Genomics & Informatics Vol. 9(1) 39-43, March 2011

MiRPI: Portable Software to Identify Conserved miRNAs, Targets and to Calculate Precursor Statistics

Dhandapani Vignesh[1¶], Paul Parameswari[1¶], Su Bin Im[1], Hae Jin Kim[2] and Yong Pyo Lim[1*]

[1]Molecular Genetics and Genomics Laboratory, College of Agriculture and Life sciences, Chungnam National University, Daejeon 305-764, Korea, [2]Ensoltek Co., LTD, Daejeon 305-510, Korea

Abstract

MicroRNAs (miRNAs) are recently discovered small RNA molecules usually resulting in translational repression and gene silencing. Despite the fact that specific cloning of small RNA's is a method in practice, computational identification of miRNA's has been a major focus recent days, since is a rapid process following AB initio and sequence alignment methods. Here we developed new software called MiRPI that aims to identify the highly conserved miRNAs without any mismatches from given fasta formatted gene sequences by using non-repeated miRNA dataset of the user's interest. The new window embedded with the software is used to identify the targets for inputted mature miRNAs in the mRNA sequences. Also MiRPI is designed to measure the precursor miRNA statistics, majorly focusing the Adjusted Minimum Folding free Energy (AMFE) and Minimum Folding free Energy Index (MFEI), the most important parameters in miRNA confirmation. MiRPI is developed by PERL (Practical Extraction and Report Language) and Tk (Tool kit widgets) scripting languages. It is user friendly, portable offline software that works in all windows OS, sized to 3 MB.

Availability: http://code.google.com/p/mirpi

Keywords: miRNA, portable software, MiRPI, miRNA targets

Introduction

MicroRNAs (miRNAs) are newly discovered class of non-coding small RNAs that are generally 18～24 nucleotides in length, that regulate gene and protein expression in plants and animals (Bartel, 2004; 2009). They were first described in the year 1993 by Lee RC and his colleagues, and the term miRNA was coined in later 2001 (Ruvkun, 2001). MiRNA's have so far been identified mostly by specific cloning of small RNA molecules, complemented by computational methods. MiRNA are partially complementary to one or more messenger RNA (mRNA) molecules, and function to downregulate the gene expression in a variety of manners, including translational repression, mRNA cleavage, and deadenylation. MiRNA genes reside in regions of the genome as distinct transcriptional units, in clusters of polycistronic units - carrying the information of several miRNAs (Lagos-Quintana, 2001). Studies suggest that approximately half of the known miRNA inhabit in non-protein coding RNAs (intron and extron) or within the intron of protein coding genes (Rodriguez, 2004). RNA polymerase II transcribes miRNA genes, producing long primary transcripts (pri-miRNAs) (Kim, 2005). However, the process to yield mature miRNAs involves two steps involving RNase-III enzymes and companion double-stranded RNA-binding domain (dsRBD) proteins. Researches suggest that about one-third of the human genes are possibly regulated by miRNAs (Lim, 2003). This occurs when mature miRNAs couples with a multiple-protein nuclease complex called the RNA-induced silencing complex (RISC). Once incorporated into a RISC, the miRNA is positioned to regulate the target genes by degradation of the mRNA through direct cleavage or by inhibiting protein synthesis. MiRNAs can bind to sequences on the target mRNAs by exact or near-exact complementary base pairing and thereby direct cleavage and destruction of the mRNA (Rhoades, 2002) similar to the mechanism employed in RNA interference (RNAi), the cleavage of a single phosphodiester bond on the target mRNA occurs between bases 10 and 11 (Elbashir *et al*., 2001). In contrast, nearly all animal miRNAs studied so far are usually not exactly complementary to their mRNA targets, and seem to inhibit protein synthesis while retaining the stability of the mRNA target (Ambros, 2004). It has been suggested that transcripts may be regulated by multiple miRNAs and an individual miRNA may target numerous transcripts.

MiRNA identification is an essential requirement for understanding the mechanisms of post-transcriptional regulation. Consequently the prediction of the potential targets also becomes an essential parameter in the un-

[¶]Authors are equally contributed.
*Corresponding author: E-mail yplim@cnu.ac.kr
Tel +82-42-8215739, Fax +82-42-8218847
Accepted 10 March 2011

derstanding cleavage and destruction mechanism (Rhoades *et al*., 2002). In this study we developed new software MiRPI to identify the highly conserved miRNAs without any mismatches, their targets and also to calculate the statistics of precursor miRNAs. Statistical analyses of predicted precursor miRNA, like MFEI, AMFE are one of the most important factors to confirm the miRNA through precursor secondary structure analysis. MiRPI is portable software that means a computer program which is able to run independently without the installation of any files for its use. Functions and importance of the software MiRPI is described further.

Methods

Interface

MiRPI has a user friendly interface with easy workflow for biologist. It' s structured into three windows (Fig. 1), the main window was designed for identification of miRNAs with two upload buttons for gene sequences file and non-repeated miRNAs file. In the second window, statistics of precursor can be measured by uploading the precursor sequence file and Minimum Folding free Energy (MFE) values file. Additionally, it's also possible to identify the targets by inputting mRNA sequence file and identified miRNAs in separate file in the third window. The detailed workflow of MiRPI is explained in Fig. 2. The current version of MiRPI was developed by using Perl/Tk.

Input

MiRNA identification window has options to upload two sequence files, one for gene sequences and another for miRNAs file, using the submitted mature miRNAs the software finds miRNA gene sequence from the inputted gene sequences. MFE values must be calculated by external software and should be organized into separated lines, for easy arrangement copy all the MFE values in one column in excel sheet and save it as text document. Equal number of MFE values and sequences must be given in the input files since MiRPI assumes the first MFE value of the input file as the first sequence's value and uses for further analysis. Target identification window also requires two input files, one with mRNA sequences as target and the other with identified mature miRNAs. Gene, mRNA (for targets), and miRNA sequence files must contain the symbols "＞", "/" at the end of each file in separate lines. This helps to read the sequences quickly and proceed to the next process. MiRPI only accepts fasta formatted sequence file, also the sequence description line must contain the accession number in between first thirteen

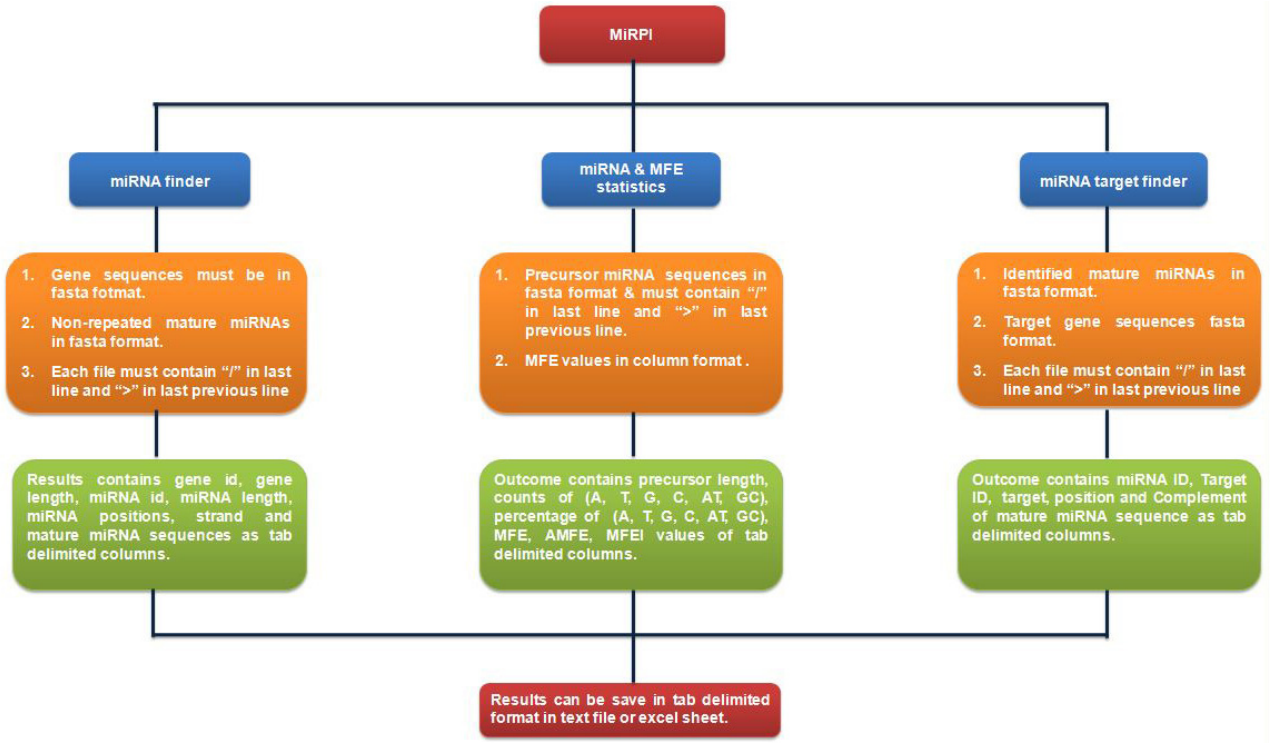

Fig. 1. Screenshot represents the structure of the software and results which analyzed.

letters. After the first and second space in the fasta description line will recognize as an organisms full name. In the case of mRNA sequence file after the third space will read as target information. For example the input sequence file must be given as default GenBank, NCBI fasta formats as follows >gi|168828718|gb|EU482897.1| *Pinus pinaster* R2R3-Myb14 transcription factor (Myb14) gene.

Results and Discussion

MiRPI identifies the highly conserved mature miRNA se-

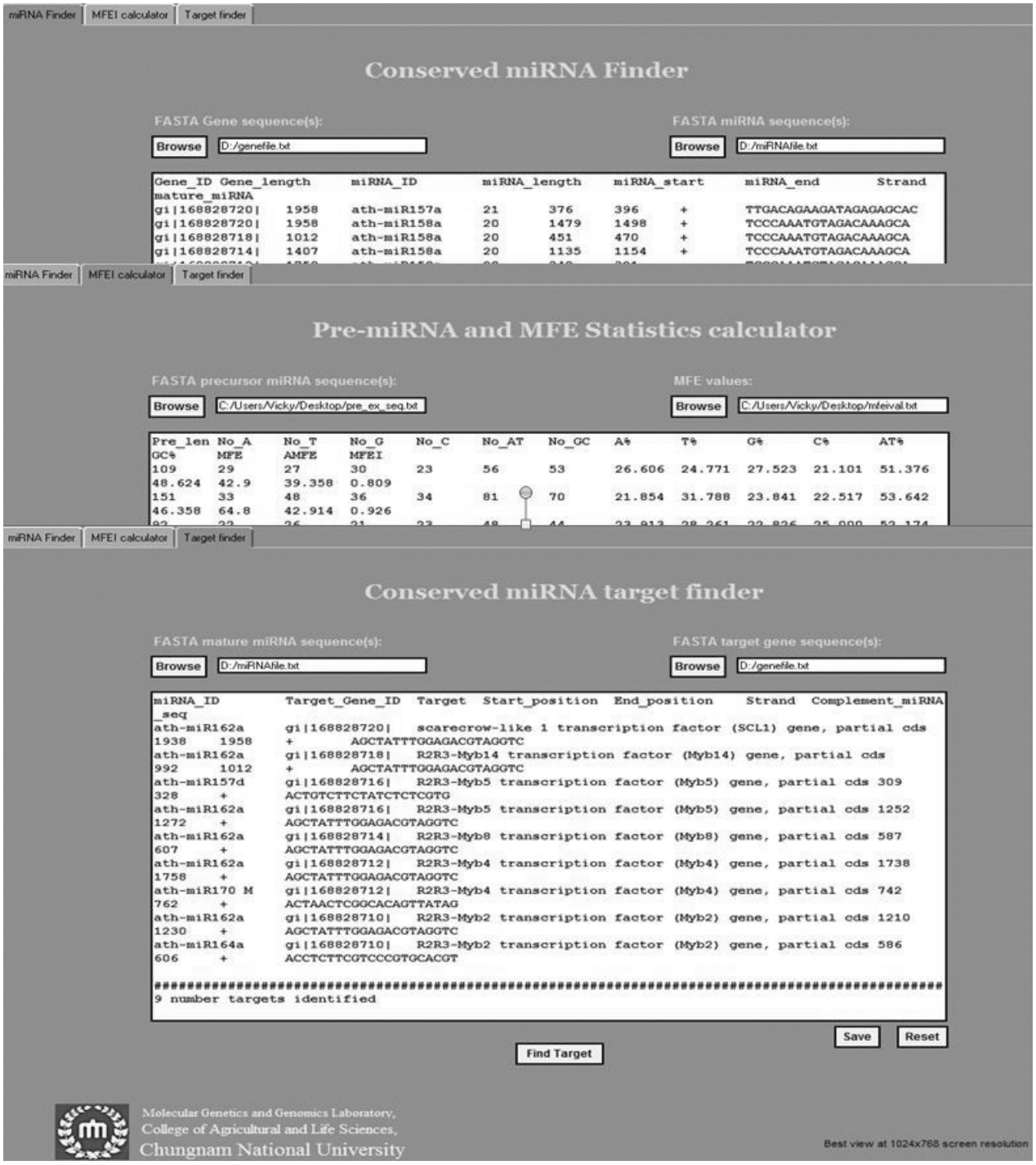

Fig. 2. The workflow explains the process of MiRPI software.

quences of the user's interest without any mismatches against the given input gene sequence file with highly trustable data. We utilized MiRPI for internal research to test the sample data which were downloaded from public databases such as National Centre for Biotechnology Information (NCBI) and plant miRNA database (PMRD) (Zhang *et al*., 2010). The output retrieves required information and frames into columns as follows, the gene id, their length, matching miRNA id, start and end position, strand, and conserved mature miRNA sequence. The total counts of identified miRNA were displayed at the end.

Calculation of precursor miRNA statistics for larger data set manually is a time consuming process and hence we included the MFE statistics calculator in the tool. Outcomes of precursor statistics calculator consist the precursor length, counts of each base pair (A, T, G, and C), percentage of each four bases, percentage of AT and GC, given MFE value, AMFE, and MFEI in tab delimitation format. MFEI and AMFE calculation formulas (Fig. 3) were utilized as described by Zhang and their team (2009). This mainly helps in the confirmation of precursor miRNA in secondary structure analysis. Targets were checked with plant mRNA sequences which were downloaded from NCBI GenBank (www.ncbi.nlm.nih.gov/genbank) database and miRNAs from PMR database (Zhang *et al*., 2010). Identified target results were outlined into seven columns, and organized as follows miRNA id, target gene id, target, start and end position of mature miRNA, strand, and compliment mature miRNA sequences which binds to identified miRNA. The results were compared with the published data and verified.

Efficiency of the tool was evaluated with standard example sets of data which were downloaded from the NCBI database and plant miRNA database (Zhang *et al*., 2010). Overlapping miRNAs were removed by in home developed scripts and resulted 1191 non-repeated mature miRNAs. All these miRNAs are searched against all plants gene sequences by MiRPI software and retrieved 2163 miRNAs, they are accessible at http://168.188.15.78/db/all_miRNA. Outcome of each analysis was framed into simple structure for easy understand and documentation, which are separated by tab delimitation into a column (Fig. 1). All the results can be saved in a text document or copied to excel sheet for better understanding.

$$\text{AMFE} = \text{MFE} \div (\text{Length of a pre-miRNA}) * 100$$

$$\text{MFEI} = (\text{MFE} \div (\text{Length of a pre-miRNA}) * 100) \div ((\text{G+C})\ \%)$$

Fig. 3. Formulas used for calculating Adjusted Minimum Folding free Energy (AMFE) and Minimum Folding free Energy Index (MFEI).

Conclusion

MIRPI is user friendly, fast and accurate software developed by using the PERL and Tk. miRNAs and their targets were identified for more than 100,000 plant sequences against 1,000 plant miRNAs using MiRPI. It is open source portable software that can be used in all windows Operating systems. MiRPI successfully identifies the miRNAs from the given gene sequences, targets from mRNA sequence and calculates the statistics for submitted MFE and precursor sequences.

Acknowledgments

This research was carried out under "Human Resource Development Center for Economic Region Leading Industry" project (Project No. 2010-1228), supported by the Ministry of Education, Science & Technology (MEST) and the National Research Foundation of Korea(NRF). Also the research was partially supported by Chinese Cabbage Molecular Marker Research Center (CMRC), grants for the Technology Development Program for Agriculture and Forestry (Grant No. 607003-05), Ministry of Agriculture, Forestry and Fisheries, Republic of Korea.

References

Ambros, V. (2004). The functions of animal microRNAs. *Nature* 431, 350-355.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215-233.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21 and 22 nt RNAs. *Genes & Development,* 15, 188-200.

Kim, V.N. (2005). MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* 6, 376-385.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *Caenorhabditis elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843-854.

Lim, L.P., Lau, N.C., and Weinstein, E.G. (2003). The microRNAs of *Caenorhabditis elegans*. *Genes & Development* 17, 991-1008.

Lim, L.P., Matthew, W., Rhoades, M.W., Reinhart, B.J., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant MicroRNA targets. *Cell* 110, 513-520.

Quintana, M.L., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science* 294, 853-858.

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.
Rodriguez, A., Jones, S.G., Ashurt, J.L., and Bradley, A. (2004). Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Research* 14, 1902-1910.
Ruvkun, G. (2001). Molecular Biology: Glimpses of a Tiny RNA World. *Science* 294, 797-799.
Zhang, B., Stellwag, E.J., and Pan, X. (2009). Large-scale genome analysis reveals unique features of miRNAs. *Gene*. 443, 100-109.
Zhang, Z., Yu, J., Li, D., Zhang, Z., Liu, F., Zhou, X., Wang, T., Ling, Yi., and Su, Z. (2010). PMRD: plant microRNA database. *Nucl. Acids Res.* 38(suppl 1), D806-813.