COMPUTATIONAL PREDICTION OF miRNAs IN BANANA (*Musa* spp.) AND EVALUATION OF THEIR ROLE IN VIRUS INFECTION

by

KOKILA SAJEEV ANURAG MATHEW

(2016-11-078)

THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN AGRICULTURE

Faculty of Agriculture

Kerala Agricultural University

DEPARTMENT OF PLANT BIOTECHNOLOGY

COLLEGE OF AGRICULTURE

VELLAYANI, THIRUVANANTHAPURAM-695 522

KERALA, INDIA

2018

DECLARATION

I, hereby declare that the thesis entitled **"COMPUTATIONAL PREDICTION OF miRNAs IN BANANA (*Musa* spp.) AND EVALUATION OF THEIR ROLE IN VIRUS INFECTION"** 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, fellowship or other similar title, of any other University or Society.

Place: Vellayani,

Date: 21/11/18

K.S. Anurag Mathew

Kokila Sajeev Anurag Mathew

(2016-11-078)

CERTIFICATE

Certified that this thesis, entitled **"COMPUTATIONAL PREDICTION OF miRNAs IN BANANA (*Musa* spp.) AND EVALUATION OF THEIR ROLE IN VIRUS INFECTION"** is a record of research work done independently by **Mr. Kokila Sajeev Anurag Mathew (2016-11-078)** under my guidance and supervision and that it has not previously formed the basis for the award of any degree, fellowship or associateship to him.

Place: Vellayani,
Date: 21/11/18

Dr. K. B. Soni
(Major advisor, Advisory Committee)
Professor
Dept. of Plant Biotechnology
College of Agriculture
Kerala Agricultural University
Vellayani, Thiruvananthapuram -- 695 522

CERTIFICATE

We, the undersigned members of the advisory committee of Mr. Kokila Sajeev Anurag Mathew, a candidate for the degree of **Master of Science in Agriculture** with major in Plant Biotechnology, agree that the thesis entitled **"COMPUTATIONAL PREDICTION OF miRNAs IN BANANA (*Musa* spp.) AND EVALUATION OF THEIR ROLE IN VIRUS INFECTION."** may be submitted by Mr. Kokila Sajeev Anurag Mathew, in partial fulfilment of the requirement for the degree.

Dr. K. B. Soni
(Chairman, Advisory Committee)
Professor
Department of Plant Biotechnology
College of Agriculture, Vellayani

Dr. Swapna Alex
(Member, Advisory Committee)
Professor and Head
Department of Plant Biotechnology
College of Agriculture, Vellayani

Dr. Joy M.
(Member, Advisory Committee)
Associate Professor and Head
Department of Plant Pathology
College of Agriculture, Vellayani

Dr. J. Sreekumar
(Member, Advisory Committee)
Principal Scientist
ICAR-Central Tuber Crops Research Institute
Sreekariyam, Thiruvananthapuram

EXTERNAL EXAMINER
(Name and Address)

Dr. J. Ramalingam
Professor & Head
Dept. of Biotechnology
Ag. college & Research Institute
Madurai

ACKNOWLEDGEMENT

First and foremost, I bow my head before and would like to thank my loving creator for making me a curious being who loves to explore his creation and for giving me the opportunity to travel my life journey in the field of agriculture and without him I can do nothing.

It was my good fortune to have ***Dr. K. B. Soni.*** *Professor, Department of Plant Biotechnology as Chairman of my advisory committee. I feel the deep sense of gratitude and thank her for her meticulous guidance, valuable advice, constructive criticisms and unfailing patience throughout my post-graduate programme without which I would not have completed this research work.*

I am particularly grateful to ***Dr. Swapna Alex,*** *Professor and Head, Department of Plant Biotechnology for her thought provoking suggestions, affectionate guidance, immense interest, valuable counselling and sincere help during all stages of my study.*

My sincere gratitude to ***Dr. Joy M., Associate*** *Professor, Department of Plant Pathology and* ***Dr. J. Sreekumar,*** *Principal Scientist, ICAR-Central Tuber Crops Research Institute for their valuable guidance, critical evaluation, helpful suggestions, wholehearted effort in my research and interpretation of the results and advice rendered throughout the degree programme.*

I wish to express my genuine thanks to ***Dr. Kiran A. G.****, Assistant Professor, Department of Plant Biotechnology and* ***Dr. Radhika,*** *Assistant Professor, Department of Plant Pathology for their crucial suggestions and help during the investigation of the work.*

I am ineffably thankful to my dear batch mates ***Elizabeth, Bijula and Nagamani*** *with a deep regret on the impossibility of repayment of their blissful presence, moral support, discussions, assistance, love and affection showered on me at each and every stage of my work.*

I wish to place on record the help rendered and moral support to me by my loving seniors ***Sachin sir, Smitha cheichi, Pritam cheichi, Lekshmi cheichi,***

Harshita akka, Afna cheichi, Deepa akka and Jancy cheichi. *I am also thankful for the help and support of my loving juniors* ***Athira, Arathy, Monisha, Sowndarya and Nazreen.***

I wish to place my gratefulness to ***Ambu vijayan chetta,*** *Research Associate, Statistics and Bioinformatics Lab, CTCRI for his advice, timely support and co-operation rendered throughout the course of this research endeavor.*

I gratefully remember my department non-teaching staff ***Ajitha cheichi and Ajila cheichi,*** *whose co-operation and affection helped me during my research work.*

Special credits to my best friends ***Kishore sir and Madhu Gunda*** *who encouraged me during my falls and for being good companions during my PG life No words can explain their unending love, care and moral support.*

I wish to place my gratitude to my dear friends ***Hampaiah, Reshma mohan, Ravi Kiran, Ivy Mary Rajan, Fathima, Reshma, Aswathy, Athul*** *and my lovely seniors* ***Sainath sir, Vinay sir, Gayathri cheichi and Manasa akka*** *and juniors* ***Sree pavan, Manikanta, Vinod and Santhosh*** *for their love, help and moral support during my study.*

No words to explain my gratitude to my dear friends ***Jaswanth Naik, Ramanjineyulu, Balaji Reddy, Vinay Yadiki, Venkatesh*** *for their blissful presence, friendship and help rendered during every stage of my work.*

Words cannot express my indebtedness to my family. I express my deep sense of gratitude and affection to my parents ***Pradeep Kumar, Prema Latha*** *and sister* ***Anusha Mathew*** *for their unconditional love and immense support throughout my life.*

Finally my heartfelt gratitude to you! Yes to YOU, for your patience and time invested to read this acknowledgement.

Kokila Sajeev Anurag Mathew

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

LIST OF APPENDICES

LIST OF ABBREVATIONS

%	Percentage
3^{l} - UTR	3^{l} – Un-translated region
A	Adenine
A_{260}	Absorbance at 260 nm wavelength
A_{280}	Absorbance at 280 nm wavelength
AMFE	Adjusted minimal folding free energy
BBrMV	Banana bract mosaic virus
BBTV	Banana bunchy top virus
BLAST	Basic local search alignment algorithm
bp	Base pair
BSV	Banana streak disease
C	Cytosine
cDNA	Complementary DNA
C_T	Threshold cycle
CTAB	Cetyl trimethyl ammonium bromide
DCL 1	Dicer like enzyme 1
DNA	De oxy ribonucleic acid
DNase	Deoxy ribonuclease
dNTPs	Deoxy nucleotide tri phosphates
DH	Doubled haploid

EBV	Epstein-Barr virus
ELISA	Enzyme linked immune sorbent assay
EST	Expressed sequence tags
F	Forward primer
G	Guanine
g	Gram
GO	Gene ontology
GSS	Genome survey sequences
ha	Hectares
h	Hours
LAF	Laminar air flow chamber
LiCl	Lithium chloride
M	Molar
Mb	Mega base pair
MFE	Minimal folding free energy
MFEI	Minimal folding free energy index
min	Minutes
miRNA	Micro RNA
mRNA	Messenger RNA
MT	Million Tonnes
NCBI	National center for biotechnology information
nt	Nucleotide
nm	Nano meters
NTC	No template control

^{O} C	Degree Celsius
OD	Optical density
Pre-miRNA	Precursor miRNA
psRNATarget	Plant small RNA target analysis server
PTGS	Post-transcriptional gene silencing
qRT-PCR	Quantitative reverse transcriptase-polymerase chain reaction
R	Reverse primer
RDV	Rice dwarf virus
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	Revolution per minute
RT-PCR	Reverse Transcription-Polymerase Chain Reaction
S	Stem-loop primer
siRNA	Small Interfering RNA
sp.	Species
spp.	Species
sRNA	Small RNA
T	Tonnes
TBE	Tris-borate EDTA buffer
T_m	Melting temperature
TMV	Tobacco mosaic virus
U	Uracil

V	**Volt**
v/v	**Volume/volume**
V1	**Version 1**
vi-miRNA	**Viral miRNA**
µg	**Microgram**
µl	**Microlitre**
µM	**Micromolar**

INTRODUCTION

1. INTRODUCTION

Banana is one of the most important tropical crops in many developing countries. It is the fourth most important food crop after rice, wheat and maize in the world. It is a herbaceous plant which belongs to the genus *Musa*. Banana is grown in many tropical and subtropical regions of India. Tamilnadu, Maharashtra, Gujarat, Andhra Pradesh and Karnataka are the leading banana growing states. In India, banana is grown in 8.58 lakh ha with annual production of 29.1 MT during 2016-17 (Department of Agriculture, cooperation & farmers welfare). In Kerala, banana is the third most important crop and is grown in 57.158 ha during 2016-17.

Production and productivity of banana is adversely affected by various biotic stresses, viruses being one of them. Most important viruses affecting banana are Banana bract mosaic virus (*BBrMV),* Banana bunchy top virus (*BBTV)* and Banana streak virus (*BSV*) (Kumar *et al.,* 2015). Triploid nature of most of the cultivated bananas made the traditional methods difficult to develop resistant varieties. Several non-conventional methods were applied for the development of virus resistance. Genetic engineering tools tried are mostly based on coat protein, movement protein and resistance genes mediated resistance.

The discovery of RNA interference (RNAi) mechanism by Small interfering RNAs or MicroRNAs has led to the development of small RNA based technologies for imparting resistance towards biotic and abiotic stresses in plants and animals. In the recent years, role of miRNAs has been established in various stress responses of plants.

miRNAs are non-coding RNAs of 18-22 nt in length, that play a key role in regulation of gene expression. Plant miRNAs target various transcription factors involved in plant growth and development. They act as post-transcriptional gene regulators, which help in plant growth and development under stress conditions (Sunkar *et al.,* 2012). A better understanding of role of miRNA in post-

transcriptional gene silencing will help to develop plant's resistance against various stresses.

Various experimental methods have been used for the identification of miRNAs in plants including the cloning based techniques. With the advancement of computational approaches, research on miRNA has advanced tremendously over the last decade.

Several computational tools have been developed which complemented the experimental approaches for the identification and validation of novel miRNAs from high throughput platforms such as NGS (Huang *et al.*, 2015). Most of them uses the characteristic features of the miRNAs like their length, high sequence conservation among species, and structural features like hairpin and minimal folding free energy for distinguishing from other RNAs (Li *et al.*, 2010). These computational approaches have been found efficient and cost-effective. miRNAs have been identified in many crops using computational prediction tools. In banana, many miRNAs have been predicted computationally from ESTs, but miRNAs predicted from the banana genome using computational tools are scarce. In this context, the present study was undertaken with an objective to computationally predict miRNAs and their targets in banana genome and to validate and analyze their expression during Banana bract mosaic virus (*BBrMV*) infection.

REVIEW OF LITERATURE

2. REVIEW OF LITERATURE

2.1 BANANA

2.1.1 Morphology, Origin and Genome Distribution of Bananas

Banana (*Musa* spp.) is one of the world's top ten food crops and most important horticultural crop belonging to order *Zingiberales* and family *Musaceae*. It is vegetatively propagated by using suckers. Banana cultivars include *Musa acuminata* (A genome) and *Musa balbisiana* (B genome). All cultivated varieties of banana are parthenocarpic, produced from intra and inter crosses of two species, *Musa acuminata* and *Musa balbisiana* (Ortiz, 2013). Banana includes triploid cultivars like desert banana (AAA), cooking banana and plantain (AAB or ABB), diploid cultivars (AA or AB) and synthetic tetraploids (AAAA, AAAB, AABB and ABBB) (Heslop-Harrison and Schwarzacher, 2007). In banana, genome structure, protein-coding genes, non-coding sequences, regulatory elements, transposable elements and repetitive DNA were identified by whole genome sequencing. Banana genome usually consists of 11 chromosomes, with genome size of 523 Mb, whose assembly consist of 24,425 contigs and 7,513 scaffolds, which represent 90% of DH-Pahang genome (Heslop-Harrison and Schwarzacher, 2007). A total of 36,542 protein-coding genes are present in banana (D'Hont *et al.,* 2012). Banana consists of around 20 varieties, which include Dwarf Cavendish, Robusta, Monthan, Poovan, Nendran, Nyali, Red banana, Basarai, Safed velchi, Rasthali, Karpuvalli, Ardhapuri, Grandnaine and Karthali *etc.* Bananas are very much prone to viruses, which leads to yield reduction and barriers to germplasm exchange between species (Kumar *et al.,* 2015). Molecular studies of banana will help in identifying genes and their levels of expression in growth and development stages of banana and their response to different environmental stresses. Small RNAs like miRNAs and siRNAs play an important role in modulating their gene expression. Thus, genetic information of banana will help in crop improvement of banana plants with desirable agronomic characters.

2.2 TYPES OF VIRUSES IN BANANA

There are more than 2000 viruses and those affecting plants include 21 families and 8 unknown genera (Hull, 2013). Many of the plant viruses have been associated with non-cultivated plants.

Majority of the viruses were transmitted to host plants by vectors. The interactions between viruses and vector vary in duration and specificity. Based on various vector species and viruses, different patterns of vector–virus interactions were found by scientists. Brault *et al.* (2010) reported three modes of transmission. In non–persistent mode, virus was acquired within seconds and retained for only a few minutes by the vector. In semi–persistent mode, virus was acquired within minutes to hours and retained for several hours by the vector. In persistent mode, vector requires minutes to hours for acquisition and it was retained for a long period of time. Plant viruses encode structural proteins on the surface of the host that are essential for their transmission. Viruses bind to specific sites on vectors and are retained until they are transmitted to their plant hosts (Whitefield *et al.,* 2015).

Most important viruses affecting banana are Banana bract mosaic virus (*BBrMV*), Banana bunchy top virus (*BBTV*) and Banana streak virus (*BSV*) (Kumar *et al.,* 2015). In India and Philippines, *BBrMV* is more important.

2.2.1 Banana Bract Mosaic Disease (*BBrMV*)

Banana bract mosaic disease is caused by Banana bract mosaic virus (*BBrMV*). It was first reported in Philippines, 1979 by Magnaye and Espino (1990). It belongs to the genus *Potyvirus* and family *Potyviridae* (Rodoni *et al.,* 1997). Symptoms include spindle shaped reddish streaks on bracts and pseudostem. Affected plants usually contain very short peduncle and leaves will fold like travellers palm (Balakrishnan *et al.,* 1996). This viral disease causes major losses to banana cultivars in four states of southern India. *BBrMV* was first identified in Nendran variety in Kerala in 1966. It is also called as kokkan disease. This viral disease was also reported in north east region of India by Indian Council of

Agricultural Research–National research centre for banana, Tiruchirappalli (Selvarajan and Jeyabaskaran, 2006). The virus is transmitted non-persistently by three aphid species *viz.*, *Pentalonia nigronervosa, Rhopalosiphum maidis, Aphis gossypii (*Magnaye and Espino, 1990). Aphids were considered as most important vectors so far, as they are transmitting 30% of all plant virus species (Brault *et al.,* 2010). *BBrMV* genome consist of single stranded RNA of about 9711 nt long (Bateson and Dale, 1995). Many molecular techniques like ELISA and RT-PCR are used for *BBrMV* detection. Due to lack of host resistance in *Musa* spp., virus free planting material and phytosanitory measures were used for virus control (Kumar *et al.*, 2015). Recently, abaca cvs. Tinawagan Pula and Tangongon having resistance to *BBrMV* were developed by in-vitro mutagenesis by gamma irradiation (Dizon *et al.,* 2012).

2.2.2 Banana Bunchy Top Disease (*BBTV*)

Banana bunchy top disease is one of the most destructing virus disease of banana. *BBTV* was found in almost 36 countries. This virus disease was first noticed in Cavendish banana (AAA) in Fiji, 1889 (Magee, 1927). It causes discrete dark green flecks and streaks on leaf sheath, midrib, leaf veins and petioles (Nelson, 2004). It is spread through suckers, corms, tissue culture plants and transmitted by banana aphid, *P. nigronervosa* (Anhalt and Almeida, 2008). In infected banana plant, aphids were observed on leaf at the top and near the base of plants (Robson *et al.,* 2006). Hu *et al.* (1996) reported that aphids transmit *BBTV* virus, after 4 h of acquisition period and 15 min of inoculation period. *BBTV* was first detected by using enzyme linked immunosorbent assay (ELISA) technique using monoclonal and polyclonal antibodies (Thomas and Dietzgen, 1991). Infection was detected only after 12–25 days based on the stage of infection and genotype of plant (Hooks *et al.*, 2008). Techniques like Real-Time PCR have been developed for quantitative detection of viral particles in both plants and aphid tissues (Bressan and Watanabe, 2011). As banana plants resistant to *BBTV* are not available, plants with B genome (ABB and AAB) which are tolerant or express symptoms slowly compared to A genome were selected (Ngatat *et al.*, 2013). Other practices of *BBTV* management

were removal of infected plants, production of virus free plants through meristem culture and use of disease resistant cultivars (Bouhida and Lockhart, 1990).

2.2.3 Banana Streak Disease (*BSV*)

Banana streak disease is caused by Banana streak virus (*BSV*) which causes chlorotic streaks on the plant. Lockhart and Jones (2000) reported the disease in nikey valley on ivory coast in 1958. It belongs to the genus *Badnavirus* and family *Caulimoviridae*. Genome includes double stranded-closed circular DNA of about 7.2-7.8 kb. Transmission is usually done by infected plant material like suckers and vectors like mealy bugs. Banana streak virus (*BSV*) was reported over 43 countries (Daniells *et al.,* 2001). *BSV* was detected by using a polyvalent polyclonal antiserum against *BSV* and *Sugarcane bacilliform virus* spp. Main method used to control *BSV* infection was to produce healthy banana plants and as mealy bugs are slowly moving, disease can be controlled by removing infected plants.

2.3 STRATEGIES FOR DEVELOPING VIRUS RESISTANCE

Plants encounter various biotic stresses, which affects their growth, development and yield. Among these, plant viruses usually impose serious threats to wide range of crops in modern agriculture and it was estimated that economic loss caused by viral pathogens ranks second compared to those caused by other pathogens (Simon-Mateo and Garcia, 2011). Due to their devastating threat to crop production, plant viruses were extensively studied since the first virus, tobacco mosaic virus was discovered. Currently, more than 6000 viruses were identified according to ninth report of international committee on taxonomy of viruses, of which 1300 were plant viruses (King *et al.*, 2012).

Several strategies were applied for development of virus resistance like conventional breeding using natural source of resistance *i.e* by breeding resistant cultivars and non-conventional methods like genetic engineering techniques based on coat protein mediated resistance, movement protein mediated resistance and Resistance genes (R genes).

The discovery of RNA interference (RNAi) mechanism or gene silencing in plants during the past decade has led to the development of small RNA based technologies for imparting resistance against biotic stress in plants. Gene silencing is defined as regulation of gene expression in a cell to prevent the expression of certain gene. It occurs either during transcription or translation. Gene silencing is often considered as gene knockdown. It is also known as quelling in fungi or post-transcriptional gene silencing (PTGS) in plants. It was discovered by Andrew Fire and Craig Mello, who won Nobel prize for their discovery in 2006 in physiology or medicine for their work on RNA interference in nematode *Caenorhabditis elegans* (Andrew *et al.*, 1998). RNA silencing mainly occurs by SmallRNAs (sRNAs). sRNAs include siRNAs and miRNAs which represent subset of sRNAs. RNA mediated virus resistance is also called as PTGS. Napoli *et al.* (1990) first reported PTGS in petunia hybrid transgenically expressing the chalcone synthase gene.

Plants have evolved three basic RNA silencing pathways, which include miRNA mediated gene silencing, siRNA directed RNA degradation pathway and siRNA directed DNA methylation. Among them, miRNA mediated gene silencing triggers with the entry of dsRNA into the cell, which is made into small fragments by Dicer like enzyme 1 (DCL1). These dsRNA fragments are either siRNAs or miRNAs. miRNAs integrate into a multi protein subunit complex called RNA induced silencing complex (RISC) which contain Argonaute proteins, as essential component of RISC. Argonaute proteins contain a single stranded RNA binding site (Yan *et al.*, 2003) to which miRNAs will bind. One strand acts as guide strand, which binds to RISC and the other acts as passenger strand, which is degraded by the helicase enzyme. The guide strand directs the silencing of target mRNA. Thus, miRNA mediated gene silencing approach helps in developing stress resistant crops for better crop improvement. There are several advantages of using miRNAs in gene silencing instead of siRNAs: fewer off-target effects, highly RNA promoter compatible, environmental biosafety, stable in vivo (Lu *et al.*, 2008).

During the course of evolution, plants have developed diverse mechanisms against viruses and viruses have evolved multiple mechanisms to counter act RNA silencing. Hence miRNA mediated interactions between host and viruses include:

1. Endogenous host miRNAs targets specific host genes
2. Host miRNAs targets viral genes
3. Virus encoded miRNAs targets specific host genes
4. Virus encoded miRNAs targets viral genes to mediate immune evasion or maintenance of latency.

In plants, usually there are two roles of miRNAs in antiviral defense (i) Direct mode through targeting viral miRNAs (ii) Indirect mode through targeting the biogenesis of miRNAs responsible for viral response.

Viruses produce non-coding RNAs (ncRNAs) called viral miRNAs (vi-miRNAs), to regulate cellular mechanisms and viral replication. First vi-miRNA was reported from Epstein-Barr virus (EBV) in 2004 (Pfeffer *et al.*, 2004). Like eukaryotic miRNAs, vi-miRNAs were also processed by DROSHA and Dicer enzymes. To date, more than 200 vi-miRNAs were identified. VIRmiRNA database consist of information about 9133 vi-miRNAs, which include 1308 validated vi-miRNAs and 7283 target genes encoded by 44 viruses (Qureshi *et al.,* 2014). Some of the studies suggested the existence of plant vi-miRNAs, but their function need to be elucidated. In Hibiscus chlorotic ringspot virus (HCRSV) genome, five putative vi-miRNAs were predicted using vi-miRNAs prediction database (Gao *et al.*, 2012). In Sugarcane streak mosaic virus (SCSMV) genome, 93 miRNAs from 25 families and 454 novel miRNAs were predicted using deep sequencing technology (Viswanathan *et al.*, 2014). As plant miRNAs are conserved among species or generations, no conservation has been observed in case of vi-miRNAs (Cullen, 2004). vi-miRNAs that mimic host miRNAs are referred as ''Analogs'', which result in gene silencing (Kincaid and Sullivan, 2012). For example, vi-miRNA encoded by *Tobacco mosaic virus* (TMV) which contains a stretch of

phytoene desaturase (PDS) silenced the transcription of PDS mRNA (Carr *et al.*, 1992).

Functions of vi-miRNAs include host mRNA/miRNA degradation, inhibition of maturation of host miRNAs, induction of transcription of host miRNAs to enhance viral replication and regulation of viral life cycles.

Pathogenic viruses, as a result of their long term interaction with plants, develop viral suppressors of RNA silencing (VSRs) as a counter defense strategy. For example, 2b of cucumber mosaic virus (CMV) (Feng *et al.*, 2013), P1/HC-pro of the Poty virus, p19 of Tomato bushy stunt virus (Lakatos *et al.*, 2006), ORFs AC-4 and AC-2 encoded by Gemini virus.

2.3.1 Applications of miRNA in Plant Antivirus Defense

Usually plants do not possess an antibody-based immune system analogous to that of animals (Waterhouse *et al.*, 2001). Hence, miRNA mediated silencing was newly discovered strategy to suppress plant viruses. miRNAs are used as antiviral agents for repressing the function of viral suppressors in the process of gene silencing. miRNAs were used in several ways in anti-viral defense: analysing the function of viral suppressor in gene silencing process, developing new miRNA mediated gene therapy, transforming plant physiological properties to increase their anti-virus capacity, developing loss-of-function transformed plants (Lu *et al.*, 2008).

2.3.2 Artificial miRNA (amiRNA) Technology

Artificial miRNA technology is used for silencing of genes in plants. They were designed for targeting the protein coding mRNA, by using pre-miRNA as structural backbone to replace the original 18-22 nt long miRNA sequence with a region complementarity to the target viral genome. These were used to create transgenic plants for improving crop tolerance to abiotic and biotic stresses (Schwab *et al.*, 2006).

Sun *et al.* (2014) constructed stem loop RNA or hairpin RNA (hpRNA), precursor of miRNA, from a segment of rice dwarf virus (RDV). The stem loop RNA was inserted into a vector. Plants, which are transformed, confer resistance to the virus. miRNAs target engineered plant viruses containing target sequences and cause gene silencing.

Simon-Mateo and Garcia (2011) found that *Plum pox virus,* member of Potyvirus family was engineered to have *Arabidopsis* miRNA target sequences. As a result, engineered virus targets were affected by tobacco miRNAs and confer resistance to *Plum pox virus*.

In order to control infestation of *Helicoverpa armigera* (cotton bollworm), which infest several crops like cotton, tomato, chickpea, pigeon pea etc., an artificial miRNA (amiRNA) was constructed by Yogindran and Rajam (2016), which replaced the original 21 nt sequence in insect *let-7a* pre-miRNA. The pre-miRNA was cloned by inserting into a bacterial vector. Larvae when fed on bacteria, amiRNA was expressed and caused reduction in expression of target gene and leads to mortality.

amiRNAs were constructed in *Nicotiana benthamiana* by Wagaba *et al.* (2016) to control cassava brown streak disease (CBSD) caused by Ugandan cassava brown streak virus and cassava brown streak virus. amiRNAs were transformed through vector. When amiRNAs were expressed, they cause degradation or inhibition of gene expression, which control virus infection in tobacco.

2.4 MicroRNA

MicroRNAs (miRNA) are non-coding regulatory RNAs with short sequences of about 18-22 nt in length, having a hairpin or stem loop structure, which play a central role in plant development and environmental stress responses (Xia *et al.*, 2018).

miRNAs are gene regulators which are originated internally, encoded by a MIR gene (Bartel, 2004). In plants, MIR gene is transcribed by RNA polymerase II (Pol II) enzyme to form pri-miRNA (Figure 1). pri-miRNA is cleaved by a member of the RNase – III class enzyme, Dicer-like enzyme 1 (DCL1) to generate precursor miRNA (pre-miRNA), which contains miRNA/miRNA* sequences with a hairpin or stem loop structure (Bartel, 2004; Kurihara and Watanabe, 2004). miRNA/miRNA* duplex is sliced out from the pre-miRNA by using an enzyme, DCL1 and transferred from nucleus to cytoplasm by HASTY, the plant orthologue of Exportin 5 protein (Park *et al.*, 2005). In cytoplasm, miRNAs are unwound into single strand mature miRNAs by helicase enzyme (Bartel, 2004). Later, mature miRNAs cause gene regulation by entering into RNA-induced silencing complex (RISC), a ribonucleoprotein complex (Hammond *et al.*, 2000; Martinez *et al.*, 2002; Schwarz *et al.*, 2002; Bartel, 2004).

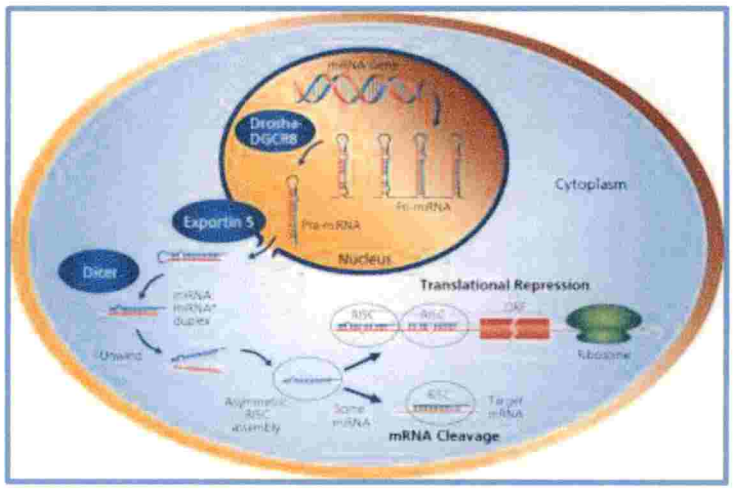


Fig. 1: miRNA Biogenesis pathway

Usually, biogenesis of animal miRNAs differs from plant miRNAs. In Animal miRNA biogenesis, pri-miRNA is cleaved to pre-miRNA by Drosha RNase III endonuclease whereas in plants by Dicer like Enzyme 1 (DCL1). pre-miRNAs are transported from nucleus to cytoplasm by Exportin 5 and then form miRNA:miRNA duplex in animals.

2.4.1 Characteristics of miRNAs

miRNAs, usually 18-22 nt in length, have a well predicted stem loop or hairpin structure having low minimal folding free energy (MFE). miRNAs are evolutionarily conserved *i.e.,* from worm to human. In plants, only mature miRNAs are conserved. miRNAs control gene expression by regulating mRNA stability and translation (Pillai *et al.,* 2006). They are encoded by both hosts and viruses *i.e.,* vi-miRNAs.

2.4.2 Seed Region

Seed region (Figure 2) is the main characteristic feature of miRNA. The specificity of target mRNA recognition is mostly determined by the seed region (2-8 nt) present at the 5^{1} end of the miRNA. Most of the functional interactions between RNA induced silencing complex (RISC) and mRNA require full complementarity between mRNA and seed region of miRNA (Bartel, 2009).

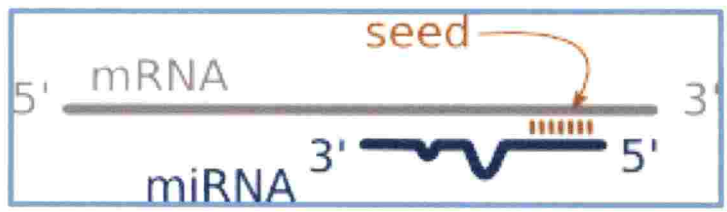


Fig. 2: Seed Region

2.4.3 Functions of Plant miRNAs

Although miRNAs are relatively small, they have versatile functions either by overexpression or under expression. They are involved in various development processes like leaf morphogenesis and polarity (Emery *et al.*, 2003); floral differentiation (Chen, 2004); root initiation (Mallory *et al.*, 2004); vascular development (Floyd and Bowman, 2004); transition of plant growth from vegetative phase to reproductive phase (Achard *et al.*, 2004); signal transduction (Sorin *et al.,* 2005); environmental stress responses; plant disease control (Ding, 2000); regulate miRNA and siRNA biogenesis (Vaucheret *et al.,* 2004).

2.5 PREDICTION OF miRNAS

As many species or organisms of phylogenetic importance were non-model organisms without complete genome sequences, miRNA discovery was limited. In order to overcome this limitation, several computational tools have been developed based on reference genome for the prediction of stem loop structures (Wen *et al.*, 2014). Sequence conservation of miRNAs simplified the prediction of miRNAs in non-model plants.

There are two approaches used to predict miRNAs: a) genetic approach by direct cloning and sequencing and b) computational approach by comparative genomics. Computational approach includes (i) Homology based search (ii) Bioinformatic or computational prediction. Deep sequencing and direct cloning were predominantly used apart from the homology-based analysis methods *i.e* expressed sequence tags (EST) and genomes survey sequences (GSS) (Zhang *et al.*, 2006c).

2.5.1 Computational Approach

2.5.1.1 *Homology Based Search*

By using ESTs and GSS, homology based search or comparative genomics search helps in identification and characterization of miRNAs and their target genes in plants (Dehury *et al.*, 2013). Identification of mature miRNAs by using ESTs had advantages over other method (Zhang *et al.*, 2008). Search algorithm uses the genome hits matching between known miRNA and query sequences. Homology based search were classified as GSS based search or EST based search. BLAST analysis was done with plant unique miRNAs, downloaded from miRBase against EST or GSS sequences, downloaded from EST databases (dbEST) or GENBANK. BLAST Result were the sequences with 0-3 mismatches against known miRNAs. These sequences were blasted (BLASTX) against protein database to remove non protein-coding sequences. Resulted sequences were considered as miRNA candidates only if they fit the following criteria: (1) at least 18 nt length between

the query and mature miRNAs (2) allowed to have 0–3 nt mismatches in sequence with all previously known plant mature miRNAs (Pani and Mahapatra, 2013). Secondary or hairpin structures of RNA sequences were predicted by using MFold 3.2, minimal folding free energy (MFE) of potential miRNAs were identified.

Homology or sequence similarity search was done with a set of 349 previously known miRNAs of Arabidopsis, Rice and Poplar, downloaded from NCBI against 14, 20,579 GSS sequences of tobacco (NCBI) and found 259 potential miRNAs in tobacco genome (Frazier *et al.*, 2010).

Four miRNAs were identified from ESTs of *Tea* (*Camellia sinensis)* by Prabhu and Mandal (2010). For this, 299 EST sequences of tea were downloaded from dbEST database and blasted against all known plant mature miRNA sequences by using default parameters.

In *Linum usitatissimum*, miRNAs were isolated from leaves using NuceloSpin miRNA kit. miRNA sequences obtained were blasted against ESTs of flax obtained from NCBI. Twenty sequences were considered as miRNAs belonging to 13 different families (Neutelings *et al.,* 2012).

Dehury *et al.* (2013) reported 8 conserved miRNA sequences representing 3 different families viz. miR168, miR2911, and miR156 by blasting 23,406 ESTs in sweet potato with 3,227 non reductant mature miRNA sequences of kingdom *Viridiplantae*.

Pani and Mahapatra (2013) identified two potential miRNAs in *Catharanthus roseus*. Computer software BLAST-2.2.14 was used to predict secondary structure or hairpin loop structure of miRNAs by using 328 miRNAs from *Arabidopsis thaliana* obtained from miRNA registry database against EST database of dbEST.

Akter *et al.* (2014) used EST sequences, obtained from GenBank of NCBI and 989 predicted mature miRNAs of *Arabidopsis thaliana, Oryza sativa* from miRBase to find out miRNAs in *Coffea arabica*. They reported one potential

miRNA by using mpiBLAST-1.6.0 software (Darling *et al.*, 2003) which was similar to BLAST-2.2.22 of NCBI.

In garlic, 6 potential miRNAs belonging to 6 different miRNA families were predicted by Panda *et al.* (2014) through computational means, where 3,227 mature miRNAs of plant kingdom *Viridiplantae* were blasted for homology against 21,637 EST sequences of garlic.

Singh and Sharma (2014) reported 9 miRNA sequences belonging to 7 different families in *Ocimum basilicum*. A set of 23,260 ESTs obtained from NCBI were used for prediction of miRNAs by using a bioinformatic tool C-mii version 1.11 (Numnark *et al.*, 2012) from 3436 contigs.

Huang *et al.* (2015) predicted 23 miRNAs in five cyprinidae fishes by BLAST analysis of ESTs and GSS of cyprinidae fishes in NCBI against previously known mature miRNAs in miRBase. Out of 23 miRNAs, 21 miRNAs were identified as novel.

In *Phaseolus vulgaris,* all the *Viridiplantae* miRNAs from miRBase were used as query in BLAST search against ESTs and GSS of *P. vulgaris* by Nithin *et al.* (2015). After removing all the protein coding sequences from BLAST results, 310 sequences were considered as potential pre-miRNAs. Out of 310, 208 were novel miRNAs, which belong to 118 different miRNA families.

Bi *et al.* (2015) reported miRNAs in Cavendish banana by using homology based search. Small RNA sequences of banana were mapped against annotated miRNAs of miRBase. A total of 125 potential miRNAs belonging to 39 different families were found.

Chai *et al.* (2015) found 32 miRNAs in banana genome, belonged to 13 miRNA families by using ESTs and GSS. BLAST analysis was done between known plant miRNA sequences obtained from miRBase (Kozomara and Griffith-

Jones, 2014) and 46,111 banana ESTs and 31,544 GSS obtained from NCBI GenBank (Benson *et al.*, 2013).

In Stevia, 5,548 ESTs and 29,874 nt sequences were downloaded from NCBI and searched for pre-miRNAs by doing BLAST analysis against stevia sequences. The result indicated 12 sequences as potential pre-miRNAs (Mandhan and Singh, 2015).

Homology based search was done in *Cannabis sativa* to identify miRNAs by Das *et al.* (2015). 18 conserved miRNAs belonging to 9 different families were found by blasting previously known 7,385 miRNAs of various plants in miRBase and local nucleotide sequences from NCBI.

22 miRNAs were found in *Humulus lupulus* belonging to 17 miRNA families by Mishra *et al.* (2015). EST–based homology search was done using mpiBLAST-1.6.0 algorithm (Darling *et al.*, 2003) between 25,692 ESTs available in NCBI database and known plant miRNA sequences of *Arabidopsis*, Brassica, Glycine, Saccharum, Sorghum, Vitis, Oryza, Solanum, Triticum, Chlamydomonas and other plant species obtained from miRBase.

Hasan *et al.* (2016) reported two potential miRNAs in *Cannabis sativa* by using an in-silico EST-based homology search. Expressed sequence tags (EST) of *Cannabis sativa* were blasted against identified plant miRNA sequences by following a series of filtration criteria.

miRNA identification was achieved by C-mii (version 1.11) (Numnark *et al.*, 2012) in ginger by Singh *et al.* (2016). 16 miRNAs belonging to 16 different families (miR167, miR407, miR414, miR5015, miR5021, miR5644, miR5645, miR5656, miR5658, miR5664, miR827,miR838, miR847,miR854, miR862 andmiR864) were identified by using ESTs of *Zingiber officianale*. The predicted miRNA sequences were blasted against reported *Arabidopsis* miRNAs in miRBase in order to remove protein- coding sequences.

Din *et al.* (2016) used EST-based algorithm and found 88 miRNAs belonging to 81 miRNA families in chilli. 8,749 known precursor and mature miRNA sequences obtained from miRBase were used as reference and chilli ESTs of about 1,18,572 obtained from dbEST were blasted and potential miRNAs were obtained.

Zinati *et al.* (2016) found two putative miRNAs in saffron stigma (*Crocus sativus)* by homology search. EST sequences of about 6,202 were downloaded from NCBI and searched against available mature miRNAs in miRBase.

In *Ravoulfia serpentine,* Prakash *et al.* (2016) identified miRNAs by BLAST analysis with 7,201 mature miRNA sequences from 53 plant species, obtained from miRBase against 99,673 nucleotide sequences, downloaded from GenBank. The result includes 15 potential miRNAs belonging to 13 different families.

In *Cichorium intybus*, Srivastava *et al.* (2017) reported 28 miRNAs by applying BLASTn analysis between ESTs, GSS and nucleotide sequences of *C. intybus* from NCBI against reported miRNAs of *Arabidopsis thaliana and Oryza sativa* from miRBase.

Homology search was performed for miRNA identification in *Eclipta prostrata* by Sahoo *et al.* (2018). BLAST analysis was performed with 84 GSS sequences of *E. prostrata* against 4802 known mature miRNAs of *Viridiplantae* plant kingdom. 8 potential miRNAs belonging to 7 different families were found by comparative analysis.

In switch grass (*Panicum virgatum* L.), ESTs and GSS sequences were subjected to BLAST analysis against 5,744 known plant precursor and mature miRNA sequences of miRBase. The result includes 158 new miRNAs belonging to 83 families (Barozai *et al.*, 2018).

Hussain *et al.* (2018) identified five novel miRNAs in pigeon pea by BLAST analysis. Homology search was done between EST sequences of pigeon pea, mined from dbEST database and mature plant miRNA sequences obtained from miRBase.

2.5.1.2 Bioinformatic or computational prediction

With the help of bioinformatic or computational tools, miRNAs were predicted efficiently with high degree of accuracy, as miRNAs cannot be detected easily through direct cloning because of their low expression level (Dong *et al.,* 2012).

Several computational tools have been developed to identify novel miRNAs. Main characteristic features used by different computational tools for miRNA prediction were length, high sequence conservation among species and structural features like hairpin and minimal folding free energy (Li *et al.*, 2010). Nowadays, several databases help in insilico prediction of miRNAs. Bioinformatic tools used to predict miRNAs were NovoMIR, HHMMiR, MiPred, MIRFINDER, MiR-PD.

HHMMiR

HHMMiR calculates the minimal folding free energy of sequences using RNAfold that possess at least 10 base pairs in stem-loop region and classifies them via Hierarchical Hidden Markov Model (HHMM). The sensitivity of HHMMiR was reported as 0.865 for *Oryza sativa,* 0.973 for *Arabidopsis thaliana* (Kadri *et al.,* 2009).

NovoMIR

NovoMIR is a widely used bioinformatic tool that uses a series of filter steps followed by a statistical model to identify pre-miRNAs from all the other RNAs. For detection, NovoMIR takes a genomic sequence as input. It was written in Perl script and was tested under Linux. This tool was trained using *Arabidopsis thaliana* miRNAs as positive controls and tRNA, mRNA, noncoding RNA and genomic sequences as negative controls. The specificity and sensitivity of NovoMIR was ~0.83 and ~0.99 respectively for prediction of pre-miRNAs. NovoMIR relies on two programs *i.e* RNAfold and RNAshapes for prediction of RNA secondary structure (Teune and Steger, 2010).

Teune and Steger (2010) searched for pre-miRNAs in *Arabidopsis* genome by using NovoMIR bioinformatic tool. All intergenic and intronic regions of Arabidopsis genome obtained from The Arabidopsis information resource (TAIR) were analysed by NovoMIR and reported 1,477 sequences as potential pre-miRNAs.

Moss and Collis (2012) downloaded *Linum usitatissimum* (Flax) genome from NCBI. Genome was searched for pre-miRNAs by using NovoMIR bioinformatic tool and predicted 2,324 potential pre-miRNAs and 10,692 mature miRNAs respectively.

2.5.2 Genetic Approach

2.5.2.1 Cloning and Sequencing

In order to identify miRNAs that are difficult to predict by In silico analysis, genetic approach by direct cloning and sequencing was developed. Total RNA was isolated from plant samples and RNA fragments were size fractioned in agarose gel. Ligation was done with 3^{I} adapter followed by 5^{I} adapter. RNA was reverse transcribed to cDNA by reverse transcriptase and was cloned by PCR. PCR products were sequenced by Illumina HiSeq 2000.

Zhang *et al.* (2011) reported 13 novel miRNAs belonging to 29 different miRNA families in sweet sorghum. Small RNAs of 18-28 nt were isolated from 3 week old sweet sorghum plants. RNA fragments were size fractionated and ligated with 5^{I} and 3^{I} RNA adapters. Ligated small RNAs were reverse transcribed and PCR was done. The final PCR product was purified and sequenced by using 454-pyrosequencing technology.

A total of 86 miRNAs belonging to 25 families were predicted by Wang *et al.* (2015) in Chinese cabbage by cloning and sequencing. Two sRNA libraries were constructed by using RNA samples from healthy and turnip mosaic virus infected

leaves of cabbage. RNA sequences were sequenced by Illumina–Solexa high throughput sequencing technology.

Jin and Wu (2015) found 3 novel and 39 known miRNAs in cucumber in response to *Pseudo peronospora cubensis* infection. RNA was isolated from healthy and infected leaves using Trizol reagent. RNA sequences extracted were sequenced by sRNA deep sequencing method. miRNAs were identified with their secondary structure by using RNAfold package (version 1.8.2).

Song *et al.* (2016) identified miRNAs, which were differentially expressed in fusarium wilt resistant and susceptible banana cultivars by cloning and sequencing. RNA samples of banana were used for library construction. After adaptor ligation, reverse transcription was done. Resulted PCR products were sequenced by Illumina Hi-seq and found 139 miRNAs belong to 38 different families were found.

A total of 218 miRNAs belonging to 21 different miRNA families were identified by Kuruvilla (2017) in *Hevea brasiliensis* by cloning and sequencing approach. Total RNA was isolated and the fragments were size fractioned in 1% agarose gel. Then they were ligated with 3^{1} adapter followed by 5^{1} adapter. RNA was converted to cDNA by reverse transcription followed by PCR. PCR products were sequenced by Illumina HiSeq 2000.

Drought responsive miRNAs in tef (*Eragrostis tef* (Zucc.) Trotter) were identified by Martinelli *et al.* (2018). RNA was isolated from roots and shoots of tef under drought stress and sequencing was performed at Fasteris, using Illumina Hiseq 2500 technology. After adaptor ligation, sequences were searched for pre-miRNAs against known miRNAs in miRBase and found 147 sequences as novel miRNAs.

2.5.3 miRBase

miRBase is a biological database that acts as a repository of all miRNA sequences and annotations. The latest miRBase release 21 (released on 26 June,

2014) contains 28,645 entries representing hairpin precursor miRNAs, expressing 35,828 mature miRNAs products in 223 species (Kozomara and Griffith-Jones, 2014). miRBase was available at http://microrna.sanger.ac.uk/. Five aims of miRBase were to provide consistent naming system for miRNAs, provide a central place for all miRNAs, provide computer and human readable information for miRNAs, provide evidences for miRNAs, to link and aggregate to miRNA target information. Validated or annotated miRNAs can be searched in miRBase by knowing their name or keyword or genomic location on chromosome or their tissue expression or by their sequence. It was initially used to assign names to miRNAs. It allows users to search for miRNAs in a wide range of species and to download the file containing the miRNA sequence information. It also allows to submit newly identified miRNAs to the database (Watanabe *et al.,* 2007). Argonaute database usually provides comprehensive information regarding miRNAs and their targets (Shahi *et al.,* 2006).

2.6 PREDICTION OF POTENTIAL miRNA TARGETS

Biological role of individual miRNAs were detected by the mRNAs they regulate. Identification of miRNA target genes helps in understanding the role of miRNAs in gene regulatory networks (Watanabe *et al.,* 2007). Regulation of protein expression was done by binding to one or more target sites on mRNA transcript by miRNAs and causing cleavage or repression of translation (Chaudhuri and Chatterjee, 2007). miRNAs target mRNAs through complementary base pairing by binding to 3^{1}- UTR of target transcripts in two ways (Rajewsky, 2006). In the first approach, target transcripts have perfect complementarity to the 5^{1} end of miRNA *i.e.,* seed region present at 2-8 nt of miRNA. In the second approach, target transcripts have imperfect complementarity to the 5^{1} end of miRNA. A single miRNA can have multiple target sites on a transcript and a transcript can have target sites for many miRNAs. This many-to-many relationship between miRNAs and mRNAs lead to complex miRNA regulatory mechanisms (Liu *et al.,* 2012). In plants, miRNAs shows near perfect complementarity to their targets. But in case of

animal, miRNAs display partial complementarity to their targets (Carrington and Ambros, 2003; Millar and Waterhouse, 2005; Axtell *et al.,* 2011).

Bentwich (2005) reported four main considerations for computational prediction of miRNA targets: 1) extraction of rules related to formation of miRNA–mRNA duplex 2) introduction of rules in computational algorithms 3) prediction of novel miRNA targets using these algorithms 4) validation of targets. But, in case of animal miRNAs, miRNA–mRNA duplex is not perfectly complementary as miRNAs are very short. A number of algorithms were developed to search miRNA target genes (Li *et al.,* 2010). Computational methods developed for target prediction of miRNAs include miRanda, psRNATarget, TargetScan, RNAhybrid, PITA (Kertesz *et al.,* 2007), PicTar (Krek *et al.,*2005), RNA22 (Miranda *et al.*, 2006), mirSVR (Betel *et al.*, 2010), DIANA-microT (Maragkakis *et al.,* 2009). These algorithms were used to predict targets based on base pairing, degree of hybridization, target accessibility, evolutionary conservation of target site. Thermodynamic properties of miRNA–mRNA duplex were analysed by calculating free energy in most of the target prediction algorithm. Experimental validation of targets helps in identification of regulated targets of miRNAs. miRBase target database provides information regarding the predicted target genes of miRNAs of various species (Watanabe *et al.,* 2007).

miRanda (http://www.microrna.org)

This software was designed to predict targets in *Drosophila melanogaster*. It predicts targets based on sequence complementarity, evolutionary conservation and binding energy of miRNAs (John *et al.,* 2004). This algorithm was also used to predict human miRNA targets (Enright *et al.,* 2003).

TargetScan (http://www.targetscan.org)

This algorithm predicts targets by searching for the presence of conserved octamer and heptamer sites that match the seed region (Lewis *et al.*, 2003).

RNAhybrid

This algorithm was used to identify miRNA secondary structure by hybridization between miRNA and target mRNA (Rehmsmeier *et al.,* 2004). It is an extension of RNA secondary structure prediction tool RNAfold and Mfold (Mathews *et al.,* 1999).

psRNATarget

Many algorithms were developed for target prediction of animal miRNAs. An online target server psRNATarget, a plant small RNA target analysis server was used for plant miRNA target prediction. It relies on two functions: reverse complementary matching between miRNA and target transcript using a proven scoring schema technology and target site accessibility evaluation by calculating unpaired energy (Dai and Zhao, 2011). The psRNATarget server was freely available at http://plantgm.noble.org/psRNATarget/. The server consists of three interfaces: user-submitted small RNA sequences, user-submitted target candidates and user-submitted small RNA and target sequences.

In user-submitted small RNAs interface, RNA may be miRNA or siRNA. Sequence must be in FASTA format and only ATCGUN were valid sequence letters. This interface will search for possible sRNAs in submitted RNA sequences that were complementary to validated target transcripts.

In user-submitted target interface, target sequence may be cDNA, EST, Unigene, mRNA or genomic segment. This interface will search for possible target sites in submitted sequences for validated RNA sequences. The sequence must be in FASTA format.

In user-submitted small RNA and target interface, both target sequences and small RNA sequences were submitted by the user in FASTA format. Based on reverse complementarity matching, targets were found for those RNA sequences.

After submitting the miRNA and target sequences, target prediction was done based on scoring schema technology. It consist of three schemas *i.e* Schema V1 (2011 release), Schema V2 (2017 release) and User-customized schema.

Scoring Schema V1 (2011 release) was developed based on animal model. Features include expectation value: 3 and seed region ranges from 2-8 nt. Scoring Schema V2 (2017 release) include features like expectation value: 5 and seed region ranges from 2-12 nt. In user-customised Schema, there is no limit for expectation value and seed region.

In each Schema, default parameters were loaded. Users can adjust the parameters according to the schema selected. Output file will be in the form of tab-delimited text file, which is useful for large scale data analysis.

Frazier *et al.* (2010) found 1,225 putative target genes belonging to different families in tobacco by using BLASTn software. 259 potential miRNAs were searched against EST sequences and protein coding sequences by using BLASTn to find target genes.

To identify potential miRNA targets in expressed sequence tags of Tea, Prabhu and Mandal (2010) blasted miRNA sequences against mRNA/cDNA sequences of tea obtained from NCBI and found 30 targets for four potential miRNAs.

Zhang *et al.* (2011) reported 125 genes as potential targets for sorghum miRNAs. Annotated sorghum coding sequences were used in BLAST analysis against predicted novel or complementary sequences of miRNAs to find targets based on complementary base pairing analysis.

In *Linum usitatissimum,* Neutelings *et al.* (2012) used psRNATarget server to find targets of flax miRNAs. Predicted miRNAs were searched against EST database of flax with default parameters and found 112 miRNA targets.

psRNATarget server with default parameters was employed to find target genes in sweet potato by Dehury *et al.* (2013). ESTs of sweet potato were analysed against *Arabidopsis thaliana gene sequences* obtained from DFCI gene index and found 42 targets for 3 miRNA families.

Pani and Mahapatra (2013) used potential miRNAs as query for BLAST analysis against mRNA sequences to find candidate target genes. Seven mRNA sequences were predicted as targets with default parameters.

In *coffee arabica*, potential targets of novel miRNAs were reported by Akter *et al.* (2014). psRNATarget server with default parameters, was used to identify 6 targets using predicted coffee miRNAs as query against cDNA sequences obtained from *A. thaliana* DFCI Gene index.

Panda *et al.* (2014) found 33 targets for 6 miRNA sequences in garlic by using psRNATarget server. Garlic miRNA sequences were used as query sequences against Ref Seq mRNAs and assembled ESTs of garlic.

Target genes of novel miRNAs in stevia were predicted by using psRNATarget program (Mandhan and Singh, 2015). Target prediction was executed with default parameters. Newly identified miRNA sequences were used as query and *Stevia rebaudiana* transcript library (EST and nt databases) as transcript database for target prediction.

Mishra *et al.* (2015) used psRNATarget server to find target genes in *Humulus lupulus* (hop). 47 potential miRNA targets were reported by using predicted miRNAs in hop as query against target transcripts of hop in DFCI plant gene index.

miRNA targets in *Cannabis sativa* were identified by using psRNATarget server. Predicted mature miRNA sequences were used as query to find complementary sequences in *Arabidopsis thaliana* Unigene (DFCI gene index, version 15) using default parameters (Das *et al.*, 2015).

Wang *et al.* (2015) reported new miRNA targets in Chinese cabbage by BLAST analysis. 271 target genes were identified by searching 86 miRNA sequences against EST and cDNA sequences of *Brassica rapa* genome database.

In *Phaseolus vulgaris,* Nithin *et al.* (2015) found putative targets of miRNAs by using psRNATarget server. EST sequences of *P.vulgaris* obtained from NCBI GenBank were used as target transcripts against miRNA sequences. 1,303 target sequences were found for 310 miRNAs with default parameters.

In banana, putative target genes of novel miRNAs were reported by Chai *et al.* (2015). Using miRNA sequences as query, psRNATarget server with default parameters was used to predict 244 target pairs in banana against *Musa* spp. all gene coding sequences.

Biswas *et al.* (2016) used psRNATarget server to understand the biological functions of predicted miRNAs of *p. hexandrum.* Target gene search was conducted using default parameters and found that miRNAs downregulate the expression of target genes.

To identify the targets of newly identified miRNAs in *Cannabis sativa,* Hasan *et al.* (2016) used psRNATarget server with default parameters. Newly identified miRNAs were used as query against gene sequences of *A. thaliana* DFCI gene index and *A. thaliana* Tar10, cDNA removed miRNA genes. Based on scoring schema technology, 14 mRNA targets were found.

Capsicum annum (Chilli) miRNAs were subjected to psRNATarget analysis to found miRNA targets with default parameters by Din *et al.* (2016). Newly identified miRNAs were used as query and the selected target transcript or library was Chilli, unigene, DFCI gene index (CAGI), version 4.

Prakash *et al.* (2016) found the targets of predicted *Ravoulfia serpentina* miRNAs by using psRNATarget analysis sever. A total of 32 genes were identified

as targets by using user submitted small RNA / user submitted transcript schema with default parameters.

Eight Potential mRNA targets were found, as miRNA sequences from *Boutelo gracilis* were subjected to target prediction by using psRNATarget server (Baquera *et al.*, 2017). miRNA sequences were used as query and annotated rice mRNA transcript database was used as reference genome.

miRNA targets of *Eclipta prostrata* were found by perfect or near complementary matching between predicted miRNAs and *Helianthus annuus, unigene,* DFCI gene index using psRNATarget server with default parameters. A total of five target genes were found by Sahoo *et al.* (2018).

In pigeon pea, *Cajanus cajan mRNA sequences* obtained from GenBank were used as reference genome against predicted mature miRNAs, used as query sequence in BLASTn analysis for the prediction of 12 mRNA targets (Hussain *et al.*, 2018).

mRNA or protein coding genes that are targeted by miRNAs in tef were found by Martinelli *et al.* (2018) using psRNATarget, a miRNA target analysis server for plants. Based on complementary base pairing/scoring schema technology, six gene sequences were reported as potential targets.

Targets for switch grass miRNAs were found by using psRNATarget server by Barozai *et al.* (2018) based on scoring schema technology. Newly identified miRNAs were subjected against *Panicum virgatum* (switch grass), unigene, DFCI gene index and found 894 gene sequences as mRNA targets.

2.7 FUNCTIONAL ANNOTATION OF miRNA TARGETS

Radivojac *et al.* (2013) reported that gene products which include proteins and RNAs play a crucial role in plant metabolism, signal transduction and hormonal regulation. Functional annotation of these protein genes helps in disease analysis, discovery of drugs and enrichment analysis of genes.

Gene ontology analysis acts as a leading tool for computational prediction of gene functions. It provides a set of terms to identify the functions of gene products from all the organisms. Usually, it includes three ontologies: cellular component, molecular function, biological function. Blast2Go (Gotz *et al.,* 2011) is the newly identified and most commonly used tool for functional annotation or gene ontology analysis of miRNA targets. Identified miRNA targets were done BLASTX against nr database of NCBI. The best hits obtained were searched against GO and KEGG databases using default settings.

Gotz *et al.* (2011) reported several functions of targets by using Blast2Go analysis. Targets includes transcription factors and several enzymes which were involved in plant metabolic pathways.

In *Catharanthus roseus,* Pani and Mahapatra (2013) reported the functions of predicted miRNA targets by using Blast2Go analysis. Maximum number of targets encode enzymes that regulate synthesis of terpenoid indole alkaloid (TIA) and also include transcription factors, which were involved in cell growth and development, signaling and metabolism.

Panda *et al.* (2014) employed gene ontology analysis in *Allium sativum* by using Blast2Go server. Target genes of garlic include transcription factors and genes involved in metabolism, stress response, plant growth and development.

In chilli (*Capsicum annum* L.), functional annotation of miRNA targets is done by Blast2Go analysis (Din *et al.,* 2016). Most of the predicted targets were categorized based on their function. They include: hypothetical protein 31%, metabolism 18%, transcription factor 27%, transporter 6%, signal transduction 4%, stress related 4%, growth and development 2%.

2.8 EXPERIMENTAL VALIDATION OF miRNAS

Computer-based approaches for miRNA identification and target prediction were being considered as indispensable in miRNA research. Similarly, effective

experimental techniques for validating in-silico predictions were crucial for testing of computational algorithms. Interactions between in-silico and experimental methods were playing a central role in the biology of miRNAs (Chaudhuri and Chatterjee, 2007). Watanabe *et al.* (2007) reported two strategies for validation of miRNA target genes: evaluation of known miRNA target genes; evaluation using artificial miRNA sequences. As computational methods are not perfect and there is a risk of false positive prediction, experimental validation was used to find novel miRNAs. Experimental validation of miRNAs include designing stem loop primers for miRNAs and gene specific primers for target genes. RNA will be isolated from plant samples and RT-PCR was carried out. Real-Time PCR was carried out using these primers for quantification of gene expression of miRNAs.

2.9 EXPRESSION ANALYSIS OF miRNAS

Expression analysis of genes helps in knowing the biological regulatory and metabolic activities of all organisms. By understanding the expression of genes under different stresses, biological functions of these genes can be predicted (Chen *et al.,* 2014). Several approaches have been developed for expression analysis of genes. They include Northern blot, Quantitative Real Time Polymerase Chain Reaction, microarrays hybridization, serial analysis of gene expression and RNA-seq of Next-Generation Sequencing (NGS).

qRT-PCR (Real-Time PCR) is most reliable method to determine the presence and quantify the expression levels of miRNAs (Chen *et al.,* 2005).

Due to small size of miRNAs and different expression levels, quantification and detection of miRNAs has become a challenging task. But, Real-Time PCR technique developed by Chen *et al.* (2005) provide an accurate and sensitive method for miRNA detection and quantification.

In Real-Time PCR, absolute and relative quantification assays were present. For expression analysis, relative quantification of samples was done. In this assay, relative expression of a gene in one sample is compared to expression of same gene

in another sample. Results are expressed as fold change (increase or decrease) in expression. A normal or house keeping gene is used as control for experimental viability.

Two types of controls are used in Real-Time PCR. No template control (NTC) contain all the components except cDNA. If amplification occurs, it is due to primer dimers and contamination of nucleic acids. No reverse transcriptase control (No-RT) contain all the components except reverse transcriptase. If amplification occurs, genomic DNA was amplified rather than cDNA.

MATERIALS AND METHODS

3. MATERIALS AND METHODS

The study entitled "Computational prediction of miRNAs in banana (*Musa* spp.) and evaluation of their role in virus infection" was conducted at the Department of Plant Biotechnology, College of Agriculture, Vellayani, Thiruvananthapuram and Central Tuber Crops Research Institute (ICAR-CTCRI), Sreekariyam, Thiruvananthapuram during 2016–2018. Details regarding the experimental materials used and the methodology followed for various experiments are presented in this chapter.

3.1 PREDICTION OF NOVEL miRNAS IN BANANA

3.1.1 Retrieval of Banana Chromosome Sequences

Banana (*Musa acuminata*) genome consist of 11 chromosomes with genome size of 523 Mb, which was estimated by flow cytometry (D'Hont *et al.,* 2012). The reference genome assembly was generated from DH-Pahang, a double-haploid genotype (2n = 22), which belongs to *Musa acuminata* ssp. *malaccensis* of accession named 'Pahang'. Chromosome sequences of *Musa acuminata* DH Pahang v1 (released in July 2012) were obtained from Banana Genome Hub (Figure 3) (Available at http://www.bananagenomehub.com/). Banana Genome Hub contains genetic and genomic data of *Musa acuminata*, developed by Cirad and Bioversity International and supported by South Green Bioinformatics Platform (Droc *et al.,* 2013). All the gene-coding sequences of *Musa acuminata* DH Pahang v1 were downloaded in nucleotide FASTA format. Each chromosome consist of numerous scaffold sequences (Table 1). As NovoMIR cannot read large sized sequences, each scaffold sequence of a chromosome was made into a FASTA file with an extension *.fa*. Each scaffold sequence was subjected to computational analysis for pre-miRNA prediction by using a bioinformatic tool NovoMIR.

Table 1. Number of scaffolds per chromosome of banana genome

Chr No.	1	2	3	4	5	6	7	8	9	10	11
No. of Scaffolds	21	21	22	13	20	30	21	27	37	33	15

3.1.2 Prediction of pre-miRNAs

Prediction was done using the tool NovoMIR. It uses a series of filter steps followed by a statistical model for prediction of plant pre-miRNAs by discriminating pre-miRNA from other RNAs. It analyses all the scaffold sequences of banana genome for stem loop structure. Command used for sequencing by NovoMIR (Figure 4): perl novomir.pl –f chr1.fa –R >Rchr1.fa. Various options used by NovoMIR are as follows (Figure 5):

- f Sequence-file - sequences must be on one single line
- g genome-wide search for potential candidates
- e energy-value for RNA shapes
- m minimum length of stem-loop-region
- t threshold for prediction
- h print the usage (Boolean)
- R use RNAshapes version 3.1.4 (default: 2.1.6)
- l long output (short output is default)
- v show version.

Out of various options, option -f was used for sequencing, -R for RNAshapes version, >R for creating the output file. -f could take either a single-seq-file or a directory with EMBL/GENBANK.

Each pre-miRNA sequence in the output file of NovoMIR started with hashes ("#"). NovoMIR replaces each character of a sequence by a minus sign except lower

case (a-z), upper case (A-Z) alphabets and numbers (0-9). In NovoMIR result, the forward strand starts at 5^1 end and the reverse strand starts at 3^1 end. The forward and reverse strands of pre-miRNA were connected by a hairpin loop. The two lines in between the forward and reverse sequence strands showed the secondary structure representation as:

" + + " base pair

" – " mismatch (or symmetric part of a loop)

" I- or -I " asymmetric (part of a) loop

The positions of the predicted miRNA/miRNA* complex is marked by stars followed by the start and end position in the forward sequence.

3.1.3 Secondary Structure Prediction of pre-miRNAs

Secondary or hairpin loop structure of pre-miRNAs was determined by using an online application called RNAfold web server (Figure 6), which is included in Vienna RNA software package version 2.1.5 (Lorenz *et al.*, 2011). It is available at http://www.RNAfoldwebserver.com/ (Hofacker, 2004). RNAfold predicts the secondary structure of single stranded RNA or DNA sequences, along with MFE, expressed in Kcal/mol. Secondary structure was predicted by submitting the pre-miRNA sequence and selecting the minimum free energy fold algorithm. Output is the graphical representation of secondary structure along with MFE, downloaded in PDF format.

For the characterization of pre-miRNAs, along with MFE, Adjusted minimal folding free energy (AMFE) and Minimal folding energy index (MFEI) values were also considered. AMFE is calculated by dividing MFE with sequence length of pre-miRNA and multiplying the result with 100 (Zhang *et al.*, 2006b).

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

MFEI = [(MFE/ length of RNA sequence) × 100] / (G+C) %

Fig. 3. Home page of Banana Genome Hub

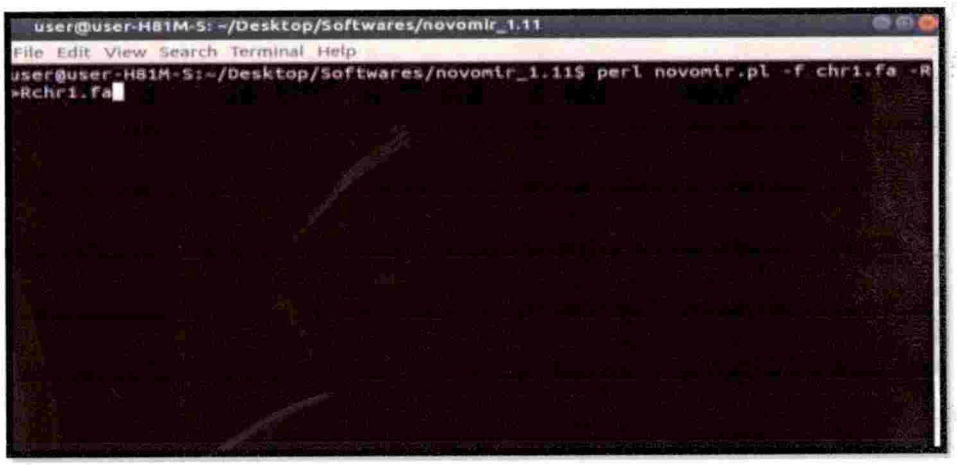


Fig. 4. Command used for NovoMIR

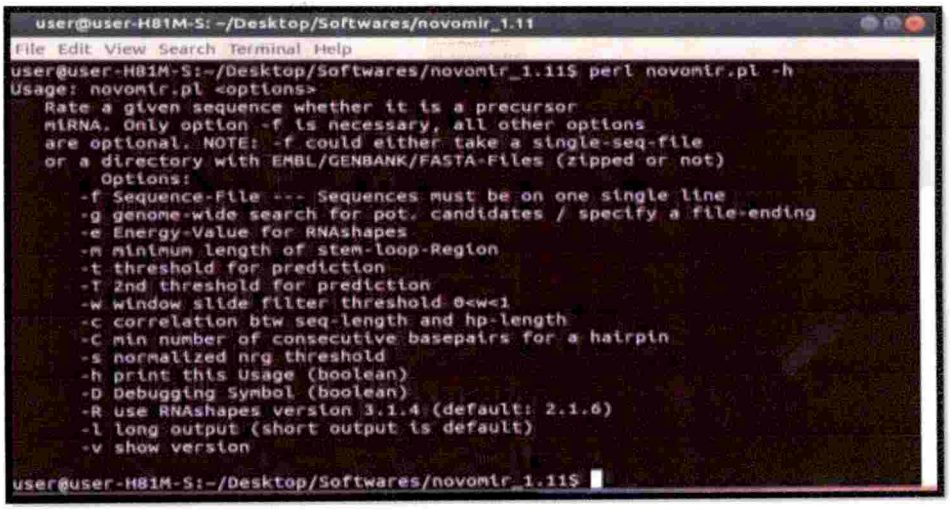


Fig. 5. Options used by NovoMIR

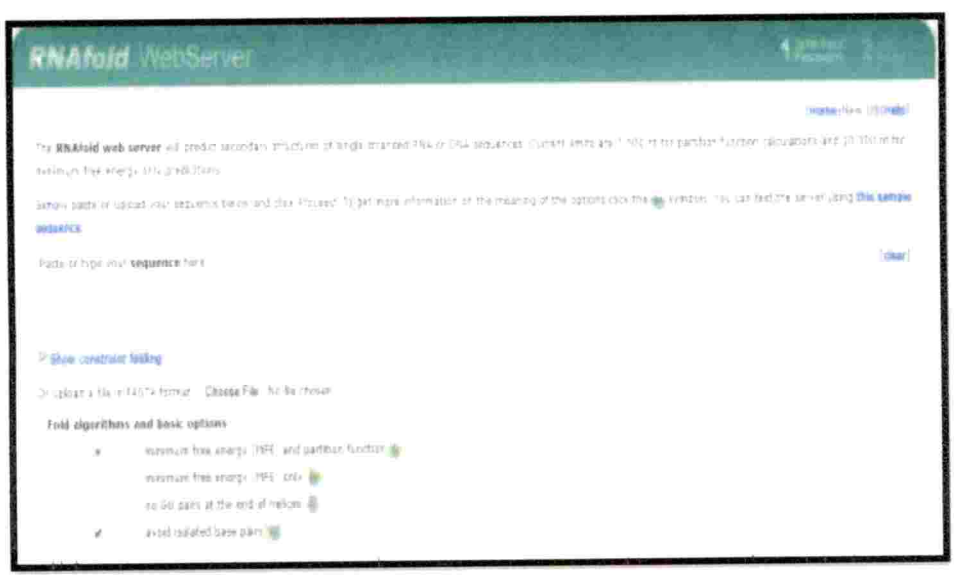


Fig. 6. Homepage of RNAfold webserver

3.1.4 Identification of Mature miRNAs

In order to obtain mature miRNAs, pre-miRNA sequences were analysed by BLAST algorithm (Figure 7) against annotated mature miRNAs of miRBase. miRBase is a biological database that acts as an archive of miRNA sequences and annotations.

Parameters included E value, which indicates number of alignments that are expected to occur in the search. Lower the E value, more significant the score is. BLAST score determines the overall degree of alignment between sequences. Higher score indicates more similarity. Default parameters used were E value cut-off – 10; maximum no. of hits – 100; score: > 70.

The results were evaluated based on the following parameters: candidate miRNA should contain at least 18 nt in length; number of mismatches between known miRNAs and pre-miRNA was taken as ≤ 3; score of the BLAST result: > 70 (Panda *et al.*, 2014).

3.1.5 Properties of Mature miRNAs

To calculate the properties of single stranded and double stranded DNA or RNA, an online tool called OligoCalc (Figure 8) was used. OligoCalc is available at http://basic.northwestren.edu/biotools/OligoCalc.html. After submitting the nucleotide sequence and selecting the sequence option as single stranded RNA, it calculated the properties like miRNA molecular weight, G+C %, A+U % and melting temperature (T_m). It also include other options like swapping the sequence for its complement, submitting the sequence to NCBI, calculating self-complementarity between two molecules (Kibbe, 2007).

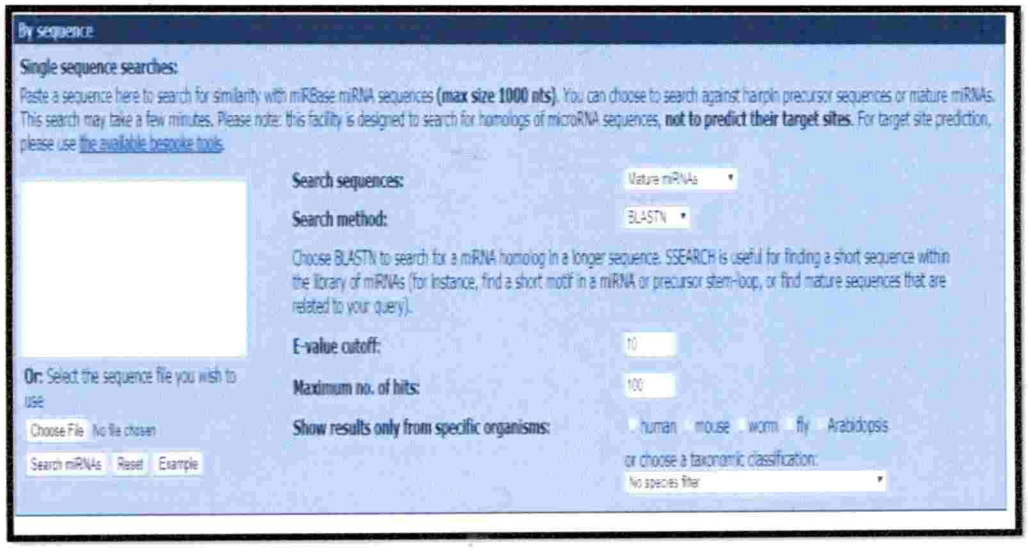


Fig. 7. BLAST page of miRBase

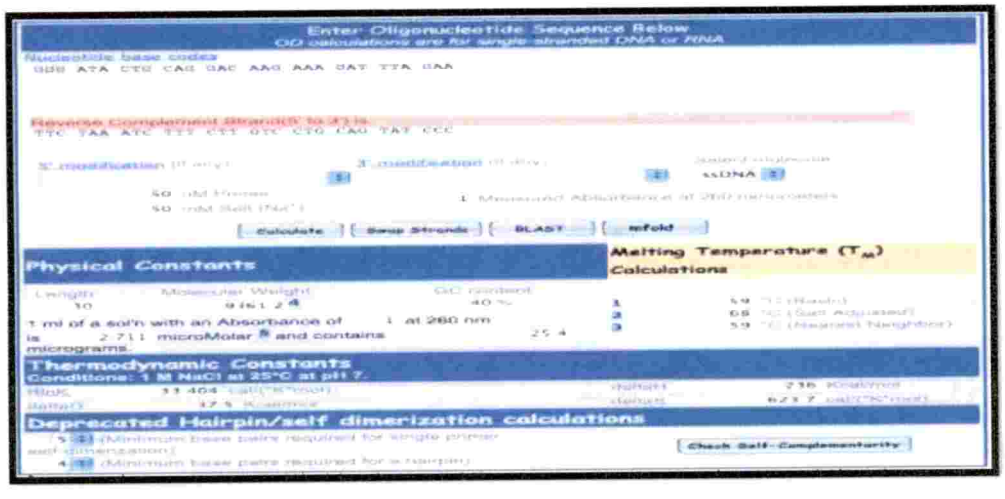


Fig. 8. Home page of Oligo Calc

3.2 IDENTIFICATION OF THE miRNA TARGETS USING TARGET PREDICTION SOFTWARE

Potential targets of the predicted miRNAs in host genome and *BBrMV* genome were predicted by reverse complementary matching between miRNA sequences and target transcripts. An online target server called psRNATarget (plant small RNA target analysis server) (Figure 9) was used for the prediction of targets based on scoring schema technology (Dai and Zhao, 2011). psRNATarget is available at http://plantgrn.noble.org/psRNATarget/. It consists of three interfaces: user-submitted small RNAs, user-submitted target candidates and user-submitted small RNAs and targets.

After submitting the mature miRNA sequence as query and all the *Musa acuminata* gene-coding sequences obtained from Banana genome hub, version 1 as target transcript library, target prediction was done by psRNATarget based on scoring schema technology. It consist of three schemas: Schema V1 (2011 release), Schema V2 (2017 release) and User-customised Schema.

In User-customised Schema, default parameters were loaded as translation inhibition range: 9–11 nt, maximum expectation: 3.0, maximum energy to unpair the target site (UPE): 25.0, flanking length around target site for target accessibility analysis of 17 bp in upstream/13 bp in downstream and length for complementarity scoring (HSP size): 18–22 (Dai and Zhao, 2011).

Output file was in the form of tab-delimited text file, which is useful for large-scale data analysis.

3.3 FUNCTIONAL ANNOTATION OF THE IDENTIFIED miRNA TARGETS

miRNA targets predicted by psRNATarget, were annotated using gene ontology databases. For functional annotation of identified miRNA targets, Blast2Go software (Figure 10) was used (Gotz *et al.*, 2011). This is the most integrated approach for functional annotation of DNA or protein sequences based on gene ontology vocabulary. This software was developed by The gene ontology consortium, 2008. Blast2Go analysis consist of 5 steps: blasting, mapping, annotation, statistical analysis and visualization. Features of Blast2Go include many graphical features such as GO-graph visualization, sequence management features, high throughput capabilities and sequence management (Gotz *et al.*, 2008).

First, predicted targets of miRNAs were subjected to BLASTX analysis against nr (non-redundant) database of NCBI. Protein or amino acid sequences of all miRNA targets were downloaded from Banana genome hub and made into a FASTA file with the extension *.fasta, .fnn* or *.faa.* After loading the fasta file, NCBI mode of BLAST was selected to start search. Sequences which are blasted turned orange.

After BLAST analysis, mapping of targets was done. Mapping is a process of assigning GO terms associated to the BLAST hits. Mapping process was done by clicking on the mapping icon and those sequences which are mapped turn green.

After mapping, target sequences were annotated by using annot option. GO terms obtained by mapping, were assigned to the query sequences. The sequences annotated turn blue. GO annotation characterizes miRNA targets in terms of molecular function, biological process and cellular component. Results were obtained in the form of charts. Blast2Go results were exported and saved as HTML file with an extension *.b2g* file.

Fig. 9. Home page of psRNATarget server

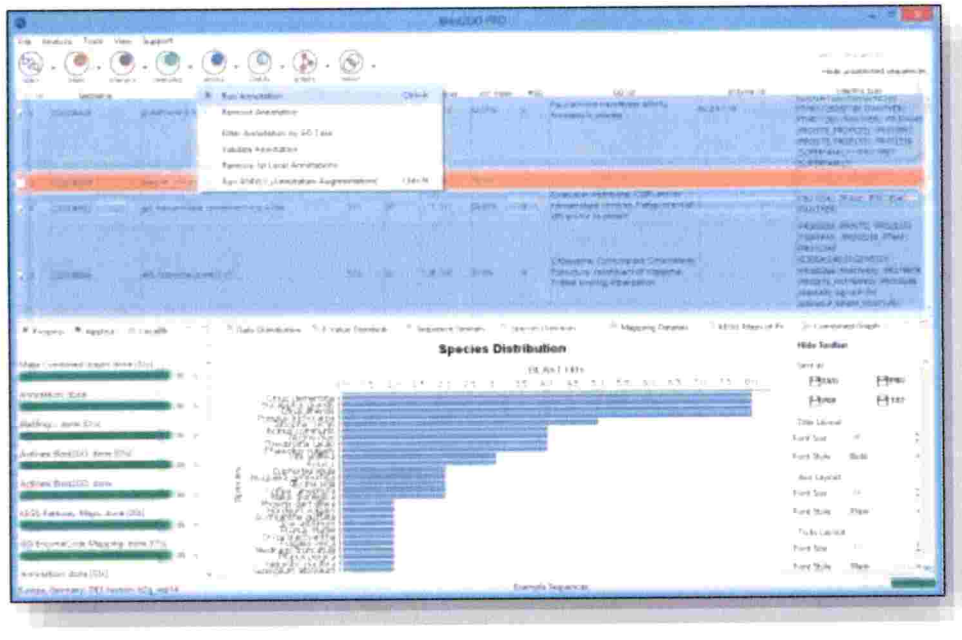


Fig. 10. Home page of Blast2Go server

3.4 EXPERIMENTAL VALIDATION OF THE NOVEL PREDICTED miRNAS

After functional annotation of miRNA targets, five miRNAs and target genes were selected, based on their biological role in the host. List of miRNAs and targets selected are summarized in Table 2.

For experimental validation of predicted miRNAs and targets, stem-loop primers for miRNAs and gene specific primers for targets were designed and synthesized.

3.4.1 Designing of Stem-Loop Primers

Stem-loop primers were designed for expression analysis of miRNAs in Real-Time PCR (qPCR). Stem-loop primer consisted of stable stem-loop structure that lengthens the cDNA. They were designed by combining 44 nt stem loop sequence of Chen *et al.* (2005) (5^{I}– GTC GTA TCC AGT GCA GGG TCC GAG GTA TTC GCA CTG GAT ACG AC –3^{I}) with reverse complement of 6 nt at 3^{I} end of mature miRNA. Along with Stem-loop RT primer, miRNA specific forward and reverse primers were also designed for amplification and quantification of miRNAs.

Forward primers, which are miRNA specific, were designed by taking the first 12 to 17 nt at 5^{I} end of miRNA and adding an extension of 5–7 nt (rich in G+C% and Tm of 60 ± 1^{O} C) to 5^{I} end of miRNA. Based on 44 nt stem-loop sequence, a universal reverse primer (5^{I}–CCAGTGCAGGGTCCGAGGTA-3^{I}) (Kramer, 2011) was also synthesized for all the miRNAs. As a reference in expression analysis of miRNAs, primers for a house keeping gene *i.e* β-actin gene was also synthesized.

3.4.2 Designing of Target Gene Primers

Primer3Plus was used for designing primers for target genes (Untergasser *et al.*, 2007). It is available at http://www.bioinformatics.nl/primer3plus. Primer3Plus web interface (Figure 11) consisted of an input box for sequence and option for selecting target region for amplification. The target sequence was submitted in

FASTA or EMBL format. Target region was indicated in [ ]. Default parameters used are (Figure 12): Length: 20-22 nt, T_m (Melting temperature): 50-65^{O}C, GC %: 55-60 %. After adjusting the default parameters, primers were designed.

To find out the specificity of primers, an online software called NCBI Primer-BLAST was used. It was developed by NCBI to design primers that are specific to target sequences. It uses BLAST software and global alignment algorithm to find primers against the PCR template *i.e.*, host or virus genome in order to avoid primer pairs that cause non-specific amplifications. Both forward and reverse primer sequences of all miRNAs were blasted against banana genome to find the score of complementarity and the best scoring pair was selected and synthesized (Ye *et al.*, 2012).

Table 2: List of miRNAs and their targets selected for expression analysis

miRNA	Sequence (5^{I}-3^{I})	Target
miR-6928-5p	GGGGAUUUUCAAGUACUGCA	FAD dependent oxidoreductase
miR-3900-5p	CAAGGAGGUGUCACCGCC	F-box/kelch-repeat protein
miR-3900-5p	CAAGGAGGUGUCACCACC	Armadillo repeat-containing kinesin-like protein 1
miR-9112	GGACCAGAGAGCAUCAUG	cytochrome c oxidase subunit 5B protein
miR-5417	AGGGGAGAAAUGGGGAUG	Glyceraldehyde-3-phosphate dehydrogenase

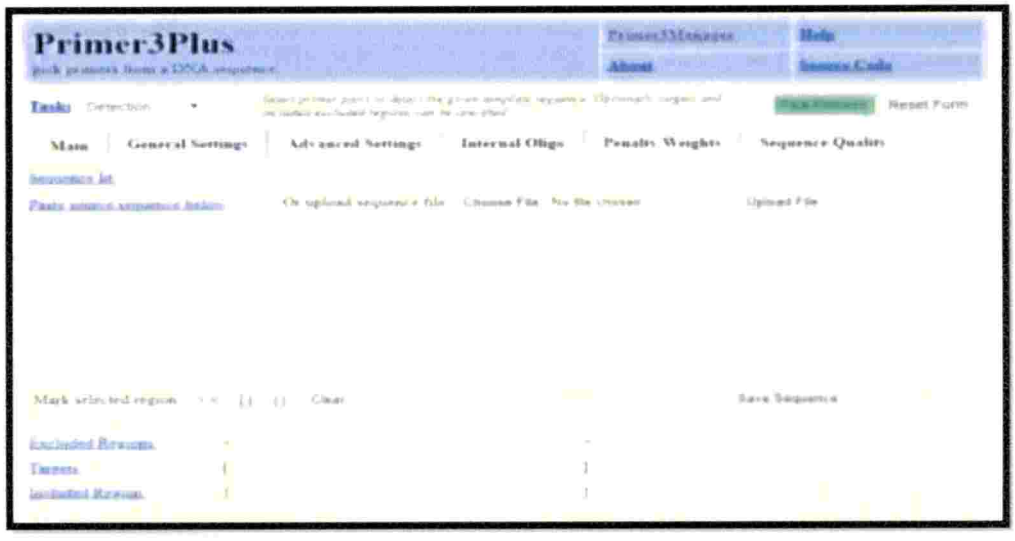


Fig. 11. Home page of Primer3Plus

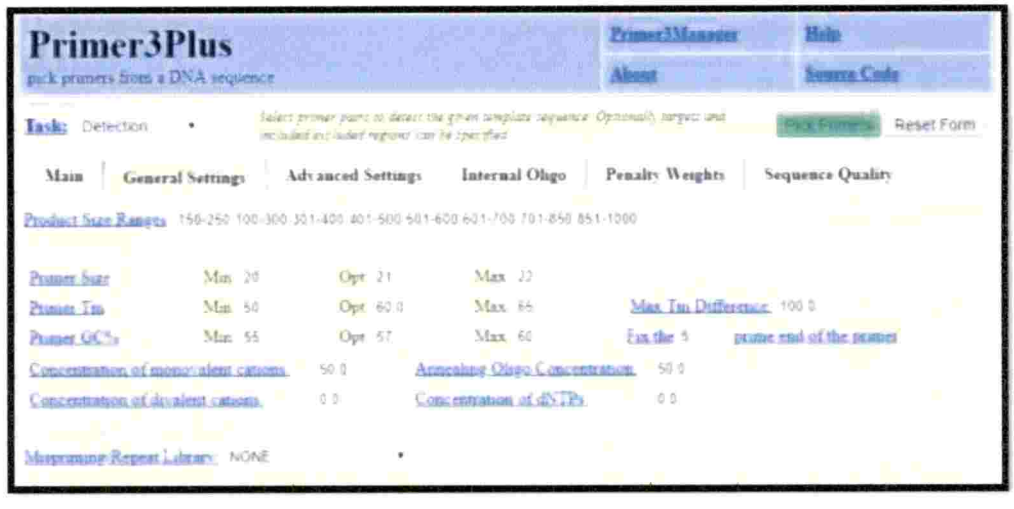


Fig. 12. Default parameters of Primer3Plus

3.5 EXPRESSION ANALYSIS OF miRNAS AND THEIR TARGET GENES DURING VIRUS INFECTION

3.5.1 Infection of Banana Plants by Banana Bract Mosaic Virus (*BBrMV*) using Aphids

Two to three months old tissue culture banana plants of variety Nendran, were infected with *BBrMV* by using banana aphids (*Pentalonia* nigronervosa), which belong to non-persistent mode of transmission. *BBrMV* infected banana suckers were collected from Instructional farm, College of agriculture, Vellayani. Aphids were collected from the infected suckers. Both infected suckers and aphids were maintained in glass house of Department of plant pathology, College of Agriculture, Vellayani, Thiruvananthapuram.

Aphids were maintained on healthy banana suckers. To start the infection process, twenty aphids were released on *BBrMV* infected sucker for acquisition feeding for 30 min, after starving them for 10 min. After acquisition feeding, the aphids were transferred from infected sucker to healthy banana plants for transmission of the virus *i.e* infection feeding for 14 hrs (Herradura *et al.,* 2003). The infection process was repeated twice a day for 7 days.

3.5.2 RNA Isolation

RNA was isolated from healthy and infected plants, using CTAB RNA extraction method, standardized by Tsai and Harding, (2003).

5 ml of extraction buffer (Appendix 5) was pre-warmed at 65° C in water bath in 15 ml centrifuge tubes. 0.5 g of leaf tissue was ground to fine powder in pestle & mortar with liquid nitrogen and was transferred to centrifuge tubes containing pre-warm extraction buffer and mixed by inverting. Under Laminar air flow chamber (LAF), 2% (v/v) β – mercaptoethanol was added to the tubes and mixed. The tubes were incubated at 65° C for 10 min. Then lysate was extracted by adding an equal volume of Chloroform/Isoamyl alcohol (24:1) and vortex it. Centrifugation

was done at 15,000 rpm for 20 min at room temperature. Supernatant was transferred to a fresh tube and re-extracted the supernatant with equal volume of Chloroform/Isoamylalcohol (24:1). To this, 1/3rd volume of 8M LiCl was added and vortex it. RNA was allowed to precipitate at -20° C for overnight. RNA was harvested by centrifugation at 4° C for 20 min at 16,000 rpm. The supernatant was removed carefully and pellet was suspended twice in 500µl absolute ethanol. Ethanol was removed and pellet was air dried under LAF. Pellet was dissolved in 50µl of DEPC water.

3.5.2.1 Qualitative Analysis of RNA

Gel electrophoresis was used to determine RNA integrity. It was carried out by using horizontal electrophoresis unit. 2g of agarose powder was melted in 1X TBE electrophoresis buffer (Tris base 10.8g, Boric acid 5.5g, 0.5M EDTA 4ml). After cooling it to room temperature, ethidium bromide (EtBr) was added. The gel was poured into casting tray and comb was inserted based on the number of samples. After solidification, comb was removed and gel was placed in electrophoresis tank, which contained 1X TBE buffer. RNA samples were loaded along with 6X gel loading dye (Bromo phenol blue) to increase the density of the RNA sample. Voltage was set (15 V/cm) and RNA was allowed to run through the gel. As RNA is negatively charged, it moves towards anode. After electrophoresis was done, gel was documented by using Gel Doc Unit (Bio-Rad) using Quantity One Software.

3.5.2.2 RNA Quantification

Absorbance of RNA sample was taken to determine the quality and quantity of RNA. RNA quantification was done by using Nano drop at 260 & 280 nm. Absorbance value of 1 at 260 nm indicates presence of 40 ng μl^{-1} of RNA. Concentration of RNA was calculated by using the formula:

Conc. of RNA ($\mu g\ ml^{-1}$) = $A_{260} \times 40\ \mu g\ ml^{-1} \times$ dilution factor

Ratio of A_{260} and A_{280} indicated the quality of RNA. Ratio ranging from 1.8–2, indicates good quality RNA.

3.5.3 cDNA Synthesis

cDNA was synthesized from isolated RNA samples by using Verso cDNA synthesis kit. A 10µl reaction mix was prepared for reverse transcription of RNA using stem-loop primers. To a sterile microfuge tube, following components were added in the following order:

5X cDNA synthesis buffer (1X)	2 µl
dNTP mix (500 µM each)	1 µl
Stem loop primers (10 µM)	0.5 µl
RNA sample	2.5 µl
Verso Reverse transcriptase enzyme	0.5 µl
Nuclease free water	3.5 µl
Total Volume	10 µl

The contents were mixed well by spinning or vortexing and PCR was carried out by using Master cycler of Eppendorf.

Thermal profile used for reverse transcription of miRNAs:

Incubation at 16° C – 30 min

Step 1: 30° C for 30 s

Step 2: 42° C for 30 s

Step 3: 50° C for 1 s

Step 4: 85° C for 5 min (to inactivate RT enzyme). Steps 2-4 were repeated for 60 cycles. cDNA samples were stored at -20° C.

A 20 µl reaction was prepared for reverse transcription of RNA using oligo-dT primers for target genes. To a sterile microfuge tube, following components were added in the following order:

5X cDNA synthesis buffer (1X)	4 µl
dNTP mix (500 µM each)	2 µl
Oligo-dT primer (10 µM)	1 µl
RT enhancer	1 µl
RNA sample	5 µl
Verso Reverse transcriptase enzyme	1 µl
Nuclease free water	6 µl
Total volume	20 µl

Thermal profile used for reverse transcription of target genes:

Step 1: 42° C for 30 min

Step 2: 95° C for 2 min

3.5.4 Confirmation of cDNA Synthesis

PCR analysis was done to confirm the cDNA synthesis. cDNA was amplified with the specific primers of house keeping gene *i.e* actin gene to confirm successful conversion of RNA to cDNA. A standard PCR mix was prepared for 20 µl total volume containing:

10X reaction buffer (1X)	2 µl
dNTP mix (100 µM each)	1 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
Taq DNA polymerase	1 µl
Template DNA	1 µl
Nuclease free water	13 µl
Total volume	20 µl

cDNA samples of both healthy and infected plants were amplified by initial denaturation at 95 ° C for 3 min, followed by 30 cycles of denaturation at 95° C for 15 s, annealing at 60° C for 15 s and extension at 72° C for 45 s. Final extension was done at 72° C for 5 min. The steps from 2-4 were repeated for 30 cycles.

PCR product was analysed in agarose gel (1.5 per cent) in a horizontal gel electrophoresis unit. 2 µl of 100 bp ladder was loaded to one of the wells as marker. Electrophoresis was carried out at 50 V until the loading dye reached three fourth of the gel and the gel was documented by using Gel Doc unit (Bio-Rad), using Quantity One Software.

3.5.5 Confirmation of *BBrMV* Infection

Confirmation of *BBrMV* infection in banana plants was done by checking the presence of *BBrMV* replicase gene. PCR was done for amplification of cDNA from healthy and infected plants, with replicase gene primers of *BBrMV*. The sequences of replicase gene are given below:

Forward – 5' AGCAATGTACGCTGGGAAGA 3'

Reverse – 5' TCCGTTCCATATGCCTAAGTG 3'

PCR mix of 20µl volume was prepared, which included:

10X reaction buffer (1X)	2 µl
dNTP mix (100 µM each)	1 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
Taq DNA polymerase	1 µl
Template DNA	1 µl
Nuclease free water	13 µl
Total volume	20 µl

cDNA samples of both healthy and infected plants were amplified by initial denaturation at 95 ° C for 3 min, followed by 25 cycles of denaturation at 95° C for 15 s, annealing at 60° C for 15 s and extension at 72° C for 45 s. Final extension was done at 72° C for 5 min. The steps from 2-4 were repeated for 25 cycles.

PCR product was analysed in agarose gel (1.5 per cent) in a horizontal gel electrophoresis unit. 2 µl of 100 bp ladder was loaded to one of the wells as marker.

Electrophoresis was carried out at 50 V until the loading dye reached three fourth of the gel and the gel was documented by using Gel Doc unit (Bio-Rad), using Quantity One Software.

3.5.6 Expression Analysis using Real-Time PCR

Real-Time PCR is a cost effective technique to clone and quantify miRNAs as well as other SmallRNAs. RT-qPCR is highly specific to mature miRNAs and not for pre-miRNAs. This analysis was carried out by using Applied Bio systems SYBR Green Master Mix available at Rajiv Gandhi Centre for Biotechnology, Poojappura, Thiruvananthapuram.

For expression analysis of miRNAs, relative quantification assay was done. In this assay, relative expression of miRNA in healthy sample was compared to expression of same miRNA in infected sample. RT-qPCR reactions were carried out by using FastStart Universal SYBR Green Master (ROX) (Chai *et al.*, 2015). A Reaction volume of 20µl was prepared, which include:

cDNA	2 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
SYBR Green Master mix (2X)	10 µl
Nuclease free water	6 µl
Total volume	20 µl

cDNA samples of miRNAs of both healthy and infected plants were amplified by initial denaturation for 5 min at 95° C, followed by 45 cycles of denaturation at 95° C for 5 s, annealing at 60° C for 10 s and extension at 72° C for 8 s.

cDNA samples of target genes of both healthy and infected plants were amplified by initial denaturation at 95 ° C for 2 min, followed by 40 cycles of denaturation at 95° C for 15 s, annealing at 55° C for 15 s and extension at 72° C for

30 s. Final extension was done at 72° C for 5 min. This was followed by Melt-curve analysis *i.e* fluorescence signals at 530 nm wavelength were monitored from 65° C to 95° C at the rate of 0.2° C per second to determine the product specificity.

Results were expressed as fold change (increase or decrease) in expression. A house keeping gene *i.e.*, β-Actin was used as reference gene for normalisation of RT-qPCR data. All the reactions were conducted in two replications along with No Template Control (NTC). Threshold cycle (Cq) values were determined for all the reactions and fold changes were calculated by using ΔΔCq method (Rao *et al.*, 2013). Threshold cycle (Cq) is the cycle at which the fluorescence value of a sample reaches an arbitrary threshold florescence. The difference between C_q values of control or infected sample and reference gene was considered as ΔC_q value.

$$\Delta C_q = C_q \text{ (Reference gene)} - C_q \text{ (Control/Infected)}$$

The difference between ΔC_q values of control and infected samples were considered as $\Delta\Delta C_q$ value. The fold change in expression of miRNAs is calculated by using $2^{\wedge(-\Delta\Delta C_q)}$.

$$\Delta\Delta C_q = \Delta C_q \text{ (Infected Sample)} - \Delta C_q \text{ (Healthy Sample)}$$

RESULTS

4. RESULTS

The study entitled "Computational prediction of miRNAs in banana (*Musa* spp.) and evaluation of their role in virus infection" was conducted at the Department of Plant Biotechnology, College of Agriculture, Vellayani, Thiruvananthapuram and Central Tuber Crops Research Institute (ICAR-CTCRI), Sreekariyam, Thiruvananthapuram during 2016–2018. The results of the study are presented in this chapter.

4.1 PREDICTION OF NOVEL MIRNAS IN BANANA

4.1.1 Retrieval of Banana Chromosome Sequences

To predict pre-miRNAs in banana (*Musa acuminata*), all the gene coding sequences of *Musa acuminata* DH Pahang v1 (released in July, 2012) were downloaded from Banana Genome Hub. A total of 260 FASTA files (472.9 Mb) containing each scaffold sequence were created for 11 chromosomes. Each FASTA file containing scaffold sequence was analysed with NovoMIR.

4.1.2 Prediction of pre-miRNAs

NovoMIR splits the scaffold sequence of all the chromosomes into small reads (Figure 13) and predicted pre-miRNAs based on their secondary structure using RNAFOLD and RNASHAPES package.

The predicted secondary structures of pre-miRNAs were represented in dot-bracket notation (Figure 14), where bracket indicates base pairing between sequences and dot indicates mispairing or as-symmetric part of loop.

By sequencing the total 260 scaffolds of 11 chromosomes by NovoMIR, 85 pre-miRNA sequences were obtained. Out of 85 pre-miRNAs, maximum number of pre-miRNAs (22) were obtained from chromosome 9 (Figure 15). Ten pre-miRNAs were obtained from chromosome 2 and nine pre-miRNAs each were obtained from chromosome 1 and 10. All other chromosomes contained less than

eight pre-miRNAs. List of all predicted pre-miRNAs were summarized in Appendix 1.

Length of pre-miRNAs ranged from 65 nt to 347 nt (Figure 16). Maximum number of pre-miRNAs (25%) ranged from 140-180 nt. In pre-miRNA sequences, G+C % ranged from 24–77 % (Figure 17) and A+U % ranged from 23–76 % (Figure 18). GC content of pre-miRNAs was much lower than AU content.

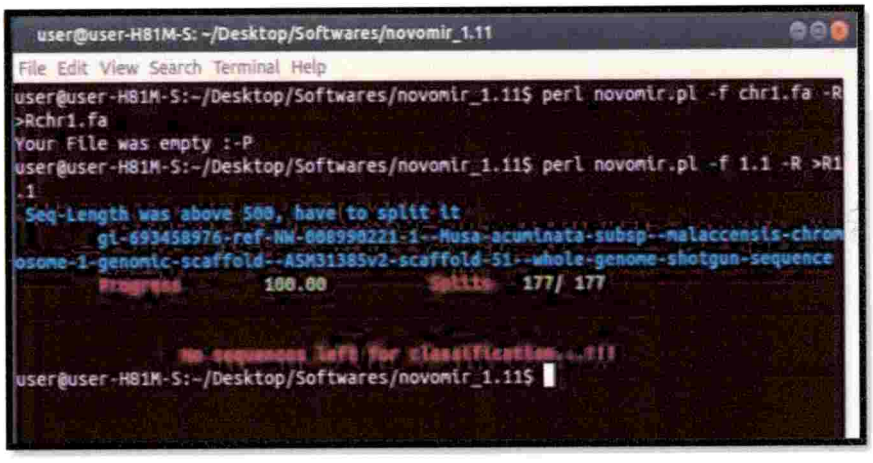


Fig. 13. Prediction using NovoMIR

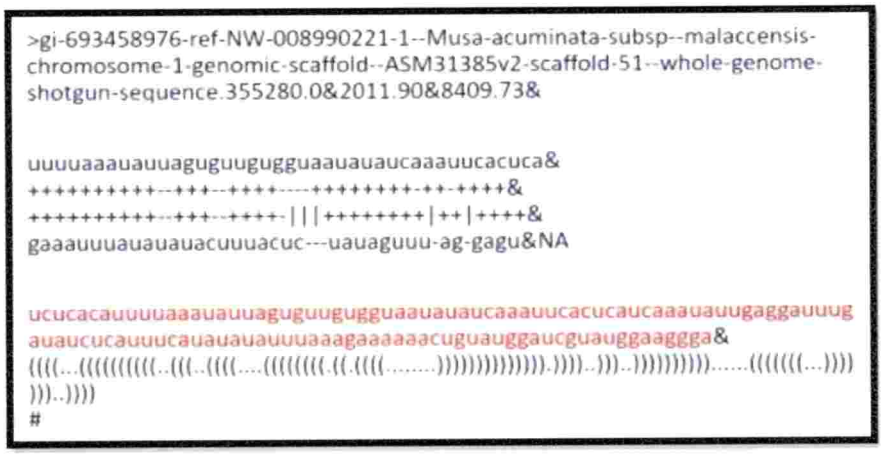


Fig. 14. Result obtained from NovoMIR

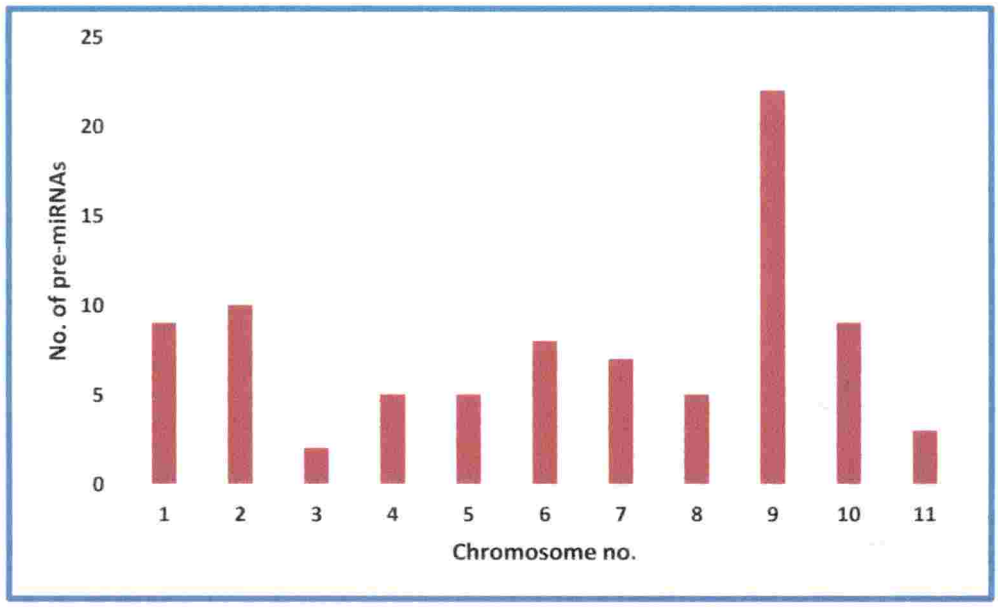


Fig. 15. Distribution of pre-miRNAs per chromosome

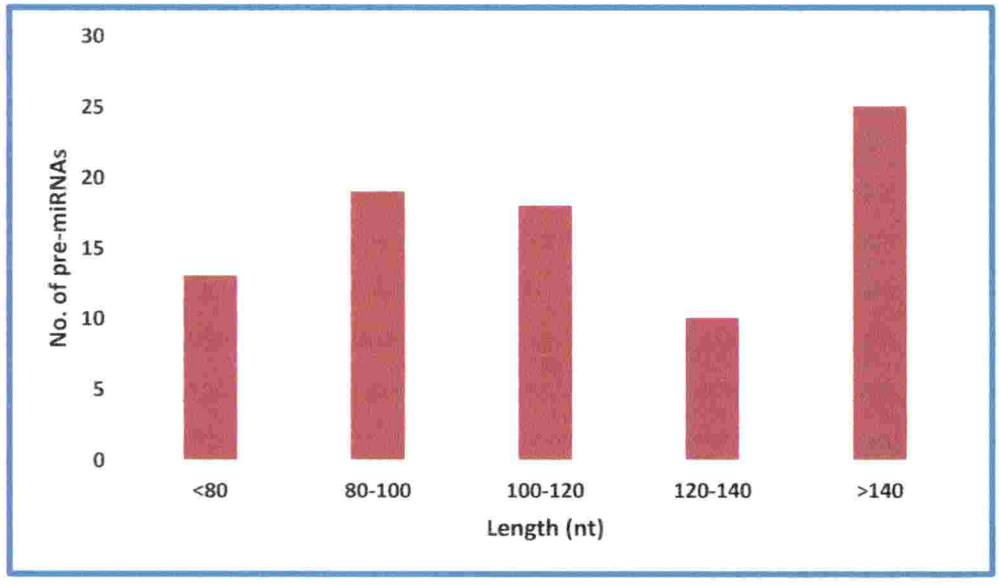


Fig. 16. Length distribution of pre-miRNAs

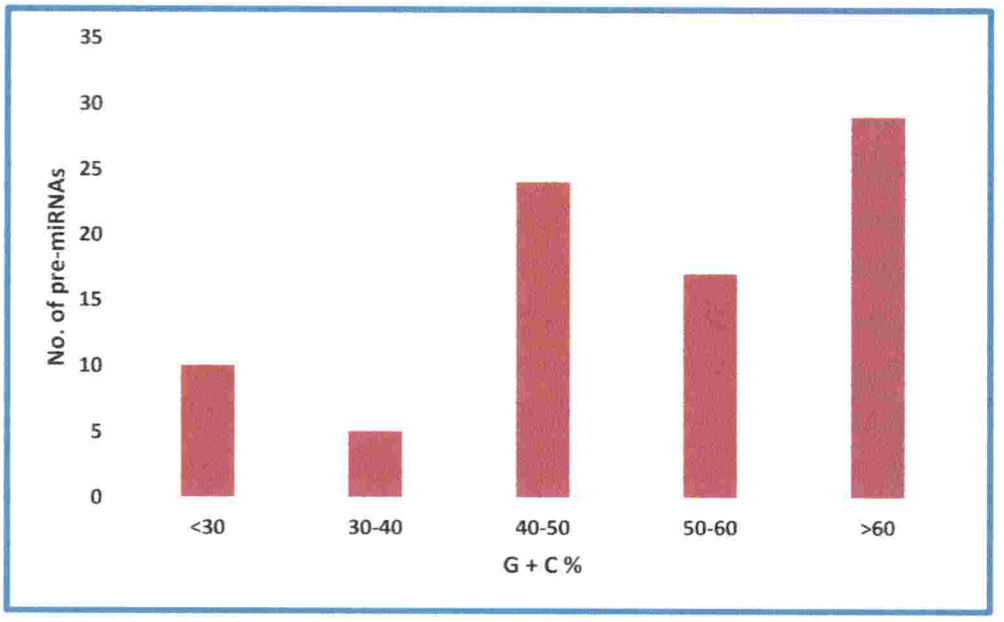


Fig. 17. Distribution of (G +C) % in pre-miRNAs

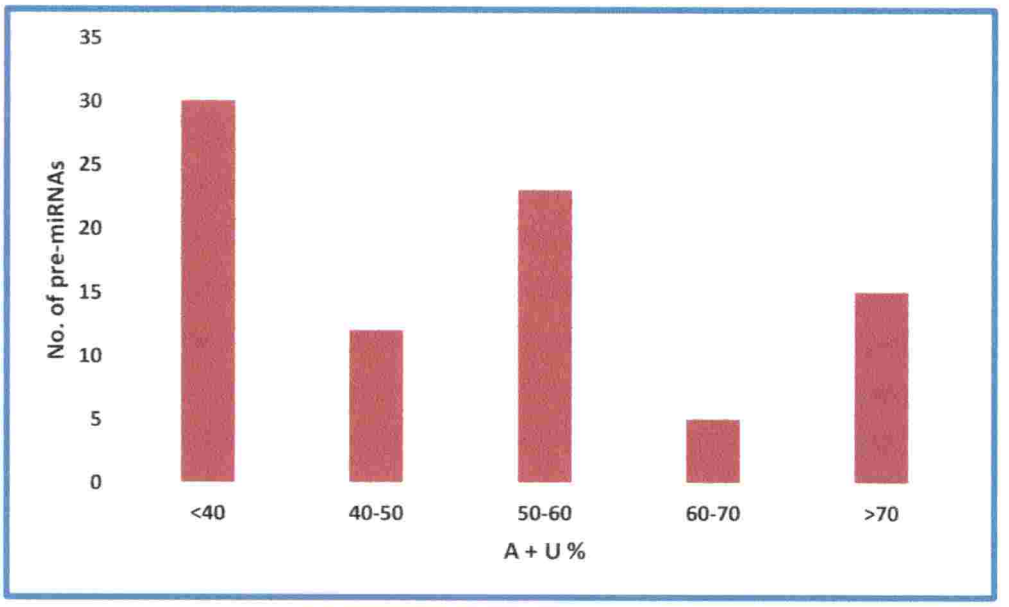


Fig. 18. Distribution of (A+U) % in pre-miRNAs

4.1.3 Secondary Structure Prediction of pre-miRNAs

Secondary or hairpin loop structure of pre-miRNAs were determined by using RNAfold web server. The average minimal folding free energy (MFE) of pre-miRNAs ranged from -20 Kcal/mol to -194.4 Kcal/mol (Figure 19). The adjusted minimal folding free energy (AMFE) of pre-miRNAs ranged from -18.9 Kcal/mol to -85.4 Kcal/mol (Figure 20), which was lower when compared to MFE values of pre-miRNAs. Most of the pre-miRNAs showed AMFE values within -50 Kcal/mol to -90 Kcal/mol. MFEI of pre-miRNAs ranged from 0.85 to 1.82 (Figure 21). This suggests that all the pre-miRNAs are more likely to be miRNAs than any other type of RNA molecules that form stem-loop structure. List of all pre-miRNAs along with their MFE are summarised in Appendix 1.

4.1.4 Identification of Mature miRNAs

All the pre-miRNAs obtained from NovoMIR were analyzed by BLAST analysis against mature miRNA sequences of various species annotated in miRBase. Based on the BLAST results, 111 sequences were considered as potential miRNAs as they are about 14 – 25 nt in length. Out of 111 potential miRNAs, 52 miRNAs having a length of 18-22 nt, were considered as mature miRNAs. Most of the predicted mature miRNAs (46%) were about 20 nt in length followed by 18 nt (21%), 22 nt (23%), 21 nt (6%) and 19 nt (4%) (Figure 22).

4.1.5 Properties of Mature miRNAs

In order to find out the properties of mature miRNAs, an online tool called "Oligo Calc" was used. G+C % of mature miRNAs ranged from 0-89 % (Figure 23), A+U % of mature miRNAs ranged from 11-100 % (Figure 24), Tm of mature miRNAs ranged from 31.3° C to 66.2° C (Figure 25) and molecular weight of mature miRNAs ranged from 5353.4 gm/mole to - 6724.3 gm/mole (Figure 26). List of all mature miRNAs along with their properties are summarized in Appendix 2.

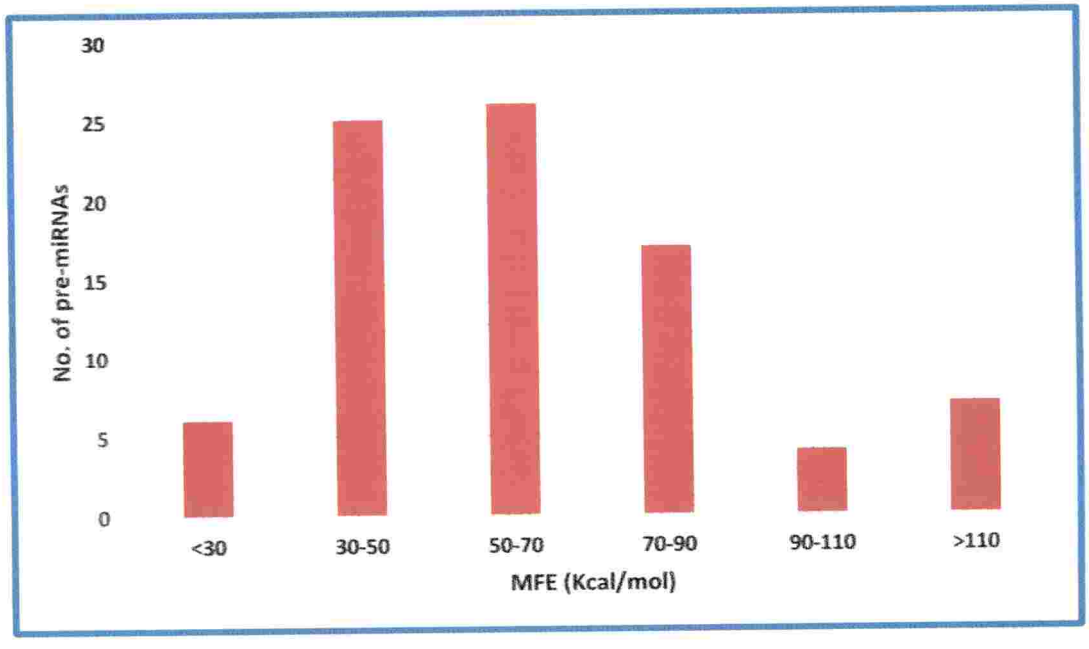


Fig. 19. Distribution of MFE of pre-miRNAs

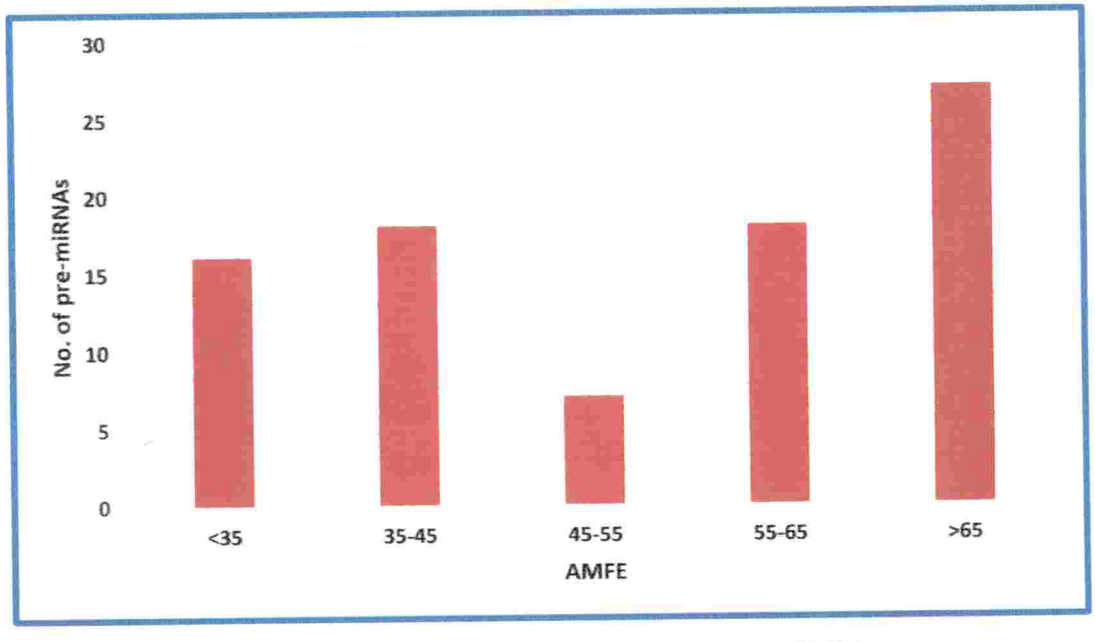


Fig. 20. Distribution of AMFE of pre-miRNAs

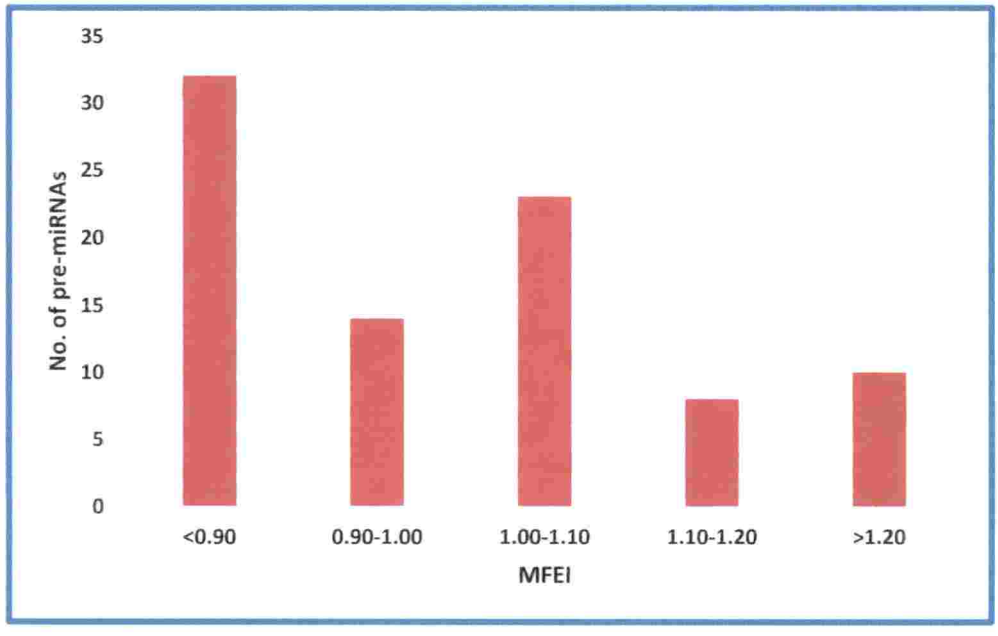


Fig. 21. Distribution of MFEI of pre-miRNAs

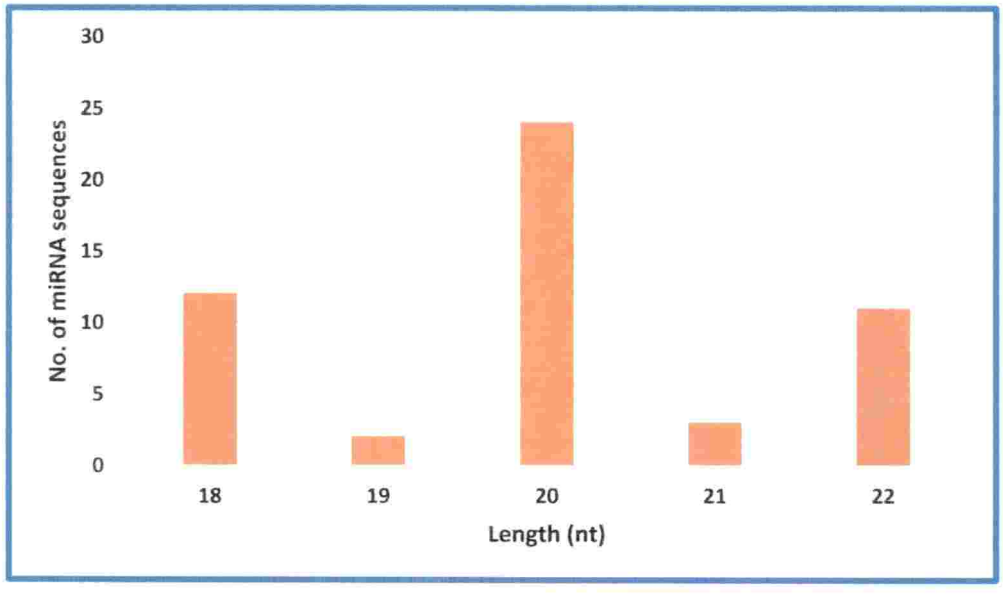


Fig. 22. Distribution of length of mature miRNAs

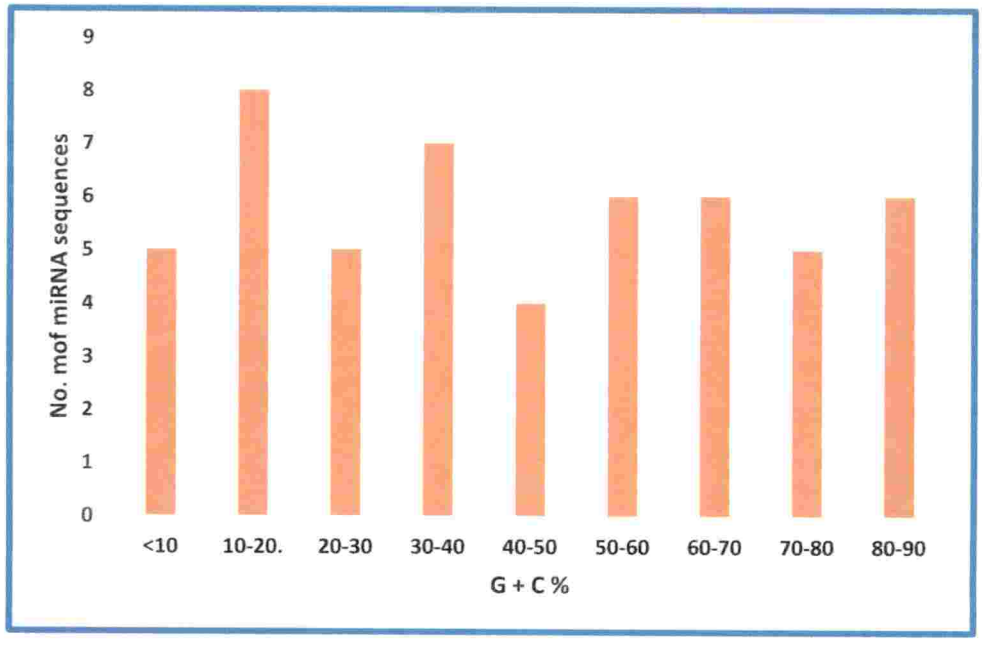


Fig. 23. Distribution of G + C % in mature miRNAs

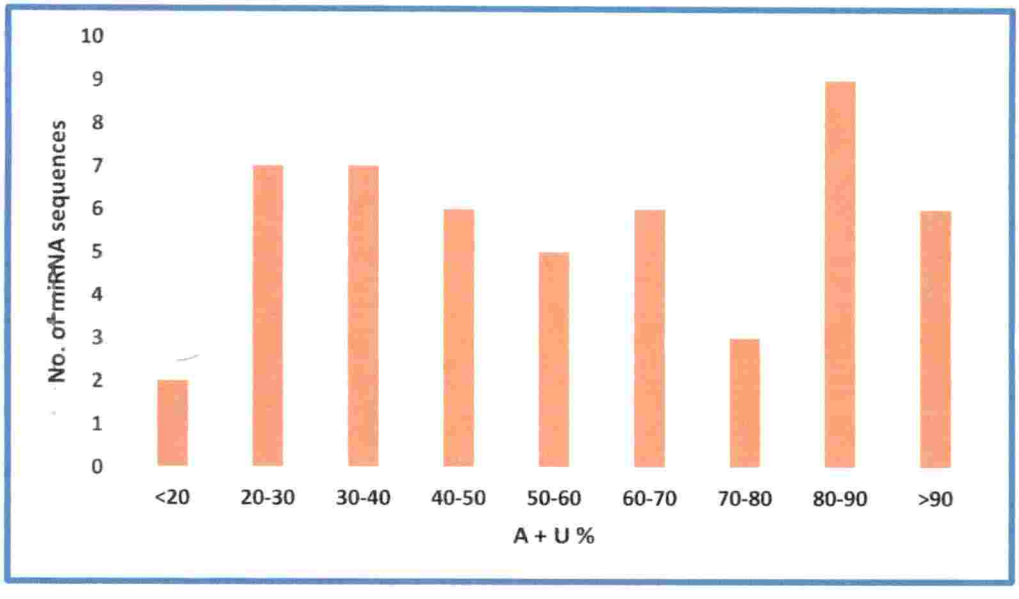


Fig. 24. Distribution of A + U % in mature miRNAs

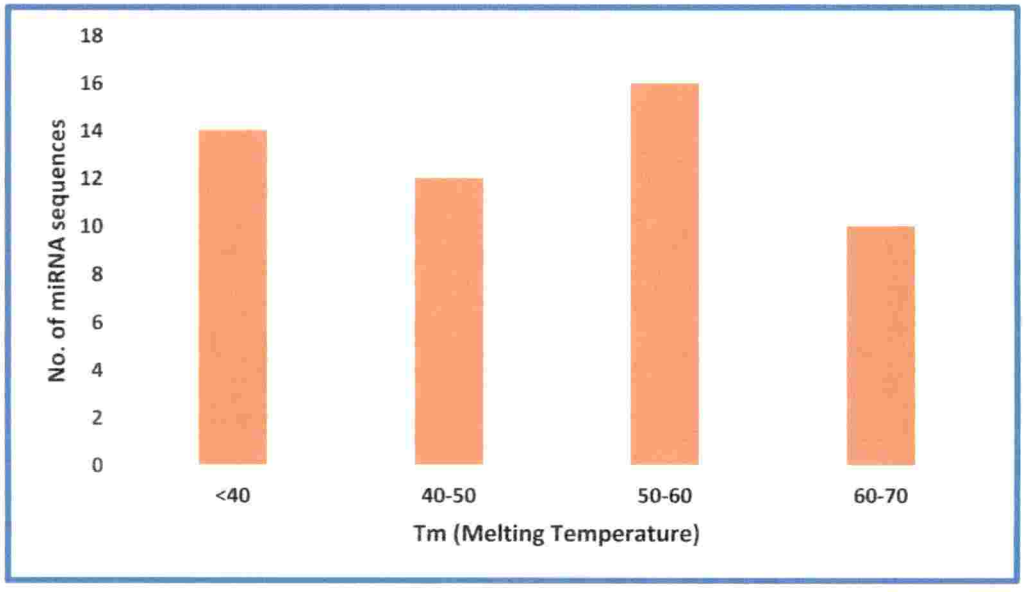


Fig. 25. Distribution of T_m of mature miRNAs

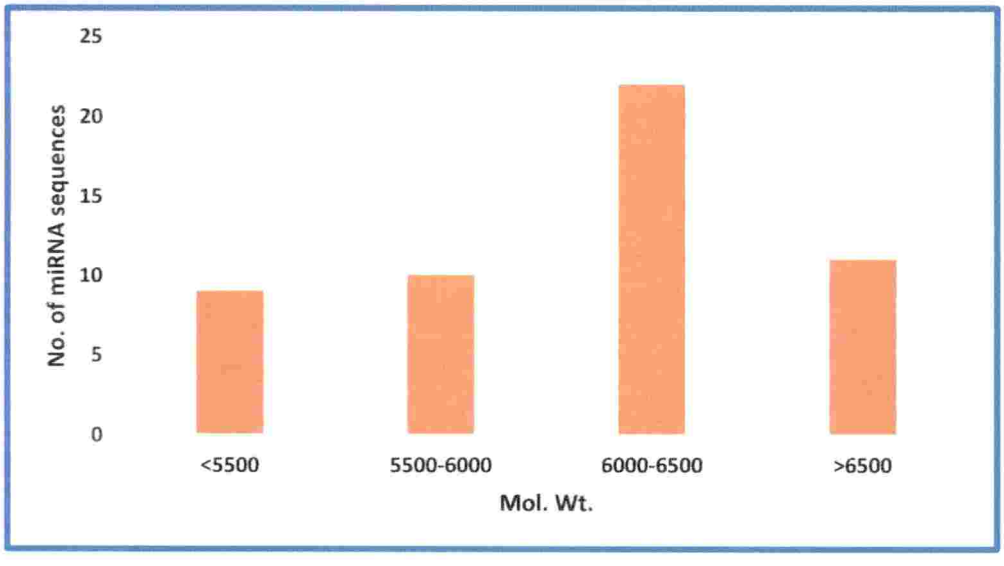


Fig. 26. Distribution of Mol. wt. of mature miRNAs

4.2 IDENTIFICATION OF THE miRNA TARGETS USING TARGET PREDICTION SOFTWARE

Potential targets of the predicted mature miRNAs were identified by using an online target server psRNATarget (plant small RNA target analysis server).

Out of 52 mature miRNAs, only 40 miRNAs have found mRNA targets in banana genome. Based on scoring schema technology (Figure 27), each miRNA was found with more than one target. A total of 124 protein coding sequences were found as miRNA targets in banana genome. List of all miRNA targets are summarised in Appendix 3.

4.3 FUNCTIONAL ANNOTATION OF THE IDENTIFIED miRNA TARGETS

For functional annotation of identified miRNA targets, Blast2Go software was used.

Results of BLAST analysis were saved with an extension of *.b2g*. BLAST result statistics were represented in the form of charts based on Top-Hit species distribution, Species distribution and Sequence similarity distribution. Top-Hit Species Distribution chart (Figure 28) showed species distribution with top-blast hits. Maximum number of blast-hits were obtained for *Musa acuminata* ssp. *Malaccensis,* followed by *Dendrobium catenatum, Musa acuminata, Populus trichocarpa, Nelumbo nucifera etc.* Species Distribution chart (Figure 29) gave a list of different species to which most sequences were aligned during the BLAST analysis. By blast analysis, maximum number of sequences aligned for *Musa acuminata* subsp. *malaccensis*, followed by *Elaeis guineensis, Phoenix dactylifera, Ananas comosus, Asparagus officianalis etc.* Sequence Similarity Distribution chart (Figure 30) displayed the distribution of all calculated sequence similarities (percentages), which helps to adjust the annotation score in the annotation step *i.e* number of sequences per annotation score.

After BLAST analysis, mapping of target gene sequences was done. Those sequences which were mapped turned green in colour. After mapping, GO annotation was done to understand miRNA function. Sequences that were annotated turned blue.

Annotation results of miRNA targets were exported in the form of charts. Out of 124 miRNA targets, 118 target genes were found to be involved in various molecular functions like transferase activity, ATP binding, Nucleic acid binding, protein binding, cofactor binding, transporter activity, oxidoreductase activity and metal ion binding (Figure 31) and 84 target genes were found to have role in various biological processes like anatomical structure development, lipid metabolism, drug metabolic process, cellular response to chemical stress, gene expression, RNA metabolic process, response to stress, transmembrane transport, carbohydrate biosynthetic process, oxidation-reduction process, protein modification process, phosphorylation etc. (Figure 32). List of all annotated miRNA targets along with their GO terms and molecular function, biological process are summarized in Appendix 4.

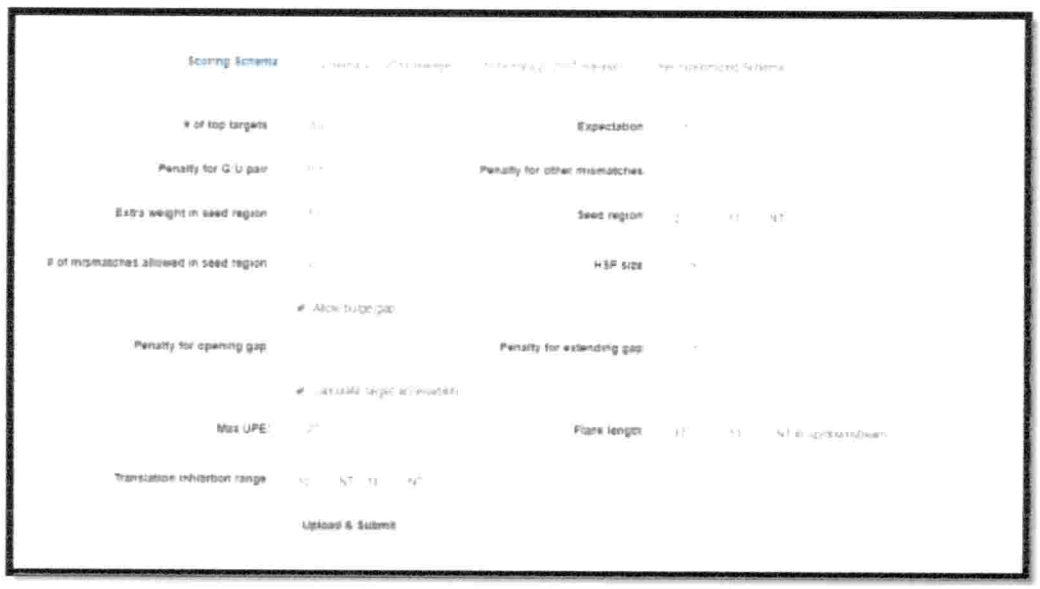


Fig. 27. Scoring schema technology of psRNATarget server

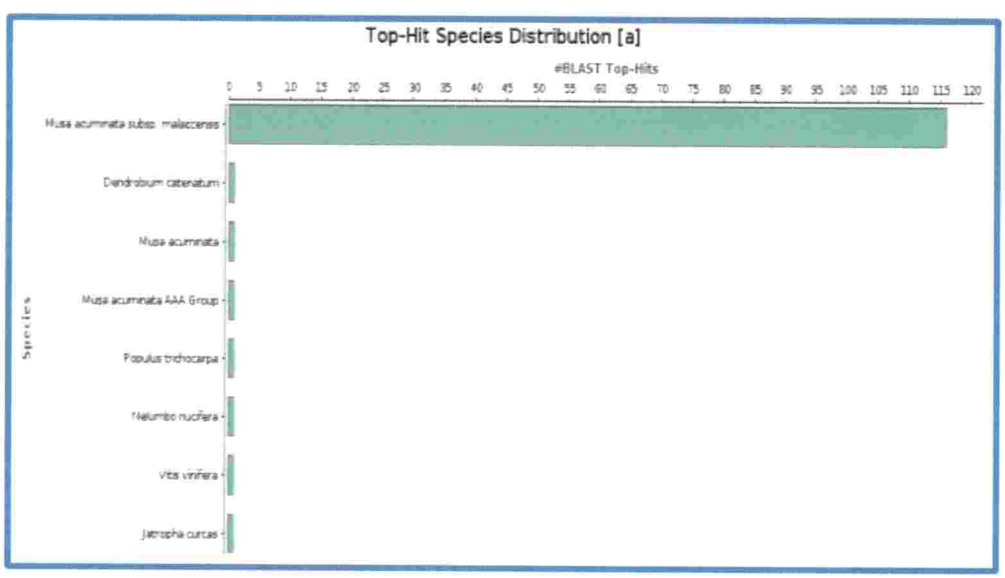


Fig. 28. Top-Hit species distribution chart

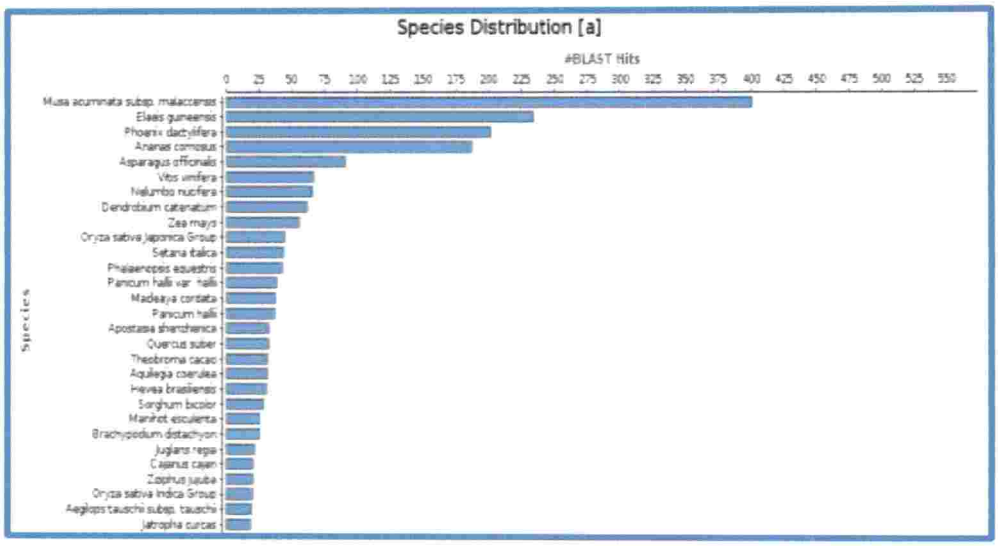


Fig. 29. Species distribution chart

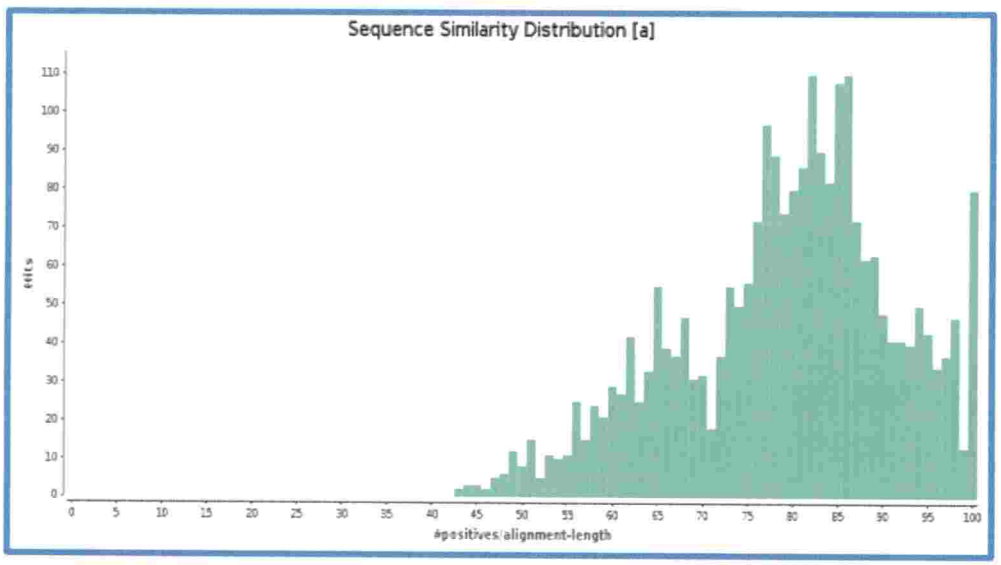


Fig. 30. Sequence similarity distribution chart

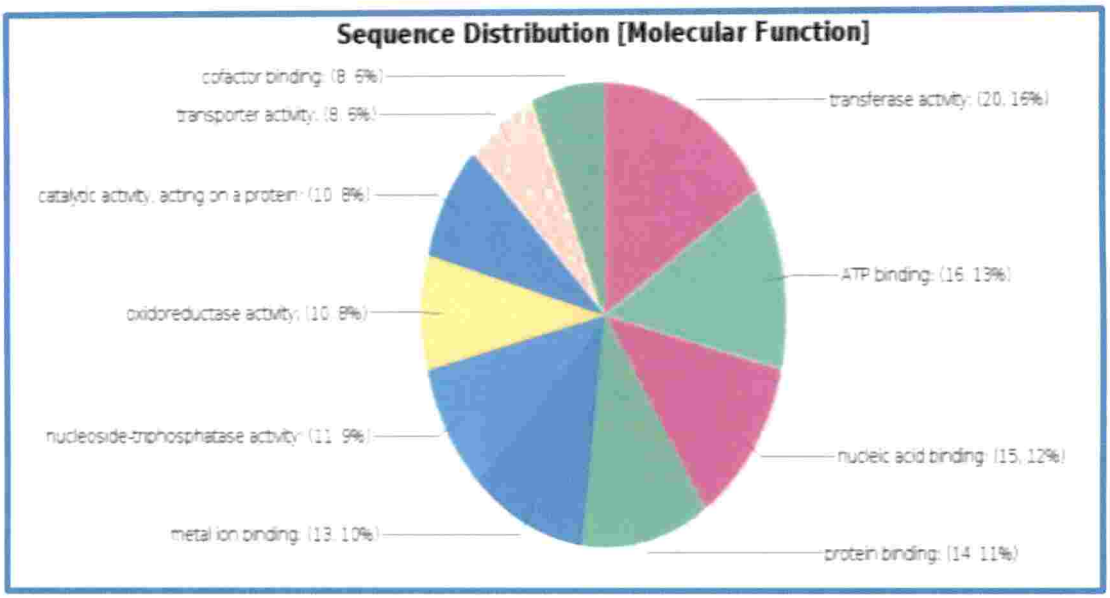


Fig. 31. Molecular functions of targets identified

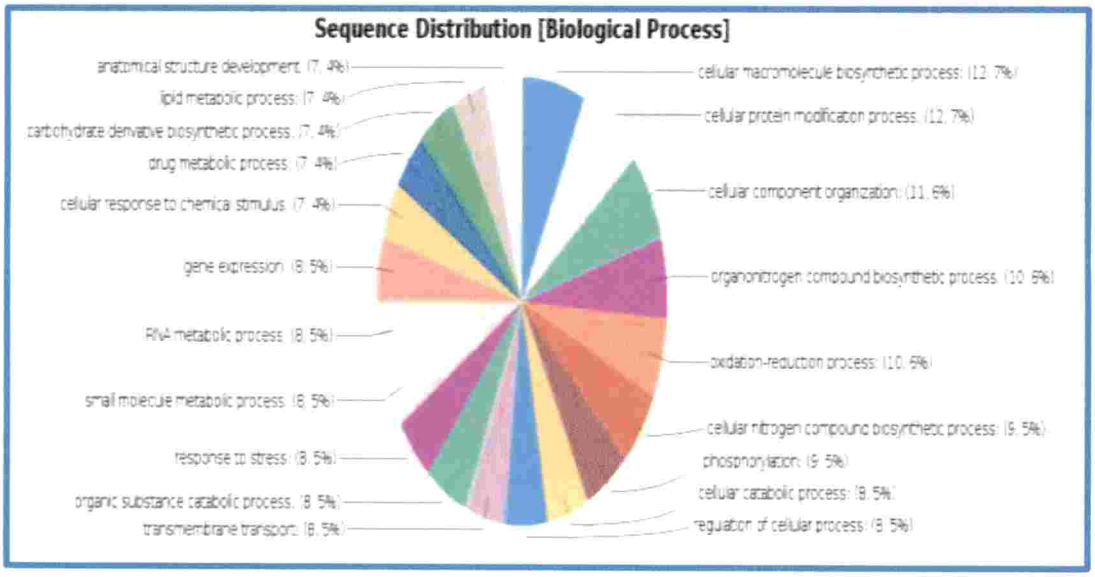


Fig. 32. Biological process of targets identified

4.4 EXPERIMENTAL VALIDATION OF THE NOVEL PREDICTED miRNAS

After functional annotation of miRNA targets, stem-loop primers were designed for miRNAs, which were selected for expression analysis. miRNA specific forward primer and reverse primer were also designed. List of stem-loop primers for selected miRNAs are summarized in Table 3.

Primers were also designed to target genes by using an online software Primer3Plus. Both forward and reverse primers of about 20-22 nt in length were designed for each target gene using default parameters. The result of Primer3Plus gives the list of primers along with their characteristics like length, G+C %, Tm of both forward and reverse primers. The primer pair, which was scored best, was selected and synthesized. Primer sequences designed were blasted against banana genome and virus genome by using NCBI Primer BLAST to find out the primer specificity. All the primer sequences showed complementarity to banana genome only. Best primer sequences that were complementary to banana genome were synthesized and used for validation experiments. List of primers for target genes are summarized in Table 4.

Table 3. Stem-loop primers designed for miRNAs

(S: Stem-loop, F: Forward, UR: Universal Reverse)

miRNA	Primers	Sequence (5^{l}-3^{l})	Length (nt)	Tm
miR-6928-5p	GTCGTATCCAGTGCAGGGTCCGAGG TATTCGCACTGGATACGACTGCAGT	S	50	61.1
	GCGGGCGGGGATTTTCAAGT	F	20	
miR-3900-5p	GTCGTATCCAGTGCAGGGTCCGAGG TATTCGCACTGGATACGACGGCGGT	S	50	62.3
	GCGGCGCAAGGAGGTGTC	F	18	
miR-3900-5p	GTCGTATCCAGTGCAGGGTCCGAGG TATTCGCACTGGATACGACGGTGGT	S	50	64.4
	GCGGGCGCAAGGAGGTGTC	F	19	
miR-9112	GTCGTATCCAGTGCAGGGTCCGAGG TATTCGCACTGGATACGACCATGAT	S	50	63.7
	CGGCGGGGACCAGAGAGC	F	18	
miR-5417	GTCGTATCCAGTGCAGGGTCCGAGG TATTCGCACTGGATACGACCATCCC	S	50	60.9
	CGGGCGGAGGGGAGAAATG	F	19	
	CCAGTGCAGGGTCCGAGGTA	UR	20	

Table 4. Primers designed for target genes

Target description	Sequence (5^{l}-3^{l})	Primers
FAD dependent oxidoreductase	F	CCTAGTGCCACAGCAGAAAAG
	R	GACCCAAGTAACCAGTCACCA
F-box/kelch-repeat protein	F	ATGAGCATCACTCTGGTGGAC
	R	GCCTGAAGAACTTGACAGTGG
Armadillo repeat-containing kinesin-like protein 1	F	ATGGCCACTAGCTCTGCTACA
	R	GCGAGGTTGCGAGTATCTTCT
Glyceraldehyde-3-phosphate dehydrogenase	F	GGTATCAGGAACCCTGAGGAG
	R	GAGGAGCAAGACAGTTGGTTG
cytochrome c oxidase subunit 5B protein	F	GTTCTTCTCCTCCGTTTCAGG
	R	GTTCATCCTCACCTTCACCAC

4.5 EXPRESSION ANALYSIS OF miRNAS AND THEIR TARGET GENES DURING VIRUS INFECTION

4.5.1 Transfer of Banana Bract Mosaic Virus (*BBrMV*) by Aphids

As *BBrMV* belongs to non-persistent mode of transmission, virus was acquired for few minutes by the aphids. Virus was transmitted from infected suckers to four raised tissue culture plants of variety Nendran (Plate 1). After acquisition feeding for 30 min on infected suckers (Plate 2), aphids were transferred to tissue culture plants for infection feeding for 14 h (Plate 3). After infecting the plants twice a day for 7 days, virus infection was checked in banana tissue culture plants.

4.5.2 RNA Isolation

RNA was isolated from healthy and infected banana plants by using CTAB RNA extraction method, standardized by Tsai and Harding, (2003). RNA separated by electrophoresis on 2% agarose gel showed rRNA bands with little genomic DNA contamination (Plate 4). Quantity of RNA was determined by using Nanodrop at 260 nm and 280 nm. Due to presence of genomic DNA contamination, RNA samples were treated with DNase enzyme by using Invitrogen DNase kit (Plate 5).

4.5.3 cDNA Synthesis

cDNA synthesis was done by using Verso cDNA synthesis kit and it was confirmed by doing PCR amplification with house keeping gene specific primers *i.e.*, β-Actin. PCR product obtained showed an amplicon of expected size (700 bp) in both healthy and infected samples (Plate 6).

4.5.4 Confirmation of *BBrMV* Infection

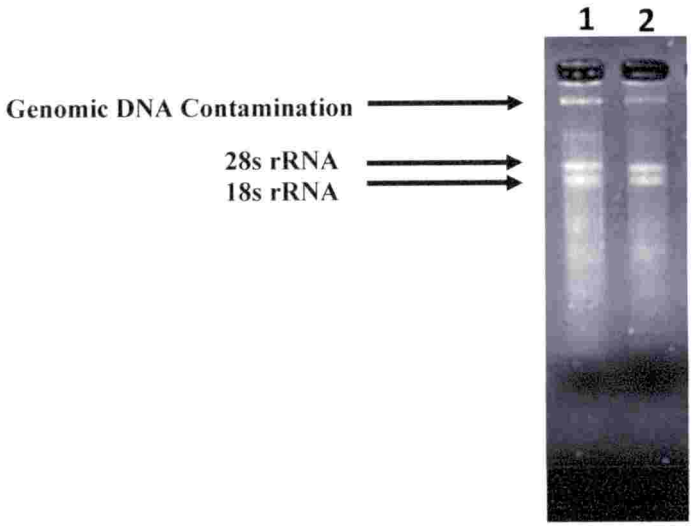
PCR analysis with replicase gene specific primers of *BBrMV* showed amplicon of expected size (500 bp) in only infected sample (Plate 7) indicating the presence of virus infection.

Plate 1. Tissue culture plants of variety Nendran

a. Aphids released for acquisition feeding **b. *BBrMV* infected sucker**

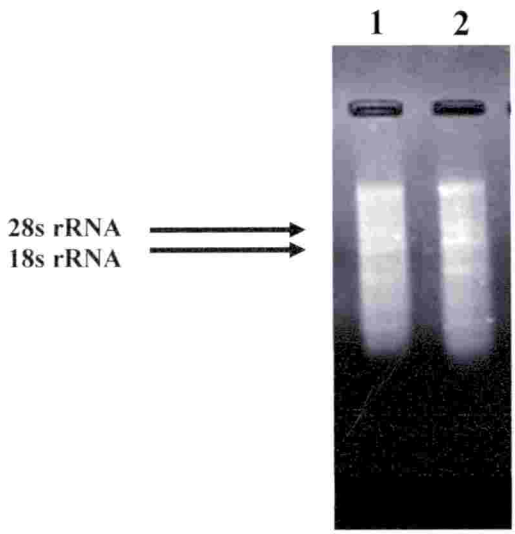
Plate 2. *BBrMV* infection using aphids

Plate 3. Banana plants inoculated with infectious aphids



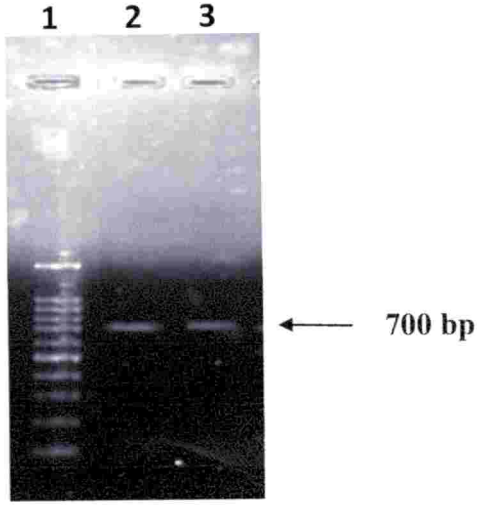
Lane 1: Healthy plant
Lane 2: Infected plant

Plate 4. Gel image of RNA isolated from healthy (1) and *BBrMV* infected (2) banana leaf tissue



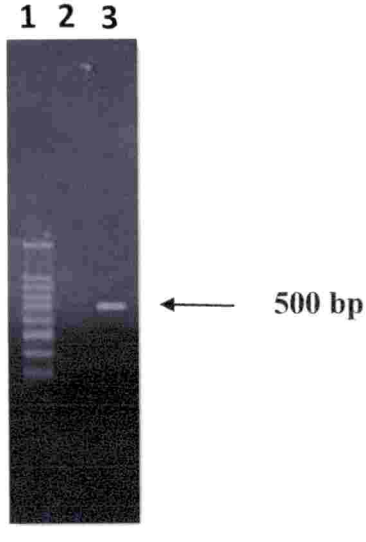
Lane 1: Healthy plant
Lane 2: Infected plant

Plate 5. Gel image of RNA samples after DNase treatment



Lane 1: 100bp DNA ladder
Lane 2: Healthy plant
Lane 3: Infected plant

Plate 6. Gel image of cDNA amplified with β-Actin gene primers



Lane 1: 100bp DNA ladder
Lane 2: Healthy plant
Lane 3: Infected plant

Plate 7. Gel image of cDNA amplified with replicase gene primers

4.5.5 Expression Analysis of miRNAs using Real-Time PCR

The expression of miRNAs and their targets were analyzed in both healthy and *BBrMV* infected samples by using RT-qPCR. The results showed the expression of all the five predicted miRNAs and their target genes in both healthy and *BBrMV* infected samples. Threshold cycle values (C_q) of miRNAs (Table 5) and their target genes (Table 6) were calculated by using Bio-Rad CFX Manager Software. All the five miRNAs exhibited different levels of expression (Figure 33). Out of them, miR-3900-5p, miR-9112 and miR-5417 showed higher expression levels compared to uninfected healthy plants with 3.4, 7.6 and 9.4 fold increase in expression respectively. On the other hand, miR-6928-5p showed 0.4 fold decrease in expression and there was no significant change in expression level of miR-3900-5p.

The expression patterns of target genes were positively correlated with that of miRNAs (Figure 34). Among them, glyceraldehyde-3-phosphate dehydrogenase, F-box/kelch-repeat protein and cytochrome c oxidase subunit 5B gene targeted by miR-5417, miR-3900-5p and miR-3900-5p respectively showed higher expression levels compared to uninfected healthy plants with 24, 1.8 and 1.2 fold increase in expression under *BBrMV* infection. FAD dependent oxidoreductase targeted by miR-6928-5p was downregulated with 0.4 fold decrease in expression. The expression level of Armadillo repeat-containing kinesin-like protein 1 remained unchanged during *BBrMV* infection. Figure 35 shows the Melt-curve analysis of the miRNAs.

The results showed that four of the selected miRNAs were up-regulated during *BBrMV* infection and their target genes were also highly expressed. Differential expression of selected miRNAs and their targets indicates their involvement in biological stress conditions in banana.

Table 5. Mean C_q values of miRNAs in qPCR amplification

miRNA	C_q Value	
	Control	Infected
miR-6928-5p	30.65652	28.04974
miR-3900-5p	36.11503	29.82070
miR-3900-5p	35.05632	27.83598
miR-9112	36.23363	30.80227
miR-5417	29.20857	23.97655

Table 6. Mean C_q values of target genes in qPCR amplification

Target gene	C_q Value	
	Control	Infected
FAD dependent oxidoreductase	37.07	36.97
F-box/kelch-repeat protein	30.48	28.44
Armadillo repeat-containing kinesin-like protein 1	36.98	35.84
cytochrome c oxidase subunit 5B protein	34.94	33.47
Glyceraldehyde-3-phosphate dehydrogenase	35.66	29.91

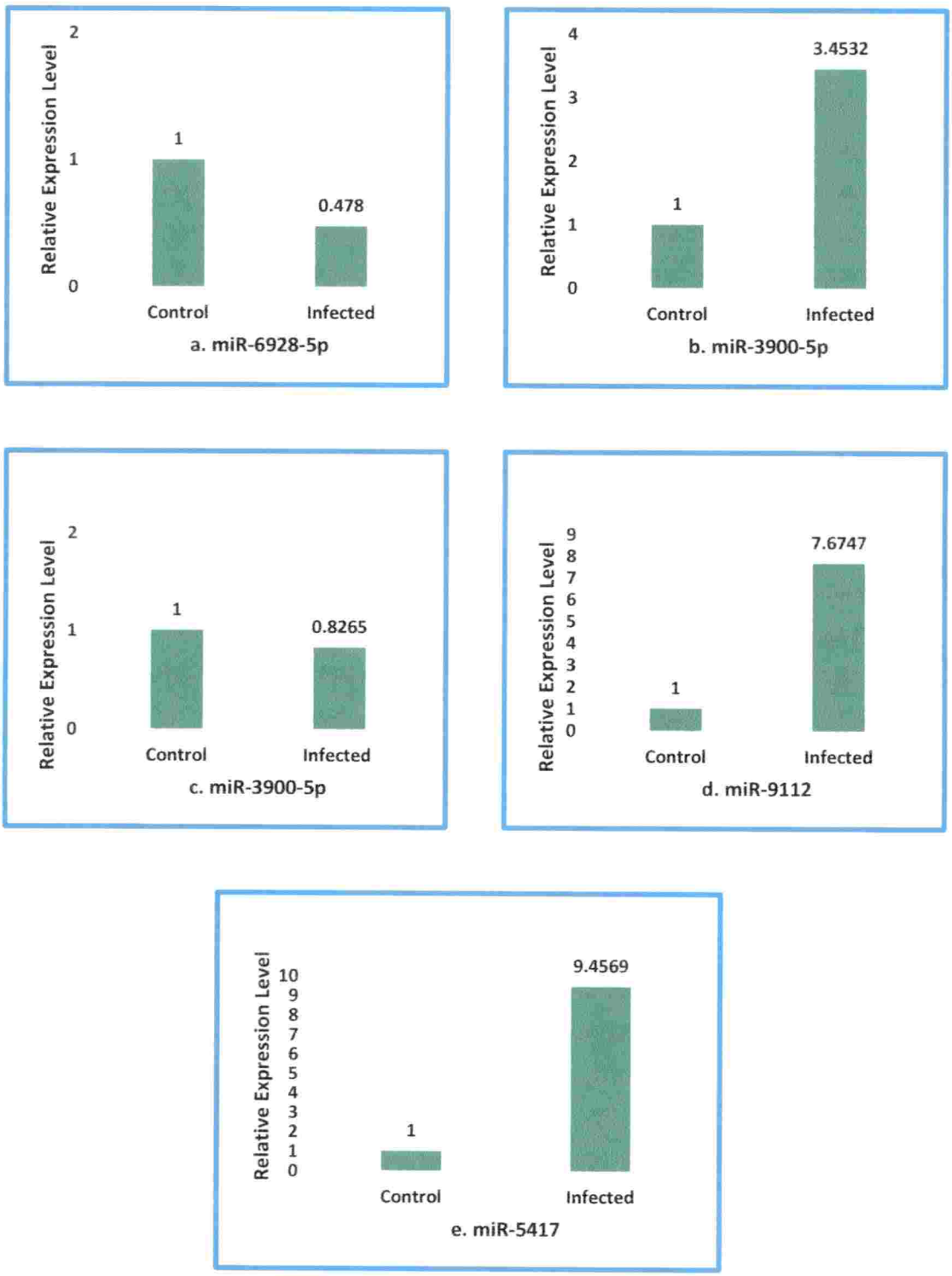
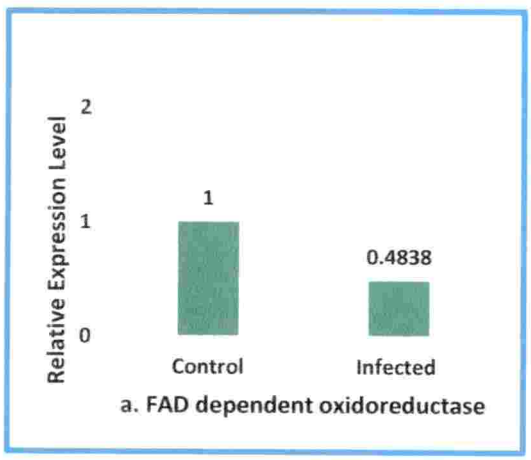
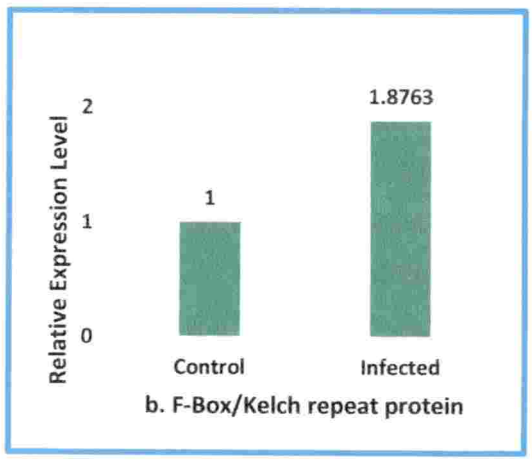
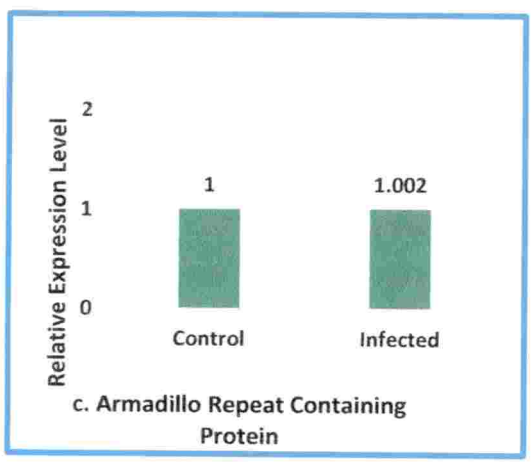
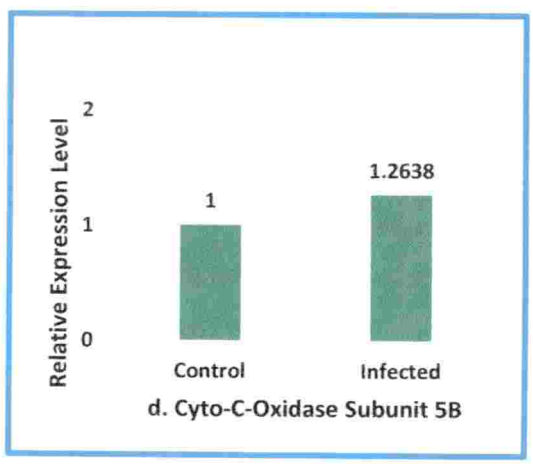


Fig. 33 (a-e). Relative expression levels of miRNAs in banana









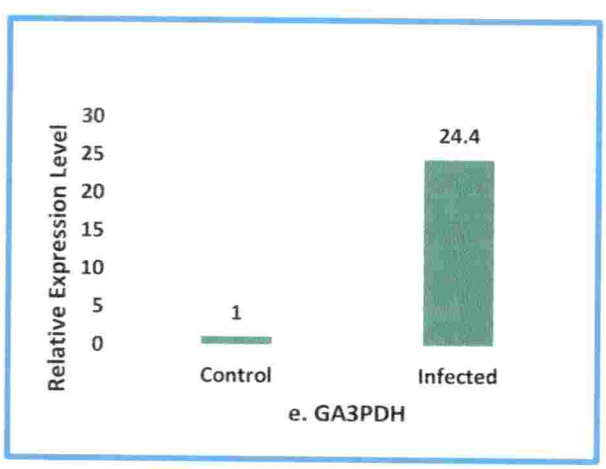
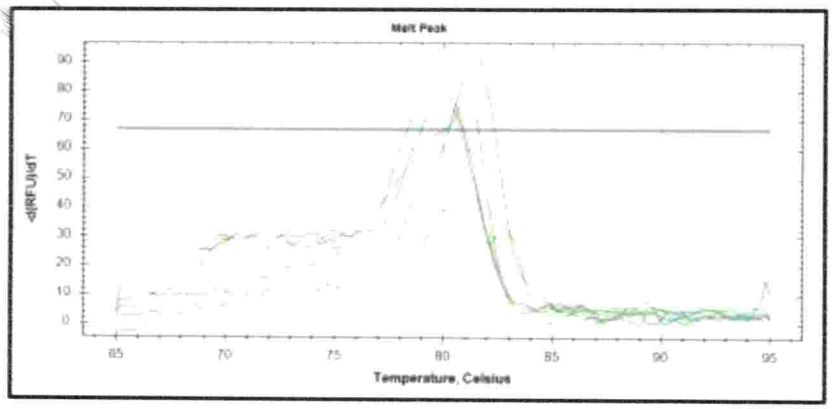
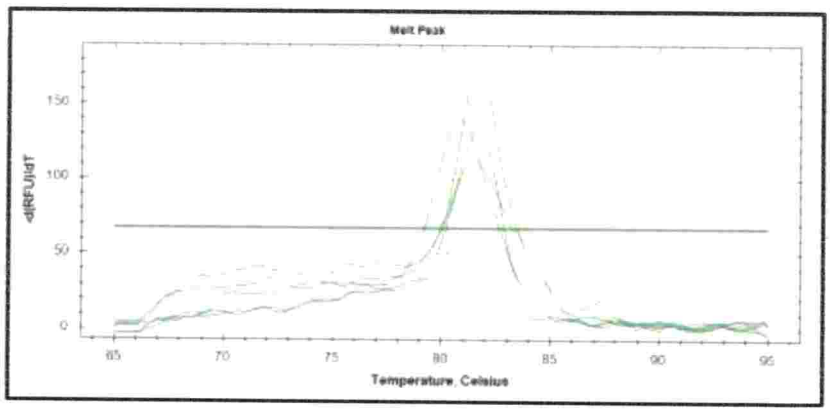


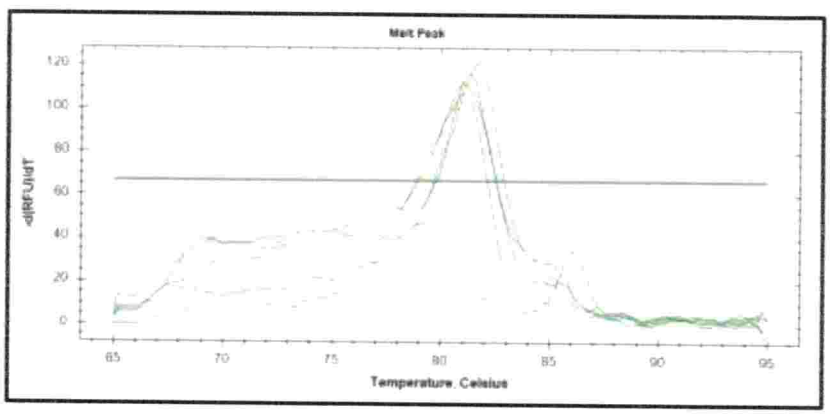
Fig. 34 (a-e). Relative expression levels of target genes in banana



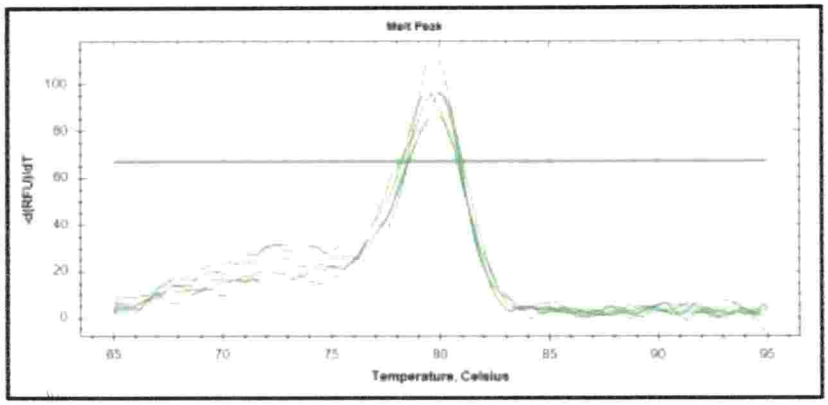
a.miR-6928-5p



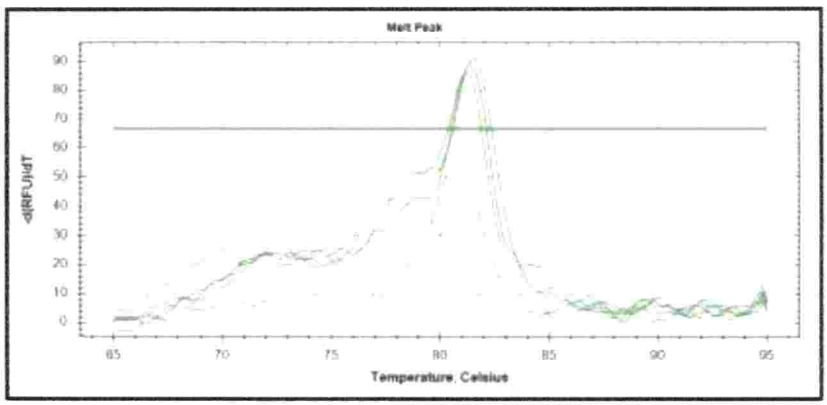
b. miR-3900-5p



c.miR-3900-5p



d. miR-9112



e. miR-5417

Fig. 35 (a-e). Melt curve analysis of miRNAs

DISCUSSION

5. DISCUSSION

Banana is the world's fourth most important food crop in terms of gross value production. It is grown in a wide range of environments and production systems. It provides nutritious staple food and is a significant source of revenue all year round. Among various biotic stresses, plant viruses usually impose serious threats to banana cultivation, leading to serious losses of yield and quality. Since most of the edible bananas are triploids and natural resistance is not available, traditional breeding for virus resistance is difficult. Several non-conventional methods like genetic engineering techniques based on coat protein mediated resistance, movement protein mediated resistance and resistance genes (R genes) have been applied for developing virus resistance. The discovery of RNA interference (RNAi) mechanism in plants during the past decade has led to the development of small RNA based technologies for imparting resistance towards biotic stress in plants. RNA mediated virus resistance mainly occurs *via* sRNAs, which include siRNAs and miRNAs. They act as regulators of gene expression, mostly by PTGS. In recent years, role of miRNAs has been established in various stress responses of plants.

miRNAs are an extensive class of endogenous small non-coding single stranded RNAs that are found in almost all eukaryotes. As plants do not possess an antibody-based immune system analogous to that of animals, miRNA mediated resistance is considered as the newly discovered strategy to suppress plant viruses (Waterhouse *et al.*, 2001). miRNAs are used as antiviral agents for repressing the function of viral suppressors in the process of gene silencing and modify the plant physiological properties to enhance their anti-virus capacity and developing loss-of-function transgenic plants.

For miRNA mediated resistance, it is important to identify miRNAs involved in defense response or resistance mechanism in the host. Many approaches have been developed for the identification of miRNAs. Although various high-throughput, time consuming and expensive approaches like direct

cloning and forward genetics are employed to predict miRNAs, novel bioinformatic tools have shown to facilitate efficient and cost effective prediction of miRNAs in plants (Panda *et al.*, 2014). A number of bioinformatic tools have been developed to detect miRNAs in plants. One of the major issues in the prediction of miRNAs with bioinformatic tools is selection of right tool for in-silico prediction of novel microRNAs in a given set of sequences.

The present study envisaged identification of miRNAs and their targets in banana genome and to validate and analyze their expression during virus infection. A bioinformatics tool namely NovoMIR is used in this study to predict miRNAs in banana.

5.1 PREDICTION OF NOVEL miRNAS IN BANANA

Homology based search or comparative genomics approach are widely used for miRNA identification based on sequence conservation among species. Using these approaches, miRNAs have been identified in various crops like soybean, wheat, cotton, potato, apple, switch grass and citrus. Advancement in the development of new computational tools make prediction of miRNAs easy and with high degree of accuracy. Many of the tools rely on the special features of miRNAs. Some of the tools are based on the phylogenetic conservation of pre-miRNA structure and miRNA sequence. These tools are not able to predict non-conserved and evolutionarily new miRNAs.

NovoMIR, the tool used for the prediction of miRNAs in this study, does not rely on comparative genomics or prior knowledge of miRNA targets. NovoMIR uses a series of filter steps followed by a statistical model for prediction of plant pre-miRNAs by discriminating pre-miRNA from other RNAs. Teune and Steger (2010) searched for pre-miRNAs in *Arabidopsis* genome by using NovoMIR. All intergenic and intronic regions of *Arabidopsis* genome obtained from The Arabidopsis information resource (TAIR) were analysed by NovoMIR and reported 1,477 sequences as potential pre-miRNAs. Moss and

Collis (2012) searched for pre-miRNAs in *Linum usitatissimum* (Flax) genome, retrieved from NCBI and using NovoMIR predicted 2,324 potential pre-miRNAs.

The present study is the first report on NovoMIR, used for prediction of novel miRNAs in banana genome. Banana genome consist of 11 chromosomes, with genome size of 523 Mb. Chromosome sequences were downloaded from Banana Genome Hub. It contained genetic and genomic data of *Musa acuminata*, developed by Cirad and Bioversity International and supported by South green bioinformatics platform (Droc *et al.*, 2013). As NovoMIR cannot read large sized sequences, all the scaffold sequences of 11 chromosomes were made into 260 FASTA files. From them, NovoMIR predicted 85 pre-miRNAs, in which chromosome 9 showed maximum number of pre-miRNA sequences (22 No.). Length of pre-miRNAs ranged from 67 to 347 nt. Maximum number of pre-miRNAs (25%) ranged from 140 to 180 nt. In pre-miRNA sequences, G+C % ranged from 24–77 % and A+U % ranged from 23–76 %. GC content of pre-miRNAs was much lower than AU content. Higher AU content make the secondary structure of pre-miRNA less stable and leads to mature miRNA processing by RISC complex (Zhang *et al.*, 2006b). These are the parameters, which can contribute for the formation of secondary structures. Zhang *et al.* (2006b) reported that AU content of pre-miRNAs was higher than other non-coding RNAs. They found out 513 pre-miRNAs, belonging to seven plant species, having higher AU content compared to 237 non-coding RNAs.

Computational prediction of miRNAs rely on many criteria. NovoMIR is developed to predict pre-miRNAs based on their characteristic features like relative thermodynamic stability of structure, number and size of loops and length of helices. The sensitivity and specificity by which NovoMIR detected pre-miRNAs in *A. thaliana* genome have been reported as 80% and 90% respectively (Teune and Steger, 2010). pre-miRNAs possess a specific range of AU content and maintain MFE and distinct secondary structure.

In this study, secondary structures of pre-miRNAs were determined by using an online application called RNAfold web server. MFE is an important criterion for determining the stability of RNA secondary structure, since they form a stable stem-loop structure. Bonnet *et al.* (2004) reported that pre-miRNAs have lower MFE, when compared to other non-coding RNAs. In their study, they found out that 506 pre-miRNA sequences downloaded from RFAM database exhibited MFE considerably lower than other non-coding RNAs. Zhang *et al.* (2006b) also reported a similar observation.

In this study, MFE of predicted pre-miRNAs ranged from -20 to -194.4 Kcal/mol. AMFE of pre-miRNAs ranged from -18.9 to -85.4 Kcal/mol, which was lower when compared to MFE values of pre-miRNAs. Zhang *et al.* (2006b) reported that pre-miRNA sequences can form potential miRNAs, if they have MFEI value of ≥0.85. In this study, MFEI of the predicted pre-miRNAs ranged from 0.85 to 1.82, suggesting that all the pre-miRNAs can form mature miRNAs.

Though numerous miRNAs have been reported in various plant species, only a few are reported in *Musa* spp. Chai *et al.* (2015) computationally predicted 32 miRNAs belonging to 13 families in banana genome, based on sequence similarity with known miRNAs in miRBase. Sequences of mature plant miRNAs, obtained from miRBase, were used as query against 46,111 banana ESTs and 31,544 GSS, obtained from NCBI GenBank.

In this study, all the pre-miRNA sequences obtained from NovoMIR were blasted against mature miRNAs of miRBase and a total of 111 sequences were identified as potential miRNAs. The length of them ranged from 14-25 nt in length. Out of 111 potential miRNAs, 52 miRNAs having a length of 18-22 nt were considered as mature miRNAs.

miRNAs predicted by NovoMIR are assigned to different miRNA families on the basis of sequence conservation against predicted miRNAs in miRBase or Plant miRNA database (PMRD) (Moss and Cullis, 2012). In the present study, BLAST analysis resulted in 52 mature miRNAs belonged to 38 different miRNA

families. Properties of mature miRNAs like T_m, G+C %, Molecular weight and A+U % were found by using an online tool called Oligo Calc. Molecular weight of mature miRNAs ranged from -5353.4 gm/mole to -6724.3 gm/mole. T_m of mature miRNAs ranged from 31.3^{O} C to 66.2^{O} C. G+C% content of mature miRNAs ranged from 9-29% and A+U% content of mature miRNAs ranged from 11-91%. GC content of mature miRNAs was much lower than AU content, which indicates their potential as miRNAs.

5.2 IDENTIFICATION OF THE miRNA TARGETS USING TARGET PREDICTION SOFTWARE

Prediction of miRNA target genes is important for understanding the role of miRNAs in gene-regulation. Plant miRNAs showed perfect or near perfect complementary to their target mRNAs and help to regulate post-transcriptional gene expression (Bartel, 2004). Usually, each miRNA regulate many target genes. These target genes encode various transcriptional factors or functional enzymes that have important role in various biological processes or stress responses. Various computational tools were used to identify miRNA targets *i.e* evolutionarily conserved miRNA binding sites (Pinzon *et al.,* 2017). Many algorithms are developed for target prediction of plant and animal miRNAs. In this study, an online target analysis server psRNATarget was used for predicting miRNA targets. This server uses a proven scoring schema technology for finding the reverse complementary matching between miRNA and target transcript *i.e* mRNA. It also evaluates target-site accessibility by calculating unpaired energy (UPE) required to open the secondary structure around the centre of miRNA/mRNA duplex. It distinguishes translation inhibition and transcription regulation (Dai and Zhao, 2011). Less UPE shows a possibility of an effective target site.

Dai and Zhao (2011) have done the performance analysis of miRNA target prediction by psRNATarget. They predicted target genes for 10 reported *Arabidopsis thaliana* miRNAs in TAIR9 cDNA library with default parameters.

psRNATarget found 92 potential targets, of which 46 were validated target genes with 50% false positive prediction rate. By putting a more stringent cut-off threshold (Expectation ≤ 2), psRNATarget identified 52 potential targets. Out of them, 38 genes were validated by 5^{1}-RACE technology. In chilli, Din *et al.* (2016) identified 409 potential targets for 88 novel miRNAs by using psRNATarget server. They reported that almost all the predicted targets were functionally validated protein sequences in various organisms. Zhang *et al.* (2011) reported 125 protein-coding genes as potential targets for 13 novel miRNAs in sorghum and reported that a few targets were experimentally validated.

In the present study, based on scoring schema technology, using mature miRNA sequences as query and *Musa acuminata* (banana), cds and all the gene-coding sequences obtained from Banana Genome Hub, Version 1 as target transcript, potential targets were predicted for 52 mature miRNAs. Out of 52 mature miRNAs, only 40 miRNAs have found potential targets in banana genome. A total of 124 protein-coding sequences were considered as potential targets for 40 miRNAs, in which each miRNA was found to have more than one target (Lai *et al.*, 2012).

5.3 FUNCTIONAL ANNOTATION OF THE IDENTIFIED miRNA TARGETS

To understand the role of miRNAs, it is necessary to know the functions of the target genes in various metabolic pathways. Functional annotation determines the biological processes of mRNA targets regulated by miRNAs. For functional analysis of miRNAs, gene ontology annotations associated with the genes or gene products are used and many bioinformatic tools are widely used for functional annotation of target genes.

Gene ontology analysis acts as a leading tool for computational prediction of gene functions. Blast2Go (Gotz *et al.*, 2011) is the most commonly used tool, which provides a platform for the functional annotation and analysis of novel sequence data (Conesa *et al.*, 2005). It uses BLAST algorithm to find similar sequences against input sequences and annotation of GO terms was done based on

hits obtained in BLAST analysis. Blast2GO has been reported in more than 300 functional annotation projects involving a wide range of model and non-model species ranging from bacteria and fungi to plants.

In this study, functional annotation of identified miRNA targets was done using Blast2Go software. Blast2Go software proceeds in three steps mainly blasting, mapping and annotation to yield the result. It yields blast result statistics based on top-hit species distribution, species distribution and sequence similarity distribution (Gotz *et al.*, 2011). Analysis indicated maximum number of blast-hits for *Musa acuminata* ssp. *malaccensis*.

miRNA regulated genes control a variety of biological and metabolic processes. Functional annotation of target genes helps to understand miRNA function in terms of molecular function and biological process. In this study, out of 124 miRNA targets predicted, 84 target genes were found to be involved in various biological processes like anatomical structure development, lipid metabolism, drug metabolic process, cellular response to chemical stress, gene expression, RNA metabolic process, response to stress, transmembrane transport, carbohydrate biosynthetic process, oxidation-reduction process, protein modification process and phosphorylation and 118 target genes were found to be involved in various molecular functions like transferase activity, ATP binding, nucleic acid binding, protein binding, cofactor binding, transporter activity, oxidoreductase activity and metal ion binding. All these miRNAs were considered as potential miRNAs, as their target genes were involved in various molecular functions and biological processes.

5.4 EXPERIMENTAL VALIDATION OF THE NOVEL PREDICTED miRNAS

Computational prediction of miRNAs and their targets are considered as indispensable in miRNA research. Similarly, effective experimental techniques for validating in silico predictions are crucial for testing of computational algorithms. Interactions between in silico and experimental validation methods were playing a central role in the biology of miRNAs (Chaudhuri and Chatterjee,

2007), as there is a risk of false positive prediction in computational methods. miRNAs have been shown to have variable expression patterns with regard to tissue differentiation (Zhang *et al.*, 2006c). This study was envisaged to find out miRNAs differentially expressed in banana during virus infection.

Since virus infection affects many normal biological processes of the host, for experimental validation, out of 52 mature miRNAs predicted, five miRNAs and their targets were selected, based on their biological role in the host.

5.5 EXPRESSION ANALYSIS OF miRNAS AND THEIR TARGET GENES DURING VIRUS INFECTION

Several differentially regulated miRNAs have been reported in various crops under various stresses. In wheat, leaf rust disease caused by *Puccinia graminis* f. sp. *tritici* led to differential expression of 22 miRNAs across resistant and sensitive wheat cultivars (Kumar *et al.*, 2014) and powdery mildew disease, caused by *Blumeria graminis* f. sp. *tritici,* showed differential expression patterns of several miRNAs. Of them, miR156, 159, 164, 171, 396 were downregulated and miR393, 444, 827 were up-regulated (Xin *et al.*, 2010). Mishra *et al.* (2016) reported 67 conserved and 49 novel miRNAs in *Humulus lupulus.* Among them, 36 conserved and 37 novel miRNAs were found to be differentially expressed during *Citrus bark cracking viroid* (CBCVd) infection. In Tomato, miR159 was found up-regulated and miR164, miR171 downregulated, due to Tomato leaf curl virus (ToLCV) (Naqvi *et al.*, 2008). In rice, Southern rice black-streaked dwarf virus (SRB-SDV) infection altered expression profiles of 56 miRNAs belonging to 6 families (Xu *et al.*, 2014). Jin and Wu (2015) reported that 3 novel and 39 known miRNAs were found to be differentially expressed during *Pseudoperonospora cubensis* infection in cucumber.

Several methods have been developed for expression analysis of miRNAs, which include RT-qPCR and northern blot hybridization. In this study, expression analysis of predicted miRNAs during *BBrMV* infection was done in *in vitro* raised banana plants of variety Nendran using Real-Time PCR. Two common methods,

used to detect PCR products by Real-Time PCR are by using non-specific fluorescent dyes and sequence specific DNA probes. It monitors the amplification of miRNAs and their targets. Frazier *et al.* (2010) successfully used RT-qPCR to confirm the expression of 11 potential miRNAs in tobacco. Chai *et al.* (2015) applied RT-qPCR to detect expression levels of 12 putative miRNAs and 6 target genes in roots, leaves, flowers and fruits of banana.

In this study, three months old tissue culture plants of variety Nendran were infected with *BBrMV* virus by using banana aphids (*Pentalonia nigronervosa).* Twenty healthy aphids were released on *BBrMV* infected sucker for 30 min for acquisition feeding, after starving them for 10 min. The aphids were transferred on tissue culture plants for infection feeding, for 14 hours. The infection process was repeated twice a day for 7 days. PCR analysis with replicase gene specific primers of *BBrMV* indicated the presence of virus infection in all the infected plants. *BBrMV* is transmitted in a non-persistent manner by *Pentalonia nigronervosa, Aphis gossypii* and *Rhopalosiphum maidis*. Herradura *et al.* (2003) reported that optimum acquisition feeding time of 30 min and infection feeding time of 24 hr produced symptom on different *Musa* spp. In this study, starvation period, acquisition feeding and infection feeding time given are found sufficient for *BBrMV* infection within 7 days as indicated by the PCR result.

For the experimental validation of predicted miRNAs and their targets. RNA was isolated from the leaves of healthy and infected plants and reverse transcribed to cDNA, using stem-loop primers for miRNAs and oligo-dT primers for target genes. In RT-qPCR, gene expression levels need to be normalized using stably expressed gene. In this study, β-actin gene was used as a reference gene for normalizing the results. RT-qPCR was done to confirm the expression of five selected miRNAs and their targets in both healthy and *BBrMV* infected banana plants. All the five miRNAs and their target genes showed their presence in healthy and *BBrMV* infected samples. The expression levels between miRNAs and their target genes showed positive correlation.

In this study, F-box protein, targeted by miR-3900-5p showed higher expression level. F-box/kelch-repeat protein is an important component of E3 ubiquitin ligase Skp1-Cullin-F box protein (SCF) complex. It is involved in various cellular processes like regulation of cell cycle, hormone signalling, protein ubiquitination and defense pathway in plants (Correa *et al.*, 2013). F-box proteins were previously reported as targets of MiR394 (Wang *et al.*, 2009) and miR4995 (Chai *et al.*, 2015). In *Arabidopsis,* it is reported that silencing suppressor protein P0 of poleroviruses interacted with *Arabidopsis thaliana* orthologues of S-phase kinase-related protein 1 (SKP1) complex by means of F-box motif (Pazhouhandeh *et al.*, 2006). Point mutations in F-box domain inhibited interaction between P0 and SKP1 and rendered resistance to polerovirus infection. This suggest that F-box protein, targeted by miR-3900-5p may have a role in virus infection in banana.

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), a cellular enzyme, involved in oxidation-reduction process and glucose metabolic process. Prasanth *et al.* (2011) reported that over expression of GAPDH inhibits the accumulation of *Bamboo Mosaic Virus* (*BaMV*) and satellite *BaMV* RNAs in *Nicotiana benthamiana.* It leads to regulation of viral replication and translation by binding to the poly A tail of *BaMV* and stem-loop-C poly A tail of satellite *BaMV* 3^{I} UTR RNAs. In this study also GAPDH, targeted by miR-5417 was highly expressed and it may plays a crucial role in regulation of *BBrMV* replication and translation by binding to the 3^{I} UTR of *BBrMV*, which needs further investigation.

Armadillo repeat-containing kinesin-like protein 1 (ARM) was involved in root development, micro-tubular movement and stress response. Mandal *et al.* (2018) reported that there was a three-fold higher expression of ARM in tomato cultivars compared to tomato leaf curl New Delhi virus (ToLCNDV) infected cultivars. This indicates the involvement of ARM in defense response against viruses. In the present study, expression level of ARM, targeted by miR-3900-5p remains unchanged during *BBrMV* infection.

In this study, cytochrome c oxidase subunit 5B gene, targeted by miR-9112 was highly expressed during *BBrMV* infection. It plays an important role in mitochondrial electron transport and ATP synthesis. Hafez and Moustafa (2011) mechanically inoculated *Tomato bushy stunt virus* (*TBSV*) into three tomato cultivars (TY20, TY70/84 and TY70/70) for differential expression analysis of genes in infected plants. Sequence analysis revealed that the expression level of cytochrome c oxidase subunit 5B gene was higher in two resistant cultivars than in control and susceptible one, conferring resistance to *TBSV*. This suggest that cytochrome c oxidase subunit 5B gene may have a role in stress in banana.

FAD dependent oxidoreductase was targeted by miR-6928-5p family. It is involved in oxidoreductase activity, electron transport and respiratory chain. In this study, a positive correlation was observed between miR-6928-5p and FAD dependent oxidoreductase.

In plants, expression level of miRNAs are mostly found negatively correlated with expression level of their target genes (Jin and Wu, 2015). However, in this study, miRNA expression levels were positively correlated with their target genes. Further studies need to be performed at different intervals of virus infection, by including more replications to determine the regulation pattern of banana miRNAs and their targets.

The present study concludes that prediction tools can be effectively used for identification of the potential miRNAs in banana. It is also supported by the validation experiments. Differentially expressed miRNAs found in this study can be further validated to study their role in virus infection. Further, validation and expression analysis of the remaining miRNAs predicted in this study may help to get an understanding about the role of miRNAs in various biological process and stress conditions in banana.

SUMMARY

6. SUMMARY

The study entitled "Computational prediction of miRNAs in banana (*Musa* spp.) and evaluation of their role in virus infection" was conducted at the Department of Plant Biotechnology, College of Agriculture, Vellayani, Thiruvananthapuram and Central Tuber Crops Research Institute (ICAR-CTCRI), Sreekariyam, Thiruvananthapuram during 2016–2018. The objective of the study was to predict miRNAs in banana using bioinformatics tools and to validate and analyze their expression during *BBrMV* infection. The salient findings of the study were summarized below:

Banana is the world's fourth most important crop in terms of gross value production. It is the second most important crop in India next to mango. In India, Tamilnadu, Maharashtra, Gujarat, Andhra Pradesh and Karnataka are the leading banana cultivars. Banana is grown in 8.58 lakh ha with annual production of 29.1 MT. Among various biotic stresses, plant viruses impose serious threats to wide range of banana cultivation in modern agriculture. Among them, Banana bract mosaic disease, caused by *BBrMV,* was first reported in Nendran variety of banana in Kerala. Selvarajan and Jeyabaskaran (2006) reported that *BBrMV* infection cause 30 % yield reduction in Nendran cultivars. Due to lack of host resistance in *Musa* spp., no varieties having virus resistance have been developed. Virus free planting materials, phytosanitory measures and several non-conventional methods like genetic engineering techniques based on coat protein mediated resistance, movement protein mediated resistance and resistance genes (R genes) have been applied for developing virus resistance (Kumar *et al.*, 2015). The discovery of RNA interference (RNAi) mechanism in plants during the past decade has led to the development of small RNA based technologies for imparting resistance towards biotic stress in plants. Hence, this study was conducted to identify miRNAs in banana, which are differentially expressed during *BBrMV* infection and to select those differentially expressed miRNAs under virus infection, which can be used as marker to develop virus resistance in banana.

In this study, plant miRNA prediction tool called NovoMIR was used for prediction of novel miRNAs in banana genome. It is reported to have better accuracy and specificity in plant miRNA prediction. NovoMIR predicted 85 pre-miRNA sequences from 11 chromosomes of banana. Maximum number of pre-miRNAs ranged from 140 to 180 nt. Minimal folding free energy (MFE) of pre-miRNAs ranged from -20 to -194.4 Kcal/mol. AMFE of pre-miRNAs ranged from -18.9 to -85.4 Kcal/mol. Analysis of pre-miRNA sequences identified 52 mature miRNAs *via.* homology search against annotated miRNAs of miRBase. Majority of the identified mature miRNAs were 20 nt in length.

The targets for the predicted miRNAs were identified by using a web tool server psRNATarget and were functionally annotated. Based on scoring schema technology, 124 protein-coding genes were identified as potential targets in banana genome for 40 predicted mature miRNAs and each miRNA was found to have more than one target. Functional annotation of target genes revealed that all the miRNA target genes belong to *Musa acuminata* ssp. *malaccensis* and are involved in various molecular functions and biological processes.

For experimental validation, five predicted miRNAs and their targets were selected based on their biological role in the host. They included miR-6928-5p (Target: FAD dependent oxidoreductase), miR-3900-5p (Target: F-box/kelch repeat protein), miR-3900-5p (Target: Armadillo repeat containing kinesin like protein), miR-9112 (Target: cytochrome c oxidase subunit 5B) and miR-5417 (Target: Glyceraldehyde 3 phosphate dehydrogenase). Three months old tissue culture plants of variety Nendran were infected with *BBrMV,* by using banana aphids (*Pentalonia nigronervosa). BBrMV* infection was confirmed by RT-PCR for the presence of virus specific replicase gene.

Expression analysis using qPCR showed the presence of all the five predicted miRNAs in healthy and *BBrMV* infected leaf samples. Out of them, miR-3900-5p, miR-9112 and miR-5417 were found up-regulated, miR-6928-5p downregulated

and miR-3900-5p remain unchanged in infected samples and there was a positive correlation with the expression of their corresponding target genes.

In the present study, 52 mature miRNAs have been computationally predicted in banana genome and five of them were validated in banana variety Nendran. The study suggest the efficient use of computational prediction tools for identifying miRNAs in banana. Identification of differentially expressed miRNAs during virus infection open up the possibility of employing these miRNAs for crop improvement program in banana.

REFERENCES

7. REFERENCES

Achard, P., Herr, A., Baulcombe, D. C., and Harberd, N. P. 2004. Modulation of floral development by a gibberlin-regulated microRNA. *Development.* 131: 3357-3365.

Akter, A., Islam, M. M., Mondal, S. I., Mahmud, Z., Jewel, N. A., Ferdous, S., Amin, M. R. and Rahman, M. M. 2014. Computational identification of miRNA and targets from expressed sequence tags of coffee (*Coffea Arabica). Saudi J. Biol. Sci.* 21: 3-12.

Andrew, F., SiQun, X. U., Mary, K. M., Steven, A. K., Samuel, E. D. and Craig, C. 1998. Potent and specific genetic interference by double stranded RNA in *Caenorhabditis elegans. Nature.* 391: 806-811.

Anhalt, M. D. and Almeida, R. P. P. 2008. Effect of temperature, vector life stage and plant access period on transmission of banana bunchy top virus to banana. *Phytopathology.* 98: 743-748.

Axtell, M. J., Westholm, J. O. and Lai, E. C. 2011. Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol.* 12:221-234.

Balakrishnan, S., Gokulapalan, C. and Paul, S. 1996. A widespread banana malady in Kerala, India. *Infomusa.* 5: 28-29.

Baquera, P. L. O., Rodriguez, E. G., Santacruz, G. A. A., Cruz, Q. R., Conesa, A., Brito, V. M., Echavarria, R. and Viveros, J. D. 2017. Identification of miRNA from *Bouteloua gracilis,* a drought tolerant grass, by deep sequencing and their in-silico analysis. *Comput. Biol. chem.* 66: 26-35.

Barozai, M. Y. K., Ye, Z., Sangireddy, S. R. and Zhou, S. 2018. Bioinformatics profiling and expressional studies of microRNAs in root, stem and leaf of the bioenergy plant switch grass (*Panicum virgatum* L.) under drought stress. *Agri. gene.* 8: 1-8.

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

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

Bateson, M. F. and Dale, J. L. 1995. *Banana bract mosaic virus:* characterization using Potyvirus specific degenerate PCR primers. *Arch. Virol.* 140: 515-527.

Bej, S. and Basak, J. 2014. MicroRNAs: The potential biomarkers in plant stress response. *Am. J. Plant Sci.* 5: 748-759.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J. and Ostell, J. 2013. GenBank. *Nucleic Acids Res.* 41: 36–42.

Bentwich, I. 2005. Prediction and validation of microRNAs and their targets. *FEBS J.* 579: 5904-5910.

Betel, D., Koppal, A., Agius, P., Sander, C. and Leslie, C. 2010. Comprehensive modelling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* 11: 90-95.

Bi, F., Meng, X., Ma, C. and Yi, G. 2015. Identification of miRNAs involved in fruit ripening in Cavendish bananas by deep sequencing. *BMC Genomics.* 16: 776-786.

Biswas, S., Hazra, S. and Chattopadhyay, S. 2016. Identification of conserved miRNAs and their putative target genes in *Podophyllum hexandrum* (Himalayan Mayapple). *Plant gene.* 6: 82-89.

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.

Bouhida, M. and Lockhart, B. E. L. 1990. Increase in importance of cucumber mosaic virus infection in greenhouse grown bananas in morocco. *Phytopathology.* 80: 981-987.

Brault, V., Uzest, M., Monsion, B., Jacquot, E. and Blanc, S. 2010. Aphids as transport devices for plant viruses. *C. R. Biol.* 333: 524-538.

Bressan, A. and Watanabe, S. 2011. Immunofluorescence localization of banana bunchy top virus within the aphid vector, *Pentalonia nigronervosa*, suggests a virus tropism distinct from aphid transmitted *luteoviruses*. *Virus Res.* 155: 520-525.

Carr, J. P., Marsh, L. E., Lomonossof, G. P., Sekiya, M. E. and Zaitlin, M. 1992. Resistance to tobacco mosaic virus induced by the 54-k Da gene sequence requires expression of the 54-k Da protein. *Mol. Plant Microbe Interact.* 5: 397-404.

Carrington, J. C. and Ambros, V. 2003. Role of MicroRNAs in plant and animal development. *Science*. 301: 336-338.

Chai, J., Feng, R., Shi, H., Ren, M., Zhang, Y. and Wang, J. 2015. Bioinformatic identification and expression analysis of banana microRNAs and their targets. *PLoS One*.10: 1-15.

Chaudhuri, K. and Chatterjee, R. 2007. MicroRNA detection and target prediction: integration of computational and experimental approaches. *DNA Cell Biol.* 26: 5-15.

Chen, C. M., Lu, Y. L., Sio, C. P., Wu, G. C., Tzou, W. S. and Pai, T. W. 2014. Gene ontology based housekeeping gene selection for RNA-seq normalization. *Methods.* 67: 354-363.

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., Lvak, K. J. and Guegler, K. J. 2005. Real time quantification of microRNAs by stem loop RT-PCR. *Nucleic Acids Res.* 33: 20-25.

Chen, X. 2004. A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development *Science.* 303: 2022-2025.

Conesa, A., Gotz, S., Gomez, J. M. G., Terol, J., Talon, M. and Robles, M. 2005. Blast2Go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 21: 3674-3676.

Correa, R. L., Bruckner, F. P., Cascardo, R. S. and Zerbini, P. A. 2013. *Int. J. Mol. Sci.* 14: 4030-4049.

Cullen, B. R. 2004. Transcription and processing of human microRNA precursors. *Mol. Cell.* 16: 861-865.

D'Hont, A., Denoeud, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., *et al.* 2012. The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. *Nature*. 488: 213-217.

Dai, X. and Zhao, P. X. 2011. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39: 155-159.

Daniells, J. W., Geering, A. D. W., Bryde, N. J. and Thomas, J. E. 2001. The effect of banana streak virus on the growth and yield of desert bananas in tropical Australia. *Ann. Appl. Biol.* 139: 51-60.

Darling, A., Carey, L. and Feng, W., 2003. The Design, implementation and evaluation of mpiBLAST. ClusterWorld conference & Expo and the 4th international conference on linux clusters: *The HPC Revolution*, pp.14.

Das, A., Chaudhury, S., Kalita, M. C. and Mondal, T. K. 2015. Insilico identification, characterization and expression analysis of miRNAs in *Cannabis sativa* L. *Plant Gene.* 2: 17-24.

Dehury, B., Panda, D., Sahu, J., Sahu, M., Sarma, K., Barooah, M., Sen, P. and Modi, M. K. 2013. In-silico identification and characterization of conserved miRNAs and their target genes in sweet potato (*Ipomoea batatas* L.) Expressed Sequence Tags (ESTs). *Plant Signal Behav.* 8: 12-20.

Department of Agriculture, cooperation & farmers welfare. 2017. *Horticultural Statistics at a Glance* 2017. Ministry of Agriculture & farmer's welfare, Government of India, 514p.

Din, M., Barozai, M. Y. K. and Baloch, I. A. 2016. Profiling and annotation of microRNAs and their putative target genes in chilli (*Capsicum annum* L.) using ESTs. *Gene Reports*. 5: 62-69.

Ding, S. W. 2000. RNA silencing. *Curr. Opin. Biotechnol.* 11: 152-156.

Dizon, O. T., Damasco, O. P., Lobina, I. T., Pinili, M. S., Lalusin, A. G. and Natsuaki, K. T. 2012. Induction of putative resistant lines of abaca (Musa textiles Nee) to banana bunchy top virus and banana bract mosaic virus through in vitro mutagenesis. *J. ISSAAS.* 18: 87-99.

Dong, Q. H., Han, J., Yu, H. P., Wang, C., Zhao, M. Z., Liu, H., Ge, A. J. and Fang, J. G. 2012. Computational identification of microRNAs in strawberry expressed sequence tags and validation of their precise sequences by miRRACE. *J. Hered.* 103: 268-277.

Droc, G., Lariviere, D., Guignon, V., Yahiaoui, N., This, D., Garsmeur, O., Dereeper, A., Hamelin, C., Argout, X., Dufayard, J. S., Lengelle, J., Baurens, F. C., Cenci, A., Pitollat, B., Hont, A. D., Ruiz, M., Rouard, M. and Bocs, S. 2013. The banana genome hub. *Database*. 35: 1-5.

Emery, J. F., Floyd, S. K., Alvarez, J., Eshed, Y., Hawker, N. P., Izhaki, A., Baum, S. F., and Bowman, J. L. 2003. Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes. *Curr. Biol.* 13: 1768-1774.

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

Feng, L., Duan, C. G. and Guo, H. S. 2013. Inhibition of in vivo slicer activity of argonaute protein 1 by the viral 2b protein independent of its dsRNA binding function. *Mol. Plant Pathol.* 14: 617-622.

Floyd, S. K. and Bowman, J. L. 2004. Ancient microRNA target sequences in plants. *Nature.* 428: 485-486.

Frazier, T. P., Xie, F., Friestaedter, A. and Zhang, B. 2010. Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabaccum). Planta.* 10: 2-9.

Gao, R., Liu, P. and Wong, S. M. 2012. Identification of a plant viral RNA genome in the nucleus. *PLoS One.* 7: 10-15.

Gotz, S., Arnold, R., Leon, P. S., Rodrigiez, S. M., Tischler, P., Jehl, M. A., Dopazo, J., Rattei, T. and Conesa, A. 2011. B2G-FAR, a species-centered GO annotation repository. *Bioinformatics.* 27: 919-924.

Gotz, S., Gomez, J. M. G., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talon, M., Dopazo, J. and Conesa, A. 2008. High-throughput functional annotation and data mining with the Blast2Go suite. *Nucleic Acids Res.* 36: 3420-3435.

Hafez, E. E. and Moustafa, M. F. M. 2011. Differential expression of cytochrome oxidase and ALY-family genes in resistant and susceptible tomato cultivars (*Solanum lycopersicum)* inoculated with *Tomato bushy stunt virus. J. Genet. Eng. Biotechnol.* 9: 43-49.

Hammond, S. M., Bernstein, E., Beach, D. and Hannon, G. J. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in drosophila cells. *Nature.* 404: 293-296.

Hasan, M. A., Hussain, M. H., Chowdhury, A. S., Dhar, S. B., Abedin, M. and Fima, I. N. 2016. Computational identification of potential microRNAs and their targets from expressed sequence tags of marijuana (*Cannabis sativa*). *Meta. Gene.* 10: 45-55.

Herradura, L. E., Magnaye, L. V., Alforque, M. A. and Lobres, M. A. 2003. Transmission of banana bract mosaic virus by three aphis species in the Philippines (2003). *AGRIS*. 39: 58-65.

Heslop-Harrison, J. S. and Schwarzacher, T. 2007. Domestication, genomics and the future for banana. *Ann. Botany.* 100: 1073-1084.

Hofacker, I. L. 2004. RNA secondary structure analysis using the Vienna RNA package. *Curr. Protoc. Bioinformatics*. 4: 12-25.

Hooks, C. R. R., Wright, M. G., Kabasawa, D. S., Manandhar, R. and Almeida, R. P. P. 2008. Effect of banana bunchy top virus infection on morphology and growth characteristics of banana. *Ann. Appl. Biol.* 153: 1-9.

Hu, J. S., Wang, M., Sether, D., Xie, W. and Leonhardt, K. W. 1996. Use of polymerase chain reaction (PCR) to study transmission of banana bunchy top virus by the banana aphid (*Pentalonia nigronervosa*). *Ann. Appl. Biol.* 128: 55-64.

Huang, Y., Ren, H. T., Zou, Q., Wang, Y. Q., Zhang, J. L. and Yu, X. L. 2015. Computational identification and characterization of miRNAs and their target genes from five cyprinidae fishes. *Saudi J. Biol. Sci.* 10: 15-20.

Hull, R. 2013. Plant Virology. *Academic Press*, pp. 100-110.

Hussain, K., Mungikar, K., Kulkarni, A. and Kamble, A. 2018. Identification, characterization and expression analysis of pigeon pea miRNAs in response to *Fusarium wilt. Gene*. 653: 57-64.

Jin, W. and Wu, F. 2015. Identification and characterization of cucumber microRNAs in response to *pseudoperonospora cubensis* infection. *Gene.* 569: 225-232.

John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C. and Marks, D. S. 2004. Human microRNA targets. *PLoS Biol.* 2: 11-22.

Kadri, S., Hinman, V. and Benos, P. V. 2009. "HHMMiR: efficient de novo prediction of microRNAs using hierarchial hidden Markov models". *BMC Bioinformatics*. 10: 1-10.

Kertesz, M., Lovino, N., Unnerstall, U., Gaul, U. and Segal, E. 2007. The role of site accessibility in microRNA target recognition. *Nature Genet.* 39: 1278-1284.

Kibbe, W. A. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* 35: 43-46.

Kincaid, R. P. and Sullivan, C. S. 2012. Virus-Encoded microRNAs: An overview and a look to the Future. *PLoS Pathog.* 16: 175-179.

King, A., Adams, M., Carstens, E., and Lefkowitz, E. 2012. Virus taxonomy: *Ninth Report of the International Committee on Taxonomy of Viruses*. Amsterdam: *Elsevier Academic Press*. 1327p.

Kozomara, A. and Griffiths-Jones, S. 2014. 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: 10-20.

Krek, A., Grun, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., Amin, P. M., Piedade, I. D., Gunsalas, K. C., Stoffel, M. and Rajewsky, N. 2005. Combinatorial microRNA target predictions. *Nature Genet.* 37: 495-500.

Ku, Y. S., Wong, J. W. H., Mui, Z., Liu, X., Hui, J. H. L. and Chan, T. F. 2015. Small RNAs in plant responses to abiotic stresses: regulatory roles and study methods. *Int. J. Mol. Sci.* 16: 24532-24554.

Kumar, D., Singh, D., Kanodia, P., Prabhu, K. V., Kumar, M. and Mukhopadhyay, K. 2014. Discovery of Novel Leaf Rust Responsive microRNAs in wheat and Prediction of their Target Genes. *J. Nucleic Acids*. 15: 1154-1199.

Kumar, P. L., Selvarajan, R., Caruana, M. L. I., Chabannes, M. and Hanna, R. 2015. Biology, etiology and control of virus diseases of banana and plantain. *Advances in virus research*. 258p.

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

Kuruvilla, L. 2017. MicroRNAs of Hevea brasiliensis: Role in abiotic stress responsive gene regulation. Ph.D. Thesis, Rubber research institute of India, Kottayam, Kerala. 2017.

Lai, X., Schmitz, U., Gupta, S. K., Bhattacharya, A., Kunz, M., Wolkenhauer, O. and Vera, J. 2012. Computational analysis of target hub gene repression regulated by multiple and cooperative miRNAs. *Nucleic Acids Res.* 40: 8818-8834.

Lakatos, L., Csorba, T., Pantaleo, V., Chapman, E. J., Carrington, J. C. and Liu, Y. P. 2006. Small RNA binding is a common strategy to suppress RNA silencing by several viral suppressors. *EMBO. J.* 25: 2768-2780.

Lewis, B. P., Shih, I. H., Rhoades, M. W. J., Bartel, D. P. and Burge, C. B. 2003. Prediction of mammalian microRNA targets. *Cell.* 115: 787-798.

Li, Y. F., Zheng, Y., Addo-Quaye, C., Zhang, L., Saini, A., Jagadeeswaran, G., Axtell, M. J., Zhang, W. and Sunkar, R. 2010. Transcriptome wide identification of microRNA targets in Rice. *Plant J.* 62: 742-759.

Liu, B., Li, J. and Cairns, M. J. 2012. Identifying miRNAs, targets and functions. *Brief. Bioinform.* 15: 1-19.

Lockhart, B. E. L. and Jones, D. R. 2000. Banana mosaic. Diseases of banana, abaca and enset. 544p.

Lorenz, R., Bernhart, S. H., Siederdissen, C. H. Z., Tafer, H., Flamm, C., Stadler, P. F. and Hofacker, I. L. 2011. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6: 26-29.

Lu, Y., Gan, Q. and Chi, X. 2008. Roles of microRNA in plant defense and virus offense interaction. *Plant Cell Rep.* 27: 1571-1579.

Magee, C. J. P. 1927. Investigations of the bunchy top disease of bananas. *CSIR.* 30: 64-65.

Magnaye, L. V. and Espino, R. R. C. 1990. Banana bract mosaic, a new disease of banana I. Symptomatology. *Philipp. Agri. Sci.* 73: 55-59.

Mallory, A. C., Reinhart, B. J., Jones-Rhoades, M. W., Tnag, 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 region 5I region. *EMBO. J.* 23: 3356-3364.

Mandal, A., Mishra, A. K. and Dulani, P. 2018. Identification, characterization, expression profiling and virus-induced gene silencing of armadillo repeat-containing proteins in tomato suggest their involvement in tomato leaf curl New Delhi virus resistance. *Funct. Integr. Genomics.* 18: 101-111.

Mandhan, V. and Singh, K. 2015. Identification of novel microRNAs and their target prediction in stevia rebaudiana. *Transcriptomics.* 3: 1-8.

Maragkakis, M., Alexiou, P., Papadopolus, G., Reczko, M., Dalamagas, T., Giannopoulus, G., Goumas, G., Koukis, E., Kourtis, K., Simossis, V. A., Sethupathy, P., Vergoulis, T., Koziris, N., Sellis, T. and Tsanakakas, P. 2009. Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics.* 10: 295-300.

Martinelli, F., Cannarozzi, G., Balan, B., Siegrist, F., Weichert, A., Blosch, R. and Tadele, Z. 2018. Identification of miRNAs linked with the drought response of tef (*Eragrostis tef* (Zucc.) Trotter). *J. Plant Physiol.* 224: 163-172.

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. and Tuschl, T. 2002. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell.* 110: 563-574.

Mathews, D. H., Sabina, J., Zuker, M. and Turner, D. H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288: 911-940.

Millar, A. A. and Waterhouse, P. M. 2005. Plant and animal microRNAs: similarities and differences. *Funct. Integr. Genomics.* 5: 129-135.

Miranda, K. C., Huynh, T., Tay, Y., Ang, Y. S., Tam, W. L., Thomson, A. M., Lim, B. and Rigoutsos, I. 2006. A pattern based method for the identification of microRNA binding sites and their corresponding hetero duplexes. *Cell.* 126: 1203-1217.

Mishra, A. K., Duraisamy, G. S., Tycova, A. and Matousek, J. 2015. Computational exploration of microRNAs from expressed sequence tags of *Humulus lupulus*, target prediction and expression analysis. *Comput. Biol. Chem.* 59: 131-141.

Mishra, A. K., Duraisamy, G. S., Matousek, J., Radisek, S., Javornik, B. and Jakse, J. 2016. Identification and characterization of microRNAs in *Humulus lupulus* using high-throughput sequencing and their response to *Citrus bark cracking viroid* (CBCVd) infection. *BMC Genomics.* 17: 919-939.

Moss, T. Y. and Cullis, C. A. 2012. Computational prediction of candidate microRNAs and their targets from the completed *Linum usitatissimum* genome and EST database. *J. Nucleic Acids.* 3: 10-20.

Napoli, C., Lemieux, C. and Jorgensen, R. 1990. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co suppression of homologous genes in trans. *Plant Cell.* 2: 279-289.

Naqvi, A. R., Choudhury, N. R., Haq, Q. M. R. and Mukherjee, S. K. 2008. MicroRNAs as Biomarkers in Tomato Leaf Curl Virus (ToCLV). *Nucleic Acids Symp. Ser.* 52: 507-508.

Nelson, S. C. 2004. Banana bunchy top: Detailed signs and symptoms. *UH-CTAHR Cooperative extension service.* 22p.

Neutelings, G., Fenart, S., Danila, A. L. and Hawkins, S. 2012. Identification and characterization of miRNAs and their potential targets in flax. *J. Plant Physiol.* 169: 1754-1766.

Ngatat, S., Hanna, R., Kumar, P. L., Gray, S. M., Cilia, M. and Fontem, A. 2013. Long term evaluation of the susceptibility of 16 *Musa* genotypes to banana bunchy top disease in Cameroon, central Africa. *Phytopathology.* 103: 105-107.

Nithin, C., Patwa, N., Thomas, A., Bahadur, R. P. and Basak, J. 2015. Computational prediction of miRNAs and their targets in *Phaseolus vulgaris* using simple sequence repeat signatures. *BMC Plant Biol.* 15: 140-150.

Numnark, S., Mhuantong, W., Ingsriswang, S. and Wichadakul, D., 2012. C-mii: a tool for plant miRNA and target identification. *BMC Genomics.* 13: 16-19.

Ortiz, R. 2013. Conventional banana and plantain breeding. *Acta Hortic.* 986: 77-194.

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

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

Pazhouhandeh, M., Dieterle, M., Marrocco, K., Lechner, E., Berry, B., Brault, V., Hemmer, O., Kretsch, T., Richards, K. E., Genschik, P. and Graff, V. Z. 2006. F-box-like domain in the polerovirus protein P0 is required for silencing suppressor function. *Proc. Natl. Acad. Sci.* 103: 1994-1999.

Pfeffer, S., Zavolan, M., Grasser, F. A., Chien, M., Russo, J. J., and Ju, J. 2004. Identification of virus-encoded microRNAs. *Science.* 304: 734-736.

Pillai, R. S., Bhattacharya, S. N., and Filipowicz, W. 2006. Repression of protein synthesis by miRNAs: how many mechanisms?. *Trends Cell Biol.* 17: 118-126.

Prabhu, G. R. and Mandal, A. K. A. 2010. Computational identification of miRNAs and their target genes from expressed sequence tags of tea (*Camellia sinensis). Genomics Proteomics Bioinformatics.* 8: 113-121.

Prakash, P., Rajakani, R. and Gupta, V. 2016. Transcriptome-wide identification of *Ravoulfia serpentine* microRNAs and prediction of their potential targets. *Comput. Biol. Chem.* 61: 62-74.

Prasanth, K. R., Huang, Y. W., Liou, M. R., Wang, R. Y. L., Hu, C. C., Tsai, C. H., Meng, M., Li, N. S. and Hsu, Y. H. 2011. Glyceraldehyde 3-phosphate dehydrogenase negatively regulates the replication of *Bamboo mosaic virus* and its associated satellite RNA. *J. Virol.* 85: 8829-8840.

Pinzon, N., Li, B., Martinez, L., Sergeeva, A., Presumey, J., Apparailly, F. and Seitz, H. 2017. microRNA target prediction programs predict many false positives. *Genome Res.* 27: 234-245.

Qureshi, A., Thakur, N., Monga, I., Thakur, A. and Kumar, M. 2014. VIRmiRNA: a comprehensive resource for experimentally validated miRNAs and their targets. *Database.* 10: 1-10.

Radivojac, P., Clark, W. T., Oron, T. R., Schnoes, A. M., Wittkop, T., Sokolov, A., Graim, K., Funk, C., Verspoor, K. and Benhur, A. 2013. A large scale evaluation of computational protein function prediction. *Nature Methods*. 10: 221-227.

Rajewsky, N. 2006. microRNA target predictions in animals. *Nature Genet.* 38: 8-13.

Rao, X., Huang, X., Zhou, Z. and Lin, X. 2013. An improvement of the 2^(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. *Biostat. Bioinforma. Biomath.* 3: 71-85.

Rehmsmeier, M., Steffen, P., Hochsmann, M. and Geigerich, R. 2004. Fast and effective prediction of microRNA/target duplexes. *RNA*. 10: 1507–1741.

Resmi, L. and Nair, A. S. 2007. Plantlet production from the male inflorescence tips of *Musa acuminata* cultivars from south india. *Plant Cell Tiss. Organ Cult.* 88: 333-338.

Robson, J. D., Wright, M. G. and Almeida, R. P. P. 2006. Within plant distribution and binomial sampling plan of *Pentalonia nigronervosa* (Hemiptera, Aphididae) on banana. *J. Econ. Entomol.* 99: 2185-2190.

Rodoni, B. C., Ahlawat, Y. S., Varma, A., Dale, J. L. and Harding, R.M. 1997. Identification and characterization of banana bract mosaic virus in India. *Plant Dis.* 81: 669-672.

Sahoo, S., Samal, R. R. and Dixit, S. 2018. Computational identification and characterization of conserved miRNAs and their putative target genes in *Eclipta prostrata*. *Gene Reports*. 11: 213-219.

Schwab, R., Ossowski, S., Riester, M., Warthmann, N. and Weigel, D. 2006. Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *Plant Cell* 18: 1121-1133.

Schwarz, D. S., Hutvagner, G., Haley, B. and Zamore, P. D. 2002. Evidence that siRNAs function as guides, not primers, in the drosophila and human RNAi pathways. *Mol. Cell.* 10: 537-548.

Selvarajan. R. and Jeyabaskaran, K. J. 2006. Effect of Banana bract mosaic virus (*BBrMV*) on growth and yield of cultivar Nendran (Plantain, AAB). *Indian Phytopath.* 59: 496-500.

Shahi, P.; Loukianiouk, S., Lang, A. B., Kenzelmann, M., Kuffer, S., Maertens, S., Eils, R., Grone, H. J., Gretz, N. and Brors, B. 2006. Argonaute – a database for gene regulation by mammalian microRNAs. *Nucleic Acids Res.* 34: 115-118.

Simon-Mateo, C. and Garcia, J. A. 2011. Antiviral strategies in plants based on RNA silencing. *Biochim. Biophys. Acta* 1809: 722-731.

Singh, N. and Sharma, A. 2014. In-silico identification of miRNAs and their regulating target functions in *Ocimum basilicum*. *Gene*. 552: 277-282.

Singh, N., Srivastava, S. and Sharma, A. 2016. Identification and analysis of miRNAs and their targets in ginger using bioinformatics approach. *Gene*. 575: 570-576.

Song, S., Chen, X., Huang, D., Xu, Y., Zeng, H., Hu, X., Xu, B., Jin, Z. and Wang, W. 2016. Identification of miRNAs differentially expressed in fusarium wilt resistant and susceptible banana varieties. *S. Afr. J. Bot.* 106: 244-249.

Sorin, C., Bussell, J. D., Camus, I., Ljung, K., Kowalczyk, M., Geiss, G., McKhann, H., Garcion, C., Vaucheret, H., Sandberg, G., Bellini, C. 2005. Auxin and light control of adventitious rooting in *Arabidopsis* require ARGONAUTE1. *Plant Cell.* 17: 1343-1359.

Srivastava, S., Singh, N., Srivastava, G. and Sharma, A. 2017. miRNA mediated gene regulatory network analysis of *Cichorium intybus* (chicory). *Agri. Gene.* 3: 37-45.

Sun, R., Du, P., Jiang, L., An, D. and Li, Y. 2014. Heterologous expression of artificial miRNAs from *rice dwarf virus* in transgenic rice. *Plant Cell Tissue Organ Cult.* 116: 353-360.

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

Teune, J. H. and Steger, G. 2010. NovoMIR: De novo prediction of microRNA coding regions in a single plant genome. *J. Nucleic Acids.* 25: 55-65.

Thomas, J. E. and Dietzgen, R. G. 1991. Purification, characterization and serological detection of virus like particles associated with banana bunchy top virus in Australia. *J. Gen. Virol.* 72: 217-224.

Tsai, C. J. and Harding, S. A. 2003. Isolation and purification of RNA. *Handbook of molecular and cellular methods in biology and medicine.* 44p.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. and Leunissen, J. A. M. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35: 71-74.

Vaucheret, H., Vanquez, F., Crete, P. and Bartel, D. P. 2004. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev.* 18: 1187-1197.

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

Wagaba, H., Patil, B. L., Mukasa, S., Alicai, T., Fauquet, C. M. and Taylor, N. J. 2016. Artificial microRNA-derived resistance to cassava brown streak disease. *J. Virol. Methods.* 231: 38-43.

Wang, X., Kong, H. and Ma, H. 2009. F-box proteins regulate ethylene signalling and more. *Genes Dev.* 23: 391-396.

Wang, Z., Jiang, D., Zhang, C., Tan, H., Li, Y., Lv, S., Hou, X. and Cui, X. 2015. Genome wide identification of turnip mosaic virus responsive microRNAs in non-heading chinese cabbage by high-throughput sequencing. *Gene.* 571: 178-187.

Wani, S. H., Sah, S. K., Hussain, M. A., Kumar, V. and Balachandra, S. M. 2016. Transgenic approaches for abiotic stress tolerance in crop plants *Advances in Plant Breeding Strategies.* 2: 345-396.

Watanabe, Y., Tomita, M. and Kanai, A. 2007. Computational methods for microRNA target prediction. *Methods Enzymol.* 427: 65-89.

Waterhouse, P. M., Wang, M. B. and Lough, T. 2001. Gene silencing as an adaptive defense against viruses. *Nature.* 411: 834-842.

Wen, J. Z., Liao, J. Y., Zheng, L. L., Xu, H., Yang, J. H., Guan, D. G., Zhang, S. M., Zhou, H. and Qu, L. H. 2014. A Contig-based strategy for the genome wide discovery of microRNAs without complete genome resources. *PLoS One.* 9: 2-8.

Whitefield, A. E., Falk, B. W. and Rotenberg, D. 2015. Insect vector-mediated transmission of plant viruses. *Virology*. 479: 278-289.

Xia, Z., Zhao, Z., Li, M., Chen, L., Jiao, Z., Wu, Y., Zhou, T., Yu, W. and Fan, Z. 2018. Identification of miRNAs and their targets in maize in response to sugarcane mosaic virus infection. *Plant Physiol. Biochem.* 125: 143-152.

Xin, M., Wang, Y., Yao, Y., Xie, C., Peng, H., Ni, Z. and Sun, Q. 2010. Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in Wheat (*Triticumaestivum* L.). *BMC Plant Biol.* 10: 123-125.

Xu, D., Mou, G., Wang, K. and Zhou, G. 2014. MicroRNAs Responding to Southern Rice Black Streaked Dwarf virus infection and their target genes associated with symptom development in Rice. *J. Virol.* 45: 227-254.

Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L. and Zhou, M. M. 2003. Structure and conserved RNA binding of the PAZ domain. *Nature.* 426: 468-474.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T. L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 13: 134-139.

Yi, M., Schultz, D. E. and Lemon, S. M. 2000. Functional significance of the interaction of hepatitis A virus with glyceraldehyde 3-phosphate dehydrogenase (GAPDH): opposing effects of GAPDH and polypyrimidine tract binding protein on internal ribosome entry site function. *J. Virol.* 74: 6459-6468.

Yogindran, S. and Rajam, M. V. 2016. Artificial miRNA-mediated silencing of ecdysone receptor (EcR) affects larval development and oogenesis in *Helicoverpa armigera*. *Insect Biochem. Mol. Biol.* 77: 21-30.

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

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

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

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

Zhang, L., Zheng, Y., Jagadeeswaran, G., Li, Y., Gowdu, K. and Sunkar, R. 2011. Identification and temporal expression analysis of conserved and novel microRNAs in sorghum. *Genomics*. 98: 460-468.

Zinati, Z., Dashtpagerdi, R. S. and Behpouri, A. 2016. Insilico identification of miRNAs and their target genes and analysis of gene co-expression network in saffron (*Crocus sativus L.*) stigma. *Mol. Biol. Res. Commun.* 5: 233-246.

ABSTRACT

COMPUTATIONAL PREDICTION OF miRNAs IN BANANA (*Musa* spp.) AND EVALUATION OF THEIR ROLE IN VIRUS INFECTION

by

KOKILA SAJEEV ANURAG MATHEW

(2016-11-078)

ABSTRACT

of the thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN AGRICULTURE

Faculty of Agriculture

Kerala Agricultural University

DEPARTMENT OF PLANT BIOTECHNOLOGY

COLLEGE OF AGRICULTURE

VELLAYANI, THIRUVANANTHAPURAM-695 522

KERALA, INDIA

2018

ABSTRACT

The study entitled "Computational prediction of miRNAs in banana (*Musa* spp.) and evaluation of their role in virus infection" was conducted at the Department of Plant Biotechnology, College of Agriculture, Vellayani, Thiruvananthapuram and Central Tuber Crops Research Institute (ICAR-CTCRI), Sreekariyam, Thiruvananthapuram during 2016–2018. The objective of the study was to predict miRNAs in banana using bioinformatics tools and to validate and analyze their expression during *BBrMV* infection.

Computational miRNA prediction tool NovoMIR was used for the prediction of miRNAs in banana genome. Analysis was performed with all the gene coding or nucleotide sequences of banana and 85 pre-miRNAs were predicted from 11 chromosomes. Most of the pre-miRNAs ranged from 140–180 nt. G+C% content of pre-miRNAs ranged from 24-77% and A+U% content ranged from 23-76%. MFE of pre-miRNAs were found by using an online application RNAfold web server. MFE of pre-miRNAs ranged from -20 Kcal/mol to -194.4 Kcal/mol. AMFE of pre-miRNAs ranged from -18.9 Kcal/mol to -85.4 Kcal/mol. All the predicted pre-miRNAs by NOVOMIR had MFE $\leq$ -20 Kcal/mol. MFEI of pre-miRNAs ranged from 0.85 to 1.82.

BLAST analysis of 85 pre-miRNAs against annotated mature miRNAs in miRBase resulted in 52 mature miRNAs. The targets for the 52 mature miRNAs were predicted using the web tool psRNATarget and were functionally annotated by Blast2Go analysis server.

Targets were identified for 40 miRNAs in banana genome. A total of 124 targets were found, with each miRNA having more than one target.

Validation of the predicted miRNAs were done in *in vitro* banana plants of variety Nendran. Three months old tissue culture plants were infected with *BBrMV* virus by using aphids. Twenty healthy aphids were released on *BBrMV* infected

sucker for 30 min for acquisition feeding, after starving for 10 min. These aphids were further released on tissue culture plants for 12-14 h for infection. The infection process was repeated twice a day for 7 days.

For experimental validation, five miRNAs and their target genes having role in biological process were selected and stem-loop/gene specific primers were designed. RNA was isolated from healthy and infected leaf samples and reverse transcribed to cDNA. Expression analysis using RT-qPCR showed the presence of all the five miRNAs in healthy and *BBrMV* infected leaf samples. Out of them, miR-3900-5p, miR-9112 and miR-5417 were found up-regulated, miR-6928-5p downregulated and miR-3900-5p remain unchanged in infected samples and there was a positive correlation with the expression of their corresponding target genes.

In the present study, 52 mature miRNAs have been predicted using bioinformatics tools in banana genome. Targets for 40 miRNAs were identified in banana genome. The five miRNAs selected were validated along with their targets. Expression analysis showed regulation of four of them during virus infection, indicating the possibility of their role in stress response in banana. The remaining miRNAs predicted need validation.

APPENDICES

Appendix I - List of all predicted pre-miRNAs by NovoMIR									
Chr. No.	**Scaffold no.**	**Pre-miRNA sequence**	**MFE**	**LP (nt)**	**Mol. Wt.**	**G+C %**	**A+U %**	**AMFE**	**MFEI**
1	241	gguugcaaaaauuucgaaagcaaaugcaagauagcuuucuuguguaagcuuaugauucuugcuuccuauug	-20.4	71	22893	35	75	-28.7	0.82
	239	ggugguacuaucgcuaagucugggcgguaccacugccugauagagcucggagacugagcucuagcaguaccaccgcuugacaggggccguaccacca	-50.7	97	31445	59	41	-52.2	0.88
	234	cugggcggugcaaccucuccuaacaagaguuggcaccgcuugagcucggucuucgagcucugguaggaggugcaaccgccucagucaggagguggcaccgccugagcucggucuucgagcucuggcaggaggugcaaucccccccuaacagagguggcacugcccaga	-92.4	167	54022	62	38	-55.3	0.89
	232	gggcgguacaucagggcguaccgagcgauacauccuaguguaucgagcgguacacccuaguguaacgaaca	-36.9	71	21779	55	45	-51.9	0.94
		gucuggcauacugaagcguaccgcccguaccgggcgguacguacugguccgaugggguuucgguacgcggaccgcccgcuaccgggugguacgccccaaaaaccccguaucgagcgauacggggcaaaucgggcgguaacggucgaaauuucgaccguuaccgucuggcaccacucgauaccgcccgguaccacucgguaacgguugaaaucgaucguuaccgcacuguagcagugcuacaguguuccagcg	-146.6	253	81606	60	40	-57.9	0.96
		ggugccaccugggcgcucgcccgagcccaggcgccaggugcuucgagcgagcgccuaggguaaacca	-43.3	67	20507	70	30	-64.6	0.92
	229	gcucagauugcguuuaagcgcaguuugcgucuaagcgcaguuuacgucuaaacucagauugcgucuaagcgcagauugcguuuaagcgcaguuugcgucugagcg	-58.1	105	33923	50	50	-55.3	1.1
	224	ucaaaaagugaaugucauuauaucacagagaugucuucuauguauguaucgaauuuugcaucggacaaaagcauugaucaucaguauaugggagcuauauucuaccaauggggauuuucaaguacugcagaauuuuguugucuuuuaaaggcacuuugauugaucuaaugggacuuucaagugcaaguaccagugagucccaugggggguuuggaucaaucaaagugccuuuaaaagacaag	-78.9	244	78679	38	62	-32.3	0.85

	221	ucucacauuuuaaauauuaguguugugguaauauaucaaauucacucaucaaauauugaggauuugauaucucauuucauauauauuuaaagaaaaaacuguauggaucguauggaaggga	-23.6	121	38948	26	74	-19.5	0.75
2	255	cggugcaaccuccugacugaggcgguugcaccgcucagccagugcucggagaucgagcccaggcggugccaccuccugucaggggagguugcaccgcccagucuugcucggagacugagccccaagcgggugccacugcuuggcuggagcgguugcaccu	-94.4	158	51104	66	34	-59.7	0.90.
		gguacuaccagacuggguggugguaccacuggcaguuaaugcugcuaggcgguaguaccgcccuacuaggugguggca	-44.2	78	23953	58	42	-56.6	0.97
	253	cgguugcaccgccuagcagagcucgaagucugagccucagggcggugcuaccuccugucaggggcgguugcaccucuugacagagcucgaagaucgagcucaggcgguugcaccugcugacugaggcgguugcaccgccuggcagagcucgaagaccgagcucaagcggugccaccucuugucaggggagguugcaccgcccagccucgcucggagacugagcccaggugguugccaccgccuggcuggagcgguuucaccucccagucuagcucggagacugagcccaggugguugcaaccu	-174.7	301	97261	65	35	-57.9	0.89
	250	agggacucccuuguggcuaggaauccuaggucguagggcugucuuguguggcuaggagacccuagccacuagggaguccu	-48.6	80	26010	58	42	-60.75	1.04
		cccuuguggcuaggaauccuaggucguagggcugucuuguguggcuaggagacccuagccacuagggaguccuuuguagcuaggaaucccuagcuacagggc	-60.9	102	31197	56	44	-59.7	1.06
	245	ucuuuaugcaaaacugcauauaauauuuggcacacacuugaaaucacacacaaaggcacuaguagaguuaaaaacucuaaaauaaaaucuuuugauguauugauuuuaaguuguuucauaugaaacaagcga	-32.1	131	42148	29	71	-24.4	0.84
		uggcacacacuugaaaucacacacaaaggcacuaguagaguuaaaaacucuaaaauaaaaucuuuugauguaugauuuuaaguuguuucauaugaaacaaagcgaaauuguuggagacauguacacuuauuuuacggaugucgucaaugauuuaaaaagucuuaguaaaauauuuuuucuaauuuugaacuuguuaacaaaa	-43.4	200	60845	28	72	-21.7	0.77
		gggccauggauucgagcaucgagccaagaagaacagauauugcaucaaggauauuggaguuacggaagccaacugguagauuaggcaauaggccacaagagaggacgaugcaucgaagaau	-23.4	121	37424	47	53	-19.3	0.41
	243	cagucuccgagauugugucaagcggugguaccaccagucuggguggugguaccacccagauucagucuccaagacua	-42.9	77	24924	55	45	-55.7	1.01
		uaccgccuagugucaggcuacaggcagugguaccgccuagugcaggcggugguacugcccguacccugaaauuuuggggguu	-47.7	82	25087	59	41	-58.1	0.98

3	276	uguuaggacucuagcuaggagagucuuaauugagagggacauucuguuaggacucuaguuaggagaguccuaauu	-33.8	75	24392	41	59	-45	1.09
	278	acuucauuaccuuuauugaugauuucucaagguauggauauguguacuuaaugaaguacaaguccgaggccuuugagaaauucagagaguauaagaauaaagug	-29.6	104	33653	34	66	-28.4	0.83
4	289	uucuguuaggacucuagcuaggagaguccuaauugagagggacacucuauuaggacucuaacuaggagaguccuaauugaaa	-43.9	82	26607	41	59	-53.5	1.30.
		aggcagugauaccgccagaccuaggcgaugguaccacuagaccugggcgggugguaucgccaggguagugucaacgucagacugacacuggacgggugguacugccc	-54.8	106	32563	60	40	-51.6	0.86
	291	ccgauguucagaugauguaccggaagugugucgggaacucucggcgagcucccgacacacuuccgguacaucauuugaaucuuuga	-62.4	86	27764	51	49	-72.5	1.42
		ggcggugccaccgcccgaccucucgggugcugggcggugccaccgcucaguccaguaguggccaccacucaguucucgggguucugggcggugccaccgccu	-65.1	100	30436	71	29	-65.1	0.91
		gcaccgccuggcagagcucgaagaccgagcucaggcggugcuaccucucugccagggagauugcaccgcccagucuugcucgaagacugagcucaggcggugccaccucucugucagggaggguugcaccgcccagucuagcucgaagacugagcucaggcgguguu	-65.1	166	50775	63	37	-39.2	0.62
5	319	gcaaguggaaugguguacauaugaaugguuuaaucacuugcgaccauucauaugucacaccauuccacuuga	-51.5	73	23564	41	59	-70.5	1.71
	315	cucuccugacaggagguagcaccgccugagcucggucuucgagcucuggcaggcagugcaaccgcccuagucaggaggu	-47.8	79	25620	63	37	-60.5	0.96
	309	agcauagccucccaggugguguacugcccagguuggggcgguguaccgccuaguguacaacuggcacuggcaguguaucgcccagcgcaggcgguguaucgcucgug	-59.7	110	35671	64	36	-54.2	0.84
	305	gacacagucugagagacuguguuugggcgguacaacugcccaaucuggcgguacuaccaccugacacagucuuagagauugugucu	-51.5	86	27827	51	49	-59.8	1.17
	304	aggcgguugcaccucuugacuggagagguugcaccgccuggcagagcucgaagaccgagcucaggcggugccaccucucugucagggagguugcaccgccc	-72.9	101	32679	65	35	-72.1	1.1
6	321	ugaauaaaagaaagaagauaagaacaauuauugaagaaguuuuauugaagaaaugaaaaugcuucaauacauuaacauguucccucuccuauuuacacaaauuaggagagaaaguauuucccucgauagaauagaggauuuuucugagauuucc	-34.9	154	49693	29	71	-22.6	0.77

	329	gauaucucaucuuaaaguaguuaaaagaauucuuagauaucuuaaagguaccauuaaucu aggauuauaguaucuaaaaucugagaacuuugaacuaauugcauaugcugauguguauuu ugcuggauauagauuagauagaaaaagcacauccggaauaugucaauuuuuaggacacgc acuuguuucuuggucuuccaagaaaca	-52	207	66532	31	69	-25.1	0.80.
	335	gcucaggcgguugcaccuccugacuggggcgguugcaccgccuggcagagcucgaagacc gagcucaggcggugccaccuccugucaagggagguugcaccgcccagucucgaucggaga cugggcccaggcggugccaccgccuggcuggagcgguugcacu gcccagucucacucggagacugagcccaggcggugcaaccuccuaccuuggggcaguucaac ccgcuuggca	-157	233	75276	67	33	-67.3	1
	341	ucggucgccuugguuuaccuaggcgcucgcccgaagcgcccagcgccugggcucgggcga gcgcccaggccaaacaggcgccgc	-58.2	84	27230	74	26	-69.2	0.93
		cgaggcgcucgggccucaccucaccucgcccgagcgccuaggcgagcgcccgagcgccuag gcgagcgcccgagcgccuuu	-50.8	81	24691	77	23	-62.7	0.81
	345	uaacuagggaugagagguuauaggccucaaguugauucaaacuuggagccuaaaaaaguc ucauccuggguuu	-32.3	73	23707	41	59	-44.2	1.07
		guugcccgcccucgaucugcgugccuucggcuacgcuacacgcacgcagaucgagggcag caacu	-45.1	65	19731	65	35	-69.3	1.06
	347	ggcggugccaccgcccagcgcucgggugcugggcggugcaaccgcccagcccaggaggug ucaccgccuagcucucgggugcugggcggugcuacugccc	-81.9	100	32453	75	25	-81.9	1.09
7	361	uggcggugccaccgcccagcgcucgggugcuggacggugcaaccgcccagccaaggaggu gucaccgcucagcacucgggugcuggacggugccaucgcuu	-70.9	101	32775	70	30	-70.1	1
		gccugggcugggcgguggcaccgcccagcacccgagagcugggcggugacaccuccuagg cugggcgguugcaccguccagcacccgagcgcugggcgguggcaccgccagcaucggga	-86.3	119	36630	75	25	-72.5	0.96
		cgguggcaccgccagauccggcgguagcaccgccugagcgcucgggcgccaggcggugac accgcuuggguuuggcgguggcaccga	-73.5	86	26444	73	27	-85.4	1.16
		gcgguggcaccgccagauccggcgguagcaccgcuugagcgcucgggagccaggcgguga caccgcucgguuuggcgguggcaccga	-66.4	87	26798	72	28	-76.3	1.05
	356	cagugcugggugguccacugccuagucaagcggugccacugcccagcucucgggugcug ggcggugccaccgcccagccuaggcggugccaccgccuaggcua	-74	104	33669	70	30	-71.1	1.01
	353	aucuaacauuggaccagagagcaucauguaagauuuugcacuaauucacccccagaggaua gauucuggggggugaauuagugcaugauacucucugguccaacguguuaga	-72.2	111	35874	43	57	-65	1.51

		cuggaggugccaccgcccagccaagcggugccaccgcccagcgcucgggugcuggacggugcaaccgcccagccaaggaggugucaccgcccagcucucgggugcugggcggugccaccgccuggcucuccga	-91.5	133	40689	74	26	-68.7	0.92
8	378	agcaucuuucagcacaugucaaaugguugaagucccacaacaauucuacaguuccguaaacauauauauauauauauauauauauauauauauauauauauauauauauauauauauggacuggaccgagaagauaa	-41.2	137	44000	26	74	-30	1.15
	382	aaugcuucugauaagauccuaaagucuuaucguaaaaaugaagaagagaaaggggagaaauggggaugauaaugaucgcuucgaggggaucggccuuccuugaucucuucgaggggaucggcccuccuugaucgcuucgaagagaucaaggagggucgauccccuugaagcgaucauuaucaucccaauuucuucccuuucucuucuucuuuuuuacgauaagacuuuaggguuuuaucauuuggyau	-194.4	248	79717	43	57	-78.3	1.82
	388	uuuugacguaaacugcauaaacugaguuuagacguaaaacuauguuuuagacguaaaacuacguuuaaacguaaaacugcguuuagacgcaaacugcguuuagacguaaaacuacauuuagacguaaaacuguguuuagacguaaaacugcguuuagacgcuaacugcacuacgc	-56.9	175	56362	36	64	-32.5	0.90.
	390	gcuagcggugccaccgcccagcgcucggguguugggcggugcaaccgcccagcccaggaggugucaccgcccagcucucgggugcugggcggugccaccgccuag	-83.6	105	34067	74	26	-79.6	1.07
	396	caagaaugauguuuguacauaaagaaacugacaaauacccauuauuuguuaugaaucuuuaugcaacagaucuuca	-27.9	76	24474	29	71	-23.5	0.81
9	403	gcgauuaaaaccgaaagcugcguucguucaaacaaaacuauuaugaugcaaaaacuaauucucaguuuguaucuaaguucaguuugcgucuaagcgcagauugcgucuaagcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacgucuaaacgcaguuuuacgucuaaacgca	-83.6	238	76463	42	58	-35.1	0.83
		ccgaaagcugcguucguucaaacaaaacuauuaugaugcaaaaacuaauucucaguuuguaucuaaguucaguuugcgucuaagcgcagauugcgucuaagcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgu	-67.7	183	55659	42	58	-36.99	0.88
		gaaagcugcguucguucaaacaaaacuauuaugaugcaaaaacuaauucucaguuuguaucuaaguucaguuugcgucuaagcgcagauugcgucuaagcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuua	-66.3	178	54172	41	59	-37.2	0.9

		guuuguaucuaaguucaguuugcgucuaagcgcagauugcgucuaagcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacgucuaaacg	-60.4	154	46872	45	55	-39.2	0.87
		guuuguaucuaaguucaguuugcgucuaagcgcagauugcgucuaagcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacgucuaaacg	-60.4	154	46872	45	55	-39.2	87
		guucaguuugcgucuaagcgcagauugcgucuaagcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacg	-60	134	40816	47	53	-44.7	0.95
		caguuugcgucuaagcgcagauugcgucuaagcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuu	-59.2	128	38975	47	53	-46.25	0.98
		uugcgucuaagcgcagauugcgucuaagcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacgucuaaacgcagauuuuacgucuaaacgca	-62.8	155	47166	46	54	-40.5	0.88
		gcgucuaagcgcagauugcgucuaagcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacgucuaaacgcaguuuuacgucuaaacgca	-62.7	153	46586	46	54	-40.9	0.88
		uugcgucuaagcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacgucuaaacgca	-49.4	120	36499	46	54	-41.1	0.89
		gcgucuaagcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacgucuaaacgca	-49.3	118	35918	47	53	-41.77	0.88
		agcgcaguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacgucuaaacgca	-42.3	103	31337	47	53	-41	0.87
		uugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacgucuaaacgca	-33.5	86	26144	45	55	-38.9	0.86
		caguuugcgucuaaacacaguuugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuu	-40.3	94	28566	45	55	-42.8	0.95
		uugcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacgucuaaacgca	-33.5	86	26144	45	55	-38.9	0.86
		gcgucuaagcgcaguuugcgucuaaacgcagauugacgucuaaacgcaguuuuacgucuagacgcagauuuacgucuaaacgca	-33.4	84	25564	46	54	-39.7	0.86

	408	uaggucgaaacaccgguacgguauuacgaaccuugagauuauauauauauauauauauau auauauauauauauauauauauauauauauaucgaacaauauaccgcucggaauacaguau cguaccguacug	-40.3	133	42738	28	72	-30.3	1.08
	416	ggcgguggcaccgcccagcacccgagagcugggcggugaaaccuccugggcugggcgguu gcaccgcccagcaccugagagcugggcggugacaccuccuggguugggcgguugcacugc ccagcacccgagcgcugggcgguggcaccgcca	-119.3	153	49737	73	27	-77.9	1.06
	418	aggcgguggcaccgccugggcucagucuccgagcgagacugggcggugcaaucgccccua ucaggcgguggcaccgccuagagguucggucucugagcugccaagcgguuguaccgccc	-78.9	119	38586	68	32	-66.3	0.97
		ccuccugacugaggcgguugcaccgccuggcaaagcucaaagacugagcucaggcggugc caccuccugucaggaga	-51.6	77	23522	62	48	-67	1.08
	429	ggugauaccgucauggcacagucucugagacugugucaagcggugguaccgucggucuag gugaugguaccacucggacacagucucuaagaggcuacaggcggugguaccgcccagacu aggcggugguaccgcccagacuggguggugguacugccuagugguagggugcuaggcgg ugguaccg	-120.3	187	60806	60	40	-64.3	1.07
	430	ccgagcucaggcggugccaccucucugucugggagguugcaccgcccagucuugcucgaa gacugagcucaggcggugccaccucucugucagggagguugcaccgcccagucuagcucg a	-80.1	121	39024	64	36	-66.1	1.03
10	442	aacuucgaugaaaaucauuuucaagucaagugauaagggaugauuuucacuaugacagga gauaaagauuuucaacaugaauuuca	-20	86	27824	30	70	-26.1	0.86
		uugaaaauaucucuaaagcuuuuccuuuucucuuuucaacaagugauuguuagugaugaa ucuauuuuauguuauuucuugaauuaagaaaagugauuuugaugauuugaauuuaaugg guuuuaucaaaaucaugaucuugauuuuucgagauuuauaacuugaaaucu	-32.2	170	51362	24	76	-18.9	0.78
		gaaguaaaauugcuaacuuucacaucucuaaaacuugcuagcuuacauguucuaaagugu uguaaaacuuguauguuuuaaaaaacuugauaguugcacuuuucuccuuuuuguugaug acaaagggggagaaguuaugauaauuuaaaauuugaaucaugcaugguguacuuugauau uuugaaauuaugauuaugcauaaaauacucaugauuuuaugaugcaugcaaugauaaaau acucaugauuuuuugaauucaaaauuucuuuuaauaugucauauugauaggggaguuu aguuuaaacuccaucaucaauugguugucaucaucaaaaaggggagau	-88.1	347	105540	29	71	-25.3	0.87
	444	cagccaggcggugccaccgcccagccaggcggugccaccgcccagcgcucgggugcuggac ggugcaaccgcccaaccaaggaggugucaccacccagcacucgggugcugggcgguuuca ccgcccagccaagcgguuccaccgccuggcuu	-118	153	49408	72	28	-77.1	1.07

	445	agcggugccaccgcccagcacucgggugcuggaugguacaaccgcccagcccaggaggug ucaccgcccagcucucgggugcugggcggugccaccgcc	-74	99	32082	73	27	-74.74	1.02
		agcggugccaccgcccagcacucgggugcuggacgguacaaccgcccagcccaggaggug ucaccgcccagcucucgggugcugggcgaugccaccgcc	-74	99	32082	73	27	-74.74	1.02
	455	caccgcccagucugguggugcuacuaccugauagagccucagagauugagcucucguggu accaucgccugacugggcggua	-48.9	82	26523	59	41	-59.6	1.01
		cccgagagcuaggcgguagcaccgccugggcugggcgguggcaccgcccaacacugucag ugucugacauugacaaguggguggcaccgccacugacaggcgguggcaccgccagcaucgg ga	-82.6	122	37508	67	33	-67.6	1
	461	cuggcgguaccaccgcuuguguuggguggguaccacagcuuguagcacuaacacuaggugg uaccaccgcccagucuggcgguaccacugcccagauuggugguacc	-53.6	106	34229	59	41	-50.5	0.85
11	474	gcuggcggugccaccgcccagcgcucggguguuggggcggugcaaccgcccagcccaggag gugucaccgcccagcucucgggugcugggcggugccaccgcccagc	-88.8	106	34387	76	24	-83.7	1.1
	475	ggaagggagugaguuggggucgguuccaccgucggcccuagcuaaagguuaaaaagaggca gaauuguggggcggcggcugucacguguugcaagcgguugcaccgcuagcuggggcggug caaccgcuugcaacacgugacagccggaggugcuaccucca	-97.7	160	52148	61	39	-61.1	1
	479	ggcgguaccaucgccagugucagucugacauuaggugguaccaccaccugugcuggguggg uaccaccgccagugucagucugacacuaggcgguaccacca	-67.6	101	32644	59	41	-66.9	1.13
Note: LP (nt) – Length of pre-miRNA, MFE – Minimal folding free energy, Mol. wt. – Molecular weight, AMFE – Adjusted minimal folding free energy, MFEI – minimal folding free energy index									

Appendix II - List of all identified mature miRNAs

miRNA Family	Chr. No.	Scaffold	Sequence	Seq. no.	Score	E value	LM (nt)	Mol. Wt.	G+C%	Tm	A+T%
miR-6928-5p	1	224	ggggauuuucaaguacugca	110-129	73	0.96	20	6087.9	45	49.7	55
miR-4096-3p	1	241	ugcaagauagcuuucuugug	25-44	73	0.78	20	6025.8	40	47.7	60
miR-961-5p	1	224	uugaucaucaguauauggga	74-93	73	0.96	20	6072.9	35	45.6	65
miR-3343	1	232	ucgaucguuaccgcacuguagc	212-233	74	0.82	22	6602.2	55	56.7	45
miR-9786-3p	2	250	gaauccuaggucguaggg	14-31	72	0.96	18	5523.6	56	50.3	44
miR-1634	2	245	aagagaggacgaugcaucga	151-170	73	1.8	20	6196.1	50	51.8	50
miR-2164-3p	2	245	acacuugaaaucacacacaa	7-26.	73	0.77	20	6004.9	35	45.6	65
miR-2164-3p	2	245	acacuugaaaucacacacaa	34-53	73	1.1	20	6004.9	35	45.6	65
miR-7748a-3p	2	245	aauauuuuuucuaauuuugaac	167-188	74	0.63	22	6520.1	14	39.9	86
miR-968-3p	2	245	aucuuuugauguaugauuuuaa	59-80	74	0.63	22	6576.1	18	41.8	82
miR-968-3p	2	245	aucuuuugauguaugauuuuaa	86-107	74	0.88	22	6576.1	18	41.8	82
miR-7909-5p	2	245	ggcaugucauugggucaucgau	224-245	74	1.4	22	6683.3	50	54.8	50
miR-1a-2-5p	3	278	uauguguacuuaaugaagua	39-58	73	0.81	20	6057.9	25	41.5	75
miR-395d-3p	4	291	aagugugucgggaacucucgg	24-44	78	0.38	21	6432.1	57	56.3	43
miR-67	5	305	gacacagucuuagagauugugu	63-84	74	0.83	22	6691.3	41	51.1	59
miR-1199-5p	6	335	cugagcccaggcggugca	183-200	72	1.1	18	5497.6	72	57.2	28
miR-4783-5p.	6	341	gcgcccagcgccugggcu	36-53	72	1.2	18	5449.5	83	61.7	17
miR-7706	6	341	agcgccugggcucgggcgag	42-61	73	0.97	20	6172	80	64.1	20
miR-6312-5p	6	335	ggagacugagcccaggcggu	178-197	73	0.91	20	6180	70	60	30
miR-6312-5p	6	335	ggagacugggcccaggcggu	116-135	73	0.91	20	6196	75	62	25
miR-518f-5p	6	321	gagaaaguauuucccucgauag	108-129	74	1.1	22	6674.3	41	51.1	59
miR-3900-5p	7	353	caaggaggugucaccgcc	74-91	72	1.3	18	5481.6	67	54.9	33
miR-9112	7	353	ggaccagagagcaucaug	11-28.	72	1.1	18	5529.6	56	50.3	44
miR-H17	7	361	gcgcucgggcgccaggcgg	38-56	77	0.47	19	5857.8	89	66.2	11
miR-3745-3p	7	361	cugggcugggcgguggcacc	3-22.	73	0.95	20	6148.9	80	64.1	20
miR-5417	8	382	aggggagaaaugggggaug	51-68	72	1.2	18	5689.7	56	50.3	44
miR-9546-5p	8	396	auuuguuaugaaucuuua	44-61	72	1	18	5353.4	17	34.4	83

miR-1277-5p	8	378	auauauauauauauauauau	62-81	73	1.1	20	5971.8	0	31.3	100
miR-1277-5p	8	378	auauauauauauauauggac	102-121	82	0.2	20	6025.9	15	37.4	85
miR-1277-5p	8	378	uauauauauauauauaugga	101-120	73	1.1	20	6026.9	10	35.4	90
miR-1277-5p	8	378	uccguaaacauauauauaua	53-72	73	1.1	20	5984.9	20	39.5	80
miR-1277-5p	8	378	auauauauauauauauauaua	96-116	78	0.43	21	6285	0	32.9	100
miR-6312-5p	9	418	accgccugggcucagucucc	11-30.	73	0.95	20	5973.8	70	60	30
miR-67	9	430	ccucucugucugggagguug	20-39	73	0.97	20	6032.8	60	55.9	40
miR-1277-5p	9	408	auauauauauauauauauaua	71-91	78	0.42	21	6285	0	32.9	100
miR-1277-5p	9	408	auauauauauauauauauaucg	73-94	74	0.91	22	6590.2	9	38.1	91
miR-3900-5p	10	444	caaggaggugucaccacc	78-95	72	1.6	18	5465.6	61	52.6	39
miR-2172-5p	10	442	gaaauuaugauuaugcau	183-200	72	1.7	18	5438.5	22	36.7	78
miR-1419e-3p	10	444	ggugucaccacccagcac	84-101	72	1.6	18	5401.5	67	54.9	33
miR-167f-3p	10	444	gugcugggcgguuucacc	106-123	72	1.6	18	5452.5	67	54.9	33
miR-971-5p	10	442	uuugaugauuugaauuua	98-115	72	1.7	18	5393.4	17	34.4	83
miR-9773	10	442	uuuuauguuauuucuugaa	65-83	77	0.67	19	5620.5	16	36	84
miR-2c-5p	10	442	caucaucaauugguugucau	310-329	73	1.4	20	5969.8	35	45.6	65
miR-3745-3p	10	455	cugggcugggcgguggcacc	26-45	73	0.98	20	6148.9	80	64.1	20
miR-7666-5p	10	455	gcgguagcaccgccugggcu	13-32	73	0.98	20	6092.9	75	62	25
miR-3745-3p	10	444	ggugccaccgcccagccagg	10-29	73	1.3	20	6074.9	80	64.1	20
miR-2293	10	442	ugauuuugaugauuugaauu	94-113	73	1.4	20	6012.8	20	39.5	80
miR-449c-3p	10	442	uugauaguugcacuuuucuc	87-106	73	1.4	20	5923.7	35	45.6	65
miR-7b-5p	10	442	aagaaaagugauuuugaugauu	86-107	74	1.2	22	6724.3	23	43.7	77
miR-7990b	10	442	aaguuaugauaauuuaaaauuu	132-153	74	1.2	22	6630.2	9	38.1	91
miR-8217-5p	11	475	guuaaaaagaggcagaauug	47-66	73	1.3	20	6182	35	45.6	65
miR-4620	11	475	gucgguuccaccgucggcccua	19-40	74	1.1	22	6593.2	68	62.3	32

Note: Chr. No. – Chromosome number, LM (nt) – Length of mature miRNA, Mol. wt. – Molecular weight

Appendix III – List of all identified targets							
miRNA Family	**Target I.D**	**Expectation**	**UPE**	**miRNA Aligned Fragment**	**Target Aligned Fragment**	**Target Description**	**Inhibition**
miR-6928-5p	GSMUA_Achr8T03160_001	2.5	19.5	GGGGAUUUUCAAGUACUGCA	UGCAGCAUUUGAAGAUCCUC	FAD dependent oxidoreductase	Cleavage
miR-968-3p	GSMUA_Achr4T24940_001	3	15.8	AUCUUUUGAUGUAUGAUUUUAA	UGGAAAUGGUACAUCAAAAGAU	G-patch domain containing protein	Cleavage
miR-1634	GSMUA_Achr6T13230_001	3	18.5	AAGAGAGGACGAUGCAUCGA	UCCGUCCAUCGUCUUCUCUU	Glucan endo-1	Cleavage
miR-1634	GSMUA_Achr6T21820_001	2.5	13.1	AAGAGAGGACGAUGCAUCGA	CCGAUCUGUCGUCCUCUUUU	patatin-like phospholipase family protein	Cleavage
miR-1634	GSMUA_Achr5T21320_001	2	5.4	AAGAGAGGACGAUGCAUCGA	UUCAUGCAUCGUCUUCUCUU	pectinesterase	Cleavage
miR-1634	GSMUA_Achr1T07730_001	2.5	16.4	AAGAGAGGACGAUGCAUCGA	CCGAUGCAUCCUCUUUUCUU	Putative 1-acylglycerophosphocholine O-acyltransferase 1	Translation
miR-9786-3p	GSMUA_Achr5T23420_001	3	23.3	GAAUCCUAGGUCGUAGGG	GCCCAUGACCUGGGAUUC	DUF292 domain containing protein	Cleavage
miR-1199-5p	GSMUA_Achr2T01510_001	3	21.7	CUGAGCCCAGGCGGUGCA	UGGACUGACUGGGCUCAG	Putative E3 ubiquitin-protein ligase RF298	Translation
miR-4783-5p	GSMUA_Achr8T12070_001	3	13.5	GCGCCCAGCGCCUGGGCU	ACCCCAAGCGCUGGGCGU	C4-dicarboxylate transporter/malic acid transport protein domain containing protein	Cleavage
miR-4783-5p	GSMUA_Achr2T05810_001	3	13.5	GCGCCCAGCGCCUGGGCU	AGUCCAGGUGCUGAGUGC	ethylene-responsive protein related	Cleavage
miR-9112	GSMUA_Achr10T27090_001	3	23.6	GGACCAGAGAGCAUCAUG	CAUGAUGUUGUCUGGUUU	cytochrome c oxidase subunit 5B	Translation
miR-3900-5p	GSMUA_Achr2T17100_001	3	23.2	CAAGGAGGUGUCACCGCC	GGCGGCGGCACUUCUUUG	F-box/kelch-repeat protein At1g55270	Cleavage
miR-3900-5p	GSMUA_Achr7T08300_001	3	18.9	CAAGGAGGUGUCACCGCC	CGCGGUGAUGUCUUCUUG	Nitrate transporter 1.1	Cleavage
miR-3900-5p	GSMUA_Achr10T29560_001	3	18.3	CAAGGAGGUGUCACCGCC	GGCUGUGGCACCUCUUUA	plant viral response family protein	Cleavage
miR-3900-5p	GSMUA_Achr4T31030_001	3	13.7	CAAGGAGGUGUCACCGCC	GUCGGUGACAUCUCGUUG	Putative disease resistance protein RGA3	Cleavage

miR-3900-5p	GSMUA_Achr2T17480_001	3	21.1	CAAGGAGGUGUCACCGCC	CGCGGUGGCGUCUCUUUG	Putative Non-specific lipid-transfer protein-like protein At5g64080	Cleavage
miR-H17	GSMUA_Achr7T07440_001	3	24.8	GCGCUCGGGCGCCAGGCGG	UCACCUGGCGCACGAGCGC	Peroxisomal acyl-coenzyme A oxidase 1	Cleavage
miR-1277-5p	GSMUA_Achr3T03000_001	0	14.2	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	40S ribosomal protein S25	Cleavage
miR-1277-5p	GSMUA_Achr7T22580_001	0	20.9	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	40S ribosomal protein S27-2	Cleavage
miR-1277-5p	GSMUA_Achr3T13390_001	1	17.3	AUAUAUAUAUAUAUAUAUAU	AUGUGUAUAUAUAUAUAUAU	60S ribosomal protein L39	Cleavage
miR-1277-5p	GSMUA_Achr1T09610_001	3	3.75	AUAUAUAUAUAUAUAUAUAU	CUCUAUAUAUAUAUAUAUAC	autophagy protein 5	Cleavage
miR-1277-5p	GSMUA_AchrUn_randomT29220_001	0	10.8	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	Caffeic acid 3-O-methyltransferase	Cleavage
miR-1277-5p	GSMUA_Achr3T19300_001	0	9.6	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	Calcium-dependent protein kinase 7	Cleavage
miR-1277-5p	GSMUA_Achr6T07880_001	0	17.1	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	Cyclic nucleotide-gated ion channel 2	Cleavage
miR-1277-5p	GSMUA_Achr9T16070_001	0.5	21.3	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUGUAU	Cysteine proteinase 2	Cleavage
miR-1277-5p	GSMUA_Achr8T33070_001	2.5	12.5	AUAUAUAUAUAUAUAUAUAU	UUAUAUAUAUAUAUAUGUAA	Cytochrome P450 (Fragment)	Cleavage
miR-1277-5p	GSMUA_Achr1T07790_001	1	22.4	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAC	division protein	Cleavage
miR-1277-5p	GSMUA_Achr1T04110_001	3	20.4	AUAUAUAUAUAUAUAUAUAU	AUAUACAUAUAUGUAUAUGC	Electron transfer flavoprotein-ubiquinone oxidoreductase	Cleavage
miR-1277-5p	GSMUA_Achr3T14710_001	3	24.7	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAU-UAU	Endochitinase 1	Cleavage
miR-1277-5p	GSMUA_Achr10T13830_001	0	22.1	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	Glycerol-3-phosphate acyltransferase 6	Cleavage
miR-1277-5p	GSMUA_Achr5T00280_001	0	12.7	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	GPI mannosyltransferase 1	Cleavage
miR-1277-5p	GSMUA_Achr7T06610_001	0	11.6	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	HEAT repeat family protein	Cleavage

miR-6312-5p	GSMUA_Achr10T23630_001	0	22.1	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	hypro1	Cleavage
miR-1277-5p	GSMUA_Achr5T20410_001	0	17.6	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	KAZ1 - Kazal-type serine protease inhibitor precursor	Cleavage
miR-1277-5p	GSMUA_Achr8T21900_001	0	14.1	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	lumenal PsbP	Cleavage
miR-1277-5p	GSMUA_Achr8T26880_001	0	21.5	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	Meiotic recombination protein DMC1 homolog	Cleavage
miR-1277-5p	GSMUA_Achr8T25770_001	2.5	19	AUAUAUAUAUAUAUAUAUAU	AUGUAUAUAUAUAUAUGUGC	Metal ion binding protein	Cleavage
miR-1277-5p	GSMUA_AchrUn_randomT07890_001	0.5	21.9	AUAUAUAUAUAUAUAUAUAU	GUAUAUAUAUAUAUAUAUAU	MTA/SAH nucleosidase	Cleavage
miR-1277-5p	GSMUA_Achr7T26040_001	3	22.4	AUAU-AUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUUAUAU	Myosin-2 heavy chain	Cleavage
miR-1277-5p	GSMUA_Achr10T16960_001	2	10.1	AUAUAUAUAUAUAUAUAUAU	GUAUAUGUAUAUAUGUGUAU	Omega-hydroxypalmitate O-feruloyl transferase	Cleavage
miR-1277-5p	GSMUA_Achr10T15670_001	2.5	10.8	AUAUAUAUAUAUAUAUAUAU	AUACAUACAUAUAUAUAUAU	phospholipid-transporting ATPase	Cleavage
miR-1277-5p	GSMUA_AchrUn_randomT19820_001	0	18.8	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	Probable LRR receptor-like serine/threonine-protein kinase At1g56130	Cleavage
miR-1277-5p	GSMUA_Achr4T20720_001	1	17.9	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUGUAUAUGUAUAU	Protein STRUBBELIG-RECEPTOR FAMILY 5	Cleavage
miR-1277-5p	GSMUA_Achr5T26470_001	0	16.5	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	Putative acyl-CoA synthetase yngI	Cleavage
miR-1277-5p	GSMUA_Achr9T11970_001	0	10.5	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	Putative cysteine-rich receptor-like protein kinase 23	Cleavage
miR-1277-5p	GSMUA_Achr7T10760_001	0	24.5	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	Putative DNA repair protein radA homolog	Cleavage
miR-1277-5p	GSMUA_Achr6T02080_001	0	23.5	AUAUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAUAU	Putative Eukaryotic translation initiation factor 3 subunit A	Cleavage

miR-1277-5p	GSMUA_AchrUn_randomT27350_001	0	17.4	AUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAU	Putative Galactoside 2-alpha-L-fucosyltransferase	Cleavage
miR-1277-5p	GSMUA_Achr11T11700_001	3	13.9	AUAUAUAUAUAUAUAUAU	AUAUAUAUAUACAUACAUAU	Putative GDSL esterase/lipase At1g28590	Translation
miR-1277-5p	GSMUA_Achr2T20550_001	2.5	4.29	AUAUAUAUAUAUAUAUAU	CUAUAUAUAUAUGUAUAUAC	Putative Glutamyl-tRNA(Gln) amidotransferase subunit A	Cleavage
miR-1277-5p	GSMUA_Achr11T18190_001	2	12.2	AUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUGCAUAUAU	Putative GTP-binding protein ERG	Cleavage
miR-1277-5p	GSMUA_Achr1T07880_001	0	16.1	AUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAU	Putative Heat stress transcription factor B-2b	Cleavage
miR-1277-5p	GSMUA_Achr10T04780_001	0.5	20.4	AUAUAUAUAUAUAUAUAU	AUGUAUAUAUAUAUAUAU	Putative Histone deacetylase HDT1	Cleavage
miR-1277-5p	GSMUA_Achr9T22100_001	2	15	AUAUAUAUAUAUAUAUAU	AUAGAUUUAUAUAUAUAU	Putative LRR receptor-like serine/threonine-protein kinase FLS2	Cleavage
miR-1277-5p	GSMUA_Achr6T13130_001	0	18.8	AUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAU	Putative Momilactone A synthase	Cleavage
miR-1277-5p	GSMUA_Achr10T04400_001	0	21	AUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAU	Putative PHD finger protein At5g26210	Cleavage
miR-1277-5p	GSMUA_Achr11T23900_001	2.5	11.3	AUAUAUAUAUAUAUAUAU	AUAUCUAUAUGUAUAUAA	Putative respiratory burst oxidase homolog protein H	Cleavage
miR-1277-5p	GSMUA_Achr2T02860_001	0	10.8	AUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAU	Putative S-linalool synthase	Cleavage
miR-1277-5p	GSMUA_Achr3T08580_001	2.5	12.9	AUAUAUAUAUAUAUAUAUA	UGUAUGUAUGUAUGUAUGUAU	Putative Villin-2	Cleavage
miR-1277-5p	GSMUA_Achr6T19700_001	2.5	24.5	AUAUAUAUAUAUAUAUGGAC	AUAUAUAUAUAUAUAUAUAU	Retrotransposon-like protein	Cleavage
miR-1277-5p	GSMUA_Achr4T02810_001	0	23.9	AUAUAUAUAUAUAUAUAU	AUAUAUAUAUAUAUAUAU	Secologanin synthase	Cleavage
miR-1277-5p	GSMUA_AchrUn_randomT05500_001	3	16	AUAUAUAUAUAUAUAUAU	AUAUAUAUACAUAUAUACAU	SPX domain-containing membrane protein Os04g0573000	Translation
miR-1277-5p	GSMUA_Achr5T07900_001	3	16	AUAUAUAUAUAUAUAUAU	AUGUGUAUAUAUAUACAUGU	UBA/TS-N domain containing protein	Cleavage

miR-5417	GSMUA_Achr8T20910_001	2.5	5.45	AGGGGAGAAAUGGGGAUG	AAUCUCCGUUUCUCCUCU	Aspartyl-tRNA synthetase	Cleavage
miR-5417	GSMUA_Achr11T09120_001	3	7.55	AGGGGAGAAAUGGGGAUG	CAUAUUCAUUUUUCCUCU	Probable mannan synthase 9	Cleavage
miR-5417	GSMUA_Achr9T26610_001	3	12	AGGGGAGAAAUGGGGAUG	CAUCCCCAUUGCUCCUCA	F-box family protein	Cleavage
miR-5417	GSMUA_Achr2T13690_001	2	11.2	AGGGGAGAAAUGGGGAUG	CAUCCCUAUUUCUCCUCG	Cytochrome b-c1 complex subunit 6	Cleavage
miR-5417	GSMUA_Achr2T11770_001	3	8.69	AGGGGAGAAAUGGGGAUG	CAUCCCUCUUUCUCUCUU	GDSL esterase/lipase At5g33370	Translation
miR-5417	GSMUA_Achr11T19110_001	3	12.8	AGGGGAGAAAUGGGGAUG	CGCCUCCAUUUCUCUUCU	Glyceraldehyde-3-phosphate dehydrogenase	Cleavage
miR-5417	GSMUA_Achr10T08300_001	3	19.6	AGGGGAGAAAUGGGGAUG	CGCCGCCGUUUCUCCCCU	Protein TRIGALACTOSYLDIACYLGLYCEROL 2	Cleavage
miR-5417	GSMUA_Achr2T04710_001	1	0.52	AGGGGAGAAAUGGGGAUG	CCUCCCCAUUUCUCCCCU	Putative Pentatricopeptide repeat-containing protein At1g11290	Cleavage
miR-5417	GSMUA_Achr6T28410_001	3	22	AGGGGAGAAAUGGGGAUG	CAUCCUCUUUUUUCUCCU	Kelch repeat-containing protein At3g27220	Translation
miR-5417	GSMUA_Achr8T15940_001	3	12.5	AGGGGAGAAAUGGGGAUG	CGUUCUUGUUUCUUCCCU	kinesin motor domain containing protein	Cleavage
miR-5417	GSMUA_AchrUn_randomT14060_001	2.5	20.1	AGGGGAGAAAUGGGGAUG	UCUCCCUAUUUCUCCCCU	LOB domain-containing protein 15	Cleavage
miR-5417	GSMUA_Achr2T09970_001	3	10.9	AGGGGAGAAAUGGGGAUG	CACCACCAUUUCUCUCUU	Putative Dof zinc finger protein DOF4.6	Cleavage
miR-5417	GSMUA_Achr3T29920_001	3	17.7	AGGGGAGAAAUGGGGAUG	CGUCUCCGUCUCUCCCCU	Stress-associated endoplasmic reticulum protein 2	Translation
miR-9546-5p	GSMUA_Achr6T24240_001	3	17.9	AUUUGUUAUGAAUCUUUA	UUAAGCUUCAUAGCAAAU	Potassium channel AKT2/3	Cleavage
miR-5417	GSMUA_Achr1T03340_001	3	11.9	AUUUGUUAUGAAUCUUUA	UCAAGAUUUAUGAUGAAU	myosin	Cleavage
miR-9546-5p	GSMUA_Achr4T19460_001	3	19	AUUUGUUAUGAAUCUUUA	UAUGGAAUCAUAACAAAU	Probable galacturonosyltransferase 4	Cleavage
miR-9546-5p	GSMUA_Achr6T17770_001	3	20.4	AUUUGUUAUGAAUCUUUA	UGAACAUUCAUGGCAAGU	MLO-like protein 4	Cleavage

miR-9546-5p	GSMUA_Achr8T05700_001	3	16.9	AUUUGUUAUGAAUCUUUA	UGGGGAUUCAUAACAAUU	Protein phosphatase 2C	Cleavage
miR-9546-5p	GSMUA_Achr1T22610_001	3	15.1	AUUUGUUAUGAAUCUUUA	GAGAGAUUCAUAACAACU	dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDasubunit precursor	Cleavage
miR-9546-5p	GSMUA_Achr8T00660_001	3	12.5	AUUUGUUAUGAAUCUUUA	UAGUAAUUCAUAGCAAAU	Putative dnaJ domain containing protein	Cleavage
miR-9546-5p	GSMUA_Achr6T22440_001	2.5	15.7	AUUUGUUAUGAAUCUUUA	GAAAAAUUUAUAACAAAU	Putative Exocyst complex component 3	Cleavage
miR-9546-5p	GSMUA_Achr6T18560_001	2.5	7.58	AUUUGUUAUGAAUCUUUA	AAAACAUUUAUAACAAAU	vacuolar sorting protein	Cleavage
miR-9546-5p	GSMUA_Achr9T05410_001	2.5	16.9	AUUUGUUAUGAAUCUUUA	UCAAGAUUCAGAACAAAU	AAA-type ATPase family protein	Cleavage
miR-9546-5p	GSMUA_Achr9T08610_001	2.5	22.7	AUUUGUUAUGAAUCUUUA	UGAAGGUUCAUGACAAAG	WD40 repeat-containing protein SMU1	Cleavage
miR-67	GSMUA_Achr3T10480_001	2.5	20.2	CCUCUCUGUCUGGGAGGUUG	CAAUCUCCCAGUUAGAGAGG	Probable xyloglucan endotransglucosylase/hydrolase protein 33	Translation
miR-971-5p	GSMUA_Achr11T09270_001	3	20	UUUGAUGAUUUGAAUUUA	UAAGAUCAAGUCGUCAGA	20 kDa chaperonin	Cleavage
miR-449c-3p	GSMUA_Achr5T16420_001	3	12.8	UUGAUAGUUGCACUUUUCUC	UAGAAAACUGCAGCUAUCAA	ABC transporter C family member 13	Cleavage
miR-971-5p	GSMUA_Achr5T09680_001	2	12.2	UUUGAUGAUUUGAAUUUA	UAAACUCAGAUCAUCAGA	Beta-galactosidase 8	Cleavage
miR-971-5p	GSMUA_Achr3T22440_001	3	8.8	UUUGAUGAUUUGAAUUUA	UAUGUUCCAAUCAUCAAA	DNA polymerase alpha catalytic subunit	Translation
miR-9773	GSMUA_Achr6T19800_001	3	14.5	UUUUAUGUUAUUUCUUGAA	UUCCAGGAAUGACAUGAGA	SART-1 family protein	Cleavage
miR-971-5p	GSMUA_Achr8T21930_001	1.5	8.95	UUUGAUGAUUUGAAUUUA	UGACUUCAAAUCAUCAAA	Serine/threonine-protein kinase PBS1	Cleavage
miR-971-5p	GSMUA_Achr6T16510_001	3	14.4	UUUGAUGAUUUGAAUUUA	CAAAUUCAAGUCAUUGAG	Transport inhibitor response 1-like protein Os04g0395600	Cleavage
miR-971-5p	GSMUA_Achr3T06980_001	3	14	UUUGAUGAUUUGAAUUUA	CAAGUUCAGAUCAUUGAA	Glycoprotein 3-alpha-L-fucosyltransferase A	Cleavage
miR-971-5p	GSMUA_Achr8T28030_001	3	14.5	UUUGAUGAUUUGAAUUUA	UAAUUUUGAGUCAUCAGA	Histone-lysine N-methyltransferase ATX4	Cleavage

miR-971-5p	GSMUA_Achr3T13970_001	3	17	UUUGAUGAUUUGAAUUUA	UAGAAUCAAAUUAUCAAC	Protein argonaute 1A	Cleavage
miR-971-5p	GSMUA_Achr4T04830_001	3	15.3	UUUGAUGAUUUGAAUUUA	UAAAUUCCACUCAUCAAA	MCM10 - Putative minichromosome maintenance MCM protein 10	Translation
miR-91-5p	GSMUA_Achr5T21790_001	3	21.1	UUUGAUGAUUUGAAUUUA	UGCAUUCAGAUUAUCAAG	Probable DNA gyrase subunit A	Cleavage
miR-971-5p	GSMUA_Achr8T20820_001	2.5	17.7	UUUGAUGAUUUGAAUUUA	UAGAUUCAAAUCAUGAAG	Protein kinase APK1A	Cleavage
miR-449c-3p	GSMUA_Achr7T02660_001	3	15.6	UUGAUAGUUGCACUUUUCUC	GACAAAGGAGCAACUAUCAA	GYF domain containing protein	Cleavage
miR-9773	GSMUA_Achr6T10830_001	3	12.1	UUUUAUGUUAUUUCUUGAA	UCCAAGAAAAAACAUAGAA	pumilio-family RNA binding repeat containing protein	Translation
miR-2293	GSMUA_Achr7T02560_001	3	15.9	UGAUUUUGAUGAUUUGAAUU	AAGUCAGAUUAUCAGAGUCA	Putative Zinc finger protein VAR3 chloroplastic	Cleavage
miR-2172-5p	GSMUA_Achr5T24060_001	3	17.4	GAAAUUAUGAUUAUGCAU	AUG-AUAAUCAUGAUUUU	Putative Ethylene-responsive transcription factor 1	Cleavage
miR-971-5p	GSMUA_Achr9T27850_001	3	5.81	UUUGAUGAUUUGAAUUUA	CAAACUCAAGUCAUCGAA	Putative fasciclin domain containing protein	Cleavage
miR-971-5p	GSMUA_Achr3T20360_001	3	12.5	UUUGAUGAUUUGAAUUUA	UGAGUUCAAUUCAUCAGA	Putative GPI mannosyltransferase 3	Translation
miR-971-5p	GSMUA_Achr6T11790_001	2.5	15.9	UUUGAUGAUUUGAAUUUA	UAAGUUCAGAUCAUGAAA	chitin-inducible gibberellin-responsive protein	Cleavage
miR-3900-5p	GSMUA_Achr5T19730_001	2	10.9	CAAGGAGGUGUCACCACC	GGUCUUGACACCUCCUUG	Armadillo repeat-containing kinesin-like protein 1	Cleavage
miR-2172-5p	GSMUA_Achr10T12230_001	3	20.8	GAAAUUAUGAUUAUGCAU	AUGCCUAAUCAUGAUUAC	DAG protein	Cleavage
miR-3900-5p	GSMUA_Achr8T05320_001	2.5	16.6	CAAGGAGGUGUCACCACC	GGUGGUGAUAACUCUUUG	Dead box ATP-dependent RNA helicase	Cleavage
miR-3900-5p	GSMUA_Achr7T05820_001	3	16.9	CAAGGAGGUGUCACCACC	AAUGGUGACACUUUCUUG	Heat shock 70 kDa protein	Cleavage
miR-3900-5p	GSMUA_Achr9T18240_001	3	20.6	CAAGGAGGUGUCACCACC	GGCUGUGAUGCCUCCUUG	Peroxidase 65	Cleavage
miR-3900-5p	GSMUA_Achr9T02950_001	3	22.1	CAAGGAGGUGUCACCACC	GGUGGUGGCGUUUCCUUU	Pleiotropic drug resistance protein 3	Cleavage

miR-1419e-3p	GSMUA_Achr11T12290_001	2.5	22.9	GGUGUCACCACCCAGCAC	GAGCUUGGUGGUGACACC	Probable ATP-dependent RNA helicase DHX35	Cleavage
miR-167f-3p	GSMUA_Achr11T17440_001	3	19.5	GUGCUGGGCGGUUUCACC	CGUCAAACUGCCCAGUAC	Probable E3 ubiquitin-protein ligase XBOS33	Cleavage
miR-1419e-3p	GSMUA_Achr3T32320_001	2.5	20.3	GGUGUCACCACCCAGCAC	GUCUUGGGUGGUGGCACU	Protein HOTHEAD	Cleavage
miR-1419e-3p	GSMUA_Achr10T15580_001	2.5	23.1	GGUGUCACCACCCAGCAC	GUGCAGGGUGGUGACAUG	Protein SELF-PRUNING	Cleavage
miR-3900-5p	GSMUA_Achr9T18150_001	3	11	CAAGGAGGUGUCACCACC	GCUGGUGAUAUUUUCUUG	Putative KH domain-containing protein At4g18375	Cleavage
miR-1419e-3p	GSMUA_Achr3T14550_001	2.5	20.3	GGUGUCACCACCCAGCAC	GUGGUGGGUGGUGACAGC	Putative Lipase member N	Cleavage
miR-3900-5p	GSMUA_Achr8T01710_001	2.5	21.3	CAAGGAGGUGUCACCACC	GGUGGUGGCACUUCCGUG	Receptor-like protein kinase FERONIA	Cleavage
miR-3745-3p	GSMUA_Achr6T07200_001	3	22.6	CUGGGCUGGGCGGUGGCACC	GGCGCCACCGUCCAGCUCAU	exo70 exocyst complex subunit	Cleavage
miR-8217-5p	GSMUA_Achr4T23740_001	3	17	GUUAAAAAGAGGCAGAAUUG	CAAUUAUGCUUCUUUUUGAA	3'-5' exonuclease	Cleavage
miR-8217-5p	GSMUA_Achr11T06020_001	3	10.1	GUUAAAAAGAGGCAGAAUUG	AAAUUCUGUCUCUUCUUAAC	Putative U-box domain-containing protein 44	Cleavage

Appendix IV – List of all annotated target sequences				
S. No.	**Target Name**	**Target I.D**	**Target function**	**GO annotation (Biological process)**
1	FAD dependent oxidoreductase	GSMUA_Achr8T03160_001	Oxidoreductase, Electron transport, respiratory chain, transport	
2	G-patch domain containing protein	GSMUA_Achr4T24940_001	RNA binding, mRNA processing, mRNA splicing	
3	Glucan endo-1	GSMUA_Achr6T13230_001	Hydrolase activity	
4	patatin-like phospholipase family protein	GSMUA_Achr6T21820_001	Hydrolase activity	Lipid catabolic process
5	Pectin esterase	GSMUA_Achr5T21320_001	Enzyme inhibitor activity, pectin esterase activity, aspartyl esterase activity	Cell wall modification, pectin catabolic process
6	Putative 1-acylglycerophosphocholine O-acyltransferase 1	GSMUA_Achr1T07730_001	Acyl transferase activity, calcium ion binding	cellular lipid metabolic process
7	DUF292 domain containing protein	GSMUA_Achr5T23420_001	Protein transport	
8	Putative E3 ubiquitin-protein ligase RF298	GSMUA_Achr2T01510_001	Catalytic activity	
9	C4-dicarboxylate transporter/malic acid transport protein domain containing protein	GSMUA_Achr8T12070_001	Malate Transmembrane transport activity	
10	ethylene-responsive protein related	GSMUA_Achr2T05810_001	DNA binding,	Transcription regulation

11	cytochrome c oxidase subunit 5B	GSMUA_Achr10T27090_001	Cytochrome c oxidase activity	Mitochondrial electron transport, Mitochondrial ATP synthesis
12	F-box/kelch-repeat protein At1g55270	GSMUA_Achr2T17100_001	Ubiquitin protein transferase activity	Protein ubiquitination
13	Nitrate transporter 1.1	GSMUA_Achr7T08300_001	Nitrate transmembrane transporter activity	Oligopeptide transport, transmembrane transport
14	plant viral response family protein	GSMUA_Achr10T29560_001	Hypersensitive response, plant defense, ATP binding	
15	Putative disease resistance protein RGA3	GSMUA_Achr4T31030_001		Response to stimulus
16	Putative Non-specific lipid-transfer protein-like protein At5g64080	GSMUA_Achr2T17480_001	Lipid transport	
17	Peroxisomal acyl-coenzyme A oxidase 1	GSMUA_Achr7T07440_001	Acyl CoA oxidase activity, FAD binding	fatty acid beta-oxidation
18	40S ribosomal protein S25	GSMUA_Achr3T03000_001	RNA binding, translation initiation	
19	40S ribosomal protein S27-2	GSMUA_Achr7T22580_001	Metal ion binding, structural constituent of ribosome	Translation
20	UBA/TS-N domain containing protein	GSMUA_Achr5T07900_001	Transferase, Ubl conjugation pathway	
21	Putative Eukaryotic translation initiation factor 3 subunit A	GSMUA_Achr6T02080_001	Protein biosynthesis, initiation factor, RNA binding	translation regulation
22	60S ribosomal protein L39	GSMUA_Achr3T13390_001	RNA binding, structural component of ribosome	Translation
23	autophagy protein 5	GSMUA_Achr1T09610_001	Apoptosis, Autophagy, immunity	

24	Caffeic acid 3-O-methyltransferase	GSMUA_AchrUn_randomT29220_001	Methyl transferase activity, protein dimerization activity	Methylation
25	Calcium-dependent protein kinase 7	GSMUA_Achr3T19300_001	Serine/threonine protein kinase activity, ATP binding, calmodulin binding, Calcium ion binding	Abscisic acid-activated signalling pathway, peptidyl-serine phosphorylation
26	Cyclic nucleotide-gated ion channel 2	GSMUA_Achr6T07880_001	Ion channel activity	Ion transmembrane transport
27	Cytochrome P450 (Fragment)	GSMUA_Achr8T33070_001	Oxidoreducatse activity, heme ion binding, iron ion binding	Oxidation-reduction process
28	division protein	GSMUA_Achr1T07790_001	Cell division, Cell cycle	
29	Cysteine proteinase 2	GSMUA_Achr9T16070_001	Cysteine type endopeptidase activity	immune response, proteolysis
30	Electron transfer flavoprotein-ubiquinone oxidoreductase	GSMUA_Achr1T04110_001	Electron transport activity, Oxidoreductase activity, metal ion binding	Leucine catabolic process, response to absence of light, respiratory electron transport chain
31	Endochitinase 1	GSMUA_Achr3T14710_001	Chitin binding, chitinase activity	Polysaccharide catabolic process, chitin catabolic process, defense response, cell wall catabolic process
32	Glycerol-3-phosphate acyltransferase 6	GSMUA_Achr10T13830_001	Phosphatase activity, glycerol-3-phosphate O-acyltransferase activity	dephosphorylation, cutin biosynthetic process
33	GPI mannosyltransferase 1	GSMUA_Achr5T00280_001	Glycosyltransferase activity, transferase activity	GPI anchor biosynthetic process

34	HEAT repeat family protein	GSMUA_Achr7T06610_001	Ribosome biogenesis, rRNA processing	Transcription, ribonucleoprotein, transcription regulation
35	hypro1	GSMUA_Achr10T23630_001	Protein glycosylation, glycosyltransferase activity	
36	lumenal PsbP	GSMUA_Achr8T21900_001	Calcium ion binding	Photosynthesis
37	Putative acyl-CoA synthetase yngI	GSMUA_Achr5T26470_001	Catalytic activity	Metabolic process
38	Putative cysteine-rich receptor-like protein kinase 23	GSMUA_Achr9T11970_001	Protein kinase activity, ATP binding	Protein phosphorylation
39	Putative DNA repair protein radA homolog	GSMUA_Achr7T10760_001	DNA binding, ATP binding, DNA dependent ATPase activity	DNA repair
40	Meiotic recombination protein DMC1 homolog	GSMUA_Achr8T26880_001	DNA binding, ATP binding, DNA dependent ATPase activity	DNA repair, chiasma assembly
41	Metal ion binding protein	GSMUA_Achr8T25770_001	Metal ion binding	Metal ion transport
42	Putative PHD finger protein At5g26210	GSMUA_Achr10T04400_001	Metal ion binding, histone binding	Transcription regulation
43	Putative respiratory burst oxidase homolog protein H	GSMUA_Achr11T23900_001	Peroxidase activity, calcium ion binding, oxidoreductase activity	Oxidation reduction process
44	Putative S-linalool synthase	GSMUA_Achr2T02860_001	Lyase activity	
45	Putative Villin-2	GSMUA_Achr3T08580_001	Actin filament binding	Actin filament bundle assembly
46	Retrotransposon-like protein	GSMUA_Achr6T19700_001	Nucleic acid binding, zinc ion binding	
47	Secologanin synthase	GSMUA_Achr4T02810_001	Monooxygenase activity, oxidoreductase activity	Oxidation reduction process

48	SPX domain-containing membrane protein Os04g0573000	GSMUA_AchrUn_randomT05500_001		Transmembrane transport
49	Putative Galactoside 2-alpha-L-fucosyltransferase	GSMUA_AchrUn_randomT27350_001	Fucosyltransferase activity	Cell wall biogenesis, membrane
50	Putative GDSL esterase/lipase At1g28590	GSMUA_Achr11T11700_001	Hydrolase activity	
51	Putative Glutamyl-tRNA(Gln) amidotransferase subunit A	GSMUA_Achr2T20550_001	Amidase activity	
52	Putative GTP-binding protein ERG	GSMUA_Achr11T18190_001	RNA binding, GTP binding	
53	Putative Heat stress transcription factor B-2b	GSMUA_Achr1T07880_001	DNA binding	Transcription regulation
54	Putative Histone deacetylase HDT1	GSMUA_Achr10T04780_001	Nucleic acid binding	
55	Putative LRR receptor-like serine/threonine-protein kinase FLS2	GSMUA_Achr9T22100_001	ATP binding, serine/threonine kinase activity	
56	Putative Momilactone A synthase	GSMUA_Achr6T13130_001	Oxidoreductase activity	Oxidation reduction process
57	Probable LRR receptor-like serine/threonine-protein kinase At1g56130	GSMUA_AchrUn_randomT19820_001	ATP binding, serine/threonine protein kinase activity	Protein phosphorylation
58	KAZ1 - Kazal-type serine protease inhibitor precursor	GSMUA_Achr5T20410_001	Protease inhibitor	

59	MTA/SAH nucleosidase	GSMUA_AchrUn_randomT07890_001	Amino acid biosynthesis, methionine biosynthesis, Hydrolase	
60	Protein STRUBBELIG-RECEPTOR FAMILY 5	GSMUA_Achr4T20720_001	Protein kinase activity, ATP binding	Protein phosphorylation
61	Cytochrome b-c1 complex subunit 6	GSMUA_Achr2T13690_001	Ubiquinol cytochrome c reductase activity	Mitochondrial electron transport
62	Stress-associated endoplasmic reticulum protein 2	GSMUA_Achr3T29920_001		Protein glycosylation, endoplasmic reticulum unfolded protein response
63	Aspartyl-tRNA synthetase	GSMUA_Achr8T20910_001	ATP binding, aspartyl tRNA amino acylation, nucleic acid binding	tRNA amino acylation for protein translation
64	Putative Pentatricopeptide repeat-containing protein At1g11290	GSMUA_Achr2T04710_001	RNA modification, endonuclease activity, zinc ion binding,	RNA modification
65	Putative Dof zinc finger protein DOF4.6	GSMUA_Achr2T09970_001	DNA binding	Transcription regulation
66	F-box family protein	GSMUA_Achr9T26610_001	Carbohydrate binding, DNA damage, DNA Repair, Unfolded protein response	
67	GDSL esterase/lipase At5g33370	GSMUA_Achr2T11770_001	Hydrolase activity, alpha-L-fucosidase activity	Metabolic process
68	Glyceraldehyde-3-phosphate dehydrogenase	GSMUA_Achr11T19110_001	Glyceraldehyde-3-phosphate dehydrogenase activity, phosphate ion binding, NAD binding	Oxidation-reduction process, response to anoxia, glucose metabolic process
69	Kelch repeat-containing protein At3g27220	GSMUA_Achr6T28410_001	Ubiquitin protein ligase binding	Regulation of proteolysis

70	kinesin motor domain containing protein	GSMUA_Achr8T15940_001	Microtubule Motor activity, microtubule binding, ATP binding, ATPase activity	Microtubule based movement
71	LOB domain-containing protein 15	GSMUA_AchrUn_randomT14060_001	Developmental protein	
72	Protein TRIGALACTOSYLDIACYLGLYCEROL 2	GSMUA_Achr10T08300_001	Lipid transport activity, phospholipid binding	intracellular lipid transport
73	dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDasubunit precursor	GSMUA_Achr1T22610_001	Transferase activity	Plant type cell wall organisation, Protein glycosylation
74	AAA-type ATPase family protein	GSMUA_Achr9T05410_001	ATP binding, ATPase activity	
75	MLO-like protein 4	GSMUA_Achr6T17770_001	Calmodulin binding	Defense response, response to biotic stimulus
76	myosin	GSMUA_Achr1T03340_001	Actin binding, Calmodulin binding, Motor protein, Cell adhesion, cell shape	
77	Myosin-2 heavy chain	GSMUA_Achr7T26040_001	Actin binding, nucleotide binding, microfilament motor activity	pollen tube growth, organelle localization, actin filament organisation
78	Omega-hydroxypalmitate O-feruloyl transferase	GSMUA_Achr10T16960_001	Acyl transferase activity	
79	phospholipid-transporting ATPase	GSMUA_Achr10T15670_001	Lipid transport, hydrolase	

80	Potassium channel AKT2/3	GSMUA_Achr6T24240_001	Voltage gated potassium channel activity	Regulation of ion membrane transport, potassium ion membrane transport
81	Probable galacturonosyltransferase 4	GSMUA_Achr4T19460_001	Cell wall biogenesis, degradation, Glycosyltransferase	
82	Protein phosphatase 2C	GSMUA_Achr8T05700_001	Protein binding, metal ion binding, serine/threonine phosphatase activity	Protein dephosphorylation, Abscisic acid signalling pathway
83	Putative dnaJ domain containing protein	GSMUA_Achr8T00660_001	Oxidoreductase, Cell redox homeostasis, ubiquitin dependent ERAD pathway, response to endoplasmic reticulum stress	
84	Putative Exocyst complex component 3	GSMUA_Achr6T22440_001		Exocytosis
85	vacuolar sorting protein	GSMUA_Achr6T18560_001	Receptor, protein transport	
86	WD40 repeat-containing protein SMU1	GSMUA_Achr9T08610_001		RNA splicing
87	Probable xyloglucan endotransglucosylase/hydrolase protein 33	GSMUA_Achr3T10480_001	Xyloglucan activity, hydrolase activity	plant cell wall modification, cell wall biogenesis
88	20 kDa chaperonin	GSMUA_Achr11T09270_001	ATP binding, Transition metal ion binding, unfolded protein binding, Chaperone binding	Response to unfolded protein, positive regulation of SOD
89	ABC transporter C family member 13	GSMUA_Achr5T16420_001	Calmodulin binding, ATP binding, ATPase activity	Transmembrane transport
90	Protein Argonaute 1A	GSMUA_Achr3T13970_001	Nucleic acid binding	gene silencing by RNA

91	Histone-lysine N-methyl transferase ATX4	GSMUA_Achr8T28030_001	Methyl transferase activity	Methylation
92	Probable DNA gyrase subunit A	GSMUA_Achr5T21790_001	DNA binding, DNA topoisomerase type II activity	DNA topological change
93	MCM10 - Putative minichromosomal maintenance MCM protein 10	GSMUA_Achr4T04830_001	DNA replication origin binding, single-stranded DNA binding	
94	Beta-galactosidase 8	GSMUA_Achr5T09680_001	Beta galactosidase activity	Carbohydrate metabolic process
95	chitin-inducible gibberellin-responsive protein	GSMUA_Achr6T11790_001	Transcription regulation, Transcription	
96	Protein kinase APK1A	GSMUA_Achr8T20820_001	ATP binding, kinase activity	Protein phosphorylation
97	DNA polymerase alpha catalytic subunit	GSMUA_Achr3T22440_001	DNA binding, Nucleotide binding, nucleoside binding, DNA directed DNA polymerase activity	DNA biosynthetic process, DNA replication
98	Glycoprotein 3-alpha-L-fucosyltransferase A	GSMUA_Achr3T06980_001	Glycoprotein 3-alpha-L-fucosyltransferase activity	Protein glycosylation, ubiquitination
99	GYF domain containing protein	GSMUA_Achr7T02660_001	RNA binding, cadherin binding, mRNA destabilisation	Posttranscriptional gene silencing, cellular protein metabolic process, post embryogenic development.
100	pumilio-family RNA binding repeat containing protein	GSMUA_Achr6T10830_001	RNA binding	
101	Putative Ethylene-responsive transcription factor 1	GSMUA_Achr5T24060_001	Ethylene signalling pathway, DNA binding	transcription regulation

102	Putative fasciclin domain containing protein	GSMUA_Achr9T27850_001	Cell adhesion	
103	SART-1 family protein	GSMUA_Achr6T19800_001	Developmental protein, leaf development, root development	mRNA splicing
104	Serine/threonine-protein kinase PBS1	GSMUA_Achr8T21930_001	Nucleotide activity, protein kinase binding	Phosphorylation
105	Transport inhibitor response 1-like protein Os04g0395600	GSMUA_Achr6T16510_001	Inositol hexakisphosphatase binding	Auxin signalling pathway
106	Putative Zinc finger protein VAR3 chloroplastic	GSMUA_Achr7T02560_001	Metal ion binding	Chloroplast organisation, chloroplast mRNA modification
107	Putative GPI mannosyltransferase 3	GSMUA_Achr3T20360_001	Mannosyl transferase activity	Mannosylation, Pollen tube guidance, embryo development
108	Armadillo repeat-containing kinesin-like protein 1	GSMUA_Achr5T19730_001	Microtubule Motor activity, microtubule binding	Microtubule based movement, root development
109	Protein SELF-PRUNING	GSMUA_Achr10T15580_001	Transcription coregulatory activity	Negative regulation of flower development, vegetative to reproductive phase of transition
110	Receptor-like protein kinase FERONIA	GSMUA_Achr8T01710_001	Protein kinase activity, ATP binding, GTP binding	
111	Peroxidase 65	GSMUA_Achr9T18240_001	Peroxidase activity, metal ion binding, heme binding	response to oxidative stress, oxidation-

				reduction process, hydrogen peroxide catabolic process
112	Protein HOTHEAD	GSMUA_Achr3T32320_001	oxidoreductase activity, lyase activity	Cell-cell signalling, embryo sac development, fatty acid omega oxidation
113	Putative KH domain-containing protein At4g18375	GSMUA_Achr9T18150_001	RNA binding	
114	Putative Lipase member N	GSMUA_Achr3T14550_001	Hydrolase activity	Lipid catabolic process
115	Probable E3 ubiquitin-protein ligase XBOS33	GSMUA_Achr11T17440_001	Catalytic activity	
116	Pleiotropic drug resistance protein 3	GSMUA_Achr9T02950_001	ATP binding, Antibiotic resistance, cycloheximide resistance, transport	
117	Probable ATP-dependent RNA helicase DHX35	GSMUA_Achr11T12290_001	Helicase activity, ATP binding	
118	DAG protein	GSMUA_Achr10T12230_001	Actin binding, calcium ion binding, SH2 domain binding	Host virus interaction
119	Heat shock 70 kDa protein	GSMUA_Achr7T05820_001	Chaperone, ATP binding	Response to stimulus
120	Dead box ATP-dependent RNA helicase	GSMUA_Achr8T05320_001	Helicase activity, RNA binding	Transcription, transcription regulation
121	exo70 exocyst complex subunit	GSMUA_Achr6T07200_001	DNA binding	Exocytosis
123	3'-5' exonuclease	GSMUA_Achr4T23740_001	3' - 5' Exonuclease activity, metal ion binding, ribosome binding, rRNA binding	Metabolic process
124	Putative U-box domain-containing protein 44	GSMUA_Achr11T06020_001	Ubiquitin protein transferase activity	

Appendix V - CTAB Extraction Buffer

CTAB	-	2.5%
Tris – HCL (pH-8)	-	100 mM
EDTA	-	25 mM
NaCl	-	1.5 M
β-mercaptoethanol	-	0.2 % (v/v)
PVP	-	4% (w/v)