Available online at www.sciencedirect.com
ScienceDirect
www.sciencedirect.com/science/journal/16720229
GENOMICS PROTEOMICS & BIOINFORMATICS

Article

Computational Analysis of Drought Stress-Associated miRNAs and miRNA Co-Regulation Network in *Physcomitrella patens*

Ping Wan[1#*], Jun Wu[1#], Yuan Zhou[1], Junshu Xiao[1], Jie Feng[2], Weizhong Zhao[2], Shen Xiang[1], Guanglong Jiang[1], and Jake Y. Chen[3,4,5*]

[1]*College of Life Sciences, Capital Normal University, Beijing 100048, China;*
[2]*School of Mathematical Sciences, Capital Normal University, Beijing 100048, China;*
[3]*School of Informatics, Indiana University-Purdue University Indianapolis, IN 46202, USA;*
[4]*Department of Computer & Information Science, Purdue University, IN 47907, USA;*
[5]*Indiana Center for Systems Biology and Personalized Medicine, IN 46202, USA.*

Genomics Proteomics Bioinformatics 2011 Apr; 9(1-2): 37-44 DOI: 10.1016/S1672-0229(11)60006-5

Received: Aug 31, 2010; Accepted: Dec 15, 2010

Abstract

miRNAs are non-coding small RNAs that involve diverse biological processes. Until now, little is known about their roles in plant drought resistance. *Physcomitrella patens* is highly tolerant to drought; however, it is not clear about the basic biology of the traits that contribute *P. patens* this important character. In this work, we discovered 16 drought stress-associated miRNA (DsAmR) families in *P. patens* through computational analysis. Due to the possible discrepancy of expression periods and tissue distributions between potential DsAmRs and their targeting genes, and the existence of false positive results in computational identification, the prediction results should be examined with further experimental validation. We also constructed an miRNA co-regulation network, and identified two network hubs, miR902a-5p and miR414, which may play important roles in regulating drought-resistance traits. We distributed our results through an online database named *ppt*-miRBase, which can be accessed at http://bioinfor.cnu.edu.cn/ppt_miRBase/index.php. Our methods in finding DsAmR and miRNA co-regulation network showed a new direction for identifying miRNA functions.

Key words: miRNA, drought stress, co-regulation network, *Physcomitrella patens*

Introduction

Plant growth and productivity are severely affected by drought stress. A large number of genes that respond to drought stress at the transcriptional level have been identified (*1, 2*). The products of these genes play important roles not only in protecting cells from drought stress (*3*), but also in regulating genes for signal transduction in the drought stress response (*4, 5*).

miRNAs are 21~24 nt non-coding small RNAs. In plant, mature miRNAs are generated from stem–loop regions of longer RNA precursors mainly by an endoribonuclease III-like enzyme, dicer like-1 (DCL1). The processed and methylated miRNA/miRNA* duplex is then exported to the cytosol. miRNAs that are incorporated into an argonaute protein containing RNA-induced silencing complex can affect the target gene expression (*6, 7*). miRNA-mediated regulations

# Equal contribution.
*Corresponding authors.
E-mail: wanp_cnu@yahoo.com.cn; jakechen@iupui.edu

rely on specific complementary base pairing of miRNAs to target mRNA (*8*), which can degrade the target transcript (*9*) and/or attenuate translation (*10*). Plant miRNAs bind to target mRNAs with near-perfect complementarity (*11*). Based on this observation, several groups have developed computational methods to predict miRNA targets in *Arabidopsis thaliana* and rice (*11-13*).

miRNAs serve critical roles in plant development, including adaptation to biotic and abiotic stresses (*6, 14, 15*). It is now considered that, developmental processes are primarily regulated by RNA regulatory networks (*16*), and miRNAs are key components of large gene regulatory networks (*6, 17*). However, our understanding of miRNA regulation network in plant is still poor.

Unlike *A. thaliana*, the model moss *Physcomitrella patens* occupies a key phylogenetic position with almost equal distance between green algae and flowering plants (*18, 19*). It is highly tolerant to drought (*20*) and serves as a valuable source for identifying the mechanism for plant drought resistance. However, currently little is known about the molecular traits underlying the drought resistance. Therefore, identifying gene regulatory processes involved in the drought resistance of *P. patens* may ultimately help design genetically modified crops with drought tolerance.

In this work, we adopted a gene-oriented strategy to identify drought stress-associated miRNAs (DsAmRs) in *P. patens* and discovered their new relationships with gene regulations, using bioinformatics techniques. We share our results through an online *P. patens* miRNA database named *ppt*-miRBase to promote community development of drought stress response.

Results and Discussion

Potential DsAmRs in *P. patens*

By implementing the strategy described in Materials and Methods, we identified 6 miRNA families (including 6 miRNAs) that target 8 drought down-regulated genes (**Table 1**) and 10 miRNA families (including 19 miRNAs) that target 11 drought up-regulated genes (**Table 2**).

Table 1 miRNAs and their down-regulated gene targets under drought stress

miRNA	Target protein ID	Protein annotation
miR1039-5p	105818	Oxyanion transporter-like protein
miR1054	194139	Glutamate synthase
miR1028a-3p	204491	Chloroplast ribosomal protein L5
miR1050	226242	Chloroplast translation elongation factor EF-G
miR2083-5p	109367	Rubisco small subunit
miR2083-5p	205107	Rubisco small subunit
miR2083-5p	210883	Rubisco small subunit
miR902a-5p	235275	Glycine decarboxylase

Note: Target protein IDs were retrieved from JGI's annotation (http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html).

Table 2 miRNAs and their up-regulated gene targets under drought stress

miRNA	Target protein ID	Protein annotation
miR1023a/b/c/d/e-5p	163734	No homology
miR1043-5p	176638	"SOUL" haem-binding protein
miR1035	177939	ATP-binding chaperone
miR1059	212537	Ubiquitin associated protein
miR1025	228672	Little protein 1
miR1221-3p	228678	WCOR413 cold acclimation protein
miR414	228695	Sucrose transporter
miR414	228697	Group 3 LEA
miR537a/b/c/d	228711	Organellar nucleoside-diphosphate-sugar epimerase
miR1031a/b	228736	Group 3 LEA
miR477f/g-5p	228741	19S proteasome subunit 9

Note: Target protein IDs were retrieved JGI's annotation (http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html).

Most down-regulated target genes encode chloroplast proteins, especially photosynthesis-associated proteins (*2*). For example, Ribulose-1,5-bisphosphate carboxylase/owygenase (Rubisco) small subunit is the target of miR2083-5p. Previous studies observed that during the drought stress, the activity of Rubisco decreased (*21*). However, the underlying molecular mechanism was unclear. Our results suggest that miR2083-5p might inhibit the translation of Rubisco small subunit and consequently decrease the activity of Rubisco.

miRNAs may also up-regulate translation during stress (*22-24*) and cell cycle (*25*), depending on how miRNA ribonucleoprotein complexes (miRNPs) are associated with other RNA-binding proteins (*26, 27*). Recently, a new study demonstrated that the miRNA target site may not be cleaved and instead negatively regulates miRNA activity through mimicry (*28*). In *Arabidopsis*, miR398 targets two Cu/Zn-superoxide dismutase (*CSD2* and *CSD1*) transcripts. During oxidative stress, miR398 is down-regulated and releases its suppression of *CSD1* and *CSD2* genes. The up-regulation of *CSD1* and *CSD2* genes also depends on expression levels of miR398 (*29*). Our results raise the possibility that miRNAs might also be activators of certain genes in *P. patens*.

miRNA co-regulation network in *P. patens*

Biological systems are driven by complex biomolecular interactions. Multiple miRNAs could target the same gene (*30-33*) and their target sites may partially overlap (*34*), suggesting that these miRNAs could combine together to regulate the expression of miRNA target genes (*35*). In this study, we identified 98 genes that were each targeted by more than one miRNA family.

General description

In the miRNA co-regulation network, each node represents a different miRNA family or the same miRNA family member from opposite arms of the miRNA stem–loop structure (for example, miR902-3p and miR902-5p). Two are connected, if they share at least one target gene (*36*). The *P. patens* miRNA co-regulation network (**Figure 1**) contained 82 nodes and 98 edges.

Targets of miR902-5p and miR414

The most intriguing feature of the *P. patens* miRNA co-regulation network was that miR902-5p and miR414 serve as two network hubs (nodes with a high number of edges connected), suggesting that they may play important roles in regulating drought resistance for *P. patens*.

In this study, both miR902-5p and miR414 were DsAmRs. miR902a-5p had 86 target genes. Gene Ontology (GO) term enrichment analysis (Table S1) indicated that targets of miR902a-5p are primarily involved in lipid metabolism. Target glycine decarboxylase (GDC) (Table 1) is down-regulated under drought stress. GDC is a key enzyme of photorespiration pathway (*37*), which could protect plants against drought stress (*38*).

miR414 had 145 predicted targets. They mainly involve in transportation of protein or sugar (Table S1). Sucrose transporter and group 3 late embryogenesis abundant (LEA) proteins are up-regulated under drought stress (Table 2). Sugar transporters have been recognized as pivotal targets for regulatory roles of distribution and partitioning of carbon resources in plants (*39*). Sucrose metabolism is extremely responsive to internal and external signals and consequently alters development and stress adjustment (*40*). LEA proteins are accumulated in plants during drought stress and have been considered to be associated with drought tolerance (*41*).

Here we must mention, due to moderately poor precursor hairpin structure, low Northern blot signal, the lack of identified targets, and the lack of conservation in genomes other than *Arabidopsis* and rice, the status of miR414 sequence as a miRNA has been questioned. Therefore, it is still suspicious whether miR414 is a potential DsAmR in *P. patens*.

Modules in *P. patens* miRNA co-regulation network

Many cellular processes are performed not by individual molecules, but by groups of functionally associated molecules, often referred to as functional modules. In the *P. patens* miRNA co-regulation network,

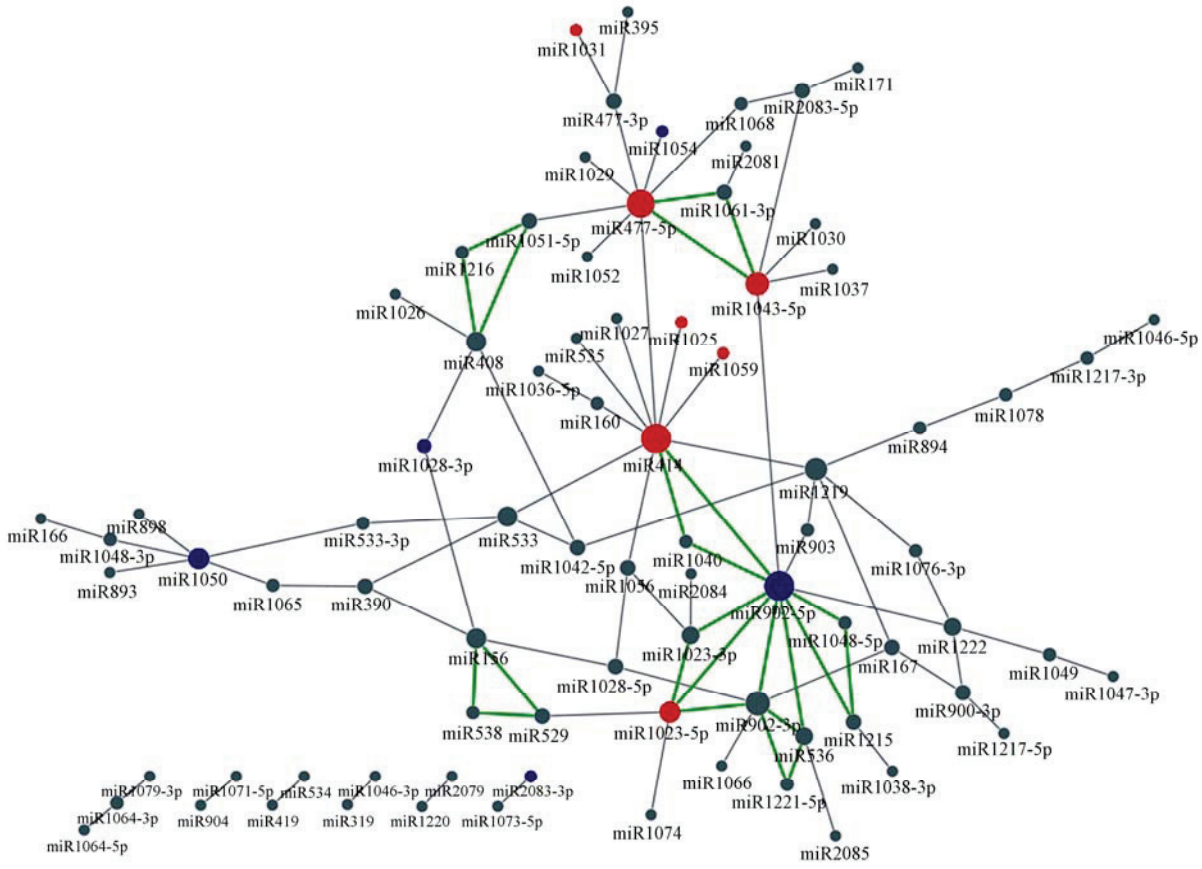

Figure 1 miRNA co-regulation network in *P. patens*. Each pair of connected nodes indicates that they share at least one common target. A red node represents that at least one of its targets is up-regulated under drought stress; whereas a blue node indicates that at least one of its targets is down-regulated under drought stress. The node size is proportional to the number of neighboring nodes. The group of nodes connected with green lines represents that they form a module.

modules appear as groups of densely interconnected nodes. Some modules overlap with each other and form a network of their own.

We identified six modules in the *P. patens* miRNA co-regulation network (**Figure 2**). Among them, three modules overlap with each other and form a sub-network (Figure 2A). Target genes of this module are mainly involved in cellular component organization. No enriched GO terms were found in target genes of module B (Figure 2B); targets of module C (Figure 2C) regulated translational initiation; and targets of module D (Figure 2D) played roles mainly in DNA replication.

ppt-miRBase database

To share our discoveries with the plant research community, we developed a *P. patens* microRNA database named *ppt*-miRBase, which could be accessed online from http://bioinfor.cnu.edu.cn/ppt_miRBase/index.php.

Here we show how to use the query capabilities supported by *ppt*-miRBase to explore the function of miRNAs included in the database (Figure S1). A user can start at the home page of *ppt*-miRNA, select a query miRNA, such as "miR902a-5p", from the pull-down list of miRNA and click the "Query" button (Figure S1A). On the query result page (Figure S1B), the user could browse 86 predicted target genes and 26 co-regulation miRNAs for miR902a-5p, which is listed as a DsAmR. By clicking the linked numbers "86" and "26", the user can also navigate into the detailed target page (Figure S1C) or the co-regulation miRNA page (Figure S1D). On the detailed target page (Figure S1C), the user can read that miR902a-5p had one drought-responsive target, protein 171132. According to GO annotation, protein 171132 involves in glycine metabolism.

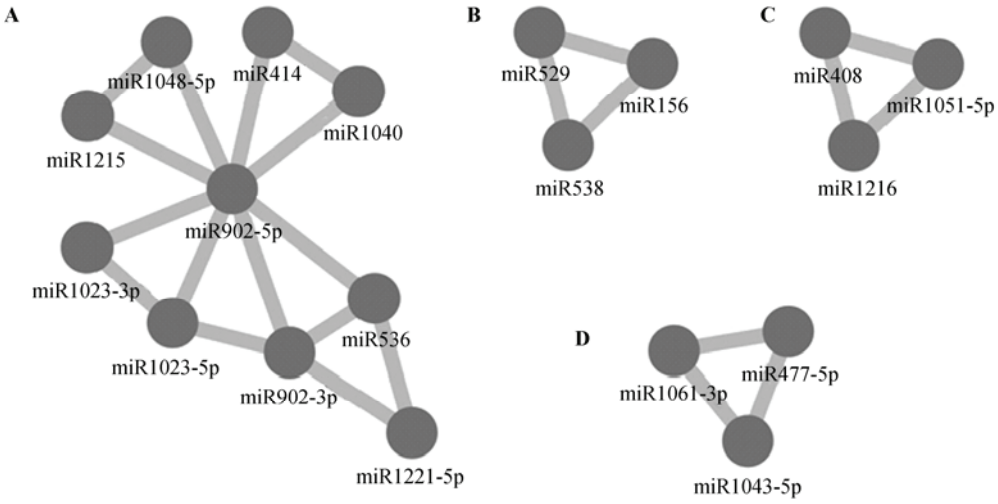

Figure 2 Modules in *P. patens* miRNA co-regulation network. Three modules share nodes to form a sub-network (**A**), their target genes mainly play roles in cellular component organization. No enriched GO terms were detected in target genes of module **B**. The functions of module **C** and module **D** target genes are translational initiation regulation and DNA replication, respectively.

Materials and Methods

Data preparation

P. patens EST sequences were retrieved from NCBI. Of all 381,669 EST sequences, 227,942 are 5' ESTs, and 153,727 are 3' ESTs. A total of 275 mature *P. patens* miRNA sequences were downloaded from miRBase (http://microrna.sanger.ac.uk/sequences/) (*42, 43*). *P. patens* genome sequence was retrieved from the *Physcomitrella* genome sequence assembly, version 1 (JGI Genome Portal; http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html). The genome encodes 35,938 proteins. The DNA sequences of drought stress responsive genes were downloaded from JGI according to Cuming's annotations (*2*). There are 82 down-regulated genes and 125 up-regulated genes (Tables S2 and S3).

Prediction of miRNA target

We modified the criteria that were developed for plant miRNA target prediction (*8, 44, 45*) to predict *P. patens* miRNAs targets. We mapped all 381,669 EST sequences onto *P. patens* genome using BLAT (*46*) and clustered EST sequences into groups according to their genome locations before assembling EST sequences of each group into contigs using Phrap program (*47*).

An initial pool of 275 *P. patens* miRNA candidate targets was created in these contigs using PatScan (*48*), with at most four unpaired nucleotides. Using Perl scripts, we aligned all 275 *P. patens* mature miRNA sequences against the candidate targets. This method leads to identification of all potential target sites of a single miRNA on the same target sequence, including false positive target sites. Consequently, a filter was applied to these targets, and the parameters were used according to Schwab *et al* (*8*): no mismatch at positions 10 and 11, no more than one mismatch at positions 2-12, and no more than two consecutive mismatches downstream of position 13. Finally, a mispair scoring system was applied to these filtered targets according to Wu *et al* (*49*): mismatches and single-nucleotide bulges were each scored as 1, and G:U pairs were each scored as 0.5. Candidate targets with a mispair score less than 3 were selected as putative miRNA targets.

Discovery of DsAmRs

We adapt a target-guided strategy to discover DsAmRs in *P. patens*. Since many drought stress responsive genes had been identified, our starting point is the drought stress-induced genes instead of related miRNAs.

Microarray analysis by Cuming *et al* revealed 207 differentially expressed genes in *P. patens* under drought stress (*2*). After predicting targets of all 275 *P. patens* miRNAs, we mapped all the predicted target genes onto the *P. patens* genome. If a predicted target gene is located within the same genomic region, in which a known drought stress-induced gene co-exists, we trace the predicted target gene back to its miRNA, and consider its targeting miRNA as a potential DsAmR (Figure S2).

Construction of miRNA co-regulation network

Two miRNAs are co-regulation miRNAs if they share at least one target gene (*36*). We constructed the miRNA co-regulation network using Cytoscape (*50*). In the network, nodes represent different miRNA families or the same miRNA family member from opposite arms of the miRNA stem–loop structure, and links indicate the sharing of at least one target gene by two nodes. Network modules in the network were identified using CFinder (*51*).

GO term enrichment analysis

To understand the function of an individual miRNA, we performed GO term enrichment analysis on target genes using topGO (*52*). We implemented a weighted counting method and Fisher's exact test to assess the significant of over-representation of GO terms. The enriched GO term ($P<10^{-5}$) represented the function of the miRNA (Table S1).

Construction of *ppt*-miRBase

The *ppt*-miRBase was implemented on a HP Core 32 Linux computer using MySQL as the database management system. Apache web server and PHP code were implemented to generate data-driven dynamic web contents.

Acknowledgements

We thank Dr. Yong Hu for his helpful discussion. We also thank the anonymous reviewers for their valuable comments on the manuscript. This work was supported by Beijing Municipal Education Commission Science and Technology Development Project (Grant No. KM200710028013) and PHR Project (Grant No. PHR201008078).

Authors' contributions

PW and JW conceived the idea. PW, JW, YZ, JX, JF, WZ, SX and GJ carried out the experiments. PW and JYC supervised the work and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors have declared that no competing interests exist.

References

1 Seki, M., *et al.* 2002. Monitoring the expression profiles of 7,000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* 31: 279-292.
2 Cuming, A.C., *et al.* 2007. Microarray analysis of transcriptional responses to abscisic acid and osmotic, salt, and drought stress in the moss, *Physcomitrella patens*. *New Phytol.* 176: 275-287.
3 Dure, L., *et al.* 1989. Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Mol. Biol.* 12: 475-486.
4 Uno, Y., *et al.* 2000. *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *Proc. Natl. Acad. Sci. USA* 97: 11632-11637.
5 Haake, V., *et al.* 2002. Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. *Plant Physiol.* 130: 639-648.
6 Jones-Rhoades, M.W., *et al.* 2006. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* 57: 19-53.
7 Voinnet, O. 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell* 136: 669-687.
8 Schwab, R., *et al.* 2005. Specific effects of microRNAs on the plant transcriptome. *Dev. Cell* 8: 517-527.
9 Llave, C., *et al.* 2002. Endogenous and silencing-associated small RNAs in plants. *Plant Cell* 14: 1605-1619.
10 Chen, X. 2004. A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science* 303: 2022-2025.
11 Rhoades, M.W., *et al.* 2002. Prediction of plant mi-

croRNA targets. *Cell* 110: 513-520.
12 Wang, X., *et al.* 2004. Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. *Genome Biol.* 5: R65.
13 Archak, S. and Nagaraju, J. 2007. Computational prediction of rice (*Oryza sativa*) miRNA targets. *Genomics Proteomics Bioinformatics* 5: 196-206.
14 Shukla, L.I., *et al.* 2008. The role of microRNAs and other endogenous small RNAs in plant stress responses. *Biochim. Biophys. Acta* 1779: 743-748.
15 Jian, X., *et al.* 2010. Identification of novel stress-regulated microRNAs from *Oryza sativa* L. *Genomics* 95: 47-55.
16 Mattick, J.S. 2007. A new paradigm for developmental biology. *J. Exp. Biol.* 210: 1526-1547.
17 Meister, G. 2007. miRNAs get an early start on translational silencing. *Cell* 131: 25-28.
18 Lang, D., *et al.* 2008. Exploring plant biodiversity: the *Physcomitrella* genome and beyond. *Trends Plant Sci.* 13: 542-549.
19 Rensing, S.A., *et al.* 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64-69.
20 Frank, W., *et al.* 2005. *Physcomitrella patens* is highly tolerant against drought, salt and osmotic stress. *Planta* 220: 384-394.
21 Parry, M.A., *et al.* 2002. Rubisco activity: effects of drought stress. *Ann. Bot.* 89: 833-839.
22 Leung, A.K. and Sharp, P.A. 2007. MicroRNAs: a safeguard against turmoil? *Cell* 130: 581-585.
23 Leung, A.K., *et al.* 2006. Quantitative analysis of Argonaute protein reveals microRNA-dependent localization to stress granules. *Proc. Natl. Acad. Sci. USA* 103: 18125-18130.
24 Bhattacharyya, S.N., *et al.* 2006. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* 125: 1111-1124.
25 Vasudevan, S., *et al.* 2007. Switching from repression to activation: microRNAs can up-regulate translation. *Science* 318: 1931-1934.
26 Vasudevan, S. and Steitz, J.A. 2007. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell* 128: 1105-1118.
27 Buchan J.R. and Parker R. 2007. Molecular biology. The two faces of miRNA. *Science* 318: 1877-1878.
28 Franco-Zorrilla, J.M., *et al.* 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39: 1033-1037.
29 Sunkar, R., *et al.* 2006. Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in *Arabidopsis* is mediated by downregulation of miR398 and important for oxidative stress tolerance. *Plant Cell* 18: 2051-2065.
30 Grün, D., *et al.* 2005. MicroRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS Comput. Biol.* 1: e13.
31 John, B., *et al.* 2004. Human microRNA targets. *PLoS Biol.* 2: e363.
32 Reinhart, B.J., *et al.* 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403: 901-906.
33 Ambros, V. 2003. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell* 113: 673-676.
34 Doench, J.G. and Sharp, P.A. 2004. Specificity of microRNA target selection in translational repression. *Genes Dev.* 18: 504-511.
35 Peter, M.E. 2010. Targeting of mRNAs by multiple miRNAs: the next step. *Oncogene* 29: 2161-2164.
36 Boross, G., *et al.* 2009. Human microRNAs co-silence in well-separated groups and have different predicted essentialities. *Bioinformatics* 25: 1063-1069.
37 Vauclare, P., *et al.* 1996. Regulation of the expression of the glycine decarboxylase complex during pea leaf development. *Plant Physiol.* 112: 1523-1530.
38 Wingler, A., *et al.* 2000. Photorespiration: metabolic pathways and their role in stress protection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355: 1517-1529.
39 Antony, E. and Borland, A.M. 2009. The role and regulation of sugar transporters in plants with Crassulacean acid metabolism. *Prog. Bot.* 70: 127-143.
40 Koch, K. 2004. Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr. Opin. Plant Biol.* 7: 235-246.
41 Tunnacliffe, A. and Wise, M. 2007. The continuing conundrum of the LEA proteins. *Naturwissenschaften* 94: 791-812.
42 Griffiths-Jones, S., *et al.* 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34: D140-144.
43 Griffiths-Jones, S., *et al.* 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36: D154-158.
44 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.
45 Zhao, T., *et al.* 2007. A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev.* 21: 1190-1203.
46 Kent, W.J. 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12: 656-664.
47 Gordon, D., *et al.* 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* 8: 195-202.
48 Dsouza, M., *et al.* 1997. Searching for patterns in genomic data. *Trends Genet.* 13: 497-498.
49 Wu, L., *et al.* 2009. Rice microRNA effector complexes and targets. *Plant Cell* 21: 3421-3435.
50 Shannon, P., *et al.* 2003. Cytoscape: a software environment for integrated models of biomolecular interaction

networks. *Genome Res.* 13: 2498-2504.

51 Adamcsek, B., *et al.* 2006. CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* 22: 1021-1023.

52 Alexa, A., *et al.* 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22: 1600-1607.

Supplementary Material

Figures S1 and S2; Tables S1-S3

DOI: 10.1016/S1672-0229(11)60006-5