

# **COMPARATIVE EVALUATION OF PLANT *miRNA* DETECTION TOOLS**

**KHUSHNUMA TANVEER  
ROLL NO.: 16MBI002**



**Department of Computer Science  
Faculty of Natural Science  
Jamia Millia Islamia  
May, 2018**

# **COMPARATIVE EVALUATION OF PLANT *miRNA* DETECTION TOOLS**

**KHUSHNUMA TANVEER**  
**Enrollment No.: 16-0011**

Submitted

in partial fulfillment of the requirements of the award of the degree of

**Master of Science (Bioinformatics)**

**Under the Supervision of**

**Dr. Shailesh Kumar**  
**(*Staff Scientist*)**



**DEPARTMENT OF COMPUTER SCIENCE**  
**Faculty of Natural Sciences**  
**Jamia Millia Islamia**  
**MAY, 2018**

## **Declaration**

I, Khushnuma Tanveer., student of M.Sc.(Bioinformatics) hereby declare that the project report entitled "**Comparative evaluation of plant *miRNA* detection tools.**" which is submitted by me to the Department of Computer Science, Jamia Millia Islamia, New Delhi, in partial fulfilment of the requirement of the degree of **M.Sc.(Bioinformatics)**, have not been submitted in part or full to any other university or institute for the award of any degree or diploma.

(Khushnuma Tanveer)  
Roll No.:16MBI002  
Enrollment No.:16-0011

# Certificate

On the basis of declaration made by the student **Khushnuma Tanveer**, I hereby certify that the project report entitled "**Comparative evaluation of plant miRNA detection tools.**" submitted by **Khushnuma Tanveer** to the Department of Computer Science, Jamia Millia Islamia, New Delhi, for the partial fulfilment of the requirements of the degree of **M.Sc.(Bioinformatics)**, is carried out by her under my guidance and supervision. The report has reached the requisite standards for submission.

**Dr. Shailesh Kumar**  
**Staff Scientist**  
**(Supervisor)**

## **Acknowledgements**

First of all, I would like to thank my supervisor Dr. Shailesh Kumar, Staff Scientist, NIPGR for his guidance and support throughout my time as a trainee. Thanks for giving me the opportunity to work in the interesting bioinformatics field. His systematic and rhythmic guidance is the main cause for the successful completion of this work. He had always provided an alternate to the problem I faced throughout my project. I am also thankful to Prof. Khurram Mustafa, Internal Supervisor for his time in guiding and evaluating this report before submission.

I would like to thank Head of the Department, Dr. SAM Rizvi and our Course coordinator, Dr. Syed Zeeshan Hussain. Teaching Faculty of the department had always helped, guided, and educated us. I thank Dr. Khalid Raza who has always guided me in the flow of literature and writing. Other office staff members were at all times there on the need. They had always provided me working computer systems. In addition,

I would like to express humble gratitude to one of our seniors Mr. Ajeet Singh, who is working as a PhD scholar at NIPGR, New Delhi. He has always devoted and guided me in the project as my elder brother. Many thanks go to colleagues and friends, especially Shahre Alam, Nikita Sharma, Shafaque Zehra and the group for the comfortable breaks during work and for keeping the coffee flowing.

I would like to thank my family members for their constant support, backup, and encouragements, and for putting up with me over the past years and for bringing joy and spirit into my life. At last but not the least, I thank Almighty Allah who has given this beautiful life, guided us to follow the path of truth and gave enough courage to complete this dissertation.

*This thesis is dedicated to my parents.*

**Khushnuma Tanveer**

# **Contents**

**Declaration**

**Certificate**

**Acknowledgements**

**List of Figures**

**List of Tables**

**Abstract**

<b>1. Introduction</b>	<b>1-7</b>
1.1 Background	2
1.2 Motivation of work	5
1.3 Objective	6
1.4 Organization of the chapters	6
<b>2. Literature Review</b>	<b>8-12</b>
<b>3. Materials &amp; Methods</b>	<b>13-26</b>
3.1 Datasets	12
3.2 miRNA Detection Tools	13
3.2.1 C-mii (Computational miRNA Identification)	13
3.2.2 miRPlant (An Integrated PLant miRNA detection tool)	14
3.2.3 MIRINHO (An efficient pre-miRNA detector)	17
3.2.4 MIRENA (Software to predict miRNAs with high accuracy)	18
3.2.5 miRDeep-P	19
3.2.6 miRExpress	21
3.2.7 miReader	23
3.3 Performance metrics	23
<b>4. Results</b>	<b>27-31</b>
<b>5. Conclusion and Future Work</b>	<b>32-33</b>
<b>References</b>	<b>34-36</b>

# List of Figures

**Fig. 1.1** MicroRNA maturation & function

**Fig. 1.2** The complexity and volume of data generated by NGS technologies and the need for bioinformatics tools for miRNAs detection (Akhtar et al., 2016)

**Fig. 1.3** Diagrammatic view of flow of miRNA analysis and available bioinformatics tools. List of available tool in each category are also presented here

**Fig. 3.1** miRNA detection pipeline of C-mii

**Fig. 3.2** Work Flow of miRPlant

**Fig. 3.3** Snapshot of parameter settings in miRPlant

**Fig 3.4** (a) Genome sequence and mapping file (mapped file of small RNA reads against genome sequence using Bowtie) are given to PROCINE. (b) Region with reads are obtained using SAM file from the genome sequence. (c) Genome has screened to predict potential miRNAs. (d) Assessing the potential miRNAs by using Smith-waterman algorithm and Nearest neighbour model.

**Fig 3.5** Working Pipeline of Mirinho

**Fig 3.6** Schema of MIRENA algorithm. MIRENA can be utilized by four ways. Four different work paths are depicted as various colors (black, red, brown and blue). The square boxes describe input and output data, while arrows show flow of information

**Fig 3.7** Workflow of miRDeep-P

**Fig 3.8** Systematic work-flow of miRExpress including miRNA expression profiling. It accepts NGS data and constructs miRNA expression profiles using mapped alignment of known miRNAs.

**Fig. 3.9** Systematic workflow of miRExpress tool consisting of three steps: i) NGS data preprocessing, ii) read alignment against known miRNAs sequences, and iii) miRNA expression profiling

**Fig. 3.10** Work flow of miReader algorithm

**Fig. 4.1** Sensitivity, Specificity and Precision of each tools

**Fig. 4.2** Performance comparison of various tools

# List of Tables

**Table 2.1** Characterization of Common Features for Selected Tools

**Table 3.1** TAIR10 cDNAs mature miRNA sequences of *Arabidopsis thaliana*

**Table 4.1** C-mii tool

**Table 4.2** MIRINHO tool

**Table 4.3** miRPlant tool

**Table 4.4** miReader Software Tool

**Table 4.5** miRExpress Software Tool

**Table 4.6** MiReNA Software Tool

**Table 4.7** miRDeep-P Software Tool

**Table 4.8** mireap Software Tool

# **Abstract**

MicroRNAs are short non-coding RNA (~22 nt in length) which are highly conserved and found in both plants and animals. They are indicated from long transcripts originate in plants, viruses, and single-celled eukaryotes. MicroRNAs have been found to be involved in various pathways and plays important role in Biogenesis and disease diagnosis including cancers and neurological disorders. From the last two decades, several computational techniques have been evolved for the detection of miRNAs, its targets and functions from sequencing and deep sequencing data. However, accurate finding these is still challenging. Among the plethora of available software tools for miRNA detections, it is very difficult of the biologist to use the best performing and reliable tool for detection of novel miRNAs. The reason is that these software tools are equipped with different algorithms and their performance greatly varies.

In this study, we have evaluated the performance of popularly known miRNA detection tools for plants using benchmark datasets consisting of 223 mature miRNA sequences of *Arabidopsis thaliana* (TAIR10 cDNAs) and miRBase release 16 dataset having 213 numbers of sequences. We have presented the results of considered tools in terms of sensitivity, specificity, precision, etc. This study will help the biological to choose a better tool for novel miRNA detection from *Arabidopsis thaliana*.

# 1

# Introduction

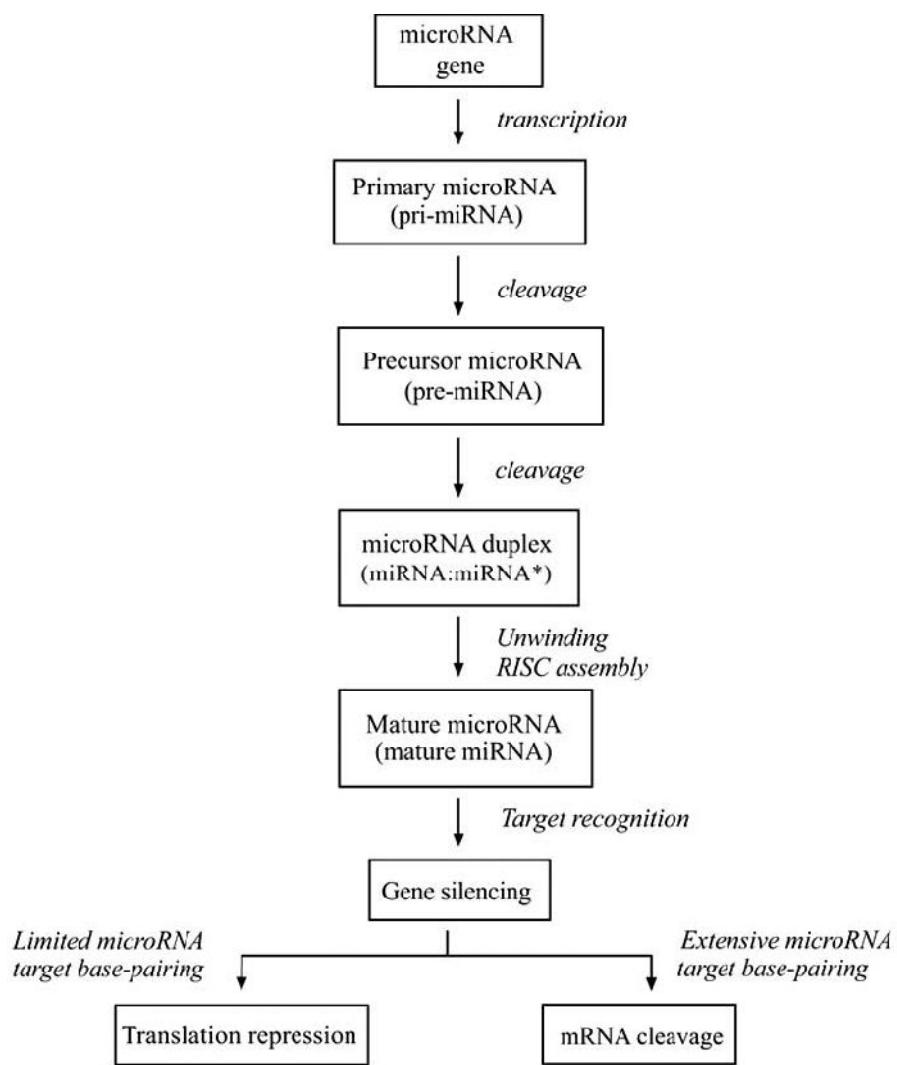
MicroRNAs are short non coding RNA molecules that are highly conserved (containing around ~22 nucleotides length), found in both plants and animals. miRNA functions in RNA silencing & control post transcriptional processes such as gene expression by complimentary base pairing (He and Hannon, 2004). Endogenous RNA have certain classes such as small transfer RNA (tRNA), small interfering RNA (siRNA), ribosomal RNA (rRNA), small nucleolar RNA (snoRNA) and microRNA (miRNA) (Bartel, 2004). miRNAs were first distinguished in *Caenorhabditis elegans* in the year 1993 utilizing hereditary techniques (Lee et al., 1993). MicroRNAs involve in different basic pathways for resolving of cell life, for example, advancement, ecological adjustment, and stress reaction. miRNAs are indicated from long transcripts originate in plants, viruses, and single-celled eukaryotes (Liu et al., 2012). miRNAs have turned into the focal point of various research because of their important role in Biogenesis and disease diagnosis and function to control certain biological procedures, for example, homeostasis (Liu et al., 2012). Cancer is the most widely recognized disease caused by miRNAs and their differential expressions produce various types of cancer, for example, lung disease (Yanaihara et al, 2006), prostate growth (Porkka et al., 2007), and ovarian disease (Yang et al., 2008). miRNAs also cause neurological disorders, such as Alzheimer's infection (Hébert et al., 2009), Schizophrenia (Beveridge et al., 2010), and various sclerosis (Cox et al., 2010). In the last few years, a large amount of miRNA information has been produced so that they can identify those miRNA targets that are responsible for diseases which

are difficult to find through biological methods. In this manner, Computational tools have been developed to provide statistical approaches. In the recent years, several tools have been developed for the detection of miRNA and their targets using deep sequencing data. Very few plant specific computational tools are available for the prediction of miRNA. Due to the different algorithms and features of tools, prediction results vary among themselves which create difficulty to recognize the best tool for miRNA prediction. Along these lines, it is troublesome for a researcher to pick the best plant miRNA prediction tools. In this paper, we have assessed the performance of 12 plant miRNA detection tools for *Arabidopsis thaliana* and *Zea mays* dataset giving the extensive outline of the considered tools, their miRNA prediction evaluation based on the different metrics involving number of miRNA detected, accuracy, sensitivity, specificity, false positive rates, true positive rates and so on. Several methodologies have been generated in the previous years to survey and assess the performance of existing miRNA detection tool.

## 1.1 Background

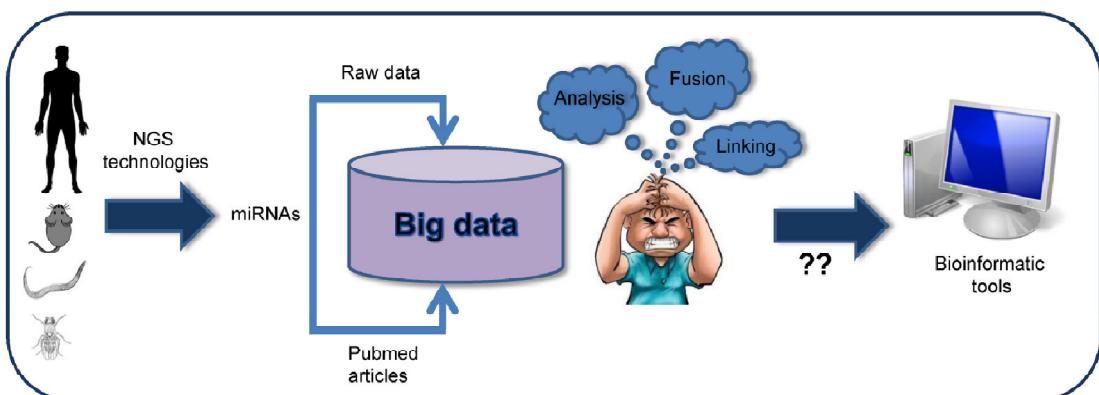
MicroRNA (miRNA), initially detected in *Caenorhabditis elegans*, is found in many eukaryotes, including human and plants. It is anticipated that miRNA represent 1-5% of the human genome and manage no less than 30% of protein-coding qualities. Accordingly to recent release of miRBase database, 1982 precursors and 2693 mature miRNAs have been recognized inside the human genome. It is obvious that miRNA assumes a significant part in the direction of gene expression controlling distinct cell and metabolic pathways. MiRNA are short, transformative preserved, single-stranded, non-coding RNA molecules tie target mRNA to check protein production by one of two different process.

Mature miRNA is produced by two-advance cleavage of essential miRNA (pri-miRNA), including effector complex RNA-induced silencing complex (RISC). The miRNA capacities as a guide by base-matching with target mRNA to negatively control its expressions (Lee et al., 2012). The level of complementarity between the guide and mRNA target figures out which silencing method will be utilized; cleavage of target mRNA with consequent degradation or translation inhibition the level of complementarity between the miRNA and the mRNA target (**Fig. 1.1**).



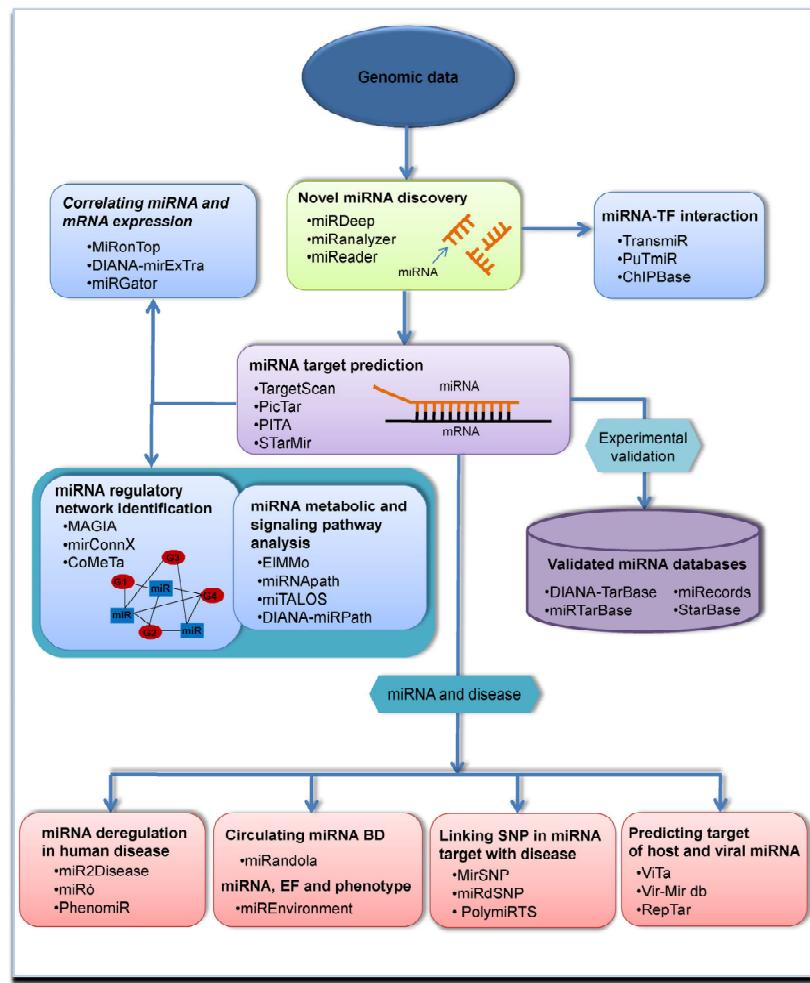
**Fig. 1.1 MicroRNA maturation & function (MacFarlane & Murphy, 2010)**

It has been found that miRNAs affect almost all kinds of cellular pathways. These effects may range from development to oncogenesis. The clinical implications of miRNAs are hypothesized to be an important tool for the detection of biomarkers for diagnostic, prognostic and therapeutic for various human related disease, specially cancers. The involvement of miRNAs in the gene regulation and disease processes, and development of high-throughput technologies such as Next Generation Sequence, and published literature increased size of data at super-linear rate (**Fig. 1.2**). As a result, a plethora of bioinformatics tools have been developed to manage the data flow.



**Fig. 1.2 The complexity and volume of data generated by NGS technologies and the need for bioinformatics tools for miRNAs detection (Akhtar et al., 2016)**

The **Fig. 1.3** presents a complete work of miRNA analysis, including correlating miRNA and mRNA expression, novel miRNA detection and miRNA-transcription factor interaction. It also includes list of available tools in each category. Fig. 1.3 also depicts the association and application of miRNA with disease, including miRNA deregulation in human disease, circulating miRNA BD, linking SNP in miRNA target with disease and predicting target of host and viral miRNA. The **Fig. 1.3**, also presents list of tools and databases used for each types of studies and analysis.



**Fig. 1.3 Diagrammatic view of flow of miRNA analysis and available bioinformatics tools. List of available tool in each category are also presented here (Akhtar et al., 2016)**

## 1.2 Motivation of Work

miRNAs are highly involved in gene regulation and disease processes in most eukaryotic organisms, therefore, microRNA detection is important for both disease diagnosis and target identification. Several traditional methods have been proposed by researchers, but traditional method is expensive and time consuming. Currently most researchers depend on computational programs to identify miRNA for successive validation. Therefore, computational prediction of miRNA targets is a fundamental step towards the characterization of miRNA targets prediction. There are number of

tools developed for the purpose of prediction of novel miRNA and their targets during the last 20 years. Computational tools with different algorithms and features have not tendency to give accurate and reliable result, there is still notable room for upgrading algorithm. Our major research question is “*Which among the available tools perform well?*”. Here, we ranked miRNA detection tools for plant.

### **1.3 Objectives**

The objectives of our studies are as follows:

- To identify list of tools for the plant miRNA detection.
- Compare and summarize them according to algorithmic features such as type of input datasets, algorithm employed, etc.
- Comprehensive evaluating performance of tools on benchmark datasets and rank them accordingly the accuracy.

### **1.4 Organization of Chapters**

First chapter of the thesis includes introduction along with the historical background of miRNA and their targets. All the facts related to discovery and findings of the miRNA and related their targets are covered in this portion under the heading of background. How do the miRNA mature and function is shown by a diagram (**Fig. 1.1**). Similarly, the need of bioinformatics tools for the analysis of miRNAs is depicted in **Fig. 1.2**). Apart from basic concepts related to MicroRNA, the introduction also covers the motivation and objective of the work.

Second chapter is on literature review, which is an elaborative version of the background discussed in the introduction part of the thesis. It is comprehensive collection of the information related to the miRNA and their targets present in

Arabidopsis are discussed in detail within this chapter. Other important things, covered in this chapter were the characteristics of the miRNA.

Third Chapter, methodology includes the step by step details, involved in this work. There is sub division of the chapter on the basis of the steps involved in series. At the beginning we have mentioned the source of the raw data, i.e. RNA sequences in form of FASTA file format. There is brief description of the tools, miRPlant, MiRENA, C-mii, MIRINHO, miReader, mireap, miRExpress and miRDeep-P; utilizing in this work. This was followed my mentioning the steps involved in the detection of miRNA, was described.

Fourth chapter is on results. It is an important part of the chapter illustrating the results, which were obtained through the comparative evaluation of miRNA prediction tools.

Final chapter of the thesis includes all the discussions related to the results obtained, along with the limitations that we discovered in the work are mentioned here in this chapter. In this work, we have also mentioned, our future work too.

# 2

# Literature Review

MicroRNAs are a kind of short RNAs having a modest number of nucleotides, i.e. 18–25 bp, which have significant role in cell formation processes, for example, digestion, immunological reactions organism development furthermore, tumorigenesis. miRNAs can either quell mRNA translation or prompt the cleavage of mRNA targets by means of hybridization with the 3' - untranslated portion of target mRNAs. Short RNA cloning strategies are utilized for distinguishing novel miRNAs, yet sequencing innovation advanced rapidly & deep sequencing through next generation sequencing (NGS) seems, by all accounts, to be exceptionally promising for miRNAs identification, since it gives the significant target points of high-throughput sequencing with rapid and lessened cost. A few researchers have effectively utilized NGS for the disclosure of novel miRNAs, particularly for those some of the species which are complex to recognize at a lower amplitude. Owing to the utilization of NGS data for miRNA prediction, several sequencing based software tools have been developed to identify miRNA and its analysis (Mendes et al., 2009). Some of the important list of tools are miRDeep-P, MIRINHO, miRExpress, miReader, C-mii, miRPlant, MIReNA, and mireap. Furthermore, mireap software tool can be downloaded from <http://sourceforge.net/ventures/mireap/>. miRDeep and mireap were current computational tools utilized to analyze deep sequencing data produced by NGS. In fact, these tools support fewer organism and species specially in plants. All tools support *Arabidopsis thaliana* which reference genome is accessible. **Mirinho** can be apply broadly in various organism. **miRExpress** is utilized if we don't have

access to genome sequencing (6). **MiReader** can be utilized in case of annotation file format. **Mireap** is a multitask standalone software. **C-mii** is a stand-alone Java based software tool which has been exclusively developed to detect plant miRNAs and their targets. At long last, **MIReNA** detects both miRNAs and pre-miRNAs within the dataset such as known miRNA sequence; profound sequencing data; putative pre-miRNAs, potentially including miRNA hopefuls; and long sequencing, including potential miRNAs. But, current analysis is only constrained to deep sequencing data. **miRDeep-P** is used for the quantification of known and new miRNA in plants from deep sequencing data. **miRPlant** (An et al., 2014) is a software program with improved accuracy of prediction which provide user friendly interface. The most common features and input type of these tools are characterized in **Table 2.1**.

**Table: 2.1 Characterization of Common Features for Selected Tools**

S. No.	TOOL	INPUT	FEATURES	ONLINE/ LOCAL
1	miRExpress	* Tab-delimited * FASTQ	* Structure * Conserevation	Local
2	MIReNA	* FASTA	*Structure	Local
3	Mirinho	* SAM	* Sequence * Structure	Local
4	MiReader	* FASTA * FASTQ	* Sequence * Conservation	Local
5	MIREAP	* FASTA	* Structure	Local
6	mirPlant	* FASTQ * BAM	* Structure * Conservation	Local
7	miRDeep*	* FASTA	* Structure	Local
8	C-mii	* FASTA	* Sequence	Local

Given this concise depiction of the individual computational tools of available, it is helpful to consider each program's ability as far as computational time, affectability, and precision and its importance for detecting novel miRNAs. Along these lines, we planned to estimate miRNA sequencing software tools to evaluate their abilities. The eight programming tools were tried utilizing deep sequencing dataset got from plant genome, i.e. *Arabidopsis thaliana* (Thale cress). This examination gives helpful data to researchers while choosing the software programming tools for miRNA detection relying upon their specific prerequisites and it gives a reference to computational scholars creating novel software tools. miRNA programming tools need to deal with two imperative issues while distinguishing miRNAs, namely, mapping NGS reads of the reference genomes and detecting 2\* structures of all mapped locus. Their feature details of miRNAs detection using each programming tools are summarize in **Table 2.1**. To deal with first issue, prediction algorithms are treated as crucial part of miRNA computational tools. MIReNA, miRDeep-P utilized MegaBLAST/BLAST calculation, so as to be classified together. MIReNA utilizes the methodology presented in miRDeep (An et al., 2012) in view of mapping NGS sequencing peruses of the reference genome utilizing MegaBLAST. SOAP2 (Li et al., 2009) is a prerequisite in MIReNA, which can decrease PC memory utilization and increment the alignment speed. miRPlant is based on miRDeep\* algorithm (An et al., 2012) and don't need third party tool. miRExpress utilize word-match and Smith-Waterman calculation, and therefore these are set in a similar class. This calculation is quite suitable for dealing with large volume of sequences on web-servers. The C-mii and miRPlant both are a stand-alone Java based software tool which has been exclusively developed to detect plant miRNAs and its targets too. C-mii is developed by Information Systems Laboratory, National Center for Genetic Engineering and

Biotechnology (<http://www3a.biotec.or.th/c-mii/>). Thailand. The calculation utilized for detecting structures of mapped loci was also considered. MiReNA utilize RNAfold and Bayes' hypothesis to assess RNA candidate's secondary structure. Up to 100 and 150nt of the genomic succession flanking the mapped locus in singular peruses is separated and collapsed utilizing RNAfold by MiReNA. Bayes' hypothesis is utilized by miRDeep-P to score probable miRNA candidate. miRExpress recovers cross-species grouping data using UCSC Genome Browser (Kuhn et al., 2009) deciding the protection of putative miRNAs. MiReNA software tool looks for miRNA successions by searching a multi-dimensional space specified by utilizing 5 important parameters to portray worthy miRNA candidate.

## COMMON FEATURES OF PLANT miRNA PREDICTION TOOLS

Computational strategies are utilized to recognize that how miRNAs particularly focuses on the mRNAs. Some common features on which generally plant miRNA detection tools are based on following as (Peterson et al., 2014; Faiza et al., 2017).

**SEED MATCH:** The section of miRNA beginning from 5'- end to the 3'- end comprising of initial 2-8 nucleotides is called the seed succession (Lewis et al., 2003). It is supposed as Watson-Crick (WC) coordinate between a miRNA and its target by plant miRNA detection tools. An arrangement between the miRNA and its target existing in the WC coordinating without any space in between is supposed as the ideal seed match. If algorithms are different then it considers different types of seed match. The most usually thought about seed matches are as per the following (Lewis et al., 2003):

a. *6-mer*: a flawless seed coordinating for six nucleotides between the miRNA seed and the mRNA.

*b. 7-mer-m8:* a perfect seed coordinate between 2-8 nucleotides of miRNA seed succession.

*c. 7mer-A1:* an ideal seed coordinate between 2-7 nucleotides of miRNA seed grouping in expansion to An over the miRNA first nucleotide.

*d. 8-mer:* a flawless seed coordinate between nucleotides 2-8 of miRNA seed succession moreover to an over the miRNA first nucleotide.

**FREE ENERGY:** It is a Gibb's free vitality which is utilized as estimate the stability of miRNA structure by different computational tools. At that point when a miRNA binds to the target mRNA conclusion to a stable structure, it is considered as the most likely target of that miRNA. The responses with more negative delta-G are less receptive, and consequently, have greater stability. The hybridization of miRNA inside its target mRNA give data about the high and low free vitality areas and delta-G predicts the quality of holding between the miRNA what's more, its target mRNA.

After literature survey we found that a plethora of software tools are being developed in the last two decades for the detection of miRNAs from sequences and deep sequencing data. Some of these tools support many species, while other set of tools are specific to a particular species. Also, some of them are based on sequence data (FASTA), while other takes deep sequencing data (FASTQ) or sequence aligned mapped (SAM) file. One of the important questions about these tools is their prediction accuracy. *How a biologist would choose a particular tool for the detection of the novel miRNAs?* Therefore, it is necessary to benchmark these tools and ranked them on the basis of prediction accuracy. The benchmarking of the tools would help the biologist and biotechnologist to select an appropriate tool for their studies.

# 3

# Materials & Methods

## 3.1 Datasets

For the benchmarking of various plant miRNA detection tools, we have considered two different datasets:

- (i) 223 mature miRNA sequences of *Arabidopsis thaliana* (TAIR10 cDNAs) having 33,602 number of sequences, and
- (ii) miRBase release 16 (only *Arabidopsis* precursor miRNAs) dataset having 213 number of sequences.

In this study, miRBase (Kozomara & Griffiths-Jones, 2011) is utilized as a kind of perspective dataset to depict thresholds and for building testing datasets of pre-miRNA/miRNA sets. The miRBase contains list of pre- and mature-miRNAs for various organisms including plants. Brief details about the considered datasets are shown in **Table 3.1**.

**Table 3.1** TAIR10 cDNAs mature miRNA sequences of *Arabidopsis thaliana*

S.N o.	Dataset Names	# of sequences	# of mature miRNAs	Source of mature miRNAs	Download link
1.	TAIR10 cDNAs (all sequences)	33,602	223	miRBase	<a href="http://www3a.biotech.or.th/c-mi/download/TAIR10_cdna.fasta">http://www3a.biotech.or.th/c-mi/download/TAIR10_cdna.fasta</a>
2.	miRBase Release 16 ( <i>Arabidopsis</i> only)	213	195	miRBase	<a href="http://www3a.biotech.or.th/c-mi/download/Arabidopsis_miRNA_hairpin_miRBase_16.fasta">http://www3a.biotech.or.th/c-mi/download/Arabidopsis_miRNA_hairpin_miRBase_16.fasta</a>

## 3.2 miRNA Detection Tools

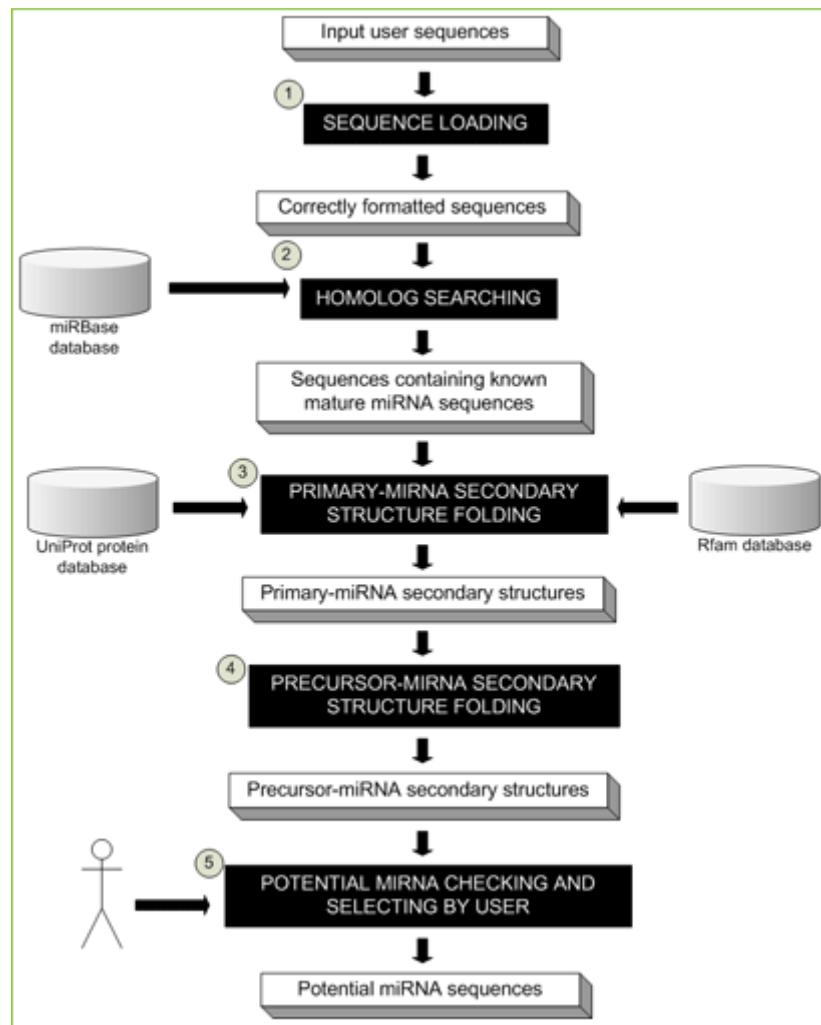
In this dissertation work, we have considered the following tools for the benchmarking.

### 3.2.1 C-mii (Computational miRNA Identification)

The C-mii (Numnark et al., 2012) is a stand-alone Java based software tool which has been exclusively developed to detect plant miRNAs and its targets too. It is developed by Information Systems Laboratory, National Center for Genetic Engineering and Biotechnology (<http://www3a.biotec.or.th/c-mii/>). Thailand. It consists of several software tools and databases which are installed standalone for the proper functioning. Some of the dependent software tools are standalone BLAST, UNAFold, Python, BioPerl. The pre-installed databases are miRBase, UniProt, and Rfam. C-mii tool follows a well-defined pipeline consisting of five steps, as shown in **Fig. 3.1**. The five steps are described as follows:

- (i) *Sequencing loading*: Input sequence can be loaded in FASTA format and checked for its correctness.
- (ii) *Homolog searching*: Sequences are searched for miRNA homolog against known plant mature miRNAs from miRBase database.
- (iii) *Primary miRNA secondary structure folding*: In this step, sequences which contains mature miRNAs of plant is optionally search against both UniProt and Rfam database in order to filter protein-coding sequences and other kinds of ncRNAs. After filtering, remaining sequences are used to predict the secondary structure of primary miRNAs by UNAFold tool.
- (iv) *Precursor miRNA secondary structure folding*: In this step, pre-miRNAs and short stem-loop structures are cleaved from primary miRNAs.

(v) *Potential miRNA detection:* The last step is the detection of miRNAs which can be decided by users based on various parameters such as minimum free energy, number of mismatches between known and predicted mature miRNAs, etc.



**Fig. 3.1 miRNA detection pipeline of C-mii**  
<http://www3a.biotecc.or.th/c-mii/>

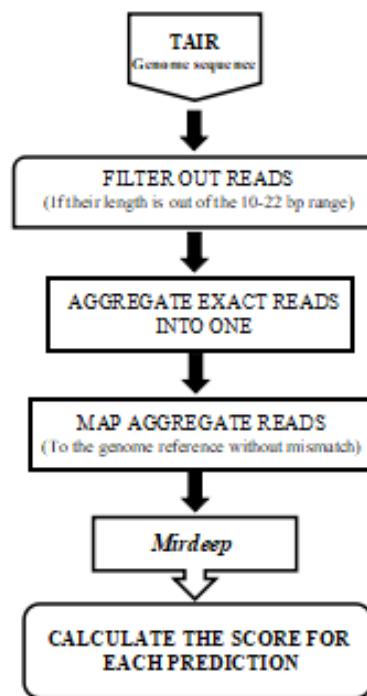
### 3.2.2 miRPlant (An Integrated PLant miRNA detection tool)

miRPlant (An et al., 2014) is an integrated software tool which is utilized to identify plant miRNA from deep sequencing data. It is totally based on JAVA programming language that freely available on <http://sourceforge.net/projects/mirplant/> site. It takes Fastq or BAM file format as a input for the identification of plant miRNAs. miRPlant

is based on miRDeep\* algorithm and don't require third party tools because they have been integrated into java library of miRPlant; such as genome mapping and RNA secondary structure prediction. miRDeep has extended to different plant species by receiving suitable strategies to predict hairpin excision region and hairpin structure for plants. miRPlant is a first software which predict hairpin structure for the detection of novel miRNAs. It has made easy to visualize new pre-miRNA structures and the location of sRNA reads to biologists. The significant benefit point of miRPlant is that it can easily used with limited bioinformatics skills.

### 3.2.2.1 Implementation

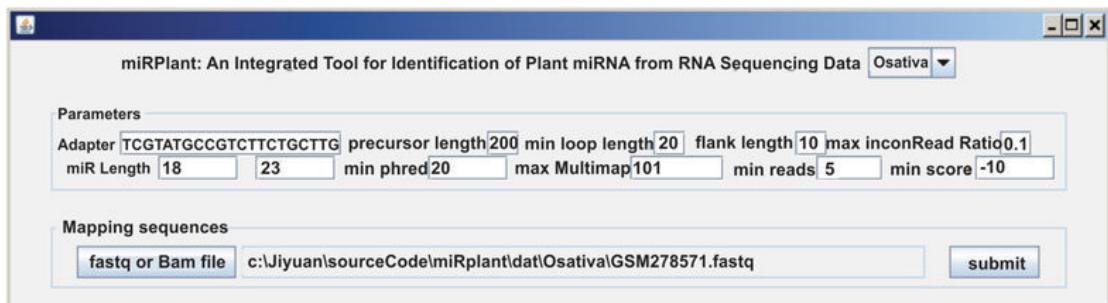
miRPlant work can be divided into five stages as following in the diagram:



**Fig. 3.2 Work Flow of miRPlant**

### 3.2.2.2 Parameter Setting

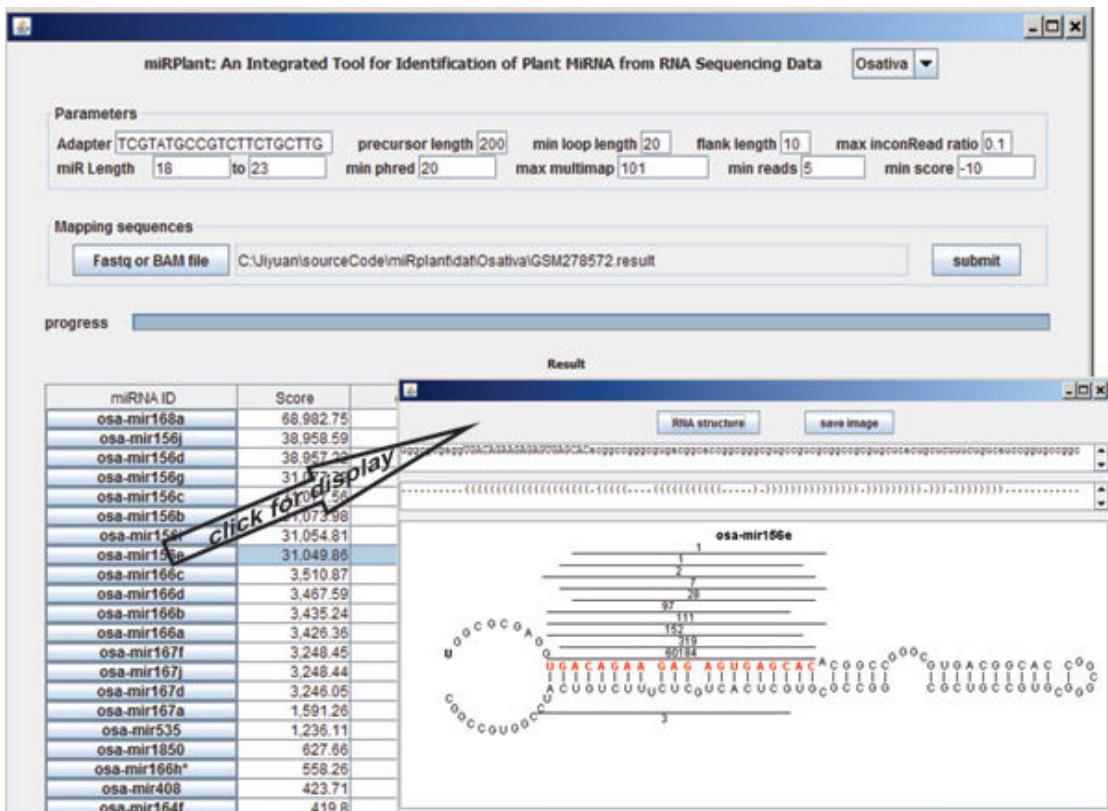
Different plant species may have various miRNA biogenesis so miRPlant allow users to customize parameters according to need, shown in **Fig 3.3**. The precursor length is set to 200bp by default in mirplant.



**Fig. 3.3 Snapshot of parameter settings in miRPlant**  
<https://doi.org/10.1186/1471-2105-15-275>

Precursor length shows length between mature microRNA and mature \*microRNA.

Six outputs generate in miRPlant which is similar to miRDeep\* (Fig. 3.4).

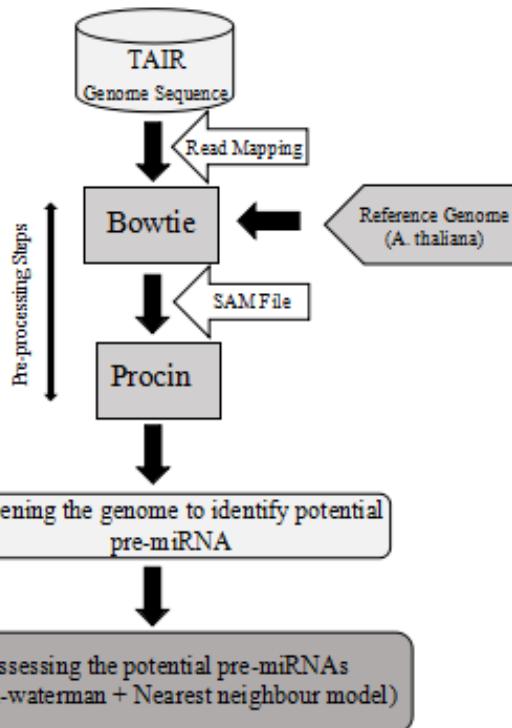


**Fig 3.4** (a) Genome sequence and mapping file (mapped file of small RNA reads against genome sequence using Bowtie) are given to PROCINE. (b) Region with reads are obtained using SAM file from the genome sequence. (c) Genome has screened to predict potential miRNAs. (d) Assessing the potential miRNAs by using Smith-waterman algorithm and Nearest Neighbour model.

<http://www.australianprostatecentre.org/research/software/mirplant>

### 3.2.3 MIRINHO (An efficient pre-miRNA detector)

Mirinho (Higashi et al., 2015) has developed to predict pre-miRNA hairpin structures which are an essential step in detection of miRNA (<http://mirinho.gforge.inria.fr/down.html>). Although cubic complexity algorithm is used for the identification of pre-miRNA hairpins structures but in the genome-wide scale it can become very slow. Mirinho is more considerable and quicker method with exceptionally sensitivity, specificity and precision which can be connected without any special adaptation. It uses genome sequence file or sRNA seq data as input (Fig. 3.5).



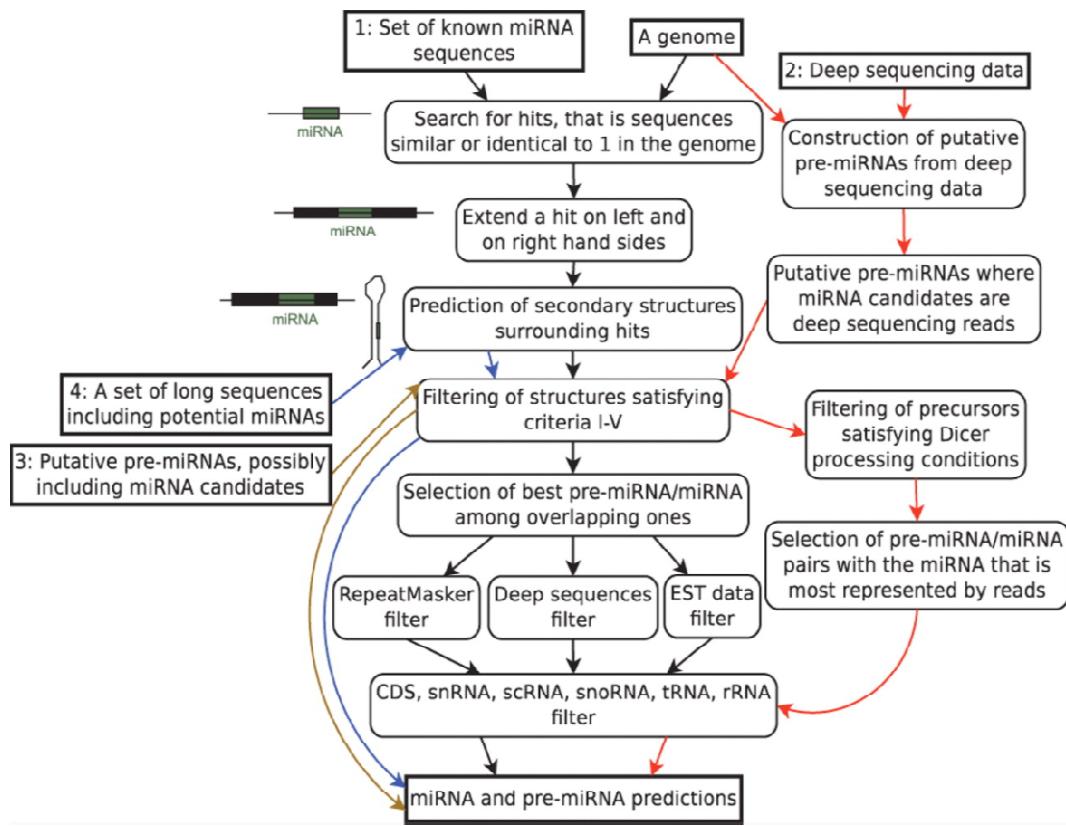
**Fig 3.5 Working Pipeline of Mirinho**

The thermodynamic Nearest- Neighbour (NN) model is used to compute free energy and detect the classical hairpin structures of pre-miRNAs. It shows that free energies thus connected well with RNAFOLD, this is known as MIRINHO. In place of cubic

complexity, it has quadratic which is more efficient in practice. The core algorithm is utilized for the identification of pre-miRNAs.

### 3.2.4 MIReNA (Software to predict miRNAs with high accuracy)

MIReNA (Mathelier & Carbone, 2010) is known as a genome wide algorithm which is found at <http://www.ihes.fr/~carbone/data8/>. It is developed in bash shell, C, perl and python. MIReNA searches for miRNA sequences by investigating a multi-dimensional space characterized by just five parameters describing acceptable pre-miRNAs. MIReNA approves pre-miRNAs having higher sensitivity and specificity values, and recognizes new miRNAs using homology search within previously known miRNAs. An execution correlation amongst this tool and four accessible predictive systems has been carried out. MIReNA method is prominently straightforward yet it ends up being effective at any cost as much as more complex algorithmic strategies. MIReNA gets preferred outcomes over three known algorithms which approve pre-miRNAs. It shows that machine-learning is not a vital computational algorithm for pre-miRNAs. Specifically, machine learning algorithm can just affirm pre-miRNAs which is identical to known ones, this being a restriction while investigating species with unknown pre-miRNAs. The likelihood to adjust the pursuit to particular species, potentially portrayed by particular properties of their miRNAs and pre-miRNAs, is an important property feature of this tool. A parameter customization computes specificity and sensitivity, a key component for prescient systems, that is absent in machine learning based algorithms. A comparison of MIReNA with miRDeep utilizing NGS data to anticipate miRNAs features is a very particular detection feature of MIReNA. The complete work-flow of MIReNA tool is depicted in Fig. 3.6. It is clear from the Fig. 3.6 that MIReNA software tool can be used in four ways.



**Fig 3.6** Schema of MIRENA algorithm. MIRENA can be utilized by four ways. Four different work paths are depicted as various colors (black, red, brown and blue). The square boxes describe input and output data, while arrows show flow of information <https://doi.org/10.1093/bioinformatics/btq329>

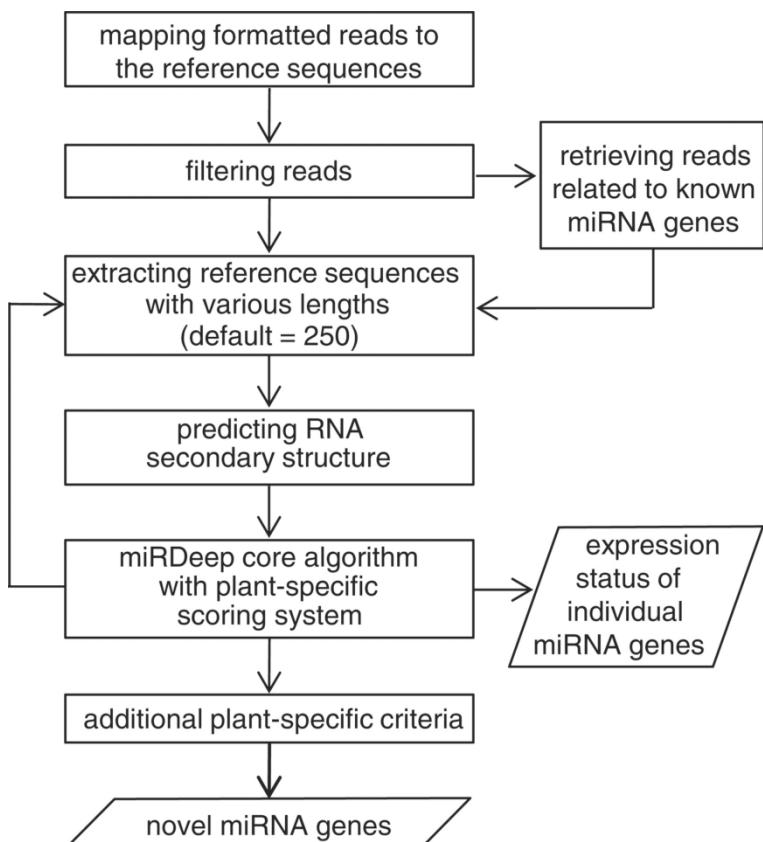
### 3.2.5 miRDeep-P (Software for analyzing microRNA transcriptome in plants)

miRDeep-P (Yang & Lee, 2011) is a modifying version of miRDeep, which is depend on a probabilistic model of miRNA biogenesis in animals, with a plant-specific scoring system and filtering criteria. miRDeep-P is an effective and easy-to-use tool for characterizing the miRNA transcriptome in plants.

#### Working description of miRDeep-P

This tool is based deep sequencing small RNA sequencing data from NGS. It allows the users to analyze expression patterns of annotated miRNAs as well as detect novel

miRNAs from it. **Fig. 3.7** depicts the working flow of miRDeep-P. Before using this software tool, short reads coming from NGS experiment need to go through quality control and data preprocessing steps including adapters trimming, discarding extremely short reads less than 15 nt, and further converting them to standard FASTA file with their copy number.



**Fig 3.7 Workflow of miRDeep-P**

<https://doi.org/10.1093/bioinformatics/btr430>

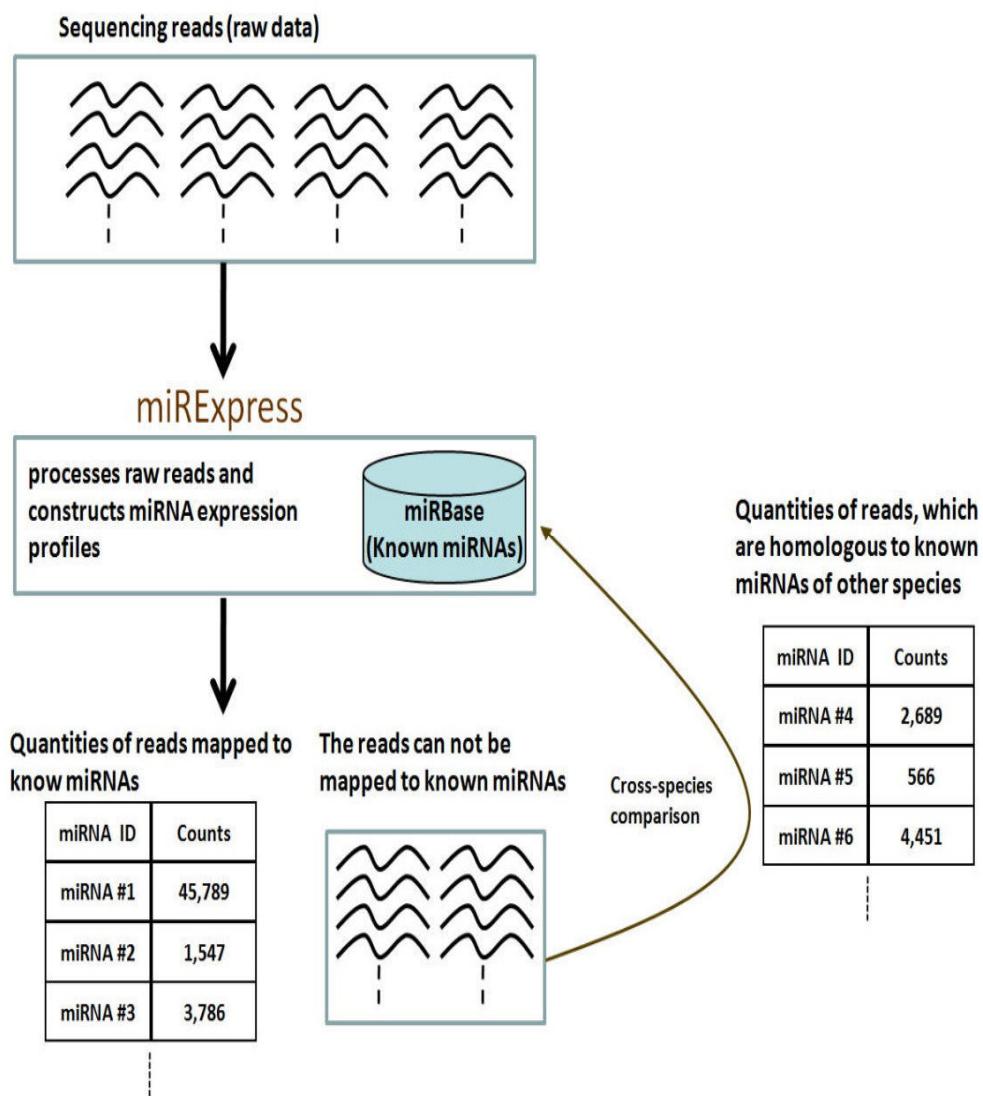
After giving correct file format as input to the tool, first it do the mapping tasks where reads are mapped to the reference genome using Bowtie aligner (Langmead et al., 2009). After mapping step, selection of an optimal window size is done for the extraction of reference sequences to prediction secondary structure of RNA which is 250 bp long (Yang & Lee, 2011). The miRDeep-P has a module which allows the users to empirically find a suitable window size. It uses miRDeep (Friedlander et al.,

2008) with a specific scoring system suitable for plant to extract secondary structures from the mapped reads. Its output is further filtered to extract plant-specific miRNA genes using some criteria. Overall, the tool quantifies signature reads related to small RNA which helps to identify reliable information for transcription and pre-miRNA processing. Thus, miRDeep-P effectively utilizes all these important information for miRNA transcriptome profiling.

### **3.2.6 miRExpress (Analyzing high-throughput sequencing data for profiling microRNA expression)**

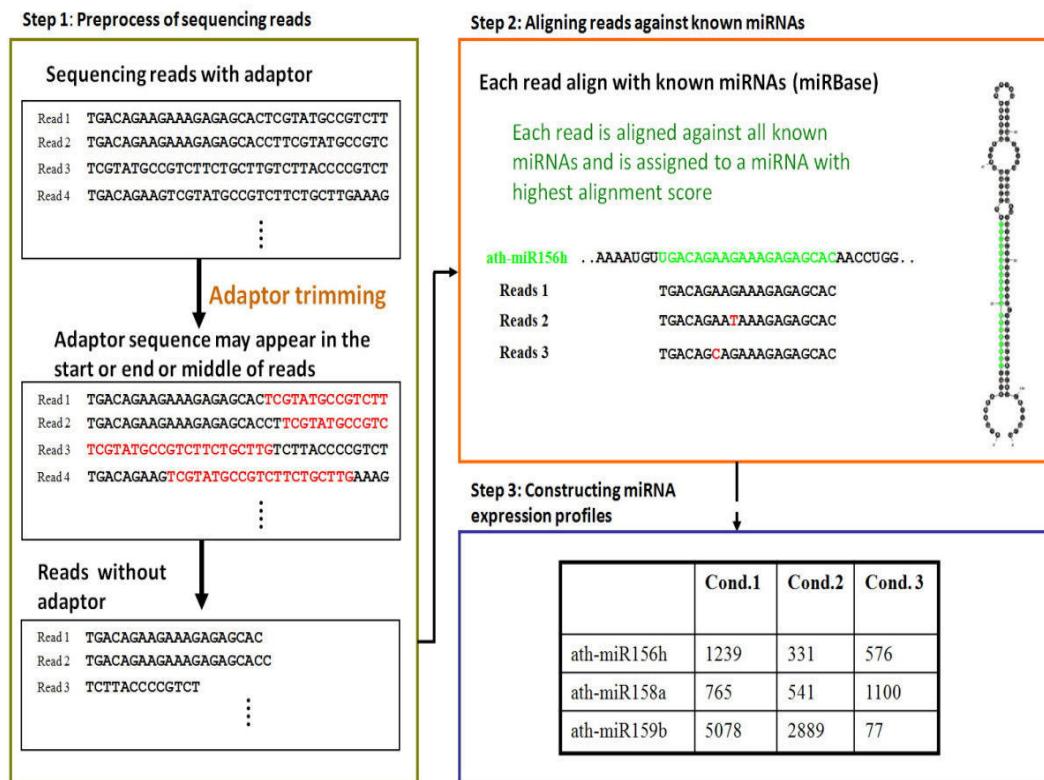
miRExpress (Wang et al., 2009) (<http://mirexpress.mbc.nctu.edu.tw/>) adopts a systematic approach to extract miRNA expression profiles from deep sequencing data from NGS experiment. It implements a stand-alone package to generate miRNA expression profiles without any sequenced genomes. miRExpress also maintains its own database, and it is very efficient and flexible for miRNA regulation investigation. This tool is found very effective for miRNA expression profiling and also shown its utility to detect novel miRNAs.

**Fig. 3.8** presents work flow of the tools including miRNA expression profiling. It takes NGS data as input consisting of one or more experimental conditions and performs miRNA expression profiling by aligning reads to known miRNAs. The miRNA expression profiling has three steps as shown in **Fig. 3.9**. Step-1 is NGS data preprocessing such as adapter trimming, while step-2 aligns all reads to known mature miRNAs. The step-3 constructs miRNA expression profiles using mapped alignment.



**Fig 3.8** Systematic work-flow of miRExpress including miRNA expression profiling. It accepts NGS data and constructs miRNA expression profiles using mapped alignment of known miRNAs. <https://doi.org/10.1186/1471-2105-10-328>

## miRExpress System Flow



**Fig. 3.9** Systematic workflow of miRExpress tool consisting of three steps: i) NGS data preprocessing, ii) read alignment against known miRNAs sequences, and iii) miRNA expression profiling. <https://doi.org/10.1186/1471-2105-10-328>

### 3.2.7 miReader (Discovering Novel miRNAs in Species without Sequenced Genome)

miReader (Jha & Shankar, 2013) is novel miRNA detection software tool which is downloadable from <http://scbb.ihbt.res.in/2810-12/miReader.php>. Most of the available tools to detect miRNAs rely upon the availability of reference genome which restricts its applications to the species of known reference genome. One of the distinguishing features of miReader tool is that it can predict mature miRNAs directly from NGS data without any need of reference genome. Hence, mature miRNAs can be detected from deep sequencing NGS data for those species too whose reference genome is not available. This method was tested and found to detect 21 novel mature

miRNA duplex candidates in a plant without any reference genome. This is expected that reference genome free approach to detect novel mature miRNAs will extend the application of miRNA biology to several new species. The workflow of miReader algorithm is shown in **Fig. 3.10**.

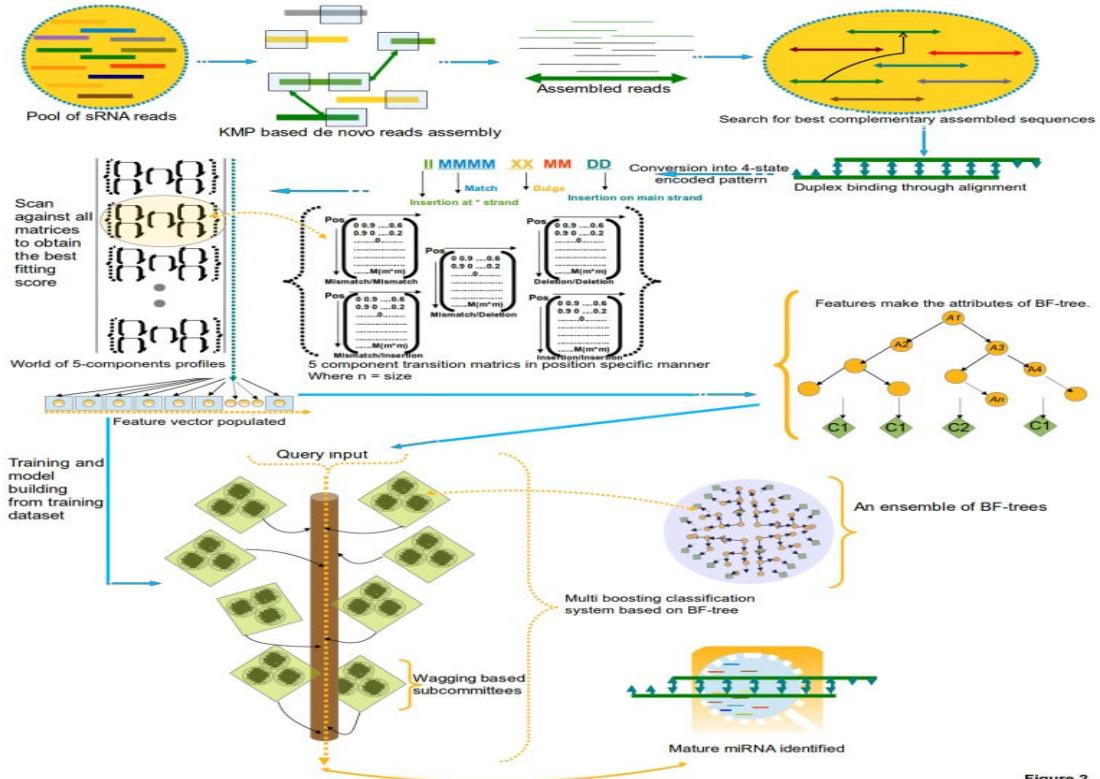


Figure 2

**Fig. 3.10 Work flow of miReader algorithm**

<http://dx.doi.org/10.1371%2Fjournal.pone.0066857>

### 3.3 Performance metrics

#### **Sensitivity and Specificity**

Sensitivity computes if miRNA recognition using various plant miRNA detection detection tools, we examine their execution when we used to two deep sequencing datasets of plants. All the miRNAs belong to different species stored in miRBase were characterized as positive set, while remaining entries were considered as negative set. This measure enabled us to look at the affectability of various prediction tools. It helped us comparing the sensitivity, specificity and precision of various prediction

tools in a stable and unbiased way. It was sensible to group the candidate as genuine set of miRNAs in case these are detected as miRNA candidates in the plants.

### ***Precision***

Precision is also a critical issue when detecting miRNAs. Clearly, anticipated miRNAs having less number of FPs are favored. miRBase was utilized as a source of perspective standard. The quantity of detected miRNAs was contrasted and aggregate number of detected miRNAs and level of known miRNAs distinguished by all considered detection tool.

To assess the performance of all the considered miRNAs detection tools, we considered the following performance metrics: the number of miRNAs accurately detected (true positives, TP), the number of false-miRNAs inaccurately detected as genuine miRNAs (pseudo-miRNA positives, FP) and the number of miRNAs inaccurately anticipated as pseudo-miRNAs (false negatives, FN). We utilized the following equations to compute the performance of the miRNA detection tools.

$$Sensitivity = \frac{TP}{(TP + FN)}$$

$$Specificity = \frac{FP}{(FP + TN)}$$

$$Precision = \frac{TP}{(TP + FP)}$$

# 4

# Results & Discussion

All the tools discussed in the previous chapter and their dependencies have been installed under Linux/Windows environment on Core i7 desktop having 32 GB of RAM. All the tools were executed using their default parameter settings. Results of each individual tool in terms of True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN), Sensitivity (Sn) and Specificity (Sp) are shown in the following sections (**Table 4.1** to **Table 4.8**).

**Table 4.1 C-mii tool**

S.N	Dataset Names	# of sequences after filtering	# of predicted miRNAs	TP	FP	TN	FN	S <sub>n</sub> (%)	S <sub>p</sub> (%)
1.	TAIR10 cDNAs <i>(all sequences)</i>	30,707	223	164	59	30,472	12	164/176 = <b>93.1</b>	30,472/30531 = <b>99.8</b>
2.	miRBase Release 16 <i>(Arabidopsis)</i>	213	195	195	0	0	18	195/213 = <b>91.5</b>	<b>0</b>

**Table 4.2 MIRINHO tool**

S.N	Dataset names	# of sequences after filtering	# of predicted miRNAs	TP	FP	TN	FN	S <sub>n</sub>	S <sub>p</sub>
1.	TAIR10 cDNAs <i>(all sequences)</i>	30,707	246	175	71	30,461	18	175/193 = <b>90.67</b>	30,461/30514 = <b>99.8</b>
2.	miRBase Release 16 <i>(Arabidopsis)</i>	213	210	210	0	0	3	210/213 = <b>98.59</b>	<b>0</b>

**Table 4.3 miRPlant tool**

S.N	Dataset names	# of sequences after filtering	# of predicted miRNAs	TP	FP	TN	FN	S <sub>n</sub>	S <sub>p</sub>
1.	TAIR10 cDNAs (all sequences)	30,707	310	190	120	30,397	29	190/219 <b>=86.75</b>	30,397/ 30,488 <b>=99.70</b>
2.	miRBase Release 16 (Arabidopsis)	213	174	150	34	0	39	150/189 <b>=79.36</b>	<b>0</b>

**Table 4.4 miReader Software Tool**

S.N	Dataset names	# of sequences after filtering	# of predicted miRNAs	TP	FP	TN	FN	S <sub>n</sub>	S <sub>p</sub>
1.	TAIR10 cDNAs (all sequences)	30,707	250	204	46	30,457	20	204/224 <b>=91.07</b>	30,397/ 30,488 <b>=99.70</b>
2.	miRBase Release 16 (Arabidopsis)	213	206	206	0	0	7	206/213 <b>=96.71</b>	<b>0</b>

**Table 4.5 miRExpress Software Tool**

S.N	Dataset names	# of sequences after filtering	# of predicted miRNAs	TP	FP	TN	FN	S <sub>n</sub>	S <sub>p</sub>
1.	TAIR10 cDNAs (all sequences)	30,707	271	211	60	30,436	33	211/244 <b>=86.47</b>	30,436/ 30,463 <b>=99.90</b>
2.	miRBase Release 16 (Arabidopsis)	213	190	190	0	0	23	190/213 <b>=89.20</b>	<b>0</b>

**Table 4.6 MIReNA Software Tool**

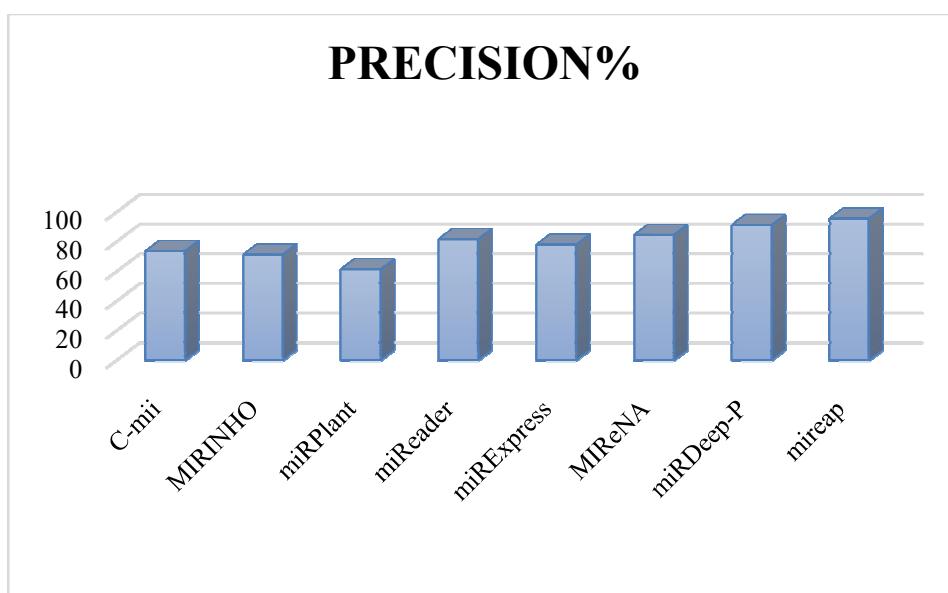
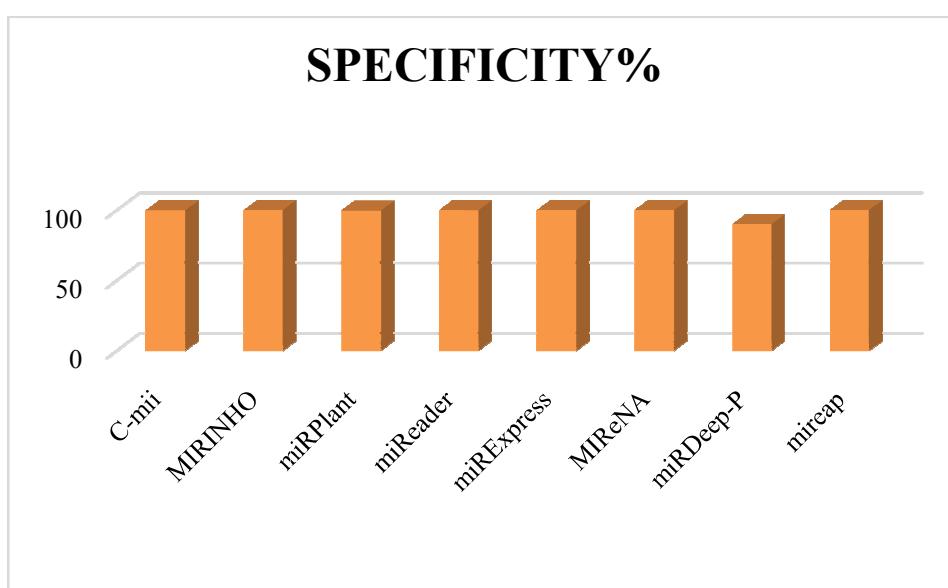
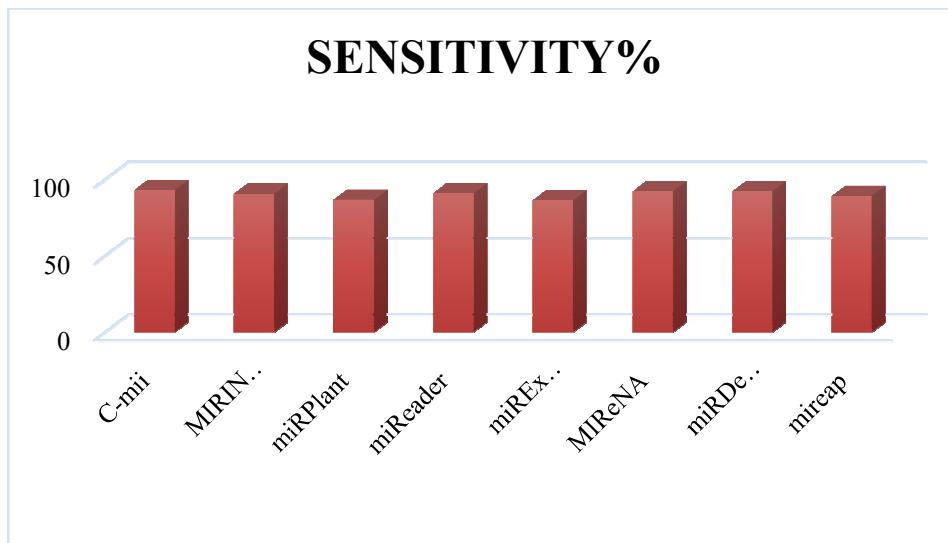
S.N	Dataset names	# of sequences after filtering	# of predicted miRNAs	TP	FP	TN	FN	S <sub>n</sub>	S <sub>p</sub>
1.	TAIR10 cDNAs (all sequences)	30,707	154	154	24	30,553	13	154/167 <b>=92.21</b>	30,553/ 30,563 <b>=99.96</b>
2.	miRBase Release 16 (Arabidopsis)	213	197	192	5	0	16	192/208 <b>=92.30</b>	<b>0</b>

**Table 4.7 miRDeep-P Software Tool**

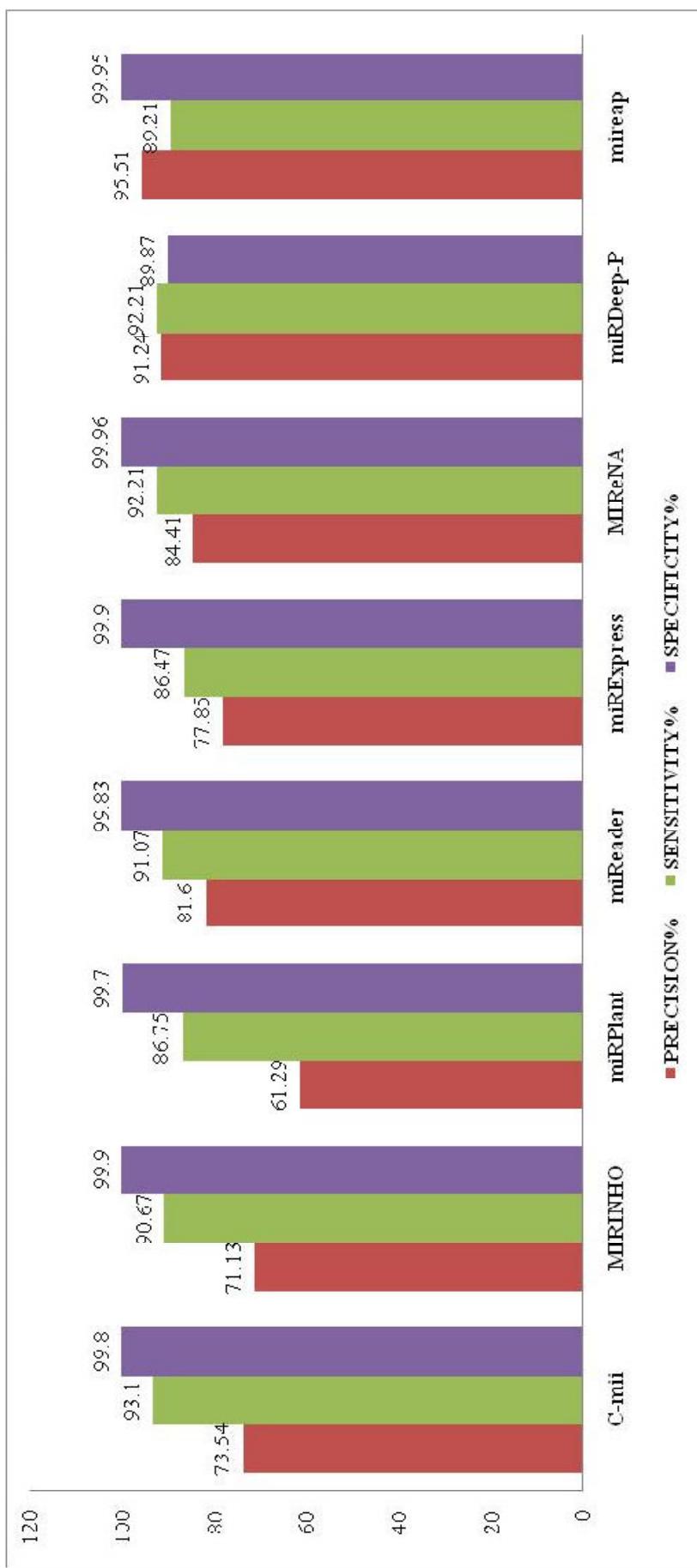
S.N	Dataset names	# of sequences after filtering	# of predicted miRNAs	TP	FP	TN	FN	S <sub>n</sub>	S <sub>p</sub>
1.	TAIR10 cDNAs (all sequences)	30,707	217	154	19	30,490	13	154/167 <b>=92.21</b>	30,553/ 30,563 <b>=99.96</b>
2.	miRBase Release 16 (Arabidopsis)	213	180	180	0	0	33	180/213 <b>=84.50</b>	<b>0</b>

**Table 4.8 mireap Software Tool**

S.N	Dataset names	# of sequences after filtering	# of predicted miRNAs	TP	FP	TN	FN	S <sub>n</sub>	S <sub>p</sub>
1.	TAIR10 cDNAs (all sequences)	30,707	312	298	14	30,395	38	298/336 <b>=88.69</b>	30,395/ 30,409 <b>=99.95</b>
2.	miRBase Release 16 (Arabidopsis)	213	167	162	5	0	46	162/208 <b>=77.88</b>	<b>0</b>



**Fig. 4.1** Sensitivity, Specificity and Precision of each tools



**Fig. 4.2** Performance comparison of various tools

# Conclusion & Future Work

Last three decades has shown a tremendous growth in miRNA biology research due to the fact that these have been found to be very crucial constituent involved in gene regulation processes, disease pathways including several different cancer types and neurological disorder. Therefore, it has become very important for the biological research community to know novel miRNAs, their targets and overall their biological functions. Last two decades have witnessed tremendous growth in biological databases for miRNAs, their targets and annotations such as miRBase, miRTarbase to name a few. Detection of miRNAs using biological experimental techniques are expensive, time-consuming, labour intensive and also not very much reliable to the limitation of experimental techniques. Hence, the researchers started looking forward for sophisticated computational approach of solving these problems.

Today, several miRNA detection software tools are available which works on different algorithmic principles. Some of these tools support multiple species, while others support specific species. After the deep sequencing techniques such as next-generation sequencing (NGS), a large number of new tools have been developed which have been found to perform well. In this dissertation, we have evaluated the performance of miRNA detection tools and ranked them accordingly to their performance metrics such as sensitivity, specificity, precision, etc. Among all the considered tools, C-mii is found to perform the best having sensitivity of 93.1% and

specificity of 99.8% for TAIR10 datasets, while it has sensitivity of 91.5% for miRBase release 16 database. After C-mii, the next well performing tools are MIRENA and miReader.

Our results suggest that most of these tools suffer from large number of false positives which makes the specificity low. Therefore, we need better filtering technique within the algorithm to filter out false positive detections. The accuracy of these tools are also dependent on other factors such as quality of sequence reads, accuracy of read mapping algorithms, and quality of reference genome or transcriptome. The limitation of our work is that we have selected few tools which supports normal sequences while some other tools supports deep sequencing data from NGS.

To enhance our work, some of the future directions are as follows:

- (1) Separate evaluation for normal sequence supported tools and deep sequencing based tools.
- (2) Consider more number of software tools for plants.
- (3) For rigorous testing of these tools, we can consider more benchmark datasets for its unbiased evaluation.
- (4) miRNA target detection is also very crucial. Therefore, we also need to evaluate and rank miRNA target finding tools.

# References

- Akhtar, M. M., Micolucci, L., Islam, M. S., Olivieri, F., & Procopio, A. D. (2015). Bioinformatic tools for microRNA dissection. *Nucleic acids research*, 44(1), 24-44.
- An, J., Lai, J., Lehman, M. L., & Nelson, C. C. (2012). miRDeep\*: an integrated application tool for miRNA identification from RNA sequencing data. *Nucleic acids research*, 41(2), 727-737.
- An, J., Lai, J., Sajjanhar, A., Lehman, M. L., & Nelson, C. C. (2014). miRPlant: an integrated tool for identification of plant miRNA from RNA sequencing data. *BMC bioinformatics*, 15(1), 275.
- Bartel, D. P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*. [https://doi.org/10.1016/S0092-8674\(04\)00045-5](https://doi.org/10.1016/S0092-8674(04)00045-5)
- Beveridge, N. J., Gardiner, E., Carroll, A. P., Tooney, P. A., & Cairns, M. J. (2010). Schizophrenia is associated with an increase in cortical microRNA biogenesis. *Molecular psychiatry*, 15(12), 1176.
- Cox, D., Brennan, M., & Moran, N. (2010). Integrins as therapeutic targets: lessons and opportunities. *Nature reviews Drug discovery*, 9(10), 804.
- Faiza, M., Tanveer, K., Fatihi, S., & Raza, K. (2017). Comprehensive assessment of miRNA target prediction tools for human and drosophila melanogaster. *arXiv:1711.01632*
- He, L., & Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews Genetics*, 5(7), 522.
- Hébert, S. S., Horré, K., Nicolaï, L., Bergmans, B., Papadopoulou, A. S., Delacourte, A., & De Strooper, B. (2009). MicroRNA regulation of Alzheimer's Amyloid precursor protein expression. *Neurobiology of disease*, 33(3), 422-428.
- Higashi, S., Fournier, C., Gautier, C., Gaspin, C., & Sagot, M. F. (2015). M irinho: An efficient and general plant and animal pre-miRNA predictor for genomic and deep sequencing data. *BMC bioinformatics*, 16(1), 179.
- Jha, A., & Shankar, R. (2013). miReader: Discovering novel miRNAs in species without sequenced genome. *PloS one*, 8(6), e66857.
- Kozomara, A., & Griffiths-Jones, S. (2011). MiRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, 39(SUPPL. 1). <https://doi.org/10.1093/nar/gkq1027>
- Kuhn, R. M., Karolchik, D., Zweig, A. S., Wang, T., Smith, K. E., Rosenbloom, K. R., ... Kent, W. J. (2009). The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res*, 37(Database issue), D613–D619.

- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3), R25.
- Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *cell*, 75(5), 843-854.
- Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., & Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, 115(7), 787-798.
- Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., & Wang, J. (2009). SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics*, 25(15), 1966–1967. <https://doi.org/10.1093/bioinformatics/btp336>
- Li, Y., Zhang, Z., Liu, F., Vongsangnak, W., Jing, Q., & Shen, B. (2012). Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis. *Nucleic acids research*, 40(10), 4298-4305.
- MacFarlane, L. A., & Murphy, P. (2010). MicroRNA: biogenesis, function and role in cancer. *Current genomics*, 11(7), 537-561.
- Mathelier, A., & Carbone, A. (2010). MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. *Bioinformatics*, 26(18), 2226-2234.
- Mendes, N. D., Freitas, A. T., & Sagot, M. F. (2009). Survey and Summary: Current tools for the identification of miRNA genes and their targets. *Nucleic Acids Research*, 37(8), 2419–2433. <https://doi.org/10.1093/nar/gkp145>
- Numnark, S., Mhuantong, W., Ingsriswang, S., & Wichadakul, D. (2012). C-mii: a tool for plant miRNA and target identification. *BMC Genomics*, 13(Suppl 7), S16. <https://doi.org/10.1186/1471-2164-13-S7-S16>
- Peterson, S. M., Thompson, J. A., Ufkin, M. L., Sathyanarayana, P., Liaw, L., & Congdon, C. B. (2014). Common features of microRNA target prediction tools. *Frontiers in genetics*, 5, 23.
- Porkka, K. P., Pfeiffer, M. J., Walterling, K. K., Vessella, R. L., Tammela, T. L., & Visakorpi, T. (2007). MicroRNA expression profiling in prostate cancer. *Cancer research*, 67(13), 6130-6135.
- Wang, W. C., Lin, F. M., Chang, W. C., Lin, K. Y., Huang, H. D., & Lin, N. S. (2009). miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC bioinformatics*, 10(1), 328.
- Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., ... & Calin, G. A. (2006). Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer cell*, 9(3), 189-198.

Yang, N., Kaur, S., Volinia, S., Greshock, J., Lassus, H., Hasegawa, K., ... & Johnstone, C. N. (2008). MicroRNA microarray identifies Let-7i as a novel biomarker and therapeutic target in human epithelial ovarian cancer. *Cancer research*, 68(24), 10307-10314.

Yang, X., & Li, L. (2011). miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics*, 27(18), 2614-2615.

# FullChapter

## ORIGINALITY REPORT



## PRIMARY SOURCES

- |   |  |     |
|---|--|-----|
| 1 | Li, Y., Z. Zhang, F. Liu, W. Vongsangnak, Q. Jing, and B. Shen. "Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis", Nucleic Acids Research, 2012.<br>Publication | 4%  |
| 2 | Leigh-Ann MacFarlane. "MicroRNA: Biogenesis, Function and Role in Cancer", Current Genomics, 11/01/2010<br>Publication   | 2%  |
| 3 | Akhtar, Most Mauluda, Luigina Micolucci, Md Soriful Islam, Fabiola Olivieri, and Antonio Domenico Procopio. "Bioinformatic tools for microRNA dissection", Nucleic Acids Research, 2016.<br>Publication            | 1 % |
| 4 | <a href="http://www.chinapubmed.net">www.chinapubmed.net</a><br>Internet Source  | 1 % |
| 5 | Yang, X., and L. Li. "miRDeep-P: a computational tool for analyzing the microRNA   | 1 % |