., Zhou,H., Ni,F., Wu,X. and Qi,Y. (2009) Rice MicroRNA effector complexes and targets. *Plant Cell*, **21**, 3421–3435.

38. Vazquez,F. (2006) Arabidopsis endogenous small RNAs: highways and byways. *Trends Plant Sci.*, **11**, 460–468.

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

40. Smith,T.F. and Waterman,M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.

41. Li,L., Stoeckert,C.J. Jr and Roos,D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.*, **13**, 2178–2189.

42. 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.

43. Pantaleo,V., Szittya,G., Moxon,S., Miozzi,L., Moulton,V., Dalmay,T. and Burgyan,J. (2010) Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *Plant J*, **62**, 960–976.

44. Zhou M, G.L., Li P, Song X, Wei L, Chen Z, Cao X. (2010) Degradome sequencing reveals endogenous small RNA targets in rice (Oryza sativa L. ssp. indica). *Front. Biol.*, 5,67–90.

45. Xie,Z., Allen,E., Wilken,A. and Carrington,J.C. (2005) DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 12984–12989.

46. Willmann,M.R., Mehalick,A.J., Packer,R.L. and Jenik,P.D. (2011) MicroRNAs regulate the timing of embryo maturation in Arabidopsis. *Plant Physiol.*, **155**, 1871–1884.

47. Gustafson,A.M., Allen,E., Givan,S., Smith,D., Carrington,J.C. and Kasschau,K.D. (2005) ASRP: the Arabidopsis Small RNA Project Database. *Nucleic Acids Res.*, **33**, D637–D640.