# Computational Prediction of Rice (*Oryza sativa*) miRNA Targets

Sunil Archak and J. Nagaraju*

*Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, Hyderabad 500076, India.*

**Bioinformatic approaches have complemented experimental efforts to inventorize plant miRNA targets. We carried out global computational analysis of rice (*Oryza sativa*) transcriptome to generate a comprehensive list of putative miRNA targets. Our predictions (684 unique transcripts) showed that rice miRNAs mediate regulation of diverse functions including transcription (41%), catalysis (28%), binding (18%), and transporter activity (11%). Among the predicted targets, 61.7% hits were in coding regions and nearly 72% targets had a solitary miRNA hit. The study predicted more than 70 novel targets of 34 miRNAs putatively regulating functions like stress-response, catalysis, and binding. It was observed that more than half (55%) of the targets were conserved between *O. sativa indica* and *O. sativa japonica*. Members of 31 miRNA families were found to possess conserved targets between rice and at least one of other grass family members. About 44% of the unique targets were common between two dissimilar miRNA prediction algorithms. Such an extent of cross-species conservation and algorithmic consensus confers confidence in the list of rice miRNA targets predicted in this study.**

**Key words: miRNA, target prediction, conservation, consensus, rice**

## Introduction

MicroRNAs (miRNAs), a class of ∼22-nucleotide noncoding transcripts, have been shown to play a significant role in plant biology as negative regulators of gene expression (*1*, *2*). Understanding the functions of these miRNAs needs identification and characterization of their target sequences as well as the affected phenotype. Presently, miRNA targets are known in *Arabidopsis thaliana* (*3–7*), *Oryza sativa* (*8–10*), *Zea mays* (*2*), *Brassica napus* (*11*), and *Populus trichocarpa* (*12*). Experimentally, miRNA functions (not mere target sequences) are studied either in mutants or by generating knockdown lines, both of which are difficult and complicated; moreover, such phenotypes are pleiotropic and the systems are not optimized in plants except *Arabidopsis*. Furthermore, miRNAs and their targets do not exist as 1:1 pairs, and the pairs are not constant across tissues and cell types and along the developmental stages. Hence, computational prediction, incorporating as many factors as possible that influence miRNA–mRNA interaction, assists in generating a set of miRNA targets upon which wet experiments can be planned.

Ever since plant miRNAs and their targets were first identified and characterized, bioinformatic approaches to plant miRNA target prediction, exemplified by miR171:SCL, have been considered straightforward owing to virtually perfect base pairing between miRNA and target sequence (*6*, *13*, *14*). As a result of the stringent base pairing and the phylogenetic conservation employed in their prediction, both of which were considered absolutely essential, most of the plant miRNA target predictions have turned out be true and have been validated experimentally (*1*). As a consequence, it has been deduced that nearly 70% of the plant miRNA targets are transcription factors (TFs) and most plant miRNA targets are possibly all identified (*1*, *15*). On the flipside, however, it is likely that we may have overlooked targets with less stringent sequence match as well as those miRNA–target pairs that are species specific. Hence, it is essential to revisit the computational methodologies employed in plant miRNA target prediction, principally to assess the implications of stringent sequence match and to analyze the influence of cross-species conservation on the target repertoire.

The challenge, therefore, is to optimize target prediction algorithms to predict plant miRNA–target pairs with less extensive sequence match without devi-

***Corresponding author.**
**E-mail: jnagaraju@cdfd.org.in**

ating from the established principles of plant miRNA–target interaction. For instance, a pattern scan for 10 miRNAs of *Arabidopsis* detected 23 targets (*16*), whereas another algorithm, miRU, predicted as many as 203 potential targets (*17*). The downside of predicting non-canonical plant targets is the occurrence of false hits. Under such circumstances, there can be two *in silico* filters for target validation. The first filter is to ensure that the algorithm is not generating either lopsided targets or false hits by comparing the results of more than one target prediction algorithms. In plants, since the focus has been on stringent sequence match between miRNA and target, the need to develop and compare different algorithms was rarely perceived. The second filter is to ensure that targets are "conserved" across taxa to increase the confidence in the predicted targets.

Genetic and molecular approaches for the improvement of rice have helped establish rice as a model for plant functional genomic studies. We also know that how the growth and development of rice could be influenced by miRNA-mediated regulation (*8–10*). However, despite the availability of whole genome sequences of two subspecies (*indica* and *japonica*), and robust and abundant genomic resources from rice as well as a number of species belonging to the same Poaceae family, a complete repertoire of rice genes regulated by miRNA mediation is yet to be established. The objective of the present study was to generate a comprehensive list of rice miRNA targets by carrying out computational prediction, internalizing some of the above-mentioned key factors like minimum sequence match (*18*, *19*), conservation across taxa (*1*), and algorithmic consensus (*20*) to ascertain the influence of each of these components on the number and repertoire of rice miRNA targets. Our results support the prospect of predicting additional plant miRNA targets and we report more than 70 such novel miRNA–target pairs in rice that could have been ignored by an archetypal plant miRNA target prediction algorithm.

# Results and Discussion

## Validation of the computational algorithm

The miRanda scanning algorithm has been successfully used earlier (*21–23*). However, the suitability of this algorithm to detect miRNA targets in plants was never verified. It was, therefore, critical for the present analysis to ascertain how miRanda could be employed in plants and what kinds of modifications are necessary. Based on known principles of plant miRNA–target interactions, we arrived at a set of filters to minimize false hits (see Materials and Methods for details). The reliability of this approach was tested on *Arabidopsis*, which has computationally and experimentally well worked out miRNA target lists. Our analysis predicted 582 *Arabidopsis* miRNA targets including multiple splice forms of the target transcripts (Table S1). The hits included all the 66 known miRNA–target pairs of *Arabidopsis* reported by 7 different studies (Table S2). Besides, it is equally important to ascertain that a prediction algorithm does not generate redundant and false hits. Our analysis produced only 330 unique miRNA–target pairs, which is equivalent to 1.14% of the input sequences (2.8 targets/miRNA). These observations showed that the algorithm employed in the study ensured adequate stringency while additional targets to the existing ones were generated.

## Prediction and analysis of rice miRNA targets

Open access rice sequence data include nucleotide sequence entries, amino acid sequences, and unigenes. Since our work was confined to computational analysis, we wanted to avoid the input that might contain predicted mRNAs and false joining of expressed sequence tags (ESTs). Hence, we opted for the experimentally derived set of rice full-length mapped and annotated cDNA sequences. Out of 242 rice miRNA sequences available in the miRBase database (*24*), the miRanda-based methodology predicted 228 miRNA sequences to have targets among 32,127 full-length cDNA sequences explored. The hits of the rest miRNAs did not qualify the algorithm criteria or could not get through the filters, or the target sequences could be absent in the cDNA collection, since they do not represent the entire rice transcriptome. The predicted targets comprised of 684 unique cDNA sequences (2.13% of the total sequences scanned) with an average minimum energy of the duplex structure $\leq -30$ kcal/mol and an average homology $\geq 89\%$. A list of these miRNAs and comprehensive annotations of the corresponding targets including chromosomal locations, mRNA and protein lengths, source tissue, start–stop positions of the alignment, location of hit, hit sequence, and putative functions are given