EVALUATION OF miRNA PREDICTION TOOLS AND IN *SILICO* ANALYSIS OF MICRO AND LONG NON CODING RNAs IN SWEET POTATO

By

ASWATHY M. B.

(2013-09-105)

THESIS

Submitted in partial fulfilment of the

requirements for the degree of

B. Sc. – M. Sc. (INTEGRATED) BIOTECHNOLOGY

Faculty of Agriculture

Kerala Agricultural University, Thrissur

B.Sc. - M.Sc. (INTEGRATED) BIOTECHNOLOGY
DEPARTMENT OF PLANT BIOTECHNOLOGY
COLLEGE OF AGRICULTURE
VELLAYANI, THIRUVANANTHAPURAM- 695 522
KERALA, INDIA

2018

DECLARATION

I, hereby declare that this thesis entitled **"Evaluation of miRNAs prediction tools and *in silico* analysis of micro and long non coding RNAs in sweet potato"** is a bonafide record of research work done by me during the course of research and the thesis has not previously formed the basis for the award to me of any degree, diploma, associateship, fellowship or other similar title, of any other University or Society.

Vellayani,

Date: 07/12/2018

Aswathy M. B.

(2013-09-105)

एक कदम स्वच्छता की ओर

भा.कृ.अनु.प- केंद्रीय कन्द फसल अनुसंधान संस्थान

(भारतीय कृषि अनुसंधान परिषद, कृषि और किसान कल्याण मंत्रालय, भारत सरकार)

श्रीकार्यम, तिरुवनन्तपुरम-695 017, केरल, भारत

ICAR- CENTRAL TUBER CROPS RESEARCH INSTITUTE

(Indian Council of Agriculture Research, Ministry of Agriculture and Farmers Welfare, Govt. of India)

Sreekariyam, Thiruvananthapuram-695 017, Kerala, India

ISO 9001:2008

3

CERTIFICATE

This is to certify that this thesis entitled **"Evaluation of miRNA prediction tools and *in silico* analysis of micro and long non coding RNAs in sweet potato"** is a record of research work done independently by **Ms. Aswathy M. B.** (2013-09-105) under my guidance and supervision and that this has not previously formed the basis for the award of any degree, diploma, fellowship or associateship to her.

Dr. J. Sreekumar
(Chairperson, Advisory Committee)
Principal Scientist (Agrl. Statistics),
Section of Extension and Social Sciences, ICAR-CTCRI. Sreekariyam,
Thiruvananthapuram- 695 017

Place: Sreekariyam
Date: 07/12/2018

डॉ. जे. श्रीकुमार/ Dr. J. SREEKUMAR
प्रधान वैज्ञानिक (कृषि सांख्यिकी)
Principal Scientist (Agricultural Statistics)
एक्सटेंशन और सामाजिक विज्ञान अनुभाग
Section of Extension and Social Sciences
भा कृ अनु प - केंद्रीय कंद फसल अनुसंधान संस्थान
I C A R - Central Tuber Crops Research Institute
श्रीकार्यम / Sreekariyam
तिरुवनंतपुरम / Thiruvananthapuram- 695 017

Phone : +91-471-2598551 to 2598554
Director (Per) : +91-471-2598431
(Res) : +91-471-2597211
Sr. Admn. Officer : +91-471-2598193

Fax : +91-471-2590063
E-mail : director.ctcri@icar.gov.in
ctcritvm@gmail.com
Web : http://www.ctcri.org

CERTIFICATE

We, the undersigned members of the advisory committee of Ms. Aswathy M. B. (2013-09-105), a candidate for the degree of B. Sc. – M. Sc. (Integrated) Biotechnology, agree that the thesis entitled **"Evaluation of miRNA prediction tools and *in silico* analysis of micro and long non coding RNAs in sweet potato"** may be submitted by Ms. Aswathy M. B. in partial fulfilment of the requirement for the degree.

Dr. J. Sreekumar
(Chairperson, Advisory Committee)
Principal Scientist (Agrl. Statistics),
Section of Extension and Social Sciences, ICAR-CTCRI. Sreekariyam,
Thiruvananthapuram- 695 017

Dr. Swapna Alex
(Member, Advisory Committee)
Professor & Head
Department of Plant Biotechnology
College of Agriculture, Vellayani
Thiruvananthapuram – 695 522

Dr. Shirly Raichal Anil
(Member, Advisory Committee)
Senior Scientist
Division of Crop improvement
ICAR-CTCRI. Sreekariyam
Thiruvananthapuram - 695 017

Dr. K. B. Soni
(Member, Advisory Committee)
Professor & Course Director
B. Sc. – M. Sc. (Integrated) Biotechnology
Department of Plant Biotechnology
College of Agriculture, Vellayani
Thiruvananthapuram – 695 522

Dr. M. K. Rajesh
(External Examiner)
Principal Scientist
(Biotechnology)
ICAR-CPCRI
Kasaragod – 671 124, Kerala

ACKNOWLEDGEMENT

It gives me immense pleasure to express deep sense of gratitude to my major advisor ***Dr. J. Sreekumar, Principal Scientist, ICAR- CTCRI*** *for the identification of research project and his immaculate guidance, patience and sympathy throughout the study.*

I wish to place on record my sincere thanks to ***Dr. Swapna Alex, Professor & HOD, Department of Plant Biotechnology*** *for her valuable suggestions in the preparation of the thesis.*

I am deeply indebted to and take this opportunity to express my boundless gratitude to ***Dr. Shirly Raichal Anil, Senior Scientist, ICAR- CTCRI and Dr. C. Mohan, Principal Scientist, ICAR-CTCRI*** *for their unstinting efforts towards my wet lab experiments while devoting their precious time in the midst of their busy schedule.*

My sincere gratitude to ***Dr. Archana Mukherjee, Director, ICAR-CTCRI*** *for allowing me to be a part of ICAR-CTCRI, Sreekariyam.*

I am deeply indebted to ***Dr. K. B. Soni, Professor and Course Director, B. Sc. - M. Sc. (Integrated) Biotechnology course,*** *for her valuable suggestions during the project formulation and patient scrutiny of the thesis.*

I find it difficult to express my gratitude to ***Dr. Senthilkumar K. M., Scientist ICAR-CTCRI, Prakash chettan, Athulya chechi*** *who was always with me rendering whatever help I needed during my wet lab experiments.*

My heart fills with happiness while I express my deep sense of gratitude to ***Ambu chettan, Priya chechi, Rekha chechi, Vishnu chettan, Rajeev chettan, and Shyam*** *for their warm welcome and sisterly affection.*

I thank my fellow classmates ***Bimal, Sahla, Reshma, Adithya, Vishnu, Sruthi, Sabari, Shahana, Athul, Alina, Reshma R, Arya A, Bhaghyalekshmi, Anjitha, Arundhathy, Arya RS, Aryalekshmi, Achuth and Ajith*** *for all the fun we have had in the last five years.*

Also I thank my besties ***Jithu, Gayathri, Sruthi, Akshay, Vinod and Haritha*** *who are always beside me encouraging and providing moral support.*

Last but not the least; I would like to thank my family: ***Achan Amma and Appu***

for supporting me spiritually throughout my life.

Once again, I thank ***God*** *for making the project completion a reality.*

ASWATHY M. B.

CONTENTS

LIST OF TABLES

LIST OF TABLES

9

Continued

LIST OF FIGURES

LIST OF FIGURES

Continued

LIST OF PLATES

12

LIST OF APPENDICES

13

LIST OF ABBREVIATIONS

A	Adenine
Acc	Accuracy
AGO	Argonaute
AMFE	Adjusted minimal folding free energy
ASCO	Alternative splicing competitor
BLAST	Basic Local Alignment Search Tool
C. elegans	*Caenorhabditis elegans*
c. DNA	Complementary DNA
CDS	Coding sequences
CIP	International Potato Center
COLDAIR	COLD ASSISTED INTRONIC NONCODING RNA
COOLAIR	COLD INDUCED LONG ANTISENSE INTRAGENIC RNAs
CPC	Coding Potential Calculator
CSF	Codon Substitution Frequency
CTCRI	Central Tuber Crops Research Institute
C_T	Threshold cycle
DCL1	Dicer like 1
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
ELENA1	ELF18 INDUCED LONG-NONCODING RNA1
EST	Expressed sequence tags
FAOSTAT	Food and Agriculture Organization Corporate Statistical Database
FASTA	fast-all
FIS	Functional information score
FLC	*FLOWERING LOCUS C*

LIST OF ABBREVIATIONS

15

Continued

FN	False negative
FP	False positive
G	Guanine
GCC	GNU compiler collection,
GNU	GNU's not unix
GO	Gene ontology
GSS	Genome survey sequence
Ha	Hectare
HMM	Hidden Markov Model
HST	Hasty
HYL1	Hyponastic leaves
Iba	*Ipomoea batatas*
ICAR	Indian Council of Agricultural Research
IPS1	*Induced by phosphate starvation 1*
lncRNA	long non coding RNA
Max	Maximum
MCC	Matthews Correlation Coefficient
MFE	Minimum folding energy
MFEI	Minimum folding energy index
miRNA	microRNA
miRdDM	miRNA-directed DNA methylation
mRNA	messenger RNA
NCBI	National Centre for Biotechnology Information
ncRNA	non coding RNA
nt	nucleotide
NSR	nuclear speckle RNA-binding protein

LIST OF ABBREVIATIONS

16

Continued

ORF	Open reading frame
PCR	Polymerase chain reaction
Pre-miRNA	Precursor microRNA
pht	phosphate transporter
PPV	Positive predictive value
PTI	Pattern triggered immunity
RNA	Ribonucleic Acid
rRNA	ribosomal RNA
RNA-Seq	RNA- sequencing
RT	Reverse transcription
RTPCR	Reverse transcription polymerase chain reaction
RTqPCR	Reverse transcription quantitative polymerase chain reaction
SAGE	Serial analysis of gene expression
SE	Serrate
Se	Sensitivity
Sp	Specificity
ssRNA-Seq	Strand specific RNA sequencing
SVM	Support vector machine
T	Thymine
Tm	Melting temperature
TN	True negative
TP	True positive
tRNA	transfer RNA
U	Uracil
UBC	Ubiquitin conjugase

LIST OF ABBREVIATIONS

17

Continued

UPE	Unpaired energy
URL	Uniform resource locator
UTR	Untranslated region
µl	microlitre

INTRODUCTION

1. INTRODUCTION

19

Sweet potato (*Ipomoea batatas* L.) is a staple food ranking seventh among food crops which can remove malnutrition, a major challenge in many developing nations by providing wealthy carbohydrates, vitamins, minerals, dietary fibers, antioxidants etc. for a better health. It is a hexaploid plant with chromosome number 2n=90 that belongs to *Convolvulaceae* family, a native to tropical America. In line with Food and Agriculture Organization Corporate Statistical Database (FAOSTAT, 2016) world's production of sweet potato in 2016 is 105.2 million tonnes and the largest producer is China with 70.6 million tonnes. In India sweet potato cultivation is spread over 2 lakh hectares of land and the stateswith highest production states are Orissa, West Bengal, Uttar Pradesh and Madhya Pradesh. In Kerala, sweet potato is cultivated in an area of 220 ha and Malappuram district stands 1st with an area of 61ha (27.73%).

There are some constraints that limit the production of sweet potato worldwide which includes drought, poor soil fertility, low yielding varieties, viral diseases and attacks of insect pests (Mwanga *et al.*, 2011). Plant breeding and genetics were focussed on developing improved cultivars worldwide, but this approach seems to be very difficult due to the self and cross incompatibility in sweet potato. Biotechnological advances offer different strategies in order to make smart crops. One of the widely accepted and economically important strategies at present is the exploitation of non coding RNAs (ncRNAs). The regulative ncRNAs embrace each short (<30nt length) and long (>200nt length) RNA species with epigenetic functions. These ncRNAs, as a powerful tool of genetic engineering and functional genomics can have great potential to generate crop cultivars with modified agronomic traits.

microRNAs (miRNA) candidates are 18-22nt short, endogenous non coding RNA with potent action in genome regulation in many biological processes in diverse species. As a potent gene regulator they mediate their action by negatively regulating their target of interest through either translation repression or mRNA cleavage based on perfect or near perfect complementarity (Bartel, 2004). The first plant miRNA was discovered in *Arabidopsis thaliana*

(Reinhart *et al.*, 2002). Plant microRNAs are shown to be act as a master ribo regulator in controlling the growth and development, cell signalling, stress responses in plants complex molecular network.

Another important candidate in regulatory non coding RNA family is long non coding RNAs with versatile functions in different species. lncRNAs identity are maintained by having more than 200 nucleotide(nt) length with either no functional polypeptides or with short open reading frames. The functional importance of this long genome regulators were reported in early 1990s, with the discovery of lncRNAs involved in epigenetic regulation, such as *H19* (Brannan *et al.,* 1990) and *Xist* (Brockdorff *et al.,* 1992). lncRNAs can have prominent role in plant gene regulatory network dealing with growth cycle and environmental stress regulations.

The future agricultural sustainable developments will surely depend on miRNA and lncRNA based biotechnology. So it is pre requisite to identify and characterize them precisely. Such a brilliant effort will really made robust contributions in the field of crop improvement and thus fulfil the major crop improvement goals like enhanced tolerance to biotic and abiotic stresses, resistance to herbicides, increased yield, and crops with superior nutritional value.

Innovative digital invasion offers next generation crop improvement in combination with next generation sequencing technologies and bioinformatics. Bioinformatic tools and databases are crucial for the efficient handling of huge datasets that produced via experimental approaches. Computational biology is now became a prerequisite factor in ncRNA research, as a means to organize, integrate, and analyse the huge amount of biological data for brilliant discoveries (Lagana *et al.*, 2015).

The present scientific study is undertaken for the evaluation of miRNA prediction tools and *in silico* analysis of micro and long non coding RNAs in sweet potato, followed by its validation with RT-qPCR. MicroRNA prediction tools were evaluated and compared to understand their performance.

REVIEW OF LITERATURE

2. REVIEW OF LITERATURE

2.1 SWEET POTATO: THE POOR MAN'S FOOD

Sweet potato is one of the most important starchy, sweet tasting, hexaploid (2n=90), dicotyledonous root vegetable belongs to *convolvulaceae* family with higher economic and nutritional value. This important staple food is native to tropical America and which suites tropical and temperate climates for its wider cultivation. Sweet potato is characterized with long and trailing stem, lobed or unlobed leaves with diverse shapes, size, and colours and mostly with high protein content, flowers appears to be funnel shaped with pale or purple colour along with coloured edible tubers. A recent research finding reveals the more power of sweet potato as an anti-diabetic, anti-oxidant and anti-proliferative agent due to the presence of highly valuable nutritional and mineral components (Jaarsveld *et al.*, 2005).

Sweet potato which is the seventh largest world food crop in developing countries. The largest sweet potato gene bank in the world is in International Potato Center (CIP) in Peru with more than 6,500 wild, traditional, and improved varieties. Asia is credited with the world's largest sweet potato-producing region, with 88.51million tonnes of annual production and China, the leading producer of sweet potatoes in the world, supplies about 76% of the world's production (Tavva and Nedunchezhiyan, 2011). In India, though sweet potato production is largely focussed on food and nutritional security, the Indian Council of Agricultural Research (ICAR) has initiated efforts to promote value addition of the crop and explore the potential for alternatives uses like animal feed and starch production.

2.1.1 Health benefits of sweet potato

Recent studies shown that sweet potato has anti-cancer effect because the tubers are enriched with antioxidants and similar valuable compounds that prevent cancer by fighting free radical damage. Boiled sweet potato is characterized with lower glycemic index that means there is no sudden rise in blood glucose level. Higher vitamin B6 content in sweet potato can helps in the breakdown of homocysteine- the substance that hardens the walls of arteries and blood vessels the ultimate cause of heart attacks.

2.1.2 Sweet potato production constraints

Viral diseases, pests and various environmental variations such as low temperature, drought and salinity generally limits the production of sweet potato worldwide (Fan *et al.,* 2012; Motsa *et al.*, 2015). Damage by insects and diseases is usually manifested in the leaves, stems and tubers.

2.1.3 Sweet potato production strategies

Last several decades practices traditional breeding techniques for producing improved cultivars in sweet potato (Ngailo *et al.,* 2013). But due to the high levels of male sterility, self and inter-specific incompatibility and hexaploid nature of the sweet potato the procedure remains very difficult (Gurmu *et al.,* 2013). With the advances in biotechnology now it is possible to genetically engineer a crop with useful traits for sustainable agriculture development. Recently, an efficient *Agrobacterium tumefaciens*-mediated transformation systems for several farmer-preferred cultivars of sweet potato were produced (Brown *et al.,* 2016).

The non coding RNA mediated technology has been extensively applied to alter the gene expression in plants with an aim to achieve modified agronomic characters. It has been used for enhancing the crop yield and productivity. It has also been applied for developing resistance against various biotic and abiotic stresses (Kamthan *et al.,* 2015). By the joint efforts of both next-generation sequencing technologies and *in silico* approaches is provided with discovery and functional characterization of novel ncRNAs in plants, especially miRNAs and lncRNAs (Liu *et al.,* 2017).

Diverse miRNA candidates are shown to be involved in various gene regulatory pathways, flowering, plant defence (immunity), drought response, abiotic stress tolerance, leaf senescence, nutrient homeostasis and many more, (Rhoades *et al.,* 2006). In plants, although a large number of lncRNA transcripts have been predicted and identified in few species, our current knowledge of their biological functions is still limited. lncRNA studies have become one of the new hotspots in current molecular biology.

2.2 MICRORNAs: THE SMALL GENOME REGULATOR

MicroRNAs (miRNA) are 18-22nt small, endogenous non coding RNA that has prominent role in many biological processes which negatively regulates the target gene (Mishra and Mukherjee, 2007). They specifically bind to the 3'UTR, exons, or 5'UTR of the target protein coding mRNA and leads to either translation repression or mRNA cleavage. miRNA mediated mRNA decay is a new type of gene regulation that is based on partial or little similarity between miRNA and its target, the miRNA removes the polyA tail from the 3' end of mRNA and decays the target. miRNAs have important roles in regulating plant growth cycle, signal transduction, organogenesis and response to environmental stimulus. A current research reveals that ancient miRNAs are characterized with high expression and evolutionary conservation whereas in case of young miRNA there are low levels of expression or their induction is controlled by certain specific conditions and also is specific for limited species (Qin *et al.*, 2014).

2.2.1 Historical approach to miRNA

By the joint efforts of a team of members led by Victor Ambros and Gary Ruvkun, the first miRNA was discovered in a nematode C*aenorhabditis elegans* in 1993. In *C. elegans,* for the transition of larval stage from L1 to L2 requires under expression of *LIN-14* along with the transcription of another gene *lin-4.* It was found that the *lin-4* gene does not synthesize any functional peptide but transcribe it into 21 and 61 nucleotides length RNAs.

In 2000, the second miRNA *let-7* which is approximately 21 nucleotide long RNA was identified in the same nematode *C. elegans* at Ruvkuns lab (Reinhart *et al.*, 2000). *let-7* was identified to controls the L4 to adult transition of larval development. They found that the *let-7* could partially suppress the expression of *lin-41* gene by interacting with 3' UTR region.

The first plant miRNA was discovered in the model plant *Arabidopsis thaliana.* These start the exploitation of more miRNA in diverse plant species. Computational analysis of miRNAs predicts the evolutionary conservation of many miRNA families (Zhang *et al.,* 2005; Zhang *et al.*, 2006b).

2.2.2 miRNA biogenesis

In plants these small transcripts are involved in almost all biological processes but are importantly considered as regulators controlling the gene expression in growth and development. miRNA genes are located in separate miRNA loci throughout the genome (Reinhart *et al.*, 2002). Most plant species are enriched with over 100 miRNA (MIR) genes. miRNA genes are transcribed through different series of steps for its final appearance as mature miRNA (Bartel, 2004).

The biogenesis starts with the transcription of miRNA gene to primary miRNA (pri-miRNA) which is characterized with more than 100 nucleotide length of sequences catalyzed by pol II enzymes (Bartel, 2004; Kurihara and Watanabe, 2004). Then the second step is the formation of precursor miRNA (Pre-miRNA), a stem loop intermediate by Dicer like 1 enzymes (DCL1) along with two other proteins, HYPONASTIC LEAVES (HYL1) and SERRATE (SE) in plants (Kurihara and Watanabe, 2004) and the similar case in animals is by Drosha RNase III endonuclease (Bartel, 2004). Then in animals, by exportin5/RanGTP translocate the pre-miRNA from nucleus to cytoplasm in order to generate 21 nt miRNA duplex and mature miRNA. This step is regulated by another enzyme called Dicer (Bartel, 2004). By DCL1 plant miRNA are cleaved into miRNA : miRNA*duplex in the nucleus, then its translocation to cytoplasm by HASTY (HST) the plant orthologue of exportin5, which is a member of nucleocytoplasmic transporter family of proteins. Helicase another important enzyme unbound the plant miRNA and animal miRNA in the cytoplasm into single stranded mature miRNAs (Bartel, 2004) and the guide strand is incorporated into ribonucleoprotein complex simply known as RISC along with Argonaute (AGO) proteins and thereby mediate the targeted gene expression either through mRNA cleavage or translational repression.

2.2.3 Molecular functions of miRNA

miRNAs is involved in the regulation of target gene expression through either mRNA cleavage or translation inhibition which is based on the complementarity between miRNAs and its specific targets (Bartel, 2004). Target

mRNA cleavage is the result of perfect pairing between miRNA and its target which mostly shown by plants and imperfect pairing results in translation repression which follows in animals (Brodersen *et al,* 2008). Plants are characterized with miRNAs guide sequence specific DNA methylation through a phenomenon referred to as miRNA-directed DNA methylation (miRdDM) (Bao *et al.,* 2004).

2.2.4 Versatile functions of plant microRNA

2.2.4.1 Role of miRNAs in plant development

miRNAs, the most powerful gene regulators shown to play an important role in the regulation of gene expression. There are evidences for the involvement of miRNAs in many biological processes like growth and development, stress responses and tumorigenesis in animals (Ambros, 2003). Although plant miRNA expression platform are behind animals there are certain experimental proof for its importance in the regulation of plant growth cycle, environmental adaptation, signalling cascades, maintaining homoeostasis (Jones-Rhoades *et al., 2002*). Now miRNAs are shown to be involved in the regulation of root initiation and development (Montgomery *et al.*, 2008; Marin *et al.*, 2010), leaf morphogenesis (Mallory *et al.*, 2004), reproductive growth (Yang *et al .,* 2007). A major category of miRNA target genes are transcription factors or other regulatory proteins that functions in plant development or signal transduction.

2.2.4.2 Role of plant miRNAs in biotic stress

Plants are sessile in nature therefore; they need evolved subtle mechanisms and economical methods to withstand the environmental stresses that are caused by various biotic or abiotic factors for better yield worldwide. Advanced molecular biology research platform provided with a new regulator molecules called non coding RNAs mostly microRNAs to regulate stress responses in plants via target mRNA cleavage or the inhibition of translatory mechanisms. miRNAs either its under or over expression regulates the genes that is involved in stress responses post transcriptionally. miR393 was the first miRNA identified in *Arabidopsis* which negatively regulates auxin signalling cascade proves to be important in plant antibacterial pattern triggered immunity (PTI)

(Navarro *et al.*, 2006). In wheat, powdery mildew disease caused by fungus *Blumeria graminis f.sp.tritici* (Bgt) have showed under expression of miR156, miR159, miR164, miR171 and miR396, and the over expression of miR393, miR444 and miR827 (Xin *et al.*, 2010). In rice, miRNAs (miR160, miR166, miR167, miR171, and miR396 family of miRNAs) had been found to be gathered in response to rice stripe virus contamination (Du *et al.*, 2011). Nematodes are now becoming a major challenge in successful crop production worldwide. miR159 family were shown to be involved in plant response to root knot nematodes *(RKN) Meloidogyne incognita* infection (Medina *et al., 2017).*

2.2.4.3 Role of plant miRNAs in abiotic stress

In depth case studies demonstrated that the expression pattern of some miRNAs were altered in response to different environmental changes mostly abiotic stresses like drought, soil salinity, nutrient deprivation etc. during the plant growth cycle. These stressors may have a great negative impact on various phases of plant growth, its yield and respective production quality (Shriram *et al.,* 2016). The screening of small RNAs library isolated from *Arabidopsis* (Liu *et al.,* 2008), rice (Zhou *et al.*, 2010), and sugar cane (Gentile *et al.*, 2015) leads to the identification and over expression of several drought responsive miRNAs. Osa-miR820 plays a role in salt, high temperature, and drought stress responses were recently reported (Sharma *et al.,* 2015a). Aluminium stress contributes the differential expression of 23 miRNAs in *M. truncatula* seedlings (Chen *et al.,* 2012). In a study identified the expression pattern of several UV responsive miRNA in wheat and identified *miR159, miR167a*, and *miR171* were over expressed and miR156, miR164, miR395 were down regulated in accordance with UV-B treatment (Wang *et al* ., 2013) .

2.2.5 Identification of microRNA

In order to explore more about novel miRNAs and its regulatory mechanisms we have to identify it precisely through efficient methodologies. The interaction between potential miRNAs and its corresponding targets presents valuable information's about the accurate functions of miRNA. Generally two category of approaches are there for miRNA detection, one is the experimental

28

point of view which deal with the extraction of small RNA candidates followed by its *in silico* analysis and the other one is the computational approach which at first predict the mature miRNAs using available sequences and then go for experimental validation.

Earlier miRNA detection focuses mainly on direct cloning, genetic screening, and *in silico* analysis (Schommer *et al.*, 2012). miRNAs were firstly identified through genetic screening (Lee *et al.*, 1993; Wightman *et al.*, 1993). The success rate of identification of miRNA through genetic screening is less compared to other approaches due to the smaller size of the small RNAs, expensive nature and time consuming.

It was in A*rabidopsis thaliana,* the model plant where the first miRNAs (Reinhart *et al.*, 2002). This approach is characterized by direct cloning after isolation of small RNAs populations (Lu *et al.*, 2005a).

In silico analysis is widely used to predict thousands of novel miRNAs in different plant species based on the criteria that plant miRNAs have perfect complementarity to its target of interest and its nature of evolutionary conservation (Sunkar *et al.*, 2005). There are lots of computational tools available for the efficient prediction of miRNAs based on diverse algorithms (Lim *et al.*, 2003a, Lim *et al.*, 2003b, Wang *et al.*, 2005). Among the computational methods mostly preferred one is the homology based comparative genomic method, which mostly exploit the evolutionarily conserved nature of miRNAs in various species. EST analysis is credited for identifying conserved miRNAs, for species whose genome is less studied (Zhang *et al.*, 2005).

2.2.5 *in silico* biology in plant microRNA genomics

Computational biology is an indispensable tool in the microRNA research world. After cloning and sequencing, the miRNA sequences are taken for computational analysis. The importance bioinformatics in miRnomics is due to the challenges faced by cloning and genetic screening strategies which include, the small size of miRNA, less expressed miRNAs identification difficulty, less stability of RNA molecules, sequence composition or post translational modification, tissue specificity (Prabhu and Mandal, 2010) and also the

evolutionary conservation of microRNAs. Low cost, high accuracy and new technologies made computational strategies more efficient and reliable (Wegener *et al.*, 2013). EST based homology search strategy is one of the most commonly used method now a days for miRNA prediction. Different homologous miRNAs were identified in different plant species based on EST analysis, which includes maize (Zhang *et al.*, 2006b), tobacco (Frazier *et al.*, 2010), potato (Xie *et al.*, 2011) soybean (Zhang *et al.*, 2008), wheat (Han *et al.*, 2013), cotton (Zhang *et al.*, 2007). Northern blotting or by quantitative or semi-quantitative real-time PCR based assays will be used for the validation of identified miRNAs expression (Zhang and Wang, 2015).

2.2.6 Plant miRNA databases

2.2.6.1 *miRBase* (*www.mirbase.org/*)

A web platform for all miRNA sequence along with annotation. The database is available to browse and search by using sequences and key words for the mature and hairpin sequences of microRNA for research studies. miRBase 22 release contains 38589 entries.

2.2.6.2 *PmiRExAT*(*http://pmirexat.nabi.res.in*)

It's an open and easy accessible online database for gathering miRNA expression data which is now enriched with miRNA expression profile and query tool for 1859 wheat, 2330 rice and 283 maize miRNA.

2.2.6.3 *MiRNEST 2.0* (*mirnest.amu.edu.pl/*)

It's an integrative repository of animal, plants and viral computationally analysed microRNA data. It is also depends upon other miRNA databases and publications for information. The database provided the user with data regarding sequences, polymorphism, expression, promoters, miRNA gene structures, degradome data and more.

2.2.6.4 *PMTED* (*pmted.agrinome.org/*)

It is plant microRNA target expression database. The user friendly database which provides the microRNA target gene expression data thereby allowing the users to predict the functions of the microRNA. The database currently comprises of 12 plant species with 1897 miRNAs and 5449 target gene data.

30

***2.2.6.5 PNRD** (http://structuralbiology.cau.edu.cn/PNRD)*

It is plant non coding RNA database, the new edition of PMRD (plant microRNA database) just for plant species with 25739 entries of 11 differing types of ncRNAs from a 150 plant species. Targets of miRNAs were extended to 178138 pairs in 46 species.

***2.2.6.6 PASmiR** (http://hi.ustc.edu.cn:8080/PASmiR)*

A literature curated and web accessible database that consist of microRNA involved the regulation of abiotic stress response in plants. miRNA- stress regulatory gene is easily accessible for the users. Currently the database is provided with data from 200 published studies.

Table 1. Bioinformatic tools in plant miRNA research.

Tools	Prediction features	URL	Reference
miRBase	Database of published miRNAs and annotation	*http://www.mirbase.org/*	(Kosomara and Jones, 2013).
miRPlant	Novel plant miRNA Identification, known miRNA identification, precursor prediction	*http://www.australianprostatecentre.org/research/software/mirplant*	(An *et al.*, 2014).
NOVOMIR	Sequence, structure, miRNA/miRNA* position prediction	*wwww.biophys.uniduesseldorf.de/novomir*	(Teune and Steger, 2010).
C-mii	Sequence, structure, target prediction, GO annotation	*www3a.biotec.or.th/c-mii/*	(Numnark *et al.*, 2012).
psRNATarget	Reverse complementary matching between small RNA and target transcript, and target site accessibility evaluation	*http://plantgrn.noble.org/psRNATarget/.*	(Dai and Zhao, 2011).
miRanda	seed match, conservation, free energy, G:U pairs allowed in the seed	*http://www.microrna.org/*	(Archak and Nagaraju, 2007).
Targetscan	seed match, conservation, free energy, site accessibility, target-site abundance	*www.targetscan.org/*	(Agarwal *et al.*, 2015).
RNAmFold	Secondary structure prediction	*http://mfold.bioinfo.rpi.edu*	(Zuker, 2003).

2.3 lncRNAs: THE LONG GENOME REGULATOR

Deep transcriptome analyses have showed that up to 90 % of eukaryotic genomes are transcribed, but the fact is that only 1–2 % of the genome is functionally stable. This data suggests the existence of a plethora of non coding transcripts in the eukaryotic genome, namely non-coding RNAs (ncRNAs). Among them short ncRNA (20-30nt length) mediated gene regulation is well understood but the functional characterization of long non coding RNA (200nt) is still in the beginning stage (Kimura *et al.,* 2015). Research results reported that more than thousands of lncRNA transcripts were identified that did not appear to be coding for proteins (Cabili *et al.,* 2011).

lncRNAs are greater than 200 nt length RNA candidates belongs to non coding RNA family characterized with inability to synthesize functional polypeptides but are polyadenylated, spliced, and mostly localized in the nucleus (Wierzbicki *et al.*, 2012). Like protein coding mRNA molecules lncRNAs are transcribed by RNA polymerase II or III and by IV or V in case of plants. This large and functionally heterogeneous group are generally expressed at low levels and also the lncRNA sequences are not well conserved between species. Experimental evidence suggests that lncRNAs are potent gene regulators of gene expression and are involved in multiple levels of the gene regulation in animals (Zhu and Wang, 2012). When compared to animals, a few lncRNAs have been identified and functionally characterized in plants (Amor *et al.,* 2009; Heo and Sung, 2011; Swiezewski *et al.,* 2009).

2.3.1 Historical approach to lncRNA

First lncRNAs, *Xist* and *H19* were discovered during 1980s but remained exceptions until the early 2000s. The completion of human genome project during 2001 published a surprising evidence of the invasion of non coding nature of the genome (Rao, 2017). Further studies have shown that many non coding RNAs were pervasively transcribed from the genome. These start the journey of lncRNAs and were shown to be act as gene regulator in many biological processes in diverse species. The advances in transcriptome analysis is accompanying with the rapid exponential increase of novel lncRNAs.

33

2.3.2 lncRNA biogenesis

lncRNAs are transcribed from non coding region in the genomic DNA by the combined action of RNA pol II/III in animals and IV/V in case of plants and specific transcription factors, mediator complex, histone modification complex and transcription elongation factor complex (Di *et al.,* 2014; Guttman *et al.,* 2009; Heo and Sung, 2011; Wang *et al.,* 2014e). Later the individual lncRNA is subjected to 5' capping, pre-lncRNA splicing and polyadenylation similar to protein coding genes (Guttman *et al.,* 2009). Finally the formation of a stable secondary and tertiary structure occurs, which confer the individual lncRNA its unique functional roles.

2.3.3 Classification of lncRNAs

The candidate lncRNA molecule is categorized based upon the genomic location and context, the effects exerted on DNA sequences, functional mechanisms and targeting mechanisms (Ma *et al.,* 2013).

2.3.3.1 Genomic location and context

According to genomic location and context lncRNAs are positioned by intergenic (lincRNA) and intronic, sense and antisense finally, bidirectional lncRNAs. This lincRNA category is the largest among ncRNA family which are 3' polyadenylated, 5'capped and transcribed from the non coding region of the genome between the protein coding genes. Intronic RNA is transcribed from the intron parts of protein coding genes so their expression pattern is somewhat similar to protein coding genes. The sense and antisense classes are distinguished by their transcription from both sense and antisense strand of the protein coding gene. The final class bidirectional lncRNA candidates are transcribed from the opposite strand of the protein-coding gene whose transcription initiation site is less than 1000 base pair away (Adossa, 2016).

2.3.3.2 Effects exerted on DNA sequences

cis and trans lncRNAs are the two classes base on the effects exerted on DNA sequences in that cis lncRNAs regulates gene expression in the close proximity while the trans-lncRNAs involved in remote regulation of gene expression (Ma *et al.,* 2013).

34

2.3.3.3 Cellular molecular mechanism

lncRNAs can also be classified into those that affect the transcriptional regulation, post-transcriptional regulation and other functions like protein localization, telomerase replication and RNA interference in accordance with cellular molecular mechanism (Ma *et al.*, 2013).

2.3.3.4 Molecular mechanism

lncRNAs can act as a signal, decoy, guide and scaffold lncRNAs based on their molecular mechanisms. Act as a signalling molecule as the transcription of lncRNA takes place at the specific time and place to integrate the developmental cues, interpret cellular context, or respond to diverse stimuli. After transcription, decoy lncRNAs regulate gene expression by binding and titrating away the transcription factors, chromatin modifiers and other regulatory factors from its target without exerting any additional function. lncRNAs can serves as a guide for the localization of ribonucleoprotein complexes to a specific target. lncRNAs serve as molecular platforms upon which relevant molecular components are assembled to form new functions.

2.3.4 Versatile functions of plant lncRNA

2.3.4.1 Role of plant lncRNAs in phosphate signalling

Phosphate is an important mineral nutrient for the plant growth and development. miR399 were identified to participate in phosphate homoeostasis regulatory network pathways in plants by targeting PHO2 gene which codes for E2 ubiquitin conjugase-related enzyme (UBC24). During phosphate starved conditions, this miRNAs expression level will increase and target cleavage occurs that in turn leads to the over expression of phosphate transporter genes, Pht1; 8 and Pht1; 9 resulting in higher phosphate uptake by plants. During phosphate starved condition another candidate called *induced by phosphate starvation 1 (IPS1)* a member of TPS1/Mt4 gene family is expressed without any protein coding function, that partially complement to miR399 and acts as an endogenous target mimic of miR399 there by blocking the activity of mir399 (Zorrilla *et al.*, 2007).

35

2.3.4.2 Role plant lncRNA in flowering

lncRNA were shown to epigenetically regulate flowering time in *Arabidopsis* (Heo and Sung, 2011). *FLOWERING LOCUS C* gene (*FLC*) is an essential regulator of flowering time in *Arabidopsis* which inhibit flowering at cold temperature. Studies show that two lncRNAs named COLD INDUCED LONG ANTISENSE INTRAGENIC RNAs (COOLAIR) and COLD ASSISTED INTRONIC NONCODING RNA (COLDAIR), transcribed from the FLC locus play a role in the repression of FLC expression by recruiting PHD-PRC2 complex to permit histone modifications of FLC via epigenetic regulation (Heo and Sung, 2011).

2.3.4.3 Role of plant lncRNAs in lateral root formation

lncRNAs are involved in the formation of lateral roots in plants. During alternative splicing in *Arabidopsis* lncRNAs acts as an alternative splicing competitior (ASCO) which forms a complex with nuclear speckle RNA-binding protein (NSR) called alternative splicing regulatory module. Development of lateral roots occurs due to the expression of AtNSR in primary and lateral root meristems (Bardou *et al.,* 2014; Kornblihtt, 2014; Liu *et al.*, 2015).

2.3.4.4 Role of lncRNA in plant immunity

lncRNA is expressed by translation elongation factor Tu (elf18), a pathogen associated factor that induces defence responses in *Arabidopsis thaliana*. Further analysis on this by Seo *et al.* (2017) caught ELF18 INDUCED LONG-NONCODING RNA1 (ELENA1) a 589-nucleotide length lncRNA. ELENA1 over expressed in response to elf18 and flagellin (flg22), but not in response to the defence hormone salicylic acid. Mutants affecting the receptors for elf18 and flg22 did not show a higher expression of ELENA1 levels in accordance with these inducers. Transgenic plants were produced using an artificial miRNA with reduced ELENA1 levels to examine the actual function of ELENA1. Further results shown that plants with lower levels of ELENA1 showed higher susceptibility to bacterial pathogen *Pseudomonas syringae* pv tomato DC3000 and over expression of ELENA1 is credited with disease resistance (Seo *et al.,* 2017).

36

2.3.5 Identification of long non coding RNA

Both experimental and computational strategies are available for the efficient identification of lncRNAs.

2.3.5.1 Experimental methods

2.3.5.1.1 Tilling array technology

Tiling arrays are a type of microarray chips which involves the hybridization of labelled DNA or RNA target molecules to probes fixed onto a solid surface. Tiling arrays is characterized with the probing of sequences that exist in a contiguous region. This is useful for efficient sequencing of the region. The constraints that limits the use of this method is dependency on existing knowledge of genome sequences, they also requirement of a large amount of starting material (RNA) to for conducting the experiment.

2.3.5.1.2 Serial analysis of gene expression (SAGE)

SAGE is one of the efficient methods for the identification of lncRNAs, which is based on the generation of short stretches of cDNA sequence tags containing restriction enzyme sites at the 3′end. The tags are concatenated followed by Sanger sequencing.

2.3.5.1.3 RNA-Sequencing

With the advances of high throughput sequencing, RNA-seq became most widely used technique in the lncRNA biology. It is based on the conversion of transcripts into a pool of cDNAs that will constitute the sequencing library. The library is prepared by RNA fragmentation, adapter ligation, cDNA synthesis, size selection and limited cycles of amplification. In rice, through whole transcriptome strand specific RNA sequencing (ssRNA-seq), 2224 lncRNAs involved in the reproductive process were verified (Li *et al.*, 2014).

2.3.5.2 in silico biology in lncRNA genomics

The main challenge faced by lncRNA research in plants is the lack of best prediction method. Therefore, there is a great need for the precise and efficient computational methods to predict novel lncRNAs in plants for further research.

37

2.3.5.2.1 ORF Length Strategy

lncRNAs are characterized with 200nt length RNA molecule with or without protein coding potential. ORF in lncRNAs, if there exists; it is short in length, not more than 100 codon. In this method, transcripts without open reading frame (ORF) or have short ORF are computationally identified.

2.3.5.2.2 Sequence and secondary structure conservation strategy

Less conserved nature of lncRNA is more prone to mutation. This mutation is measured by codon substitution frequency (CSF). So it could be used as a way to predict lncRNAs.

2.3.5.2.3 Machine learning strategy

Different tools are available for the easier prediction of lncRNAs based on diverse machine learning algorithm like RNAplonc (Negri *et al.*, 2017). But compared to animals, the availability of tools for plant lncRNA prediction is limited. Most of the tools utilize the coding potential feature of a sequence to distinguish lncRNAs from mRNAs.

2.3.6 Plant lncRNA databases

In depth identification and characterization of plant lncRNAs using accessible computational aids generates more data that in turn may leads to the creation of repositories for arranging plant lncRNA related information. Functional annotations of candidate lncRNAs become easier through the available databases.

2.3.6.1 NONCODE v4

It is an integrated eukaryotic non coding RNA database except tRNAs and rRNAs with data provided by GenBank, specialized databases, and the literature. All the information of lncRNAs like location, length, strand, class, isoforms, coding potential, sequence and its conservation, disease relation and the related research article are available to browse, search, and download in a user friendly manner for convenient use. It does not focuses only on plants, provided only *Arabidopsis* lncRNA information.

38

2.3.6.2 LncRNAdb

lncRNAdb are manually curated lncRNA repository from literature-supported evidence. The updated version lncRNAdb v2.0 comprises nearly 283 entries spanning 71 different organisms. It is organised for similarity search to any known lncRNA and can retrieve all the basic details regarding the genomic features, expression pattern, and functional annotation.

2.3.6.3 PlantNATsdb

PlantNATsdb or plant NAT database is a natural antisense transcript database which is an identified isoform of long non coding RNAs for the functional annotation of all NATs in plants. The database comprises approximately, 2138,498 NATs from 70 plant species along with GO (gene ontology) annotation and high-throughput small RNA sequencing data. It is a user-friendly platform.

2.3.6.4 PLNlncRbase

PLNlncRbase is an easy-to-use resource that provides information exclusively for plant lncRNAs. In fact, it enables a user to browse through the repository based on diverse plant species and/or lncRNA category (biotypes such as NATs, intergenic, intronic lncRNAs etc.). Currently, 1187 plant lncRNAs in 43 plant species have been manually curated from over 200 published studies.

Table 2. Bioinformatic tools in lncRNA research

Tools	Prediction features	URL	Reference
AnnoLnc	Full spectrum of annotations covering genomic location, RNA secondary structure, expression, transcriptional regulation, miRNA interaction, protein interaction, genetic association and evolution	*http://annolnc.cbi.pku.edu.cn*	(Hou *et al.*, 2016).
CPC	ORF prediction quality	*http://cpc.cbi.pku.edu.cn.*	(Kong *et al.*, 2007).
lncFunTK	Functional InformationScore (FIS)	*http://sunlab.cpy.cuhk.edu.hk/lncfuntk*	(Zhou *et al.*, 2018).
RNAplonc	Pattern recognition analysis	*rnaplonc.cp.utfpr.edu.br/*	(Negri *et al.*, 2017).
LncRNApred	Random forest classification	*http://mm20132014.wicp.net:57203/LncRNApred/home.jsp.*	(Pian *et al.*, 2016).
lncRNA-screen	Quality assessment, transcript filtration, novel lncRNA identification, coding potential estimation, expression level quantification, histone mark enrichment profile integration, differential expression analysis, annotation	*https://github.com/NYU-BFX/lncRNA-screen*	(Gong *et al.*, 2017).
FEELnc	Multi k-mer frequencies and relaxed open reading frames	*https://github.com/tderrien/FEELnc.*	(Wucher *et al.*, 2017).
Lncident	ORF prediction quality	*http://csbl.bmb.uga.edu/mirrors/JLU/Lncident/index.php*	(Han *et al.*, 2016).

40

2.4 VALIDATION OF ncRNAS BY QUANTITATIVE REAL-TIME REVERSE TRANSCRIPTION PCR (qRT-PCR)

Non coding RNA candidate genes possess specific functions at RNA level in the cell even if it is ineligible to synthesize stable proteins (Huttenhofer and Vogel, 2006). Deep transcriptome analysis in association with computational algorithm results in many ncRNAs bearing important functions in different species.

Validation of identified ncRNAs is essential for further analysis. Different methodologies are available like qPCR, Northern blotting, micro-array analysis etc. for detection and quantitation. Due to high sensitivity and efficiency qPCR is the widely adopted one for expression analysis.

Real-time PCR (RT-PCR) is also known as quantitative PCR or qPCR. In qPCR amplification of cDNA is detected in real time as PCR is in progress by the use of fluorescent reporter for RNA expression studies. Fluorescent probes mostly used are sequence-specific TaqMan probe and generic non-sequence-specific double-stranded DNA-binding dye such as SYBR green. The principle behind this technique is that the intensity of fluorescence emitted by the probe at each cycle is directly proportional to the template quantity.

MATERIALS AND METHODS

3. MATERIALS AND METHODS

The study entitled "Evaluation of miRNA prediction tools and *in silico* analysis of micro and long non coding RNAs in sweet potato (*Ipomoea batatas* L.)" was conducted at ICAR- CTCRI during October 2017 to August 2018. This chapter is presented with the detailed information about experimental materials and methodologies adopted for handling various experiments.

3.1 REVIEW OF miRNA PREDICTION TOOLS AND COMPARATIVE EVALUATION

Available miRNA identification and target prediction tools were reviewed. The most widely used plant specific miRNA identification tools like NOVOMIR *(wwww.biophys.uniduesseldorf.de/novomir) and* miRPlant *(www.australianprostatecentre.org/research/software/mirplant)* along with target prediction tools like psRNAtarget *(https://plantgrn.noble.org/psRNATarget/),* miRanda *(http://www.microrna.org/)* were selected to compare and evaluate their potential in miRNA research. A noted fact is that the majority of tools widely exploited the whole genome of *Arabidopsis thaliana* for its creation.

3.1.1 Datasets used

3.1.1.1 For miRNA identification

- Positive datasets: Pre-miRNA sequence of *Arabidopsis thaliana* from miRBase (326 sequences)
- Negative datasets: pseudomiRNA sequences created using *Arabidopsis* coding sequences (CDS) (328 sequences)

3.1.1.2 For miRNA target prediction

- Mature miRNA sequences from miRBase (15 sequences)
- Positive datasets: Experimentally validated *Arabidopsis thaliana* miRNA-mRNA interaction sequences from miRTarBase (35 sequences)
- Negative datasets: Experimentally validated negative sequences used by (Heikham and Shanker, 2010) (36 sequences)

3.1.1.3 Creation of pseudo pre-miRNAs datasets 43

- Chop the *Arabidopsis* CDS sequences in FASTA file
- Construct shorter sequences library from CDS.
- Fold and filter shorted sequences by UNAFold (Hairpinstructure MFE<10

3.1.2 Evaluation of miRNA identification tools

Both positive and negative miRNA sequence datasets of *Arabidopsis thaliana* were given as input to both NOVOMIR and miRPlant miRNA identification tool to compare and evaluate their performance in miRNA research.

NOVOMIR is a program for the identification of pre-miRNA sequences in plant genomes which is available at *(wwww.biophys.uniduesseldorf.de/novomir)* (Teune and Steger, 2010). It uses a series of filter steps and a statistical model to discriminate a pre-miRNA from other RNAs and does rely neither on prior knowledge of a miRNA target nor on comparative genomics.

The parameters used in NOVOMIR miRNA prediction tool is as follows;

-f Sequence-File --- Sequences must be on one single line

-g genome-wide search for pot. candidates / specify a file-ending

-e Energy-Value for RNAshapes

-m minimum length of stem-loop-Region

-t threshold for prediction

-T 2nd threshold for prediction

-w window slide filter threshold 0<w<1

-c correlation btw seq-length and hp-length

-C min number of consecutive basepairs for a hairpin

-s normalized nrg threshold

-h print this Usage (boolean)

-D Debugging Symbol (boolean)

-R use RNAshapes version 3.1.4 (default: 2.1.6)

-l long output (short output is default)

-v show version

Only option -f is necessary, all other options are optional.

miRPlant a miRNA identification program is available at (*www.australianprostatecentre.org/research/software/mirplant)*. The objective of miRPlant program is to predict novel plant miRNA, while providing a user-friendly interface with improved accuracy of prediction. miRPlant does not require any third party tools such as mapping or RNA secondary structure prediction tools. All the parameters were set as default for the prediction of miRNA such as precursor length: 200, min loop length: 20, flanking length: 10, miR length: 18-23, min phred: 20, max multi map: 101, min reads: 5.

The output was analysed to calculate the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) (An *et al.*, 2014).

3.1.3 Evaluation of miRNA target prediction tools

The input sequences such as *Arabidopsis* mature miRNA sequences, positive and negative datasets were given to selected miRNA- taregt prediction tools like psRNATarget *(https://plantgrn.noble.org/psRNATarget/) and* miRanda *(http://www.microrna.org/)* and go for comparative evaluation.

All the parameters in psRNAtarget were kept at default for miRNA-mRNA interaction analysis. That is # of top targets: 200, Expectation: 5, Penalty for G: U pair: 0.5, Penalty for other mismatches: 1, Extra weight in seed region: 1.5, Seed region: 2-13, # of mismatches allowed in seed region: 2, Translation inhibition range: 10-11nt, Flank length: 17/13nt in up/downstream (Dai and Zhao, 2011).

miRanda which is an algorithm for the prediction of potential microRNA target sites in the FASTA formatted input genomic sequences (Betel *et al.*, 2008). The algorithm which proceeds with dynamic-programming local alignment and thermodynamics to identify the miRNA-mRNA targets.

Here also all the parameters for the prediction of potential microRNA target sites by miRanda were kept at default score> 95; energy <-20 kcalmol-1 (Archak and Nagaraju, 2007).

The output was analysed to calculate the number of true positives (TP), true negatives (TN), false positives (FP), false negatives (FN).

45

- True positives (TP) are the number of experimentally supported datasets that are predicted by the program.
- False negatives (FN) are the number of experimentally supported datasets that are not predicted by the program.
- False positives (FP) are the number of all negatives that are predicted by the program.
- True negatives (TN) are the number of all negatives that are not predicted by the program.

For further validation several statistical parameters, viz., Sensitivity (Se), Specificity (Sp) and Accuracy (Acc), a summary statistic: Matthews correlation coefficient (MCC) and Positive predictive value (PPV) were used for the evaluation of the performance of selected prediction tools. These parameters are based on TP, FN, TN and FP and are calculated using the given equations:

- Accuracy (Acc) = (TP+TN)/(TP+TN+FP+FN)*100
- Specificity (Sp) = TN/(TN+FP)*100
- Sensitivity (Se) / Recall = TP/(TP+FN)*100
- MCC= ((TP*TN)-(FP*FN))/ ((TP+FP)*(TN+FN)*(TP+FN)*(TN+FP))1/2
- Positive Predictive value (PPV) / Precision = TP/ (TP+FP)*100

3.2 IDENTIFICATION OF SWEET POTATO miRNA

46

3.2.1 Mining of sweet potato EST dataset from NCBI

All 37698 *Ipomoea batatas* ESTs were downloaded from NCBI *(https://www.ncbi.nlm.gov/)*. 'Ipomoea batatas', as search key for *Ipomoea batatas* ESTs dated on November 2, 2017 and create the sequence file in FASTA format.

3.2.2 Treating EST datasets for assembly

CAP3 (http://seq.cs.iastate.edu/CAP3.html) sequence assembly software is performed for the assembly of *Ipomoea batatas* ESTs to remove the redundancy in the dataset. The EST sequences are treated to overlapping sequences known as contigs and non-overlapping sequences known as singleton.

➢ **Usage: cap3 File_of_reads [options]**

If the file of reads is named 'xyz', then the file of quality values must be named 'xyz.qual' and the file of constraints named 'xyz.con'.

Cap3 options (default values):

-a N specify band expansion size N > 10 (20)

-b N specify base quality cutoff for differences N > 15 (20)

-c N specify base quality cutoff for clipping N > 5 (12)

-d N specify max qscore sum at differences N > 100 (250)

-e N specify extra number of differences N > 10 (20)

-f N specify max gap length in any overlap N > 10 (300)

-g N specify gap penalty factor N > 0 (6)

-h N specify max overhang percent length N > 5 (20)

-m N specify match score factor N > 0 (2)

-n N specify mismatch score factor N < 0 (-5)

-o N specify overlap length cutoff > 20 (30)

-p N specify overlap percent identity cutoff N > 65 (75)

-r N specify reverse orientation value N >= 0 (1)

-s N specify overlap similarity score cutoff N > 100 (500)

-t N specify max number of word occurrences N > 30 (500)

-u N specify min number of constraints for correction N> 0 (4)

-v N specify min number of constraints for linking N > 0 (2)

-w N specify file name for clipping information (none)

-x N specify prefix string for output file names (cap)

-y N specify clipping range N > 5 (250)

-z N specify min no. of good reads at clip pos N > 0 (2)

If no quality file is given, then a default quality value of 10 is used for each base.

3.2.3 Sweet potato pre-miRNA prediction by NOVOMIR

NOVOMIR is a pre-miRNA and miRNA/miRNA* sequence prediction tool in plant genome, which is written in PERL and is tested under Linux platform. The tool which uses a series of filter steps and different statistical analysis to find out potential pre-miRNA sequences with a sensitivity of about 96.01 and a specificity of about 98.47 from all other kinds of RNA classes without any comparative genomic approaches. The program which is available at *(wwww.biophys.uniduesseldorf.de/novomir).*

3.2.3.1 Prediction Features:

- Sequence
- Structure
- miRNA/miRNA* position prediction

➢ ***Usage: perl novomir.pl <options>***

Potential pre-miRNAs of *Ipomoea batatas* are identified and selected by NOVOMIR using available contigs and singletons sequences through heuristic filters and paired HMM Algorithm.

3.2.4 Identification of potential miRNA

miRBase: the home of all identified miRNAs is a searchable online resource for published microRNA sequences and annotation. miRBase release 22 comprises of 271 species with 38589 entries representing hair pin structures and corresponding 48885 mature miRNA sequences. All the plant miRNAs are evolutionarily conserved. So homology based prediction is an efficient strategy for miRNA prediction. All the pre-miRNAs identified from both contigs and singletons by novomir is subjected to BLAST similarity comparison against *viridiplantae* with default parameters in miRBase which is an online repository of published miRNA sequences.

48

3.2.4.1 Selection criteria for potential miRNAs

- Length of mature miRNAs should be 18-22nt
- Score should be >70
- Allowed mismatch value is 2-4

3.2.4.2 Default parameters for similarity search in miRBase

- Search sequences: mature miRNAs
- Search method: BLASTN
- E-value cutoff: 10
- Max. No. of hits: 100

All the information regarding known miRNAs are freely available to the users through the web platform at *http://www.mirbase.org/* and in a flat file form *ftp://mirbase.org/pub/mirbase/.*

3.2.5 Secondary structure representation of miRNA using RNA::HairpinFigure

Secondary structure representation of miRNA was performed using RNA::HairpinFigure. RNA::Hairpin Figure uses RNA sequences to draw hairpin-like text figure along with secondary structure in dot-bracket notation. This program is freely available and is written in PERL language. Pre-miRNA sequence along with the dot bracket notation from NOVOMIR is used as input for RNA::HairpinFigure.

- ***Usage: perl example.pl***

3.2.6 Phylogenetic analysis of potential miRNAs in sweet potato

The pre-miRNA sequences of the known miRNAs from miRBase, the known miRNA registry, followed by its alignment with the precursors of newly identified miRNAs of sweet potato by Environment for Tree Exploration (ETE) available at GenomeNet (*https://www.genome.jp/tools/ete/).*

3.2.7 Potential target prediction of miRNA by psRNATarget

psRNATarget (a plant small RNA target analysis server) tool *(http://plantgrn.noble.org/psRNATarget/)* is widely used for the prediction of potential miRNA targets. The miRNAs target genes could be identified based on their perfect or nearly perfect complementarity between them and their target

genes through homology algorithm. Many studies reveals that there exists a perfect binding between plant miRNAs and its target genes than animal miRNAs and later the action rely on either the cleavage of mRNA or the inhibition of translation mechanisms.

3.2.7.1 Prediction features

- Reverse complementary matching between small RNA and target transcript
- Target-site accessibility evaluation by calculating unpaired energy (UPE)

3.2.7.2 psRNATarget scoring schema V2(2017 release)

- The seed region has been extended to 2-13 bp
- The maximum number of mismatches (excluding G-U) allowed in seed region has been restricted to two.
- The analysis of target accessibility has been disabling since its value didn't change the final output.
- The default maximum expectation is set to 5.0

The targets were predicted by comparing newly identified miRNAs against ESTs of *Ipomoea batatas*. All the parameters are set as default except allowed e: value which is taken as 3. Potential targets of putative miRNAs of sweet potato are predicted.

3.2.7.3 Default parameters

Penalty for G: U pair: 0.5; Penalty for other mismatches: 1; Extra weight in seed region: 1.5; Seed region: 2-13nt; # of mismatches allowed in seed region: 2; HSP size: 19; Penalty for opening gap: 2; Penalty for extending gap: 0.5; Translation inhibition range: 10-11nt

3.2.8 Functional annotation of targets using BLASTX and BLAST2GO

Once the potential targets of putative miRNAs were identified, BLASTX search against the NCBI protein database *(https://www.ncbi.nlm.gov/)* and BLAST2GO analysis was performed for the functional characterization of predicted target genes.

50

3.3 SWEET POTATO lncRNA PREDICTION BY RNAplonc

RNAplonc *(https://github.com/TatianneNegri/RNAplonc/)* is a powerful tool to predict lncRNAs from a non redundant transcript dataset in plants. The tool which rely on almost 16 features that are robustly selected from about 5468 features to discriminate potential lncRNAs from protein coding genes in a user friendly manner with few computational resources. RNAplonc robustly and accurately selects features using the best and computationally accessible method called REPTree in weka platform.

3.3.1 Prediction features:

(1) GC content, (2) AACG, (3) CCGT, (4) CGCA, (5) CGCT, (6) CGGG, (7) CGTA, (8) TACC, (9) TACG, (10) TCCG, (11) TCGC, (12) Sequence length, (13) score, (14) cdsStop, (15) cdsSizes and (16) cdsPercent

3.3.2 RNAplonc pipeline

3.3.2.1 Dataset.fasta

All 37698 *Ipomoea batatas* ESTs were downloaded from NCBI (https://www.ncbi.nlm.gov/) in FASTA format and taken as input dataset for lncRNA prediction.

3.3.2.2 200nt.pl

LncRNA molecules are characterized with sequences greater than 200nt length. So 200nt.pl was performed for the removal of sequences smaller than 200 nucleotides to findout potential lncRNAs.

- ***Usage: perl 200nt file.fasta***

3.3.2.3 Cd-hit-est

CD-HIT is a program that is used for the clustering of biological sequences in order to remove the sequence redundancy for better analysis. Currently the CD-HIT package has many programs for sequence clustering in that cd-hit-est is to cluster nucleotide sequences that meet a user defined similarity threshold. The tool which accepts DNA/RNA sequences in fasta format.

- ***Usage: cd-hit-est -i dataset_.fasta -o result.fasta -c 0.8***

dataset_.fasta = result file from step2

result.fasta = output file name

-i = Name of the output file from step2

-o = Output

-c = Percentage cut used of 80% similarity

3.3.2.4 txCdsPredict

txCdsPredict program from the UCSC genome browser, which predicts the potential ORFs from a given sequence.

3.3.2.5 System requirements

- Linux/Ubundu/centOS/MacOSX operating system
- gnu gcc – *http://www.gnu.org/software/gcc/*
- gnu make – *http://www.gnu.org/software/make/*
- MySQL development system and libraries – *http://dev.mysql.com/*
- libpng runtime and development packages – *http:// www.libpng.org/*
- libssl runtime and development packages – *http:// www.openssl.org/*
- 'git' soure code management: *http://git-scm.com/downloads*

➢ ***Usage:cd/kentUtilis/src/hg/txCds/txCdsPredict/./txCdsPredict result.fasta result.cds***

result.fasta = Output file from step3

result.cds = Output file name

3.3.2.5 Feature_extraction.pl

About 16 features were selected for accurate prediction of lncRNA which includes (1) GC content, (2) AACG, (3) CCGT, (4) CGCA, (5) CGCT, (6) CGGG, (7) CGTA, (8) TACC, (9) TACG, (10) TCCG, (11) TCGC, (12) Sequence length, (13) score, (14) cdsStop, (15) cdsSizes and (16) cdsPercent in order to avoid complex classification model, false correlations and increased computational time.

➢ ***Usage: perl feature_extraction.pl result.fasta result.cds***

result.fasta = Output file from step3

result.cds = Output file from step4

52

3.3.2.6 RNAplonc.model

➢ **Usage: java -Xmx10G -cp weka.jar weka classifiers.trees.J48 -l RNAplonc.model -T result.arff -p 0 >resultado_end.txt**

result.arff = output from step5

53

3.4 EXPERIMENTAL VALIDATION

The experimental validation of computationally predicted miRNAs and lncRNAs were conducted by randomly choosing two predicted miRNAs (Iba-miR1, Iba-miR2) and lncRNAs (Iba-lnc1, Iba-lnc2). RT-PCR and RT-qPCR were performed as described below using total RNA isolated from leaves samples of three different varieties of sweet potato available at ICAR-CTCRI.

3.4.1 Selected sweet potato varieties

- Sree Kanaka : dark orange coloured tuber
- ST13 : purple coloured tuber
- Khanjakad : white coloured tuber

3.4.2 Primer synthesis

3.4.2.1 miRNA sequences and primers

Two predicted sweet potato's miRNAs were taken for experimental verification. Stem loop RT primer combines 44 nt of stem loop sequence of Chen *et al.*, (2005), 5'- GTC GTA TCC AGT GCA GGG TCC GAG GTA TTC GCA CTG GAT ACG AC -3′ with the complement of the six 3' nt of the mature miRNA sequence. Forward primers are specific to the miRNA sequence but exclude the last six nucleotides at the 3' end of the miRNA. A 5' extension of 5–7 nucleotides is added to each forward primer to increase the melting temperature to 60° C; these sequences were chosen randomly. Standard primer design software *(http://www.idtdna.com/analyzer/Applications/oligoAnalyzer)* was used to assess the quality of forward primers. By using the 44 nt stem loop sequence for all RT primers, a universal primer can be derived from sequences within the stem loop. This is used as reverse primer. The universal reverse primer used is of sequence 5′-CCAGTGCAGGGTCCGAGGTA-3′ (Kramer, 2011).

3.4.2.2 lncRNA sequences and primers

Two predicted sweet potato's lncRNAs were taken for experimental verification. lncRNAs specific primers were designed using primer3plus software which is available at *www.bioinformatics.nl/primer3plus.*

3.4.2.3 Parameters for primer synthesis

Primer size: 20-22; Tm: 50-60; GC content: 50-80

54

3.4.2.4 Default parameters

Max Tm Difference: 100; Concentration of monovalent cations: 50; Annealing Oligo Concentration: 50; Concentration of divalent cations: 0.0; Concentration of dNTPs: 0.0

3.4.3 RNA isolation

Total RNA isolation was performed from leaves of three different varieties of sweet potato such as Sree Kanaka, ST13 and Khanjakad available at ICAR-CTCRI using Gene JET RNA Purification Kit of Thermo Fisher Scientific in accordance with manufacturer's protocol.

The integrity of the RNA was verified by 1% agarose gel electrophoresis. The gel was viewed under G: BOX gel documentation system (SYNGENE). RNA samples were kept at -80^{0}C.

3.4.4 RNA quantification

The concentration of RNA was determined by using a Nano-drop (using 1 OD260= 40µgRNA). A260/A280 ratios were also calculated for each sample.

3.4.5 Reverse transcription

Verification was done for the expression of two predicted miRNAs (Iba-miR1, Iba-miR2) and lncRNAs (Iba-lnc1, Iba-lnc2). For miRNAs specifically designed stem loop RT primers were used to obtain cDNA from total RNA. The components of the mixture were optimized as listed below for 20µl reverse transcription reaction.

Table 3. miRNA reverse transcription profile

Components	Volume (µl)
Verso Enzyme Mix	1µl
5X cDNA synthesis Buffer	4µl
dNTP Mix (5mM each)	2µl
RT Enhancer	1µl
Stem loop RT primer	1µl
Template	1.5µl
Nuclease free water	Variable

The RT reaction was performed by loading thermal cycler (Bio-Rad) and incubating for 30 min at 16^0C, followed by pulsed RT of 60 cycles at 30^0C for 30s, 42^0C for 30s and 50^0C for 1s. Finally the components were incubated at 85^0C for 5min for the inactivation of reverse transcriptase. The amplified products were separated on 1.5% agarose gel. The gel was viewed under G: BOX documentation system (SYNGENE).

Table 4. lncRNA reverse transcription profile

Components	Volume (µl)
Verso Enzyme Mix	1µl
5X cDNA synthesis Buffer	4µl
dNTP Mix (5mM each)	2µl
RT Enhancer	1µl
RT primer	1µl
Template	1.5µl
Nuclease free water	Variable

56

The RT reaction was performed by loading thermal cycler (Bio-Rad) and incubating for 30 min at 16^{0}C, followed by RT of 60 cycles at 30^{0}C for 30s, 42^{0}C for 30s and 50^{0}C for 1s. Finally the components were incubated at 85^{0}C for 5min for the inactivation of reverse transcriptase. The amplified products were separated on 1.5 agarose gel. The gel was viewed under G: BOX documentation system (SYNGENE).

3.4.7 RT-qPCR validation

Real-time quantitative polymerase chain reaction (RT-qPCR) is a sensitive technique for gene expression studies. The qPCR reaction was performed with forward and reverse primers (specific to predicted miRNAs Iba-miR1, Iba-miR2 and lncRNAs Iba-lnc1, Iba-lnc2) using cDNA samples from the young leaves of three different sweet potato varieties Sree Kanaka, ST13 and Khanjakad available at ICAR-CTCRI.

Table 5. RT-qPCR reaction profile

Components	Volume(µl)
Diluted cDNA	1.5µl
Forward primer (F)	1µl
Reverse primer (R)	1µl
DyNAmo ColorFlash SYBR Green Qpcr	5µl
Double distilled water	1.5µl

3.4.7.1Thermal profile

Initial denaturation : 95^{0}C 5min

Denaturation : 95^{0}C 10s

Annealing : 55^{0}C 10s

Extension : 72^{0}C 15-30s

Number of cycles : 35-45 cycles, step2-4

57

After the completion of the real time reactions, the threshold cycle (C_T) was recorded and gene expression level was calculated using delta- delta C_T method (Livak and Schmittgen, 2001).

The relative gene expression level of each miRNA and lncRNAs in different sweet potato varieties are represented as: $2^{-\Delta\Delta Ct}$

$$\Delta C_t = C_t \text{ (target gene)} - C_t \text{ (reference gene)}$$

$$\Delta\Delta C_t = \Delta C_t \text{ (sample)} - \Delta C_t \text{ (control)}$$

RESULTS

59

4. RESULTS

4.1 REVIEW OF microRNA PREDICTION TOOLS AND COMPARATIVE EVALUATION

Available plant miRNA identification and target prediction tools were selected for comparative evaluation based on several criteria. Most of the plant miRNA prediction tools are developed either in the form of a web server or as a standalone package. A review of all the plant miRNA identification and target prediction tools in accordance with their prediction features is presented in table 6, 7 and 8 respectively.

The miRNA identification tools: NOVOMIR and miRPlant and miRNA target prediction tools: psRNATarget and miRanda were taken for statistical comparison. The results obtained by miRNA identification tools: NOVOMIR and miRPlant and miRNA target prediction tools: psRNATarget and miRanda were organized in terms of percent count of TP, FN, FP, TN, Sensitivity, Specificity, Accuracy, MCC and PPV and is presented in table 9 and table 10.

Of the two tools considered for plant miRNA identification specificity, accuracy and sensitivity of NOVOMIR are as high as 98.47, 97.54, and 96.01 compared to 66.68, 77.26, and 87.85 of miRPlant. Hence proved that NOVOMIR is a best tool with 99.06 positive predictive values (PPV). The MCC value of NOVOMIR (0.95) is also high when compared to miRPlant (0.55) showing the high efficacy of NOVOMIR in miRNA identification. Similarly for target prediction tools comparison specificity, accuracy and sensitivity of psRNATarget is as high as 83.33, 77, 71.43 compared to 77.79, 71.87, 65.71 of miRanda. Hence proved that psRNATarget is a best tool with 87.88 positive predictive values (PPV). The MCC value of psRNATarget (0.55) is also high when compared to miRanda (0.15) showing the high efficiency of psRNATarget in miRNA-target prediction. Graphical analysis of both miRNA identification and target prediction tools were displayed in figure1 and 2 respectively.

Table 6. Review of miRNA identification tools

Program	Algorithm	Availability	Requirements	Reference
BayesmiRNAfind	Naive Bayesian model	Only web server	Linux/Unix	(Yousef *et al.*, 2009).
Mirfinder	Support vector machine	Source code and web server	Linux /Windows / MacOsX Python	(Huang *et al.*, 2007).
miRPlant	Java coded bowtie alignment algorithm	Source code and web server	Java	(An *et al.*, 2014).
miRA	Support vector machine	Only source code	Java 1.6+, LaTeX, gnuplot	(Evers *et al.*, 2015).
NOVOMIR	Heuristic filters and paired hidden markov model	Only source code	Perl	(Teune *et al.*, 2010).

Table 7. Review of miRNA- target prediction tools

Program	Algorithm	Availability	Requirements	Reference
miRanda	Local alignment	Source code and web server	C++, Java	(Enright *et al.*, 2003).
psRNATarget	Smith-Waterman	Only web server	FASTA sequence	(Archak and Nagaraju, 2007).
p-TAREF	Support vector regression (SVR)	Source code and web server	Linux 64 bit O.S. with Qt 4	(Jha and Shanker, 2011).
Target scan	Custom made	Source code and web server	Perl	(Agarwal *et al.*, 2015).
TAPIR	FASTA/ RNAhybrid	Source code and web server	Bioperl	(Bonnet *et al.*, 2010).

Table 8. Prediction features used in different miRNA-target prediction tools

Program	Website	SP	TSA	MS	CF	TI	Input
miRanda	*http://www.microrna.org/*	+	+	+	-	+	User-supplied miRNA sequence and UTR sequence for command line
psRNATarget	*(https://plantgrn.noble.org/psRNATarget/)*	-	+	+	-	+	User-supplied miRNA sequence and UTR may be in the tool itself or can be supplied by the user
p-TAREF	*scbb.ihbt.res.in/new/p-taref/form1.html*	-	+	+	-	-	User-supplied UTR sequence
Target scan	*http://www.targetscan.org*	+	-	+	-	+	miRNA name, gene name or miRNA family
TAPIR	*bioinformatics.psb.ugent.be/webtools/tapir/*	+	+	+	-	-	User-supplied miRNA sequence and UTR sequence

SP: Seed pairing, **TSA:** Target site accessibility, **MS:** Multiple sites, **CF:** Conservation filter, **TI:** Translation inhibition; **'+'** indicates the prediction feature used by the tool and '-' indicates that these prediction feature is not used by the tool.

63

Table 9. Performance of miRNA identification tools

Parameters	NOVOMIR	miRPlant
True positive (%)	96.01	87.85
False negative (%)	3.98	12.15
False positive (%)	0.91	33.32
True negative (%)	98.47	66.68
Specificity (%)	98.47	66.68
Sensitivity (%)	96.01	87.85
Accuracy (%)	97.54	77.26
MCC	0.95	0.55
PPV	99.06	72.50

Table 10. Performance of miRNA-target prediction tools

Parameters	PsRNATarget	MiRanda
True positive (%)	71.43	65.71
False negative (%)	28.57	34.29
False positive (%)	16.67	22.22
True negative (%)	83.33	77.79
Specificity (%)	83.33	77.79
Sensitivity (%)	71.43	65.71
Accuracy (%)	77	71.87
MCC	0.55	0.15
PPV	87.88	74.72

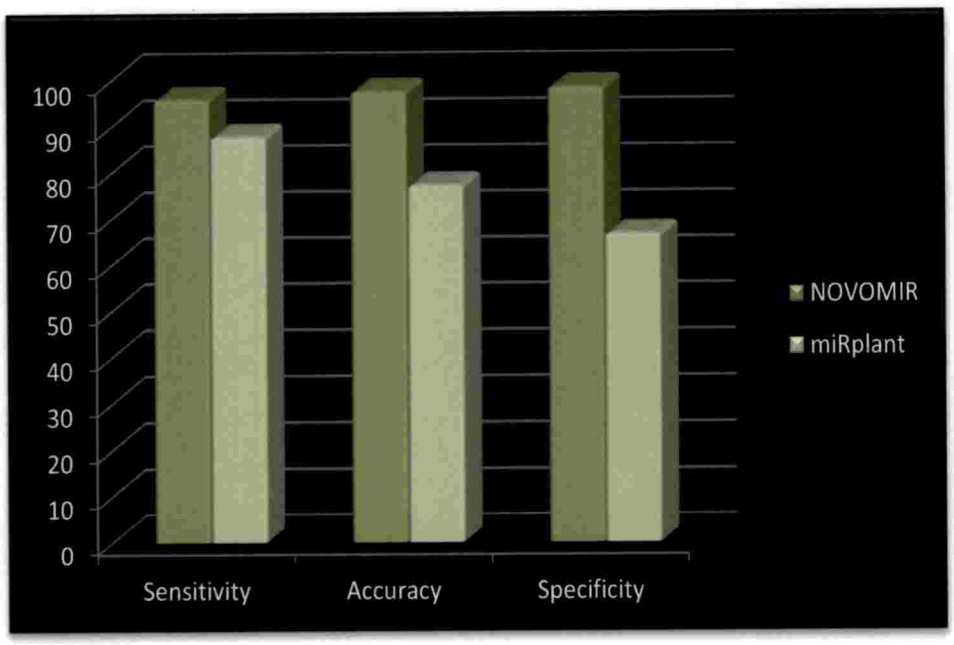


Figure 1. Graphical representation of the efficiency of miRNA identification tools such as NOVOMIR and miRplant based on statistical parameters Sensitivity, Accuracy, Specificity

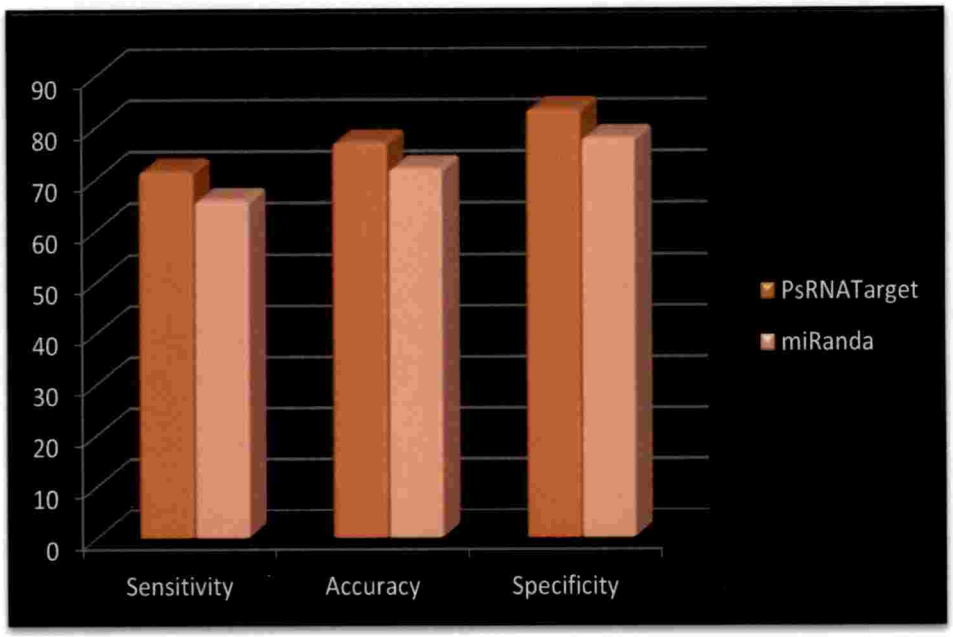


Figure 2. Graphical representation of the efficiency of miRNA-target prediction tools such as psRNATarget and miRanda based on statistical parameters Sensitivity, Accuracy, Specificity

4.2 IDENTIFICATION OF SWEET POTATO miRNA

4.2.1 Mining of sweet potato EST datasets and its assembly by CAP3

A total of 37698 ESTs of *Ipomoea batatas* downloaded from NCBI *(https://www.ncbi.nlm.gov/)* were available for analysis. The EST assembly have been performed in order to remove the redundancy in sequences by CAP3 sequence assembly program which results in 4853 over lapping sequences known as contigs and 15256 non-overlapping sequences known as singletons sequences. The outputs obtained were taken for further analysis.

4.2.2 Sweet potato pre-miRNA prediction by NOVOMIR

NOVOMIR which is a pre-miRNA identification tool have been performed for the identification of potential pre-miRNA candidates of sweet potato using input sequences as the clustered 4853 contigs and 15256 singletons sequences. The program which uses a series of filtering steps and statistical models to predict the pre-miRNA sequences in an accurate and specific manner. The output results in 343 pre-miRNAs from contigs and 698 pre-miRNA from singletons sequences.

Most of the plant miRNAs are evolutionarily conserved within species. This feature is widely used for the *in silico* analysis of potential miRNAs based on homology comparison. Mostly homology search was performed in miRBase which is an online repository of all the identified miRNAs of different species. The identified potential pre-miRNA sequences from NOVOMIR can be used for the prediction of putative miRNAs based on homology search against BLASTN in miRBase database. Homology comparison was done between the predicted pre-miRNA sequences with all the known miRNAs from *viridiplantae* in miRBase. The selection of putative miRNA is based on certain parameters like length should be between 18-22nt, mismatch permitted is 0-4nt and the score considered should be >70. Hence 13 potential miRNAs, 5 from contigs and 8 from singletons were predicted.

>Contig124, >Contig172, >Contig234, >Contig247, >Contig251 from contigs and >DC881876, >DC882009, >DV035460, >EE878362, >EE880233, >HX982495, >HX982890, >HX983556 from singlets sequences are the predicted

putative miRNAs based on homology approach. A unique name was assigned to each identified miRNA. The details regarding the predicted 13 miRNAs are given in table 11 and 12 respectively. The work flow for *in silico* prediction of sweet potato miRNAs and target is shown in figure 3.

4.2.3 Secondary structure representation of miRNA using RNA::HairpinFigure

Secondary structure analysis of identified pre-miRNA sequences was done using RNA::HairpinFigure. RNA::HairpinFigure is a program that draws the hairpin like text figures of the RNA sequences along with its secondary structure in dot bracket notation.

All the predicted pre-miRNAs and its dot bracket notation from NOVOMIR are used as the input sequences for RNA::HairpinFigure. Hence the hairpin text figures of 343 contigs and 698 singlet sequences were predicted. The secondary structures of all the identified micro-RNAs by RNA::HairpinFigure are shown in Appendix I.

4.2.4 Phylogenetic analysis of potential miRNAs in Sweet potato

Phylogenetic analysis was done with all the newly predicted 13 miRNAs and their corresponding homologous precursor miRNAs using Environment for Tree Exploration (ETE) available at GenomeNet. The output data proven that miRNA is highly conserved in nature. The constructed phylogenetic tree is displayed in figure 4.

Table 11. Characterization of the novel identified sweet potato miRNAs.

Sweet potato miRNAs	Source	Homologous miRNA	Query region	Similarity	Evalue	Score	NM	LM	LP	Strand
Iba- miR1	>Contig124	cca-miR6110-5p	5-26	18/22	1.2	74	4	22	251	+
Iba- miR2	>Contig172	stu-miR319-5p	56-76	18/21	0.60	78	3	21	136	-
Iba- miR3	>Contig234	lja-miR1117a-3p	120-140	18/21	0.75	78	3	21	165	-
Iba- miR4	>Contig247	stu-miR398b-5p	150-171	18/22	1.1	74	4	22	245	+
Iba- miR5	>Contig251	mtr-miR5278	161-182	18/22	0.82	74	4	22	183	+
Iba- miR6	>DC881876	gma-miR10407c	71-92	18/22	1.5	74	4	22	153	-
Iba- miR7	>DC882009	ath-miR829-3p.1	51-71	18/21	0.62	78	3	21	282	-
Iba- miR8	>DV035460	pab-miR1551a	150-171	18/22	0.93	74	4	22	204	+
Iba- miR9	>EE878362	ath-miR836	113-134	18/22	0.99	74	4	22	215	+
Iba- miR10	>EE880233	ath-miR5653	89-110	18/22	0.81	74	4	22	181	+
Iba- miR11	>HX982495	vvi-miR3639-5p	6-27	18/22	1.3	74	4	22	134	+
Iba- miR12	>HX982890	gma-miR4994-3p	15-35	18/21	0.54	78	3	21	125	+
Iba- miR13	>HX983556	pab-miR11551a	10-30	18/21	0.46	78	3	21	214	-

68

Table 12. Major sequence features of the novel predicted sweet potato miRNAs

Sweet potato miRNAs	Mature sequence	(A+U) %	(G+C) %	MFE	MFEI	AMFE
Iba- miR1	AUCUUGUGACAUCAGAUAAUGU	66.13	33.86	67.50	0.76	26.89
Iba- miR2	GGAUGGACUAAACACUUUCUU	60.29	39.70	39.30	0.70	28.89
Iba- miR3	CUUCAUCUUCUUCCUCCAAAA	60.60	39.39	48.70	0.73	29.51
Iba- miR4	GAGUGUGCCUUGGAACCGGGGU	50.61	49.38	77.94	0.62	31.79
Iba- miR5	AAUUCUCUUCAUGAAAUGUGUA	68.85	31.14	43.20	0.73	23.60
Iba- miR6	CGAAAAAAAUAUUCGAAUGUUA	56.07	43.92	98.80	1.04	46.16
Iba- miR7	CCUUCAUCUGGUAUUAGAGCU	64.80	35.20	30.70	0.68	24.56
Iba- miR8	UGUUUUUGUUUCCCUCUGCAUU	61.94	38.05	36.70	0.70	27.38
Iba- miR9	CUGGCUUUCCUUUGAAGUGUGG	64.64	35.35	45.40	0.70	25.00
Iba- miR10	UGUGUUGAAUUGAAUUGAAUUG	72.55	27.44	51.20	0.83	23.81
Iba- miR11	AUAGACUUCUGGAAGGGGAAAA	71.07	28.92	50.20	0.82	24.60
Iba- miR12	UGCUAUGCUUGAUCUAAUACA	69.14	30.85	64.70	0.71	22.94
Iba- miR13	UUGGGGGGGGAAAAAAAAACA	68.62	31.37	40.80	0.82	26.66

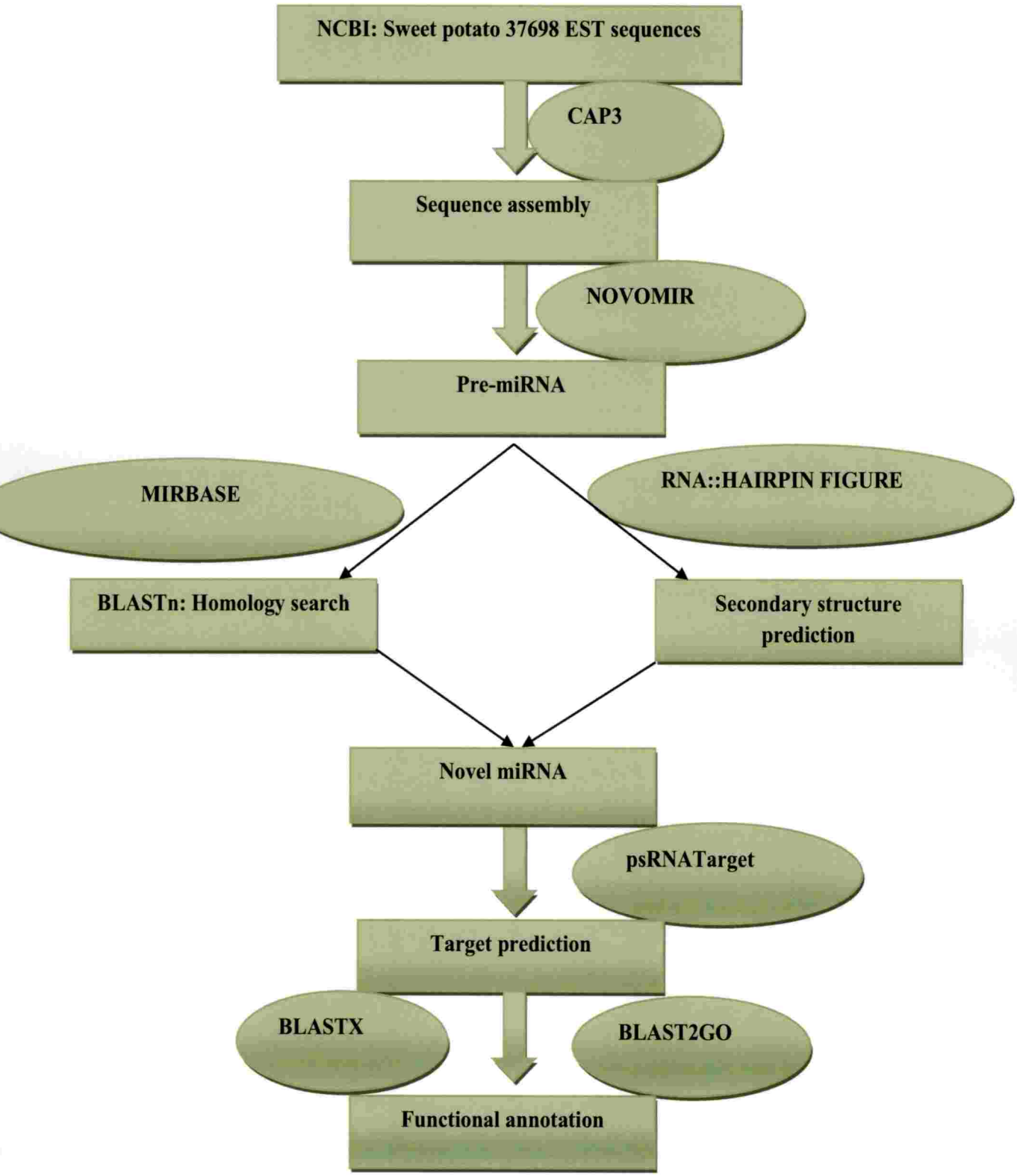


Figure 3. Workflow for the identification of miRNAs in sweet potato

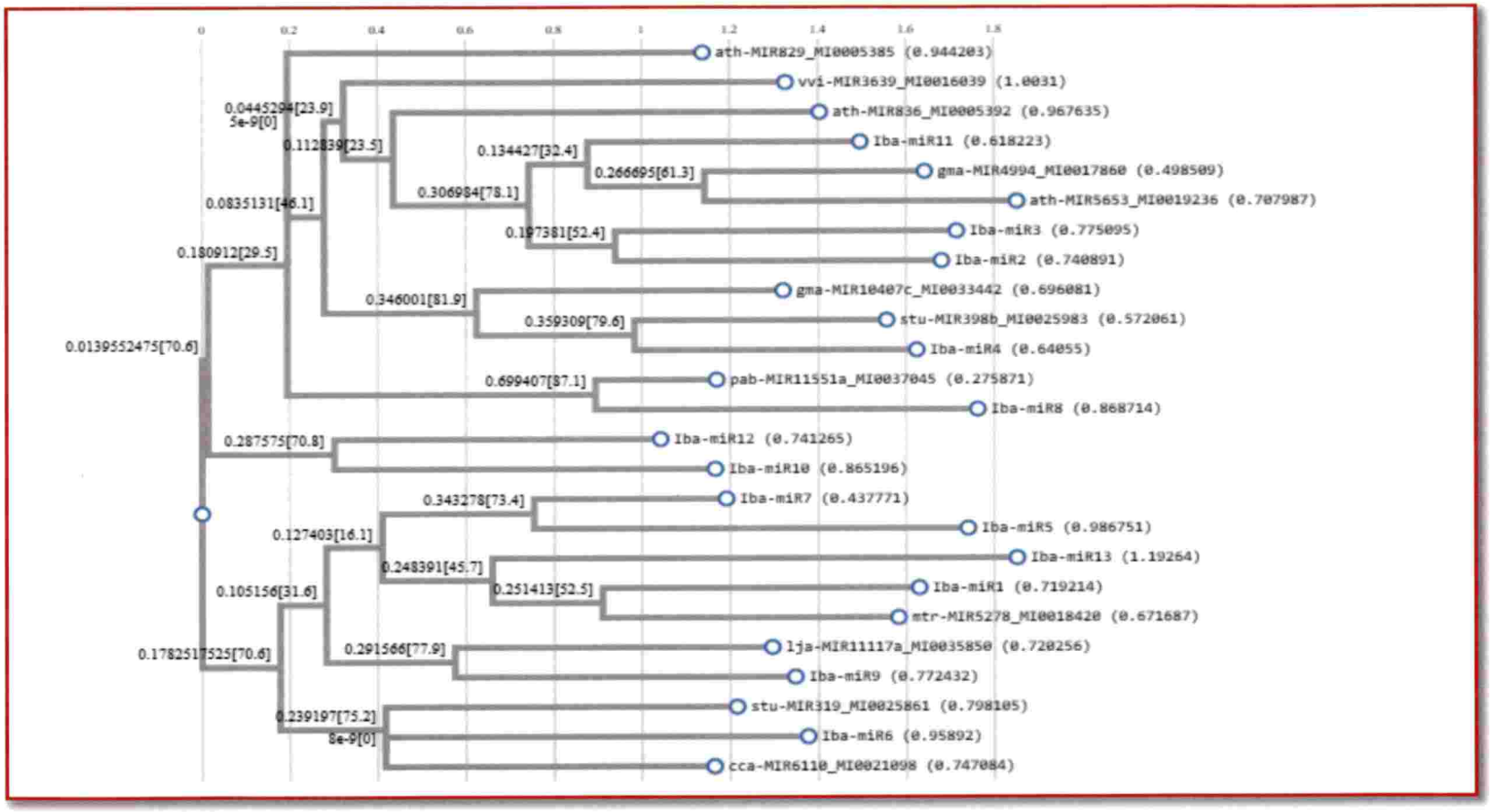


Figure 4. Phylogenetic analysis of predicted pre-miRNAs and its homologous pre-miRNAs in sweet potato

71

Most of the predicted mature miRNAs is characterized with having a length between 20-22 nucleotides, Similarly the precursor miRNA length ranges from 125 nucleotides to 282nt, having higher negative minimum fold energies (MFEs) (ΔG Kcal/mol) which ranges from -30.70 to -98.80 Kcal/mol, adjust minimum fold free energy (AMFE) ranges from 22.94 to 46.16, and minimum free energy index (MFEI) ranges from 0.6 to 1.0. The length distribution of both mature miRNAs and its precursor miRNAs are displayed in figure 5 and 6 respectively.

An important parameter that indicates the evolutionary relationship and stabilization of one specific RNA sequence cased by their secondary structure is the composition of the four nucleotides (A, G, C, and U). The predicted miRNAs shows an A+U content of 50.61 to 72.55% compared to G+C which is 28.92 to 49.38%.

The present study which shows lower GC content for predicted miRNAs in sweet potato. In general the nucleotides G and C has an important role in the formation and stabilization of the secondary structure in stem loop hairpins due to its ability to form three hydrogen bonds compared to A and U nucleotides which form only two hydrogen bonds with each other. The distribution of (A+U) % and (G+C) %, MFE, MFEI, AMFE content are displayed in figures 7 and 8 respectively.

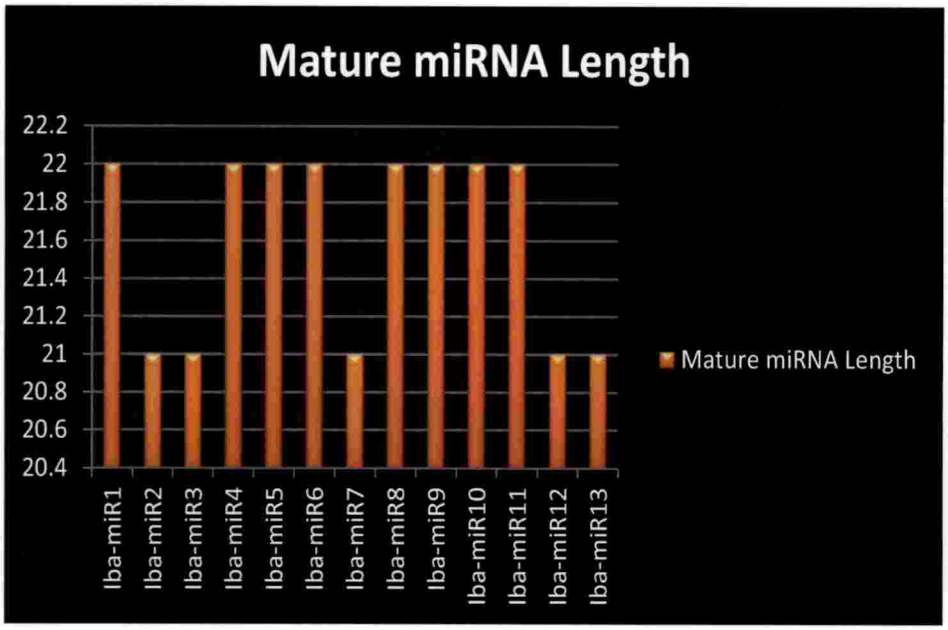


Figure 5. The length distribution of identified sweet potato mature miRNAs

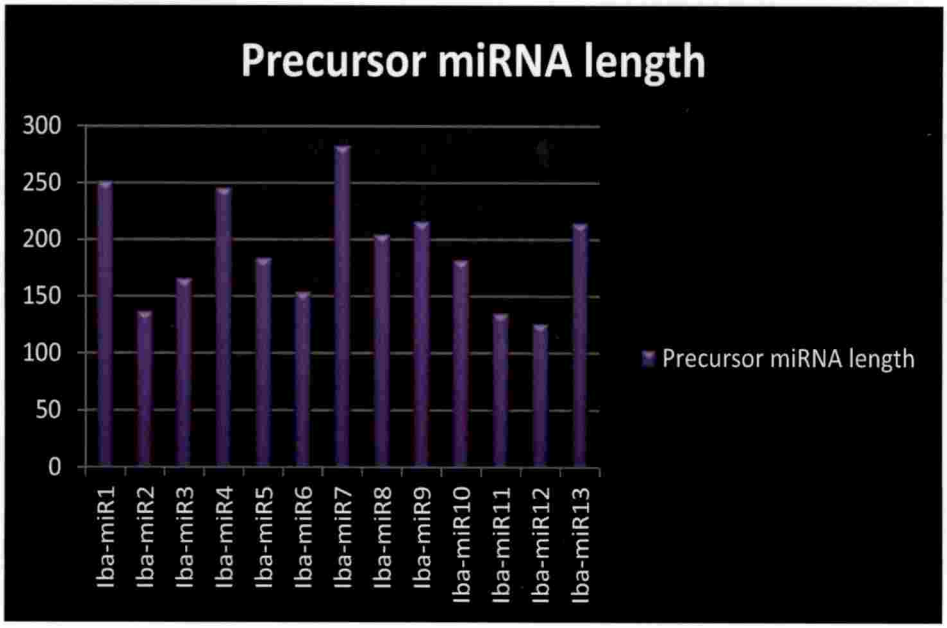


Figure 6. The length distribution of identified sweet potato pre-miRNAs

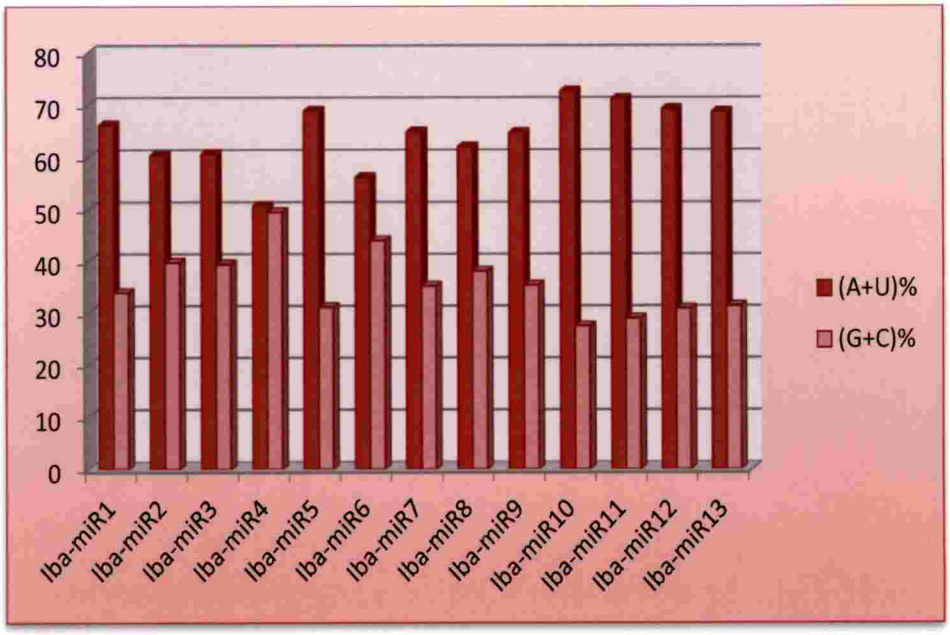


Figure 7. The distribution of (A+U) % and (G+C) % content of predicted sweet potato pre-miRNA

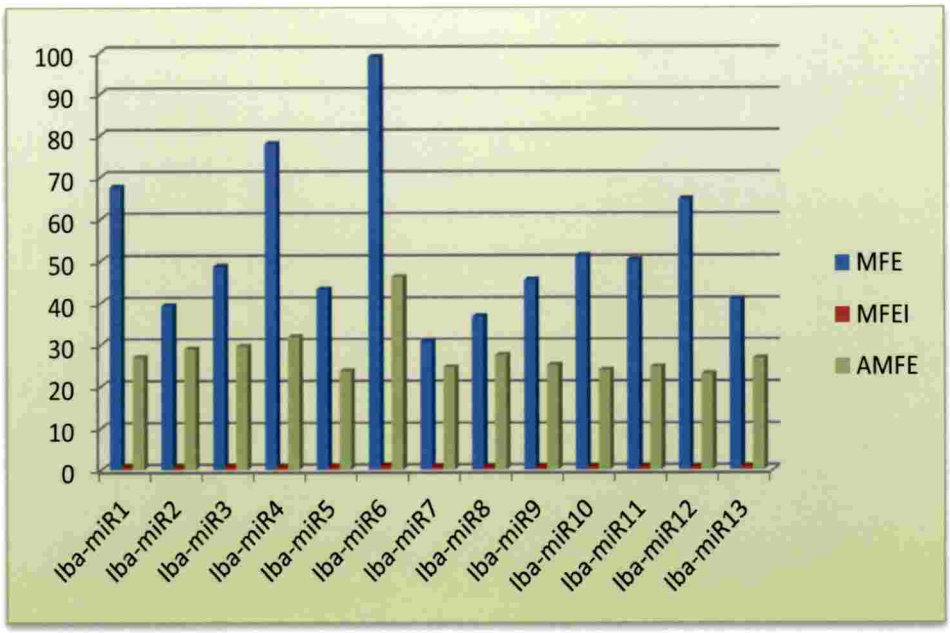


Figure 8. The distribution of MFE, MFEI and AMFE of predicted sweet potato pre-miRNA

4.2.5 Potential target prediction of miRNA by psRNATarget

Functional role of miRNAs can be achieved only through a better knowledge regarding the potential targets. MicroRNAs identify their specific target genes via perfect or near perfect complementary interaction which leads to the regulation of target gene expression. The targets were predicted by comparing newly identified miRNAs against ESTs of *Ipomoea batatas* in psRNATarget webserver.

All predicted miRNAs have target genes except Iba-miR5 and Iba-miR12 and the peculiarity is that other miRNAs are credited with more than one target genes which reveal the importance of computationally predicted miRNAs. Most of the miRNAs are regulating the target gene via cleavage mechanism. Details regarding the target gene were shown in Appendix II.

4.2.6 Functional annotation of targets using BLASTX and BLAST2GO

Functional analyses of predicted target genes were done using BLASTX against non redundant database and BLAST2GO. The analysis have shown that most of the targets predicted in this study are involved in regulation of plant growth, development, signal transduction, post-transcriptional gene silencing, protection of plants against pathogens, metabolic pathways and in stress responses. Detailed information about the functional analysis of target genes by BLASTX is shown in Appendix III. Similarly BLAST2GO analyses were done and organize the target genes in accordance with cellular function, molecular function and biological process. All the details of BLAST2GO analysis data were shown in figure 9, 10 and 11 respectively.

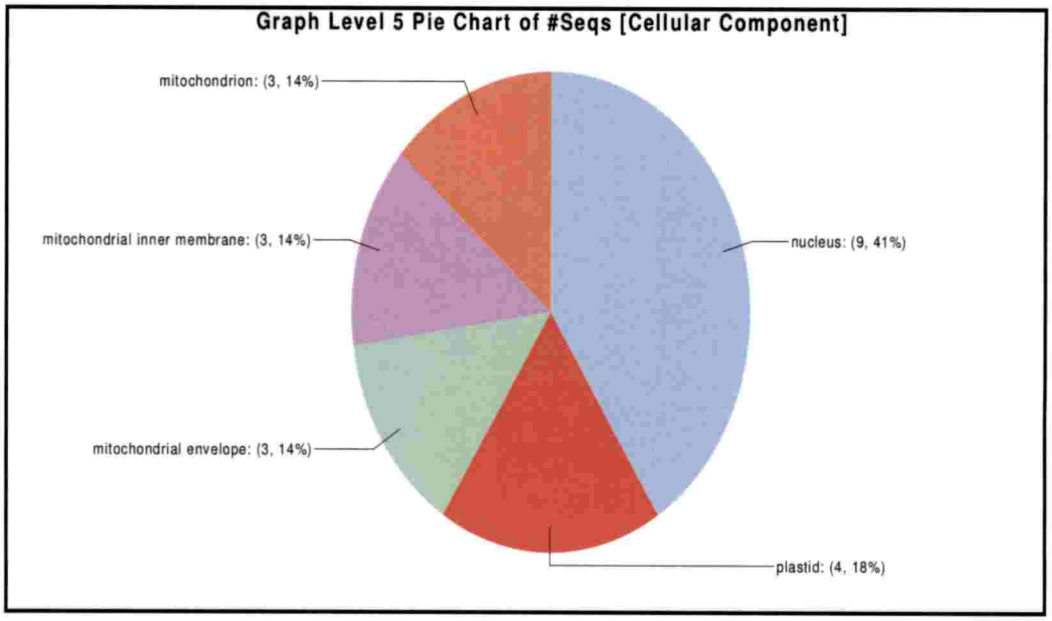


Figure 9. miRNA-target gene sequences were annotated to "cellular component" ontology

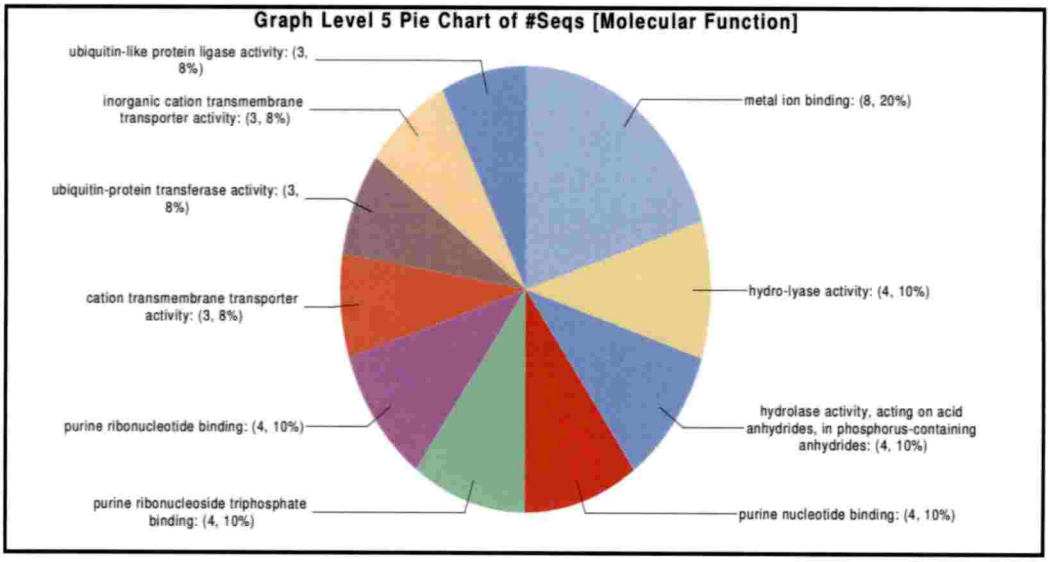


Figure 10. miRNA-target gene sequences were annotated to "molecular function" ontology

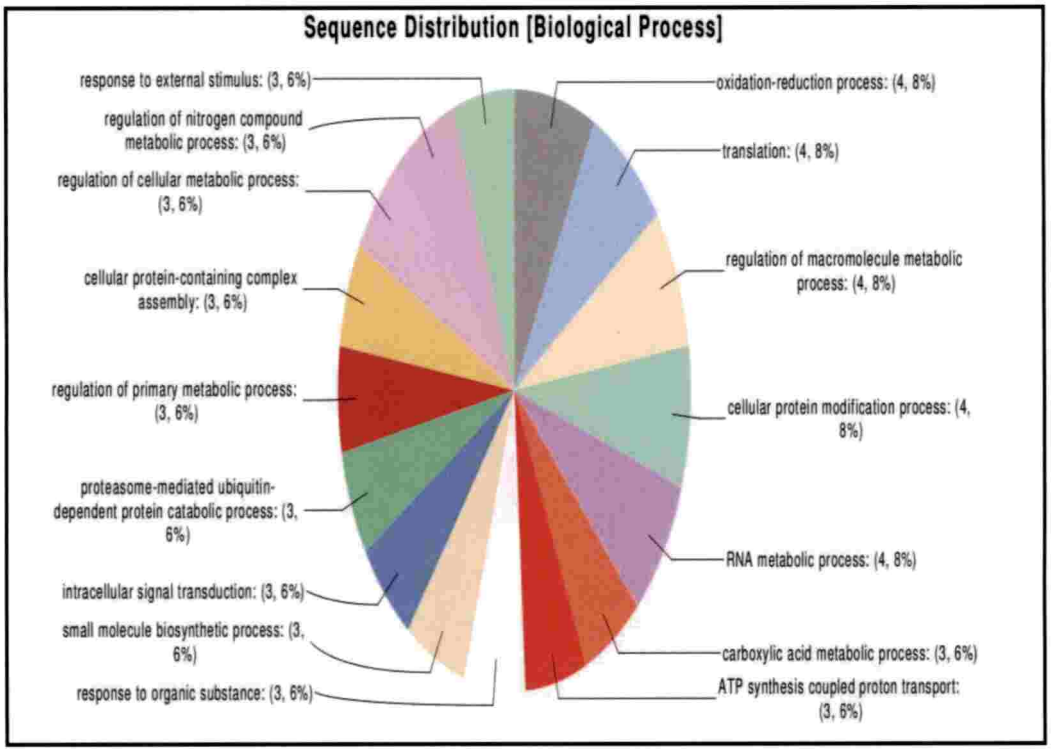


Figure 11. miRNA-target gene sequences were annotated to "biological process" ontology

4.3 SWEET POTATO lncRNA PREDICTION BY RNAPLONC

RNAplonc, a classifier approach for the identification of lncRNAs in plants from mRNA based data. The tool itself uses a series of steps to distinguish lncRNAs from mRNAs (Negri *et al.*, 2017).

4.3.1 200nt.pl

All the available 37698 *Ipomoea batatas* ESTs from NCBI were taken for lncRNA analysis. RNAplonc which removed all the sequences less than 200 nucleotides. Because lncRNA is characterized with 200 nucleotide long molecule. Thus got 35633 potential sequences in that the largest sequence length is 1496 and the shortest sequence is 200 were taken for further analysis of lncRNA prediction.

4.3.2 Cd-hit-est

All the 35633 long sequences were taken for clustering in order to remove the sequence redundancy for better analysis with a percentage cut of 80% similarity. The results after clustering analysis include 17944 sequences. The output is taken for further analysis.

4.3.3 txCdsPredict

The potential lncRNA candidates are possessing 200nt length with or without coding potential. Most of the cases lncRNA's ability to produce a functional polypeptide is very less. txCdsPredict program from the UCSC genome browser, which predicts the potential ORFs of the available 17944 sequences and filtered potential lncRNA candidates with non coding nature from all protein coding genes. The output gives 17880 sequences with predicted ORF, in that all the sequences with a score value > 800 is regarded as protein others as lncRNAs.

4.3.4 feature_extraction.pl

All the 17880 output sequences from txCdsPredict were taken for feature selection. The features includes (1) GC content, (2) AACG, (3) CCGT, (4) CGCA, (5) CGCT, (6) CGGG, (7) CGTA, (8) TACC, (9) TACG, (10) TCCG, (11) TCGC, (12) Sequence length, (13) score, (14) cdsStop, (15) cdsSizes and (16) cdsPercent. The results are available at .arff format.

78

4.3.5 RNAplonc.model

An RNAplonc.model was created with all those 17880 sequences which include both lncRNAs and mRNAs with a percentage of classification data for discriminating lncRNAs from mRNAs.

4.3.6 Result

RNAplonc efficiently predicts 9215 lncRNAs and 8665 protein coding genes in sweet potato from available ESTs sequences. Graphical representation of predicted lncRNAs and mRNAs in sweet potato by RNAplonc.were displayed in figure 12.

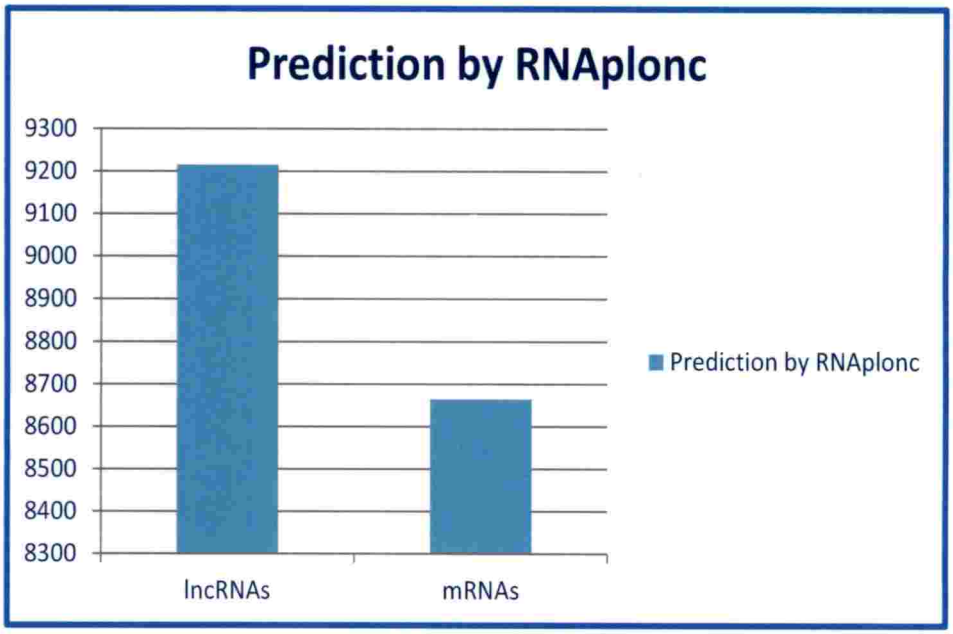


Figure 12. Graphical representation of predicted lncRNAs and mRNAs in sweet potato by RNAplonc

4.4 EXPERIMENTAL VALIDATION

The expression of two miRNAs (Iba-miR1, Iba-miR2) and two lncRNAs (Iba-lnc1, Iba-lnc2) were detected using a two step process. In the first, the RT primers were hybridized to miRNA and lncRNA molecules and then reverse transcribed in a reverse transcription reaction. In second, the RT product was amplified and quantified using DyNAmo Flash SYBR Green qPCR assay.

4.4.1 miRNA sequences and primers

Based on the mature miRNA sequences of Iba-miR1 and Iba-miR2 stem loop RT primers (miR1RT, miR2RT) and forward primers (miR1F, miR2F) were designed. A universal primer based on stem loop sequence was used as reverse primer for qPCR reaction. The primer details are given in table 13.

4.4.2 lncRNA sequences and primers

Based on lncRNA sequences of Iba-lnc1 and Iba-lnc2, forward and reverse primers were designed (Iba-lnc1F, Iba-lnc1R, Iba-lnc2F, Iba-lnc2R) were designed. The primer details are given in table 14.

4.4.3 Total RNA isolation

Total RNA isolation was performed from leaves of 3 different varieties of sweet potato Sree Kanaka, ST13 and Khanjakad available at ICAR-CTCRI using GeneJET RNA purification Kit of Thermo Fisher Scientific. A distinct or intact RNA with minimum RNA degradation and minimum genomic DNA contamination were observed on agarose gel, showing good quality total RNA extraction. Gel image of RNA isolated from leaves of 3 different varieties of sweet potato ST13, Khanjakad and Sree Kanaka were displayed in Plate 1.

4.4.4 Reverse transcription

The reverse transcription of extracted total RNA was carried out with stem loop RT primer (miR1RT, miR2RT) and with RT primer to get cDNA of corresponding miRNA and lncRNA. Positive results were viewed in the agarose gel.

Plate 1. Gel image of RNA isolated from 3 different varieties of sweet potato ST13, Khanjakad and Sree Kanaka- ST: ST13; K: Khanjakad; SK: Sree Kanaka

4.4.5 DyNAmo Flash SYBR Green qPCR Assay

82

Quantitative Real Time PCR (qPCR) is a reliable method to determine the expression of specific miRNAs and lncRNAs. qPCR was used to detect the expression of the two computationally predicted miRNAs and lncRNAs.

The result shows that all of the 2 predicted miRNAs and lncRNAs exists and are expressed in the entire selected sweet potato varieties figure 13. Figure 14 shows the melt curve assay. Here the threshold cycle (c_T) values for the miRNA and lncRNA amplification are listed in the table 15. In order to compare the expression level delta delta C_T method were used. Actin was used as a reference gene and ST13 as calibrator for normalization in sweet potato. Relative expression levels of target genes compared against a calibrator were displayed in table 16.

Table13.Synthesized primers for miRNA validation

Target miRNA	Primer name	Primer sequence (5' to 3')	No. of bases	GC content	Tm (^{0}C)
Iba-miR1	miR1RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACACATTA	50	52	69.4
	miR1F	GCGTGCATCTTGTGACATCAG	21	52.4	56.7
	miR1R	CCAGTGCAGGGTCCGAGGTA	20	65.0	61.3
Iba-miR2	miR2RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAAGAAA	50	52.0	69.6
	miR2F	GGT TGCGGATGGACTAAACAC	21	61.9	59.9
	miR2R	CCAGTGCAGGGTCCGAGGTA	20	65	61.3

Table 14. Synthesized primers for lncRNA validation

Target lncRNA	Primer name	Primer sequence (5' to 3')	No. of bases	GC content	Tm (^{0}C)
Iba-lnc1	lnc1F	CGGAGACTATGATCCGGTGT	20	55.0	56.0
	lnc1R	TACCGGCGTAGGTCTTGAAC	20	55.0	56.7
Iba-lnc2	lnc2F	GTCCTCAACATGTGGTCGTG	20	55.0	55.9
	lnc2R	AACTTTCGGGTTCAACATCG	20	45.0	53.

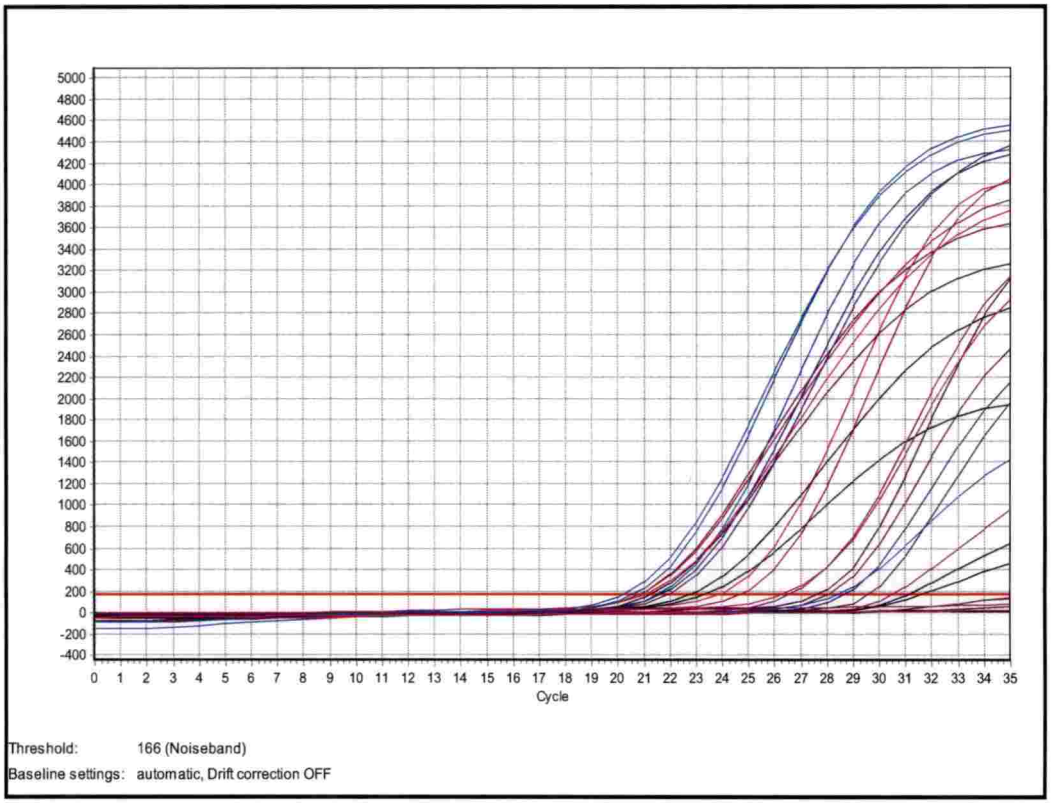


Figure 13. Real time PCR for Iba-miR1, Iba-miR2 and Iba-lnc1, Iba-lnc2 validations with designed primers

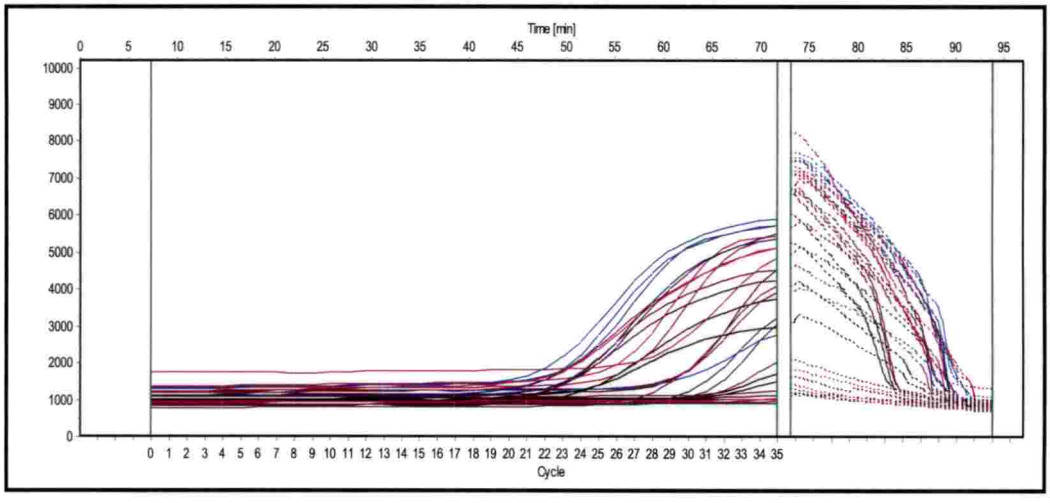


Figure 14. Melt curve analysis

Table 15. Mean C_T value in qPCR amplification

Target gene	Khanjakad	ST13	Sree Kanaka	Amplification
Iba-miR1	24.26	31.31	27.81	+
Iba-miR2	24.93	30.28	29.14	+
Iba-lnc1	20.86	22.95	20.97	+
Iba-lnc2	20.38	26.36	21.75	+

Table 16. Target gene's expression in different sweet potato varieties

Target gene	Gene expression levels		
	ST13	Khanjakad	Sree Kanaka
Iba-miR1	13.37	65.79	11.79
Iba-miR2	13.33	77.17	2.05
Iba-lnc1	5.04	7.11	6.02
Iba-lnc2	8.66	64	22.62

DISCUSSION

5. DISCUSSION

The advent of next generation sequencing is credited with the exploitation of non coding RNAs which is a crucial gene regulator of wide variety of biological processes in plants. Their functional importance at RNA level generated a great momentum for the discovery and functional annotation of massive number of ncRNAs. Among the large class of ncRNAs micro and long non coding RNAs made great attention to the researchers because of its potent regulatory action. MicroRNAs are short regulatory ncRNA molecule with 20-22nt length, performs sequence specific function in both plants and animal development. Similarly lncRNA candidates are long regulatory ncRNA molecule with 200nt length, are emerging class of highly heterogeneous molecules having action on multiple levels of gene expression in both plants and animals. Different methodologies are there for the identification of ncRNAs which includes both experimental screening and *in silico* analysis. Expensive nature of experimental strategies paves way for the establishment of efficient computational approaches. This computational analysis plays a crucial role for unravelling the complex miRNA and lncRNA mediated regulatory networks controlling growth and development, metabolism and tolerance to biotic and abiotic stresses in plants through comparative genomics approach.

Expressed sequence tags (ESTs) are an efficient tool for miRNA and lncRNA prediction due to the evolutionary conservation of sequences among plant species (Gupta *et al.,* 2015). These ESTs are partial cDNA sequences of a species that are highly expressed one and are cloned into a plasmid. The increasing data of ESTs of a species are good source for gene discovery and annotation (Ohlrogge and Benning, 2000). There are different bioinformatic tools available for the *in silico* prediction of miRNAs and lncRNAs. The tools which uses diverse algorithms for the precise prediction of ncRNA candidates. A combined effort of bioinformatics tools and comparative genomics is a powerful well proved strategy adopted by the researchers' worldwide for identification of non coding RNAs. Successful identification and validation of the new miRNAs and lncRNAs in sweet potato have been done.

88

5.1 REVIEW OF miRNA PREDICTION TOOLS AND COMPARATIVE EVALUATION

There are different tools available for the identification and target prediction of microRNAs. It is very important to analyse the diversities among all the available algorithms and the involved prediction criteria used by different tools in order to develop a novel strategy to contribute more to the miRNA research world. Thus a comparative analysis of available miRNA prediction tools is very useful for the future studies (Sinha *et al.,* 2009). Here available miRNA identification and target prediction tools were reviewed in detail to find out the role of computational biology in miRNA research.

Performance analysis of different tools was done to find out the best one in miRNA research. For comparison in miRNA identification and target prediction two tools selected for each purpose, which is selected based on the ease of availability and wide use. NOVOMIR and miRPlant miRNA identification tools were selected to check their efficacy in miRNA identification. Similarly psRNAtarget and miRanda were selected to compare their performance in miRNA-target prediction. The output from each tool was analysed by different statistical parameters for accurate results. Hence in case of miRNA identification NOVOMIR shows higher specificity, sensitivity, accuracy, MCC and PPV compared to miRPlant. In the case of target prediction psRNATarget had better accuracy, sensitivity, specificity, MCC, PPV compared to miRanda. So the study by statistical validation proves that NOVOMIR and psRNATarget was found to be a better tool in miRNA identification and target prediction in plants compared to miRPlant and miRanda respectively.

5.2 IDENTIFICATION OF SWEET POTATO miRNAs AND THEIR TARGETS

Computational identification of miRNA prediction widely exploits the evolutionary conserved nature of miRNAs in plants (Muvva *et al.*, 2012). Nucleotide sequences, genome sequences, ESTs and Genome survey sequences (GSS) are the input sequences for homology based prediction of potential miRNAs. EST analysis is credited for identifying conserved miRNAs, for species whose genome is less studied (Zhang *et al.,* 2005). High efficacy and cost

effectiveness made *in silico* analysis as best compared to experimental studies. 89

Due to the hexaploid nature of sweet potato (*Ipomoea batatas*) "the poor man's food" belonging to the family of *Convolvulaceae*, EST analysis can be widely used as an alternative strategy for the identification of potential microRNAs. This staple food crop is widely exploited by huge population due to its high nutritive and economical value. For sweet potato, a total of 37698 EST sequences have been deposited in dBEST database of NCBI (as of November 2, 2017).

The present study used all the available 37698 ESTs of sweet potato for miRNA prediction. For accurate analysis all the EST sequences were treated by CAP3 sequence assembly program to remove the sequence redundancy. Thus the resulted 4853 over lapping sequences known as contigs and 15256 non-overlapping sequences known as singletons sequences were taken as input sequences for the prediction of pre-miRNA by NOVOMIR which is a widely used highly efficient pre-miRNA identification tool in plants. The prediction tool which uses a series of filtering steps and statistical analysis for efficient pre-miRNA predicion. NOVOMIR decides 343 pre-miRNAs from contigs and 698 pre-miRNA from singletons sequences are the most potent pre-miRNA sequences holding mature miRNA candidates.

In order to find out the highly efficient mature miRNAs a homology screening was done with miRBase which is an online repository of all the published miRNAs. So all the available 343 pre-miRNAs from contigs and 698 pre-miRNA from singletons sequences were taken for filtering by exploiting the highly conserved nature of plant miRNAs. Thus 13 potential mature miRNAs belonging to 13 different families were predicted based on homology search.

There are evidences which show that the newly identified miRNA have characteristics similar to other known miRNAs due to high sequence conservation and the pre-miRNA sequence length may be diverse among species. (Sunkar *et al.*, 2005; Zhang *et al.*, 2006a,b; Zhang *et al.*, 2008; Panda *et al.*, 2014; Qu *et al.*, 2016). Similarly all the newly identified pre-miRNAs length ranges from 125nt to 282nt. This diversity may sometimes lead to diverse regulatory role.

All the newly predicted potential mature miRNAs holds 20nt to 22nt length. In order to find out potential miRNAs MFEs, AMFE and MFE index (MFEI) of the secondary structures are an important factor to determine the evolutionary relationship among miRNAs (Monavar *et al.*, 2012; Vishwakarma and Jadeja 2013). Here the predicted miRNAs having higher negative minimum fold energies (MFEs). (ΔG Kcal/mol) of ranges from -30.70 to-98.80 Kcal/mol, adjust minimum fold free energy (AMFE) ranges from 22.94 to 46.16, and minimum free energy index (MFEI) ranges from 0.6 to 1.0.

Functional analysis of miRNAs is correlated with the detailed information about their target genes (Sun *et al.*, 2011). Multiple miRNAs can control a single gene expression and a single miRNA can have multiple gene targets (Dehury *et al.*, 2013). miRNAs is involved in the regulation of target gene expression through either mRNA cleavage or translation inhibition which is based on the complementarity between miRNAs and its specific targets (Bartel, 2004). Target mRNA cleavage is the result of perfect pairing between miRNA and its target which mostly shown by plants and imperfect pairing results in translation repression which follows in animals (Hutvágner and Zamore., 2002; Brodersen *et al*, 2008. Here potential 81 target genes were predicted by psRNATarget server. All the predicted miRNAs except Iba-miR5 and Iba-miR12 find more than one potent target genes. This reveals the importance of the computationally predicted sweet potato miRNAs. Functional analysis was performed by BLASTX against non redundant protein database and BLAST2GO. The BLASTX analysis results reveals that the predicted miRNAs target different transcription factors, proteins, enzymes in diverse signalling pathways, metabolic networks that are needed for the plant growth and development. Single predicted miRNAs were characterized with more than one target. Iba-mir1 has 4 targets but 2 of them possess function in signalling and metabolism. Similarly Iba-miR2, Iba-miR3, Iba-miR4, Iba-mir6, Iba-miR7, Iba-miR8, Iba-miR9, Iba-miR10, Iba-miR11, Iba-miR13 were characterized with more than one target genes. Some target genes identified was uncharacterized.

Almost all of the predicted sweet potato miRNAs posses more than one target genes and they regulate gene expression via cleavage mechanism. Most target genes posse's potent functions in various biological processes in plant growth and development and few are uncharacterized in plants. Similarly the BLAST2GO analysis organizes the target genes according to cellular component, molecular function and biological process.

5.3 SWEET POTATO lncRNA PREDICTION BY RNAPLONC

The most important emerging powerful gene regulator among ncRNA family is long non coding RNAs. Long non coding RNAs are characterized by having 200 nucleotide long sequences with no protein coding function (Ng *et al.*, 2013). Some lncRNAs have ORF but doesn't synthesize functional polypeptides. These classes of lncRNAs possess important regulatory action in many essential biological processes in plants (Liu *et al.*, 2015).

Major challenge in lncRNA research is the availability of limited strategies for the efficient prediction. Today next generation sequencing offers wide analysis of whole transcriptome of organisms and this information can be used to identify novel lncRNAs in diverse species (Singh, *et al.*, 2017). The availability of bioinformatic tools for lncRNA prediction is very less in plants. So it is prerequisite for the development of best tool for accurate prediction of lncRNAs.

The current study which uses RNAplonc which is a powerful plant lncRNA identification tool (Negri *et al.*, 2017). The tool which uses a series of 16 features for lncRNA prediction in plants based on mRNA transcript data. The lncRNA prediction features includes (1) GC content, (2) AACG, (3) CCGT, (4) CGCA, (5) CGCT, (6) CGGG, (7) CGTA, (8) TACC, (9) TACG, (10) TCCG, (11) TCGC, (12) Sequence length, (13) score, (14) cdsStop, (15) cdsSizes and (16) cdsPercent which discriminates lncRNAs from all mRNAs.

For potential lncRNA prediction all the available 37698 ESTs of sweet potato were taken as input sequence for RNAplonc. The tool which use a series of steps for lncRNA prediction. First of all the 200nt less sequences from input sequences were removed which results in 35633 sequences followed by clustering with CD-HIT EST to remove sequence redundancy. The results after clustering

analysis include 17944 sequences. Since lncRNAs doesnt code for proteins txCdsPredict program were performed for ORF prediction which results in 17880 sequences with predicted ORF in that all the sequences with a score value > 800 is regarded as protein others as lncRNAs. Based on deep analysis by RNAplonc, potential 9215 lncRNAs were predicted from 8665 protein coding genes in a precise and cost effective manner in sweet potato. This made us clear that long non coding RNAs have that much importance in plants like messenger RNAs. Since there is lack of lncRNA annotation tool the functional analysis of predicted lncRNAs is quiet difficult. But researches till date suggested that the identified lncRNAs regulate gene expression by acting as a chromatin modifier, a scaffold, a guide, a signal and as a decoy. Recently traditional reverse genetics like over-expression and RNAi as well as popular CRISPR/cas9 genome editing may also express their willingness in functional analysis.

5.4. EXPERIMENTAL VALIDATION OF miRNA

RT-qPCR was done for the validation of computationally predicted miRNAs and lncRNAs in sweet potato which is the most efficient and reliable method in gene expression. Three different varieties of sweet potato Sree kanaka, ST13, Khanjakad which possess dark orange, purple and white coloured tubers which is available at ICAR-CTCRI, were selected for validation of Iba-miR1, Iba-miR2, Iba-lnc1 and Iba-lnc2. Young leaves samples were collected from all the varieties for expression analysis. The level of miRNA and lncRNA expression was deduced from C_T.

Here the purple variety ST13 were taken as calibrator and compare the fold of expression of predicted miRNAs and lncRNAs in Khanjakad and Sree Kanaka based on delta delta C_T method. It was found that Iba-miR1 gene is increased by 65.79 times in Khanjakad variety and 11.79 times increased in Sree Kanaka variety with respect to the calibrator. Similarly the expression of Iba-miR2 gene is increased by 77.17 times in Khanjakad variety and 2.05 times increased in Sree Kanaka. When considering the expression of Iba-lnc1 the gene is increased by 7.11 times in Khanjakad variety and 6.02 times the gene is expressed in Sree Kanaka. And finally Iba-lnc2 gene expression is increased 64

93

times in Khanjakad variety and the gene is expressed 22.62 times increased in Sree Kanaka with respect to the calibrator.

This reveals that the expression of predicted miRNAs (Iba-miR1 and Iba-miR2) and predicted lncRNAs (Iba-lnc1 and Iba-lnc2) is more in the white tuber variety Khanjakad. It was predicted here that the miRNA, Iba-miR1 has its role in metabolic pathways and signal transduction and Iba-miR2 is involved in defence mechanisms and ATP synthesis. Similarly, the validation results of lncRNAs, Iba-lnc1 and Iba-lnc2 by real time PCR proved the accurate prediction of lncRNAs by RNAplonc in sweet potato. Because it is clear that the C_T value of validated lncRNAs is low which is an indicator of higher lncRNA expression in selected sweet potato varieties. This made us clear that long non coding RNAs have that much importance in sweet potato. Due to the lack of functional annotation tool for plant lncRNAs, functional characterization of predicted lncRNAs is quiet difficult. Both this small genome regulator and long genome regulator can have great potential to contribute more in the field of crop improvement in future.

SUMMARY

95

6. SUMMARY

The study entitled "Evaluation of miRNA prediction tools and *in silico* analysis of micro and long non coding RNAs in sweet potato (*Ipomoea batatas* L.)" was conducted at the ICAR-CTCRI, Sreekariyam, Thiruvananthapuram during October 2017 to August 2018. The objectives of the study is to compare different miRNA and target prediction tools and *in silico* analysis of the miRNA and lncRNA in sweet potato. The salient findings of the study are summarized below.

MicroRNA identification tools such as NOVOMIR, miRPlant and target prediction tools such as psRNATarget and miRanda were compared. NOVOMIR was found to be a better tool compared to miRPlant for miRNA identification. Similarly psRNATarget was found to be a better tool for target prediction compared to miRanda. Both NOVOMIR and psRNATarget shows higher accuracy, specificity and sensitivity compared to miRPlant and miRanda.

A total of 37,698 available ESTs of *Ipomoea batatas* were analysed and a set of 4853 over lapping sequences known as contigs and 15256 non-overlapping sequences known as singletons sequences from the EST sequences were assembled using the CAP3 sequence assembly program. NOVOMIR which is a pre-miRNA prediction tools uses the available contigs and singletons and predicts potential 343 pre-miRNAs from contigs and 698 pre-miRNA from singletons sequences. A total of 13 potential putative miRNAs in sweet potato were found out using BLAST algorithm in miRBase and the secondary structure of all those predicted miRNAs were analysed by RNA::Hairpinfigure. All the newly identified pre-miRNAs length ranges from 125nt to 282nt. About 81 potential target genes were identified for the predicted miRNAs by psRNATarget. They include transcription factors, metabolic enzymes; defense proteins etc. by BLASTX and BLAST2GO functional annotation were done and organize the genes according to gene ontology like cellular component ontology, molecular function ontology and biological process ontology. Most of the predicted sweet potato miRNAs were shown to have multiple targets but two of the miRNAs failed to show targets in sweet potato. The predicted miRNAs having higher

negative minimum fold energies (MFEs) which ranges from -30.70 to -98.80 Kcal/mol, AMFE ranges from 22.94 to 46.16, and MFEI ranges from 0.6 to 1.0.

About 37,698 available ESTs of *Ipomoea batatas* were analysed by RNAplonc for potential lncRNA prediction. RNAplonc found out 9215 lncRNAs and 8665 mRNAs from the input sequences in sweet potato.

Two miRNAs (Iba-miR1 and Iba-miR2) and two lncRNAs (Iba-lnc1 and Iba-lnc2) were randomly selected and their presence was validated in 3 different varieties of sweet potato such as Sree Kanaka, ST13, Khanjakad available at ICAR-CTCRI by RT-qPCR. The miRNAs and lncRNAs showed amplification in both the samples during qPCR analysis indicating their presence in all the 3 varieties of sweet potato. The increased level of selected gene expression that means, the expression of Iba-miR1 and Iba-miR2, Iba-lnc1 and Iba-lnc2 was more in Khanjakad white tuber variety. Overall study proves that both miRNAs and lncRNAs can have great importance in crop improvement.

REFERENCES

7. REFERENCE

Adossa, A. N. 2016. Computational functional prediction of novel long noncoding RNA in TCGA glioblastoma multiforme sample. M. Sc. thesis, University of Tampere, 81p.

Agarwal, V., Bell, G. W., Nam, J., and Bartel, D. P. 2015. Predicting efective microRNA target sites in mammalian mRNAs. *Elife* 4: e05005.

Ambros V., Bartel B., and Bartel D. P. 2003. A uniform system for microRNA annotation. *RNA.* 9: 277–279.

Amor, B. 2009. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Res.* 19: 57–69.

An, J., Lai, J., Sajjanhar, A., Lehman, M. L., and Nelson, C. C. 2014. miRPlant: an integrated tool for identification of plant miRNA from RNA sequencing data. *BMC Bioinforma.* 15: 275.

Archak, S. and Nagaraju, J. 2007. Computational prediction of rice (Oryza sativa) miRNA targets. *Geno. Prot. Bioinforma.* 5: 196-206.

Bao, N., Lye, K. W., and Barton, M. K. 2004. MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome. *Dev. Cell.* 7: 653–662.

Bardou, F., Ariel, F., Simpson, C. G., Romero-Barrios, N., Laporte, P., Balzergue, S., Brown, J.W., and Crespi, M. 2014. Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. *Dev. Cell* 30: 166–176.

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

Betel, D., Wilson, M., Gabow, A., Marks, D.S. and Sander, C. 2008. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* 36: D149-153.

Bonnet, E., He, Y., Billiau, K. and Van de Peer, Y. 2010. TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinforma.* 26 (12): 1566–1568.

Brannan, C. I., Claire, D. E., Ingram, R. S., and Tilghman, S. M. 1990. The product of the H19 gene may function as RNA. *Mol. and cellular biol.* 10: 28-36.

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe,V. M., Norris, D. P., Cooper, P. J., Swift, S., and Rastan, S. 1992. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved orf and located in the nucleus. *Cell.* 71: 515–526.

Brodersen, P., Achard, L., Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L., and Voinnet, O. 2008. Widespread translational inhibition by plant miRNAs and siRNAs. *Sci.* 320: 1185–1190.

Brown, J. 2016. Agrobacterium tumefaciencs mediated transformation of sweet potato (*Ipomea batatas*) tuber and regeneration of transformed tissue. M. Sc. thesis. Univ. of the Witwaters- rand, Johannesburg.

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J. L. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25: 1915–1927.

Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R., Lao, K.Q., Livak, K.J. and Guegler, K.J. 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* 33: 179.

Chen, L., Wang, T., Zhao, M., Tian, Q., and Zhang, W. H. 2012. Identification of aluminum-responsive microRNAs in *Medicago truncatula* by genome-wide high-throughput sequencing. *Planta.* 235: 375–386.

Dai, X. and Zhao, P. X. 2011. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39: W155–W159.

Dehury, B., Panda, D., Sahu, J., Sahu, M., Sarma, K., Barooah, M., Sen, P., and Modi, M. 2013. *In silico* identification and characterization of conserved miRNAs and their target genes in sweet potato (*Ipomoea batatas L.*) expressed sequence tags (ESTs). *Plant Signal Behav.* 8(12): e26543.

Di, C., Yuan, J., Wu, Y., Li, J., Lin, H., Hu, L., Zhang, T., Qi, Y., Gerstein, M. B., Guo, Y., and Lu, Z. J. 2014. Characterization of stress-responsive lncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. *Plant J.* 80: 848–861.

Du, P., Wu, J., Zhang, J., Zhao, S., Zheng, H., Gao, G., Wei, L., and Li, Y. 2011. Viral infection induces expression of novel phased microRNAs from conserved cellular microRNA precursors. *PLoS Pathogens.* 7: e1002176.

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. 2003. MicroRNA targets in drosophila. *Genome Biol.* 5(1): R1.

Evers, M., Huttner, M., Dueck, A., Meister, G., and Engelmann, J. C. 2015. miRA: adaptable novel miRNA identification in plants using small RNA sequencing data. *BMC Bioinforma.* 16: 370.

Fan, W., Zhang, M., Zhang, H., and Zhang, P. 2012. Improved tolerance to various abiotic stresses in transgenic sweet potato (*Ipomoea batatas*) expressing spinach betaine aldehyde dehydrogenase. *PLoS ONE.* 7(5): e37344.

FAOSTAT [Food and Agriculture Organization Corporate Statistical Database]. 2016.

Frazier, T. P., Xie, F., Freistaedter, A., Burklew, C. E., and Zhang, B. 2010. Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum*) *Planta.* 232: 1289–1308.

Gentile, A., Dias, L. I., Mattos, R. S., Ferreira, T. H., and Menossi M. 2015. MicroRNAs and drought responses in sugarcane. *Front. Plant Sci.* 6:58.

Gong, Y., Huang, H. T. and Liang, Y. 2017. lncRNA-screen: an interactive platform for computationally screening long non-coding RNAs in large genomics datasets. 18: 434.

Gupta, P. K. 2015. MicroRNAs and target mimics for crop improvement. *current sci.*108(9): 1624-1633.

Gurmu, F., Hussein, S., and Laing, M. 2014. The potential of orange-fleshed sweet potato to prevent vitamin A deficiency in Africa. *Int. J. for Vitam. and Nutr. Res.* 84: 65-78.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J. P., Cabili, M. N., Jaenisch, R., Mikkelsen, T. S., Jacks, T., Hacohen, N., Bernstein, B. E., Kellis, M., Regev, A., Rinn, J. L., and Lander, E. S. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nat.* 458: 223–227.

Han, J., Kong, M. L., Xie, H., Sun, Q. P., Nan, Z. J., Zhang, Q. Z., and Pan, J. B. 2013. Identification of miRNAs and their targets in wheat (*Triticum aestivum L.*) by EST analysis. *Genet. Mol. Res.* 12 (3): 3793-3805.

Han, S., Liang, Y., Li, Y., and Du, W. 2016. "Lncident: A Tool for Rapid Identification of Long Noncoding RNAs Utilizing Sequence Intrinsic Composition and Open Reading Frame Information." *Int. J. of Genomics.* 11: 251-253.

Heikham, R. and Shankar, R. 2010. Flanking region sequence information to refine microRNA target prediction. *J. Biosci.* 35: 105-118.

Heo, J. B. and Sung, S. 2011. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Sci.* 331: 76–79.

Hou, M., Tang, X., Tian, F., Shi, F., Liu F., and Gao, G. 2016. AnnoLnc: a web server for systematically annotating novel human lncRNAs. *BMC Genomics* 17(1): 931.

Huang, T., Fan, B., Rothschild, M. F., Hu, Z., Li, K., and Zhao, S. 2007. MiRFinder: an improved approach and software implementation for genome-wide fast microRNA precursor scans. *BMC Bioinforma.* 8: 341.

Huttenhofer, A. and Vogel, J. 2006. Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res.* 34(2): 635–646.

Hutvágner, G. and Zamore, P. D. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Sci.* 297: 2056-2060.

Jaarsveld, P. J., Faber, M., Tanumihardjo, S. A., Nestel, P., Lombard, C. J., and Benade, A. J. 2005. β-Carotene–rich orange-fleshed sweet potato improves the vitamin A status of primary school children assessed with the modified-relative-dose-response test. *The Am. J. of Clin. Nutr.* 81: 1080–

1087.

Jha, A. and Shankar, R. 2011. Employing machine learning for reliable miRNA target identification in plants. *BMC Genomics*. 12: 636.

Kamthan, A., Chaudhuri, A., Kamthan, M., and Datta, A. 2015. Small RNAs in plants: Recent development and application for crop improvement. *Front Plant Sci.* 6: 208.

Kimura, T., Aikata, H., Takahashi, S., Takahashi, I., Nishibuchi, I., Doi, Y., Kenjo, M., Murakami,Y., Honda, Y., Kakizawa, H., Awai, K., Chayama, K., and Nagata, Y. 2015. Stereotactic body radiotherapy for patients with small hepatocellular carcinoma ineligible for resection or ablation therapies. *Hepatol. Res.* 45(4): 378–386.

Kong, L., Zhang, Y., Ye, Z., Liu, X., Zhao, S., Wei, L., and Gao, G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine, *Nucleic Acids* Res. 35: 345–349.

Kornblihtt, A. R. 2014. A Long Noncoding Way to Alternative Splicing in Plant Development. *Developmental Cell.* 30 (2): 117-119.

Kozomara, A. and Griffiths-Jones, S. 2010. miRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39: D152–157.

Kramer, M. F. 2011. Stem loop RT qPCR for miRNAs. *Curr. Protoc. Mol. Biol.* 15 (10).

Kurihara, Y. and Watanabe, Y. 2004. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci.* 101: 12753–12758.

Lagana, A., Ferro, A., and Croce, C. M. 2015. Bioinformatics of non-coding RNAs with applications to biomedicine: recent advances and open challenges. *Front. Bioeng. Biotechnol.* 3: 156.

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

103

Li, Y. F., Zheng, Y., Addo-Quaye, C., Zhang, L., Saini, A., Jagadeeswaran, G., Axtell, M. J., Zhang, W., and Sunkar, R. 2014. Transcriptome-wide identification of microRNA targets in rice. *Plant J.* 62(5): 742–59.

Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B., and Bartel, D. P. 2003a. Vertebrate microRNA genes. *Sci.* 299: 1540.

Lim, L.P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., Burge, C. B., and Bartel, D. P. 2003b.The microRNAs of *Caenorhabditis elegans. Genes Dev.* 17: 991-1008.

Liu, D., Li, Y., Luo, G., Xiao, X., Tao, D., Wu, X., Wang, M., Huang, C., Wang, L., Zeng, F., and Jiang, G. 2017. LncRNA SPRY4-IT1 sponges miR-101-3p to promote proliferation and metastasis of bladder cancer cells through up-regulating EZH2. *Cancer Lett.* 388: 281-291.

Liu, H. H., Tian, X., Li, Y. J., Wu, C. A., and Zheng, C. C. 2008. Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana. RNA-a Publ. of the RNA Soc.* 14: 836-843.

Liu, X., Hao, L., Li, D., Zhu, L., and Hu, S. 2015. Long non-coding RNAs and their biological roles in plants. *Genomics Proteomics Bioinforma.* 13(3): 137–147.

Livak, K. J. and Schmittgen, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C (T)) Method. *Methods.* 25(4): 402- 408.

Lu, C., Tej, S. S., Luo, S. J., Haudenschild, C. D., Meyers, B. C., and Green, P. J. 2005a. Elucidation of the small RNA component of the transcriptome. *Sci.* 309: 1567 –1569.

Ma, L., Bajic, V. B., and Zhang, Z. 2013. On the classification of long non-coding RNAs. *RNA Biol.* 10(6): 1-10.

Mallory, A. C., Dugas, D.V., Bartel, D. P., and Bartel, B. 2004. Micro-RNA regulation of NAC-domain targets is required for proper formation and separation of adiacent embryonic, vegetative, and floral organs. *Curr. Biol.* 14: 1035–1046.

Marin, E., Jouannet,V., Herz, A., Lokerse, A. S., Weijers, D., Vaucheret, H., Marin, E., Jouannet,V., Herz, A., Lokerse, A. S., Weijers, D., Vaucheret, H., Nussaume, L., Crespi, M. D., and Maizel, A. 2010. miR390, Arabidopsis *TAS3* tasiRNAs, and their *AUXIN RESPONSE FACTOR* targets define an autoregulatory network quantitatively regulating lateral root growth. *Plant Cell.* 22: 1104-1117.

Medina, C., Rocha, M., Magliano, M., Ratpopoulo, A., Revel, B., Marteu, N., Magnone, V., Lebrigand, K., Cabrera, J., Barcala, M., Silva, A. C., Millar, A., Escobar, C., Abad, P., Favery, B., and Jaubert-Possamai, S. 2017. Characterization of microRNAs from Arabidopsis galls highlights a role for miR159 in the plant response to the root-knot nematode Meloidogyne incognita. *New Phytol.* 216(3): 882-896.

Mishra, N. S. and Mukherjee, S. K. 2007. A Peep into the plant miRNA world. *The Open Plant Sci. J.* 1: 1-9.

Monavar, F. A., Mohammadi, S., Frazier, T. P., Abbasi, A., Abedini, R., Farsad, L. K., Ehya, F., Salekdeh, G. H., and Mardi, M. 2012. Identification and validation of Asteraceae miRNAs by the expressed sequence tag analysis. *Gene.* 493: 253–259.

Montgomery, T. A., Yoo, S. J., Fahlgren, N., Gilbert, S. D., Howell, M. D., Sullivan, C. M., Alexander, A., Nguyen, G., Allen, E., Ahn, J. H., and Carrington, J. C. 2008. AGO1-miR173 complex initiates phased siRNA formation in plants. *Proc. Natl. Acad. Sci.* 105: 20055–20062.

Motsa, N., Modi, M., Albert, T., and Tafadzwanashe, M. 2015. Sweet potato (*Ipomoea batatas* L.) as a drought tolerant and food security crop. *S. Afr. J. of Sci.* 111(12): 1-8.

Mwanga, R. M., Niringiye, C., Alajo, A., Kigozi, B., Namakula, J., Mpembe, I., Tumwegamire, S., Gibson, R.W., and Yencho, G. C. 2011. 'NASPOT 11', a sweet potato cultivar bred by a participatory plant-breeding approach in Uganda. *Hortic. Sci. 46: 317–321.*

105

Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., Voinnet, O., and Jones, J. D. 2006. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Sci.* 312: 436-439.

Negri,T. D. C., Alves, W. A. L., Bugatti, P. H., Saito, P. T. M., Domingues, D. S., and Paschoal, A. R. 2017. Pattern recognition analysis on long noncoding RNAs: a tool for prediction in plants. *Brief Bioinforma.* 34: 1–8.

Ng, S. Y., Bogu G. K., Soh B. S., and Stanton L. W. 2013. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol. Cell.* 51: 349–359.

Ngailo, S., Shimelis, H., Sibiya, J., and Mtunda, K. 2013. Sweet potato breeding for resistance to sweet potato virus disease and improved yield: Progress and challenges. *Afr. J. Agric. Res.* 8(25): 3202-3215.

Numnark, S., Mhuantong, W., Ingsriswang, S., and Wichadakul, D. 2012. C-mii a tool for plant miRNA and target identification. *BMC Genomics* 13: S16.

Ohlrogge, J. B. and Christoph Benning, C. 2000. Unraveling plant metabolism by EST analysis. *Curr. opinion in plant biol.* 3: 224–228.

Panda, D., Dehury, B., Sahu, J., Barooah, M., Sen, P., and Modi, M. K. 2014. Computational identification and characterization of conserved miRNAs and their target genes in garlic (Allium sativum L.) expressed sequence tags. *Gene* 537: 333–342.

Pian, C., Zhang, G., Chen, Z., Chen, Y., Zhang, J., and Yang, T. 2016. LncRNApred: Classification of Long Non-Coding RNAs and Protein-Coding Transcripts by the Ensemble Algorithm with a New Hybrid Feature. *PLoS ONE* 11(5): e0154567.

Prabu, G. R. and Mandal, A. K. 2010. Computational identification of miRNAs and their target genes from expressed sequence tags of tea (*Camellia sinensis*) *Genom. Prot. Bioinf.* 8(2): 113–121.

Qin, Z., Li, C., Mao, L., and Wu, L. 2014. Novel insights from non-conserved microRNAs in plants. *Front. in Plant Sci.* 5: 586.

Qu, D., Yan, F., Meng, R., Jiang, X., Yang, H., Gao, Z., Dong, Y., Yang, Y., and Zhao, Z. 2016. Identification of MicroRNAs and Their Targets Associated

106

with Fruit-Bagging and Subsequent Sunlight Re-exposure in the "Granny Smith" Apple Exocarp Using High-Throughput Sequencing. *Front. in plant sci.. 7:* 27.

Rao, M. R. S. 2017. *Long non coding RNA biology.* Springer Nat. Bangalore, 15p.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., and Ruvkun, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nat.* 403: 901-906.

Reinhart, B. J., Weinstein, E. G., Rhoades, M.W., Bartel, B., and Bartel, D. P. 2002. MicroRNAs in plants. *Genes Dev.*16: 1616-1626.

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.

Rhoades, M.W., Bartel, D. P., and Bartel, B. 2006. MicroRNAs and their regulatory roles in plants. *Annu.Rev. of Plant Biol.* 57: 19-53.

Schommer, C., Bresso, E. G., Spinelli, S. V., and Palatnik, J. F. 2012. Role of microRNA miR319 in plant development. *Signaling and Commun. in Plants.* 15: 29 – 47.

Seo, J. S., Sun, H. X., Park, B. S., Huang, C. H., Yeh, S. D., Jung, C., and Chua, N. H. 2017. ELF18-INDUCED LONG-NONCODING RNA associates with Mediator to enhance expression of innate immune response genes in Arabidopsis. *Plant Cell* 29: 1024–1038.

Sharma, D., Tiwari, M., Lakhwani, D., Tripathi, R. D., and Trivedi, P. K. 2015a. Differential expression of microRNAs by arsenate and arsenite stress in natural accessions of rice. *Metallomics* 7: 174–187.

Shriram, S., Kumar, V., Devarumath, R. M., Khare, T. S., and Wani, H. 2016. MicroRNAs as potential targets for abiotic stress tolerance in plants. *Front. Plant Sci.* 7: 817.

Singh, U., Khemka, N., Rajkumar, M. S., Garg, R., and Jain, M. 2017. PLncPRO for prediction of long non-coding RNAs (lncRNAs) in plants and its application for discovery of abiotic stress-responsive lncRNAs in rice and chickpea. *Nucleic Acids Res.* 45(22): e183.

107

Sinha, S., Vasulu, T. S., De, R. K. 2009. Performance and evaluation of microRNA gene identification tools. *J. Proteomics Bioinforma.* 2: 336-343.

Sun, Y-H., Lu, S., Shi, R. and Chiang V. 2011. Computational Prediction of Plant miRNA Targets. In: Kodama, H. and Komamine, A. (eds), RNAi and Plant Gene Function Analysis. Springer Protocols, Heidelberg, 175-186pp.

Sunkar R., Girke T., Jain P. K., Zhu, J. K. 2005. Cloning and characterization of microRNAs from rice. *Plant Cell.* 17: 1397–1411.

Swiezewski, S., Liu, F., Magusin, A., and Dean, C. 2009. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nat.* 462: 799–802.

Tavva, S. and Nedunchezhiyan, M. 2011. Global status of sweet potato cultivation. *Fruit, Veg. and Cereal Sci. and Biotechnol.* 6: 143-147.

Teune, J. H. and Steger, G. 2010. NOVOMIR: *De Novo* prediction of microRNA-coding regions in a single plant-genome. *J. Nucleic Acids* : 495904.

Vishwakarma N. P. and Jadeja V. J. 2013 Identification of miRNA encoded by Jatropha curcas from EST and GSS. *Plant Signal Behav.* 8: e23152.

Wang, B., Sun, Y. F., Song, N., Zhao, M., Wang, X. J., Tian, Q., Feng, H., and Huang, L. L. 2013. Identification of UV-B-induced microRNAs in wheat. *Genet. Mol. Res.* 12: 4213–4221.

Wang, J. W., Wang, L. J., Mao, Y. B., Cai, W. J., Xue, H. W., and Chen, X. Y. 2005. Control of root capformation by microRNA-targeted auxin response factors in Arabidopsis. *Plant Cell.* 17: 2204–2216.

Wang, Z. W., Wu, Z., Raitskin, O., Sun, Q., and Dean, C. 2014e. Antisense-mediated FLC transcriptional repression requires the P-TEFb transcription elongation factor. *Proc. Natl Acad. Sci.* 111: 7468–7473.

Wegener, D., Rossi, S., Buffa, F., Delorenzi, M., and Ruping, S. 2013. Towards an environment for data mining based analysis processes in bioinformatics and personalized medicine. *Netw. Model Anal. Health Inform. Bioinforma.* 2: 29-44.

Wierzbicki, A. T., Cocklin, R., Mayampurath, A., Lister, R., Rowley, M. J., Gregory, B. D., Ecker, J. R., Tang, H., and Pikaard, C. S. 2012. Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. *Genes & Dev.* 26: 1825–1836.

Wightman, B., Ha, I., and Ruvkun, G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell.* 75: 855–862.

Wucher, V., Legeai, F., Hédan, B., Rizk, G., Lagoutte, L., Leeb, T., Jagannathan, V., Cadieu, E., David, A., Lohi, H., Cirera, S., Fredholm, M., Botherel, N., Leegwater, P. A., Céline LeBéguec, C., Hille Fieten, H., Jeremy Johnson, J., Alföldi, J., André, C., Lindblad-Toh, K., Hitte, C., and Derrien, T. 2017. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome, *Nucleic Acids Res.* 45: 55-57.

Xie, F., Frazier, T. P., and Zhang, B. 2011. Identification, characterization and expression analysis of microRNAs and their targets in the potato (*Solanum tuberosum*) *Gene.* 473(1): 8–22.

Xin, M., Wang, Y., Yao, Y., Xie, C., Peng, H., Ni, Z., and Sun, Q. 2010. Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 10:123.

Yang, T., Xue, L., and An, L. 2007. Functional diversity of miRNA in plants. *Plant Sci.*172: 423–432.

Yousef M., Nebozhyn M., Shatkay H., Kanterakis S., Showe L. C., Showe M. K. 2009. Combining multi-species genomic data for microRNA identification using a Naive Bayes classifier. *Bioinforma.* 22: 1325–1334.

Zhang, B. and Wang, Q. 2015. MicroRNA- based Biotechnology for Plant Improvement. *J. Cell Physiol.* 230: 1-15.

Zhang, B. H., Pan, X. P., and Anderson, T. A. 2006a. Identification of 188 conserved maize microRNAs and their targets. *FEBS Lett.* 580: 3753-3762.

109

Zhang, B. H., Pan, X. P., Cannon, C. H., Cobb, G. P., and Anderson, T. A. 2006b. Conservation and divergence of plant microRNA genes. *Plant J.* 46: 243–259.

Zhang, B. H., Pan, X. P., Cox, S. B., Cobb, G. P., and Anderson, T. A. 2006b. Evidence that miRNAs are different from other RNAs. *Cell. Mol. Life Sci.* 63: 246–254.

Zhang, B. H., Pan, X. P., Wang, Q. L., Cobb, G. P., and Anderson, T. A. 2005. Identification and characterization of new plant microRNAs using EST analysis. *Cell Res.* 15: 336–360.

Zhang, B. H., Pan, X., and Anderson, T. A. 2006b. Identification of 188 conserved maizemicroRNAs and their targets. *FEBS Lett.* 580: 3753–3762.

Zhang, B., Pan, X., and Stellwag, E. J. 2008. Identification of soybean microRNAs and their targets. *Planta.* 229(1): 161–182.

Zhang, B., Pan, X., Cobb, G. P., and Anderson, T. A. 2006b. Plant microRNA: a Small regulatory molecule with big impact. *Dev Biol.* 289: 3–16.

Zhang, B., Wang, Q., Wang, K., Pan, X., Liu, F., Guo, T., Cobb, G. P., and Anderson, T. A. 2007. Identification of cotton microRNAs and their targets. *Gene.* 397(1): 26-37.

Zhou, J., Huang, Y., Ding, Y., Yuan, J., Wang, H., and Sun., H. 2018. lncFunTK: a toolkit for functional annotation of long noncoding RNAs. *Bioinforma.* 34(19): 3415-3416.

Zhou, L., Liu, Y., Liu, Z., Kong, D., Duan, M., and Luo, L. 2010. Genome-wide identification and analysis of drought-responsive microRNAs in Oryza sativa. *J Exp Bot.* 61: 4157–4168.

Zhu, Q. H. and Wang, M. B. 2012. Molecular functions of long non-coding RNAs in plants. *Genes* 3: 176-90.

110

Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J. A., and Paz-Ares, J. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39(8): 1033-7.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31: 3406-3415.

http://mfold.bioinfo.rpi.edu

http://plantgrn.noble.org/psRNATarget/

http://www.idtdna.com/analyzer/Applications/oligoAnalyzer

http://www.mirbase.org/

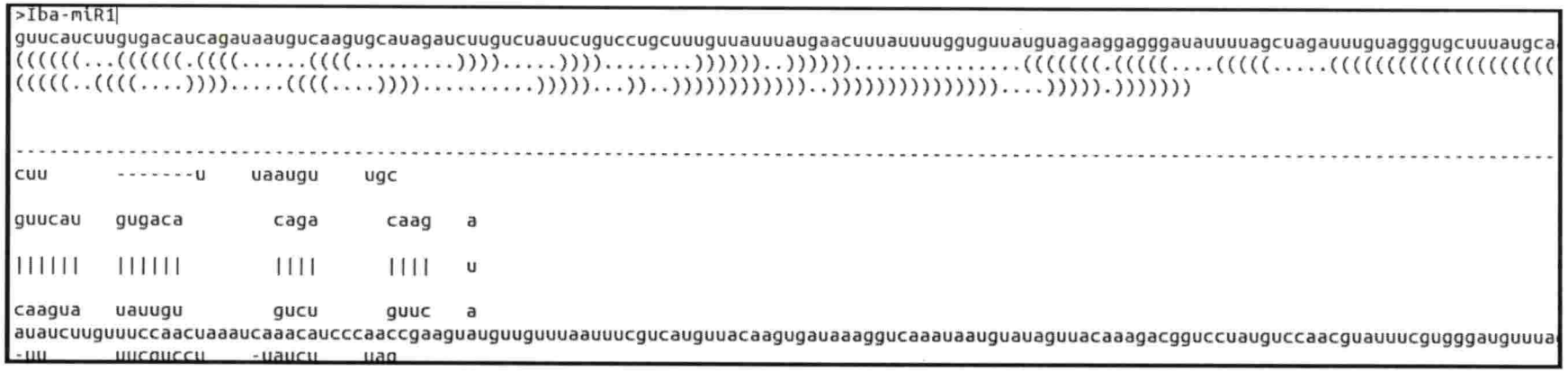
http://www.ncbi.nlm.nih.gov/

APPENDICES

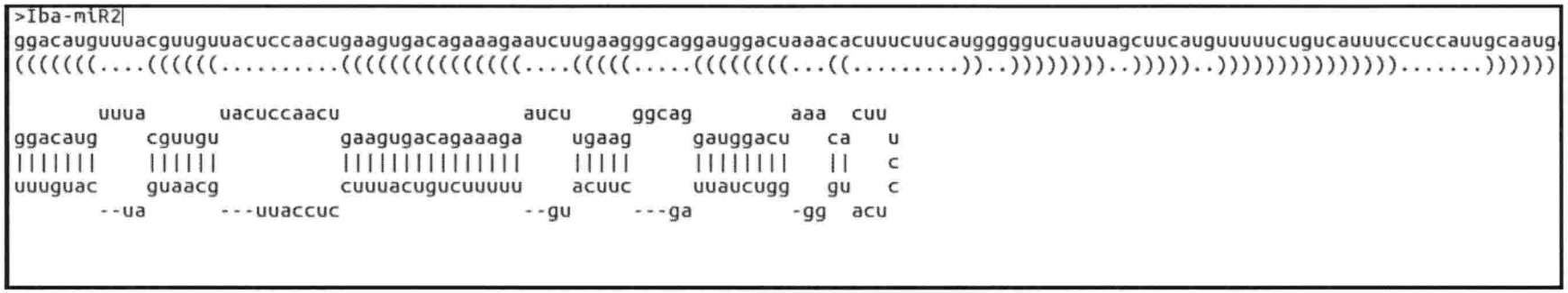
APPENDIX I

Secondary structures of predicted 13 potential miRNAs

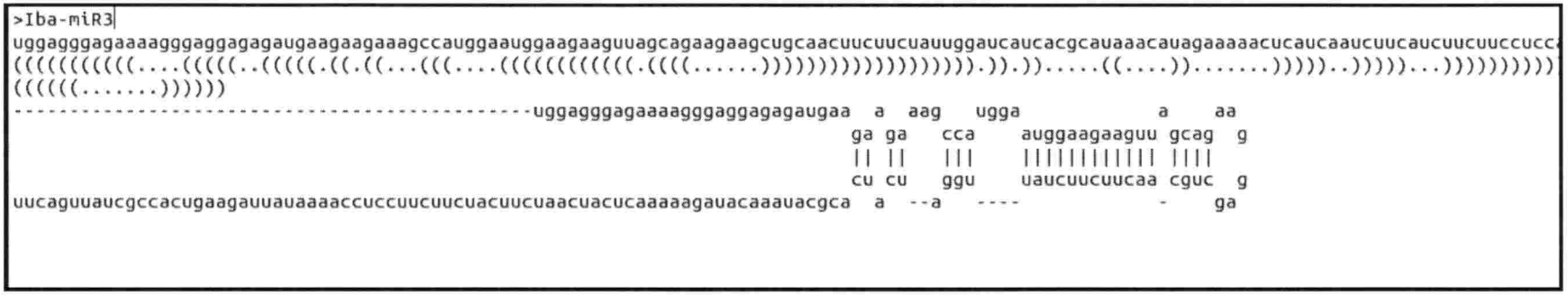
Iba-miR1



Iba-miR2


>Iba-miR2
ggacauguuuacguuguuacuccaacugaagugacagaaagaaucuugaagggcaggauggacuaaacacuuucuucaugggggucuauuagcuucauguuuuucugucauuuccuccauugcaaug
(((((((....((((((..........((((((((((((((....(((((.....(((((((...((.........))..)))))))..))))))..))))))))))))))).......))))))
uuua uacuccaacu aucu ggcag aaa cuu
ggacaug cguugu gaagugacagaaaga ugaag gauggacu ca u
uuuguac guaacg cuuuacugucuuuuu acuuc uuaucugg gu c
--ua ---uuaccuc --gu ---ga -gg acu

Iba-miR3

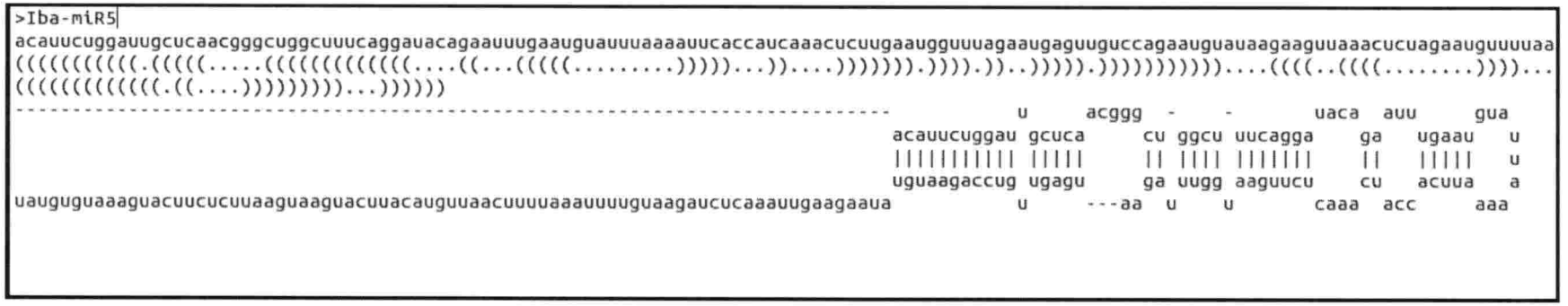

>Iba-miR3
uggagggagaaaagggaggagagaugaagaagaaagccauggaauggaagaaguuagcagaagaagcugcaacuucuucuauuggaucaucacgcauaaacauagaaaaacucaucaaucuucaucuucuuccucc
((((((((((....(((((..(((((.((.((...(((....((((((((((((.((((......)))))))))))))))))).)).))).....((....)).......)))))..)))))...))))))))))
((((((.......))))))
----------uggagggagaaaagggaggagagaugaa a aag ugga a aa
ga ga cca auggaagaaguu gcag g
|| || ||| |||||||||||| |||| |
cu cu ggu uaucuucuucaa cguc g
uucaguuaucgccacugaagauuauaaaaccuccuucuucuacuucuaacuacucaaaaagauacaaauacgca a --a ---- - ga

Iba-miR4

```
>Iba-miR4
ggaggcugagagugaugcuuguguugaugauuuuggggacuuuacuccuagccuggagcucaggaaguucauccccacuguugauacuucuggggauaucuccucuuucccacucaucauuuuccaggua
(((((((..((((.((((((((((((((((....((((((..........((((((((....(((((((((..........))).)))))).((((...........))))..........)))))))).
((((((...((((((....)))))).))))))))))))...))))).))...))))))).....)))))))))))).....)).))).))))..)))))))
       ga    g   -  -----          -uuuu      uuuacuccuagccuggagcucaggaaguucauccccacuguugauacuucuggggauaucuccucuuucccacucauc
ggaggcu  gagu aug cu     uguguugauga     ggggac                                                                                a
|||||||  |||| ||| ||     |||||||||||     ||||||
ucuccgg  cucg uac ga     acauaacuacu     ucccug                                                                                u
       ag    g   a  cuggu          uuacu      cucugugguagucuuucacauaccuuguggggccaagguuccgugugagguggugugaacuuugcccaauggaccuuu
```

115

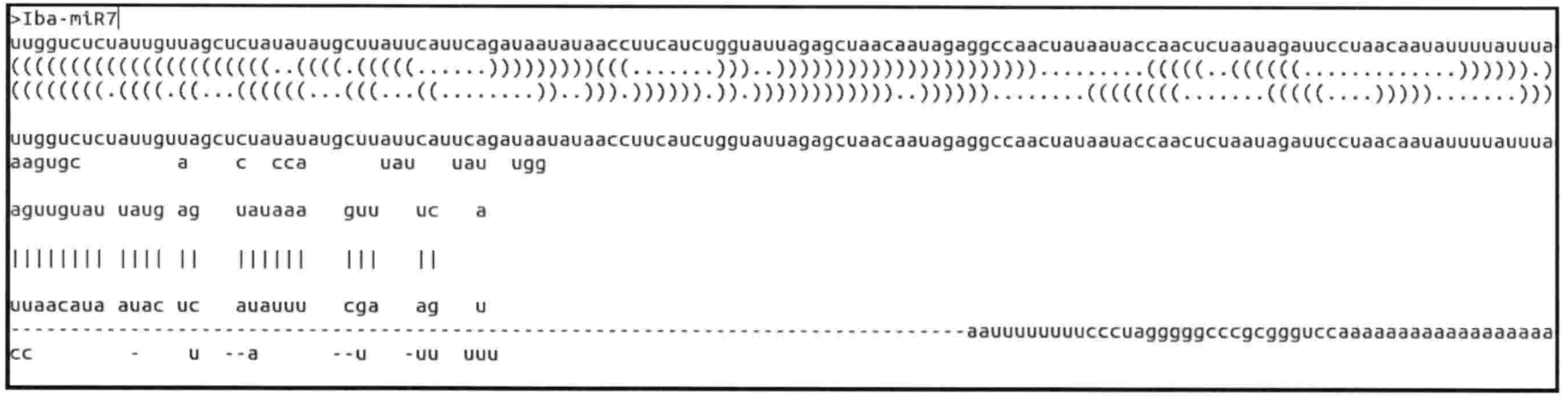
Iba-miR5


>Iba-miR5
acauucuggauugcucaacgggcuggcuuucaggauacagaauuugaauguauuuaaaauucaccaucaaacucuugaaugguuuagaaugaguuguccagaauguauaagaaguuaaacucuagaauguuuuaa
((((((((((.(((((.....((((((((((((....((...(((((.........)))))...))....)))))))..))))..)))))..)))))))))))....((((..((((........))))...
((((((((((((.((....))))))))))...))))))
u acggg - - uaca auu gua
acauucuggau gcuca cu ggcu uucagga ga ugaau u
uguaagaccug ugagu ga uugg aaguucu cu acuua a
uauguguaaaguacuucucuuaaguaaguacuuacauguuaacuuuuaaauuuuguaagaucucaaauugaagaaua u ---aa u u caaa acc aaa

Iba-miR6

```
>Iba-miR6
uucgaauauucgaauaguauucgaaucaaauauacucguauucguauucguauucgaauaguauucgaaucgaaaaaaauauucgaauguuauucaaauccaaauacaaaucaggauagauucgauu
..(((..(((((((((.((((((((((.(((.((.((..(((.(((((...(((.
((((((((((((((.........))))))))))))))).)))...)))))).)))..)).)).))).))))))))))).)))))))..)))).....

---uu   au         g          a   a  c  gu   c     cgu   c               cgaa
     cga  auucgaaua uauucgaauc aau ua uc  auu guauu   auu gaauaguauucgaau    a
     |||  ||||||||| |||||||||| ||| || ||  ||| |||||   ||| ||||||||||||||
     gcu  uaagcuuau auaagcuuag uua au gg  uaa cauaa   uaa cuuauuguaagcuua    a
gccgu   cc         g          c   g  a  ac   a     acc   a               uaaa
```

Iba-miR7


>Iba-miR7
uuggucucuauuguuagcucuauauaugcuuauucauucagauaauauaaccuucaucuggauuagagcuaacaauagaggccaacuauaauaccaacucuaauagauuccuaacaauauuuuauuua
aagugc a c cca uau uau ugg
aguuguau uaug ag uauaaa guu uc a
uuaacaua auac uc auauuu cga ag u
cc - u --a --u -uu uuu

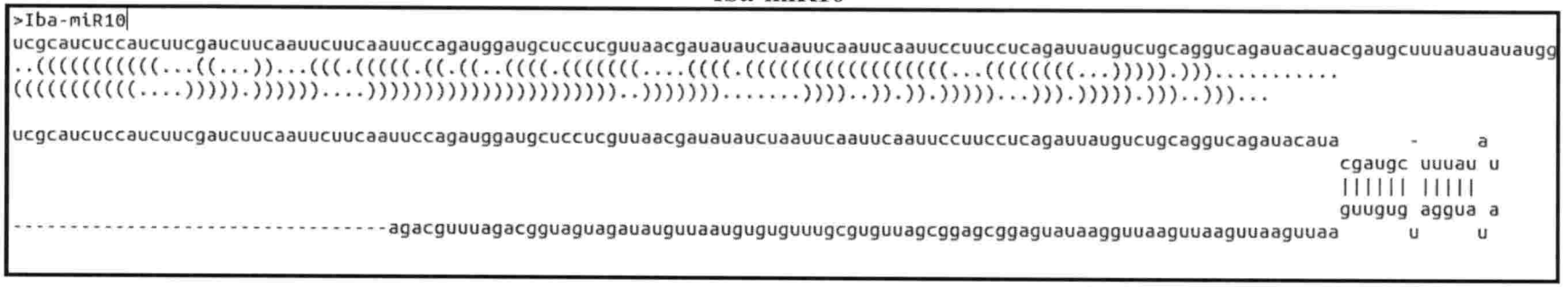
Iba-miR8

```
>Iba-miR8
auauauauauauauauauauauauauauauauauauauuaauaauaaauuaguggguaggauaaagguugguagucuugagacccgugcgucuugagcucuuaacucuguuaucuuuuauccuuuguuauuuucuccua
((((((((((((((((....))))))))))))))))(((((((((......(((((((((((((((((((.((.((..((((......))))..)))).))))).))))))))))..))))))..
(((((........))))).))))))))))............(((((.....(((((..........))))).....)))))

------------------------------auauauauauauauauauauauauauauauauauauauuaauaauaaauuagu     --         -     u  u  ug    cc
                                                                                   gggua  ggauaauag guugg ag cu  agac  g
                                                                                   |||||  ||||||||| ||||| || ||  ||||
                                                                                   ccuau  ucuauuguc caauu uc ga  ucug  c
gacguuauuuacgucuuccucuguugacguuauuuacgucucccuuuguuuuuguuguuguuaauaauccucuuuuauuguuu     uu         u     c  -  gu    cg
```

Iba-miR9

```
>Iba-miR9
caugucuuuuuuuuuucuuuuucuuuuuuuuuuccaguuuugaauugcauucaguaaauggaaaaaaaaaaagaaaaagaaaaaaaagaggucuagugugcaccuugucagucuggcuuuccuuugaagugugggaaaguauug
(((.((((((((((((((((((((((((((((((...((((((...))))))....)))))))))))))).))))))))))))))))))))....)))......(((.((((..((.(((((..........))))).......
((((((....))))))........))..)))).)))..............................

--------------------------------------------------------------------------------------------------------------------   ---g
-             -guu
                                                                                                                     cau    ucuuuuuuuuucuuuuuc
uuuuuuuuuucca    uugaauu
                                                                                                                     |||    ||||||||||||||||||
|||||||||||||    |||||| g
                                                                                                                     gug    ggagaaaaaaaagaaaaag
aaaaaaaaaaggu    gacuuau
uguucaaaaaaaaaaaaaaaaaaaaaaaaaaaacugucguucuaaccuaaguaugguucuuuaaccguuuuguuaugaaaagguguguaaguuuccuuucggucugacuguuccacgu   aucu
a             aaau
```

Iba-miR10


>Iba-miR10
ucgcaucuccaucuucgaucuucaauucuucaauuccagauggaugcuccucguuaacgauauaucuaauucaauucaauuccuuccucagauuaugucugcaggucagauacauacgaugcuuuauauauaugg
..((((((((((...((...)))...(((.(((((.((.((..((((.(((((((....((((.(((((((((((((((((...(((((((...)))))).))).....
((((((((((....)))))).))))))....))))))))))))))))))))..)))))))......))))..)).)).)))))...))).)))))..)))..)))...
ucgcaucuccaucuucgaucuucaauucuucaauuccagauggaugcuccucguuaacgauauaucuaauucaauucaauuccuuccucagauuaugucugcaggucagauacaua
cgaugc uuuau u
|||||| |||||
guugug aggua a
agacguuuagacgguaguagauauguuaaugugugcuuugcguguuagcggagcggaguauaaggguuaaguuaaguuaaguuaa

Iba-miR11

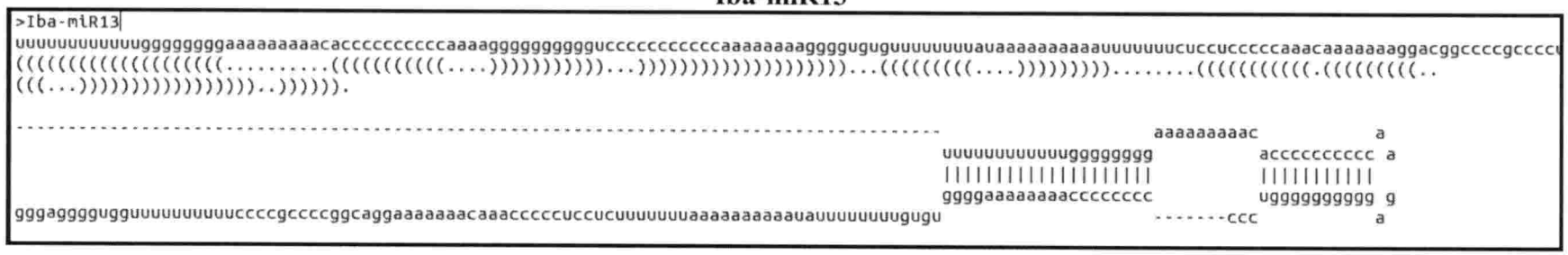
```
>Iba-miR11
gauauagucauauagacuucuggaaggggaaaaaaucccuaacggaauugcugguuaugaucuacaaauccgguugcuuugaaguaaugcuuugagcagaucuggacuguagauuagaugauguuggagaaau
((((..(((((((...(((((..(((((......)))))..)))))..((......(((((((((..(((((.(((.((((((.....)))))))))...)))))..)))))))))......)).....)
gauauagucauauagacuucuggaaggggaaaaaaucccuaacggaauu  ugguua         aa     --u   u      u
                                                  gc      ugaucuaca  uccgg   ugc uugaag a
                                                  ||      |||||||||  |||||   ||| |||||| a
                                                  ug      auuagaugu  agguc   acg aguuuc c
--------------------------uuauuguaguauauaaagaggu  uaguag         -c     uag   -      g
```

Iba-miR12

```
>Iba-miR12
ugagcacccgauuuaucccauuaccgugaaguccgaauuauaccaugcaugacuagaguaugcauugcauugugaacccuagucaaaaauagcauucauauuuuguacuugcauuugacauauuuuguaccugcaguuaacuuuuuu
..(((((.((((......(((................((((((((((...((((((.((....(((((((((((((((..(((((((...(((..((.....))....))).)))))))....((((((.(((...(((((....
((((((((.((((((((((((((..(((...((((((((.(((((((((((((....((((((.((((.(((((((((((.
(((((........))))).))))))))...)))).)))).))))))....))))))))))))).))))))))...)))..))))))))))))))).)))))))).)).)))...)))...))))))........................
(((((.((((.((....)).)))).))))).)))......)))).)))))..

ugagcacccgauuuaucccauuaccgugaaguccgaauuauaccaugcaugacuagaguaugcauugcauugugaacccuagucaaaaauagcauucauauuuuguacuugcauuuugacauau      --c   cag   -
uuuu       u                ag   auc        a            aaac      c    a    ---       a     ucu
                                                                                                                          uuugua   cug   uua ac
uuuuguac uguaaacaaauugaa  cac   acauggua cuguagauacau    uguuaa ugca guac   agaauau uuaac   u
                                                                                                                          ||||||   |||   ||| ||
|||||||| ||||||||||||||  |||   |||||||| ||||||||||||    |||||| |||| ||||   ||||||| |||||
                                                                                                                          agacau   gac   aau ug
aagacaug guauuuguuuaacuu  gug   uguauuau gauguuugugua    acaauu acgu caug   ucuuaua aauug   u
---aguugugguuguuguauguauauacggguuugggauuaccgaagugugguuaacaauaugguaccuggucccaaguaugacguaauacuugggucgaacuagauuuuuaauauauauau      uaa   aua   a  ---
u       u              cu   cau       g           cgua      a    c    uag       c     cug
```

123

Iba-miR13


>Iba-miR13
uuuuuuuuuuuuggggggggaaaaaaaaacacccccccccaaaaggggggggggucccccccccccaaaaaaaaagggguguguuuuuuuuuauaaaaaaaaaauuuuuuuucuccucccccaaacaaaaaaaggacggccccgccccu
(((((((((((((((((((.........((((((((((....))))))))))...)))))))))))))))))))...(((((((((....)))))))))........((((((((((.(((((((((..
(((...)))))))))))))))))..))))))).
aaaaaaaaac
a
uuuuuuuuuuuuggggggggg
acccccccccc a
|||||||||||||||||||
|||||||||||
ggggaaaaaaaaccccccccc
ugggggggggg g
gggaggggugguuuuuuuuuuccccgccccggcaggaaaaaaacaaaccccccuccucuuuuuuuaaaaaaaaaauauuuuuuuugugu
-------ccc
a

APPENDIX II

List of potential target genes and their regulatory action of predicted 13 miRNAs in sweet potato

miRNA Acc.	Target Acc.	Inhibition
Iba-miR1	HX983361.1	Cleavage
Iba-miR1	BU690802.1	Cleavage
Iba-miR1	BU691847.1	Cleavage
Iba-miR1	EE875746.1	Translation
Iba-miR2	JG698134.1	Cleavage
Iba-miR2	HX982678.1	Cleavage
Iba-miR3	DV035054.1	Cleavage
Iba-miR3	EE878783.1	Cleavage
Iba-miR3	DV038112.1	Cleavage
Iba-miR3	DV036794.1	Cleavage
Iba-miR3	BU692365.1	Cleavage
Iba-miR3	JG699416.1	Cleavage
Iba-miR3	DV036197.1	Cleavage
Iba-miR3	DV036251.1	Cleavage
Iba-miR3	EE882805.1	Cleavage
Iba-miR3	CB329912.1	Cleavage
Iba-miR3	EE882624.1	Cleavage
Iba-miR3	DV036703.1	Cleavage
Iba-miR3	DV034550.1	Cleavage
Iba-miR3	CO501253.1	Cleavage
Iba-miR3	EE878566.1	Cleavage
Iba-miR3	CO500906.1	Cleavage
Iba-miR3	EE876161.1	Cleavage
Iba-miR3	JG699599.1	Cleavage
Iba-miR3	DV037504.1	Cleavage

Iba-miR3	DV037246.1	Cleavage
Iba-miR4	HX983063.1	Cleavage
Iba-miR4	DC880535.1	Cleavage
Iba-miR4	JZ128761.1	Cleavage
Iba-miR4	JZ132211.1	Cleavage
Iba-miR4	HX980910.1	Cleavage
Iba-miR4	DC880162.1	Cleavage
Iba-miR4	EE884033.1	Cleavage
Iba-miR4	EE876085.1	Cleavage
Iba-miR4	EE884025.1	Cleavage
Iba-miR4	EE876106.1	Cleavage
Iba-miR4	DV035558.1	Cleavage
Iba-miR4	HX981863.1	Cleavage
Iba-miR4	HX983005.1	Cleavage
Iba-miR4	HX981671.1	Cleavage
Iba-miR4	EE878378.1	Cleavage
Iba-miR4	HX980794.1	Cleavage
Iba-miR4	EE879249.1	Cleavage
Iba-miR6	DV036190.1	Cleavage
Iba-miR6	HX982369.1	Cleavage
Iba-miR6	HX982681.1	Cleavage
Iba-miR7	EE874956.1	Cleavage
Iba-miR7	DC880168.1	Cleavage
Iba-miR7	DC881968.1	Cleavage
Iba-miR8	CB330677.1	Translation
Iba-miR8	HX983981.1	Translation
Iba-miR9	EE875269.1	Cleavage
Iba-miR9	CO500809.1	Cleavage
Iba-miR9	JZ125592.1	Cleavage

Iba-miR10	JG697743.1	Cleavage
Iba-miR10	EE878098.1	Cleavage
Iba-miR11	BU692566.1	Cleavage
Iba-miR11	BU692852.1	Cleavage
Iba-miR11	JZ132217.1	Cleavage
Iba-miR11	JZ127561.1	Cleavage
Iba-miR11	JZ129508.1	Cleavage
Iba-miR11	JZ128833.1	Cleavage
Iba-miR11	DV035135.1	Cleavage
Iba-miR13	DV037597.1	Cleavage
Iba-miR13	GO567368.1	Cleavage
Iba-miR13	DV034533.1	Cleavage
Iba-miR13	DV036153.1	Translation
Iba-miR13	DV036202.1	Cleavage
Iba-miR13	JG699359.1	Cleavage
Iba-miR13	DV034534.1	Cleavage
Iba-miR13	DV037064.1	Cleavage
Iba-miR13	HX983556.1	Cleavage
Iba-miR13	DV035361.1	Cleavage
Iba-miR13	HX982586.1	Cleavage
Iba-miR13	EE880228.1	Cleavage
Iba-miR13	DC880033.1	Cleavage
Iba-miR13	DV036103.1	Cleavage
Iba-miR13	HX982736.1	Cleavage
Iba-miR13	DV036968.1	Cleavage
Iba-miR13	DV035185.1	Cleavage
Iba-miR13	DV036227.1	Cleavage

APPENDIX III

Functional annotation of potential target genes of predicted 13 miRNAs using BLASTX

miRNA Acc.	Target Acc.	Target protein	Target function
Iba-miR1	HX983361.1	L-arabinitol 4-dehydrogenase	Metabolic pathway
	BU690802.1	Uncharacterized	Uncharacterized
	BU691847.1	Uncharacterized	Uncharacterized
	EE875746.1	Mitogen-activated protein kinase kinase 3	Signal transduction
Iba-miR2	JG698134.1	Disease resistance protein RPM1	Protection against plant pathogen
	HX982678.1	ATP synthase subunit O, mitochondrial	ATP synthesis
Iba-miR3	DV035054.1	Uncharacterized	Uncharacterized
	EE878783.1	E3ubiquitin protein ligase	Ubiquitination
	DV038112.1	Uncharacterized	Unchracterized
	DV036794.1	Uncharacterized	Unchracterized
	BU692365.1	Uncharacterized	Unchracterized
	JG699416.1	Pheophorbide oxygenase leaf spot1 like protein	Chlorophyll catabolism
	DV036197.1	Ring H2 finger protein	Plant adaptation to abiotic stress
	DV036251.1	100S acidic ribosomal protein	Transcription factor
	EE882805.1	CMP sialic acid transporter 4 like	Metabolic pathway
	CB329912.1	Gibberellin regulated protein	Gibberellin signalling
	EE882624.1	GTP galactose transporter like	GTP galactose transporter like
	DV036703.1	Uncharacterized	Unchracterized
	DV034550.1	Sporamin A	Stress tolerance
	CO501253.1	Late blight resistance protein	Protection against plant pathogen
	EE878566.1	Mg dependent phosphatase	Membrane lipid modelling
	CO500906.1	Uncharacterized	Unchracterized

Iba-miR3	EE876161.1	Uncharacterized	Unchracterized
	JG699599.1	Ring H2 Finger protein	Plant adaptation to abiotic stress
	DV037504.1	40S ribosomal protein	Transcription factor
	DV037246.1	BAG family molecular chaperone	Stress tolerance
Iba-miR4	HX983063.1	Eukaryotic translation initiation factor	Transcription factor
	DC880535.1	Eukaryotic translation initiation factor	Transcription factor
	JZ128761.1	U4/U6 small nuclear ribonucleoprotein	Pre- mRNA spilicing
	JZ132211.1	U4/U6 small nuclear ribonucleoprotein	Pre- mRNA spilicing
	HX980910.1	Chaperone protein	Stress tolerance
	DC880162.1	Uncharacterized	Unchracterized
	EE884033.1	Acidic leucine rich nuclear phospho protein	Metabolic pathway
	EE876085.1	Uncharacterized	Unchracterized
	EE884025.1	Acidic leucine rich nuclear phosphor protein	Metabolic pathway
	EE876106.1	Uncharacterized	Unchracterized
	DV035558.1	E3 ubiquitin protein ligase	Ubiquitination
	HX981863.1	Protein TIC 100	Embryo development
	HX983005.1	Uncharacterized	Unchracterized
	EE878378.1	Uncharacterized	Unchracterized
	HX980794.1	Mitochondrial amidoxime reducing component	Signal transduction
	HX981671.1	Uncharacterized	Unchracterized
	EE879249.1	Eukaryotic translation initiation factor	Transcription fator

Iba-miR6	DV036190.1	Uncharacterized	Uncharacterized
	HX982369.1	Fructose 1,6-bisphosphatase chloroplastic	Metabolic pathway
	HX982681.1	Psbp domain containing protein	Chloroplastic precursor
Iba-miR7	EE874956.1	DNA directed RNA pol I	Gene silencing
	DC880168.1	Mitochondrial processing peptidase subunit	Metabolic pathway
	DC881968.1	Uncharacterized	Unchracterized
Iba-miR8	CB330677.1	Protein CURVATURE THYLAKOID chloroplastic	Metabolic pathway
	HX983981.1	Uncharacterized	Unchracterized
Iba-miR9	EE875269.1	Fructose-1,6-bisphosphatase	Metabolic pathway
	CO500809.1	Cyclin D3	Transcription regulation
	JZ125592.1	Myosin-4	cell cycle-regulated transport of various organelles and proteins
Iba-miR10	JG697743.1	RHOMBOID-like protein 1	Intramembrane proteolysis
	EE878098.1	Uncharacterized	Unchracterized
Iba-miR11	BU692566.1	GDP mannose 4,6 dehydratase	Metabolic pathway
	BU692852.1	GDP mannose 4,6 dehydratase	Metabolic pathway
	JZ132217.1	Galactokinase	Metabolic pathway
	JZ127561.1	Galactokinase	Metabolic pathway
	JZ129508.1	Galactokinase	Metabolic pathway
	JZ128833.1	Galactokinase	Metabolic pathway
	DV035135.1	Dihydroceramide fatty acyl- 2 hydroxylase FAH1	Metabolic pathway

Iba-miR13	DV037597.1	Formin like protein	Pollen specific
	GO567368.1	H+ transporting ATP Synthase	ATP synthesis
	DV034533.1	Uncharacterized	Unchracterized
	DV036153.1	Uncharacterized	Unchracterized
	DV036202.1	Isopropyl malate dehydratase	Metabolic pathway
	JG699359.1	Tryptophan synthase	Defense response
	DV034534.1	Uncharacterized	Unchracterized
	DV037064.1	Uncharacterized	Unchracterized
	HX983556.1	Elongation factor	Transcription factor
	DV035361.1	Uncharacterized	Unchracterized
	HX982586.1	Alpha carbonic anhydrase	Photosynthesis
	EE880228.1	Uncharacterized	Unchracterized
	DC880033.1	Uncharacterized	Unchracterized
	DV036103.1	N- hydroxyl cinnamoyl benzoyl transferase	Lignin biosynthesis
	HX982736.1	Mitogen activated kinase	Signal transduction
	DV036968.1	Histidine containind phosphotransferase	Signal transduction
	DV035185.1	Metallothionein like protein	Metabolic pathway
	DV036227.1	Uncharacterized	Unchracterized

132

EVALUATION OF miRNA PREDICTION TOOLS AND *IN SILICO* ANALYSIS OF MICRO AND LONG NON CODING RNAs IN SWEET POTATO

Submitted by

ASWATHY M. B.

(2013-09-105)

Abstract of Thesis

Submitted in partial fulfilment of the

requirement for the degree of

B. Sc. – M. Sc. (INTEGRATED) BIOTECHNOLOGY

Faculty of Agriculture
Kerala Agricultural University, Thrissur

B. Sc. - M. Sc. (INTEGRATED) BIOTECHNOLOGY
DEPARTMENT OF PLANT BIOTECHNOLOGY
COLLEGE OF AGRICULTURE
VELLAYANI, THIRUVANANTHAPURAM- 695 522
KERALA, INDIA

2018

9. ABSTRACT

The study entitled "Evaluation of miRNA prediction tools and *in silico* analysis of micro and long non coding RNAs in sweet potato (*Ipomoea batatas L.*)" was conducted at the ICAR-CTCRI, Sreekariyam. The objectives of the study is to compare different miRNA and target prediction tools and *in silico* analysis of the miRNAs and lncRNAs in sweet potato.

The plant miRNA identification tools: NOVOMIR and miRPlant and miRNA-target prediction tools: psRNATarget and miRanda were compared. NOVOMIR and psRNATarget were found to be a better tool in miRNA identification and target prediction.

MicroRNAs (miRNA) are 18-22nt small, endogenous non coding RNA that has prominent role in many biological processes. In the present study, we report the computational prediction of miRNAs and targets from expressed sequence tags (ESTs) of sweet potato. We predicted 13 novel potential miRNAs and 81 potential target genes and functionally characterized by BLASTX and BLAST2GO. The predicted target genes were credited with their role in signalling cascades, metabolism, and defence and stress responses.

Another candidate that has more importance in the genome regulation is lncRNAs. lncRNAs are greater than 200 nucleotide length ncRNA candidate that holds functions at RNA level itself. RNAplonc is a plant long non coding RNA identification tool which uses 16 feature selection methods to predict long non coding RNA molecules. The present study which predicts 9215 lncRNAs and 8665 protein coding genes by RNAplonc in sweet potato for the first time using available ESTs sequences.

Since there is a lack of lncRNA functional annotation tool, the functional analysis of predicted lncRNAs is quiet difficult. From the predicted miRNAs and lncRNAs two miRNAs and two lncRNAs were randomly selected for experimental validation by real time quantitative PCR using three different sweet potato varieties Sree Kanaka, ST13 and Khanjakad available at ICAR-CTCRI and compared the target gene's expression in each variety. Validation results prove that both the miRNAs and lncRNAs show their importance in crop improvement.