EVALUATION OF PREDICTION TOOLS AND COMPUTATIONAL ANALYSIS OF microRNAs IN CASSAVA
(*Manihot esculenta* Crantz.)

By

ANJANA HARIDAS

(2009 - 09 - 102)

THESIS

Submitted in partial fulfilment of the

requirements for the degree of

MASTER OF SCIENCE (INTEGRATED) IN BIOTECHNOLOGY

Faculty of Agriculture

Kerala Agricultural University

M.Sc. (INTEGRATED) BIOTECHNOLOGY COURSE
DEPARTMENT OF PLANT BIOTECHNOLOGY

COLLEGE OF AGRICULTURE

VELLAYANI, THIRUVANANTHAPURAM – 695 522

KERALA, INDIA

2014

DECLARATION

I hereby declare that this thesis entitled "**Evaluation of prediction tools and computational analysis of microRNAs in cassava (*Manihot esculenta* Crantz.)**" is a bonafide record of research work done by me during the course of research and that 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.

Anjana

Vellayani,
Date: 28·11·2014

ANJANA HARIDAS
(2009- 09- 102)

केन्द्रीय कन्द फसल अनुसंधान संस्थान
(भारतीय कृषि अनुसंधान परिषद्)
श्रीकार्यम, तिरुवनन्तपुरम 695 017, केरल
CENTRAL TUBER CROPS RESEARCH INSTITUTE
(Indian Council of Agricultural Research)
SREEKARIYAM, THIRUVANANTHAPURAM 695 017, KERALA, INDIA

CERTIFICATE

Certified that this thesis entitled **"Evaluation of prediction tools and computational analysis of microRNAs in cassava (*Manihot esculenta* Crantz.)"** is a record of research work done independently by Ms. Anjana Haridas (2009-09-102) under my guidance and supervision and that it has not previously formed the basis for the award of any degree, diploma, fellowship or associateship to her.

Sreekariyam
Date: 28-11-2014

Dr. J. Sreekumar
(Chairman, Advisory Committee)
Senior Scientist (Agrl. Statistics)

डॉ. जे. श्रीकुमार / Dr. J. SREEKUMAR
वरिष्ठ वैज्ञानिक (कृषि सांख्यिकी)
Senior Scientist (Agricultural Statistics)
एक्सटेंशन और सामाजिक विज्ञान अनुभाग
Section of Extension and Social Sciences
केंद्रीय कंद फसल अनुसंधान संस्थान
Central Tuber Crops Research Institute
श्रीकार्यम / Sreekariyam
Thiruvananthapuram / तिरुवनंतपुरम-695 017

Telephone :
Laboratory & Office : 2598551 to 2598554
Director (Per.) : 2598431
" Residence : 2441957
Admn. Officer : 2598193
STD Code : 0471
ISD Code : 0091

Telegram : Tubersearch
Fax : (0091) 471-2590063
E-Mail : ctcritvm@yahoo.com
ctcritvm@hotmail.com
Website : http://www.ctcri.org

CERTIFICATE

We, the undersigned members of the advisory committee of Ms. Anjana Haridas, a candidate for the degree of **Master of Science (Integrated) in Biotechnology**, agree that the thesis entitled **"Evaluation of prediction tools and computational analysis of microRNAs in cassava (*Manihot esculenta* Crantz.)"** may be submitted by Ms. Anjana Haridas, in partial fulfilment of the requirement for the degree

Dr. J. Sreekumar
(Chairman, Advisory Committee)
Senior Scientist (Agrl. Statistics)
Division of Extension & Social Sciences
ICAR-CTCRI, Sreekariyam
Thiruvananthapuram

Dr. B. R. Reghunath
(Member, Advisory Committee)
Professor and Head
Department of Plant Biotechnology
College of Agriculture, Vellayani
Thiruvananthapuram

Dr. T. Makesh Kumar
(Member, Advisory Committee)
Principal Scientist
Division of Crop Protection
ICAR-CTCRI, Sreekariyam
Thiruvananthapuram

Dr. S. K. Chakrabarti
(Member, Advisory Committee)
Director
ICAR-CTCRI
Sreekariyam
Thiruvananthapuram

Dr. Deepa S. Nair
(Member, Advisory Committee)
Assistant Professor
Department of Plant Biotechnology
College of Agriculture, Vellayani
Thiruvananthapuram

ACKNOWLEDGEMENT

It gives me immense pleasure to express my deep sense of gratitude to Dr. J. Sreekumar, Senior Scientist, 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. B. R. Reghunath, Professor & HOD, Department of Plant Biotechnology for his valuable suggestions in the preparation of the thesis.

I am deeply indebted to and take this opportunity to express my boundless gratitude to Dr. T. Makesh Kumar, Principal Scientist, CTCRI for his unstinting efforts towards my wet lab experiments while devoting his precious time in the midst of his busy schedule.

My sincere gratitude to Dr. S. K. Chakrabarti, Director, CTCRI for allowing me to be a part of CTCRI and for his constructive perusal of the manuscript.

I am deeply indebted to Dr. K. Rajmohan, former Course Director, M. Sc. (Int.) Biotechnology course, for his valuable suggestions during the project formulation and patient scrutiny of the thesis.

I find it difficult to express my gratitude to Mrs Deepthi D. C. (Deepthi 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 Mrs Prejitha Rinu (Prejitha chechi) for her warm welcome and sisterly affection.

My grateful thanks to Ms. Riji and Mrs Arya Sreejith (Arya chechi) who were always there to lighten the tight schedule of my work.

I acknowledge Anuja and Anand for being always there for me in the computer cell.

I am deeply thankful to Lord Almighty for giving me my family, who are always there to assure me of a dawn after every dark night.

Thank you Mr Anil and family for a good stay during the research work.

I am thankful to all my teachers who shaped me and made me efficient for undertaking this work.

I take this opportunity to thank my friends who are always beside me encouraging and providing moral support.

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

ANJANA HARIDAS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

Continued

LIST OF PLATES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

A	Adenine
AMV	Avian myeloblastosis virus
Acc	Accuracy
BLAST	Basic Local Alignment Search Tool
bp	Base pair
C	Cytosine
CTAB	Cetyl trimethyl ammonium bromide
cDNA	Complementary DNA
CMD	Cassava mosaic disease
C_T	Threshold cycle
DEPC	Diethyl pyrocarbonate
DNA	Deoxy ribonucleic acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylene diamine tetra acetic acid
EST	Expressed sequence tag
FP	False positive
FN	False negative
G	Guanine
kcal mol^{-1}	Kilocalories per mol
M	Molar
mM	millimolar
mRNA	Messenger RNA
miRNA	Micro RNA
MCC	Mathew correlation coefficient
MFE	Minimal free folding energy
MFEI	Minimal free folding energy index
Mol. Wt.	Molecular weight

LIST OF ABBREVIATIONS

Continued

NCBI	National Centre for Biotechnology Information
nt	Nucleotide
PTGS	Post transcriptional gene regulation
PVP	Polyvinyl pyrrolidone
Pri-miRNA	Primary microRNA
Pre-miRNA	Precursor miRNA
PPV	Positive predictive value
qPCR	Quantitative Real Time Polymerase Chain Reaction
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT-PCR	Reverse Transcription- PCR
RISC	RNA induced silencing complex
siRNA	Small interfering RNA
Se	Senstivity
Sp	Specificity
T	Thymine
Tris-HCl	Tris (hydroxy methyl) aminomethane hydrochloride
TP	True positive
TN	True negative
TF	Transcription factor
U	Uracil
µl	microlitre

INTRODUCTION

1. INTRODUCTION

Cassava (*Manihot esculenta*) is grown throughout tropical Africa, Asia and the America for its starchy storage roots, and feeds an estimated 750 million people each day. Cassava is also an excellent energy source - its roots contain 20-40% starch that costs 15-30% less to produce per hectare than starch from corn, making it an attractive and strategic source of renewable energy. Farmers choose it for its high productivity and its ability to withstand a variety of environmental conditions (including significant water stress) in which other crops fail. However, it has very low protein content, and is susceptible to a range of biotic stresses. Cassava Mosaic Disease (CMD) is one such biotic stress. CMD caused by cassava mosaic geminiviruses is one of the most devastating crop diseases affecting cassava cultivation. The primary spread of the disease is through infected planting material and secondary spread is by an insect vector, white fly (*Bemisia tabaci* Genn.). Recombination and pseudorecombination between cassava mosaic geminiviruses give rise to different strains and members of novel virus species with increased virulence causing severe disease epidemics. Various approaches are currently being applied to mitigate these constraints to achieve better cassava varieties. One of the strategies is the potential application of the knowledge of microRNAs (miRNAs) in gene regulation.

MicroRNAs are a newly recognized class of endogenous gene regulators that negatively control gene expression at post transcriptional level by binding to messenger RNA (mRNAs) and either targeting them for degradation or inhibiting protein translation (Bartel, 2004; Dugas and Bartel, 2004). The first plant miRNA was discovered in *Arabidopsis thaliana* (Park *et al.*, 2002; Reinhart *et al.*, 2002) and since then computational and experimental methods have identified thousands of miRNAs in wide range of plant species. MicroRNAs have been shown to be highly evolutionarily conserved from lower mosses to higher flowering eudicots (Zhang *et al.*, 2005).

The majority of plant miRNAs target transcription factors that control gene expression during plant growth and development. Due to their function in gene regulation, they have been shown to play an important role in variety of plant metabolic and biological processes like organ maturation, signal transduction, responses to environmental stresses etc. Plants being sessile organisms have developed complex gene regulatory network to combat environmental stresses. miRNAs are involved in plant tolerance to biotic and abiotic stresses by regulating plant responses to environmental conditions.

A better understanding of the roles of miRNAs in post transcriptional gene silencing in response to biotic and abiotic stresses will be vital in attempt to develop superior stress tolerant cassava varieties. Breakthroughs in this area are likely to reveal developmental regulation and disease mechanisms related to miRNAs.

The present study is undertaken to computationally predict miRNAs and their targets in cassava and cassava mosaic virus and to understand the miRNA-mRNA interaction in cassava in biotic stress response (cassava mosaic virus). MicroRNA target prediction tools were evaluated and compared to understand their performance.

REVIEW OF LITERATURE

2. REVIEW OF LITERATURE

MicroRNAs (miRNAs) are non-coding RNAs with short sequences that negatively regulate gene expression at the post-transcriptional levels by either binding to mRNAs for degradation, or by inhibiting protein translation. MicroRNAs have high evolutionary conservation, from lower mosses to higher flowering plants, and have been shown to play an important role in plants by regulating growth and development, developmental timing, hormone signalling, organogenesis, and response to environmental stresses. The first report of miRNA was in early 1990s. In 1993, it was discovered that *lin-4*, a gene known to control the timing of *Caenorhabditis elegans* larval development, did not code for a protein, but, instead, generated two small RNAs of different size (one of 22 nt and the other of about 61 nt). The longer RNA was predicted to fold into a stem loop structure and was proposed to be the precursor of the shorter one. These *lin-4* RNAs had antisense complementarity to multiple sites in the 3'UTR of the *lin-14* gene. A reduction of the amount of LIN-14 protein without noticeable change in levels of *lin-14* mRNA was observed leading to creation of a model of action where by substantially lin-*4* RNAs (non gene product) pairs to the *lin-14* 3'UTR to specify the translational repression of the *lin-14* mRNA. This negative regulation triggers the transition from cell divisions of the first larval stage to those of the second (Lee *et al.*, 1993).

Subsequently another non-coding RNA was discovered: *let-7* which is involved in the regulation of larval development. *let-7* RNA promotes the transition from late-larval to adult cell fates in the same way that the *lin-4* RNA (Reinhart *et al.*, 2000). Furthermore homologs of the *let-7* gene were soon identified in the human and fly genomes, and *let-7* RNA itself was detected in human, *Drosophila*, and eleven other bilateral animals. Because of their common roles in controlling the timing of developmental transitions initially the *lin-4* and *let-7* RNAs were called *small temporal RNAs* (stRNAs) (Pasquinelli *et al.*, 2000) and only later was identified as members of new class of tiny (20-25 nt)

regulatory RNAs. The term microRNA was subsequently used to refer to these stRNAs and to all the other tiny RNAs with similar features but unknown functions (Lau *et al.,* 2001). Small RNA cloning efforts led to identification of many more miRNAs. The first plant miRNA was discovered in *Arabidopsis thaliana* (Reinhart *et al.,* 2002) and since then, computational and experimental methods have identified thousands of miRNAs in a wide range of plant species. The spread and importance of miRNA-directed gene regulation are coming into focus as more miRNAs and their regulatory targets and functions are discovered.

The history of miRNAs serving as gene regulators dates back to more than 400 million years ago. *Chlamydomonas reinhardtii*, a unicellular green algae, has been shown to encode miRNAs (Zhao *et al.* 2007). It is suggested that the miRNA pathway is an ancient mechanism of gene regulation and it occurred prior to the emergence of multicellularity. This also suggests that miRNAs may have a common ancestor in evolution (Zhang *et al.* 2005). Greater evolutionary conservation of miRNAs than siRNAs was proposed (Bartel and Bartel 2003). Computational prediction revealed that many miRNA families were evolutionarily conserved across all major lineages of plants (Zhang *et al.,* 2005; Zhang *et al.,* 2006b). However, the regulation of a given miRNA may not be similar in diverse plant species. The variety of miRNAs must have expanded significantly during evolution of early land plants. Thus, some miRNA families were specific to bryophyte *Physcomitrella*, whereas other miRNA families were specific to higher land plants (Isam *et al.,* 2007). It is indicated that miRNAs have evolved after the divergence between vascular plants and mosses. The evolution of miRNA genes has been accompanied with miRNA functionality change due to the process of genome-wide duplication, tandem duplication, and segmental duplication, followed by dispersal and diversification. The process is similar to the processes that drive the evolution of protein gene families (Maher *et al.,* 2006). It is assumed that in ancient times, miRNA played an important role in plant anti-virus defense, and novel functions came into being only after the basic requirements of survival were satisfied.

2.1 BIOGENESIS OF PLANT miRNAs AND ACTION

The loci encoding miRNAs, termed miRNA genes are located throughout the genomic regions not associated with known protein coding genes (Reinhart *et al.,* 2002). This indicates that most plant miRNAs are derived from their own endogenous genes and form one independent transcript (Lagos-Quintana *et al.,* 2001; Lau *et al.,* 2001). Most plants possess over 100 miRNA genes (MIR) (Nozawa *et al.,* 2012). Studies revealed that the primary single stranded RNA miRNA transcript, termed the pri-miRNA, is transcribed by RNA polymerase II enzymes. These transcripts are usually ~ 1kb in length, polyadenylated, 5' capped, contain introns and typical TATA-box motifs, characteristics of class II transcription (Kurihara and Watanabe, 2004; Xie *et al.*, 2005).

Plant miRNAs are released from pri-miRNAs through at least two sequential processing steps by RNase III enzymes - Dicer. Dicer homologues have been found in most organisms that undergo RNA silencing (Bernstein *et al.*, 2001). Plants have many DCL genes and there appears to be different role for distinct Dicer enzymes in plants. DCL1 is specifically involved in miRNA accumulation and is responsible for both cleavage steps in the nucleus (Park *et al.*, 2002; Reinhart *et al.*, 2002; Schauer *et al.*, 2002). The remaining three are involved in siRNA biogenesis. The protein structure of the DCL enzymes facilitates the processing of dsRNA molecules. They are dimeric proteins with four characteristic domains: a PIWI/ ARGONAUTE/ ZWILLE (PAZ) domain, a RNA helicase domain, two ribonuclease III motifs and one dsRNA binding domain (Blaszczyk *et al.*, 2001).

The processing by DCL1 also requires two other proteins, HYPONASTIC LEAVES1 (HYL1) and SERRATE (SE). HYL1 belongs to a family of dsRNA-binding proteins in *Arabidopsis* (Hiraguri *et al.*, 2005). It has two dsRNA binding domains. SE encodes a C_2H_2 zinc finger protein that plays a general role in biogenesis. These proteins probably facilitate the positioning of DCL1 on the transcript. The processing releases a double stranded short RNA, whose length is

determined by the distance between PAZ domain and the RNase III domains (Macrae *et al.*, 2006).

In the nucleus, the pri-miRNA molecule is cleaved by DCL on each arm of the stem loop to form smaller pre-miRNA precursor (Bernstein *et al.*, 2001). The pre-miRNA molecule folds to form an imperfect hairpin dsRNA stem-loop structure with the mature miRNA located on one arm (Lau *et al.*, 2001). Subsequent DCL cleavage of the molecule on each arm releases a miRNA duplex, containing the mature miRNA and it's near reverse complement (miRNA*), from the pre- miRNA stem loop (Lee *et al.*, 2003). The miRNA and the miRNA* remain together after cleavage, leaving 2-nucleotide 3'-overhangs with 3' hydroxyl and 5' monophosohate ends, characteristics of dicer cleavage products (Elbashir *et al.*, 2001).

It is clear that DCL1 in plants cuts preferentially at specific position in the miRNA stem-loop to release the appropriate mature miRNA molecule (Reinhart *et al.*, 2002). It is hypothesized that correct Dicer processing of miRNA is determined by the structure of the precursor molecule to yield the mature miRNA duplex. The sequence of the molecule has no role in the same (Parizotto *et al.*, 2004). This hypothesis would explain the diversity observed in the sequences of different miRNA which are processed via the same mechanism.

The *Arabidopsis* HUA ENHANCER 1 (HEN1) protein is a dsRNA methylase with two dsRNA binding domains and a nuclear localization signal (NLS) (Park *et al.*, 2002). The protein methylates the 3'- terminal nucleotide of miRNA and is predicted to protect the molecule from uridylation (Chen, 2008). After the miRNA:miRNA* duplex is formed in the nucleus, most plant miRNAs are transported to the cytoplasm. The export of miRNAs is facilitated by HASTY (HST), which is a member of nucleocytoplasmic transporter family of proteins (Park *et al.*, 2005).

Once the mature miRNA molecules are formed and transported out of the nucleus they associate with a ribonucleoprotein complex, termed the RNA

Induced Silencing Complex (RISC). It is shown that the 2 nt 3'overhangs and 5'-phosphate termini of the small RNA are essential requirements for incorporation into RISC. Once RISC is associated with miRNA duplex it is an inactive form, known as RISC Loading Complex (RLC) (Tang, 2005). RISC is activated when miRNA duplex molecules are unwound in RLC, and only one strand accumulates as the mature miRNA, the other arm miRNA* is subsequently degraded. The mature miRNA sequestered in active miRISC, guide it to the target mRNA. Proteins of Argonaute (AGO) gene family are principle components of RISCs. The protein structure includes two conserved region, the PAZ and Piwi domains. The PAZ domain is a RNA-binding domain that binds single stranded RNA at the 3' end of the molecule through a hydrophilic cleft. The Piwi domain is a putative RNase H and is thought to facilitate miRNA mediated mRNA cleavage or provide RISC with slicer activity (Song *et al.*, 2003). Large number of miRNA targets undergoes cleavage. The Piwi domain of AGO protein forms an RNaseH-like fold with a slicer endonuclease activity capable of cleaving RNA targets that are complementary to the loaded guide strand (Liu *et al.*, 2004). Plant miRNAs are highly complementary to targets throughout their length (Fahlgren and Carrington, 2010), and the high degree of complementarity is a requirement for effective target slicing by AGO proteins (Mallory *et al.*, 2004).

MicroRNA targets regulated at the protein level in the absence of noticeable changes in mRNA level have suggested that plant miRNAs also interfere with target mRNA translation. Translational repression is distinct from slicing and is more widespread in plants. The mechanism of translation repression in plants is still unknown. AGO1 and AGO10, the two AGO proteins examined to date with respect to the translational inhibition activity of plant miRNAs, have been shown to be required for this activity of miRNAs (Brodersen *et al.*, 2008). In addition to post transcriptional gene silencing, miRNAs in plants are capable of transcriptional gene silencing. In rice, DCL3-dependent long miRNAs of 24 nucleotides are sorted to AGO4 and trigger cytosine DNA methylation at both MIR and target loci (Wu *et al.*, 2010).

2.2 FUNCTIONAL ROLES OF PLANT miRNAs

Studies showed that many predicted and experimentally confirmed miRNA targets are genes encoding regulatory proteins, showing that miRNAs play a role at the core of gene regulatory networks. Functional genomics studies have shown the involvement of plant miRNAs in many developmental processes (Jones-Rhoades *et al.*, 2006; Jung *et al.*, 2009) and their diverse roles in stress responses (Sunkar *et al.*, 2012). Several developmental programs were detected that involved miRNAs, such as root initiation and development (Montgomery *et al.*, 2008; Marin *et al.*, 2010), vascular development (Yu *et al.*, 2005; Donner *et al.*, 2009), leaf morphogenesis and polarity (Mallory *et al.*, 2004), floral differentiation (Chuck *et al.*, 2008), and phase transition from vegetative growth to reproductive growth (Yang *et al.*, 2007; Yant *et al.*, 2010). Loss-of-function in some miRNA genes and miRNA complementary sites of target genes as well as in the genes related to miRNA biogenesis lead to abnormalities in plant development and growth. The *dcl1* and *hasty* are important genes for plant miRNA biogenesis, and their loss-of-function results in abnormalities in plant growth and development, such as altered leaf morphology, delayed floral transition, female sterility, and early stage embryo arrest (Dugas and Bartel 2004; Zhang *et al.*, 2006c).

Plants are exposed to serious biotic and abiotic stresses such as drought, salinity, alkalinity, cold, pathogen infections, and diseases, which are the predominant causes of decreased crop yields. Plants use adaptive responses operating at the transcriptional, post-transcriptional, translational, and post-translational levels to cope with these environmental challenges (Sunkar, 2012). As a post-transcriptional gene regulator, a number of miRNAs play roles in multiple stress responses in plants. There is evidence showing the direct link between miRNA regulation and stress response in plants. Expression of plant miRNAs has been up- and down-regulated upon treatment with diverse stress conditions (Eldem *et al.*, 2013). In several studies, the roles of small RNA in disease resistance responses were revealed (Navarro *et al.*, 2006; Fahlgren *et al.*,

2007; Jin *et al.*, 2008; Katiyar- Agarwal and Jin 2010; Li *et al.*, 2010; Zhang *et al.*, 2011). Numerous miRNAs have been predicted or validated to be involved in plant defense. For example, nine of the forty eight miRNAs are related to defense in *Physcomitrella.* MiR1-39 targets a gene coding for a mucin-like protein carrying a dense sugar coating against proteolysis, which is a pivotal step in pathogen invasion. miR160-3 acts on intracellular pathogenesis- related protein. miR408 provides defense though interaction with the genes coding for a copper ion binding protein, and with electron transporter or Phytocyanin homolog (Isam *et al.*, 2007). Approximately 70% of 130 miRNA targets were predicted to be involved in the defense response in *Populus* (Lu *et al.*, 2005). Over expression of a plant miRNA (miR393) resulted in the increased bacterial resistance in plants (Navarro *et al.*, 2006). Therefore, it is thought that plant miRNA-directed RNAi or miRNA specified mRNA destruction determines the balance in plant defense system. miR393 is the first reported responsive miRNA upon bacterial inoculation in plants (Navarro *et al.*, 2006). Pérez-Quintero *et al.* (2012) suggest that miRNAs in cassava play a role in defence against *Xanthomonas manihotis*, and that the mechanism is similar to what is known in *Arabidopsis* and involves some of the same families.

2.3 PLANT miRNAs AND VIRUS INFECTION

Plants often encounter various pathogens (bacteria, fungi, virus, and phytoplasma) invasions and show hypersensitive response through a series of resistance mechanisms. Among them, virus infection is a major threat to crops worldwide with loss of billions of US dollars in agricultural productivity every year (Thompson and Tepfer 2010). To protect from viral diseases, plants have developed pathogen- specific defense mechanisms either through pathogenesis-related (PR) proteins or by RNA interference mediated gene silencing (Pantaleo 2011). Over the course of evolution, many viruses have developed sophisticated counter-defensive mechanisms such as post-transcriptional gene silencing (PTGS) suppressor proteins and small interfering (siRNA) and microRNA (miRNA)-mediated RNAi silencing pathways (Singh *et al.*, 2010; Song *et al.*, 2011). A viral

genome can be targeted by a host miRNA, either by specific miRNAs against a particular virus or by fortuitous complementarities with the multitude of miRNAs (Simo´n-Mateo and Garcı´a, 2006). Naturally occurring miRNA in plants participate in viral infection. Indirect evidence for this originated from the observation that *Arabidopsis* mutant *dcl1* showed reduced susceptibility to RCNMV infection (Dunoyer *et al.,* 2004). The primary role of DCL1 is to process pre-miRNAs. Thus it is supposed that viruses not only suppress, but also exploit endogenous miRNA to redirect host gene expression. Plant virus infections resulted in a dramatic increase in miRNA (Du *et al.,* 2011) whereas virus infected vertebrate cells increased siRNA content (Bennasser *et al.,* 2005).

It is also reported that microRNAs (miRNAs) are involved in modulating plant viral diseases (Dunoyer *et al.,* 2004; Carmen and Juan 2006). Plants and invertebrates employ their miRNA in defense against viruses by targeting and degrading viral products (Carl *et al.,* 2013). Endogenous miRNAs exhibit preparative feature. miRNAs that have already existed within a cell before viruses invade help to serve as advance preparation to counteract the infection. miRNA mediated gene silencing exhibits several advantages over other gene silencing strategies: (1) proactive and long-acting, (2) without disruption by a non-target virus, and (3) multiple targeting.

There is a strong potential for antiviral activity of plant miRNAs and the miRNA pathway may be a support mechanism to siRNA pathway in antiviral defence (Pérez-Quintero *et al.*, 2010). Baig and Khan (2013) used bioinformatic approach to search cotton miRNA targets in genome of cotton leaf curl multan virus (DNA A) and betasatellite (DNA β). The study revealed 34 putative miRNA targets in DNA A encoded protein loci and 2 putative targets in DNA β above threshold value. The targeting miRNA may have potential to confer effective resistance against Cotton leaf curl disease infection in cotton. Tripathy and Mishra (2012) used computational approach explore the possibility of endogenous rice miRNAs having role in antiviral defense by targeting the mRNA of different

genes of Tungro viruses. The results highly support that the rice miRNA can resist tungro viruses.

The following areas have the potential for application of miRNAs in plant anti-virus defense: (1) Analyzing the function of viral suppressor in the process of gene silencing. (2) Designing and developing novel miRNA-mediated gene therapy. (3) Modifying plant physiological properties to enhance their anti-virus capacity. (4) Developing loss of function transgenic plants (Lu *et al.*, 2008).

2.4 MICRORNA IDENTIFICATION

Understanding the miRNA-mediated gene regulation is largely dependent on the availability of innovative strategies and methodological approaches for accurate detection of miRNA. Identification of differentially expressed miRNA genes in cell transcriptome directly reflects the dynamic cell behaviour under changing conditions. The miRNA and target mRNA expression level measurement presents valuable information about the miRNA functions. So far, numerous methodologies have been developed for rapid, sensitive, specific, and genome-wide detection of miRNAs. Approaches to discovering miRNAs can be split into two groups. In experiment-driven methods, the expression of small RNAs is first established, and bioinformatics is then used to identify RNAs that meet structural requirements. In computation-driven approaches, candidate miRNA are first predicted in (whole) genome sequences on the basis of structural features, and experimental techniques are then used to validate these predictions by demonstrating expression of the corresponding sequences.

Forward genetics is the classical approach where researchers have a known phenotype, but the DNA sequence (genotype) coding for that particular phenotype is unknown. Forward genetics methods were instrumental in identifying the first miRNA genes, *lin-4* and *let-7* (Berezikov *et al.*, 2006). However to date, there is only one example using a forward genetics experimental approach to identify miRNA in plants (Baker *et al.*, 2005).

Direct cloning of small RNAs from plants is one of the basic approaches of miRNA discovery. Scientists have used this methodology to isolate and clone small RNAs from various plant species such as *Arabidopsis* and rice (Reinhart *et al.,* 2002; Park *et al.,* 2002; Llave *et al.,* 2002; Sunkar *et al.,* 2005). Identification of miRNAs using the direct cloning approach basically involves the creation of a cDNA library and includes six steps: (1) isolation of total RNA from plant tissue, (2) recovery of small RNAs from gel, (3) adaptor ligation, (4) reverse transcription, RT-PCR, (5) cloning, and (6) sequencing methods. Expression of several miRNAs is broad but many of them are detected in certain environmental conditions, at different plant developmental stages and tissues. Therefore specific time points, tissues, and/or biotic and abiotic stressed induced plant samples are used for miRNA cloning (Unver *et al.*, 2009).

The emergence of next-generation high-throughput cDNA and direct-RNA sequencing techniques has revolutionized whole-transcriptome analysis at an unprecedented depth, accuracy, and resolution. New generation sequencing (NGS) technologies have been successfully applied in genome-wide identification and quantification of known and novel miRNAs and other non coding small RNAs in a single instrument run. Third-generation sequencing technologies offer significant advantages in terms of simplified library construction, small amounts of starting material, and longer read lengths. These technologies do not require the conversion of RNA into cDNA or ligation/pre amplification steps. Therefore, they can be effectively used for direct sequencing of RNA without the need for cDNA conversion process causing the cDNA synthesis-based artifacts and pre-amplification experiments leading to biases and errors (Schadt *et al.,* 2010). The deep sequencing approach can easily eliminate some technical challenges and obstacles sourced from intrinsic properties of miRNAs, such as small read size, low-abundance, instability, and contamination with other RNA fragments. *Hevea brasiliensis* (Gébelin *et al.,* 2012), soybean (Kulcheski *et al.,* 2011), *Oryza sativa* (Jian *et al.,* 2010) and sugarcane (Thiebaut *et al.,* 2012) microRNA have been identified by small RNA deep sequencing. Deep sequencing allows us to

determine the miRNAs whose expression profiles could be differentiated under a variety of stress conditions including drought (Barrera-Figueroa *et al.,* 2011; Wang *et al.,* 2011), cold (Zhang *et al.,* 2009), phosphate deficiency (Hsieh *et al.,* 2009) and sulfate deficiency (Huang *et al.,* 2010).

2.5 COMPUTATIONAL PREDICTION OF PLANT miRNA

Once potential miRNA sequences have been cloned and sequenced, the sequence data can be imported into a variety of software programs for computational analysis. These bioinformatics tools search for sequence and structure conservation of miRNAs (Lai *et al.,* 2003) using homology searches with previously known/identified miRNAs. To date a number of computational methods have been reported for the identification of plant (Laufs *et al.,* 2004; Lagos-Quintana *et al.,* 2001; Reinhart *et al.,* 2002). Research in plants has revealed that short length sequences of mature miRNAs are conserved and have high complementarities to their target mRNAs (Laufs *et al.,* 2004). Hence, candidate miRNAs can be detected using the conserved complementarities of miRNA to target mRNA, if the mRNA target sequence is known. On the other hand, it has also been shown that the secondary structures of miRNA precursor (pre-miRNA) are relatively more conserved than pri-miRNA sequences (Wang *et al.,* 2005). Recent bioinformatics tools were used to identify miRNA utilizing both sequence and secondary structure alignments. Since the characteristic patterns of the conservation of miRNAs are searched by algorithms, the major challenge is finding miRNAs which are species specific and unrelated to previously known organisms. Zhang *et al.* (2005) identified and characterized new plant microRNA using EST analysis. Some of the new identified potential miRNAs may be induced and regulated by environmental biotic and abiotic stresses. Some may be preferentially expressed in specific tissues, and are regulated by developmental switching. These findings suggest that EST analysis is a good alternative strategy for identifying new miRNA candidates, their targets, and other genes. Conserved microRNAs and their targets were computationally identified and validated by qPCR in coffee (Akter *et al.,* 2013), cassava (Patanum

et al., 2012), kodo millet (Babu *et al.,* 2013), *Catharanthus roseus* (Pani and Mahapatra, 2013), *Allium sativa* (Panda *et al.,* 2014), *Thellungiella halophila* (Panahi *et al.,* 2013), tobacco (Frazier *et al.,* 2010), tomato (Din and Barozai, 2013) and wheat (Han *et al.,* 2013).

Obviously, methods that rely on phylogenetic conservation of the structure and sequence of a miRNA cannot predict non conserved genes. To overcome this problem, several groups have developed *ab initio* approaches to miRNA prediction (Bentwich *et al.,* 2005; Sewer *et al.,* 2005; Xue *et al.,* 2005) that use only intrinsic structural features of miRNAs and not external information. Each of these methods builds classifiers that can measure how a candidate miRNA is similar to known miRNAs on the basis of several features. Once a set of features is defined, a popular machine learning approach called 'support vector machines' is used to build a model, based on positive and negative training sets, that assigns weights to different features such that their contribution to an overall score results in the optimal separation of positives and negatives. With these *ab initio* prediction methods, many non-conserved miRNAs have been discovered and experimentally verified (Berezikov *et al.,* 2006). *Linum ussitatissimum* miRNA and their targets were predicted using the prediction tool NOVOmir (Moss and Cullis, 2012). miRDeep-P was used for the identification and characterization of a subset microRNAs in wheat (Su *et al.,* 2014).

2.5.1 Some bioinformatics tools used for identifying miRNA and its target mRNA

2.5.1.1 miRBase database (http://www.mirbase.org/)

The miRBase database is a searchable database of published miRNA sequences and annotation. miRBase was established in 2002, then called the MicroRNA Registry, with the primary aim of assigning stable and consistent names to newly discovered microRNAs. Each entry in the miRBase Sequence database represents a predicted hairpin portion of a miRNA transcript (termed mir in the database), with information on the location and sequence of the mature

miRNA sequence (termed miR). Both hairpin and mature sequences are available for searching and browsing, and entries can also be retrieved by name, keyword, references and annotation. All sequence and annotation data are also available for download. The miRBase Registry provides miRNA gene hunters with unique names for novel miRNA genes prior to publication of results. The latest miRBase release (v20, June 2013) contains 24,521 microRNA loci from 206 species (primates, rodents, birds, fish, worms, flies, plants and viruses), processed to produce 30,424 mature microRNA products (Kozomara and Griffiths-Jones, 2013).

2.5.1.2 MiPred (http://www.bioinf.seu.edu.cn/miRNA/)

To distinguish the real pre-miRNAs from other hairpin sequences with similar stem-loops (pseudo pre-miRNAs), a hybrid feature which consists of local contiguous structure-sequence composition, minimum of free energy (MFE) of the secondary structure and P-value of randomization test is used. Besides, a novel machine-learning algorithm, random forest (RF), is introduced. Given a sequence, MiPred decides whether it is a pre-miRNA-like hairpin sequence or not. If the sequence is a pre-miRNA-like hairpin, the RF classifier will predict whether it is a real pre-miRNA or a pseudo one (Jiang *et al.,* 2007).

2.5.1.3 NOVOMIR

NOVOMIR (Teune and Steger, 2010) is a program for the identification of miRNA genes in plant genomes. 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. Plant pre-miRNAs are more heterogeneous with respect to size and structure than animal pre-miRNAs. Despite these difficulties, NOVOMIR is well suited to perform searches for pre-miRNAs on a genomic scale. NOVOMIR is written in Perl and relies on two additional, free programs for prediction of RNA secondary structure.

2.5.1.4 RNA mFold (http://mfold.bioinfo.rpi.edu)

The abbreviated name, 'mfold web server', describes a number of closely related software applications available on the World Wide Web (WWW) for the prediction of the secondary structure of single stranded nucleic acids. The objective of this web server is to provide easy access to RNA and DNA folding and hybridization software to the scientific community at large. By making use of universally available web GUIs (Graphical User Interfaces), the server circumvents the problem of portability of this software. Detailed output, in the form of structure plots with or without reliability information, single strand frequency plots and 'energy dot plots', are available for the folding of single sequences (Zuker, 2003).

2.5.1.5 psRNATarget (http://plantgrn.noble.org/psRNATarget/)

psRNATarget, a plant small RNA target analysis server, which features two important analysis functions: (i) reverse complementary matching between small RNA and target transcript using a proven scoring schema, and (ii) target-site accessibility evaluation by calculating unpaired energy (UPE) required to 'open' secondary structure around small RNA's target site on mRNA. The psRNATarget incorporates recent discoveries in plant miRNA target recognition, e.g. it distinguishes translational and post-transcriptional inhibition, and it reports the number of small RNA/target site pairs that may affect small RNA binding activity to target transcript. The psRNATarget server is designed for high-throughput analysis of next-generation data with an efficient distributed computing back-end pipeline that runs on a Linux cluster. The server front-end integrates three simplified user-friendly interfaces to accept user-submitted or preloaded small RNAs and transcript sequences; and outputs a comprehensive list of small RNA/target pairs along with the online tools for batch downloading, key word searching and results sorting. The psRNATarget server is freely available (Dai and Zhao, 2011).

2.5.2 Plant miRNA target prediction tools

Plant miRNAs regulate gene expression by binding to the target mRNAs through complementary base-pairing (Carthew and Sontheimer, 2009). Three modes of target repression proposed in plants are cleavage, translational inhibition and destabilization of targets. The review by Dai *et al.* (2010) focused on recent progress in plant miRNA target recognition mechanism, principles of target prediction based on these understandings, comparison of current prediction tools and algorithms for plant miRNA target analysis and the outlook for future directions in the development of plant miRNA target tools and algorithms.

One of the earliest programs used for searching complementary target sites is PatScan (Dsouza *et al.*, 1997), which has been successfully applied in studying miRNAs in rice and *Arabidopsis*. miRU (Zhang, 2005), the tool for the plant-specific miRNA target prediction, which was later upgraded to psRNATarget (Dai and Zhao, 2011), uses a dynamic programming approach, aligning sequences using a modified Smith-Waterman algorithm and applying the 'RNAup' algorithm (Lorenz *et al.*, 2011) for target site accessibility. Targetfinder (Fahlgren *et al.*, 2007) implements a 'FASTA' program along with a penalty scoring scheme for mismatches, bulges, or gaps for aligning the sequences. In 2010, two web-servers, TAPIR (Bonnet *et al.*, 2010) and Target-align (Xie and Zhang, 2010), were introduced. TAPIR is imbedded with two search options, the 'FASTA' search engine (for 'fast' searches), and the 'RNA hybrid' search engine (for 'precise' results). Target-align also employ the Smith-Waterman based scoring method for predicting the complementarities between miRNAs and mRNAs. Target-align is implemented both as a web server and as a standalone tool, but its utility for genome-wide target predictions for smRNAs has not been tested. Target_Prediction (Sun *et al.*, 2011) is based on 'scanning' targets for miRNA-patterns followed by the calculation of the minimum free energy (with the help of 'RNAhybrid') for predicting miRNA-mRNA duplexes. miRTour (Milev *et al.*, 2011), a web server based program, implements a variety of resources such as BLASTX, RNAfold and ClustalW for the prediction of targets (and thus also

involves energy minimizations). imiRTP (Ding *et al.*, 2011) is an integrated miRNA target interaction prediction tool kit only for *Arabidopsis thaliana* miRNAs. Further, machine learning has been implemented for predicting the plant miRNA targets, for instance, p-TAREF (Jha and Shankar, 2011) implements support vector regression (SVR) and uses a feature of information of 'dinucleotide density variation' around the target site from datasets of *Arabidopsis thaliana, Oryza sativa, Medicago truncatula and Solanum lycopersicum.* psRobot (Wu *et al.*, 2012) is a server hosting a toolbox for analyzing plant smRNAs: it has two modules of stem-loop prediction and smRNA target prediction. psRobot uses a modified Smith-Waterman algorithm and target site conservation to predict targets in *A. thaliana, Brachypodium distachyon, Carica papaya, O. sativa, Populus trichocarpa, Sorghum bicolor, Vitis vinifera and Zea mays.* Parallel programming is implemented to reduce the run-time during analysis of large datasets such as transcriptomes and genomes. Archak and Nagaraju (2007) carried out global computational analysis of rice transcriptome to generate a comprehensive list of putative miRNA targets using miRanda target prediction algorithm. The miRanda target prediction algorithm (http://www.microrna.org/) aligns a miRNA to target mRNA using a scoring scheme based on complementarities of nucleotides. miRanda does not require exact seed pairing and predict sites which contain either a bulge or a G:U wobble in the seed region (Betel *et al.*, 2008).

2.6 MICRORNA DETECTION AND QUANTIFICATION METHODS

Efficient and suitable miRNA detection and quantification are essential to understand miRNA function in specific conditions, cell and tissue types. Northern hybridization, cloning, and microarray analysis are widely used to detect and quantify miRNAs in plants, but these techniques are less sensitive and are not high throughput compared with effective and sensitive quantitative real-time reverse transcription PCR (qRT-PCR). Recently Varkonyi-Gasic *et al.* (2007) described a protocol for an end-point and real-time looped RT-PCR procedure. Their approach includes two steps. In the first step, a stem-loop RT primer is

designed, following the strategy developed by Chen *et al.* (2005) and is hybridized with the candidate miRNA. The second step includes the specific amplification of the miRNA, using a forward primer specific for the miRNA and a universal reverse primer, which is designed for the stem-loop RT primer sequence. The clues for designing the reverse RT primers and miRNA specific forward primers are that the specificity of stem loop RT primers for a certain miRNA is conferred by a six nucleotide extension at the 3' end. This extension is the reverse complement of the last six nucleotides at the 3' end of the miRNA. Forward RT primers are specifically designed for individual miRNA sequences. At the primer's 5' end 5–7 random and relatively GC-rich nucleotides are added to increase the template's melting temperature.

In general, there are two common qRT-PCR methods: SYBRGreen- based miRNA RT-qPCR assays and Stem-loop RT based TaqMan, which are differentiated from each other in terms of chemical reaction. The SYBR-based assays use fluorescent double-stranded DNA binding dye, which can intercalate into strands of amplification products, and measuring the increase in fluorescence during PCR cycles monitors detection of amplified miRNAs. However, the detection of expression level of any miRNA by using SYBR Green assay presents several disadvantages because dye can bind to any double-stranded DNA regardless of amplicon or nonspecific cDNAs; thus, it may generate false positive signals. However, the SYBR-based assay is widely used for detection of the expression profile of well-known miRNAs because of its cost per sample, sensitivity, and no requirement for probes (Raymond *et al.*, 2005; Varkonyi-Gasic *et al.*, 2007; Sharbati-Tehrani *et al.*, 2008). Unlike SYBR-based qRT-PCR methods, the stem-loop RT-based TaqMan method uses a target-specific fluorogenic probe that enables the rapid detection and quantification of desired miRNAs (Chen *et al.*, 2005; Mestdagh *et al.*, 2008). Figure 1 is a schematic illustration of TaqMan probe and SYBR Green real time PCR assay for quantification of mature miRNAs.

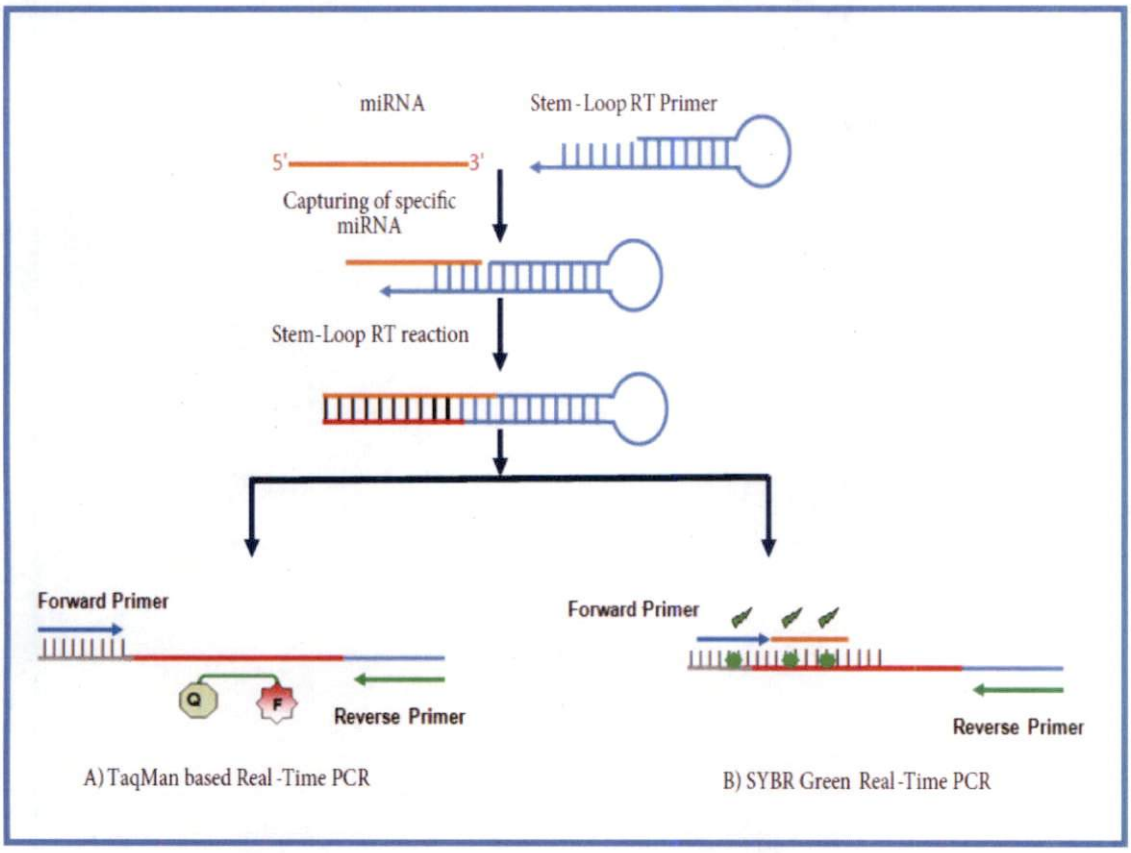


Figure 1. TaqMan probe and SYBR Green real-time PCR assay for quantification of mature miRNA (Eldem *et al.*, 2013)

Benes and Castoldi (2010) reviewed the different methodologies to estimate the expression levels of microRNAs (miRNAs) using real time quantitative PCR (qPCR). They have introduced novel technological approaches and compared them to existing qPCR profiling methodologies. Discussion on expression profiling of mature miRNAs by qPCR in four sequential sections: (1) cDNA synthesis; (2) primer design; (3) detection of amplified products; and (4) data normalization are also reviewed. Technical challenges associated with each of these are addressed and possible solutions outlined.

2.7 VIRAL miRNAs

Granted that miRNA-mediated gene silencing serves as a general defense mechanism against plant viruses, it would not be a surprise that viruses also employ miRNAs to circumvent the defense system. Viruses generate miRNAs and employ them to modulate their own gene expression as well as that of their host cells (Sullivan and Ganem 2005). Present research advances reveal that the virus encoded miRNAs are the key players in modulating the antiviral host defense machinery by regulating both host cellular and their own gene expression (Song *et al.,* 2011). Based on the diversity of virus families, it is reasonable to predict that there will be several categories of virally encoded miRNAs. Nonetheless, extensive cDNA cloning studies across many families of RNA viruses have failed to identify miRNAs (Pfeffer *et al.,* 2005) which is perhaps due to the predominant role of the DNA-dependent RNA polymerase II in biogenesis of pri-miRNAs (Sullivan and Ganem 2005). However, miRNAs may be produced by viral RNA-dependent RNA polymerases, especially for virus families in which genomic replication or transcription occurs in the host nucleus. The first virus exhibited to encode miRNA is Epstein Bar Virus, a causative agent of infectious mononucleosis (Pfeffer *et al.,* 2004) followed by many discoveries (Bennasser *et al.,* 2004; Omoto *et al.,* 2004; Cai *et al.,* 2005; Omoto and Fujii 2005; Pfeffer *et al.,* 2005; Samols *et al.,* 2005; Sullivan *et al.,* 2005). However, no conservation has been observed among the virally encoded miRNAs. Computational predictions show that these miRNAs could participate in a variety of functions:

biogenesis of other small RNAs, viral DNA polymerase synthesis, viral transcription, as well as host cell apoptosis. Virally encoded miRNAs are involved in counter-defense to circumvent plant defense system.

2.7.1 Identification of viral miRNAs

Most viral miRNAs had initially been identified by a protocol previously developed for the identification of host-encoded miRNAs, a procedure that involves RNA size fractionation, ligation of linkers, reverse transcription, concatamerization, and Sanger sequencing. There are also computational approaches that rely on commonalities in the predicted secondary structures of pre-miRNAs to identify miRNA-encoding loci specifically in viral genomes (Grundhoff, 2011; Pfeffer *et al.*, 2005; Sullivan *et al.*, 2005). While such *ab initio* prediction approaches often produce significant numbers of false positives that have to be eliminated experimentally, they have the advantage of being able to identify the less abundantly expressed miRNAs which frequently had been overlooked in the original cloning protocol. VMir, an *ab initio* prediction program was recently designed to specifically identify pre-miRNAs in viral genomes (Grundhoff, 2011). Computational miRNA prediction represents a valuable alternative which can be performed with comparably little technical effort. This is especially true for viruses, as the number of predicted candidates generally remains low and thus within a range that may be readily confirmed by experimental means. However, with the advent of massively parallel sequencing technologies it is now possible to explore libraries of cloned small RNAs with unprecedented depth (Lu *et al.,* 2008). A recent study has predicted five and experimentally demonstrated one viral miRNA (hcrsv-miR-H1- 5p) from an (+)-sense ss RNA virus, Hibiscus chlorotic ringspot virus (HCRSV, Carmovirus) infecting *Hibiscus cannabilis* L. using the vir-miRNAs prediction database (Gao *et al.,* 2012). Viswanathan *et al.* (2014) computationally predicted and experimentally validated the miRNA encoded by the Sugar Cane Streak Mosaic Virus (SCSMV) genome with detection efficiency of 99.9 % in stem-loop RT-qPCR and predicted their potential gene targets in sugarcane. These sugarcane

target genes considerably broaden future investigation of the SCSMV encoded miRNA function during viral pathogenesis and might be applied as a new strategy for controlling mosaic disease in sugarcane.

2.8. GENOMIC RESOURCES OF CASSAVA AND CASSAVA MOSAIC VIRUS

Cassava Genome Project with the goal to generate draft sequence of cassava began in 2003. A 454-based whole genome shotgun sequence has been assembled, which covers 69% of the predicted genome size and 96% of protein-coding gene space. The predicted 30,666 genes and 3,485 alternate splice forms are supported by 1.4 M expressed sequence tags (ESTs). The resulting assembly and its annotation are available through Phytozome and have also been deposited in GenBank (Prochnik *et al.,* 2012).

The main diseases affecting cassava are cassava mosaic disease (CMD), cassava bacterial blight, cassava anthracnose disease, and root rot. The cassava mosaic virus causes the leaves of the cassava plant to wither, limiting the growth of the root (Legg and Fauquet, 2004). Cassava mosaic disease is the most important disease threatening cassava production causing losses of between 20 – 80% of total yields throughout Africa and can result in complete crop failure (Fregene and Puonti-Kaerlas, 2002). Cassava Mosaic Virus (CMV) belongs to the genus *Begomovirus*, family *Geminiviridae* and is transmitted by the whitefly *Bemisia tabaci* as well as planting cuttings from diseased plants. Geminiviruses are large family of plant viruses with circular, single stranded genomes packaged within geminate particles. Members of genus *Begomovirus* have caused significant yield losses in many crops worldwide (Varma and Malathi, 2003).

The genome of CMGs contains two DNA molecules – A and B each of about 2.8 kbps (Stanley, 2004) which are coding for different proteins responsible for different functions in the infection process. Both the DNA molecules are required for infectivity, vector transmission, virus spread and for the systemic infection of susceptible host plants (Fregene and Puonti-Kaerlas, 2002). DNA A is

involved in the replication of DNA components and virus. DNA B is involved in cell-to-cell and long-distance virus spread and production of disease symptoms. Conserved genome sequence called common region (CR) located in 5' intergenic region contains modular *cis*-acting sequences involved in transcriptional regulation of certain viral genes and the sequence elements essential for virus replication (Idris and Brown, 1998). By far, the most informative of both genomic components is the DNA A that encodes two overlapping virion-sense open reading frames (ORFs) AV2 and AV1, and at least four overlapping complementary-sense ORFs AC1, AC2, AC3 and AC4. AV1 encodes the coat protein gene (CP) and is the determinant of vector transmission (Harrison *et al.,* 2002) in addition to its role in genome encapsidation. Complementary-sense genes individually and in concert, are implicated in the replication of CMVs within the host cell. ORF AC1 encodes a replication-associated protein (Rep), AC2 a transcriptional activator protein (TrAP), and AC3 a replication enhancer protein (REn). ORF AC4 plays a role as a host activation protein, which serves as an important symptom determinant implicated in cell-cycle control, and may also counteract the host response to Rep gene expression (Hull, 2002). The two ORFs of the DNA B component, BV1 and BC1, encode the nuclear shuttle protein and the movement protein, respectively. These two ORFs are non-overlapping and code for genes that play a role in intra- (BV1) and inter- (BC1) cellular movement of virions within the host plant cell (Stanley *et al.,* 2005).

Geminivirus replication relies on DNA intermediates and takes place within the nucleus via two stages: by converting the genomic ssDNA into a dsDNA intermediate and amplification of viral ssDNA through the rolling-circle replication (Gutierrez *et al.*, 2004). The genomic ssDNA is then transported to neighbouring cells and is encapsidated to form mature viral particles.

Nine species of CMGs have been identified between Africa and South Asia based on their genomic sequence and phylogenetic analysis. They include representatives of seven African and two south Asian species namely *African Cassava Mosaic Virus* (ACMV), *East African Cassava Mosaic Virus* (EACMV),

East African Cassava Mosaic Cameroon Virus (EACMCV), *East African Cassava Mosaic Kenyan Virus* (EACMKV), *East African Cassava Mosaic Malawi Virus* (EACMMV), *East African Cassava Mosaic Zanzibar Virus* (EACMZV), *South African Cassava Mosaic Virus* (SACMV), all from Africa as well as *Indian Cassava Mosaic Virus* (ICMV) and *Sri Lankan Cassava Mosaic Virus* (SLCMV) in Asia (Fauquet *et al.,* 2008; Patil and Fauquet, 2009). The number will probably grow resulting from high rate of natural recombination between geminiviruses and high transmission rate of white fly vectors (Patil and Fauquet, 2009).

MATERIALS & METHODS

3. MATERIALS AND METHODS

The study entitled "Evaluation of prediction tools and computational analysis of microRNAs in cassava (*Manihot esculenta* Crantz.)" was conducted at the Central Tuber Crops Research Institute during August 2013 to June 2014. Details regarding the experimental materials used and methodology adopted for various experiments are presented in this chapter.

3.1 REVIEW OF miRNA TARGET PREDICTION TOOLS AND COMPARATIVE EVALUATION

The various available miRNA target prediction tools were reviewed. Target prediction tools: plant small RNA (psRNA) target (http://plantgrn.noble.org/psRNATarget/) and miRanda target prediction algorithm (http://www.microrna.org/) were compared to evaluate their performance. *Arabidopsis thaliana* is a widely used model flowering plant, for which the majority of tools have been developed. *Arabidopsis* mature miRNA sequences were downloaded from miRBase (http://www.mirbase.org/). The data regarding experimentally validated *Arabidopsis* miRNA-mRNA interaction (201) was obtained (Srivastava *et al.,* 2014) and this was used as the positive dataset. Similarly, 32 experimentally validated negative sequences used by Heikham and Shankar (2010) was used as negative dataset. The miRNA sequences, both positive and negative datasets were given as input to psRNA-target and miRanda target prediction tool.

The psRNA-target with the following parameters was employed in prediction of miRNA targets: Maximum Expectation = 3.0; Length of Complementarity Scoring = 20 bp; Target Accessibility – Allowed Maximum Energy to Unpair the Target Site = 25; Flanking Length along Target Site for Target Accessibility Analysis = 17 bp in upstream/ 13 bp in downstream; Range Of Central Mismatch Leading To Translational Inhibition = 9-11 nt. The miRanda scanning algorithm (Betel *et al.,* 2008), which utilizes dynamic-programming alignment and thermodynamics to predict miRNA targets, was employed in a

stand-alone version 1.9 (http://www.microrna.org/). The parameters in miRanda were kept at default except for scaling factor = 2; score > 95; energy <= -20 kcal mol^{-1} (Archak and Nagaraju, 2007). The output was analyzed to calculate the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

True positives is defined as the number of experimentally supported miRNA targets that are predicted by a program and false negatives are those experimentally supported miRNA targets that are not predicted by the program. Similar to the above case, false positives is defined as the number of all negatives that is predicted by a program and true negative on the other hand is all negatives that are not predicted by the program.

Further, in order to evaluate the performance of these different predictive tools, we used the statistical parameters, viz., Sensitivity (Se), Specificity (Sp) and Accuracy (Acc), a summary statistic: Mathew correlation coefficient (MCC) and Positive predictive value (PPV).These parameters are based on TP, FN, TN and FP and are calculated using the following 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

Similarly to find the best maximum expectation value of target prediction tool psRNA-target, the maximum expectation value was set at 1/2/3/4/5, keeping all the other parameters at default values. The output was analyzed to calculate the number of TP, TN, FP, FN. Accuracy, sensitivity and specificity was also calculated. To determine the most suitable threshold/cut-offs, 'precision / positive predictive value' and 'recall / sensitivity' were calculated at five maximum

expectation values (1, 2, 3, 4 and 5). Scores at which precision and recall value intersect were considered optimal for the tool.

3.2 COMPUTATIONAL PREDICTION OF miRNAs

3.2.1 Cassava (*Manihot esculenta*) miRNA prediction

A computational prediction was used for predicting potential miRNAs in cassava by using homology search based on miRNA conservation among different plant species.

3.2.1.1 miRNA reference set and cassava genome

A total of 6690 mature miRNA sequences from various species of monocots (1910) and eudicots (4780) were obtained from miRBase (Release 20) (http://www.mirbase.org/) database (Griffiths-Jones *et al.,* 2008). Redundant sequences were removed from dataset using Jalview version 2.8 (www.jalview.org). The non redundant sequences (3513) were then used to probe for potential cassava miRNAs. The cassava genome database Cassava 4 consisting of 12977 scaffolds spanning 533 Mb was accessed from Phytozome (http://www.phytozome.net/cassava)

3.2.1.2 Homology search

BLASTn (Altschul *et al.,* 1990) performed with selected miRNA sequences and cassava genome in Phytozome with default parameters (Output sequences with less than 3 mismatches when compared with the query miRNA sequence, e-value < 0.01 and not less than 18 nucleotides (nt)). Precursor sequences of 400 nt. were extracted (200 nt upstream and downstream from BLAST hits). The extracted sequences were subjected to remove the protein coding sequences by using BLASTx with default parameters in Phytozome.

3.2.1.3 Prediction of miRNA

Mfold, a publicly available online application (http://mfold.bioinfo.rpi.edu) (Zuker, 2003) was used to predict the secondary structure of obtained sequences

based on thermodynamic stability. The RNA folding application was used and all parameters were kept at default. The structure with the highest score and lowest free energy was analysed and precursor sequences was predicted based on secondary folding structure. The extent of precursor was predicted by identifying any large loops with little or no nucleotide pairing that followed the end of a region with significant pairing.

Sequences that fit the following criteria were designated as potential miRNAs in cassava: (1) Mature miRNA should be 18–25 nt in length, (2) The predicted pre-miRNA folded into a perfect or near perfect stem-loop hairpin secondary structure, (3) The potential mature miRNA sequence located on one arm of hairpin structure, (4) No loops or breaks were allowed in the miRNA/miRNA* duplex, (5) 6 nt mismatches were allowed between miRNA/miRNA* duplex, (6) (A + U) content should be 30–75 per cent, (7) The predicted pre-miRNA must have a high negative minimal free – folding energy (MFE) which obtained from the negative folding free energies (Delta G) and MFE index (minimal free folding energy index or MFEI) > 0.85 in order to distinguish from other small RNAs (Zhang *et al.,* 2005).

The minimal folding energy (MFE), expressed in kcal mol^{-1}, is a method of calculating the thermodynamic stability of the secondary structure of RNA or DNA. The lower the MFE of a molecule, the more stable the secondary structure. Because MFE values are strongly correlated with the length of the sequence the adjusted MFE (AMFE) is calculated. The minimal folding free energy index (MFEI) was calculated for the *M. esculenta* miRNA precursors via the equation:

AMFE= [(MFE/length of RNA sequence) x 100]

MFEI= [(AMFE) x 100]/(G+C) %.

3.1.2 *Indian cassava mosaic virus* miRNA prediction

The *Indian cassava mosaic virus* miRNA prediction was performed using complete genome (DNA A & DNA B) sequence of virus.

(http://www.ncbi.nlm.nih.gov/). The genome size of DNA A and DNA B are 2,815 bp and 2,645 bp respectively. The viral genome was scanned for hairpin structured miRNA precursor using VMir Analyzer program (Grundhoff, 2011) with default parameters. VMir is an *ab initio* prediction program which is designed specifically to identify pre-miRNAs in viral genomes. The precursors were further identified using MiPred program (http://www.bioinf.seu.edu.cn/miRNA/) and sequences with lower minimum folding energy (<= -25 kcal/mol) were selected. The selected candidates within or antisense to protein coding regions were removed after NCBI BLASTx. At the last step, the mature sequences were predicted by Bayes-SVM-MiRNA web server v1.0 (http://wotan.wistar.upenn.edu/BayesSVMmiRNAfind/). Selected pre-miRNA candidates were used for secondary structure prediction. The candidates were checked for the following criteria: (i) Can fold into an appropriate stem loop hairpin structure, (ii) Predicted mature miRNA resides in one arm of the hairpin no matter if it is 3' or 5', (iii) No more than 8 mismatches between the predicted mature miRNA and their opposite sequence in the other arm (miRNA*), (iv) No loops or breaks in the miRNA sequence, (v) Predicted secondary structure with high MFEI and negative MFE (Viswanathan *et al.,* 2014).

3.3 COMPUTATIONAL PREDICTION OF miRNA TARGETS AND FUNCTIONAL ANNOTATION

The target genes of miRNAs could be predicted according to their perfect or nearly perfect complementarity between them and their target genes through homology algorithm.

3.3.1 Cassava / *Indian cassava mosaic virus* miRNA targets in Cassava transcripts

Either perfect or near perfect complementary binding of miRNAs to their target genes in plants enables us to identify miRNA targets. The cassava miRNA targets were identified through pair wise homolog search.

The web tool plant small RNA (psRNA)-target server (http://plantgrn.noble.org/psRNATarget/) (Dai and Zhao, 2011) was applied for predicting cassava miRNA targets in cassava transcripts. The analysis was performed in user submitted small RNAs/ user submitted transcript. Cassava transcripts were downloaded from Phytozome and predicted cassava miRNAs / Indian cassava mosaic virus miRNA were used as small RNA input. The various parameters were kept at default (maximum expectation = 3.0; length of complementarity scoring = 20 bp; target accessibility – allowed maximum energy to unpair the target site = 25; flanking length along target site for target accessibility analysis = 17 bp in upstream/ 13 bp in downstream; range of central mismatch leading to translational inhibition = 9-11 nt). Results from the analysis were individually inspected on Phytozome, where functional annotation of targets was obtained.

3.3.2 Identification of *Manihot esculenta* miRNA targets in the genome of cassava mosaic virus

A set of 153 known miRNA sequence of *M. esculenta* were downloaded from miRBase. Further 75 nucleotide sequence of DNA A and 18 nucleotide sequence of DNA B of cassava mosaic virus was retrieved from NCBI GenBank. The nucleotide sequence composed of sequences from all 9 strains of cassava mosaic virus. To identify miRNA target sites in DNA A and DNA B, miRanda target prediction algorithm (http://www.microrna.org/) was applied. Analysis was performed on Mac (OS) based computer having Intel ® core i7, 2.8 GHz processor and 4 GB RAM. The threshold sequence complementarity score was tuned at 50, free energy was adjusted at -20 kcal mol^{-1} and threshold percentage complementarity between miRNA-target duplex was selected at 60. True regulatory targets were scrutinized on the assumption that all potential miRNA targets do not have more than five mismatches. These include one or two mismatches in octameric seed region, not more than three mismatches in position 13-22 and complementarity or wobble at position 10 and 11 and not more than a single gap inserted with their corresponding miRNAs.

3.4 EXPERIMENTAL VALIDATION OF miRNA

To experimentally validate the mature cassava miRNA having target in cassava mosaic virus, reverse transcription- PCR (RT-PCR) and stem-loop end-point pulse RT-qPCR were performed as described below using the isolated total RNA from the collected healthy and cassava mosaic virus-infected cassava leaf tissues.

3.4.1 miRNA sequences and primers

Due to limited resources only two cassava miRNAs were included for verification study. Cassava miRNA mes-miR164 and mes-mir395 (having targets in cassava mosaic virus genome) were randomly selected. The mature miRNA sequences were downloaded from miRBase.

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 Total RNA isolation

Total RNA was isolated from healthy and cassava mosaic virus infected cassava leaf sample by lithium chloride method (Zeng and Yang, 2002).

About fifteen ml of extraction buffer (appendix 1) was pre-warmed at 65° C in water bath.0.5-0.7 g leaf samples were ground in liquid nitrogen. The frozen

powder was quickly transferred to pre-warm extraction buffer and mixed by inverting. This was then incubated at 65° C for 10 min with vigorous shaking and centrifuged at 15,000 rpm for 10 min at room temperature. To the supernatant an equal volume of chloroform: isoamyl alcohol (24:1) was added and shaken vigorously. Centrifugation was done at 20,000 *g* for 10 min at 4° C. Viscous supernatant was transferred to a new tube and re-extracted with equal volume of chloroform: isoamyl alcohol (24:1). To the carefully collected supernatant, 0.25 volume 10 *M* lithium chloride was added, mixed well and stored at 4° C overnight. The RNA was recovered by means of centrifugation at 30,000 *g* for 30 min at 4° C. The supernatant was discarded and the pellet was washed thrice with 30 µl ethanol (75 %) at 10,000 *g* for 10 min at 4° C. The pellet was dried at 37° C for 30 min. 50 µl of DEPC treated distilled water was added to pellet and is kept at 37° C for 30 min for dissolution.

RNA quality was determined by using agarose electrophoresis. The gel was viewed under Alpha Innotech gel documentation system. RNA samples were stored at -80° C.

3.4.3 Stem loop pulsed reverse transcription

Verification was done for the expression of two cassava miRNA *viz.*, mes-miR395 and mes-miR164. mes-mir164 and mes-mir395 RT primers are used to obtain cDNA from total RNA. The components of the mixture were optimized as listed below for reverse transcription reaction:

dNTP (10 *mM*)	: 0.50 µl
5X buffer	: 4.00 µl
AMV Reverse Transcriptase (10 U/ µl)	: 0.25 µl
RT primer (1 *µM*)	: 0.10 µl
Nuclease Free Water	: 14.15 µl
RNA	: 1.00 µl

Total volume : 20.00 µl

The RT reaction was performed by loading thermal cycler (Eppendorf Mastercycler (Germany)) and incubating for 30 min at 16° C, followed by pulsed RT of 60 cycles at 30° C for 30 s, 42° C for 30 s and 50° C for 1 s. Finally the components were incubated at 85° C for 5 min to inactivate the reverse transcriptase. Control reactions were carried out to distinguish the target products from non-target products and primer dimers. The amplified products were separated on 1 per cent agarose gel. The gel was viewed under Alpha Innotech gel documentation system.

3.4.4 miRNA SYBR Green I assay

The qPCR reaction with the forward primer (specific to mes-mir164 or mes-mir395) and the universal reverse primer was done in Genaxy (Axygen) 96 FLT-C (12.5µl) plate with the healthy and infected mes-mir164 and mes-miR395 cDNA samples in Eppendorf Realplex.

For each reaction, 12.5 µl PCR reaction mixtures were prepared and each contained 2.5 µl of RT product from the reverse-transcription reaction, 1 µl of miRNA-specific forward primers and reverse primer, 6.25 µl Mesa green qPCR master mix plus for SYBR, and 1.75 µl of nuclease-free water. The plate was then sealed with a stopper. The samples were then incubated at 95°C for 5 min, followed by 35 cycles of 95° C for 5 s, 60° C for 10 s, and 72° C for 8 s. This was followed by a melt curve analysis (rapid heating to 94° C to denature the DNA, followed by cooling to 60° C). Negative template control reaction was also performed.

After the completion of the real-time reactions, the threshold cycle (C_T) was recorded (reference gene for normalization was not included in the assay). All reactions were conducted in duplicates.

RESULTS

4. RESULTS

4.1 REVIEW OF microRNA TARGET PREDICTION TOOLS AND COMPARATIVE EVALUATION

Out of the 18 published miRNA target prediction tools, 11 are quantitatively available for sequential evaluation based on different criteria. Plant specific miRNA target prediction tools are implemented either in the form of a web server or as a standalone tool. A summary of all the tools is presented in Table 1.

The miRNA target prediction tools: psRNATarget and miRanda were compared. The results of miRNA target prediction tools: psRNATarget and miRanda are summarized in terms of per cent count of TP, FN, FP, TN, Sensitivity, Specificity, Accuracy, MCC and PPV of each tool (Table 2).

Of the two tools considered for plant miRNA target prediction, specificity of psRNATarget is as high as 61.4 per cent compared to 37.8 per cent specificity of miRanda. The sensitivity of miRanda is slightly high compared to psRNATarget. But because of the high specificity of psRNATarget the overall accuracy is more for this tool. Hence psRNATarget is a better tool when compared to miRanda which is indicated by the 86.5 per cent PPV for this tool. The MCC value of psRNATarget (0.499) is high compared to the other tool miRanda (0.397) showing the high efficiency of psRNATarget in target prediction.

The results of psRNATarget prediction tools at different maximum expectation values are summarized in terms of percent count of TP, FN, FP, TN, Specificity, Accuracy, Recall (sensitivity) and Precision (PPV) (Table 3). At the maximum expectation value of 3, values of precision (86.50 %) and recall (87.85 %) nearly intersected with minimum difference. The specificity and accuracy are also optimum at this value. Comparison of psRNAtarget at different maximum expectation value as obtained from the evaluation measures of Specificity (Sp), Sensitivity (Se) and Accuracy (Acc) are depicted in Figure 2.

Table 1. Parameters used in the different miRNA target prediction algorithms

Tool	Algorithm	SP	TSA	MS	CF	TI	Availability
Targetfinder	FASTA	+	-	-	-	-	Only source code
TAPIR	FASTA/ RNAhybrid	+	+	+	-	-	Web server and source code
Target-align	Smith-Waterman like	-	-	+	-	-	Web server and source code
Target_Prediciton	Scan for matches and RNA hybrid	-	+	-	-	-	Only source code
psRNATarget	Smith-Waterman	-	+	+	-	+	Only web server
p-TAREF	Support Vector Regression (SVR)	-	+	+	-	-	Web server and source code
psRobot	Modified Smith-Waterman	-	-	+	+	-	Web server and source code
miRanda	Local Alignment	+	+	+	-	+	Web server and source code
RNAhybrid	Intramolecular hybridization	+	+	+	-	+	Web server and source code
Targetscan 6.2	Custom made	+	-	+	-	+	Only source code

SP: Seed Pairing; **TSA**: Target site accessibility; **MS**: Multiple sites; **CF**: Conservation filter; **TI**: Translation inhibition; '+' Represent feature used, '-'indicates that these features were not used.

Table. 2 Performance of microRNA target prediction tool

Tool	Positive Data (%)		Negative Data (%)		Sp (%)	Se (%)	Acc (%)	MCC	PPV
	TP	FN	FP	TN					
psRNATarget	87.85	12.15	38.60	61.40	61.40	87.85	80.92	0.499	86.50
miRanda	95.41	04.59	62.16	37.84	37.84	95.41	74.61	0.397	73.07

Table. 3 Performance of psRNATarget prediction tool at different maximum expectation value

e-value	Positive Data (%)		Negative Data (%)		Sp (%)	Acc (%)	Recall (%)	Precision (%)
	TP	FN	FP	TN				
1	27.63	72.37	20.69	79.31	79.31	39.13	27.63	82.35
2	65.83	34.17	16.67	83.33	83.33	69.48	65.83	93.75
3	87.85	12.15	38.60	61.40	61.40	80.92	87.85	86.50
4	92.21	7.79	61.96	38.04	38.04	72.48	92.21	72.20
5	92.81	7.19	84.65	15.35	15.35	49.59	92.81	46.48

4.2 IDENTIFICATION OF CASSAVA microRNAs AND THEIR TARGETS

In order to identify the potentially conserved miRNAs in cassava, a reference set of 3,513 mature miRNA sequences from various monocots and eudicots species, after removing redundant sequences, were subjected to BLAST analysis against cassava genome database. The protein-coding sequences were then removed. Using mFOLD, secondary structure analysis of the results identified 152 potentially conserved miRNAs in cassava. The details regarding the predicted 152 miRNA are given in the Appendix II. The work flow for *in silico* prediction of cassava miRNA and target is shown in Figure 3.

The 152 potential cassava miRNAs belong to 30 families. Of these 30 miRNA families, the miR169 family in cassava was the largest family with 27 members. The two families, miR171 and miR156 were found to contain 12 and 11 members per family respectively. All other miRNA families contained fewer than 10 members and most only contained 1 or 2 miRNAs per family (Figure 4).

Of the 152 predicted cassava miRNAs, 8 miRNA sequences are newly identified and are not present in miRBase microRNA database. These new miRNA sequences show similarities with miRNA family mes-miR166, 171, 172, 390, 396, 397 and 399. The secondary structures of 8 new miRNAs are displayed in Figure 5.

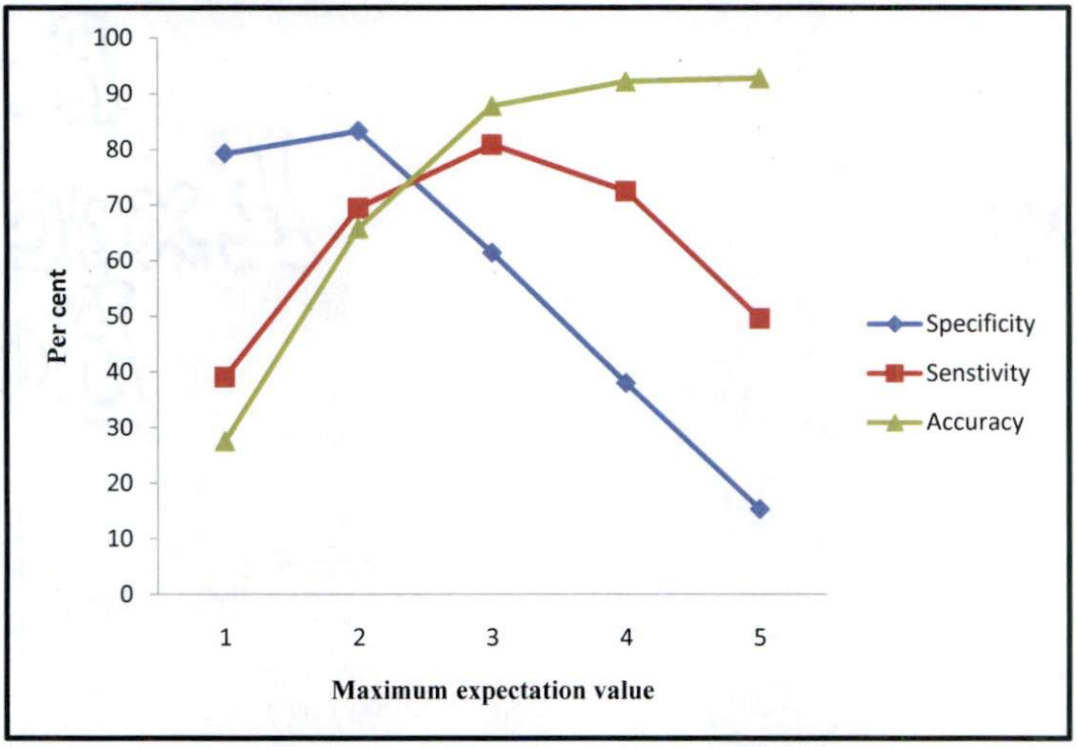


Figure 2. Comparison of psRNATarget at different maximum expectation value as obtained from the evaluation measures of Specificity (Sp), Sensitivity (Se) and Accuracy (Acc)

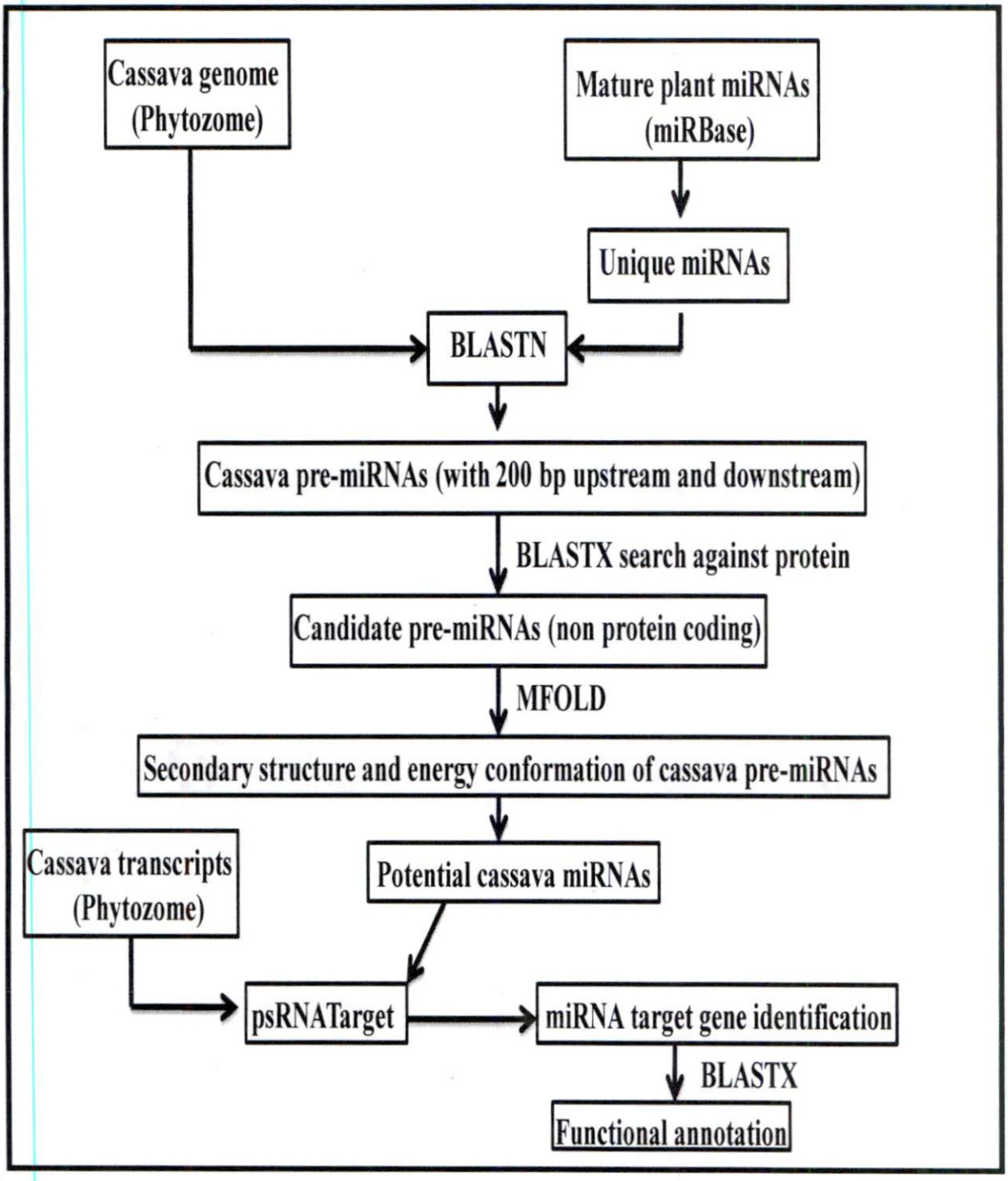


Figure 3. Workflow for the identification of potential miRNAs and targets in cassava

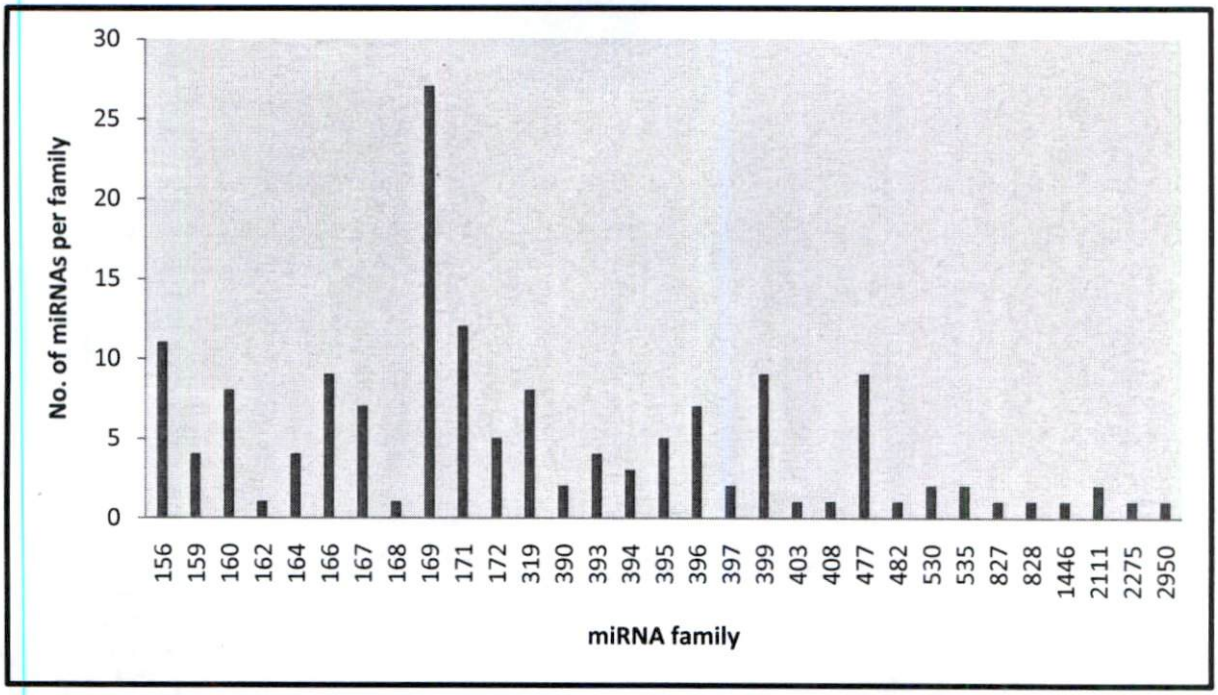


Figure 4. Number of cassava miRNAs per family

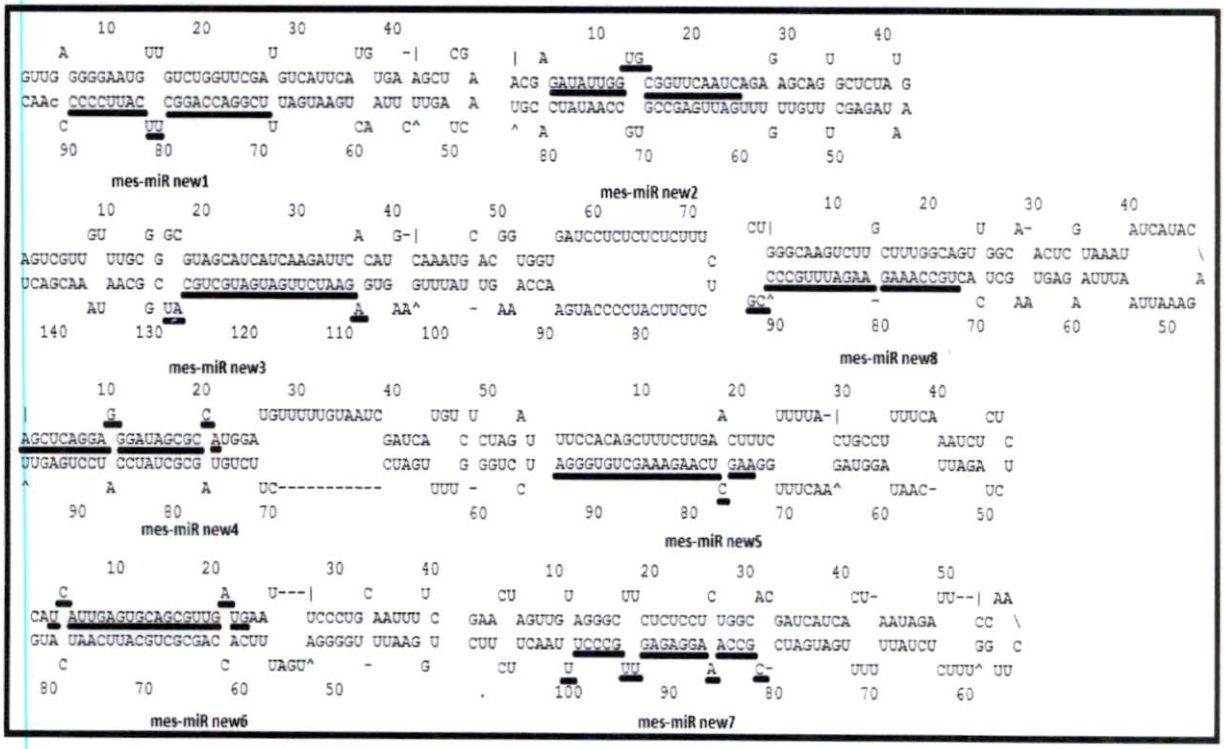


Figure 5. Predicted secondary hairpin structure for the eight new cassava miRNAs (mature miRNA sequences are underlined)

Characteristics of the potentially conserved miRNAs in cassava varied between families. Most of the identified mature miRNA sequences (71.71 %) in cassava began with the base uracil (U). The majority of identified potential cassava miRNAs (78.95 %) were 21 nt in length followed by 20 nt (11.18 %), 22 nt (7.24 %) and 23 nt (2.63 %) (Figure 6). A majority of identified miRNAs were obtained from the plus strand. However, there were several miRNAs identified from the minus strand.

A total number of 93 (61.4 %) miRNAs were predicted to be found on the 5' arm of the pre-miRNA stem-hairpin loop. In contrast, 59 miRNAs (38.6 %) were predicted to be found on the 3' arm of the pre-miRNA stem-hairpin loop. Potential cassava pre-miRNA sequences also showed great variability. The average length of a potential cassava pre-miRNA sequence was 108 ± 24 nt; however, the majority of potential cassava pre-miRNA sequences were only 85-100 nt in length. mes-miR167g exhibited the shortest precursor length of 68 nt whereas mes-miR159b exhibited the longest precursor length of 218 nt (Figure 7).

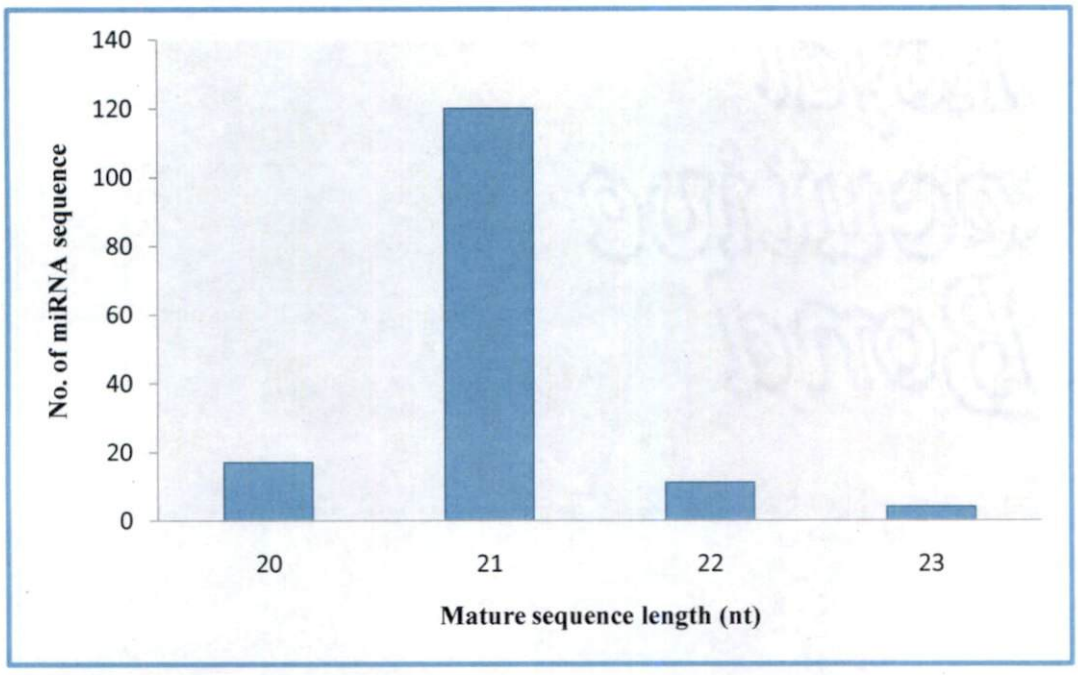


Figure 6. The length distribution of identified cassava mature miRNAs

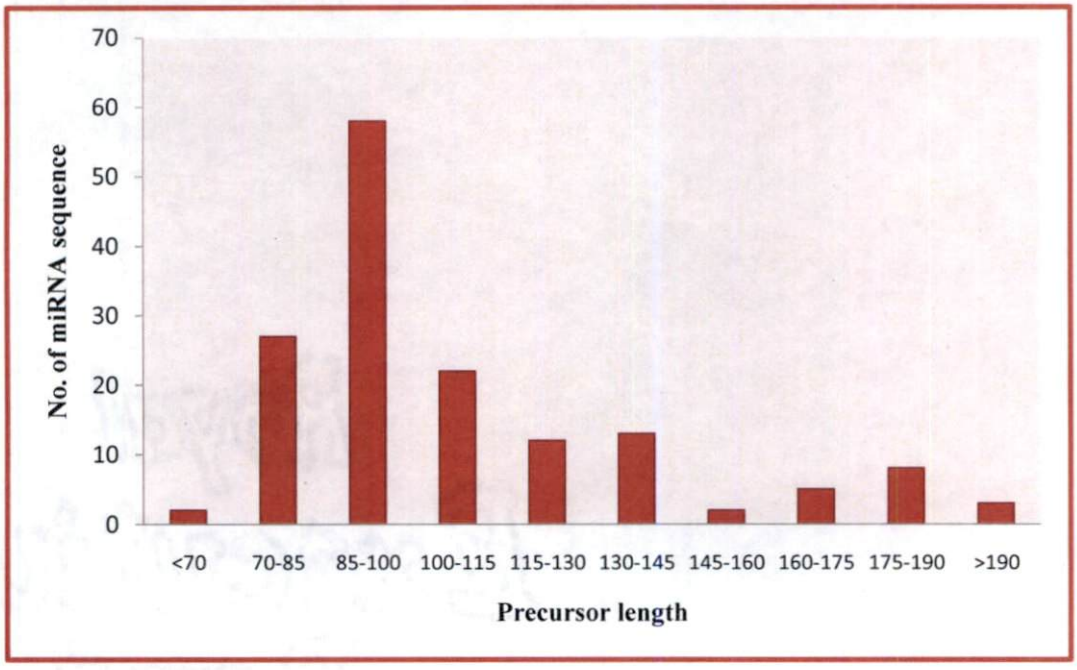


Figure 7. The length distribution of pre-miRNAs identified in cassava

The composition of the four nucleotides (A, G, C, and U) is an important parameter, which is an indicator for species evolution as well as for the stabilization of one specific RNA sequence cased by their secondary structure. The percentage composition of each nucleotide was not evenly distributed in the identified cassava pre-miRNAs. Uracil (U) is dominant in both mature miRNAs and pre-miRNAs. U content varied from 20.2 per cent to 41.5 per cent with an average of 30.7 per cent in the identified cassava pre-miRNAs which is significantly higher than the content of other nucleotides, particularly much higher than nucleotides C (21.7 %), A (23.8 %) and G (23.8 %). Uracil (U) in miRNA may serve as a signal for miRNA biogenesis.

In the identified pre-miRNA, the GC content was much lower than AU content. The nucleotide composition of the newly identified potential cassava miRNA precursor sequences had an average G+C content of 45.5 ± 4.09 per cent and an average A+U content of 54.5 ± 4.05 per cent. The average A/U nucleotide ratio of the potential cassava miRNA precursor sequences was 0.80 ± 0.16. Given that A-U and G-C form two and three hydrogen bonds, respectively, a higher A-U content may make the pre-miRNA secondary structure less stable and thus easier to be processed into mature miRNA by the RISC complex. The distribution of (G+C) % content and (A+U) % content are displayed in Figures 8 and 9 respectively.

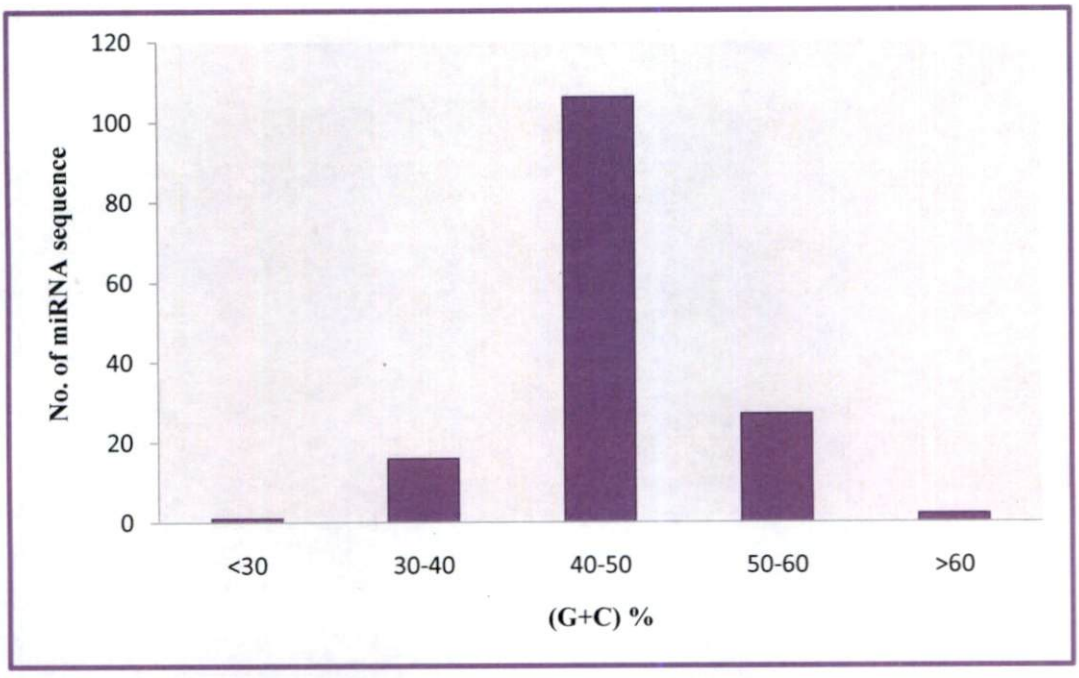


Figure 8. The distribution of (G + C) % content of predicted cassava pre-miRNA

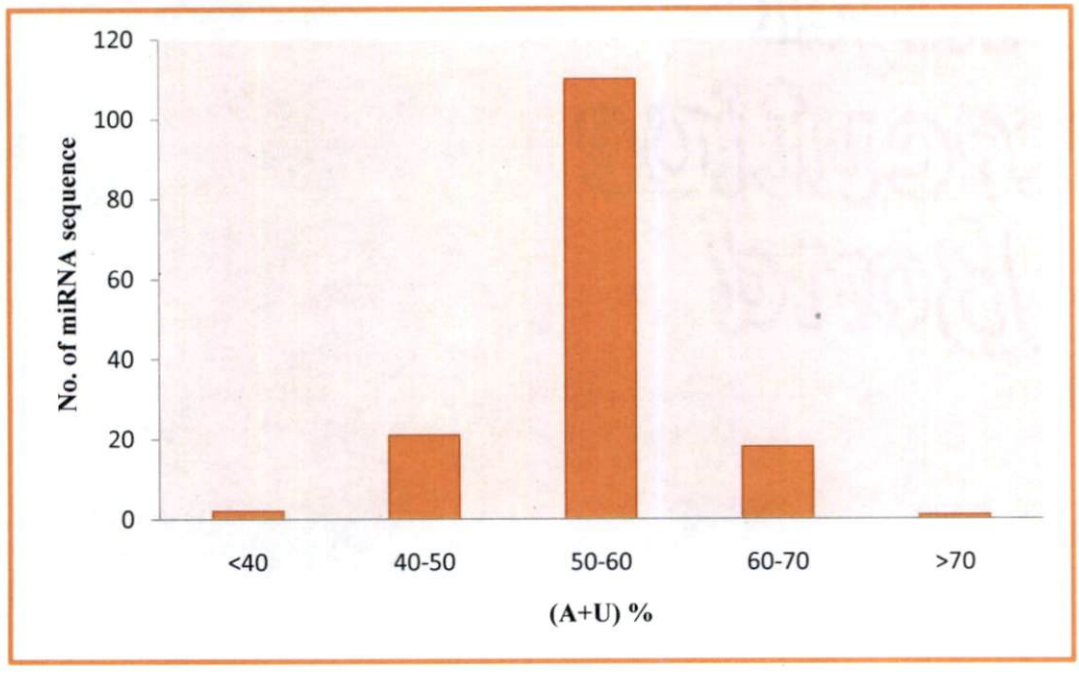


Figure 9. The distribution of (A + U) % content of predicted cassava pre-miRNA

The average minimal folding free energy (MFE) of the putative cassava pre-miRNAs was - 51.14 ± 10.35 kcal mol^{-1}. The energy varied from - 27.4 kcal mol^{-1} to - 93.5 kcal mol^{-1}. The Adjusted Minimal Folding Free Energy (AMFE) of the 152 identified cassava pre-miRNAs ranged from – 31.0 kcal mol^{-1} to – 65.93 kcal mol^{-1} with an average of – 47.92 ± 5.80 kcal mol^{-1}, which is a smaller range as compared with the MFE range. Most of the pre-miRNAs have AMFE values within – 45 kcal mol^{-1} to – 50 kcal mol^{-1}.

The stem-loop hairpin structure is not unique to miRNAs. Therefore, other measures must be used to further validate that a result could potentially be a miRNA. Minimal folding free energy index (MFEI) is a criterion that has been established to eliminate false positives and distinguish miRNAs from other types of RNA molecules (Zhang et al. 2006d). Potential miRNA sequences are more likely to be miRNAs if they have a MFEI value of 0.85 or greater (Zhang et al. 2006d). All the identified potential cassava miRNA displayed MFEI greater than or equal to 0.85. MFEI ranged from 0.85 to 1.5 with an average of 1.06 ± 0.11. This suggests that all the potential cassava miRNA sequences are more likely to be miRNAs than any other type of RNA molecule that can form the same stem-loop hairpin structure, such as tRNA. The distribution of MFE, AMFE, MFEI are given in the Figures 10, 11 and 12 respectively. The major characteristics of predicted cassava pre-miRNA are summarized in the Table 4.

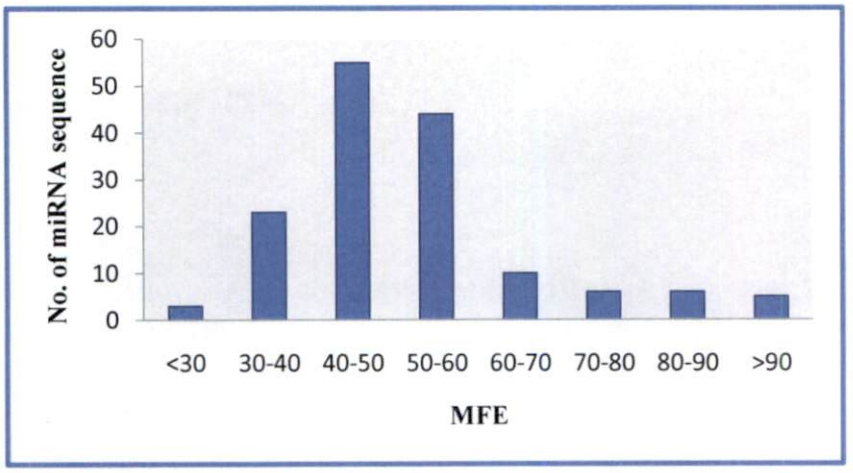


Figure 10. The distribution of MFE of predicted cassava pre-miRNA

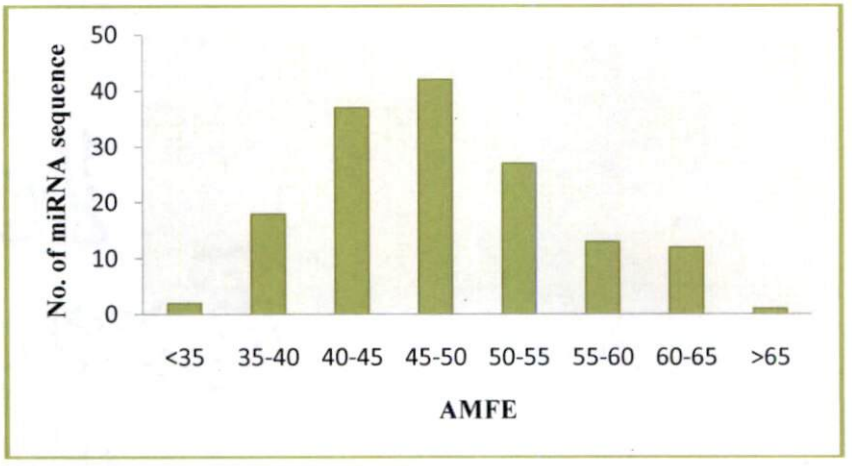


Figure 11. The distribution of AMFE of predicted cassava pre-miRNA

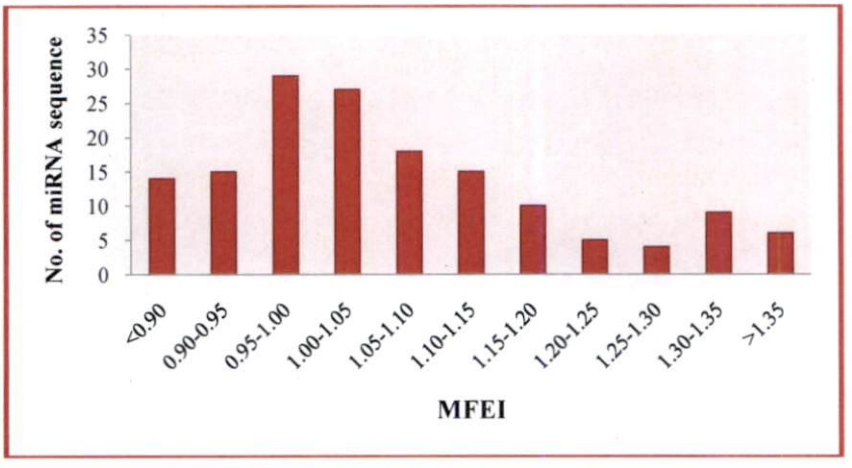


Figure 12. The distribution of MFEI of predicted cassava pre-miRNA

Table 4. Major characteristics of predicted cassava pre-miRNAs

	MINIMUM	MAXIMUM	MEDIAN	AVERAGE	ST. DEVATION
P.L.	68.00	218.00	96.00	108.07	31.60
A %	16.87	37.65	23.70	23.84	3.72
C %	16.05	33.33	21.17	21.71	3.21
G %	10.00	32.79	23.84	23.78	3.75
U %	20.22	41.46	30.95	30.66	4.87
(G+C) %	26.67	60.67	45.00	45.50	5.41
(A+U) %	39.33	73.33	55.00	54.50	5.35
A/U	0.43	1.78	0.77	0.80	0.21
MFE	-27.40	-93.50	-48.35	-51.14	14.29
MFEI	0.85	1.50	1.03	1.06	0.14

MicroRNAs in plants identify their target mRNAs through perfect or near-perfect complementarity. Based on the mechanism of miRNAs in plants, a homology search based method was used for miRNA target prediction in cassava using psRNAtarget server. The identified miRNAs (152) of cassava were used as queries in the psRNATarget with default parameters to predict the potential mRNA targets in cassava transcripts. A total of about 300 putative target genes, belonging to a variety of gene families that partake in various biological and physiological functions, were identified. These genes are involved in transcription, stress response, structural component, development and metabolism (Appendix III).

Cassava miRNAs targeted several transcription factors like squamosa promoter binding like protein, MYB transcription factors, auxin response transcription factor, APETALA2 etc. In addition to transcription factors, the other targets include various enzymes (laccase, multicopper oxidase, short chain dehydrogenase, protein kinases etc.) that play critical role in various metabolisms. Most of the miRNAs have more than one target. The biological functions of some target genes have not been known yet in cassava. Some of the miRNA families such as miR390 and miR535 failed to identify their target genes.

4.3 PREDICTION OF *INDIAN CASSAVA MOSAIC VIRUS* miRNA AND THEIR TARGETS IN CASSAVA TRANSCRIPTS

The *Indian Cassava Mosaic Virus* genome (DNA A- 2815nt and DNA B-2645 nt) was screened for hairpin structured miRNA precursors in the genome using VMir Analyzer. A total of 109 sequences (50+59) with potential hairpin like structures were predicted as candidate miRNA precursors. In order to reduce the false predictions, the predicted candidates were further screened using MiPred. Then candidates within or antisense to protein coding regions were removed. At the last step the mature miRNA sequences were predicted using Bayes-SVM-MiRNA web server v 1.0.

Three non protein coding sequences were selected as miRNA precursor candidate. The predicted precursors were designated as ICMVmiR-1, ICMVmiR-2 and ICMVmiR-3 (Table 5 and Figure 13). The mature sequences were about 21nt in length and two of them are located in the 5' arm of the precursor. The average precursor length is 78 ± 5 nt. 62.73 ± 10.59 and 37.27 ± 25.46 are the average (G + C) per cent and (A+U) per cent respectively. The MFE and MFEI averages to - 34.33 ± 3.38 kcal/mol and 0.72 ± 0.07 respectively.

A total of 12 gene targets for ICMV miRNA were identified in cassava transcript. This was performed by using psRNA Target prediction tool with default parameters. The predicted viral miRNAs targeted cassava transcripts like WD repeat protein, NADH ubiquinone oxidoreductase related, Spermidine hydroxycinnamoyl transferase, Actin related protein 2/3 (Arp 2/3) complex subunit, NB ARC domain (R protein), RNA binding protein and Lupus la protein related. The various targets identified have functional annotation as those involved in catalysis, regulation, stress response etc. MicroRNA targets in cassava along with functional annotation are presented in Table 6.

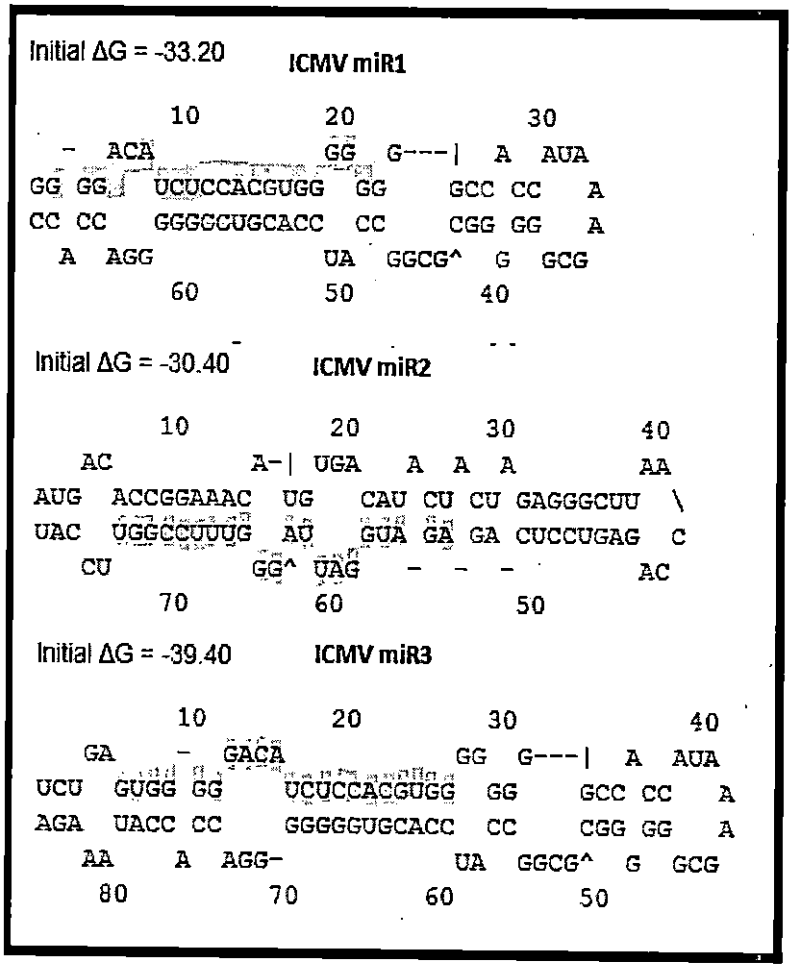


Figure 13. Predicted secondary hairpin structure for the ICMV miRNA (mature miRNA sequences are highlighted)

Table 5. Identified ICMV miRNAs

miRNA	Mature sequence	PL	ARM	(G+C) %	(A+U) %	MFE (kcal/mol)	MFEI
ICMV miR-1	GGACATCTCCACGTGGGGGGG	70	5'	74.28	25.72	-33.20	0.64
ICMV miR-2	AGATGGATTAGGGTTTCCGGT	79	3'	46.84	53.16	-30.4	0.82
ICMV miR-3	GTGGGGGACATCTCCACGTGG	85	5'	67.06	32.94	-39.4	0.69

Table 6. ICMV miRNA targets in cassava

miRNA	TARGET	TARGETED PROTEIN
ICMV miR-1	cassava 4.1_000272m	WD repeat protein
	cassava 4.1_013419m	NADH ubiquinone oxidoreductase related
ICMV miR-2	cassava 4.1_032275m	Transferase
	cassava 4.1_028642m	Protein of unknown function DUF607
	cassava 4.1_010437m	oxidoreductase
	cassava 4.1_009712m	Actin related protein 2/3 complex subunit
	cassava 4.1_031128m	NB ARC domain (R protein)
	cassava 4.1_009101m	RNA recognition motif
	cassava 4.1_009109m	RNA binding protein family isoform
ICMV miR-3	cassava 4.1_011575m	Lupus la protein related
	cassava 4.1_000272m	WD repeat protein
	cassava 4.1_023678m	Lupus la protein related

4.4 IDENTIFICATION OF *Manihot esculenta* miRNA TARGETS IN THE GENOME OF CASSAVA MOSAIC VIRUS

Bioinformatics approach was applied to identify potential cassava miRNA regulated genes in cassava mosaic virus genome. DNA A and DNA B nucleotide sequences of nine different strains of cassava mosaic virus namely *East African Cassava Mosaic Virus, African Cassava Mosaic Virus* isolate West Kenyan 844 segment, *East African Cassava Mosaic Zanzibar Virus, East African Cassava Mosaic Cameroon Virus, East African Cassava Mosaic Kenyan Virus, Indian Cassava Mosaic Virus, Srilankan Cassava Mosaic Virus, East African Cassava Mosaic Malawi Virus clone* and *South African Cassava Mosaic Virus* were downloaded from NCBI GenBank and used as target transcripts. Mature cassava miRNA from miRBase served as query in miRanda target prediction algorithm.

Potential regulatory targets, having 5 or fewer mismatches and with no gaps in full length nucleotide sequences were identified. Fourteen miRNA families, namely mes-miR159, mes-miR164, mes-miR167, mes-miR168, mes-miR171, mes-miR319, mes-miR394, mes-miR395, mes-miR397, mes-miR408, mes-miR477, mes-miR482, mes-miR1446 and mes-miR2275 were found to have the potential to target cassava mosaic virus genome with nearly perfect complementarities. Most cassava miRNAs families (85.7 %) targeted DNA A. They include mes-miR159, mes-miR164, mes-miR167, mes-miR168, mes-miR171, mes-miR319, mes-miR395, mes-miR397, mes-miR408, mes-miR477, mes-miR1446 and mes-miR2275. AC1 gene was targeted by 16 different miRNAs, followed by AV1/AC4, AC3 and AC2/AV2 which showed 12, 2 and 1 putative target respectively. BV1 of DNA B was targeted by 7 miRNAs while BC1 was targeted by no miRNA. The miRNA families mes-miR159 and mes-miR164 had targets in both DNA A and DNA B.

East African Cassava Mosaic Zanzibar Virus genome has potential targets for 16 cassava miRNAs followed by *East African Cassava Mosaic Kenyan Virus* (13), *Indian Cassava Mosaic Virus* (11) and *Srilankan Cassava Mosaic Virus* (8).

The remaining five strains are targeted by less than 5 cassava miRNAs. ICMV and SLCMV strains seen in India were targeted at BV1 by mes-miR164d. The average free energy, miRanda score and percent of sequence complementarity between miRNA-target duplex are -22.01 ± 1.02 kcal/mol, 112.23 ± 13.42 and 80.31 ± 3.51 per cent respectively. The cassava miRNA - cassava mosaic virus genome interaction is summarized in Table 7.

Table 7. Cassava miRNA targets in cassava mosaic virus

	AV1	AV2	AC1	AC2	AC3	AC4	BV1	BC1
ACMV	164d; 408; 1446	—	—	—	—	—	482	—
EACMV	1446; 408	2275	—	—	—	—	—	—
EACMZV	168a; 171g,h,i,j,k	2275	395a,b,c,d	—	—	—	159c,d; 394c	—
EACMCV	408	2275	—	395e	395e	—	—	—
ICMV	319h;408	—	—	—	—	—	164d	—
SLCMV	319f,g	—	477a,b,c,d,e	—	—	—	164d	—
EACMKV	168a	2275	319a,b,c,d,e,f,g; 408	—	—	—	159c,d; 394a,b	—
EACMMV	—	—	—	—	397	—	—	—

4.5 EXPERIMENTAL VALIDATION OF miRNA

Cassava miRNAs (mes-miR164 and mes-miR395) having targets in cassava mosaic virus genome was randomly selected. mes-miR395 and mes-miR164 had targets in cassava mosaic virus DNA A and DNA B respectively. The expression of miRNAs was detected using a two step process. In the first, stem loop RT primers was hybridized to miRNA molecule and then reverse transcribed in a pulsed reverse transcription reaction. In second, the RT product was amplified and quantified using SYBR Green I assay.

4.5.1 miRNA sequences and primers

Based on the mature miRNA sequences of mes-miR164 and mes-miR395, stem loop RT primers (164RT, 395RT) and forward primers (164F, 395F) were designed. A universal primer based on stem loop sequence was used as reverse primer for qPCR reaction. The primer details are given in the Table 8.

4.5.2 Total RNA isolation

Young leaves from healthy and cassava mosaic disease infected cassava was used for extraction of total RNA. A distinct or intact RNA with no apparent RNA degradation and minimum genomic DNA contamination were observed on agarose gel, showing good quality total RNA extraction (Plate 1).

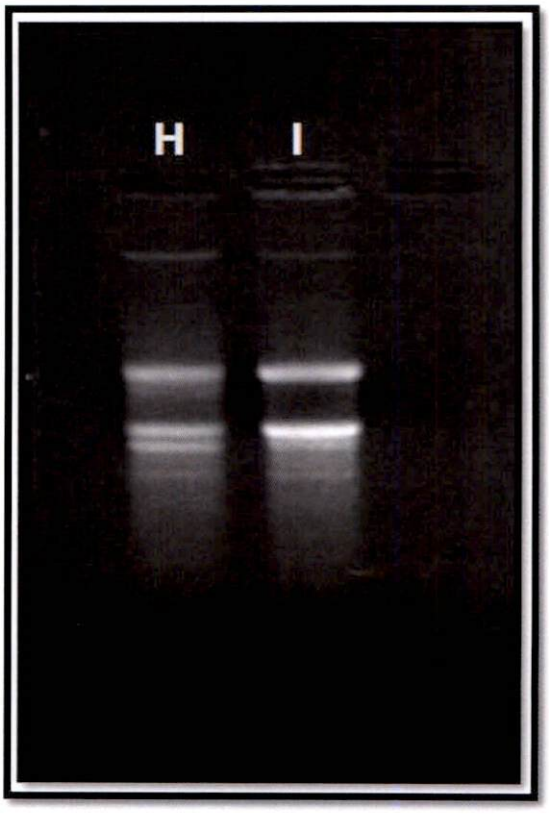
4.5.3 Stem loop pulsed reverse transcription

The reverse transcription of extracted total RNA was carried out with stem loop RT primer (164RT and 395RT) to get cDNA of corresponding miRNA. Positive results were viewed in the agarose gel. The samples yielded weak amplicons. Control reactions were negative.

4.5.4 miRNA SYBR Green I Assay

Quantitative Real Time PCR (qPCR) is a reliable method to determine the expression of specific miRNAs. qPCR was used to detect the expression of the two predicted miRNAs. The result shows that all of the 2 predicted miRNAs exist

Plate 1. Gel image of RNA isolated from healthy (H) and cassava mosaic disease infected (I) cassava leaf tissue.



and are expressed in cassava mosaic disease infected and healthy cassava leaf samples (Figure 14). Figure 15 shows the melt curve of the assay.The threshold cycle (C_T) values for the miRNA amplification are listed in the Table 9.

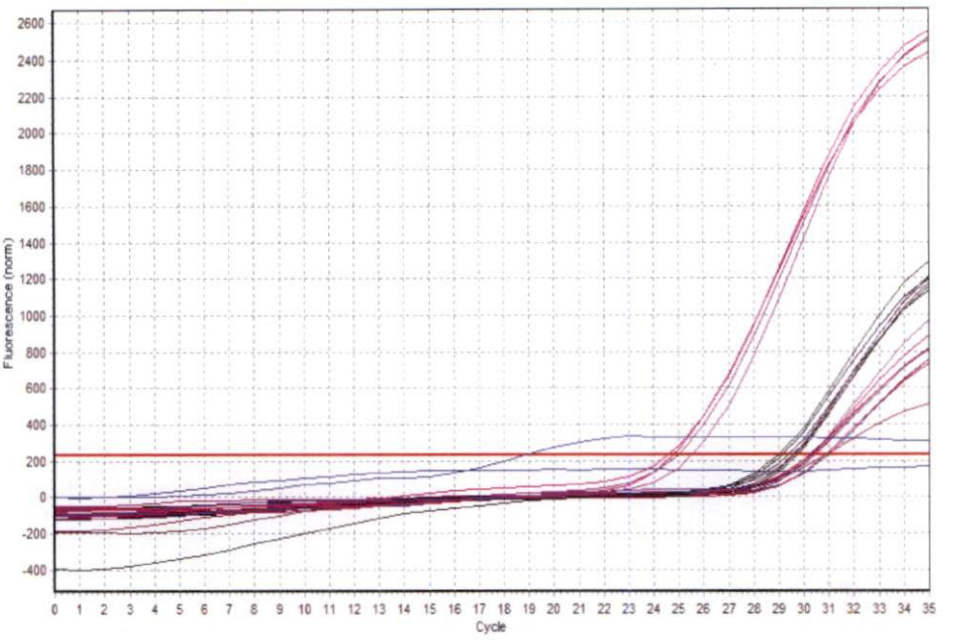
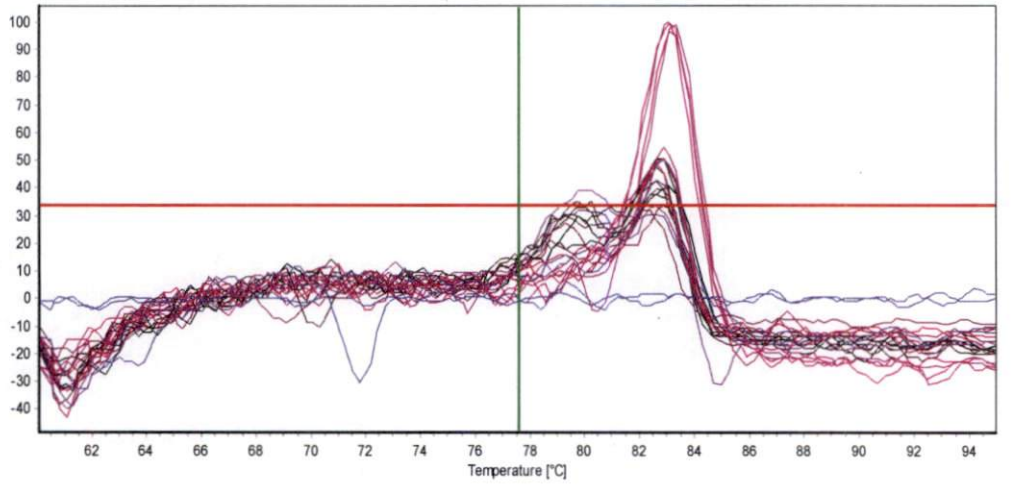


Figure 14. Real time PCR for mes-miR164 and 395 validations with designed primers



Threshold: 33%

Figure 15. Melt Curve Analysis

Table 8. Sequences of designed primers

Target miRNA	Primer name	Primer sequence (5' to 3')	No. of bases	GC content	T_m (^oC)
mes-miR164	164RT	GTCGTATCCAGTGCAGGGTCCGAGT ATTCGCACTGGATACGACAGCAG	50	56	70.2
	164F	GGTTGGAGAAGCAGGGCACA	20	60	60.3
	164R	CCAGTGCAGGGTCCGAGGTA	20	65	61.3
mes-miR395	395RT	GTCGTATCCAGTGCAGGGTCCGAGT ATTCGCACTGGATACGACGAGTC	50	56	70.5
	395F	CGGCTGAAGTGTTTGGGGGA	20	60	60.2
	395R	CCAGTGCAGGGTCCGAGGTA	20	65	61.3

Table 9. Mean C_T value in qPCR amplification

miRNA	C_T value		Amplification
	Infected (I)	Healthy (H)	
mes-miR164	29.18	29.67	Positive
mes-miR395	30.90	24.73	Positive

DISCUSSION

5. DISCUSSION

The endogenous small non-coding functional microRNAs (miRNAs) are short in size, range from ~ 21 to 24 nucleotides in length and play a pivotal role in gene expression in cassava by silencing genes either by destructing or blocking of translation of homologous mRNA. Although various high-throughput, time consuming and expensive techniques like forward genetics and direct cloning are employed to detect miRNAs in plants still, comparative genomics complemented with novel bioinformatic tools pave the way for efficient and cost effective identification of miRNAs through homologous sequence search with previously known miRNAs (Panda *et al.,* 2014). Plant viruses invade and cause infections by utilizing the biosynthetic pathways of host cell, but plants have evolved strategies to resist virus and other pathogen attacks. RNA silencing is one of the main adaptive defense mechanisms against pathogens including viruses (Pantaleo 2011). To counteract this host defense, viruses encode specific RNA silencing suppressor mechanisms. Many studies have revealed that viruses can also encode miRNAs, which are proposed to be involved in RNA silencing suppressor mechanisms, viral replication and persistence (Song *et al.,* 2011; Gao *et al.,* 2012; Vishwanathan *et al.,* 2014). Cassava mosaic virus is a pathogenic *Begomovirus* that severely affect cassava cultivation. Studies on virus encoded miRNAs and their function have profound insights for understanding the infection and the pathogenic mechanisms.

5.1. REVIEW OF miRNA TARGET PREDICTION TOOLS AND COMPARATIVE EVALUATION

The basic principle of many computational methods is to learn from known examples in order to find new ones and make better predictions. Since all the mechanisms behind miRNAs and their actions are not completely revealed, computational tasks associated with miRNA studies are often posed as a challenging with limited prior information (Yoon and Micheli, 2006). In spite of such difficulties, several algorithms have been developed but because the field of

miRNA research is still in its blooming phase and the understanding at the molecular level are yet not very clear, the process of algorithms development for miRNA target identification may not be completely exhaustive. It demands more understanding of the molecular aspect of miRNA biology and more clarity is needed to develop more accurate and efficient tools for the aforesaid purpose. Thus a comparative analysis of the performance of the various target prediction tools available might be useful to carry out further research work in this area (Sinha *et al.,* 2009).

The miRNA target prediction tools: psRNATarget and miRanda were compared for high throughput miRNA target prediction in *Arabidopsis* at optimized parameters. The selections of these tools were based on their easy availability and execution time. Execution time is the time required by the program to predict targets in transcripts for a given set of miRNAs. Tools with long execution times may not be suitable for high throughput analysis. The plant small RNA target (psRNATarget) involves a dynamic programming approach, aligning sequences using a modified Smith-Waterman algorithm. The miRanda algorithm aligns a miRNA to target based on complementarities of nucleotides. psRNAtarget had better accuracy, MCC, PPV and specificity but lower sensitivity compared to miRanda. The reason for increased sensitivity of miRanda target prediction algorithm is that it predicted a large number of targets per miRNA. Such high number of predictions indicate that the tool use algorithm that may not be relevant into plant miRNA target identification due to difference in the mechanism of target recognition in plants and animals. miRanda is routinely used for target prediction in humans and other model organisms. So psRNATarget was found to be a better tool in identification of plant miRNA targets compared to miRanda.

'Precision' and 'recall' are important evaluative parameters to measure accuracy and sensitivity of prediction. To determine the most suitable threshold/cut offs in psRNATarget, 'precision' and 'recall' were calculated at all possible maximum expectation (e) values. Scores at which 'precision' and 'recall'

value intersected were considered optimal for the tool. Optimum maximum expectation value in psRNATarget was found to be 3. At this value the parameters 'recall' and 'precision nearly intersected (Figure 16).

5.2. IDENTIFICATION OF CASSAVA miRNAs AND THEIR TARGETS

In plant kingdom, a substantial number of miRNAs are conserved in different plant species, in lineages from mosses and gymnosperms to flowering plants. Such homologous miRNA families typically have conserved and essential regulatory functions across many plants. Computational methods have been successfully used to predict hundreds of miRNAs in a wide variety of plant species (Patanum *et al.*, 2012). In the study 152 potentially conserved miRNAs in cassava belonging to 30 different families were identified. The newly identified potential cassava miRNAs exhibited a wide range of characteristics between different families and even among members of the same family. For example, while only one member was identified for a majority of the miRNA families, a total number of 27 potential miRNAs were identified for the miR169 family and 12 miRNAs were identified for the miR171 family.

Predicted cassava miRNAs have similar characteristics to the miRNAs in other plant species (Sunkar *et al.*, 2005; Zhang *et al.*, 2006a; Zhang *et al.*, 2008; Frazier *et al.*, 2010; Panda *et al.*, 2014). Potential cassava miRNA precursor sequences exhibited diversity. Most animal miRNAs have precursor lengths of ~70-100 nt (Ambros 2004), however, plant miRNA precursor sequences have been shown to be highly variable (Zhang *et al.*, 2006b). The length of the potential cassava pre-miRNAs varied greatly among members of the same family. This suggests that members of the same family can have differing expression patterns, such as in a spatiotemporal or tissue specific manner (Zhang *et al.*, 2008).

miRNA precursor sequences have a higher negative minimal folding free energy (MFE) compared to other RNAs as they form a stable stem-loop hairpin structure (Bonnet *et al.*, 2004). However, other RNAs, such as tRNAs and rRNAs,

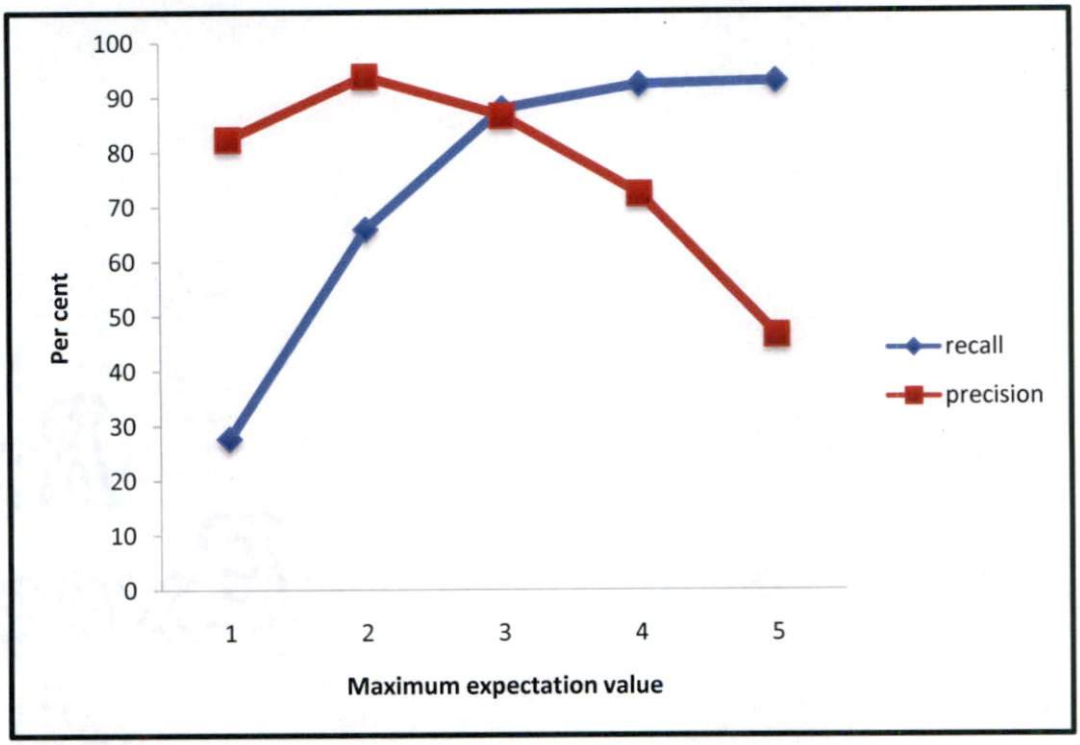


Figure 16. Recall and precision at different psRNATarget maximum expectation value

can also form the hairpin structure. Therefore, prediction of potential miRNAs cannot be based on MFE alone. A new criterion, minimal folding free energy index (MFEI), distinguishes miRNAs from other types of RNA molecules if they have an MFEI value of 0.85 or greater (Zhang *et al.,* 2006d). The potential cassava miRNAs had MFEI values of 0.85 to 1.50 with an average of 1.06 ± 0.11. This value is significantly higher than the MEFI values of other RNAs such as tRNAs (0.64), rRNAs (0.59), and mRNAs (0.62-0.66) (Zhang *et al.,* 2006d). Therefore, the predicted potential cassava miRNAs are more likely to be miRNAs than any other type of RNA molecule.

The knowledge on target function of the identified cassava miRNA will help us to gain insight into the important function and regulation of miRNAs in this plant. Plant miRNAs are a perfect or near-perfect match to their target mRNAs and help regulate post-transcriptional gene expression by binding to mRNAs and promoting mRNA degradation, or binding and inhibiting protein translation (Bartel, 2004). Most plant miRNAs that bind to mRNAs lead to transcript cleavage (Bartel, 2004), however, some miRNAs have been shown to inhibit protein translation in plants (Aukerman and Sakai, 2003). For the 152 potential cassava miRNAs, nearly 300 potential target genes were predicted. miRNAs have been shown to target transcription factors as a means of regulating plant growth and development (Jones-Rhoades and Bartel, 2004). It is predicted that miR156 in cassava targets the SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) transcription factor, a protein that is involved in regulating developmental timing. miR156 has also been shown to target SPL in *Arabidopsis* (Jones-Rhoades *et al.,* 2006) and is predicted to target SPL in other plant species (Zhang *et al.,* 2006e). MYB transcription factors, a large family of proteins functionally diverse in plants, represent key regulators for controlling plant development, metabolism as well as responses to biotic and abiotic stress response (Dubos *et al.,* 2010). This transcription factor was identified as miR159 targeted genes in cassava. Additionally, the potential cassava miR172 will target APETALA2-like (AP2) proteins. This is similar to other reports that have shown that miR172 targets AP2

and AP2-like genes, ultimately promoting floral organ maturation (Aukerman and Sakai, 2003). Almost all of the 152 potential cassava miRNAs were predicted to target multiple genes. The target genes are involved in a wide variety of biological and metabolic pathways. Two of the cassava miRNA families failed to identify targets; this may be due to the incompletion in the understanding of cassava genome.

5.3. PREDICTION OF *INDIAN CASSAVA MOSAIC VIRUS* miRNA AND THEIR TARGETS IN CASSAVA TRANSCRIPT

Viral miRNAs are less conserved than plant and animal miRNAs suggesting that they evolve rapidly. VMir Analyzer was used in identification of three *Indian cassava mosaic virus* miRNAs of about 21 nt in mature sequence length. The average precursor length is 78 ± 5 nt. This is similar to the average viral miRNA precursor length predicted by Pan *et al.* (2007). The viral miRNAs are much smaller than plant and animal miRNAs. The predicted miRNAs had a high MFEI and MFE.

To understand the biological function of cassava mosaic virus encoded miRNAs in mosaic disease infection, it is necessary to identify their targets in host genome. The predicted viral miRNAs targeted cassava transcripts like WD repeat protein, NADH ubiquinone oxidoreductase related, Spermidine hydroxycinnamoyl transferase, Actin related protein 2/3 (Arp 2/3) complex subunit, NB ARC domain (R protein), RNA binding protein and Lupus la protein related. Most of these proteins have a direct or indirect role in counteracting viral infection and inducing apoptosis. WD repeat proteins are a large family found in all eukaryotes and are implicated in a variety of function ranging from signal transduction and transcription regulation to cell cycle control, autophagy and apoptosis (Smith *et al.,* 1999). Mitochondria from plants contain NADH ubiquinone oxidoreductase related that catalyze the redox reaction in respiratory chain. It also has a role in triggering apoptosis (Chomova and Racay, 2010). Spermidine hydroxycinnamoyl transferase has the function to acylate spermidine

(polyamine). Polyamine metabolism is known to have role in plants responding to abiotic environmental stresses and undergo profound changes in plants interacting with fungal and viral pathogens (Walters, 2003). Arp 2/3 complex plays a major role in regulation of actin cytoskeleton. It is essential for regulation of intracellular motility of endosomes, lysosomes, pinocyctic vesicles and mitochondria (Mathur, 2005). NB ARC domain is a central nucleotide binding domain in resistance (R) protein. Resistance proteins in plants are involved in pathogen recoginition and subsequent activation of innate immune responses (van Ooijen *et al.,* 2008). RNA recognition motif is putative RNA binding domains that are known to bind to ssRNAs. Some of the RNA binding proteins have been involved in the inhibition of RNA virus replication, movement and specific binding. Host plant use RNA binding proteins for defense against viral infection (Huh and Paek, 2013).

5.4. IDENTIFICATION OF *Manihot esculenta* miRNA TARGETS IN THE GENOME OF CASSAVA MOSAIC VIRUS

A bioinformatics approach was applied to identify endogenous *M. esculenta* miRNAs having anti-cassava mosaic disease defense by targeting DNA A and B of various strains of cassava mosaic virus. DNA A and B were targeted at several loci by various miRNAs. The degree of complementarity determines the fate of a target site. Perfect complementarity leads to endonucleolytic cleavage, while imperfect complementarity results in translational repression leading to destabilization of miRNA (Baek *et al.,* 2008; Selbach *et al.,* 2008).

The targeted regions in DNA A were mainly associated with replication (AC1), silencing suppressor (AC4) and coat protein (AV1). AC1 is involved in replication of viruses within the host cells. The open reading frame (ORF) encodes a replication associated proteins (Hull, 2002). Targeting AC1 gene would impact viral replication by reducing viral DNA accumulation in host. For *begomo viruses*, AC1 gene in sense and antisense orientation has also been used with various success rates against *Bean golden mosaic virus* (Bonfim *et al.,* 2007) and

African cassava mosaic virus (Vanderschuren *et al.*, 2009). AC1 is targeted by 16 miRNAs. Among them mes-miR395 showed maximum complementarity (85.71 %). AC4 plays a role as host activation protein which serves as an important symptom determinant implicated in cell cycle control and may also counteract the host response to replication gene expression (Hull, 2002). AC4 protein in *African cassava mosaic virus* and *Srilankan cassava mosaic virus* act as a suppressor of gene silencing (Vanitharani *et al.*, 2004). AC4 is targeted by 12 cassava miRNAs. AV1 coding coat protein is the target for 12 cassava miRNAs. DNA B was targeted at BV1 region by mes-miR159, mes-miR164, mes-miR394 and mes-miR482. BV1 region is responsible for coding nuclear shuttle protein (NSP) that is responsible for intra cellular movement of virions within host plant cell (Stanley *et al.*, 2005).

This suggests that these miRNA families and corresponding artificial miRNA (amiRNA) constructs can be manipulated as a strategy to engineer anti-Cassava Mosaic Virus defense in cassava. Controlling viruses following degrading their mRNAs within a plant cell is a relatively straight forward process and can be effectively achieved using amiRNAs. Niu *et al.* (2006) used a 273 bp sequence of *Arabidopsis* mir159a pre-miRNA transcript expressing amiRNAs against viral suppressor genes to generate resistance against *Turnip yellow mosaic virus* and *Turnip mosaic virus infection*. Conclusively, co-expression of amiRNAs targeting different domains of Cassava Mosaic Virus genome and may lead to development of broad spectrum resistance against Cassava Mosaic Virus infection (Baig and Khan, 2013).

5.5. EXPERIMENTAL VALIDATION OF miRNA

MicroRNAs have been shown to have variable expression patterns with regard to tissue differentiation and developmental stages (Zhang *et al.*, 2006c). qRT PCR was used to confirm the expression of 2 potential cassava miRNAs in healthy and cassava mosaic disease infected cassava leaf sample. Using the method existence of mes-miR164 and mes-miR395 in young cassava leaf sample

was validated. The level of miRNA expression was deduced from C_T .The level of expression of mes-miR164 was almost same in healthy and infected leaf sample (Figure 17). There was a slight increase in mes-miR395 expression in healthy leaf sample compared to cassava mosaic disease infected (Figure 18).

During sulfate limitation, expression of miR395 is significantly up-regulated. miR395 targets two families of genes, ATP sulfurylases (encoded by APS genes) and sulfate transporter 2;1 (SULTR2;1, also called AST68), both of which are involved in the sulfate metabolism pathway. Their transcripts are suppressed strongly in miR395-over-expressing transgenic plant (Liang et al., 2010). miR164 negatively regulates through mRNA cleavage, several genes that encode NAC like transcription factors (Baker *et al.,* 2005). These genes include CUP SHAPED COTYLEDON (CUC) which are expressed in and necessary for the formation of boundaries between meristems and emerging organ primodia (Heisler *et al.,* 2005). mir164 family have been reported to prevent lateral root to prevent lateral root initiation (Guo *et al.,* 2005).

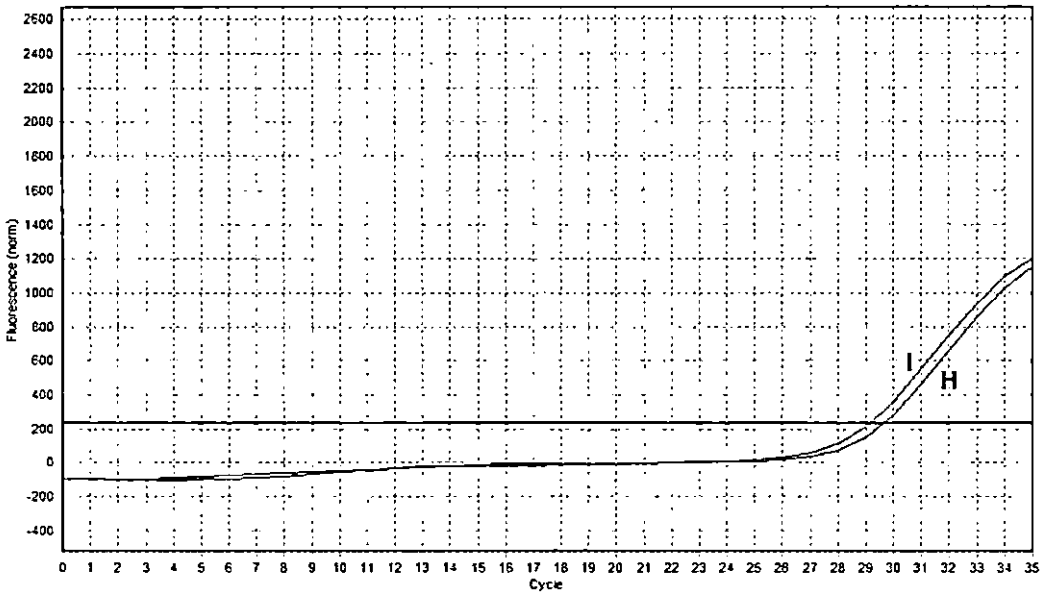


Figure 17. mes-miR164 qPCR amplification in healthy and infected cassava cDNA samples

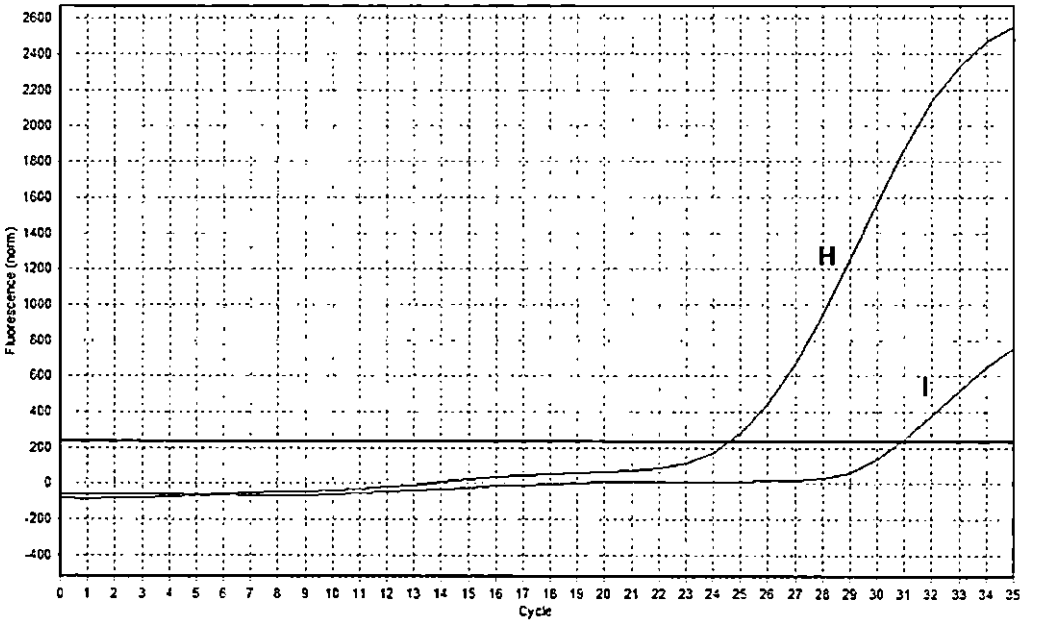


Figure 18. mes-miR395 qPCR amplification in healthy and infected cassava cDNA samples

SUMMARY

6. SUMMARY

The study entitled "Evaluation of prediction tools and computational analysis of microRNAs in cassava (*Manihot esculenta* Crantz.)" was conducted at the CTCRI, Sreekariyam, Thiruvananthapuram during August 2013 to June 2014. The objectives of the study included comparison and evaluation of miRNA prediction software, computational prediction and annotation of miRNA in cassava and understanding the role of miRNA-mRNA interaction in cassava in biotic stress responses (Cassava Mosaic Virus). The salient findings of the study are summarized below.

MicroRNA target prediction tool psRNATarget was found to be a better tool compared to miRanda target prediction tool in the prediction of plant miRNA targets. psRNATarget had better specificity and accuracy in target prediction compared to miRanda. The optimum maximum expectation value of psRNATarget was found to be 3 where the values of recall and precision nearly intersected.

About 152 potential cassava miRNAs belonging to 30 families were identified via homology search. Eight of the identified cassava miRNAs were new and not listed in miRBase microRNA database. The majority of the predicted miRNAs were 21 nt in length and found in 5' arm of stem loop hairpin secondary structure. The average pre-miRNA length was about 108 ± 24nt. The predicted cassava pre-miRNA had a high MFE, AMFE and MFEI values. The miRNAs had ahigher A+U content than G+C content. About 300 potential target genes were identified for the predicted miRNAs. They include transcription factors, metabolic enzymes etc. Most of the predicted cassava miRNAs were shown to have multiple targets within the cassava transcripts. Two of the miRNAs failed to show targets in cassava. Conserved miRNAs do exist in cassava and play an important role in cassava growth and development.

Three Indian cassava mosaic virus (ICMV) miRNAs with 21 nt in length were identified. Their targets in cassava include genes involved in catalysis,

regulation and stress response. The ICMV miRNAs may have a role in combating the plant defense.

Fourteen cassava miRNA families were found to target Cassava mosaic virus genome with nearly perfect complementarity. Most of the families targeted DNA A. AC1 involved in the viral replication was the most targeted open reading frame in DNA A. It was targeted by about 14 cassava miRNA families. Seven cassava miRNA families targeted BV1 gene in DNA B. this suggests that cassava may have some miRNAs which may pose to have action against viral transcripts and thus help the plant to recover from the same. These miRNA families can be manipulated as a strategy to engineer anti-Cassava Mosaic Virus defense in cassava.

Two of the cassava miRNAs (mes-miR164 and 395) having target in Cassava Mosaic Virus was randomly selected and their presence validated in healthy and CMD infected cassava leaf samples by qRT PCR. The miRNAs showed amplification in both the samples during PCR indicating their presence in both healthy and diseased conditions.

REFERENCES

7. REFERENCE

Akter, A., Islam, M., Mondal, S. I., Mahmud, Z., Jewel, N.A., Ferdous, S., Amin, R. and Rahman, M. 2013. Computational identification of miRNA and targets from expressed sequence tags of coffee (*Coffea arabica*). *Saudi J. Biol. Sci.* 21 (1): 3-12.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.

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

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

Aukerman, M.J. and Sakai, H. 2003. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *Plant Cell* 15: 2730-2741.

Babu, R.N., Jyothi, M.N., Sharadamma, N., Sahu, S., Rai, D.V. and Devaraj, V. R. 2013. Computational identification of conserved micro RNAs from kodo millet (*Paspalum scrobiculatum*). *Afr. Crop Sci. J.* 21: 75 – 83.

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P. and Bartel, D.P. 2008. The impact of microRNAs on protein output. *Nature* 445 (7209): 64-71.

Baig, H.S. and Khan, J.A. 2013. Identification of Gossypium hirustum miRNA targets in the genome of cotton leaf curl multan virus and betasatellite. *Indian J. Biotechnol.* 12: 336-342.

Baker, C.C., Sieber, P., Wellmer, F. and Meyerowitz, E.M. 2005. The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in Arabidopsis. *Curr. Biol.* 15 (4): 303–315.

Barrera-Figueroa, B.E., Gao, L., Diop, N.N., Wu, Z., Ehlers, J.D., Roberts, P.A., Close, T.J., Zhu, J.K. and Liu, R. 2011. Identification and comparative

analysis of drought-associated microRNAs in two cowpea genotypes. *BMC Plant Biol.* 11: 127.

Bartel, B. and Bartel, D. P. 2003. MicroRNAs: at the root of plant development?. *Plant Physiol.* 132: 709–717.

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

Benes, V. and Castoldi, M. 2010. Expression profiling of microRNA using real time quantitative PCR, how to use it and what is available. *Methods* 50: 244-49.

Bennasser, Y., Le, S.Y., Benkirane, M. and Jeang, K.T. 2005. Evidence that HIV-1 encodes an siRNA and a suppressor of RNA silencing. *Immunity* 22: 607–619.

Bennasser, Y., Le, S.Y., Yeung, M.L. and Jeang, K.T. 2004. HIV-1 encoded candidate microRNAs and their cellular targets. *Retrovirology* 1: 43.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., Sharon, E., Spector, Y and Bentwich, Z. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* 37: 766–770.

Berezikov, E., Cuppen, E. and Plasterk, R.H. A. 2006. Approaches to microRNA discovery. *Nat. Genet.* 38: S2-7.

Bernstein, E., Candy, A.A., Hammond, S.M. and Hannon, G.J. 2001. Role of bidentate ribonuclease in initiation step of RNA interference. *Nature* 409: 363-366.

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.

Blaszczyk, J., Tropea, J.E., Bubunenko, M., Routzahn, K.M., Waugh, D.S., Court, D.L. and Ji, X. 2001. Crystallographic and modelling structure of RNAseIII suggest a mechanism for double stranded RNA cleavage. *Structure* 9: 1225-1236.

Bonfim, K., Faria, J.C., Nogueria, E.D.P.L., Mendes, E.A. and Aragao, F.J.L. 2007. RNAi mediated resistance to *Bean golden mosaic virus* in genetically engineered common beans (*Phaseolus vulgaris*). *Mol. Plant Microbe In.* 20 (6): 717-726.

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. *Bioinformatics* 26 (12): 1566–1568.

Bonnet, E., Wuyts, J., Rouze, P. and Van de Peer, Y. 2004. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* 20: 2911-2917.

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

Cai, X., Lu, S. and Zhang, Z. 2005. Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. *Proc. Natl. Acad. Sci.* 102: 5570–5575.

Carl, J.W., Trgovcich, J. and Hannenhalli, S. 2013. Wide spread evidence of viral miRNAs targeting host pathways. *BMC Bioinfo.* 14 (2): S3.

Carmen, S.M. and Juan, A.G. 2006. MicroRNA-guided processing impairs plum pox virus replication, but the virus readily evolves to escape this silencing mechanism. *J. Virol.* 80: 2429–2436.

Carthew, R.W. and Sontheimer, E.J. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136 (4): 642–655.

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, X. 2008. microRNA metabolism in plants. *Curr Top Microbiol Immunol.* 320: 117-136.

Chomova, M. and Racay, P. 2010. Mitochondrial complex I in the network of known and unknown facts. *Gen. Physiol. Biophys.* 29 (1): 3-11.

Chuck, G., Meeley, R. and Hake, S. 2008. Floral meristem initiation and meristem cell fate are regulated by the maize AP2 genes *ids1* and *sid1*. *Development* 135: 3013–3019.

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

Dai, X., Zhuang, Z. and Zhao, P.X. 2010. Computational analysis of miRNA targets in plants:current status and challenges. *Brief. Bioinform.* 12 (2): 115-121.

Din, M. and Barozai, M.Y.K. 2013. Profiling microRNAs and their targets in an important fleshy fruit: Tomato (*Solanum lycopersicum*). *Gene* 535: 198–203.

Ding, J., Yu, S., Ohler, U., Guan, J. and Zhou, S. 2011. imiRTP: An Integrated Method to Identifying miRNA-target Interactions in Arabidopsis thaliana. Proceedings of an international conference on Bioinformatics and Biomedicine, Atlanta, pp. 100-104.

Donner, T.J., Sherr, I. and Scarpelia, E. 2009. Regulation of preprocambial cell state acquisition by auxin signaling in *Arabidopsis* leaves. *Development* 136: 3235–3246.

Dsouza, M., Larsen, N. and Overbeek, R. 1997. Searching for patterns in genomic data. *Trends Genet.* 13 (12): 497–498.

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 Pathog.* 7: 8.

Dubos, C., Stracke, R., Groteworld, E., Weisshar, B., Martin, C. and Lepiniec, L. 2010. MYB transcription factors in Arabidopsis. Trends Plant Sci. 15 (10): 573-81.

Dugas, D.V. and Bartel, B. 2004. MicroRNA regulation of gene expression in plants. *Curr. Opin. Plant Biol.* 7: 512–520.

Dunoyer, P., Lecellier, C.H., Parizotto, E.A., Himber, C. and Voinnet, O. 2004. Probing the microRNA and small interfering RNA pathways with virus-encoded suppressors of RNA silencing. *Plant Cell* 16: 1235–1250.

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

Eldem, V., Okay, S. and Unver, T. 2013. Plant microRNAs new players in functional genomics. *Turk J Agric For.* 37: 1-21.

Fahlgren, N., and Carrington, J.C. 2010. miRNA target prediction in plants. *Methods Mol. Biol.* 592: 51–57.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L. and Carrington, J.C. 2007. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS One* 2 (2): e219.

Fauquet, C.M., Briddon, R.W., Brown, J.K., Moriones, E., Stanley, J., Zerbini, M. and Zhous, X. 2008. Geminivirus strain demarcation and nomenclature. *Arch. Virol.* 153 (4): 783-821.

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

Fregene, M. and Puonti-Kaerlas, J. 2002. Cassava Biotechnology. In: Hillocks, R. J., Thresh, J. M. and Bellotti, A. C. (eds), *Cassava: Biology, Production and Utilization.* CAB International, Wallingford, Oxon, pp. 179 – 207.

Gao, R., Liu, P. and Wong, S. M. 2012. Identification of a plant viral RNA genome in the nucleus. *PLoS ONE* 7: e48736.

Gébelin, V., Argout, X., Engchuan, W., Pitollat, B., Duan, C., Montoro, P. and Leclercq, J. 2012. Identification of novel microRNAs in Hevea brasiliensis and computational prediction of their targets. *BMC Plant Biol.* 12: 18.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A. J. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36: D154–D158.

Grundhoff, A. 2011. Computational prediction of viral miRNAs. *Methods Mol. Biol.* 721: 143–152.

Guo, H.S., Xie, Q., Fei, J.F. and Chua, N.H. 2005. MicroRNA directs mRNA cleavage of transcription factor NAC1 to downregulate auxin signals for Arabidopsis lateral root development. *Cell* 17: 1376-86.

Gutierrez, C., Ramirez-Parra, E., Castellano, M.M., Sanz-Burgos, A.P., Luque, A. and Missich, R. 2004. Geminivirus DNA replication and cell cycle interactions. *Vet. Microbiol.* 98: 111 – 119.

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.

Harrison, B.D., Swanson, M.M., and Fargette, D. 2002. Begomovirus coat protein: serology, variations and functions. *Physiol. Mol. Plant Pathol.* 60: 257-271.

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

Heisler, M.G., Ohno, C., Das, P., Sieber, P., Reddy, G.V., Long, J.A., Meyerowitz, E.M. 2005. Patterns of auxin transport and gene expression during primordium development revealed by live imaging of the *Arabidopsis*inflorescence meristem. *Curr. Biol.* 15: 1899–1911.

Hiraguri, A., Itoh, R., Kondo, N., Nomura, Y., Aizawa, D., Murai, Y., Koiwa, H., Seki, M., Shinozaki, K., and Fukuhara, T. 2005. Specific interactions between Dicer-like proteins and HYL1/DRBfamily dsRNA-binding proteins in Arabidopsis thaliana. *Plant Mol. Biol.* 57: 173–188.

Hsieh, L.C., Lin, S.I., Shih, A.C., Chen, J.W., Lin, W.Y., Tseng, C.Y., Li, W. H. and Chiou, T.J. 2009. Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis* by deep sequencing. *Plant Physiol.* 151: 2120–2132.

Huang, S.Q., Xiang, A.L., Che, L.L., Chen, S., Li, H., Song, J.B. and Yang, Z. M. 2010. A set of miRNAs from *Brassica napus* in response to sulphate deficiency and cadmium stress. *Plant Biotechnol. J.* 8: 887–899.

Huh, S.U. and Paek, K-H. 2013. Plant RNA binding proteins for control of RNA virus infection. *Front. Physiol.* 4: 397.

Hull, R. 2002. Genome Organization. In: Hull, R. (ed.), *Matthew's Plant Virology* (4th Ed.). Academic Press, London, UK, pp. 171-224.

Idris, A.M. and Brown, J.K. 1998. Sinaloa tomato leaf curl geminiviruses: biological and molecular evidence for a new subgroups III virus. *Phytopathology* 88: 648-57.

Isam, F., Ralf, R., Wolfgang, R.H. and Wolfgang, F. 2007. Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution. *Plant Biol.* 7: 13.

Jha, A. and Shankar, R. 2011. Employing machine learning for reliable miRNA target identification in plants. *BMC Genomics* 12 (1): 636.

Jian, X., Zhang, L., Li, G., Zhang, L., Wang, X., Cao, X., Fang, X. and Chen, F. 2010. Identification of novel stress-regulated microRNAs from Oryza sativa L. *Genomics* 95: 47–55.

Jiang, P., Wu, H., Wang, W., Ma, W., Sun, X. and Lu, Z. 2007. MiPred-classification of real and pseudo microRNA precursors using random forest model with combined features. *Nucleic Acids Res.* 35: 339-44.

Jin, H., Vacic, V., Girke, T., Lonardi, S. and Zhu, J.K. 2008. Small RNAs and the regulation of cis-natural antisense transcripts in *Arabidopsis*. *BMC Mol. Biol.* 9: 6.

Jones-Rhoades, M.W. and Bartel, D.P. 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* 14: 787-799.

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

Jung, J.H., Seo, P.J. and Park C.M. 2009. MicroRNA biogenesis and function in higher plants. *Plant Biotechnol. Rep.* 3: 111–126.

Katiyar-Agarwal, S. and Jin, H. 2010. Role of small RNAs in host-microbe interactions. *Annu Rev Phytopathol.* 4: 225–246.

Kozomara, A. and Griffith-Jones, S. 2013. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42: 68-73.

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

Kulcheski, F.R., de Oliveria, L.F.V., Molina, L.G., Almerao, M.P., Rodrigues, F.A., Marcolino, J., Barbosa, J.F., Stolf-Moreira, R., Nepomuceno, A.L., Marcelino-Guimaraes, F.C., Abdelnoor, R.V., Nascimento, L.C.,

Carazzolle, M.F., Pereira, G.A.G. and Margis, R. 2011. Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics* 12: 307.

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

Lagos-Quintana, M., Rauhut, R., Lendeckel, W. And Tuschl, T. 2001. Identification of novel genes encoding for small expressed RNAs. *Science* 294: 853-858.

Lai, E.C., Tomancak, P., Williams, R.W. and Rubin, G.M. 2003. Computational identification of Drosophila microRNA genes. *Genome Biology* 4 (7).

Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. 2001. An abundant class of tiny RNAs with probable regulatory role in *C. elegans*. *Science* 294 (5543): 858-62.

Laufs, P., Peaucelle, A., Morin, H. and Traas, J. 2004. MicroRNA regulation of the CUC genes is required for boundary size control in *Arabidopsis* meristems. *Development* 131(17): 4311–4322.

Lee, R.C., Feinbaum, R.L. and Ambros, V. 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75 (5): 843-54.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. and Kim, V. N. 2003. The nuclear RNAse III Drosha initiates microRNA processing. *Nature* 425: 415-419.

Legg, J.P. and Fauquet, C.M. 2004. Cassava mosaic geminiviruses in Africa. *Plant Mol. Biol.* 56 (4): 585-99.

Li, Y., Zhang, Q., Zhang, J., Wu, L., Qi, Y. and Zhou, J.M. 2010. Identification of microRNAs involved in pathogen-associated molecular pattern-triggered plant innate immunity. *Plant Physiol.* 152: 2222–223.

Liang, G. and Yu, D. 2010. Reciprocal regulation among miR395, APS and SULTR2.1 in *Arabidopsis thaliana*. *Plant Signal Behav.* 5(10): 1257-69.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.-J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305: 1437–1441.

Llave, C., Xie, Z., Kasschau, K.D. and Carrington, J.C. 2002. Cleavage of *Scarecrow-like* mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* 297 (5589): 2053– 2056.

Lorenz, R., Bernhart, S.H., Honerzusiederdissen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. 2011. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6 (1): 26.

Lu, S.F., Sun, Y.H., Shi, R., Clark, C., Li, L. and Chiang, V.L. 2005. Novel and mechanical stress-responsive microRNAs in Populus trichocarpa that are absent from Arabidopsis. *Plant Cell* 17: 2186–2203.

Lu, Y., Gan, Q., Chi, X. and Qin, S. 2008. Roles of microRNA in plant defense and virus offense interaction: *Plant cell Rep.* 27: 1571-79.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. 2006. Structural basis for double stranded RNA processing by Dicer. *Science* 311: 195–198.

Maher, C., Stein, L. and Ware, D. 2006. Evolution of Arabidopsis microRNA families through duplication events. *Genome Res.* 16: 510–519.

Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K., and Bartel, D.P. 2004. MicroRNA control of PHABULOSA in leaf development: Importance of pairing to the microRNA 59 region. *EMBO J.* 23: 3356–3364.

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.

Mathur, J. 2005. The Arp 2/3 complex: giving plant cells a leading edge. *BioEssays* 27 (4): 377-87.

Mestdagh, P., Feys, T., Bernard, N., Guenther, S., Chen, C., Speleman, F. and Vandesompele, J. 2008. High-throughput stem-loop RT-qPCR miRNA expression profiling using minute amounts of input RNA. *Nucleic Acids Res.* 36: 143.

Milev, I., Yahubyan, G., Minkov, I. and Baev, V. 2011. miRTour: plant miRNA and target prediction tool. *Bioinformation* 6 (6): 248–249.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E. and Carrington, J.C. 2008. Specificity of ARGONAUTE7-miR390 interaction and dual functionality in *TAS3 trans*-acting siRNA formation. *Cell* 133: 128–141.

Moss, T.Y. and Cullis, C.A. 2012. Computational prediction of candidate microRNA and their targets from completed *Linum ussitatissimum* genome and EST database. *J. Nucleic Acids Invest.* 3: e2.

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. *Science* 312: 436–439.

Niu, Q.W., Lin, S.S., Reyes, J.L., Chen, K.C., Wu, H.W., Yeh, S.D. and Chua, N.H. 2006. Expression of artificial microRNAs in transgenic *Arabidopsis thaliana* confers virus resistance. *Nat. Biotechnol.* 24: 1420-1428.

Nozawa, M., Miura, S., and Nei, M. 2012. Origins and evolution of microRNA genes in plant species. *Genome Biol. Evol.* 4: 230–239.

Omoto, S. and Fujii, Y.R. 2005. Regulation of human immunodeficiency virus-1 transcription by nef microRNA. *J. Gen. Virol.* 86: 751–755.

Omoto, S., Ito, M., Tsutsumi, Y., Ichikawa, Y., Okuyama, H., Brisibe, E.A., Saksena, N.K. and Fujii, Y.R. 2004. HIV-1 nef suppression by virally encoded microRNA. *Retrovirology* 1: 44.

Pan, X., Zhang, B., Franciso, M.S. and Cobb, G.P. 2007. Characterizing viral microRNAs and its application on identifying new microRNAs in viruses. *J. Cell. Physiol.* 211: 10-18.

Panahi, B., Mohammadi, S.A. and Ebrahimie, E. 2013. Identification of miRNAs and their potential targets in halophyte plant *Thellungiella halophila. J. Biotechnol. Comp. Biol. Bionanotechnology* 94 (3) : 285-290.

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.

Pani, A. and Mahaptra, R.K. 2013. Computational identification of microRNAs and their targets in Catharanthus roseus expressed sequence tags. *Genomics Data* 1: 2–6.

Pantaleo, V. 2011. Plant RNA silencing in viral defence. *Adv. Exp. Med. Biol.* 722: 39–58.

Parizotto, E.A., Dunoyer, P., Rahm, N., Himber, C. and Voinnet, O. 2004. Invitro investigation of the transcription processing, endonucleolytic activity and functional relevance of spatial distribution of plant miRNA. *Genes Dev.* 18: 2237-42.

Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., and Poethig, R.S. 2005. Nuclear processing and export of microRNAs in Arabidopsis. *Proc. Natl. Acad. Sci.* 102: 3691–3696.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr. Biol.* 12: 1484–1495.

Pasquinelli, A.E., Reinhart, B.J., Slack, F.J., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnam, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E. and Ruvkun, G., 2000. Conservation of sequences and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408 (6808): 86-9.

Patanun, O., Lertpanyasampatha, M., Sojikul, P., Viboonjun, U. and Narangajavana, J. 2012. Computational Identification of MicroRNAs and Their Targets in Cassava (Manihot esculenta Crantz.). *Mol Biotechnol.* 53 (3): 257-69.

Patil, B.L. and Fauquet, C.M. 2009. Cassava mosaic geminiviruses: actual knowledge and perspectives. *Mol. Plant Pathol.* 10: 685-701.

Pérez-Quintero, A., Neme, R., Zapata, A. and Lopez, C. 2010. Plant microRNAs and their role in defense against viruses: a bioinformatic approach. *BMC Plant Biol.* 10: 138.

Pérez-Quintero, A., Quintero, A., Urrego, O., Vanegas, P. and Lopez, C. 2012. Bioinformatic identification of cassava miRNAs differentially expressed in response to infection by *Xanthomonas axonopodis pv. manihotis*. *BMC Plant Biol.* 12: 29.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F.A., van Dyk, L., Ho, C.K., Shuman, S. and Chien, M. 2005. Identification of microRNAs of the herpesvirus family. *Nat. Methods* 2: 269–276.

Pfeffer, S., Zavolan, M., Grasser, F.A., Chien, M., Russo, J.J., Ju, J.Y., John, B., Enright, A.J., Marks, D., Sander, C. and Tuschl, T. 2004. Identification of virus-encoded microRNAs. *Science* 304: 734–736.

Prochnik, S., Marri, P.R., Desany, B., Pablo D. Rabinowicz, P.D., Kodira, C. Mohiuddin, M., Rodriguez, F., Fauquet, C., Tohme, J., Harkins, T., Rokhsar, D.S. and Rounsley, S. 2012. The Cassava Genome: Current Progress, Future Directions. *Trop Plant Biol.* 5 (1): 88–94.

Raymond, C.K., Roberts, B.S., Garrett-Engele, P., Lim, L.P. and Johnson, J.M. 2005. Simple, quantitative primer-extension PCR assay for direct monitoring of microRNAs and short-interfering RNAs. *RNA* 1: 1737–1744.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougive, A.E., Horvitz, H.R. and Ruvkun, G. 2000. The 21 nt let-7 RNA regulates developmental timings in *C. elegans*. *Nature* 403 (6772): 901-6.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. and Bartel, D.P. 2002. MicroRNAs in plants. *Genes Dev.* 16 (13): 1616-26.

Samols, M.A., Hu, J., Skalsky, R.L. and Renne, R. 2005. Cloning and identification of a microRNA cluster within the latency-associated region of Kaposi's sarcoma-associated herpesvirus. *J. Virol.* 79: 9301–9305.

Schadt, E.E., Turner, S. and Kasarskis, A. 2010. A window into third generation sequencing. *Hum. Mol. Genet.* 19: 853.

Schauer, S.E., Jacobsen, S.E., Meinke, D.W. and Ray, A. 2002. Dicer like 1:blindmen and elephant in Arabidopsis development. *Trends Plant Sci.* 7: 487-491.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R. and Rajewsky, N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* 445 (7209): 58-63.

Sewer, A., Paul, N., Landgraf, P., Aravin, A., Pfeffer, S., Brownstein M.J., Tuschl, T., Nimwegen, E. and Zavolan, M. 2005. Identification of clustered microRNAs using an ab initio prediction method. *BMC Bioinformatics* 6: 267.

Sharbati-Tehrani, S., Kutz-Lohroff, B., Bergbauer, R., Scholven, J. and Enspanier, R. 2008. miR-Q: a novel quantitative RT-PCR approach for the expression profiling of small RNA molecules such as miRNAs in a complex sample. *BMC Mol. Biol.* 9: 34.

Simo´n-Mateo C. and Garcı´a J.A. 2006. MicroRNA-guided processing impairs Plum pox virus replication, but the virus readily evolves to escape this silencing mechanism. *J. Virol.* 80: 2429–2436.

Singh, J., Singh, C.P., Bhavani, A., Nagaraju, J. 2010. Discovering microRNAs from Bombyx mori nucleopolyhedrosis virus. *Virol.* 407: 120–128.

Sinha, S., Vasulu, T.S., De, R.K. 2009. Performance and Evaluation of MicroRNA Gene Identification Tools. *J. Proteomics Bioinform.* 2: 336-343.

Smith, T.F., Gaitatzes, C., Saxena, K. and Neer, E.J. 1999. The WD repeat: a common architecture for diverse functions. *Trends Biochem. Sci.* 24 (5): 181-5.

Song, J.J., Simth, S.K., Hannon, G.J. and Joshua-Tor, L. 2003. Crystal structure of argonaute and its implications for RISC slicer activity. *Science* 305: 1434-1437.

Song, L., Gao, S., Jiang, W., Chen, S., Liu, Y., Zhou, L. and Huang, W. 2011. Silencing suppressors: viral weapons for countering host cell defenses. *Protein Cell* 2: 273–281.

Srivastava, P.K., Moturu, T.R., Pandey, P., Baldwin, I.T. and Pandey, S.P. 2014. A comparison of performance of plant miRNA target prediction tools and characterization of features of genome wide target prediction. *BMC Genomics* 15: 348.

Stanley, J. 2004. Subviral DNAs associated with geminivirus disease complexes. *Vet. Microbiol.* 98: 121–129.

Stanley, J., Bisaro, D.M., Briddon, R.W., Brown, J.K., Fauquet, C.M., Harrison, B.D., Rybicki, E.P., and Stenger, D.C. 2005. Geminiviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U. and Ball, L.A. (eds), *Virus Taxonomy VIIIth Report of the ICTV.* Elsevier/Academic Press, London, pp. 301-326

Su, C., Yang, X., Gao, S., Tang, Y., Zhao. C. and Li, L. 2014. Identification and characterization of a subset microRNAs in wheat (*Triticum aestivum* L.). *Genomics* 103 (4): 298-307.

Sullivan, C.S. and Ganem, D. 2005. MicroRNAs and viral infection. *Mol.Cell* 20: 3–7.

Sullivan, C.S., Grundhoff, A.T., Tevethia, S.J., Pipas, M. and Ganem, D. 2005. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* 435: 682–686.

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.* Vol. 744, Springer Protocols, Heidelberg, pp. 175-186.

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

Sunkar, R., Li, Y.F. and Jagadeeswaran, G. 2012. Functions of microRNAs in plant stress responses. *Trends Plant Sci.* 17: 196–203.

Tang, G. 2005. siRNA and miRNA an insight into RISCs. *Trends Biochem. Sci.* 30: 106-114.

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

Thiebaut, F., Grativol, C., Carnavale-Bottino, M., Rojas, C.A., Tanurdzic, M., Farinelli, L., Martienssen, R.A., Hemerly, A.S. and Ferreira, P.C. G. 2012.

Computational identification and analysis of novel sugarcane microRNAs. *BMC Genomics* 13: 290.

Thompson, J.R. and Tepfer, M. 2010. Assessment of the benefits and risks for engineered virus resistance. *Adv. Virus Res.* 76: 33–56.

Tripathy, R. and Mishra, D. 2012. A computational approach of rice plant miRNA target prediction against tungro virus. *Procedia. Engineering* 38: 1357-1361.

Unver, T., Namuth-Covert, D.M. and Budak, H. 2009. Review of current methodological approaches for characterising microRNAs in plants. *Int. J. Plant Genomics.* 2009: 11.

Vaintharani, R., Chellapan, P., Pita, J.S. and Fauquet, C.M. 2004. Differential role of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. *J. Virol.* 78: 9487-98.

Vanderschuren, H., Alder, A., Zhang, P. and Gruissen, W. 2009. Dose dependent RNAi mediated gemiminivirus resistance in tropical root crop cassava. *Plant Mol. Biol.* 70 (3): 265-72.

van Ooijen, G., Mayr, G., Kaisem, M. M., Albrecht, M., Cornelissen, B. J. and Takken, F. L. 2008. Structure-function analysis of NB ARC domain of plant disease resistance proteins. *J. Exp. Bot.* 59 (6): 1383-97.

Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E.F. and Hellens, R.P. 2007. Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods* 3 (1): 12.

Varma, A. and Malathi, V.G. 2003. Emerging geminivirus problems: A serious threat to crop production. *Ann. Appl. Biol.* 142: 145–164.

Viswanathan, C., Anburaj, J. and Prabhu, G. 2014. Identification and validation of sugarcane streak mosaic virus and their targets in sugarcane. *Plant Cell Rep.* 33: 265-276.

Walters, D.R. 2003. Polyamines and plant diseases. *Phytochemistry* 64 (1): 97-107.

Wang, T., Chen, L., Zhao, M., Tian, Q. and Zhang, W.H. 2011. Identification of drought-responsive microRNAs in *Medicago truncatula* by genome-wide high-throughput sequencing. *BMC Genomics* 12: 367.

Wang, X., Zhang, J., Li, F., Gu, J., He, T., Zhang, X. and Li, Y. 2005. microRNA identification based on sequence and structure alignment. *Bioinformatics* 21(18): 3610-3614.

Wu, H.J., Ma, Y.K., Chen, T., Wang, M. and Wang, X.J. 2012. PsRobot: a web-based plant small RNA meta-analysis toolbox. *Nucleic Acids Res.* 40: W22–W28.

Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C., and Qi, Y. 2010. DNA methylation mediated by a microRNA pathway. *Mol. Cell* 38: 465–475.

Xie, F. and Zhang, B. 2010. Target-align: a tool for plant microRNA target identification. *Bioinformatics* 26 (23):3002–3003.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., and Carrington, J.C. 2005. Expression of Arabidopsis MIRNA genes. *Plant Physiol.* 138: 2145–2154.

Xue, C., Li, F., He, T., Liu, G-P. Li, Y. and Zhang, X. 2005. Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. BMC *Bioinformatics* 6: 310.

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

Yant, L., Mathieu, J., Dinh, T.T., Ott, F., Lanz, C., Wollmann, H., Chen, X. and Schmid, M. 2010. Orchestration of the floral transition and floral development in *Arabidopsis* by the bifunctional transcription factor APETALA2. *Plant Cell* 22: 2156–2170.

Yoon, S. and Micheli. G.D. 2006. Computational Identification of MicroRNAs and Their Targets. *Birth Defects Res.* 78: 118-128.

Yu, L., Yu, X., Shen, R. and He, Y. 2005. HYL1 gene maintains venation and polarity of leaves. *Planta* 221: 231–242.

Zeng, Y. and Yang, T. 2002. RNA isolation from highly viscous samples rich in polyphenols and polysaccharides. Plant Mol. Biol. Rep. 20 (4): 417.

Zhang, B.H., Pan, X.P. and Stellwag, E.J. 2008. Identification of soybean microRNAs and their targets. *Planta* 229: 161-182.

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.

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., Cobb, G.P. and Anderson, T.A. 2006c. Plant microRNA: A small regulatory molecule with big impact. *Dev. Biol.* 289: 3-16.

Zhang, B.H., Pan, X.P., Cox, S.B., Cobb, G.P. and Anderson, T.A. 2006d. 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. 2006e. Computational identification of microRNAs and their targets. *Comput. Biol. Chem.* 30: 395-407.

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

Zhang, J., Xu, Y., Huan, Q. and Chong, K. 2009. Deep sequencing of *Brachypodium* small RNAs at the global genome level identifies microRNAs involved in cold stress response. *BMC Genomics* 10: 449.

Zhang, W., Gao, S., Zhou, X., Chellappan, P., Chen, Z., Zhou, X., Zhang, X., Fromuth, N., Coutino, G., Coffey, M. and Jin, H. 2011. Bacteria responsive microRNAs regulate plant innate immunity by modulating plant hormone networks. *Plant Mol. Biol.* 75: 93–105.

Zhang, Y. 2005. miRU: an automated plant miRNA target prediction server. *Nucleic Acids Res.* 33: W701–W704.

Zhao, T., Li, G.L., Mi, S.J., Li, S., Hannon, G.J., Wang, X.J. and Qi, Y.J. 2007. A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev.* 21: 1190–1203.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31 (13): 3406-3415.

http://mfold.bioinfo.rpi.edu

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

http://wotan.wistar.upenn.edu/BayesSVMmiRNAfind/

http://www.bioinf.seu.edu.cn/miRNA/

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

http://www.jalview.org

http://www.microrna.org/

http://www.mirbase.org/

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

http://www.phytozome.net/cassava

APPENDICES

APPENDIX I

RNA Extraction Buffer (pH 8.0)

CTAB	2 %
PVP (Mol. Wt. 25000)	2 %
Tris HCl (pH 8.0)	100 *mM*
NaCl	2 *M*
Spermidine trihydrochloride	0.05 %
β – mercaptoethanol	2 %

APPENDIX II Cassava miRNAs identified by homolog search

SL.NO.	mes-miR	MATURE SEQUENCE	SCAFFOLD	ML	ARM	PL	(G+C)%	(A+U)%	A/U	MFE	AMFE	MFEI	FRAME
1	156a	UGACAGAAGAGAGUGAGCAC	10493	20	5'	84	52.38	47.62	0.74	-51.6	-61.43	1.17	-
2	156b	UGACAGAAGAGAGUGAGCAC	9822	20	5'	84	50.00	50.00	1.10	-53.1	-63.21	1.26	-
3	156c	UGACAGAAGAGAGUGAGCAC	7859	20	5'	84	47.62	52.38	0.83	-47.5	-56.55	1.19	-
4	156d	UGACAGAAGAGAGUGAGCAC	7318	20	5'	86	45.35	54.65	0.62	-52.9	-61.51	1.36	-
5	156e	UGACAGAAGAGAGUGAGCAC	3049	20	5'	90	42.22	57.78	0.79	-47.8	-53.11	1.26	+
6	156f	UGACAGAAGAGAGUGAGCAC	847	20	5'	80	41.25	58.75	0.74	-39.2	-49.00	1.19	+
7	156g	UGACAGAAGAGAGUGAGCAC	206	20	5'	86	47.67	52.33	0.80	-47.7	-55.47	1.16	+
8	156h	UUGACAGAAGAUAGAGAGCAC	12711	21	5'	85	41.20	58.80	0.85	-44.4	-52.24	1.3	-
9	156i	UUGACAGAAGAUAGAGAGCAC	8935	21	5'	87	40.20	59.80	0.93	-47.6	-54.71	1.4	+
10	156j	UUGACAGAAGAUAGAGAGCAC	6871	21	5'	84	39.30	60.70	0.89	-44.3	-52.74	1.3	-
11	156k	UGACAGAAGAGAGAGAGCACA	2421	21	5'	88	50.00	50.00	0.76	-42.9	-48.75	0.975	+
12	159a	UUUGGAUUGAAGGGAGCUCUA	3525	21	3'	183	44.81	55.19	0.68	-87.9	-48.03	1.07	-
13	159b	UUUGGAUUGAAGGGAGCUCUA	6631	21	3'	218	41.30	58.70	0.75	-90.7	-41.61	1.01	-
14	159c	AUUGGAGUGAAGGGAGCUCUG	3581	21	3'	183	51.37	48.63	0.85	-85.6	-46.78	0.91	-
15	159d	AUUGGAGUGAAGGGAGCUCUG	2658	21	3'	215	47.91	52.09	0.75	-93.3	-43.40	0.91	-
16	160a	UGCCUGGCUCCCUGUAUGCCA	10183	21	5'	86	59.30	40.70	0.94	-43.4	-50.47	0.85	-
17	160b	UGCCUGGCUCCCUGUAUGCCA	2954	21	5'	89	60.67	39.33	0.94	-55.7	-62.58	1.03	+
18	160c	UGCCUGGCUCCCUGUAUGCCG	8404	21	5'	94	60.64	39.36	0.76	-51.3	-54.57	0.89	-
19	160d	UGCCUGGCUCCCUGUAUGCCA	271	21	5'	86	55.81	44.19	0.90	-50.4	-58.60	1.05	-
20	160e	UGCCUGGCUCCCUGAAUGCCAUC	4163	23	5'	88	52.27	47.73	1.00	-55.6	-63.18	1.21	-

SL.NO.	mes-miR	MATURE SEQUENCE	SCAFFOLD	ML	ARM	PL	(G+C)%	(A+U)%	A/U	MFE	AMFE	MFEI	FRAME
21	160f	UGCCUGGCUCCCUGAAUGCCAUC	2255	23	5'	84	52.38	47.62	0.90	-45.1	-53.69	1.02	+
22	160g	UGCCUGGCUCCCUGUAUGCCAUC	2960	23	5'	86	59.30	40.70	0.94	-53.1	-61.74	1.04	-
23	160h	UGCCUGGCUCCCUGUAUGCCAUU	12004	23	5'	86	59.98	43.02	0.85	-50.4	-58.60	0.977	+
24	162	UCGAUAAACCUCUGCAUCCAG	7762	21	3'	92	45.65	54.35	1.00	-36.4	-39.57	0.87	-
25	164a	UGGAGAAGCAGGGCACGUGCA	11668	21	5'	89	55.10	44.90	0.74	-55.9	-62.81	1.14	+
26	164b	UGGAGAAGCAGGGCACGUGCA	8265	21	5'	94	51.06	48.94	0.70	-48.2	-51.28	1.004	+
27	164c	UGGAGAAGCAGGGCACGUGCA	4457	21	5'	111	55.86	44.14	0.75	-55.8	-50.27	0.89	+
28	164d	UGGAGAAGCAGGGCACAUGCU	6700	21	5'	84	46.43	53.57	0.96	-49.3	-58.69	1.26	+
29	166a	UCGGACCAGGCUUCAUUCCCC	11998	21	3'	92	47.83	52.17	0.78	-45.6	-49.57	1.04	+
30	166b	UCGGACCAGGCUUCAUUCCCC	10563	21	3'	95	49.47	50.53	0.85	-48.3	-50.84	1.03	-
31	166c	UCGGACCAGGCUUCAUUCCCC	7330	21	3'	77	45.45	54.55	0.56	-41.1	-53.38	1.17	-
32	166e	UCGGACCAGGCUUCAUUCCCC	1551	21	3'	88	50.00	50.00	1.00	-44.2	-50.23	1.00	+
33	166f	UCGGACCAGGCUUCAUUCCCC	977	21	3'	133	39.10	60.90	0.65	-50.7	-38.12	0.97	+
34	166g	UCGGACCAGGCUUCAUUCCCC	872	21	3'	125	42.40	57.60	0.71	-53.0	-42.4	1.00	-
35	166h	UCGGACCAGGCUUCAUUCCCGU	5979	21	3'	145	38.62	61.38	0.62	-63.3	-43.66	1.13	+
36	166j	UCGGACCAGGCUUCAUUCCUC	2895	21	3'	97	44.33	55.67	0.50	-42.0	-43.3	0.98	-
37	167b	UGAAGCUGCCAGCAUGAUCUA	11998	21	5'	80	37.50	62.50	1.00	-33.4	-41.75	1.10	-
38	167c	UGAAGCUGCCAGCAUGAUCUA	11998	21	5'	108	46.30	53.70	1.00	-44.4	-41.11	0.89	-
39	167d	UGAAGCUGCCAGCAUGAUCUGA	9347	22	5'	88	47.73	52.27	1.09	-40.9	-46.48	0.97	-
40	167e	UGAAGCUGCCAGCAUGAUCUGA	3802	22	5'	126	43.65	56.35	0.73	-55.1	-43.73	1.00	-
41	167f	UGAAGCUGCCAGCAUGAUCUGA	3614	22	5'	126	42.86	57.14	0.89	-59.6	-47.3	1.10	-
42	167g	UGAAGCUGCCAGCAUGAUCUU	10563	21	5'	68	47.06	52.94	0.89	-39.0	-57.35	1.22	+

SL.NO.	mes-miR	MATURE SEQUENCE	SCAFFOLD	ML	ARM	PL	(G+C)%	(A+U)%	A/U	MFE	AMFE	MFEI	FRAME
43	167h	UGAAGCUGCCAGCAUGAUCUU	1551	21	5'	68	44.12	55.88	0.81	-32.4	-47.65	1.08	-
44	168a	UCGCUUGGUGCAGGUCGGGAA	2477	21	5'	173	48.55	51.45	0.71	-81.2	-46.94	0.969	+
45	169a	CAGCCAAGGAUGACUUGCCGG	12448	21	5'	90	48.90	51.10	0.77	-39.4	-43.78	0.89	-
46	169c	CAGCCAAGGAUGACUUGCCGG	11341	21	5'	97	44.30	55.70	0.74	-46.9	-48.35	1.09	+
47	169d	CAGCCAAGGAUGACUUGCCGG	3874	21	5'	97	45.36	54.64	0.71	-47.2	-48.66	1.07	+
48	169e	CAGCCAAGGAUGACUUGCCGG	3049	21	5'	97	47.42	52.58	0.70	-44.5	-45.88	0.97	-
49	169f	UAGCCAAGGAUGACUUGCCGG	4233	21	5'	103	40.78	59.22	0.69	-47.9	-46.5	1.14	-
50	169g	CAGCCAAGGAUGACUUGCCGA	6582	21	5'	83	48.20	51.80	1.53	-40.1	-48.31	1.00	-
51	169h	UGAGCCAAGGAUGACUUGCCG	3264	21	5'	89	46.07	53.93	0.66	-46.3	-52.02	1.12	+
52	169i	GAGCCAAGAAUGACUUGCCGG	12768	21	5'	82	47.56	52.44	0.95	-33.4	-40.73	0.86	-
53	169j	GAGCCAAGAAUGACUUGCCGG	12768	21	5'	85	45.88	54.12	1.56	-37.0	-43.53	0.95	-
54	169k	GAGCCAAGAAUGACUUGCCGG	12448	21	5'	85	41.18	58.82	1.78	-31.7	-37.29	0.91	-
55	169l	CAGCCAAGAAUGACUUGCCGG	11232	21	5'	85	50.59	49.41	0.68	-39.9	-46.94	0.93	-
56	169m	CAGCCAAGAAUGACUUGCCGG	11232	21	5'	85	51.76	48.24	0.64	-42.1	-49.53	0.96	+
57	169n	GAGCCAAGAAUGACUUGCCGA	2017	21	5'	83	55.42	44.58	1.18	-42.7	-51.45	0.93	-
58	169p	UAGCCAAGGAUGACUUGCCUG	5005	21	5'	114	49.12	50.88	0.66	-57.6	-50.53	1.03	+
59	169q	UAGCCAAGGAUGACUUGCCUG	5005	21	5'	108	43.52	56.48	0.65	-52.0	-48.15	1.11	-
60	169r	UAGCCAAGGAUGACUUGCCCG	5005	21	5'	93	43.01	56.99	0.56	-35.6	-38.28	0.89	-
61	169s	UAGCCAAGGAUGACUUGCCUG	5005	21	5'	109	49.54	50.46	0.77	-54.1	-49.63	1.00	+
62	169t	UAGCCAAGGAUGACUUGCCCG	5005	21	5'	102	45.09	54.91	0.65	-50.6	-49.61	1.10	+
63	169u	UAGCCAAGGAUGACUUGCCUG	5005	21	5'	120	50.00	50.00	0.62	-60.8	-50.67	1.01	+
64	169v	UAGCCAAGGAUGACUUGCCCG	5005	21	5'	102	45.10	54.90	0.65	-50.6	-49.61	1.10	+

SL.NO.	mes-miR	MATURE SEQUENCE	SCAFFOLD	ML	ARM	PL	(G+C)%	(A+U)%	A/U	MFE	AMFE	MFEI	FRAME
65	169w	UAGCCAAGGAUGACUUGCCUG	5005	21	5'	108	44.44	55.56	0.67	-51.4	-47.59	1.07	-
66	169x	UAGCCAAGGAUGACUUGCCCG	3800	21	5'	95	43.16	56.84	0.50	-38.8	-40.84	0.95	-
67	169y	UAGCCAAGGAUGACUUGCCUG	3800	21	5'	112	42.73	57.27	0.75	-43.1	-38.48	0.92	-
68	169z	UAGCCAAGGAUGACUUGCCCG	5005	21	5'	93	43.01	56.99	0.51	-37.4	-40.22	0.94	-
69	169aa	UAGCCAAGGAUGACUUGCCUG	363	21	5'	181	43.65	56.35	0.73	-81.6	-45.08	1.03	-
70	169ab	UAGCCAAGGAUGACUUGCCUA	9876	21	5'	115	43.48	56.52	1.10	-43	-37.39	0.86	+
71	169ac	UAGCCAAGGAUGACUUGCCUA	1701	21	5'	126	46.03	53.97	0.89	-51.2	-40.63	0.88	+
72	171a	GGAUUGAGCCGCGUCAAUAUC	5214	21	3'	81	38.27	61.73	0.79	-40.5	-50	1.30	-
73	171b	UUGAGCCGUGCCAAUAUCACG	6598	21	3'	90	41.10	58.90	0.71	-55.1	-61.22	1.50	-
74	171c	UUGAGCCGUGCCAAUAUCACG	2717	21	3'	84	44.05	55.95	0.81	-53.6	-63.81	1.45	-
75	171d	AUGAGCCGUGCCAAUAUCACG	7330	21	3'	106	44.34	55.66	0.79	-43.7	-41.23	0.93	-
76	171e	AGAUUGAGCCGCGCCAAUAUC	11970	21	3'	99	41.41	58.59	1.07	-43.1	-43.54	1.05	+
77	171f	AGAUUGAGCCGCGCCAAUAUC	5760	21	3'	97	41.24	58.76	1.19	-43.4	-44.74	1.08	+
78	171g	UGAUUGAGCCGUGCCAAUAUC	9876	21	3'	75	44.00	56.00	0.83	-36.7	-48.93	1.11	+
79	171h	UGAUUGAGCCGUGCCAAUAUC	9683	21	3'	113	38.94	61.06	0.64	-46.7	-41.33	1.06	-
80	171i	UAUUGGCCUGGUUCACUCAGA	7520	21	3'	140	40.70	59.30	0.54	-60.8	-43.43	1.07	+
81	171j	UGAUUGAGCCGUGCCAAUAUC	6512	21	3'	132	44.70	55.30	0.52	-49.9	-37.8	0.85	+
82	171k	UGAUUGAGCCGUGCCAAUAUC	1701	21	3'	90	42.22	57.78	0.86	-42.2	-46.89	1.11	+
83	172b	AGAAUCUUGAUGAUGCUGCAU	5815	21	3'	119	47.06	52.94	0.91	-53.2	-44.71	0.95	+
84	172d*	AGAAUCUUGAUGAUGCUGCAU	224	21	3'	131	40.46	59.54	0.47	-58.7	-44.81	1.11	+
85	172e	GGAAUCUUGAUGAUGCUGCAG	6705	21	3'	167	47.30	52.70	0.69	-77.0	-46.11	0.97	+
86	172f	GGAAUCUUGAUGAUGCUGCAG	3293	21	3'	137	46.72	53.28	0.70	-66.2	-48.32	1.03	+

SL.NO.	mes-miR	MATURE SEQUENCE	SCAFFOLD	ML	ARM	PL	(G+C)%	(A+U)%	A/U	MFE	AMFE	MFEI	FRAME
87	319a	UUGGACUGAAGGGAGCUCCCU	10493	21	3'	177	53.11	46.89	1.08	-93.5	-52.82	0.99	-
88	319b	UUGGACUGAAGGGAGCUCCCU	10122	21	3'	176	44.32	55.68	1.13	-76.9	-43.69	0.98	+
89	319c	UUGGACUGAAGGGAGCUCCCU	6557	21	3'	201	41.29	58.71	0.82	-92.3	-45.92	1.11	+
90	319d	UUGGACUGAAGGGAGCUCCCU	3429	21	3'	174	43.68	56.32	0.92	-75.1	-43.16	0.99	+
91	319e	UUGGACUGAAGGGAGCUCCCU	3264	21	3'	174	48.28	51.72	0.8	-87.8	-50.46	1.04	-
92	319f	UUGGACUGAAGGGAGCUCCUU	6446	21	3'	178	48.88	51.12	0.98	-93.5	-52.53	1.07	-
93	319g	UUGGACUGAAGGGAGCUCCUU	3454	21	3'	188	46.81	53.19	0.85	-84.6	-45.00	0.96	-
94	319h	CUUGGACUGAAGGGAGCUCCU	2877	21	3'	180	46.11	53.89	0.73	-79.0	-43.89	0.95	-
95	390	CGCUAUCCAUCCUGAGUUUC	5375	21	3'	137	40.88	59.12	0.56	-60.6	-44.23	1.08	-
96	393a	UCCAAAGGGAUCGCAUUGAUCC	12439	22	5'	103	38.83	61.17	0.62	-39.7	-38.54	0.99	-
97	393b	UCCAAAGGGAUCGCAUUGAUCC	8265	22	5'	110	44.55	55.45	0.61	-47.6	-43.30	0.97	+
98	393c	UCCAAAGGGAUCGCAUUGAUCU	11425	22	5'	116	40.52	59.48	0.60	-57.1	-49.22	1.21	+
99	393d	UCCAAAGGGAUCGCAUUGAUCU	341	22	5'	99	37.37	62.63	0.72	-40.2	-40.61	1.09	+
100	394a	UUGGCAUUCUGUCCACCUCC	9683	20	5'	151	45.70	54.30	0.64	-69.4	-45.96	1.01	+
101	394b	UUGGCAUUCUGUCCACCUCC	2943	20	5'	98	43.88	56.12	0.45	-43.6	-44.49	1.01	+
102	394c	UUGGCAUUCUGUCCACCUCCAU	6914	22	5'	93	41.94	58.06	0.86	-54.0	-58.06	1.40	-
103	395a	CUGAAGUGUUUGGGGGAACUC	11495	21	3'	73	46.58	53.42	0.63	-40.3	-55.21	1.18	-
104	395b	CUGAAGUGUUUGGGGGAACUC	11495	21	3'	83	51.81	48.19	0.54	-43.5	-52.41	1.01	+
105	395c	CUGAAGUGUUUGGGGGAACUC	7690	21	3'	80	45.00	55.00	0.69	-37.2	-46.50	1.03	+
106	395d	CUGAAGUGUUUGGGGGAACUC	1497	21	3'	103	55.34	44.66	0.64	-60.8	-59.03	1.07	+
107	395e	CUGAAGGGUUUGGAGGAACUC	80	21	3'	120	41.17	55.83	0.86	-44.4	-37.00	0.90	-
108	396a	UUCCACAGCUUUCUUGAACUG	5760	21	5'	120	45.00	55.00	0.53	-52.1	-43.42	0.96	-

SL.NO.	mes-miR	MATURE SEQUENCE	SCAFFOLD	ML	ARM	PL	(G+C)%	(A+U)%	A/U	MFE	AMFE	MFEI	FRAME
109	396b	UUCCACAGCUUUCUUGAACUG	5297	21	5'	130	40.00	60.00	0.50	-53.2	-40.92	1.02	+
110	396c	UUCCACAGCUUUCUUGAACUU	12498	21	5'	96	40.62	59.38	0.54	-50.9	-53.02	1.30	-
111	396d	UUCCACAGCUUUCUUGAACUU	5760	21	5'	96	38.54	61.46	0.74	-43.2	-45.00	1.20	+
112	396e	UUCCACAGCUUUCUUGAACUU	5297	21	5'	76	43.42	56.58	0.72	-43.0	-56.58	1.30	-
113	396f	UUCCACAGCUUUCUUGAACUU	4151	21	5'	123	40.65	59.35	0.43	-50.6	-41.14	1.01	+
114	397	UUUGAGUGCAGCGUUGAUGA	5338	21	5'	81	34.57	65.43	0.66	-27.4	-33.83	0.98	-
115	399a	UGCCAAAGGAGAAUUGCCCUG	4003	21	3'	121	43.80	56.20	1.06	-50.2	-41.49	0.95	+
116	399b	UGCCAAAGGAGAUUUGCCCGG	10535	21	3'	91	45.05	54.95	0.79	-41.9	-46.04	1.02	-
117	399c	UGCCAAAGGAGAUUUGCCCGG	10535	21	3'	91	48.35	51.65	0.81	-41.7	-45.82	0.95	+
118	399d	UGCCAAAGGAGAUUUGCCCGG	10535	21	3'	91	42.86	57.14	0.93	-36.0	-39.56	0.92	+
119	399e	UGCCAAAGGAGAUUUGCUCGG	10535	21	3'	92	44.60	55.40	0.76	-37.5	-40.76	0.914	+
120	399f	UGCCAAAGGAGAGUUGCCCUG	11425	21	3'	100	50.00	50.00	0.79	-48.4	-48.40	0.968	+
121	399g	UGCCAAAGGAGAUUUGCCCGG	4003	21	3'	97	49.48	50.52	0.88	-41.60	-42.89	0.87	-
122	403b	UUAGAUUCACGCACAAACUCG	80	21	3'	94	36.17	63.83	0.88	-33.4	-35.53	0.98	-
123	408	AUGCACUGCCUCUUCCCUGGC	80	21	3'	134	50.00	50.00	1.03	-65.7	-49.03	0.98	-
124	477a	CUCUCCCUCAAGGGCUUCUG	6716	21	5'	79	43.04	56.96	0.88	-29.1	-36.84	0.855	+
125	477b	CUCUCCCUCAAGGGCUUCUG	3175	20	5'	114	51.75	48.25	1.04	-51.2	-44.91	0.87	+
126	477c	CUCUCCCUCAAGGGCUUCUG	3175	21	5'	80	52.50	47.50	0.90	-41.0	-51.25	0.97	+
127	477d	CUCUCCCUCAAGGGCUUCUC	2226	20	5'	162	47.53	52.47	1.43	-75.2	-46.42	0.98	+
128	477e	CUCUCCCUCAAGGGCUUCUG	2226	20	5'	105	45.71	54.29	0.97	-62.4	-59.43	1.30	+
129	477f	AUCUCCCUCAAAGGCUUCCA	4043	20	5'	81	39.51	60.49	0.96	-40.6	-50.12	1.27	-
130	477g	AUCUCCCUCAAAGGCUUCCA	1701	20	5'	89	47.19	52.81	1.04	-49.1	-55.17	1.17	+

SL.NO.	mes-miR	MATURE SEQUENCE	SCAFFOLD	ML	ARM	PL	(G+C)%	(A+U)%	A/U	MFE	AMFE	MFEI	FRAME
131	477h	ACUCUCCCUCAAGGGCUUCAG	8265	21	5'	143	53.15	46.85	0.86	-71.0	-49.65	0.93	-
132	477i	ACUCUCCCUCAAGGGCUUCCG	6716	21	5'	86	56.98	43.02	0.95	-56.7	-65.93	1.15	+
133	482	UCUUCCCUACUCCACCCAUUCC	2264	22	3'	96	45.83	54.17	0.58	-58.1	-60.52	1.32	+
134	530a	UGCAUUUGCACCUGCACCUU	8316	20	5'	140	40.71	59.29	0.73	-50.6	-36.14	0.89	+
135	530b	UGCAUUUGCACCUGCACCUU	4182	20	5'	134	37.31	62.69	0.79	-50.5	-37.69	1.01	+
136	535a	UGACAACGAGAGAGAGCACGU	5782	21	5'	80	46.25	53.75	0.72	-39.3	-49.13	1.06	-
137	535b	UGACAACGAGAGAGAGCACGG	1190	21	5'	80	47.50	52.50	1.00	-40.6	-50.75	1.07	-
138	827	UUAGAUGACCAUCAACAAACA	4024	21	3'	90	26.67	73.33	0.89	-27.9	-31.00	1.16	+
139	828a	UCUUGCUCAAAUGAGUAUUCCA	6377	22	5'	137	38.69	61.31	0.53	-53.4	-38.98	1.01	-
140	1446	UUCUGAACUCUCUCCCUCAU	12262	20	5'	100	40.00	60.00	1.31	-38.7	-38.70	0.97	+
141	2111a	UAAUCUGCAUCCUGAGGUUUA	11204	21	5'	104	47.12	52.88	0.62	-63.9	-61.44	1.30	+
142	2111b	UAAUCUGCAUCCUGAGGUUUA	6244	21	5'	89	42.70	57.30	0.70	-51.1	-57.42	1.34	-
143	2275	UUUGGUUUCCUCCAAUAUCUUA	3614	22	3'	93	35.48	64.52	0.67	-33.0	-35.48	1.00	+
144	2950	UUCCAUCUCUUGCACACUGGA	3235	21	5'	104	44.23	55.77	0.71	-56.4	-54.23	1.23	-
145	new1	UCGGACCAGGCUUCAUUCCCC	2906	21	3'	95	46.32	53.68	0.70	-49.7	-52.32	1.13	+
146	new2	AGAUAUUGGUGCGGUUCAAUC	2717	21	5'	84	44.05	55.95	0.81	-53.6	-63.81	1.44	-
147	new3	AGAAUCUUGAUGAUGCUGCAU	4182	21	3'	144	44.44	55.56	0.82	-57.5	-39.90	0.89	-
148	new4	AGCUCAGGAGGGAUAGCGCCA	7117	21	5'	96	45.83	54.17	0.53	-41.1	-42.81	0.93	-
149	new5	AAGCUCAAGAAAGCUGUGGGA	5760	21	3'	94	38.30	61.70	0.71	-42.1	-44.79	1.17	+
150	new6	UCAUUGAGUGCAGCGUUGAUG	9000	21	5'	82	43.90	56.10	0.70	-37.8	-46.10	1.05	-
151	new7	CGCCAAAGGAGAGUUGCCCUU	7088	21	3'	111	42.34	57.66	0.60	-51.9	-46.76	1.10	+
152	new8	UGCCAAAGAAGAUUUGCCCCG	7108	21	3'	93	41.94	58.06	1.16	-42.9	-46.13	1.09	+

APPENDIX III

Cassava miRNA targets in cassava

Sl.no.	mes-miR	TARGETS (cassava 4.1_)	TARGETED PROTEIN
1.	156	006419,005991,005580,008947,006701,029311,009507, 009657,006155,18710	Squamosa Promoter Binding (SPB) domain
		004204	Amino acid permease
		019305,019305	60s acidic ribosomal protein
		019305	DnaJ domain
		025018	Transferase
		029763, 026570	Ankyrin repeat
		007289	Tyrosine kinase
		021390	Cellulose synthase
		013690	αβ hydrolase fold
		015061	U box domain
		019983	Ubiquitin related modifier 1
		008516	ACT domain
		021248, 030139, 018907	unknown
2.	159	004606,005846,024052, 021857, 022856	myB like DNA binding domain
		002814	αβ hydrolase
		009884	Aminotransferase class 4
		000478	Transport protein trs120
		033959	KIP1 like protein
		025907	SAM carboxyl methyl transferase
		006246	PPR repeat + CBS domain
		010517	DNA methylase
		023704	Nozzle TF
		011374	POT family
		003936, 008576	Tyrosine kinase
		011374	POT family
		025653	QLQ, WRC domain
		013636, 005962,21030, 026951	Unknown
3.	160	002980, 002668, 002684, 002960, 004122	Auxin response factor & B3 DNA binding domain
		034267	Glycosyl hydrolase family
		034244	Dirigent like protein
		000225	Sec7 domain
4.	162	010833, 010837	ER-Golgi intermediate compartment protein
		000068	Helicase related
		015268	Glutathione-s-transferase

		001291	Leucine zipper hand containing transmembrane protein
		003913	Unkown
5.	164	029889,026876,031247,020925,026590,010528,025545, 010869	No apical meristem protein (NAM)
		010996	NAD epimerase
		013375,013858, 013970, 015924, 014705	CCT motif
		024744	GDSL like lipase
		006541	Amine oxidoreductase
		002230	Inorganic proton pyrophosphatase
6.	166	001674,001672,001600,001956,001653,001649,001618, 001673	MEKHLA/ START/Homeobox domain
		001619	HEX T.F.
		029393	bZIP T.F.
		004052	Tyrosine kinase
		022415	NAM
		001098, 030978, 021054	unknown
7.	167	015606, 016753	PAP2 superfamily
		004576	Helix loop helix DNA binding domain
		020625	unknown
8.	168	002527	1,3-b-glucan synthase component
9.	169	007505,010819,012637, 014256,014279, 011576, 011590, 011636,011722, 011264	CCAAT binding TF subunit B
		021412	bZIP TF
		007595	Ferrous ion transport protein
		003550	Leo-1 like protein
		025850	Aldehyde dehydrogenase
		026967	Sec 1 family
		022992, 023878	unknown
10.	171	002034,034057	GRAS family TF
		004839	Acyltransferase
		003281,003282	ANTH domain
		008470	Aminotransferase class 1 & 2
		021738	unknown
11.	172	006005,006069,005961,025425,007151,006539,005852	AP2 domain
		002563, 023500, 002568	Protein Tyrosine kinase
		000161	Transcription elongation factor b
		025182	Ion transport protein
		002596	ThiF family
12.	319	006246	PPR repeat + CBS domain

		024052, 004606, 005846	Myb like DNA binding domain
		000738, 000737	Metallopeptidase family
		027328	Glycosyl transferase
		001661	Metallo-b-lactamase
		006620, 006618, 011082, 004517	TCP family TF
		016634	Kelch repeat domain
		004344	DnaJ domain
		027831	unknown
13.	390	NO RESULT	
14.	393	004520,004514,004294,032685	F box Domain
		013191	Ceramidase
		004395, 005103	Zn finger protein
		006180	Glycosyl hydrolase family 17
15.	394	007038,007255,008769	F box Domain
		005653	Cytochrome P450
		003485	SAM methyl transferase
		000867, 001415	Sucrose phosphate synthase
16.	395	026478	Multicopper oxidase (laccase)
		024866	Fusaric acid resistance protein
		021316	α-l- fucosidase
		030105	ABC transporter
		005020	AP2 domain
		022829	ATP sulfurylase
		009267	Decarboxylase
		012341, 013377, 012567	Myb like DNA binding domain
		025754	Glutaredoxin
		028322	Oxygenase
17.	396	027844	QLQ, WRC
		020707	RNA Polymerase
		008793	Galactosyltransferase
		002615	TPR repeat containing protein
		004733, 004736	Phosphoglycerate mutase
		033816, 010296, 029196	unknown
18.	397	006971,004761,031010,023475,004438, 012403, 004370, 004785, 004756 004799, 004450,005501, 021113	Multicopper oxidase (laccase)
		000743	Serine/threonine kinase
		034197	unknown
19.	399	007038,007255,008769	F box domain
		005653	Cytochrome p450

		003485	SAM methyl transferase
		000867, 001415	Sucrose phosphate synthase
20.	403	001070	RNA Pol 1 TF UAF
		025174	ATPase
		017019, 018686	Unknown
21.	477	032725	GRAS family TF
		028810	Replication factor C
		002858	Tetracopeptide repeat + thioredoxin
		008283, 008297	Elongation factor 1 gamma
		031366	PPR repeat
		014896, 016691	Der 1 like protein
		029176	F box domain
		025141	Na/H exchanger family protein
		008447, 010926	PLAC 8 family
		012562	Calponin homology domain
		011291, 012636	Fascicilin domain
		021439, 030289	unknown
22.	482	027893, 029503, 032178	NB ARC domain
		022477	Glycosyl hydrolase family 31
		032573	WD40
23.	530	000754	Plus 3 domain
		005909	Protein kinase domain
		004619	Starch synthase
		004795, 008242, 019308	unknown
24.	535	NO RESULTS	
25.	827	002769	Major facilitator superfamily & SPX domain
26.	828	030774, 027214, 024299, 008985	myB like DNA binding domain (TF)
		013915, 013958, 016775, 034344, 024583, 025854, 033576, 013293	unknown
27.	1446	010714	Alliinase
		007488	AP2 domain
		013704	Short chain dehydrogenase
		000452,000919, 026946, 027306	unknown
28.	2111	024575, 007360	Kelch motif
		013605	Kinesin motor domain
		001850	Terpene synthase
		000504	ATPase
		011014	unknown
29.	2275	023110	IQ calmodulin binding motif proteins
		033399	Cytochrome P450

		033843	FHA domain
		001848, 002255	Phospholipase
		019721, 019717, 019720	Stress responsive αβ barrel domain
		001779, 001808, 002854	Protein tyrosine kinase
		008283, 008297	Glutathione-s- transferase
		030801	Hydrolase
		002874	GTPase
		009078, 002852, 034349, 000217	Unknown
30.	2950	008692, 008586, 008184, 008156	F box domain
		033242, 028745	Copper transport protein
		000297, 010729, 025116	Unknown

EVALUATION OF PREDICTION TOOLS AND COMPUTATIONAL ANALYSIS OF microRNAs IN CASSAVA (*Manihot esculenta* Crantz.)

ANJANA HARIDAS

Abstract of the
thesis submitted in partial fulfilment of the requirement
for the degree of

Master of Science (Integrated) in Biotechnology

Faculty of Agriculture
Kerala Agricultural University, Thrissur

M.Sc. (INTEGRATED) BIOTECHNOLOGY COURSE

DEPARTMENT OF PLANT BIOTECHNOLOGY
COLLEGE OF AGRICULTURE
VELLAYANI, THIRUVANANTHAPURAM 695522
KERALA, INDIA

2014

ABSTRACT

The study entitled "Evaluation of prediction tools and computational analysis of microRNAs in cassava (*Manihot esculenta* Crantz.)" was conducted at the CTCRI, Sreekariyam, Thiruvananthapuram during August 2013 to June 2014. The objectives of the study included comparison and evaluation of miRNA prediction software, computational prediction and annotation of miRNA in cassava and understanding the role of miRNA-mRNA interaction in cassava in biotic stress responses (Cassava Mosaic Virus). Experiments were also conducted to validate the presence of the predicted cassava miRNA.

The plant miRNA target prediction tools: psRNATarget and miRanda were compared for their performance. Analysis was performed to identify the optimal maximum expectation value for psRNATarget. psRNATarget with an optimum maximum expectation value of 3 was found to be a better plant miRNA target prediction tool.

Homology based method was used to identify the conserved potential cassava miRNAs. The targets for the predicted miRNA were predicted using the web tool psRNATarget and these were functionally annotated. A total of 152 miRNAs belonging to 30 miRNA families were identified having multiple targets in cassava transcripts. Majority of the microRNAs were about 21 nt in length and found in the 5' arm of stem loop hairpin secondary structure. miRNAs had a high MFE and MFEI values. VMir Analyzer was used in the prediction of 3 *Indian cassava mosaic virus* miRNAs and their targets in cassava transcripts include genes involved in catalysis, regulation and stress response.

Cassava miRNA targets in cassava mosaic virus were identified using miRanda target prediction algorithm. 14 cassava miRNA families targeted Cassava Mosaic Virus with nearly perfect complementarity. Two of the cassava miRNAs having target in the viral genome was validated for their presence in healthy and CMD infected leaf sample. Primers were designed and qRT-PCR reaction was performed and their presence validated in both the samples.