Research News For U (RNFU) ISSN: 2250 –3668, Vol. 9, 2012
Available online http://www.doublehelixresearch.com/RNFU

Computational identification of microRNA in plants

C. Anuradha [1]* and Parameswari B[2]

[1]National Research Centre for Banana, Trichy, Tamil Nadu, India.
[2]Sugarcane Breeding Institute Regional Centre, Karnal- 132001.

MicroRNA

miRNAs are about 20–24 nucleotide (nt), single-stranded RNAs processed from typical stem loop precursors by the Dicer-like (DCl) family of enzymes in plants. miRNAs are known to play important regulatory roles in plants by targeting mRNAs for cleavage or translational repression. miRNAs and their targets have been found to affect diverse processes, including organ development such as leaf morphogenesis, floral organ identity, and root development. Plant miRNAs also function in feedback regulation in small RNA pathway and in the biogenesis of certain class of siRNAs, for example trans-acting siRNAs. Moreover, they are involved in various stress responses.

Origin and evolution of miRNA genes

The basic principles of miRNA biogenesis remains almost similar between plant and animal systems, there are also ample differences between plant and animal miRNA characteristics and biogenesis. Plant miRNAs are mostly generated from noncoding transcriptional units in contrast with some of the animal miRNAs which are processed from introns and protein coding genetic sequences. Compared to animals, plants have a more complex small RNA population in their transcriptomes. Due to the abundance of plant-specific RNA Polymerase IV and RNA Polymerase V-dependent siRNA and *trans*-acting siRNAs, plant miRNAs are represented in the pool of small RNAs. By contrast, animal small RNA populations are generally filled with miRNAs in their transcriptomes. Plant miRNAs have a unique 5′ end which differs from animal miRNA 5′ end sequences. To repress translation, plant miRNAs tend to bind to the protein-coding region of target mRNAs, but animal miRNAs bind to the 3′ untranslated region (3′ UTR) of their target mRNA transcripts.

Regulation of miRNA genes

Recently, several studies have shown that there are some conserved sequence motifs in the upstream regions of miRNA genes in *C. elegans* and *A. thaliana* and some factors activate miRNA gene expression. By aligning sequences upstream and downstream of orthologous *C. elegans* miRNA foldbacks, a highly conserved sequence motif, with consensus CTCCGCCC, exists in almost all *C. elegans* miRNA genes. It is located about 200 bp upstream of miRNA precursors. A TATA box-like sequence appeared to be conserved in a majority of *Arabidopsis* miRNA genes, centered at consensus position−29 from the start site. The promoters of miRNAs and transcription factors for

miRNA expression should be identified. The alignment of more miRNA sequences including longer upstream and downstream sequences to identify more common conserved motifs could be done using more computational approaches.

Identification of miRNA

miRNA identification largely relies on two main reverse genetics strategies: (1) computer-based (bioinformatics) and (2) experimental approaches. A third identification approach, forward genetics, is rarely used in miRNA discovery. miRNA identification using bioinformatics tools is one of the most widely used methods, contributing considerably to the prediction of new miRNAs in both animal and plant systems. This is largely due to the low cost, high efficiency, fast and comprehensive methodology of bioinformatics. The main theory behind this approach is finding homologous sequences of known miRNAs both within a single genome and across genomes of related organisms. Sequence and structure homologies are used for computer-based predictions of miRNAs. Computational strategies provide a valuable and efficient manner to predict miRNA genes and their targets. The software-based approach is applied to animals, human, fungi, and plants. For example Zhang et al. identified 338 new possible miRNAs in 60 different plant species and Adai et al. have predicted 43 new miRNAs in *Arabidopsis*.

In contrast, cloning and sequencing of small RNA libraries represents an experimental approach to identify and characterize miRNAs. However, in contrast with bioinformatics, such approaches for miRNA identification also have limitations. First, most of the miRNAs are tissue and time specific and generally their expression level is low. In addition, they mostly express in response to specific environmental stimuli. They also coexist with their cleaved and degraded target mRNAs, hence cloning small RNAs (miRNA and siRNA) is difficult, whereas computational approaches are effective because of no need for cloning. Since forward genetics, or the genetic screening approach, is time consuming, expensive, and less efficient, it is rarely used for plant miRNA identification. Next generation massive sequencing techniques are also applied to identify new miRNAs in plants. Here we summarize the computational approaches for identifying plant miRNAs.

Computational Approaches

The principles of computational approaches are based on the major characteristic features of miRNAs: hairpin-shaped stem loop secondary structure, high evolutionary conservation from species to species, and high minimal folding free energy index. Many computational approaches have been developed by different laboratories. All these approaches have been successfully used to identify miRNA genes in various plant and animal species. Currently, 4034 miRNAs which belong to 45 different species have been deposited in the miRNA database. A majority of these miRNAs were identified by using computational approaches and subsequently and/or concurrently verified directly or indirectly by different experimental approaches, including direct cloning technologies, Northern blotting, PCR, and/or 5'RACE. The computational approaches can be classified into five

major categories: homology search-based, gene search, neighbor stem loop search, algorithms based on comparative genomics and phylogenetic shadowing-based.

1. Homology Search-Based

Homology search-based approach is a method for identifying miRNA genes by searching nucleotide databases using BLAST program. Since the beginning of abundant miRNA identification, it was well recognized that miRNAs are evolutionarily conserved in both plants and animals. This suggests a powerful approach to identify miRNA orthologues and/or homologues by searching publicly available DNA databases against known miRNAs that are experimentally identified in model species. Homology searches can be classified as genome-based search or EST-based search. EST analysis has proven to be an economically feasible alternative for gene discovery in species lacking a draft genome sequence; many important genes have been found through EST analysis. As of June 2012, GenBank (National Center for Biotechnology Information, URL http://www.ncbi.nlm.nih.gov/) contained more than 73,327,421 entries in its EST database. Thus, EST-based homology search has become a powerful approach to identify miRNA genes in various species, especially in species whose genome sequences are not available.

2. Gene-finding approach

Gene-finding approaches are designed for predicting animal miRNA genes, which do not onsider miRNA conservation. These approaches complement homology search approaches. Gene-finding approaches do not depend on homology or proximity to previously known miRNAs, and can be used for an entire genome search. Gene-finding approaches first need to identify conserved genomic regions and then put these conserved regions into a window that can hold about 110-nt using a specific computer program, and the window is folded with a secondary structure-predication program, such as Mfold or RNAfold, then scores these hairpin-shaped stem loops for potential miRNA candidates.

3. Neighbor stem loop search

The neighbor stem loop search approach is based on miRNA clusters and secondary structures. Many miRNAs present as tandem arrays within clusters like operons. In most cases, only two or three miRNAs cluster together. However, some larger clusters have also been observed, such as the conserved miRNA cluster *miR-17* found in mammals, which has six members. This approach can only be used in animal miRNA identification due to the fact that very few plant miRNA clusters have been observed.

4. Algorithms based on comparative genomics

Comparative genomics uses sequence comparisons between species to identify different genes and regulatory elements. Recently, it has become a powerful approach to predict miRNA genes in animals and plants through the comparison of two known genomes. Although these

methods are widely used in computational identification of miRNA genes, their application is somewhat limited by the lack of genome sequences for a majority of animal and plant species.

5. Phylogenetic shadowing-based approach

Phylogenetic shadowing approach is a modified comparative genomics approach for multiple species. Phylogenetic footprinting (cross-species sequence comparison) is one approach to identify functional genetic elements. However, the sensitivity of this method decreases with increasing phylogenetic distance. Additionally, species-specific elements may be missed by this approach, especially for short sequences.

Computational identification of microRNA targets

Although thousands of miRNAs have been identified in animals, plants and viruses, the targets for a majority of these miRNAs have not been identified due to the fact that large-scale experimental detection of targets is not currently available. The known miRNA targets have a high degree of complementarity to the miRNAs, especially for plant miRNAs. This allows the prediction of miRNA targets by computational approaches. Several studies indicate that computational approaches play important role not only in the discovery of miRNA genes but also in the identification of miRNA targeted genes. Currently, several computational approaches have been used to successfully identify potential miRNA targets in mRNA sequences or to select potential targets for experimental validation. A majority of these computational approaches are based on three major characteristic properties: (1) miRNAs are perfectly or near-perfectly complementary to their target mRNAs with no bubbles or gaps at the complementary sites in plants. Although the complementarity between miRNAs and their targets were not perfect in animals, there still are some seed regions between them. These complementary sites are usually located at the 3′ untranslated region (3′ UTR) of the targeted genes in animals and any location along the targeted mRNAs in plants (2) The RNA–RNA duplex has a higher negative folding free energy. (3) Mature miRNAs, binding sites of mRNA to miRNA, and miRNA:mRNA duplex are highly conserved from species to species, particularly within the same kingdom. Thus, almost all computational approaches for plants first detect potential binding sites with a high degree of complementarity between the miRNAs and mRNAs (usually less than three mismatches and no gaps between miRNAs and mRNAs in perfect complementarity), and then remove these sequences that do not appear to be conserved in multiple species. However, these approaches can only be used to predict conserved targets or closely related species. For distant species and species-specific miRNAs, the only criterion is complementarity which will increase the chance of false positives. Thus, the computationally predicted targets need to be validated by experimental methods such as Northern blotting to detect the expression level of targeted mRNAs. At present, several web-based or non-web-based computer software programs are publicly available and commonly used to predict miRNA targets, such as TargetScan, TargetScanS, miRanda, MovingTargets, RNAhybrid, PicTar and DIAN-AmicroT. A majority of predicted and validated miRNA targets are transcription factors, which play important roles in animal and plant development, timing, and stem cell differentiation.

There is only one complementary site between miRNAs and their targets in plants, while there are several complementary sites in animals.

Conclusion and future perspectives

Computational approaches play an important role not only in the discovery of miRNA genes but also in the identification of miRNA targets. A majority of miRNA genes and their targets have been first identified by computational approaches based on their major characteristic properties and then validated by different experimental approaches. Several web-based or non web-based computer software programs are publicly available to predict miRNAs and their targets. Although computational approaches have allowed for great progress in identifying miRNA genes, lots of miRNAs, especially species-specific miRNAs, await discovery. A majority of computational programs predicting miRNAs rely on evolutionary conservation. Thus, these types of screening miss species-specific and rapidly evolved miRNAs. Current computational approaches need to be modified or new computational approaches need to be developed for identifying undiscovered miRNAs. One strategy is to use the conserved motifs described above to develop new computational approaches for identifying new miRNA genes. In addition, a majority of miRNAs have a minimal folding free energy index higher than 0.85, which is much higher than other coding and non-coding RNAs. We may combine this newly developed miRNA criteria to predict new miRNAs.

Authors:
C. Anuradha,
Scientist,
NRC for Banana,
Trichy- 620102, Tamil Nadu.
*(Corresponding author)

Parameswari, B.
Scientist,
Sugarcane Breeding Institute Regional Centre,
Karnal- 132001, Haryana.