Abstract

MicroRNAs (miRNAs) are small single-stranded non-protein-coding RNAs of about 22 nucleotides that play an important role in post-transcriptional gene regulation in animals and plants by targeting mRNAs for direct cleavage of mRNAs or repression of mRNA translation. In this study, a homology search based approach has been used for identifying the novel miRNA genes of *Musa sp*. using the BLAST program with a total of 42,978 *Musa sp.* expressed sequence tags (EST) were compared to a total of 959 previously known plant miRNA sequences. Forty-seven candidate miRNA with homology to previously known miRNAs were identified. By using the mFold3.1 program, secondary RNA structures were predicted and total of nine novel miRNA were identified in *Musa sp*. One of the most promising novel miRNA was miR156 which has the strongest characteristics of a miRNA precursor structure and was 100% identical to the mature miR156 from other plant species. It was experimentally validated to be expressed in *Musa acuminata* var. *Berangan* and is the first known report of the identification and expression of miRNA from a banana plant species.

Abstrak

MikroRNAs (miRNAs) merupakan satu kumpulan rantaian tunggal yang bukan protein berkoding RNAs yang kecil lebih kurang ~22 nukleotida dan memainkan peranan yang penting pada permulaan transkripsi dalam pengawalaturan gen binatang dan tumbuhan dengan sasaran mRNAs melalui belahan terus mRNAs atau penindasan penterjemahan mRNA. Dalam kajian ini, berdasarkan pendekatan carian homolog dengan menggunakan program BLAST, jumlah sebanyak 42,978 jujukan EST *Musa sp.* telah dibandingkan dengan 959 jujukan miRNA tumbuhan yang telah diketahui awal untuk mecari calon-calon miRNA *Musa sp.* yang baru. Sebanyak empat puluh tujuh calon miRNA telah dikenalpastikan homolog dengan miRNA yang telah diketahui awal. Dengan menggunakan program mFold3.1, struktur sekunder RNA telah diramalkan dan sebanyak sembilan calon miRNA yang baru dikenalpastikan dalam *Musa sp.* Salah satu calon baru miRNA yang paling berpotensi ialah miR156 dimana ia mempunyai ciri-ciri struktur pelopor miRNA yang kuat, ia telah disahkan dengan ujikaji experimen bahawa miR156 terdapat dalam *Musa acuminata* var. *Berangan* dan jujukan matangnya menunjukkan 100% seiras kepada jujukan matang miR156 daripada tumbuh-tumbuhan spesies yang berlainan. Ini juga merupakan laporan kali pertama mengenai miRNA dalam species tumbuhan pokok pisang.

Acknowledgements

First and foremost I would like to offer my deepest gratitude to my supervisor, Asso. Prof. Dr. Jennifer Harikrishna, who has supported me throughout my thesis with her patience, guidance and knowledge, at the same time allowing me the room to work in my own way. The level of my Masters degree is attributed to her effort and encouragement, without which, my thesis could not have been written or completed. One could not wish for a more helpful and understanding supervisor.

Prof. Rofina Yasmin Othman has provided me with all of the EST sequences covered in this thesis, with consistent quality. I am deeply thankful for all her guidance and knowledge that was generously given to me.

I am also grateful for the support and guidance provided by my seniors. I would especially like to thank my senior Lee Wan Sin for his help in certain laboratory procedures and results. He has made his support available in a number of ways even with his tight work and travel schedule. It is also appropriate to show my appreciation here for my senior, Nadirah Md. Nasaruddin, who helped and guided me on the bioinformatic knowledge with patience.

I am as ever, especially indebted to my parents, Mr. Tan How Chee and Mrs. Ee Ooi Huat for their love and support throughout my life. I would also like to thank my sisters, Tan Yew Leng and Tan Yew Mei, as well as my brother, Tan Yew Chong, for their support, encouragement and understanding during the course of my study.

Finally, I offer my regards and blessings to all those who had supported me in any respect during the completion of my thesis.

Table of Contents

List of Figures

List of Tables

Abbreviations

%	- percent sign
º C	- degree Celsius
3'UTR	- 3' untranslated region
AGO	- ARGONAUTE
AMFE	- adjusted minimal folding free energy
BAC	- bacterial artificial chromosome
BLAST	- Basic Local Alignment Search Tool
cDNA	- complementary deoxyribonucleic acid
cm	- centimeter
CTAB	- cetyl trimethylammonium bromide
DCL	- Dicer-like
DEPC	- diethyl pyrocarbonate
DGCR8	- DiGeorge syndrome critical region gene 8
DNA	- deoxyribonucleic acid
dNTP	- deoxyribonucleotide triphosphate
dsRNA	- double-stranded RNA
EDTA	- ethylenediaminetetraacetic acid
EST	- Expressed Sequence Tags
et al.	- *et alia*
EtOH	- ethanol
G	- gram
i. e.	- *id est*
m	- meter
M	- molar

Mbp	- megabasepairs
MFE	- minimal folding free energy
MFEI	- minimal folding free energy index
$MgSO_4$	- magnesium sulphate
min	- minute
miRNA	- microRNA
miRNA*	- complementary of microRNA
ml	- milliliter
mM	- milimolar
mRNA	- messenger ribonucleic acid
MW	- molecular weight
NaCl	- sodium chloride
NaOAC	- sodium acetate
Nat-siRNA	- natural-antisense-transcript siRNA
NCBI	- National Center for Biotechnology Information
Nt	- nucleotide
PCR	- Polymerase Chain Reaction
Phenol-CI	- Phenol-chloroform-isoamylalcohol
pre-miRNA	- precursor miRNA
pri-miRNA	- primary transcripts
PTGS	- post-transcriptional gene silencing
PVP	- polyvinypyrrolidone
RISC	- RNA-Induced Silencing Complex
RNA	- ribonucleic acid
RNAi	- RNA interference
RNase	- ribonuclease

rRNA	- ribosomal ribonucleic acid
RT-PCR	- Reverse Transcription Polymerase Chain Reaction
siRNA	- small interfering ribonucleic acid
ssRNA	- single-stranded RNA
sp	- species
tasiRNA	- *trans*-acting siRNA
TBE buffer	- Tris-Borate-EDTA buffer
TE buffer	- Tris EDTA buffer
tRNA	- transfer ribonucleic acid
UV	- ultraviolet
µl	- micro liter

CHAPTER 1:

INTRODUCTION

1.0 Introduction

MicroRNAs (miRNAs) are small single-stranded non-protein-coding RNAs of about 22 nucleotides that play an important role in post-transcriptional gene regulation in animals and plants by targeting mRNAs for direct cleavage of mRNAs or repression of mRNA translation (Bartel, 2004; Zhang et al., 2005a; Yang et al., 2007; Qiu et al., 2007). miRNAs arise from larger precursors that are able to form self-complementary foldback structures. These precursors are then further processed by a ribonuclease III enzyme termed Dicer in animals or Dicer-like1 (DCL1) in plants (Dugas and Bartel, 2004; Bartel, 2004; Allen et al., 2005). According to the miRNA Registry Database (Release 9.2, May 2007) (http://microrna.sanger.ac.uk/), 4584 miRNA genes have been discovered in various organisms. Out of those, 959 miRNA genes have been predicted and identified in plants.

Although thousands of miRNAs have been found in recent years, only a small number of plant miRNAs have been discovered and functionally identified. Many other plant miRNA gene functions remain to be elucidated. However, many of those studied have shown that plant miRNAs play an important role and function in many processes, including organ development, phase changes, plant disease, signal transduction and response to environment stress (Zhang et al., 2006b; Zhang et al., 2006c; Yang et al., 2007).

The largest and most widely distributed group of *Musa sp* in Malaysia is derived from *Musa acuminata* and *Musa balbisiana* either alone or in various hybrid combinations. By identifying potential miRNAs in *Musa sp* expressed sequence tags (ESTs), the genomic knowledge on this important crop can be improved. Bananas are one of the most important staple food crops in our country and an important source of

income to millions of farmers. Bananas are rich in vitamins A, B6, C and are an important source of energy and fibre (Nelson et al., 2006). Recently, bananas have been considered as a useful tool to deliver edible vaccines or medicine (Nelson et al., 2006). Besides that, bananas can grow in perennial production systems, so that they maintain cover throughout the year and thus protect against land slides.

The principles of computational identification of microRNAs are based on the major characteristic features of miRNAs including high evolutionary conservation from species to species, hairpin-shaped stem loop secondary structures and high minimal folding free energy index. Homology search based approaches, such as the ones used in this study, can either be classified as genome-based searches or EST-based searches (Zhang et al., 2006d). Expressed sequence tags (ESTs) are partial sequences of expressed genes and cDNA fragments that are cloned into a plasmid (Zhang et al., 2005a; Matukumalli et al., 2004). Reverse Transcription PCR (RT-PCR) was used to confirm the expression of a predicted miRNA in bananas.

In Malaysia, detailed information on the miRNAs of *Musa sp* and the potential application are still lacking, while many key questions remain to be answered such as: What is the origin of miRNAs and what regulates miRNA expression? Therefore, the overall objective of this study is to identify the miRNAs in *Musa sp* to help bring new approaches and strategies for the study of miRNA roles in the regulation and development of *Musa sp*.

1.1 Objectives

With improvements and advances in technology and computational approaches over the years, many plant miRNAs have now been discovered and functionally identified (Zhang et al., 2006b; Zhang et al., 2006d; Xie et al., 2007; Yang et al., 2007; Qiu et al., 2007; Nasaruddin et al., 2007). However, there are many more plant miRNAs that remain to be discovered and elucidated especially for *Musa sp*. as there have not been any published research carried out on this crop species. Thus, the aims of this project are to identify potential miRNA sequences within *Musa sp* EST sequences, the development of a simple and rapid procedure to identify the miRNAs in *Musa sp*. as well as to identify strong candidates for experimental confirmation of miRNA expression in *Musa sp*.

1.2 Literature Review

1.2.1 Banana

There are two types of bananas, including sweet or dessert bananas which make up approximately 43% of the world's production, while the other 57% are cooking bananas which are also recognized as plantains (Valmayor et al. 2000). The banana is popular as an appetizing and nutritious fruit around the world (Horigome et al. 1992). This globally major food crop is grown and consumed in more than 100 countries throughout the tropics and sub-tropics (INIBAP 2000). One thousand banana cultivars or landraces are recognized world-wide (Heslop-Harrison & Schwarzacher 2007). The modern day edible bananas are a mixture of wild and cultivated species and hybrids associated with *Musa acuminata* and *Musa balbisiana*. *Musa acuminata* is the most widespread of the species in the *Musa* family (Daniells et al. 2001).

1.2.2 Banana plant features

1.2.2.1 Taxonomy

Banana was derived from the Arabic word *banan* which means finger (Boning 2006). The genus name *Musa*, was derived from the Arabic name for the plant (mouz) which in turn, may have been applied in honour of Antonius Musa (63 – 14 BC), physician to Octavius Augustus, the first emperor of Rome (Hyam & Pankhurst 1995).

Musa is a member of the family Musaceae. All genera are monocotyledons and defined as "herbs". Most of the cultivated sweet bananas and plantains are triploid varieties that evolved from two wild diploid species, *M. acuminata*, given the genome designation "AA", and *M. balbisiana* which was given the genome designation "BB" (Simmonds & Shepherd 1995). The formation of homogenomic triploid (2n = 3x) hybrids with the AAA genotype occurred within *M. acuminata*, leading to the

development of cultivars that are mostly comprised of the sweet bananas (Daniells et al. 2001). Cross of the diploid and triploid types of *M. acuminata* with *M. balbisiana* led to the formation of heterogenomic triploid hybrids that are mostly plantains (AAB genotype) and other cooking bananas (ABB genotype). Tetraploid (2n = 4x) and other diploid combinations also exist (Pillay et al. 2004).

Pillay et al. (2004) suggest that the range of genome size for *Musa* lies between 550 and 612 megabasepairs (Mbp), a relatively small size. An analysis of the organization of the banana genome has been done through the sequencing of bacterial artificial chromosome (BAC) clones (Aert et al. 2004; Cheung & Town 2007).

1.2.2.2 Origin and cultivation

Theoretically, the origin of edible bananas was Malesia, a biogeographical region which includes the Malay Peninsula, Indonesia, the Philippines and New Guinea, and was known as the primary centre, while India was recognized as the secondary centre. However, the precise origin is unknown (Simmonds & Shepherd 1955). The centre for *M. acuminata*, the most widespread species, is thought to be either Malaysia (Simmonds 1962) or Indonesia (Horry et al. 1997). The genetic basis of parthenocarpy in *M. acuminata* has not been characterized (Heslop-Harrison & Schwarzacher 2007). Clones of the diploids have been cultivated in the wetter parts of South East Asia (Valmayor et al. 2000).

De Langhe (1995) gave a brief history of the domestication of banana. It has been claimed that there was written (Sanskrit) reference to bananas as early as 500 BC. It was also thought that traders from Arabia, Persia, India and Indonesia distributed banana suckers around coastal regions (except in Australia) of the Indian Ocean

between the 5th and 15th centuries. Between the 16th and 19th centuries, suckers were traded by the Portuguese and Spanish in tropical America. Today, the cultivation of bananas occurs throughout the tropics and sub-tropics of Asia, America, Africa and Australia.

Edible diploids of *M. balbisiana* underwent a parallel evolution in drier parts of Asia, including India, Myanmar, Thailand and Philippines, but there was some geographical overlap with *M. acuminata* and hybrids of the seeded types produced (Valmayor et al. 2000; Daniells et al. 2001). The Indian subcontinent was a major centre for hybridization (Daniells et al. 2001). East Africa and West Africa represent two major secondary centres of *Musa* diversity as a result of the long history of cultivation in these regions (De Langhe 1995). Another secondary centre of diversity is Polynesia which was where the "Maia Maoli/Popoulu" cultivars (thought to be AAB hybrids) were carried to, from the Philippines more than 4000 years ago (De Langhe 1995).

1.2.2.3 Morphology

The cultivated banana plant is a tall (2–9m) perennial monocotyledon and therefore classed as an arborescent herb (INIBAP 2000). The above ground "trunk" is called a pseudostem and consists of concentric layers of leaf sheaths rolled into a cylinder 20–50cm in diameter. Variation in pseudostem morphology exists between cultivars, especially its length, disposition and colouration. The pseudostems of Highland and sweet bananas are predominantly green to dark green with black blotches while those of plantains are yellowish green with brown blotches (Pillay & Tripathi 2007). The true stem is a large underground corm (also called a butt), and the meristem of the apical bud initially gives rise to the leaves before it elongates up through the

pseudostem and emerges some 10–15 months after planting, as a large terminal inflorescence.

The leaves of *Musa* plants emerge tightly rolled, from the centre of the pseudostem in an anticlockwise spiral manner (Barker & Steward 1962). The leaf sheaths taper on both sides to form the petiole. Leaves are more or less vertical when they emerge, becoming horizontal and growing upwards until just before flowering. These leaves then droop down until the emergence of the last leaf, immediately after which comes the inflorescence.

Musa fruits vary in size, shape and colour. They are generally elongate-cylindrical, straight to strongly curved. The fruit apex may be tapered, rounded or blunt. Skin of the fruits may be thin and tender to thick and leathery, with colours that can be yellow, green or red. The flesh of *Musa* ranges from starchy to sweet while colours range from white, cream and yellow to orange (Nelson et al. 2006).

The root system is spread out laterally as far as 5.5m, forming a dense mat mainly in the top 15cm of the soil (Australian Government – Department of Health and Ageing, 2008).

1.2.3 Usage of Banana

1.2.3.1 Staple food

Musa sp. can be consumed either raw or cooked; they are generally eaten fresh as fruit, and are one of the most significant sources of calories in the human diet worldwide (Nelson et al., 2006). Bananas are used for culinary purposes in a variety of ways all over the world. For example in Uganda, bananas can be mixed together with

peanuts and spices, wrapped in a banana leaf and cooked by steaming, while in Nigeria, *fufu* is made from a mixture of plantains and cassava. Both ripe and unripe fruits can be used to produce flour, while the flowers have been used as an ingredient in confections (Kiple and Ornelas, 2000).

Bananas have also been used in salads, breads, as garnish for meats or mashed with spices for making chutneys and sauces. They can be used in the manufacturing of dairy products like yogurts, milkshakes and ice-creams as well as in the production of baby food and sauces. Furthermore, bananas are especially popular as chips – they are peeled, sliced and fried to make this food. Bananas and plantains can usually be found in most tropical households and are often one of the first solid foods fed to infants as they are readily accepted and easily digested (Sharrock, 1997).

1.2.3.2 Medicinal uses

According to Agot (1968) and Sharaf et al (1979), the amount of protein and fat in banana pulp was found to be nutritionally insignificant. However, banana pulp was found to contain useful quantities of vitamins as well as a good source of potassium and energy (Resources Council of Japan, 1982).

Vitamin A found in banana helps aid in digestion while the boiled, mashed ripe fruits are good in treating constipation especially when mixed with other recommended plants. Because banana pulp has high levels of potassium, it has been found to be good in treating high blood pressure (Li, 2008). The ripe fruits can also be used in the treatment of asthma and bronchitis and the pounded peels from the ripe bananas can be made into poultice to treat wounds. In Nigeria, a weaning food based on plantain and soybean has been developed; it is nutritious for babies and can be used as a therapeutic

diet for the treatment of malnutrition and kwashiorkor, which results from protein deficiency. The vitamin-rich nectar sap obtained from the flower buds can be used to strengthen babies (Sharrock, 1997).

1.2.3.3 Alcohol production

In Central and East Africa, the juice from the ripe fruits of the variety known as "beer bananas" may be drunk fresh or fermented to make a beer with low alcohol content and a short shelf-life. Beer brewing has long been an important activity among various communities in these regions and it has been reported that consumption in Rwanda may reach 1.2 litres per capita per day (Stover and Simmonds 1987). The beer is important nutritionally and is rich in vitamin B due to the yeast content. In Uganda and Sudan, banana beer is distilled to produce banana alcohol, or "waragi". The production of commercial or medicinal alcohol from bananas has been carried out for many years in several countries. In terms of alcohol production per hectare, bananas are among the crops giving the highest yields (Thompson 1995).

1.2.3.4 Other uses

Starch can be extracted from banana and plantain pseudostems, and such starch from the related species *Ensete ventricosum* provides a staple food crop in some parts of Ethiopia. Banana starch has been used for producing glue used in the manufacture of cartons for exporting fresh bananas (Thompson, 1995). Pseudostems may also be used as temporary bridges, seats and supports. After harvest they provide valuable mulch, suppressing weed growth in production areas. In mixed farming systems, bananas are used as a ground shade and nurse-crop for a range of shade-loving crops including cocoa, coffee, black pepper and nutmeg in the West Indies, South East Asia, and in parts of East and West Africa. In horticulture, bananas themselves, as well as several

related species, most notably *M. ornata* and *M. velutina* are prized as ornamentals worldwide.

Other minor, but extremely varied, uses of the plant include the use of banana seeds for making necklaces and other ornaments, the use of banana sap as a dye, the use of the fruits as meat tenderizer and the use of banana ash in soap. In Indonesia, the production of floor wax and shoe polish from banana peels is also being explored (Sharrock, 1997).

1.2.4 RNA silencing (RNA interference)

RNA silencing is a type of gene suppression which occurs at the RNA level and includes post-transcriptional gene silencing (PTGS) in plants and fungi, and RNA interference (RNAi) in animals (Segers et al., 2006; Tang et al., 2003). RNA silencing is a process of sequence-specific regulation of gene expression which is triggered by double-stranded RNA (dsRNA) and these RNA silencing mechanisms have been found to be conserved in almost all eukaryotes (Segers et al., 2006; Tang et al., 2003).

RNA silencing processes are able to mediate transcriptional and/or post-transcriptional gene silencing (PTGS) which depends on the organism, as well as the source of the dsRNA trigger (Segers et al., 2006; Tang et al., 2003). Post-transcriptional RNA silencing mechanisms can be defined through mRNA degradation or translational repression. For the target mRNA to be effectively degraded, high sequence complementarity between both the target RNA and the dsRNA trigger is required (Segers et al., 2006; Tang et al., 2003).

RNA silencing plays a key role in regulating the development and control of transposition events in plants, as well as in animals (Segers et al., 2006). In fission yeast, one of the systems that help establish and maintain the heterochromatin structure of the centromere and mating type locus is RNA silencing (Segers et al., 2006). Furthermore, RNA silencing also has an antiviral role in plants and insects such as *Drosophila melanogaster* (Segers et al., 2006).

1.2.5 Mechanism of RNA interference (RNAi)

PTGS/RNAi is an adaptive immune system in plants, targeted against viruses. In response, these viruses have evolved suppressors of the silencing reaction (Dunoyer et al., 2004). The proteins produced are varied in structure, sequence, as well as the superficial effects they have on transgene silencing; this makes them likely to target the separate stages of the PTGS/RNAi process (Dunoyer et al., 2004). More research has yet to be done however, to pinpoint their exact mode of action and position in the PTGS/RNAi pathway.

A common feature of RNA silencing is the processing of structured or dsRNA into small interfering RNAs (siRNAs) of 21 to 26 nucleotides (nt) (Tang et al., 2003). Studies have shown that an RNase III-type enzyme known as Dicer in animals and Dicer-like (DCL) in plants is important in the processes of RNA silencing. Dicer enzymes bind and cleave dsRNA into 21- to 26-nt dsRNA species (Tang et al., 2003), which then function as sequence-specific small interfering RNA (siRNA) or microRNA (miRNA). These small RNA cleavage products are then incorporated into the RNA-induced silencing complex (RISC) to guide in transcript turnover, cleavage or translational control (Tang et al., 2003).

Higher plants have found ways to evolve diversified RNA silencing pathways to generate different classes of small RNAs, each with a specialized function (Qu et al., 2007). For example in *Arabidopsis thaliana*, there are at least four Dicer-like proteins (DCL1 – DCL4). DCL1 proteins are mainly involved in processing imperfectly base-paired fold-back precursors to produce miRNAs while the other three DCL proteins are responsible for the generation of various endogenous siRNAs from perfect dsRNAs. Some of these include the stress-related 24-nt natural-antisense-transcript siRNAs (nat-siRNA) produced by DCL2, 24-nt repeat associated siRNAs that guide heterochromatic formation by DCL3 and 21-nt *trans*-acting siRNAs (tasiRNA) that control the development of certain plant aspects by DCL4 (Qu et al., 2007).

1.2.6 MicroRNA (miRNA)

MicroRNAs (miRNAs) are small (22-nt long), non-coding regulatory RNA molecules encoded as ssRNA precursors within the genome. They are found in both animals and plants and function by repressing translation and/or by cleaving messenger RNA (mRNA) depending on levels of complementarity of base-pairing (Ambros, 2004). miRNAs arise from longer primary transcripts (pri-miRNAs). These pri-miRNAs fold into one or more short dsRNA hairpin structures known as pre-miRNAs (Cullen, 2004). The processing of pre-miRNAs to miRNAs occurs in the nucleus, assisted by the dsRNA-binding protein DiGeorge syndrome critical region gene 8 (DGCR8).

The miRNA duplex undergoes RNA-induced silencing complex (RISC) assembly whereby the miRNA strand is incorporated into a protein complex of which the main protein component is ARGONAUTE (AGO) protein. This mature miRNA then guides the bound AGO protein to target mRNAs based on complementarity (Chen, 2008).

Two regions define the AGO family: The PAZ and PIWI domains. The PAZ domain was found to interact with the 2-bp-3' overhangs of siRNA or miRNA duplexes; while PIWI domains interact through a cryptic RNase H-like activity to mediate the slicing of target mRNA substrates (Ronemus et al., 2006). The *AGO1*-like proteins are the only conserved components of RISC and are present in many forms in different organisms (Ronemus et al., 2006).

In the model plant *A. thaliana*, a single RNase III – Dicer-like 1 (DCL1) – generates both pri- and pre-miRNAs. This reaction produces a miRNA duplex which is composed of mature 21- to 23-nt miRNA and its complementary 21- to 23-nt strand. The structure of the miRNA duplex consists of 2-nt 3' overhangs and 5' phosphates (Lingel and Sattler, 2005). In animals, the pathway varies as there are two different RNase III enzymes that cleave the pri-miRNAs to form miRNAs (Kim, 2004).

In miRNA processing, partial loss-of-function in *DCL1* mutants causes reduced accumulation of miRNAs thereby proving the importance of DCL1. Null mutations in DCL1 were shown to cause embryonic lethality while immunoprecipitated DCL1 from *Arabidopsis* shows the ability to generate 21-nt small RNAs from long dsRNAs (Chen, 2008). The fact that mutations in other Dicer homologs (DCL2, DCL3 and DCL4) do not affect accumulation quantity of miRNAs proves that pre-miRNAs are also processed by DCL1 (Chen, 2008).

A total of 959 plant miRNA genes from ten plant species (including moss, dicots and monocots) have been recorded. Features of plant miRNA genes determined from current sets of miRNA, include paralogs, i.e. 184 *Arabidopsis* miRNAs which represent

almost 100 families of related miRNAs. Gene duplication as well as diversification which have driven the evolution of protein-coding gene families causes plant miRNA gene families to arise (Chen, 2008). Certain miRNA families appear to have been conserved in mosses to flowering plants since ancient times although it is currently impossible to determine how many miRNA families are conserved due to the incomplete genome sequencing of non-flowering land plants (Chen, 2008). Yet another feature is the fact that the majority of plant miRNA genes are localized in intergenic regions unlike animal miRNAs which are located in introns or exons of protein-coding genes. Furthermore, plant miRNAs are usually not co-expressed or arranged in tandem in the genome.

1.2.6.1 Biological function of miRNA in plants

The regulation of target gene expression through miRNAs is important for many biological processes in both animals and plants (Axtell et al., 2007).

In plants, the role of miRNAs as controllers of development in flowering plants is crucial; for example, in *Arabidopsis thaliana* many miRNA targets encode proteins which either have a known or suspected role in development control (Axtell and Bartel, 2005). Floral and leaf-patterning defects have been seen when individual miRNAs are dysfunctional and thus lose their ability to regulate their targets properly (Axtell and Bartel, 2005). Other defects like floral development and timing, loss of organ polarity and altered vascular development, reduced fertility, abortion of the shoot apical meristem and misexpression of early auxin response genes have also been observed (Axtell and Bartel, 2005). With such results, it is clear to see the crucial role of miRNAs in regulating various stages of cell growth in specific cell types for normal plant development.

The generation of dsRNA is stimulated through miRNA-directed target cleavage with the usage of the cleaved product as a template. The dsRNA which results from the process is then further subjected to a Dicer protein to obtain siRNAs known as *trans*-acting siRNAs (ta-siRNAs). These ta-siRNAs are able to repress mRNAs from loci different from the originating locus (Axtell et al., 2007).

Plant and animal miRNAs differ in biogenesis aspects: Predicted foldback of miRNA stem loops in plants are much more variable and larger in size compared with those in animals. There are also increased pairing between miRNA and the second arm of the stem loop in plants. Furthermore, plants have a tighter distribution with the 21-nt length of miRNAs as opposed to the 22- to 23-nt length in animals. Finally, plant miRNAs often recognize a single target site, while animal miRNAs recognize several target sequences in the 3'UTR of mRNAs which can inhibit translation (Bartel, 2004; Yang et al., 2007; Vazquez, 2006).

1.2.7 Computational identification of microRNA

MicroRNAs – short, endogenous non-coding RNAs – are found in animals, plants and the Epstein-Barr virus. The majority of miRNA genes exist as independent transcriptional units and have been identified through cloning as well as computational approaches. They are customized to the key features of two founding members of miRNA from *Caenorhabditis elegans* – *lin-4* and *let-7* – which also include a fold-back hairpin RNA precursor that is joined with evolutionary conservation (Sunkar et al., 2005).

It has been currently estimated that miRNA genes represent about 1% of the expressed genome in complex organisms including worms, flies and humans. However, the numbers of miRNAs have been raised significantly in primates through recent computational predictions by comparing analysis of human, mouse and rat genomes (Sunkar et al., 2005). Identification of all miRNAs and their target genes from model organisms is very important so that regulatory networks and gene silencing mechanisms can be understood (Sunkar et al., 2005).

The identification of many miRNAs has been enabled with direct cloning. However, large variations in expression levels have made it difficult for low abundance miRNAs to be cloned (Adai et al., 2005). The fact that most miRNAs are derived from precursor transcripts (70–100 nucleotides) with extended stem-loop structures is one of the reasons that suggest computational miRNA predictions approaches will be viable in identifying miRNAs. Another reason is that miRNAs are highly conserved between genomes of related species (Thomassen et al., 2006).

Studies done with bioinformatic tools and cDNA cloning were used to identify potential *C. elegans* miRNAs since 2001 (Lee and Ambros, 2001). Fifteen novel miRNAs were discovered; two of them through computational screening and the rest from cDNA cloning (Thomassen et al., 2006). According to a study by Lim et al., (2003) another computational tool – MiRscan – was designed for the identification of miRNA genes conserved between genomes and was first applied to *C. elegans* and *C. briggsae*. Computational methods to identify miRNAs in *Drosophila* were also developed (miRseeker) to search for euchromatic DNA sequences of *D. melanogaster* and *D. pseudoobscura* for transcripts that are able to form stem-loop structures (Thomassen et al., 2006).

1.2.7.1 Computational identification of microRNA in plants

In the search to identify plant miRNAs, a study was conducted to identify characteristic miRNA features as well as sequence conservation. This allowed the identification of 63% of known *Arabidopsis* miRNAs with 83 claimed novel miRNAs of which 25 were verified. Possible miRNA precursors were evaluated through a computer algorithm based on: Stem-loop structure, GC content of mature miRNA, loop length, mismatches in stem containing mature miRNA as well as conservation of mature miRNA sequence. In plants, a few mismatches can be found between the alignment of miRNA and mRNA targets, which can be used together with typical miRNA features and conservation searches (Thomassen et al., 2006). With the increasing number of plant miRNA genes identified in various species, homology-based search methods were developed in an attempt to seek complete enumeration of miRNAs in model organisms (Mendes et al., 2009).

1.2.7.2 Expressed Sequence Tags (EST) and Prediction of microRNA from ESTs

Partial complementary DNA (cDNA) sequences of expressed genes, which are cloned into a plasmid are known as Expressed Sequence Tags (ESTs) (Zhang *et al.* 2005). Computational analyses of ESTs are one of the strategies that have contributed to the existing identified plant miRNAs (Yang *et al.* 2007). ESTs provide sequence information that can be used to identify novel genes based on the highly conserved sequences in mature miRNAs as well as long hairpin structures in precursors, gene locations and intron–exon boundaries within genomic sequence assemblies (Matukumalli *et al.* 2004; Qiu *et al.* 2007).

As the use of traditional computational methods of gene identification was not that efficient and not comprehensive, ESTs provide an economical alternative tool to discover genes in species that lack draft genome sequences (Matukumalli *et al.* 2004). In addition, the systematic mining of EST databases could provide a deeper insight into the distribution and conservation of miRNAs (Zhang *et al.* 2007).

miRNAs are non-coding RNA. Many miRNAs are evolutionarily conserved from species to species; therefore miRNA genes may exist as orthologs in other species and could be present in EST data sets (Zhang *et al.* 2006a). Thus, by searching for homologs of known miRNAs in an EST database, new miRNAs in other species may be discovered (Zhang *et al.* 2006a). However, due to the inequality and non-uniformity of most EST data sets, biases could happen and affect the data mining results (Zhang *et al.* 2006a). To date, analysis of EST data has been used to successfully predict miRNAs in *Arabidopsis* (Jones-Rhoades and Bartel 2004), soybean (Zhang *et al.* 2005), maize, sorghum, medick, poplar (Dezulian et al., 2005), *Brassica napus* (Xie et al. 2007), oil palm (Nasaruddin et al., 2007), as well as cotton (Zhang et al., 2007).

1.2.7.3 Characteristics of miRNAs and their precursors

In general, there are a few characteristics that differentiate miRNAs from other RNA molecules. It has been noted however, that miRNAs and small interfering RNAs (siRNAs) function as gene silencers (Lin et al., 2006). miRNAs derived from hairpin-like precursors have been observed in almost all eukaryotes including yeast, plants, nematodes, flies, mice and humans. On the other hand, siRNAs have been found in large numbers in plants and in simple animals (i.e. worms, flies); but are rarely seen in mammals. The widespread presence of miRNAs may be able to contribute to novel therapeutics for cancers and viral infections (Lin et al., 2006).

Although the formation of a stem-loop structure is an essential step in miRNA maturation, it is not a unique characteristic of miRNAs as other RNAs can form similar hairpin structures. One way to avoid selecting other RNAs or RNA sequences as new miRNAs is that the predicted hairpin structure has the lowest folding free energy. However, there was not a significant difference in folding free energies for most miRNA precursors despite the fact that some of these precursors did have minimal folding free energies (MFEs). The MFEs were also related to RNA length: The longer the sequence, the lower the MFE. To make MFEs comparable (MFEs can be used to compare different types of RNAs as long as the lengths remain constant) a new term – adjusted MFE (AMFE) – was introduced. This further contributed to minimal folding free energy index or MFEI which consisted of a few parameters including AMFE, length of RNAs and GC content. With this, it was discovered that on average, miRNA precursors had a MFEI greater than 0.85 compared with tRNAs (0.64) and rRNAs (0.59). Hence, MFEI values can be used to distinguish miRNA sequences from other coding and non-coding RNAs (Zhang et al., 2005b).

1.2.7.4 Homology search-based approaches

The Basic Local Alignment Search Tool or BLAST, allows comparison between similar sequences from the same species or even across multiple species. It has become one of the most popular bioinformatics algorithms used by biologists which enables them to search for their queries against sequence databases and produce pair-wise alignments based on query sequences and the corresponding sequences from the database (Gollapudi et al., 2008).

BLAST, which is available via the web, allows users to search against specialized databases like NCBI and PlantGDB. Users can search these databases for novel miRNAs as mature miRNAs and long hairpin structures are highly conserved. With sequence data becoming widely accessible, additional programs like MuSeqBox and BioParser have been developed to solve some of the limitations faced by BLAST. With these programs, BLAST is able to provide information on matching regions based on percent alignment coverage and identity, alignment scores and expectation value (E-value) among others (Gollapudi et al., 2008).

1.2.7.5 RNA secondary structure prediction

Secondary structures and minimal folding free energies (MFEs) of RNA sequences, including that of miRNAs, can be predicted with the software mfold (Zucker, 2003) – which is available publicly on the web (http://www.bioinfo.rpi.edu/applications/mfold/old/rna/). For each type of RNA studied, the MFE, length of sequence, number of nucleotides (A, C, G or U) and the number of base pairs are determined and exported into an Excel file (Zhang et al., 2005b). The secondary structures are then predicted through the MFOLD software based on minimizing free energy of RNA or DNA sequences. This is done by maximising the number of favourable base-pairing interactions. Even so, biologically correct structures for RNAs are not often the calculated optimal structure. Predicted structures of rRNAs and/or tRNAs by MFOLD therefore, may be slightly different from the biologically active structure (Zhang et al., 2005b).

CHAPTER 2:

MATERIALS AND METHODS

2.0 Materials and Methods

2.1 Obtaining sequences of miRNAs and *Musa sp* EST

The sequence data for a total of 959 previously known plant miRNA sequences from: *Arabidopsis thaliana* (184), *Brassica napus* (5), *Glycine max* (22), *Medicago truncatula* (30), *Oryza sativa* (242), *Physcomitrella patens* (77), *Populus trichocarpa* (215), *Saccharum officinarum* (16), *Sorghum bicolor* (72) and *Zea mays* (96) were obtained from the miRNA Registry Database, Wellcome Trust Sanger Institute (Release 9.2, May 2007) (http://microrna.sanger.ac.uk/). The sequence data for a total of 42,978 *Musa sp*. EST data were obtained from The Global Musa Genomics Consortium (http://www.musagenomics.org/) (33,922 sequences) and from the National Centre for Biotechnology Information (NCBI) GenBank nucleotide databases (http://www.ncbi.nlm.nih.gov/) (9,056 sequences).

2.2 Removal of duplicate miRNAs sequences

The miRNA sequences were first scanned within the species mentioned above, while the second round of scanning was inter-species, to remove duplicated mature miRNA sequences. After the first round of scanning, a total of 558 sequences were retained from the original 959 miRNA sequences. After the second round of scanning a set of 397 non-redundant miRNA sequences were retained to be used as the reference set for miRNA identification.

2.3 BLAST comparison of the sequences

Comparative software BLAST-2.2.16 was downloaded from NCBI GenBank (http://www.ncbi.nlm.nih.gov/BLAST/) in order to carry out a BLAST search for *Musa sp.* homologs to the miRNA reference sequences within the *Musa sp.* EST sequences. BLAST-2.2.16 was run through Command Prompt (DOS Prompt). Steps for the BLAST comparison are shown below (Figure 2.1):

a) Firstly, the miRNA database was created in BLAST, with the command:

formatdb –i miRNA.txt –p F –o T

b) Then, the EST sequences obtained were saved in .txt file format.

c) Next, the database and EST sequences were subjected to a BLAST search for *Musa sp.* homologs, with the command:

blastall –p blastn –d miRNA.txt –i EST.txt –o EST.out

d) The parameters used were set to default at the threshold level: expected value (E-value) maximum at 10.0 and low complexity filter.

d) Output from the BLAST search was scanned and a candidate from the miRNA *Musa sp.* homologs would be selected based on less than 4 nucleotide mismatches [(n/n, (n-1)/n, (n-2)/n or (n-3)/n, n = the length of the known miRNA)] in sequence with the miRNAs. These miRNA candidates were then used in the secondary structure predictions (section 2.4).

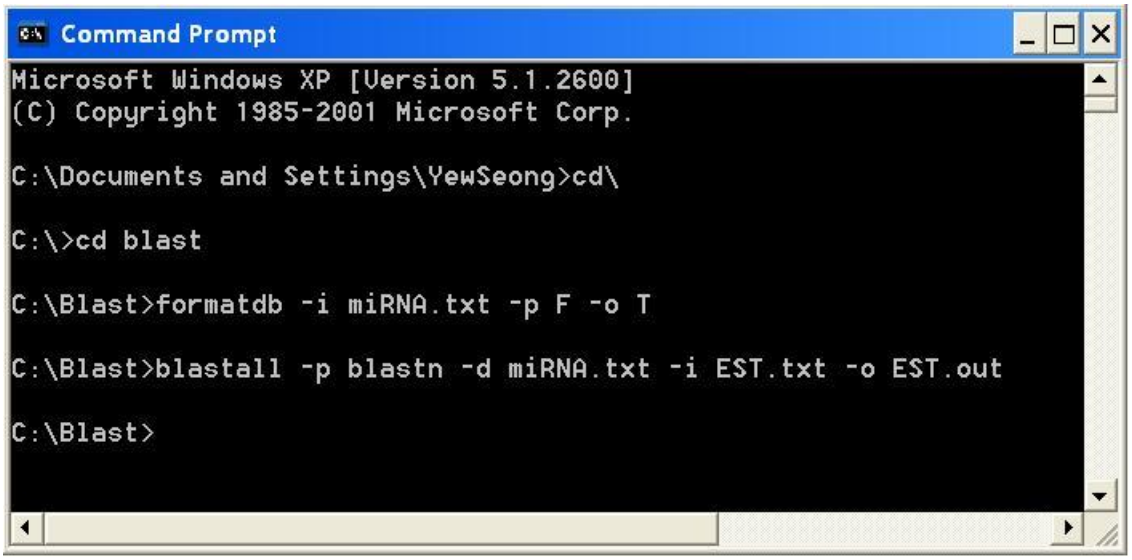


Figure 2.1: Arguments required for running a BLAST program in the Command Prompt.

2.4 Prediction of secondary structure of candidate miRNA sequences

The secondary structures of candidate miRNA sequences were predicted by using web-based software Mfold 3.2 developed by Zuker and Turner from Rensselaer Polytechnic Institute (Zuker, 2003; http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi). The following criteria, which are based on structural and energetic features of known miRNA sequences, were used for selecting the candidates of potential miRNAs (Qiu et al., 2007; Nasaruddin et al., 2007):

a) The secondary structures of miRNA candidates have high negative minimal folding free energies (MFEs).

b) Sequences of the miRNA candidates are within one arm of the hairpin secondary structure that will contain the ~22 nucleotides mature miRNA sequence.

c) No loops or breaks in the miRNA sequence were allowed.

d) Sequence of the miRNA had less than six mismatches with the miRNA* sequence in the other arm.

e) miRNA candidates have 30-70% A+U content.

f) Any loop or bulge in mature miRNA:miRNA* dimer should contain a maximum of three nucleotides that are not involved in canonical base pairing (i.e. GU, GA and AU pairing).

In the web-based software Mfold 3.1, the parameters used for prediction of the secondary structures of the miRNA candidates were default as shown below:

a) The RNA sequence was linear.

b) Folding temperature was fixed at 37°C.

c) Ionic conditions: 1M NaCl, no divalent ions.

d) The percent sub optimality number was 5.

e) An upper bound on the number of computed folding was 50.

f) The window parameter was default.

g) The maximum interior/bulge loop size was 30.

h) The maximum asymmetry of an interior/bulge loop was 30.

i) The maximum distance between paired bases was no limit.

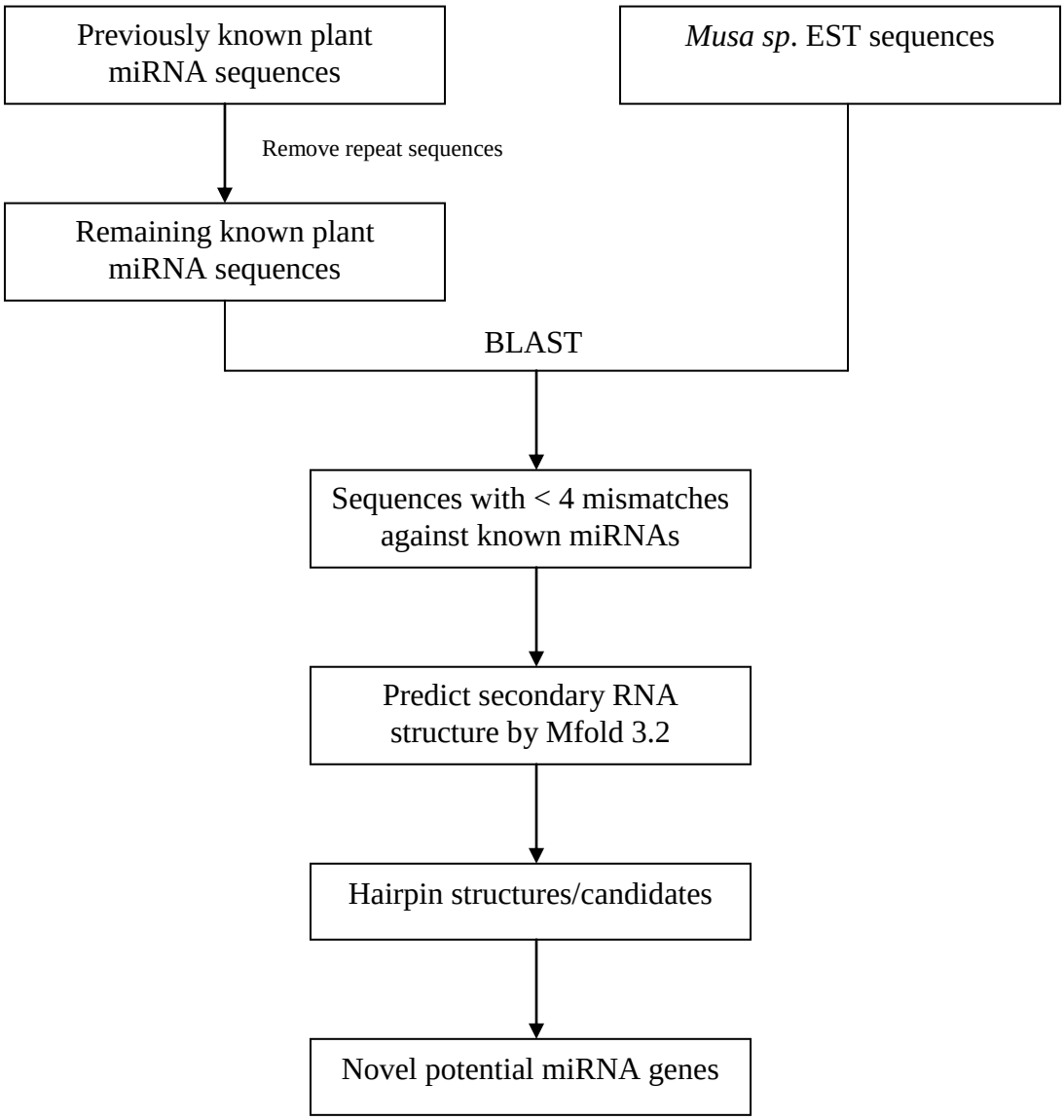


Figure 2.2: Procedure of identification potential miRNA genes using bioinformatic approaches.

2.5 Confirmation of expression of a predicted miRNA in banana

2.5.1 RNA extraction using CTAB-NETS

CTAB extraction buffer contains 2% CTAB (cetyl trimethylammonium bromide), 2% PVP (MW 40000) (polyvinypyrrolidone), 100mM Tris-HCl with pH 8.0, 25mM EDTA, 2M NaCl (sodium chloride) and 2% β-mercaptoethanol. Solution preparation for RNA extraction using CTAB-NETS were shown in Appendix 2.

A few banana leaves were pulverized in liquid nitrogen using a mortar and pestle pre-chilled to -20°C. The ground sample was then immediately transferred to a 2ml tube containing 1ml of pre-heated (65°C) CTAB buffer and 20µl of β-mercaptoethanol, and the tube was inverted. Next, 1ml of chloroform-isoamylalcohol was added; the tube was inverted and centrifuged at 13,000 X *g* for 15 minutes at room temperature (25°C).

The upper aqueous phase was extracted using a pipette, transferred to a clean tube and the procedure was repeated with 1ml of chloroform-isoamylalcohol added. The tube was inverted and centrifuged at 13,000 X *g* for 15 minutes at room temperature (25°C). The supernatant obtained was then precipitated using 0.1 volume of 3M NaOAC at a pH of 5.2 and 3 volume of pre-cooled abs EtOH.

The contents in the tube were gently mixed by inversion and then centrifuged at low speed for 5 minutes (or the fibrous DNA was spooled out using the tip of a pipette). Then, the DNA was transferred to a 50ml tube and after 20 minutes washed out with 10ml of the washing buffer; the contents were centrifuged again at low speed (or the DNA was spooled out).

The pellet obtained was air-dried and re-suspended in a mixture of 1ml TE buffer, 2ml distilled water and 1.5ml 7.5M ammonium acetate and mixed by inversion. Next, 10ml of cold absolute ethanol was added. The tube was gently inverted to precipitate DNA and then centrifuged at low speed (or the DNA was spooled out). After that, the pellet was washed with 5ml of 70% ethanol and further centrifuged at low speed. Finally, the pellet was air-dried and re-suspended in 500-750µl of TE buffer.

2.5.2 DNase I treatment

A 40µl aliquot of the sample was topped up to 200µl with DEPC-H_2O, followed by 40µl of Phenol-CI (Phenol-chloroform-isoamylalcohol) pipetted into the microcentrifuge tube. The sample was mixed well in the tube and then centrifuged at 13,000 X *g* for 15 minutes at room temperature (25°C).

The supernatant was discarded completely. A volume of 40µl of Phenol-CI was then pipetted into the microcentrifuge tube and the pellet mixed well. This mixture was centrifuged at 13,000 X *g* for 15 minutes at room temperature (25°C). The supernatant obtained was again discarded completely and the pellet gently mixed with Isopropanol to precipitate the RNA. The sample was then incubated overnight at -80°C.

2.5.3 Agarose gel electrophoresis

A 1% agarose gel [0.2g agarose powder and 20ml of Tris-Borate-EDTA (TBE buffer: Tris base, Boric acid and 0.05M EDTA at a pH of 8.0)] was run to check the RNA quality. Approximately 8µl of RNA and 2µl of loading RNA dye per well were loaded. The gel was run at 100 V for 30 min, in 1 X TBE buffer.

2.5.4 Primer design

Three pairs of forward and reverse PCR primers were designed based on the two selected EST sequences (gi|146226345|, 600092884T1 and 600078119T1) for reverse-transcription PCR. The primer sequences were designed using the web-based software Primer 3 (http://frodo.wi.mit.edu/). The primer sequences are as shown in Table 2.1:

miRNA	Primer	Sequences
bna-miR156a (gi|146226345|)	Forward (6345-F)	AGA GAA CTT CCT CCC CCA AA
	Reverse (6345-R)	GCT TCC TCC TGC ACA ATT TC
osa-miR169f (600092884T1)	Forward (884T1-F)	ATC CTT CCA AAG TCC TGC AA
	Reverse (884T1-R)	ATC CAT GGC CAC GCT ATA AA
osa-miR169f (600078119T1)	Forward (119T1-F)	ATC TGC TTC CGT CTG TGC TT
	Reverse (119T1-R)	AGA AAG GGC TTG CAT GGA G

Table 2.1: Primer sequences based on the two selected EST sequences.

2.5.5 Reverse Transcription PCR (RT-PCR)

RT-PCR was performed using the Access RT-PCR System Kit (Promega Corporation, USA). The Access RT-PCR System was designed for reverse transcription (RT) and polymerase chain reaction (PCR) amplification of a specific target RNA from either total RNA or mRNA (Miller and Storts, 1995).

A master mixture with a final volume of 300µl was prepared for six reactions. RT reaction contained 162µl of RNase-free water, 60µl of 5 X RT-PCR buffer, 6µl of 10mM dNTP mix, 6µl of AMV Reverse Transcriptase (AMV RT), 6µl of *Tfl* DNA Polymerase and 12µl of 25mM $MgSO_4$. The mixture was spun quickly and then distributed into three 0.2ml PCR tubes, each with a volume of 84µl. Next, 6µl of primer 6345-F and 6µl of primer 6345-R were added into the first tube. The same amounts of

primers 884T1-F/884T1-R and primers 119T1-F/119T1-R were added into the remaining two tubes. All three tubes were labelled and quickly spun. The tube that contained primer pair 6345-F/6345-R was then aliquoted into a new 0.2ml PCR tube. RNA template (2µl) was added into one of the tubes. The second tube was used as a negative control with 2µl of RNase-free water. The same procedure was repeated for primer pairs 884T1-F/884T1-R and 119T1-F/119T1-R. Six tubes (three negative controls of primers, as well as one with the primer pair 6345-F/6345-R, one with the primer pair 884T1-F/884T1-R and one with the primer pair 119T1-F/119T1-R) were prepared.

An Eppendorf Mastercycler Gradient was used to run RT-PCR. The reverse transcription reaction was run at 45ºC for 45 min and then incubated for 2 min at 94ºC. The PCR conditions were: 40 cycles of denaturation at 94ºC for 30 seconds, annealing at 60ºC for 1 min, extension at 68ºC for 2 min and final extension at 68ºC for 7 min.

Fragment sizes of the PCR products were visualized using ethidium bromide followed by gel electrophoresis (as mentioned in section 2.5.3). VC 100 bp Plus DNA ladder (Vivantis) was used to compare the fragment sizes. Two controls were used: primers targeting ß-actin gene, a house keeping gene and RNA sample obtained after PCR (RNA PCR).

2.5.6 Purification of PCR Products

Purification of PCR products was performed using Wizard® SV Gel and PCR Clean-Up System Kit (Promega Corporation, USA). The Wizard® SV Gel and PCR Clean-Up System was designed to purify products directly from PCR with up to 95% recovery achieved depending on the fragment size.

A 1.5ml microcentrifuge tube was weighed for each RNA sample obtained after PCR to be purified. The DNA fragments of interest were excised from the agarose gel using a clean scalpel under a long UV-wavelength lamp. The gel slices were then transferred into the weighed microcentrifuge tubes and the new weight recorded. Membrane binding solution was added in a ratio of 10µl of solution per 100mg of agarose gel slice. The mixture was incubated at 55ºC until the gel was completely dissolved.

One SV Minicolumn was placed in a Collection Tube for each dissolved gel slice. Then, the dissolved gel mixture was transferred to the SV Minicolumn assembly and incubated for 1 minute at room temperature. The SV Minicolumn assembly was then centrifuged at 16,000 X *g* for 1 minute. SV Minicolumn was removed from the Spin Column assembly and the liquid discarded from the Collection Tube. The SV Minicolumn was then returned to the Collection Tube.

Next, 700µl of Washing Buffer was added to the SV Minicolumn to wash the column and then centrifuged at 16,000 X *g* for 1 minute. Flow-through was again discarded. Centrifugation was repeated with 500µl of Membrane Wash Solution for 5 minutes at 16,000 X *g*. Flow-through was discarded. Then, the column assembly was re-centrifuged for 1 minute with the microcentrifuge lid open to allow evaporation of any residual ethanol. After that, the SV Minicolumn was transferred to a clean 1.5ml microcentrifuge tube. 50µl of Nuclease-Free Water was applied directly to the centre of the column without touching the membrane with the pipette tip. This mixture was then incubated at room temperature for 1 minute and later centrifuged for 1 minute at 16,000 X *g* to spin down the pure DNA. The microcentrifuge tube containing eluted DNA was

stored at -20ºC. Gel electrophoresis (as section 2.5.3) was run to confirm the size of DNA obtained.

2.5.7 DNA sequencing

DNA fragments of the purified PCR products were sent to the Macrogen Genome Research Centre for sequencing (Seoul, Korea).

2.5.8 Sequence alignments

DNA sequences of PCR products were analyzed by alignment with the *Musa sp* ESTs sequences that were used initially for primers design using BioEdit, a biological sequence alignment editor (http://www.mbio.ncsu.edu/BioEdit/BioEdit.html). Sequences of candidate mature miRNA for *Musa sp* were aligned with known miRNA mature sequences of various plant species. Besides that, the sequencing results of *Musa sp* were also aligned to known precursor miRNA sequences from similar miRNA families of various plant species.

CHAPTER 3:

RESULTS

3.0 Result

3.1 BLAST comparison of *Musa sp*. EST sequences

From a database of 42,978 *Musa sp*. EST sequences, a total of 47 candidate miRNA sequences were chosen for secondary structure prediction. These comprised 37 candidate miRNA sequences from The Global Musa Genomics Consortium EST dataset (33,922 sequences) and 10 candidate miRNA sequences from the NCBI GenBank nucleotide database (9,056 sequences). BLAST results of these 47 potential miRNAs of *Musa sp.* were shown in Appendix 1. The candidate miRNA *Musa sp.* homologs were selected based on having less than 4 nucleotide mismatches [(n/n, (n-1)/n, (n-2)/n or (n-3)/n, n = the length of the known miRNA)] to sequences within the mature miRNA reference dataset. The nucleotide sequence identities of the matched EST sequences ranged from 21/21, 20/21, 19/21, 20/20, 19/20, 19/19, 18/19 to 18/18 nucleotides. The 47 candidate miRNA sequences were classified into 11 miRNA families based on their nucleotide identity, as shown in Table 3.1.

miRNA family	LM (nt)	Strand	EST sequence name	Identities	NM (nt)	Candidate miRNA sequences
bna-miR156a	21	+	gi|146226345|	21/21	0	UGACAGAAGAGAGUGAGCACA
	21	-	gi|146226218|	20/21	1	UGUGCUC**U**CUCUCUUCUGUCA
	21	+	600104506T1	20/21	1	UGACAGAAGAGAG**A**GAGCACA
	21	+	600145442T1	20/21	1	UGACAGAAGAGAG**A**GAGCACA
	21	+	600081801T1	19/20	2	UGACAGAAGAGAG**A**GAGCAC
	21	+	600145360T1	19/20	2	UGACAGAAGAGAG**A**GAGCAC
	21	-	gi|146223567|	18/19	3	UGCUC**U**CUCUCUUCUGUCA
	21	+	600129804T1	18/19	3	UGACAGAAGAGAG**A**GAGCA
	21	+	600133517T1	18/19	3	UGACAGAAGAGAG**A**GAGCA
	21	+	600156376T1	18/19	3	UGACAGAAGAGAG**A**GAGCA
	21	+	600160669T1	18/19	3	UGACAGAAGAGAG**A**GAGCA
mtr-miR395b	21	+	gi|146222248|	18/19	3	UGAAGUAUUUGG**U**GGAACU
osa-miR156k	21	-	gi|146226218|	21/21	0	UGUGCUCUCUCUCUUCUGUCA
	21	+	600104506T1	21/21	0	UGACAGAAGAGAGAGAGCACA
	21	+	600145442T1	21/21	0	UGACAGAAGAGAGAGAGCACA
	21	+	gi|146226345|	20/21	1	UGACAGAAGAGAG**U**GAGCACA
	21	+	600081801T1	20/20	1	UGACAGAAGAGAGAGAGCAC
	21	+	600145360T1	20/20	1	UGACAGAAGAGAGAGAGCAC
	21	-	gi|146223567|	19/19	2	UGCUCUCUCUCUUCUGUCA
	21	+	600129804T1	19/19	2	UGACAGAAGAGAGAGAGCA
	21	+	600133517T1	19/19	2	UGACAGAAGAGAGAGAGCA
	21	+	600156376T1	19/19	2	UGACAGAAGAGAGAGAGCA
	21	+	600160669T1	19/19	2	UGACAGAAGAGAGAGAGCA
	21	+	600132974T1	18/18	3	UGACAGAAGAGAGAGAGC
	21	+	600133648T1	18/18	3	UGACAGAAGAGAGAGAGC
	21	+	600148339T1	18/18	3	UGACAGAAGAGAGAGAGC
	21	+	600174542T1	18/18	3	UGACAGAAGAGAGAGAGC
osa-miR156l	21	-	gi|146223567|	18/19	3	AUGCUC**U**CUCUCUUCUGUC
	21	+	gi|146226345|	18/18	3	GACAGAAGAGAGUGAGCA
	21	+	600129804T1	18/19	3	GACAGAAGAGAG**A**GAGCAU
	21	+	600133517T1	18/19	3	GACAGAAGAGAG**A**GAGCAU
	21	+	600156376T1	18/19	3	GACAGAAGAGAG**A**GAGCAU
	21	+	600160669T1	18/19	3	GACAGAAGAGAG**A**GAGCAU
osa-miR160e	21	-	gi|70796212|	20/20	1	GGCAUACAGGGAGCCAGGCA
osa-miR160f	21	-	gi|70796212|	19/20	2	GGCAU**A**CAGGGAGCCAGGCA
osa-miR164d	21	+	600173221T1	18/19	3	GAGAAGCAGGGCAC**C**UGCU
osa-miR169d	21	-	600081013T1	18/19	3	GGCAA**G**UCAUCCUUGGCUA
	21	-	600092447T1	18/19	3	GGCAA**G**UCAUCCUUGGCUA
	21	-	600092884T1	18/19	3	GGCAA**G**UCAUCCUUGGCUA
osa-miR169e	21	-	600081013T1	19/19	2	GGCAAGUCAUCCUUGGCUA
	21	-	600092447T1	19/19	2	GGCAAGUCAUCCUUGGCUA
	21	-	600092884T1	19/19	2	GGCAAGUCAUCCUUGGCUA
	21	-	600078119T1	18/18	3	GGCAAGUCAUCCUUGGCU
osa-miR169f	21	-	600081013T1	20/20	1	AGGCAAGUCAUCCUUGGCUA
	21	-	600092447T1	20/20	1	AGGCAAGUCAUCCUUGGCUA
	21	-	600092884T1	20/20	1	AGGCAAGUCAUCCUUGGCUA
	21	-	600078119T1	19/19	2	AGGCAAGUCAUCCUUGGCU

osa-miR169n	21	-	600081013T1	19/20	2	AGGCAAGUCAU**C**CUUGGCUA
	21	-	600092447T1	19/20	2	AGGCAAGUCAU**C**CUUGGCUA
	21	-	600092884T1	19/20	2	AGGCAAGUCAU**C**CUUGGCUA
	21	-	600078119T1	18/19	3	AGGCAAGUCAU**C**CUUGGCU
	21	+	600080802T1	18/19	3	AGC CAA GAA UGA **A**UU GCC U
	21	+	600080875T1	18/19	3	AGC CAA GAA UGA **U**UU GCC U
	21	+	600085062T1	18/19	3	AGC CAA GAA UGA **A**UU GCC U
	21	+	600087026T1	18/19	3	AGC CAA GAA UGA **A**UU GCC U
	21	+	600087712T1	18/19	3	AGC CAA GAA UGA **U**UU GCC U
	21	+	600153776T1	18/19	3	AGC CAA GAA UGA **A**UU GCC U
	21	+	600177165T1	18/19	3	AGC CAA GAA UGA **A**UU GCC U
osa-miR169q	21	+	600092884T1	19/19	2	UAGCCAAGGAGACUGCCCA
osa-miR171b	21	-	gi|146222618|	19/21	2	GATAUUG**U**C**G**CGGCUCAAUCA
osa-miR171i	21	-	gi|146222618|	19/20	2	GATAUUG**U**CGCGGCUCAAUC
osa-miR172c	21	+	600079015T1	18/19	3	GAAUC**C**UGAUGAUGCUGCA
	21	+	600084823T1	18/19	3	GAAUC**C**UGAUGAUGCUGCA
	21	+	600087966T1	18/19	3	GAAUC**C**UGAUGAUGCUGCA
	21	+	600093195T1	18/19	3	GAAUC**C**UGAUGAUGCUGCA
	21	+	600100194T1	18/19	3	GAAUC**C**UGAUGAUGCUGCA
	21	+	600103944T1	18/19	3	GAAUC**C**UGAUGAUGCUGCA
	21	+	600106033T1	18/19	3	GAAUC**C**UGAUGAUGCUGCA
	21	+	600124872T1	18/19	3	GAAUC**C**UGAUGAUGCUGCA
	21	+	600142076T1	18/19	3	GAAUC**C**UGAUGAUGCUGCA
	21	+	600160442T1	18/19	3	GAAUC**C**UGAUGAUGCUGCA
	21	-	gi|146228497|	18/19	3	UGCAGCAUCAUCA**G**GAUUC
osa-miR393	21	-	gi|146222506|	18/18	3	CAAUGCGAUCCCUUUGGA
	21	+	gi|146228486|	18/18	3	UCCAAAGGGAUCGCAUUG
	21	+	600079708T1	18/18	3	UCCAAAGGGAUCGCAUUG
	21	+	600103903T1	18/18	3	UCCAAAGGGAUCGCAUUG
osa-miR395v	21	+	gi|146222248|	18/19	3	UGAAGUAUUUGG**U**GGAACU
osa-miR396d	21	+	gi|146222802|	18/19	3	UCCACA**U**GCUUUCUUGAAC

Table 3.1: Candidate miRNA sequences chosen for secondary structure prediction. LM: Length of mature miRNAs; NM: Number of mismatch (less than 4 mismatches, 0-3); Letter (bold and underlined): Nucleotide substitution compared with the previously known mature miRNAs.

3.2 Secondary structures of candidate miRNA sequences

Strong miRNA candidates within the 47 candidate sequences were identified through prediction of the secondary structure for each. From these 47 sequences, 920

secondary structures were generated as listed in Table 3.2. After further analysis , it was found that only nine of the 920 EST sequences (Table 3.3 and Figure 3.1) fulfilled the six criteria of mature miRNA secondary structures, including: high negative minimal free energy (MFEs), ability to fold such that the mature miRNA was within one arm of the hairpin secondary structure, the absence of loops or breaks in the miRNA sequence, less than six unpaired bases and GU mismatches between the miRNA and miRNA* sequences, an A+U content of 30-70% and with any loop or bulge in the mature dimer containing a maximum of three nucleotides not involved in canonical base pairing (i.e. GU, GA and AU pairing). From the results, it can be predicted that the miRNA gene candidates are members of the miR156 as well as miR169 and miR172 families.

No	Sequence name	Number of secondary structure generated	Number of secondary structures fulfilling criteria for miRNA precursors
1	gi|70796212|	29	0
2	gi|146228497|	27	0
3	gi|146228486|	24	0
4	gi|146226345|	24	1
5	gi|146226218|	22	0
6	gi|146223567|	35	0
7	gi|146222802|	28	0
8	gi|146222618|	17	0
9	gi|146222506|	30	0
10	gi|146222248|	34	0
11	600177165T1	7	0
12	600174542T1	28	0
13	600173221T1	18	0
14	600160669T1	26	0
15	600160442T1	14	1
16	600156376T1	28	0
17	600153776T1	16	0
18	600148339T1	30	0
19	600145442T1	9	0
20	600145360T1	11	0
21	600142076T1	32	0
22	600136111T1	22	0
23	600133648T1	5	0
24	600133517T1	11	0
25	600132974T1	19	0
26	600129804T1	6	0

27	600124872T1	26	0
28	600106033T1	22	0
29	600104506T1	10	0
30	600103944T1	28	1
31	600103903T1	18	0
32	600100194T1	15	0
33	600093195T1	33	0
34	600092884T1	12	1
35	600092447T1	6	1
36	600087966T1	12	0
37	600087712T1	20	0
38	600087026T1	19	0
39	600085062T1	8	0
40	600084823T1	18	1
41	600081801T1	18	0
42	600081013T1	11	1
43	600080875T1	18	0
44	600080802T1	15	1
45	600079708T1	17	0
46	600079015T1	31	0
47	600078119T1	11	1
	Total	**920**	**9**

Table 3.2: Summary of secondary structure data analysis and results.

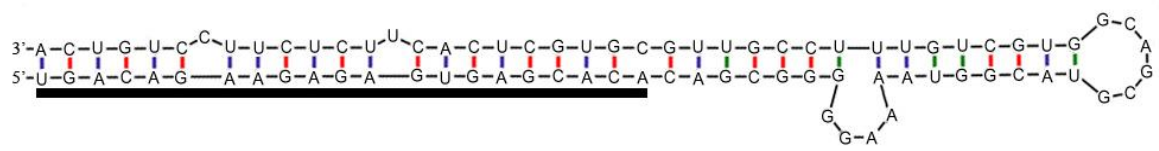
miRNA family	EST sequence name	EST length (nt)	A	C	G	U	N	(A+U) %	MFEs (kcal/mol)
miR156	gi|146226345|	789	138	171	230	250	0	49.18	-315.21
miR169	600081013T1	397	111	99	70	102	15	53.65	-97
	600092447T1	524	139	137	102	141	5	53.44	-152.41
	600092884T1	555	143	151	104	149	8	52.61	-159.38
	600078119T1	414	111	102	99	102	0	51.45	-137.82
	600080802T1	615	190	159	117	149	0	55.12	-146.49
miR172	600084823T1	497	148	112	96	132	9	56.34	-98.62
	600103944T1	567	155	132	120	159	1	55.38	-132.28
	600160442T1	397	104	90	100	101	2	51.64	-103.92

Table 3.3: Characteristics of candidate miRNA precursor sequences.

MFE: minimal free energies

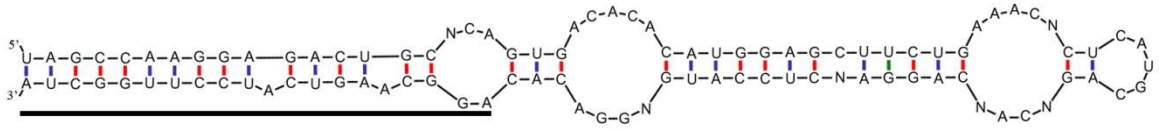
miRNA family: miR 156

(i) Sequence name: gi|146226345|

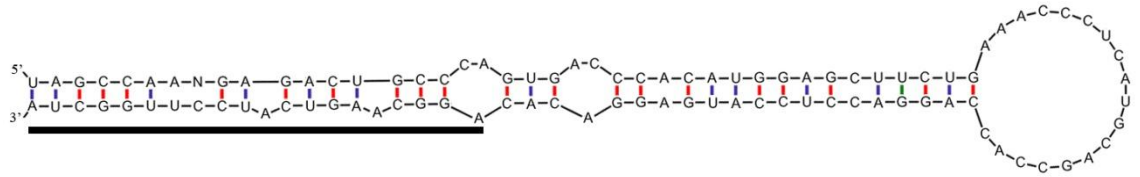


miRNA family: miR 169

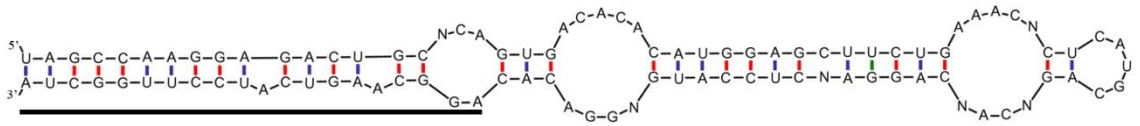
(ii) Sequence name: 600081013T1



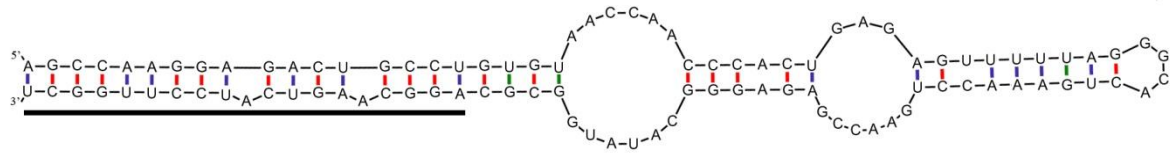
(iii) Sequence name: 600092447T1



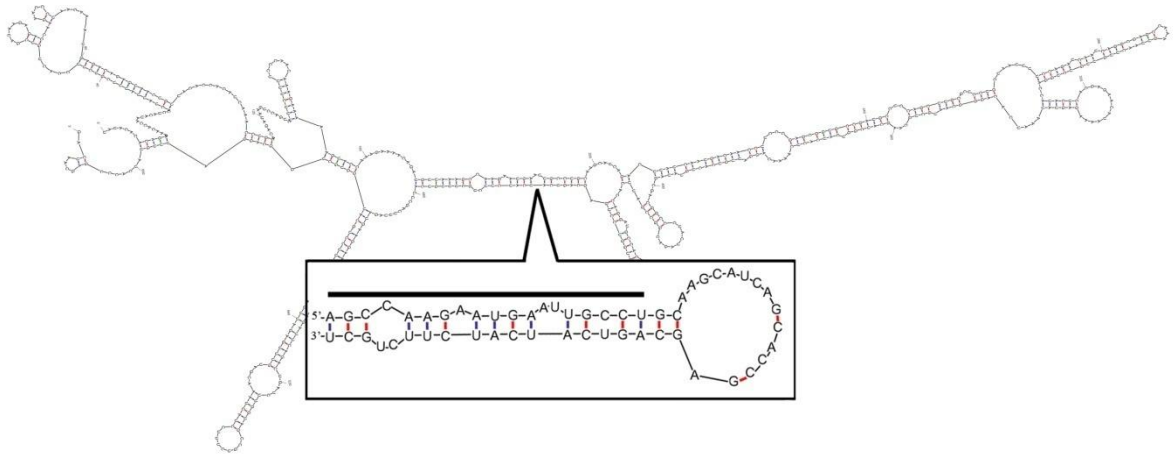
(iv) Sequence name: 600092884T1



(v) Sequence name: 600078119T1

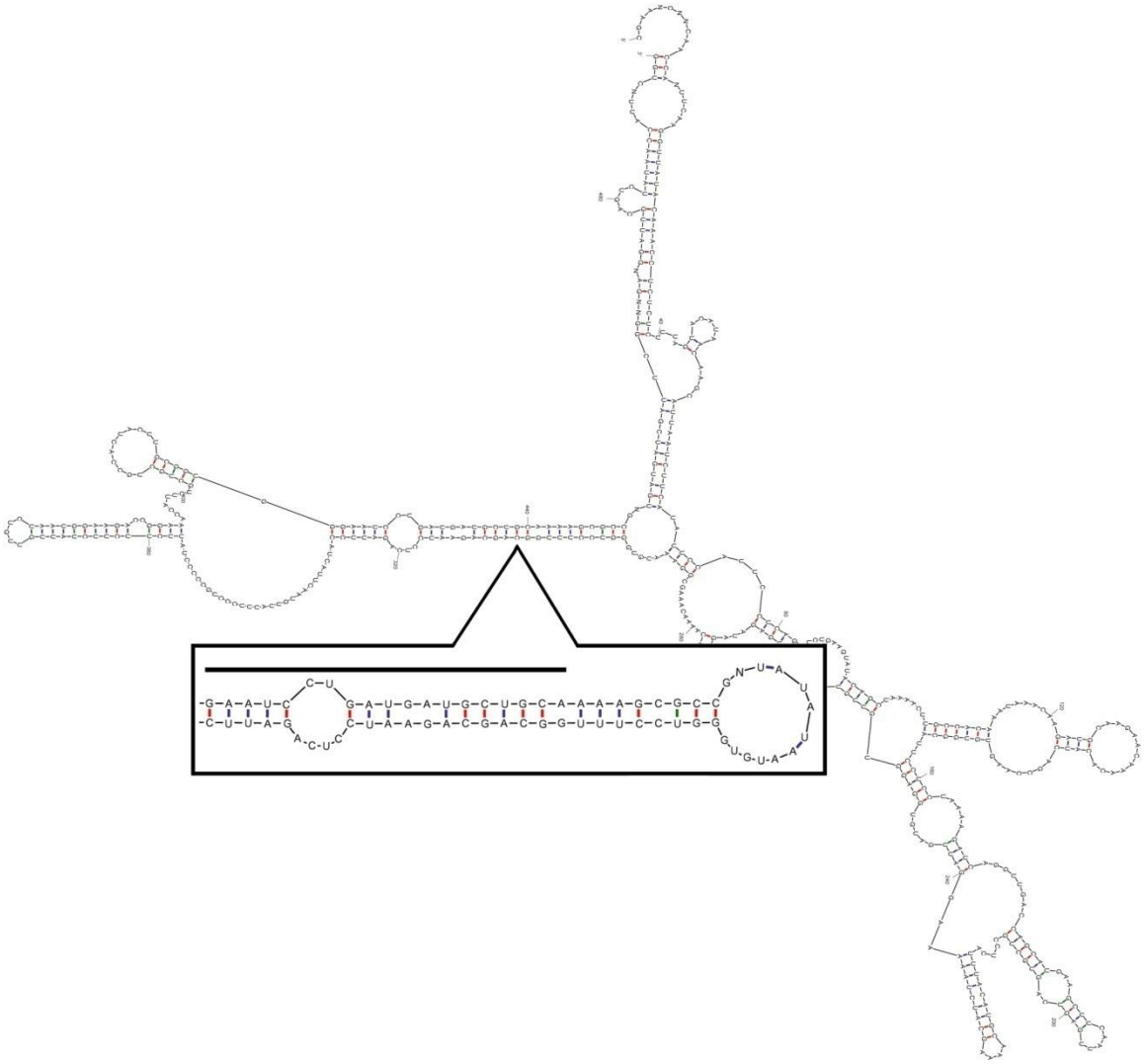


(vi) Sequence name: 600080802T1

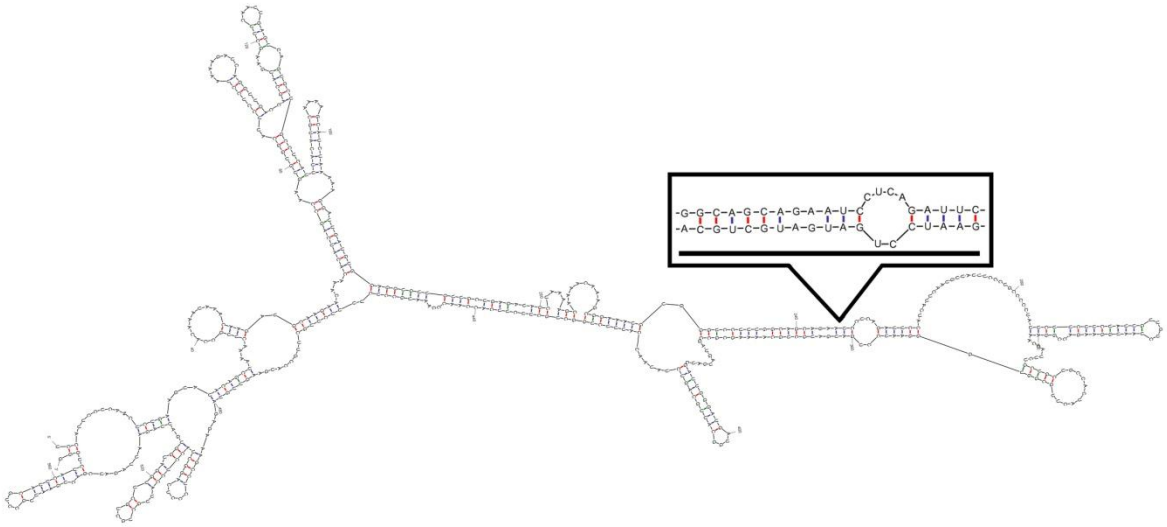


miRNA family: miR 172

(vii) Sequence name: 600084823T1



(viii) Sequence name: 600103944T1



(ix) Sequence name: 600160442T1

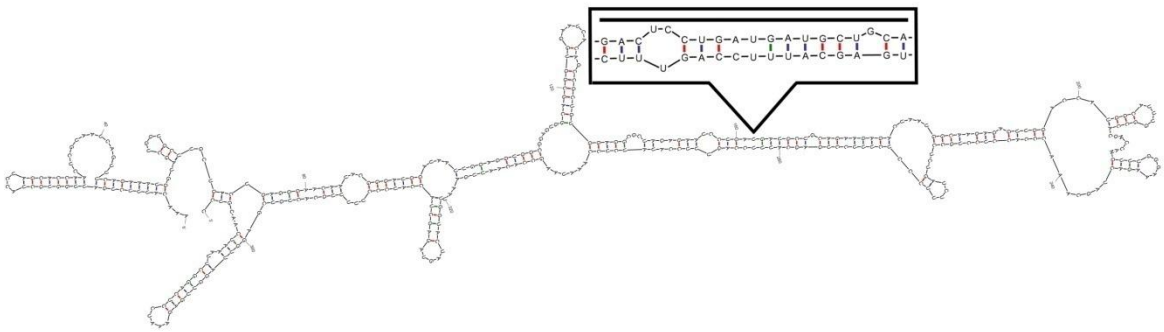


Figure 3.1 (i – ix): The predicted secondary structures of the selected nine candidate miRNA sequences using Mfold. The sequences of the predicted mature miRNA are indicated by the lines above or below the strand in which the mature miRNA is encoded.

3.3 RNA extraction of *Musa* sp.

The RNA extraction was carried out using the CTAB-NETS method (mentioned in 2.5.1 section of Materials and Methods). This step is important in order to obtain pure RNA samples which will be then further used to confirm the expression of the predicted miRNAs in *Musa* sp. Samples were ran twice to increase accuracy and precision, to make sure the samples were clean and without any sign of degradation.

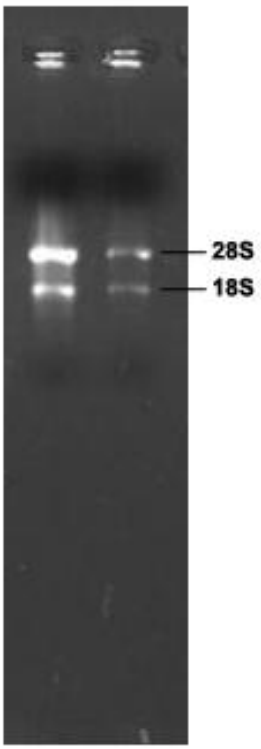


Figure 3.2: Agarose gel of RNA samples from *Musa* sp. Left: Leaf sample, Right: Root sample.

Results are as observed in Figure 3.2. These samples, which were **separated on a** 1% agarose gel, produced two visible bands in each lane that corresponded to the expected sizes of two ribosomal RNAs (rRNAs): 18 S and 28 S respectively. Observations of the clear bands indicate that the RNA samples were high quality and did not show signs of much degradation.

3.4 Reverse Transcription PCR

RT-PCR was performed using the Access RT-PCR System Kit (Promega Corporation, USA) (mentioned in section 2.5.5 of the Materials and Methods) to confirm the expression of the predicted miRNAs in *Musa* sp.

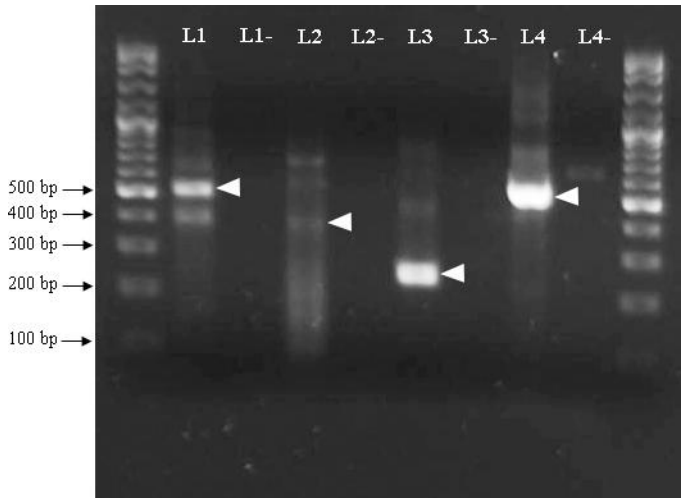


Figure 3.3: Agarose gel of RT-PCR products with RNA isolated from banana leaves.

L1: primers miR targeting region of EST gi|146226345| (with miR156),

L2: primers miR targeting region of EST 600078119T1 (with miR169),

L3: primers targeting region of EST 600092884T1 (with miR169),

L4: primers targeting ß-actin gene, a house keeping gene,

L1- to L4-: RT negative controls.

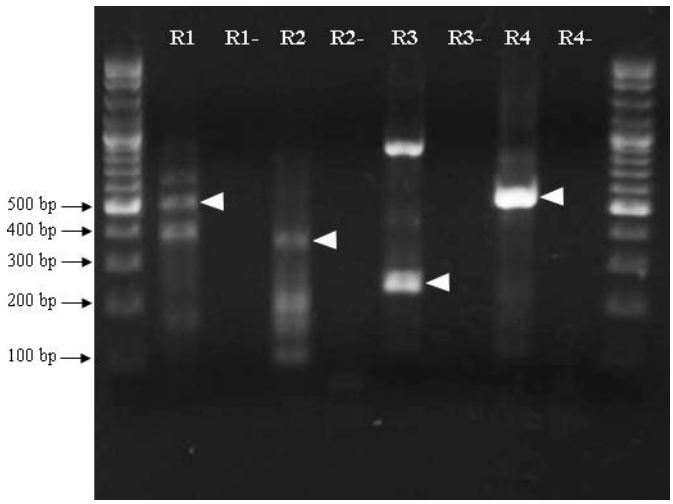


Figure 3.4: Agarose gel of RT-PCR products with RNA isolated from banana roots.

R1: primers miR targeting region of EST gi|146226345| (with miR156),

R2: primers miR targeting region of EST 600078119T1 (with miR169),

R3: primers targeting region of EST 600092884T1 (with miR169),

R4: primers targeting ß-actin gene, a house keeping gene,

R1- to R4-: RT negative controls.

Results are as observed in Figure 3.3 & 3.4. It can be observed that a band of the expected size was obtained for each of the miRNA candidates which was ~500bp, ~350bp and ~300bp for EST sequences gi|146226345, 600078119T1 and 600092884T1 respectively. These results show that the computationally predicted miRNAs of *Musa sp*. were likely expressed in the leaf and root samples.

3.5 DNA sequence analysis

Sequencing results of the RT-PCR amplified cDNAs from *Musa acuminata* var. *Berangan* were shown in Appendix 3. Sequence alignments were carried out for PCR products and ESTs of *Musa sp*. of the three families – miR156, miR169 and miR172 as shown in Figures 3.5, 3.6 and 3.7 to determine the percentages of similarities between the PCR product and EST sequences. Alignments between the PCR products and ESTs show high percentages of similarities. For miR families 156 (Figure 3.5) and 172 (Figure 3.7), the similarities were an almost perfect match with sequence identities of 99%. The miR169 family (Figure 3.6) too showed high similarity between the PCR product and EST sequence with a sequence identity of 98%.

In addition, multiple sequence alignments were carried out with other known mature and precursor of miR families from different plant species as shown in Figures 3.8, 3.9, 3.10 and 3.11 to determine the similarities and conservation of the different plant species. Sequence alignments of the miR156 family candidates from *Musa sp.* with other known mature miR156 sequences from plant species including *Arabidopsis thaliana, Brassica napus, Oryza sativa, Populus trichocarpa* and *Zea mays* (Figure 3.8) showed high similarities of around 90% across the plant species. Beside that, sequence alignments of the miR169 family candidates from *Musa sp.* with other known mature miR169 sequences from plant species including *Arabidopsis thaliana, Oryza sativa, Populus trichocarpa* and *Zea mays* (Figure 3.10) had a similarity percentage of 76%. However, both sequence alignments of the miR156 and miR169 family candidates from *Musa sp.* with precursors of respective family from the above mentioned plant species Figures 3.9 and 3.11), did not show an overall high similarity and only the mature

miRNA and complementary miRNA sequences (miRNA*) have been found to be highly conserved across all species.

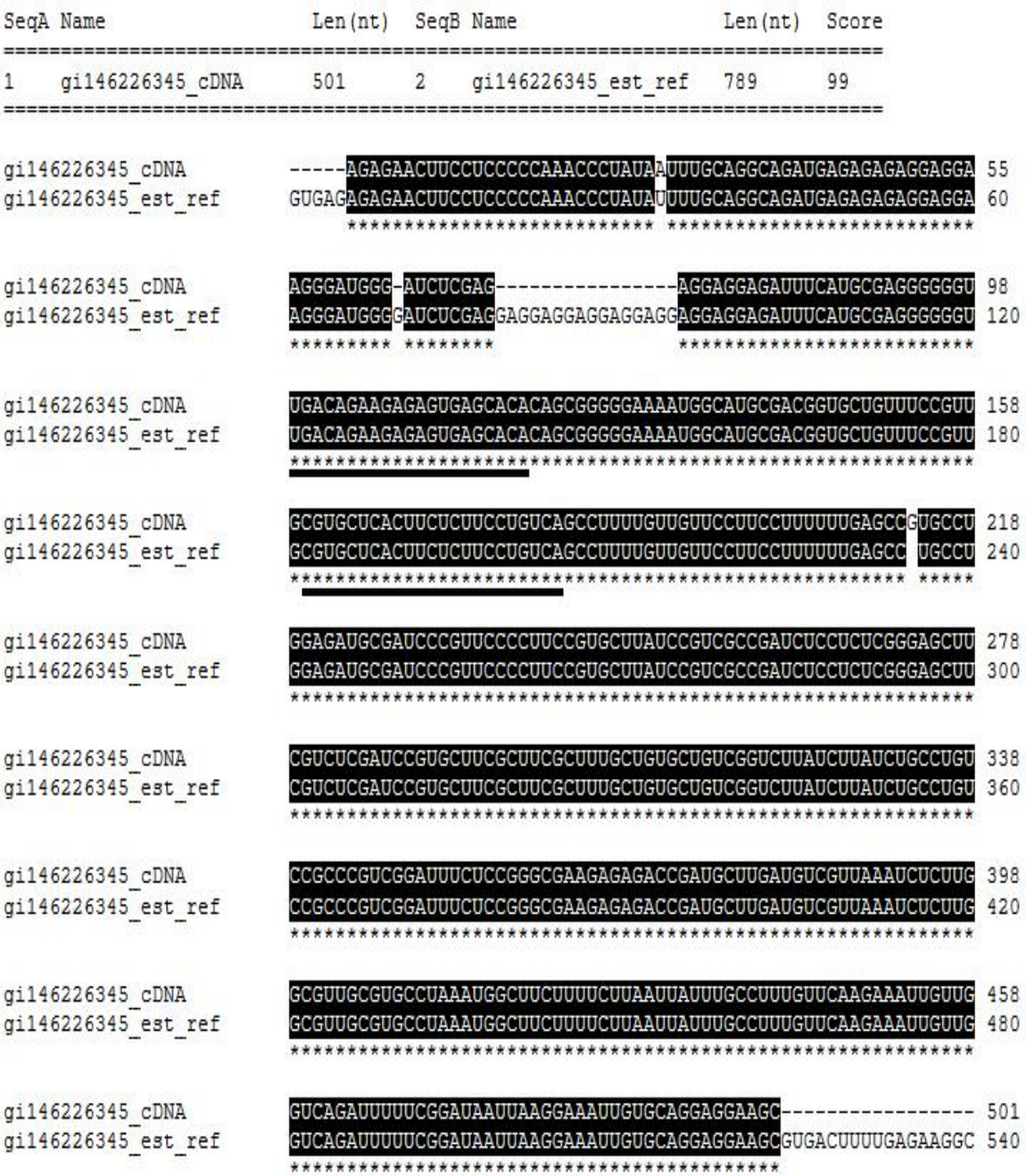


Figure 3.5: Sequence alignment of *Musa sp.* EST gi|146226345| and RT-PCR amplified cDNAs from *Musa acuminata* var. *Berangan*. The underlined areas indicate the mature miRNA156 sequence (top) and the miRNA* sequence (below).

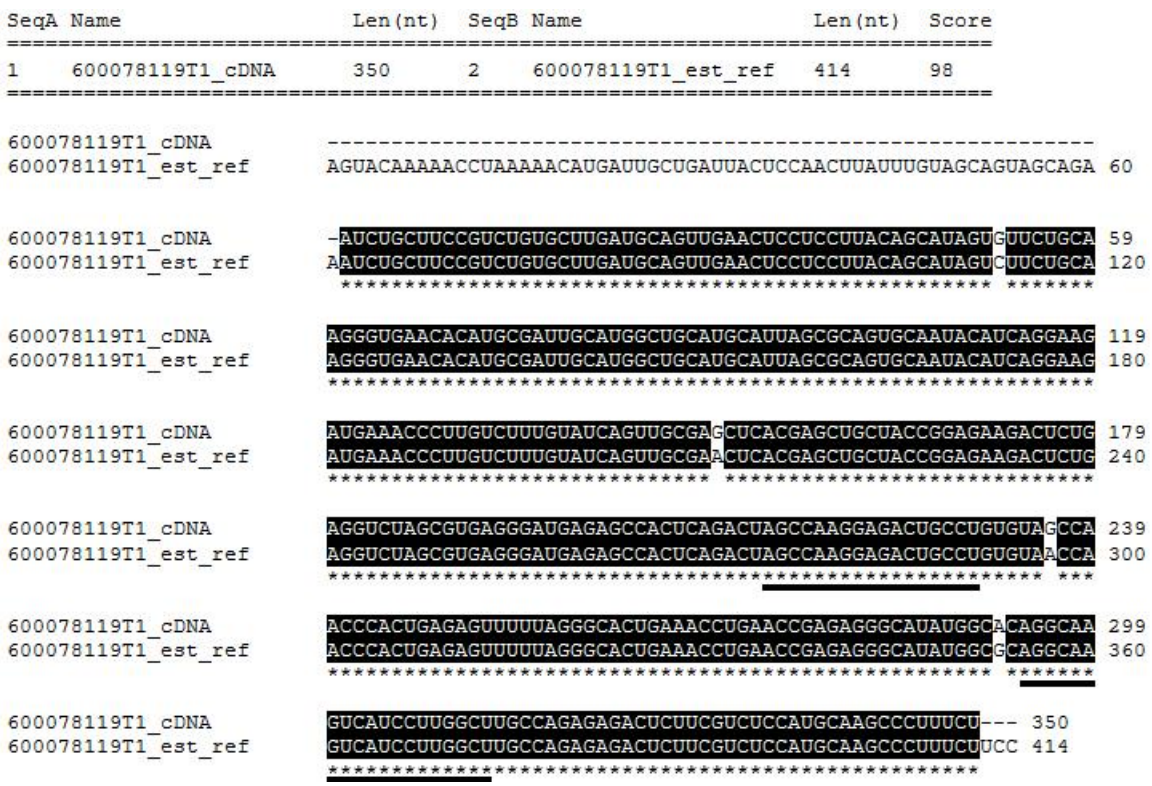


Figure 3.6: Sequence alignment of *Musa sp.* EST 600078119T1 and RT-PCR amplified cDNAs from *Musa acuminata* var. *Berangan*. The underlined areas indicate the mature miRNA169 sequence (top) and the miRNA* sequence (below).

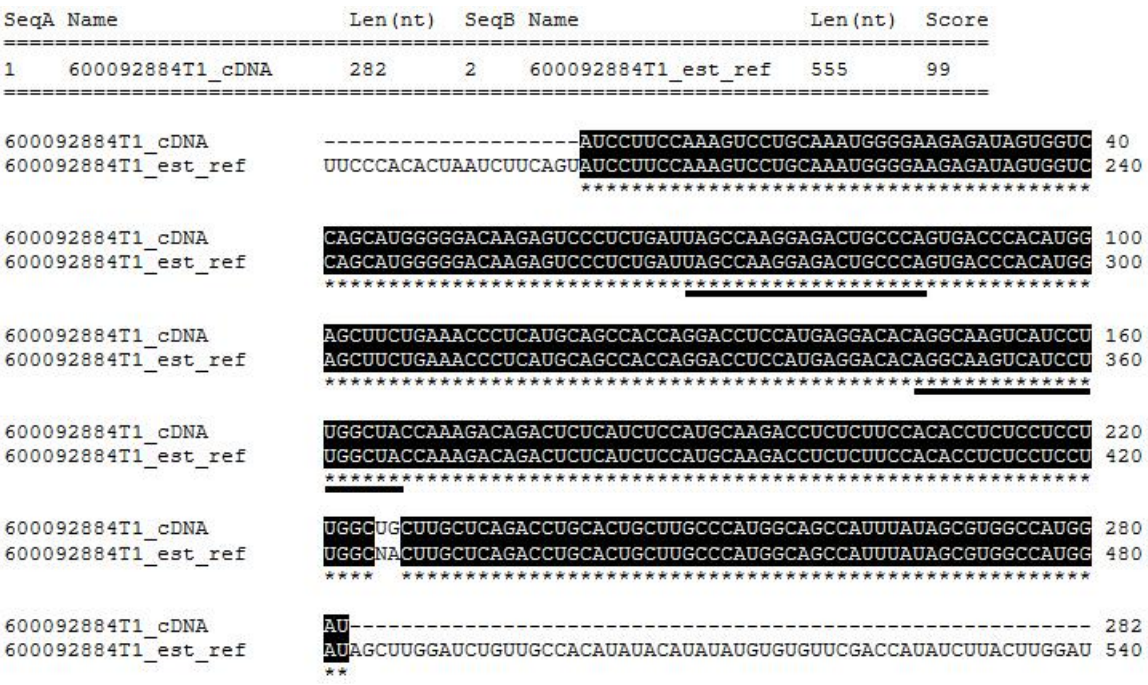


Figure 3.7: Sequence alignment of *Musa sp.* EST 600092884T1 and RT-PCR amplified cDNAs from *Musa acuminata* var. *Berangan*. The underlined areas indicate the mature miRNA169 sequence (top) and the miRNA* sequence (below).

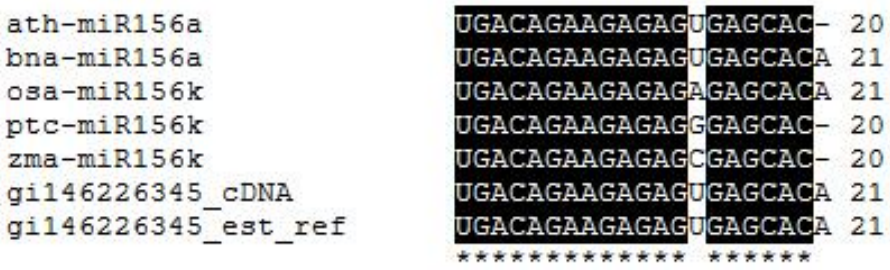


Figure 3.8: Multiple sequence alignment of mature miRNA *Musa sp.* EST gi|146226345| and RT-PCR amplified cDNAs from *Musa acuminata* var. *Berangan* with miR156 families from different plant species, ath-MIR156a: *Arabidopsis thaliana*; bna-MIR156a: *Brassica napus*; osa-MIR156k: *Oryza sativa*; ptc-MIR156k: *Populus trichocarpa*; zma-MIR156k: *Zea mays*.

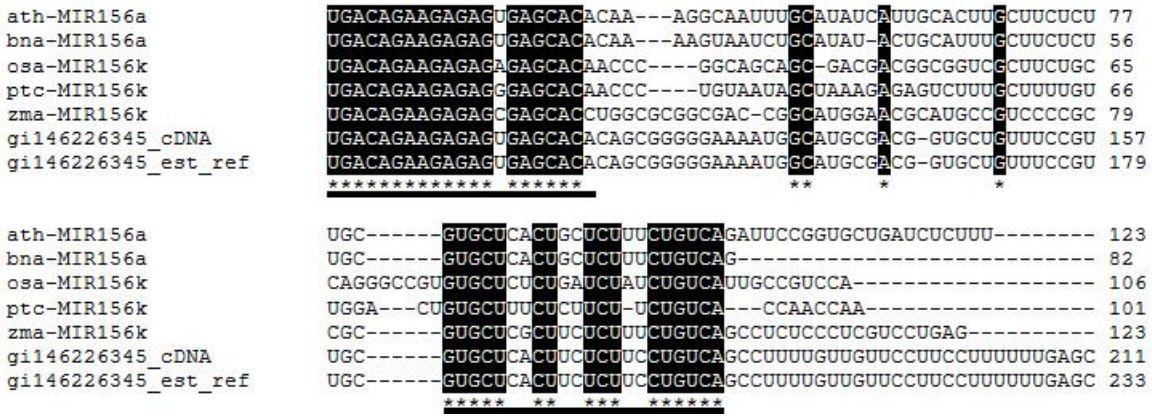


Figure 3.9: Multiple sequence alignment of *Musa sp.* EST gi|146226345| and RT-PCR amplified cDNAs from *Musa acuminata* var. *Berangan* with precursors of miR156 from different plant species, ath-MIR156a: *Arabidopsis thaliana*; bna-MIR156a: *Brassica napus*; osa-MIR156k: *Oryza sativa*; ptc-MIR156k: *Populus trichocarpa*; zma-MIR156k: *Zea mays*. The underlined areas indicate the mature miRNA sequence (top) and the miRNA* sequence (below).

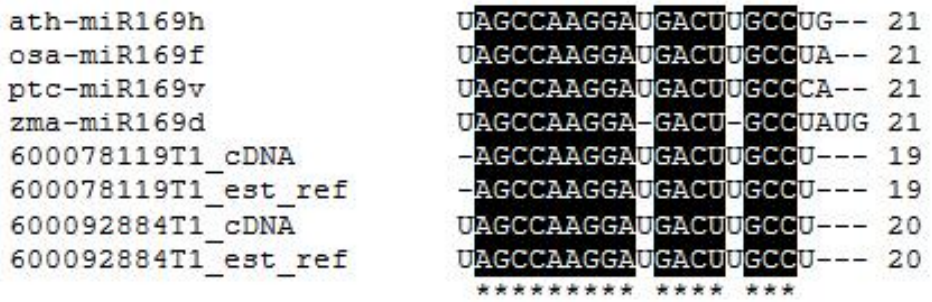


Figure 3.10: Multiple sequence alignment of mature miRNA *Musa sp.* EST 600078119T1, EST 600092884T1 and RT-PCR amplified cDNAs from *Musa acuminata* var. *Berangan* with miR169 families from different plant species, ath-MIR169h: *Arabidopsis thaliana*; osa-MIR169f: *Oryza sativa*; ptc-MIR169v: *Populus trichocarpa*; zma-MIR169d: *Zea mays*.

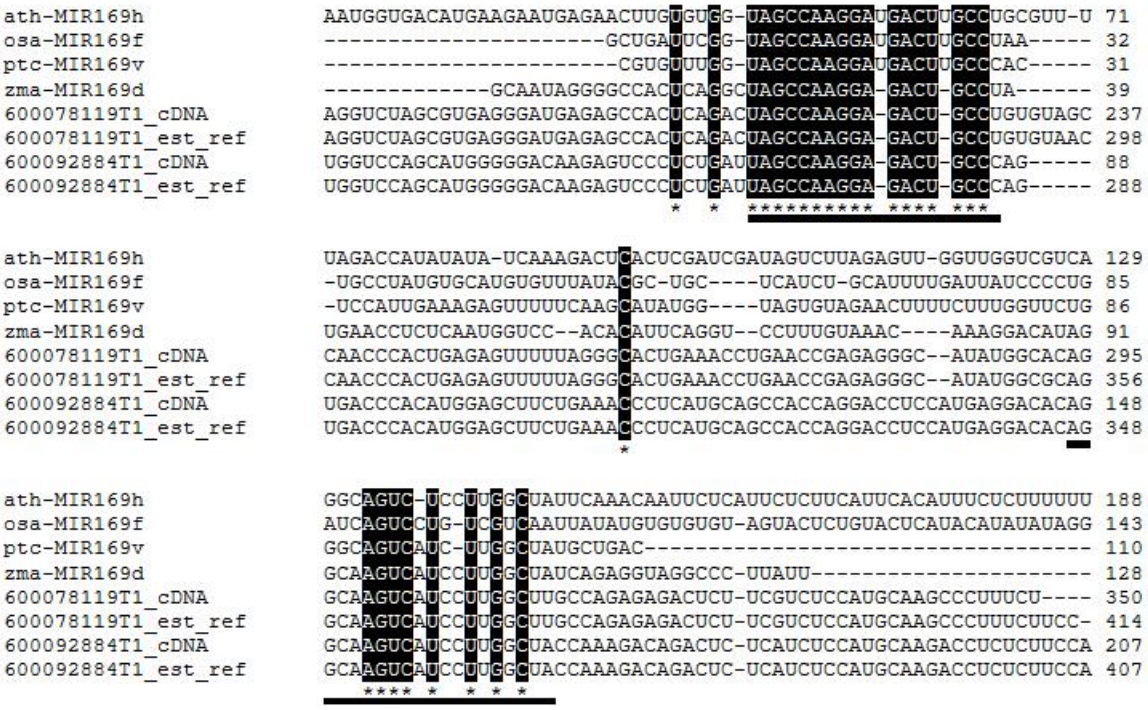


Figure 3.11: Multiple sequence alignment of *Musa sp.* EST 600078119T1, EST 600092884T1 and RT-PCR amplified cDNAs from *Musa acuminata* var. *Berangan* with precursors of miR169 from different plant species, ath-MIR169h: *Arabidopsis thaliana*; osa-MIR169f: *Oryza sativa*; ptc-MIR169v: *Populus trichocarpa*; zma-MIR169d: *Zea mays*. The underlined areas indicate the mature miRNA sequence (top) and the miRNA* sequence (below).

CHAPTER 4:

DISCUSSION

4.0 Discussion

Homology-based search methods first identify sequences that match known miRNA mature sequences. The genomic context of such sequences is then extracted and the candidates aligned with their assumed miRNA families. This step is followed by the application of certain criteria used to determine the final list of candidate homologs (Mendes et al, 2009). In this study, the model organisms used as miRNA reference sets included: *Arabidopsis thaliana* (184), *Brassica napus* (5), *Glycine max* (22), *Medicago truncatula* (30), *Oryza sativa* (242), *Physcomitrella patens* (77), *Populus trichocarpa* (215), *Saccharum officinarum* (16), *Sorghum bicolor* (72) and *Zea mays* (96) which were obtained from the miRNA Registry Database, Wellcome Trust Sanger Institute (Release 9.2, May 2007). This brings a total of 959 plant miRNA sequences.

In plants, there have been many approaches used to understand the complexity of gene expression as well as interaction. One such approach is the EST method. EST sequencing has been used as an efficient and economical approach for large-scale gene discovery and has proven to be a rapid and effective way of acquiring information. Information obtained can be from a wide variety of tissues, cell types or development stages, covering gene diversity and mRNA expression patterns (Low et al, 2008; Weng et al, 2005). Over five million plant EST sequences have been generated from more than 60 plant species and made available to the public. These species cover most of the plant kingdom including both model and crop plants like *Arabidopsis thaliana*, rice, wheat and maize (Weng et al., 2005; De Peer, 2005).

For this study, a total of 42,978 *Musa sp*. EST sequences were obtained from the Global Musa Genomics Consortium (http://www.musagenomics.org/) and the National Centre for Biotechnology Information (NCBI) GenBank nucleotide databases

(http://www.ncbi.nlm.nih.gov/). From this total, 9 predicted miRNA sequences were eventually obtained. The predicted secondary structures for these sequences were found to have high negative minimal free energies (MFEs), were able to fold such that the mature miRNA sequences were within one arm of the hairpin secondary structure and with no loops or breaks in the mature miRNA sequences. The candidate miRNAs also had at least 30-70% contents of A+U, less than six mismatches with their corresponding miRNA* sequence and no loop or bulge in mature miRNA:RNA* dimers containing a more than three nucleotides that were not involved in canonical base pairing (i.e. GU, GU and AU pairing) (Qiu et al., 2007; Nasaruddin et al., 2007).

The number of miRNA obtained through EST analysis depends on the total amount of ESTs available in the database; it has been estimated that about 10,000 ESTs contain 1 to 1.67 miRNA (Zhang et al., 2006b; Qiu et al., 2007; Xie et al., 2007). Therefore, it was reasonable to expect one or more miRNA in the collection of 42,978 *Musa sp*. EST. When queried against the miRNA reference sets using BLAST, a total of 47 candidate sequences were obtained, from which 920 EST secondary sequences were generated. From these 920 EST secondary sequences, the number of predicted miRNA sequences found was nine. Therefore, the predicted sequences in the 42,978 *Musa sp.* EST sequences was 9/42,978, showing a detection of 2.09/10,000 ESTs, a slightly higher value than that reported before (Zhang et al., 2006b; Qiu et al., 2007; Xie et al., 2007). In a previous study (Nasaruddin et al., 2007), 6.86 miRNAs per 10,000 ESTs was reported for *Elaeis guineensis* Jacq., and this higher rate of detection was due to the method and plant materials used in the study to generate the ESTs. These ESTs were isolated from the leaves and roots of *Musa sp*. that were from several different development stages, thus improving the representation of low abundance and

differentially regulated transcripts from the selected plant tissues and enriching the miRNA sequences data (Nasaruddin et al, 2007).

Computational approaches such as the ones used in this study are good approaches to predict miRNAs; however their applications are limited as they require genome sequences, which currently are only available for a limited number of plant species. ESTs are a powerful tool that has been used to identify miRNA homologs, one reason being that many identified miRNAs are evolutionarily conserved from species to species. EST analysis could actually provide a deeper insight into the distribution and conservation of miRNAs as well as being used to identify miRNAs in any species – whether their genome sequence is available or not.

4.1 miRNA genes in *Musa sp*.

Identification of miRNAs can be further used to determine the identity of the mRNAs that they regulate. The miRNA pathway is important for the proper regulation of plant development and miRNA genes play a variety of roles including the regulation of organ development such as leaf morphogenesis, floral organ identity and root development as well as in various stress responses such as dehydration, mineral-nutrient and mechanical stress (Yang et al, 2007). In this study, the nine candidate miRNA precursors potentially belong to the miR156, miR169 and miR172 families.

The predicted *Musa sp*. member of the miR156 family was EST sequence gi|146226345|. This sequence, which consists of 21 nucleotides, has the strongest characteristics of miRNA amongst other EST candidates with no mismatches. It has a minimal free energy of -315.21 and an (A+U)% of 49.18%. All 21 nucleotides are engaged in the canonical (GU, GA and AU) base pairing of the mature

miRNA:miRNA* dimer in the hairpin structure and without any bulge or loop in the secondary structure. Members of the miR156 family have been found to target Squamosa promoter Binding Protein (SBP) - like transcription factor proteins that are important in regulating flowering time (Rhoades et al., 2002; Mallory and Vaucheret, 2006)

EST Sequences found within the miR169 family included: 600081013T1, 600092447T1, 600092884T1, 600078119T1 and 600080802T1, which had minimal free energies of -97, -152.41, -159.38, -137.82 and -146.49 respectively. These sequences also had (A+U)% that fell within the range with 53.65%, 53.44%, 52.61%, 51.45% and 55.12% respectively. Of all these sequences only 600081013T1, 600092447T1 and 600092884T1 had one mismatch, which meant that the rest of the nucleotides were engaged in base pairing. The miR169 gene family, which target HAP2 transcription factors, has been found to be conserved from ferns to eudicots. Studies have shown that in alfalfa, miR169 targets a HAP2 gene which is important for root nodulation (Willmann and Poethig, 2008).

The predicted *Musa sp*. member of the miR172 family were EST sequences 600084823T1, 600103944T1 and 600160442T1, which have minimal free energies of -98.62, -132.28 and -103.92 respectively. With all three sequences having 3 mismatches each, the remaining nucleotides are engaged in the canonical (GU, GA and AU) base pairing of the mature miRNA:miRNA* dimer and have (A+U)% of 56.34%, 55.38% and 51.64% respectively. The miR172 gene family is involved in the regulation of flowering time and floral organ identity through the regulation of APETALA2 (AP2) – like genes' activity. AP2 play an important role during floral development in establishing meristem and organ identity (Chen et al., 2006; Glazinska et al., 2009).

4.2 Experimental validation of predicted miRNAs

Computational strategies are a useful method to predict miRNAs and their targets and have shown successful results for vertebrates, insects, and plants (Ambros, 2004; Adai et al., 2005; Li et al., 2005). However, many plant miRNA genes and their targets have yet to be discovered, particularly for less well studied genomes. Traditional computational approaches have been slightly inefficient and tedious causing difficulties in demonstrating the expression of predicted miRNAs (Zhang et al., 2007). Therefore, experimental approaches were further used to validate and support the results of the identification.

There are a few different techniques that are available for the detection and validation of miRNA and their candidates including Northern blots and other techniques using PCR. In this study, the RT-PCR technique was used to verify the expression of the predicted miRNAs in leaf and root samples of *Musa sp*. Here, RT-PCR was performed with three pairs of PCR primers, which were designed based on the three selected candidate miRNA EST sequences of *Musa sp*. From the RT-PCR gel results (Figures 3.3 and 3.4), it can be observed that a single band of the expected size was obtained for each of the miRNA candidates which was ~500bp, ~350bp and ~300bp for EST sequences gi|146226345, 600078119T1 and 600092884T1 respectively. These results show that the computationally predicted miRNAs of *Musa sp*. were expressed in the leaf and root samples.

4.2.1 DNA sequencing and sequence alignments

The comparisons of sequences are one of the important computational tools in molecular biology as they represent biological molecules that encode significant information which include structure, function and history. They can be used to identify

genes or regularity sites across evolutionarily related genomes as these functional sequences often show conservational patterns that are different from those seen in regions that are not functional. However, comparisons at sequence level have their limitations. Similar sequences have been reported to play similar biological functions although the opposite is not always true. This is because similar functions are encoded in higher order sequence elements that are not always conserved at the sequence level. Therefore, similar functions are often encoded by different sequences and cannot be detected by conventional sequence alignment methods (Blanco et al, 2007).

From the results obtained, it can be noted that sequences of the PCR products and ESTs of *Musa sp*. of the three families – miR156, miR169 and miR172 (Figures 3.5, 3.6 and 3.7) – were found to be highly matched. Alignments between the PCR products and ESTs show high percentages of similarities. For miR families 156 and 172, the similarities were an almost perfect match with sequence identities of 99%. The miR169 family too showed high similarity between the PCR product and EST sequence with a sequence identity of 98%.

Sequence alignments of the miR156 family candidates from *Musa sp.* with other known mature miR156 sequences from plant species including *Arabidopsis thaliana, Brassica napus, Oryza sativa, Populus trichocarpa* and *Zea mays* showed high similarities of around 90% across the plant species. Beside that, sequence alignments of the miR169 family candidates from *Musa sp.* with other known mature miR169 sequences from plant species including *Arabidopsis thaliana, Oryza sativa, Populus trichocarpa* and *Zea mays* had a similarity percentage of 76%. Result of both this sequence alignments of the miR156 and miR169 confirms that the findings of other

studies that the mature miRNA are highly conserved among land plants (Zhang et al., 2005a; Willmann and Poethig, 2008).

However, both sequence alignments of the miR156 and miR169 family candidates from *Musa sp.* with precursors of respective family from the above mentioned plant species, did not show an overall high similarity and only the mature miRNA and complementary miRNA sequences (miRNA*) have been found to be highly conserved across all species. This showed that precursor sequences are not well conserved across all species due perhaps to the different origins and selection pressure of the plant species. The divergences of miRNA sequences between species were believed to be involved in the regulation of various specific processes (Yang et al., 2007).

In addition, precursors of miRNA were found to consist of more A+U bonds, thus forming more base pairs in their secondary structures. More hydrogen bonds and base pairs formed within miRNA precursor sequences leads to more stable miRNA hairpin precursors, which in turn may prove beneficial in the transport of miRNA precursors from the nucleus to the cytoplasm. This may also help in the recognition of stem-loop structures by the Dicer-like enzymes, thus cutting miRNA precursors into miRNA:miRNA* duplexes which then form mature miRNAs (Zhang et al., 2005b).

CHAPTER 5:

CONCLUSION

5.0 Conclusion

By using publicly available EST data and the high degree of evolutionary conservation of mature miRNAs within the plant kingdom, a computational approach was used to predict and identify novel miRNA orthologs in other plant species. In this study, from a database of 42,978 *Musa sp*. EST sequences, nine novel mature miRNAs were identified and three selected novel mature miRNAs had been experimentally validated for the miRNA expression in *Musa acuminata* var. *Berangan*. The potential roles of these novel miRNAs in *Musa sp*. include regulation of flowering time, root nodulation and floral development. All this information is crucial for the development of this major crop plant. In future, these novel miRNAs should undergo further experimental identification and validation for their target in *Musa sp*.

APPENDICES

Appendix 1: BLAST results of 47 potential miRNAs of *Musa sp*.

Query= gi|146228486|gb|ES436957.1|ES436957 EST1237163 ESTSYN-L Musa acuminata cDNA clone pESTSYNL-10F14 5', mRNA sequence. Length = 961

>osa-miR393 MIMAT0000957, length = 21

Score = 36.2 bits (18), Expect = 5e-005

Identities = 18/18 (100%)

Strand = Plus / Plus

```
Query: 530 tccaaagggatcgcattg 547
           ||||||||||||||||||
Sbjct: 1   tccaaagggatcgcattg 18
```

Query= gi|146226345|gb|ES434836.1|ES434836 EST1235042 ESTSYN-L Musa acuminata cDNA clone pESTSYNL-4M20 5', mRNA sequence. Length = 789

>bna-miR156a MIMAT0004445, length = 21

Score = 42.1 bits (21), Expect = 7e-007

Identities = 21/21 (100%)

Strand = Plus / Plus

```
Query: 121 tgacagaagagagtgagcaca 141
           |||||||||||||||||||||
Sbjct: 1   tgacagaagagagtgagcaca 21
```

>osa-miR156l MIMAT0001021, length = 21

Score = 36.2 bits (18), Expect = 4e-005

Identities = 18/18 (100%)

Strand = Plus / Plus

```
Query: 122 gacagaagagagtgagca 139
           ||||||||||||||||||
Sbjct: 2   gacagaagagagtgagca 19
```

>osa-miR156k MIMAT0001020, length = 21

Score = 34.2 bits (17), Expect = 2e-004

Identities = 20/21 (95%)

Strand = Plus / Plus

```
Query: 121 tgacagaagagagtgagcaca 141
           ||||||||||||| |||||||
Sbjct: 1   tgacagaagagagagagcaca 21
```

Query= gi|146222506|gb|ES431747.1|ES431747 EST1231954 ESTSYN-F Musa acuminata cDNA clone pESTSYNF-4I7 5', mRNA sequence. Length = 925

>osa-miR393 MIMAT0000957, length = 21

Score = 36.2 bits (18), Expect = 5e-005

Identities = 18/18 (100%)

Strand = Plus / Minus

```
Query: 425 caatgcgatcccttgga 442
           ||||||||||||||||
Sbjct: 18  caatgcgatcccttgga 1
```

Query= gi|146223567|gb|ES432808.1|ES432808 EST1233015 ESTSYN-F Musa acuminata cDNA clone pESTSYNF-8N18 5', mRNA sequence. Length = 699

>osa-miR156k MIMAT0001020, length = 21

Score = 38.2 bits (19), Expect = 9e-006

Identities = 19/19 (100%)

Strand = Plus / Minus

```
Query: 524 tgctctctctcttctgtca 542
           ||||||||||||||||||
Sbjct: 19  tgctctctctcttctgtca 1
```

>bna-miR156a MIMAT0004445, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Minus

```
Query: 524 tgctctctctcttctgtca 542
           ||||| ||||||||||||
Sbjct: 19  tgctcactctcttctgtca 1
```

>osa-miR156l MIMAT0001021, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Minus

```
Query: 523 atgctctctctcttctgtc 541
           |||||| ||||||||||| 
Sbjct: 20  atgctcactctcttctgtc 2
```

Query= gi|70796212|gb|DN238918.1|DN238918 MUC4LH1013_G07_.b_052. Musa acuminata leaves HOT stress Musa acuminata subsp. burmannicoides cDNA 5', mRNA sequence. Length = 747

>osa-miR160e MIMAT0001030, length = 21

Score = 40.1 bits (20), Expect = 2e-006

Identities = 20/20 (100%)

Strand = Plus / Minus

```
Query: 332 ggcatacagggagccaggca 351
           ||||||||||||||||||||
Sbjct: 20  ggcatacagggagccaggca 1
```

>osa-miR160f MIMAT0001031, length = 21

Score = 32.2 bits (16), Expect = 6e-004

Identities = 19/20 (95%)

Strand = Plus / Minus

```
Query: 332 ggcatacagggagccaggca 351
           ||||| ||||||||||||||
Sbjct: 20  ggcattcagggagccaggca 1
```

Query= gi|146226218|gb|ES434709.1|ES434709 EST1234915 ESTSYN-L Musa acuminata cDNA clone pESTSYNL-4G9 5', mRNA sequence. Length = 947

>osa-miR156k MIMAT0001020, length = 21

Score = 42.1 bits (21), Expect = 8e-007

Identities = 21/21 (100%)

Strand = Plus / Minus

```
Query: 858 tgtgctctctctcttctgtca 878
           |||||||||||||||||||||
Sbjct: 21  tgtgctctctctcttctgtca 1
```

>bna-miR156a MIMAT0004445, length = 21

Score = 34.2 bits (17), Expect = 2e-004

Identities = 20/21 (95%)

Strand = Plus / Minus

```
Query: 858 tgtgctctctctcttctgtca 878
           ||||||| ||||||||||||
Sbjct: 21  tgtgctcactctcttctgtca 1
```

Query= gi|146228497|gb|ES436968.1|ES436968 EST1237174 ESTSYN-L Musa acuminata cDNA clone pESTSYNL-10H12 5', mRNA sequence. Length = 733

>osa-miR172c MIMAT0001071, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Minus

```
Query: 338 tgcagcatcatcaggattc 356
           ||||||||||||| |||||
Sbjct: 20  tgcagcatcatcaagattc 2
```

Query= gi|146222802|gb|ES432043.1|ES432043 EST1232250 ESTSYN-F Musa acuminata cDNA clone pESTSYNF-5D19 5', mRNA sequence. Length = 735

>osa-miR396d MIMAT0001600, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 252 tccacatgctttcttgaac 270
           |||||| ||||||||||
Sbjct: 1   tccacaggctttcttgaac 19
```

Query= gi|146222248|gb|ES431489.1|ES431489 EST1231695 ESTSYN-F Musa acuminata cDNA clone pESTSYNF-2J22 5', mRNA sequence. Length = 781

>mtr-miR395b MIMAT0001649, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 702 tgaagtatttggtggaact 720
           |||||||||||| ||||||
Sbjct: 2   tgaagtatttgggggaact 20
```

>osa-miR395v MIMAT0003876, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 702 tgaagtatttggtggaact 720
           ||||||||||| ||||||
Sbjct: 2   tgaagtatttggcggaact 20
```

Query= gi|146222618|gb|ES431859.1|ES431859 EST1232066 ESTSYN-F Musa acuminata cDNA clone pESTSYNF-4L19 5', mRNA sequence. Length = 803

>osa-miR171i MIMAT0001085, length = 21

Score = 32.2 bits (16), Expect = 6e-004

Identities = 19/20 (95%)

Strand = Plus / Minus

```
Query: 33 gatattgtcgcggctcaatc 52
          ||||||| ||||||||||||
Sbjct: 21 gatattgacgcggctcaatc 2
```

>osa-miR171b MIMAT0001063, length = 21

Score = 26.3 bits (13), Expect = 0.040

Identities = 19/21 (90%)

Strand = Plus / Minus

```
Query: 33 gatattgtcgcggctcaatca 53
          ||||||| | ||||||||||
Sbjct: 21 gatattggcacggctcaatca 1
```

Query= 600078119T1 2536 view annotation for est 600078119T1. Length = 414

>osa-miR169f MIMAT0001051, length = 21

Score = 38.2 bits (19), Expect = 6e-006

Identities = 19/19 (100%)

Strand = Plus / Minus

```
Query: 355 aggcaagtcatccttggct 373
           ||||||||||||||||||
Sbjct: 20  aggcaagtcatccttggct 2
```

>osa-miR169e MIMAT0001050, length = 21

Score = 36.2 bits (18), Expect = 2e-005

Identities = 18/18 (100%)

Strand = Plus / Minus

```
Query: 356 ggcaagtcatccttggct 373
           ||||||||||||||||||
Sbjct: 19  ggcaagtcatccttggct 2
```

>osa-miR169n MIMAT0001059, length = 21

Score = 30.2 bits (15), Expect = 0.001

Identities = 18/19 (94%)

Strand = Plus / Minus

```
Query: 355 aggcaagtcatccttggct 373
           |||||||||| |||||||
Sbjct: 20  aggcaagtcattcttggct 2
```

Query= 600079708T1 3158 view annotation for est 600079708T1. Length = 517

>osa-miR393 MIMAT0000957, length = 21

Score = 36.2 bits (18), Expect = 3e-005

Identities = 18/18 (100%)

Strand = Plus / Plus

```
Query: 449 tccaaagggatcgcattg 466
           ||||||||||||||||||
Sbjct: 1   tccaaagggatcgcattg 18
```

>osa-miR393b MIMAT0001078, length = 22

Score = 36.2 bits (18), Expect = 3e-005

Identities = 18/18 (100%)

Strand = Plus / Plus

```
Query: 449 tccaaagggatcgcattg 466
           ||||||||||||||||||
Sbjct: 1   tccaaagggatcgcattg 18
```

Query= 600103903T1 16131 view annotation for est 600103903T1. Length = 467

>osa-miR393 MIMAT0000957, length = 21

Score = 36.2 bits (18), Expect = 3e-005

Identities = 18/18 (100%)

Strand = Plus / Plus

```
Query: 442 tccaaagggatcgcattg 459
           ||||||||||||||||||
Sbjct: 1   tccaaagggatcgcattg 18
```

>osa-miR393b MIMAT0001078, length = 22

Score = 36.2 bits (18), Expect = 3e-005

Identities = 18/18 (100%)

Strand = Plus / Plus

```
Query: 442 tccaaagggatcgcattg 459
           ||||||||||||||||||
Sbjct: 1   tccaaagggatcgcattg 18
```

Query= 600132974T1 20252 view annotation for est 600132974T1. Length = 503

>osa-miR156k MIMAT0001020, length = 21

Score = 36.2 bits (18), Expect = 3e-005

Identities = 18/18 (100%)

Strand = Plus / Plus

```
Query: 210 tgacagaagagagagagc 227
           ||||||||||||||||||
Sbjct: 1   tgacagaagagagagagc 18
```

Query= 600133648T1 20548 view annotation for est 600133648T1. Length = 353

>osa-miR156k MIMAT0001020, length = 21

Score = 36.2 bits (18), Expect = 2e-005

Identities = 18/18 (100%)

Strand = Plus / Plus

```
Query: 264 tgacagaagagagagagc 281
           ||||||||||||||||||
Sbjct: 1   tgacagaagagagagagc 18
```

Query= 600148339T1 25074 view annotation for est 600148339T1. Length = 591

>osa-miR156k MIMAT0001020, length = 21

Score = 36.2 bits (18), Expect = 3e-005

Identities = 18/18 (100%)

Strand = Plus / Plus

```
Query: 216 tgacagaagagagagagc 233
           ||||||||||||||||||
Sbjct: 1   tgacagaagagagagagc 18
```

Query= 600174542T1 32465 view annotation for est 600174542T1. Length = 495

>osa-miR156k MIMAT0001020, length = 21

Score = 36.2 bits (18), Expect = 3e-005

Identities = 18/18 (100%)

Strand = Plus / Plus

```
Query: 271 tgacagaagagagagagc 288
           ||||||||||||||||||
Sbjct: 1   tgacagaagagagagagc 18
```

Query= 600081013T1 3968 view annotation for est 600081013T1. Length = 397

>osa-miR169f MIMAT0001051, length = 21

Score = 40.1 bits (20), Expect = 1e-006

Identities = 20/20 (100%)

Strand = Plus / Minus

```
Query: 342 aggcaagtcatccttggcta 361
           ||||||||||||||||||||
Sbjct: 20  aggcaagtcatccttggcta 1
```

>osa-miR169e MIMAT0001050, length = 21

Score = 38.2 bits (19), Expect = 6e-006

Identities = 19/19 (100%)

Strand = Plus / Minus

```
Query: 343 ggcaagtcatccttggcta 361
           ||||||||||||||||||
Sbjct: 19  ggcaagtcatccttggcta 1
```

>osa-miR169n MIMAT0001059, length = 21

Score = 32.2 bits (16), Expect = 3e-004

Identities = 19/20 (95%)

Strand = Plus / Minus

```
Query: 342 aggcaagtcatccttggcta 361
           |||||||||| ||||||||
Sbjct: 20  aggcaagtcattcttggcta 1
```

>osa-miR169d MIMAT0001049, length = 21

Score = 30.2 bits (15), Expect = 0.001

Identities = 18/19 (94%)

Strand = Plus / Minus

```
Query: 343 ggcaagtcatccttggcta 361
           ||||| ||||||||||||
Sbjct: 19  ggcaattcatccttggcta 1
```

Query= 600092447T1 9453 view annotation for est 600092447T1. Length = 524

>osa-miR169f MIMAT0001051, length = 21

Score = 40.1 bits (20), Expect = 2e-006

Identities = 20/20 (100%)

Strand = Plus / Minus

```
Query: 345 aggcaagtcatccttggcta 364
           ||||||||||||||||||||
Sbjct: 20  aggcaagtcatccttggcta 1
```

>osa-miR169e MIMAT0001050, length = 21

Score = 38.2 bits (19), Expect = 7e-006

Identities = 19/19 (100%)

Strand = Plus / Minus

```
Query: 346 ggcaagtcatccttggcta 364
           ||||||||||||||||||
Sbjct: 19  ggcaagtcatccttggcta 1
```

>osa-miR169q MIMAT0001062, length = 21

Score = 32.2 bits (16), Expect = 5e-004

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 267 tagccaangagactgccca 285
           ||||||| ||||||||||
Sbjct: 1   tagccaaggagactgccca 19
```

>osa-miR169n MIMAT0001059, length = 21

Score = 32.2 bits (16), Expect = 5e-004

Identities = 19/20 (95%)

Strand = Plus / Minus

```
Query: 345 aggcaagtcatccttggcta 364
           |||||||||| ||||||||
Sbjct: 20  aggcaagtcattcttggcta 1
```

>osa-miR169d MIMAT0001049, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Minus

```
Query: 346 ggcaagtcatccttggcta 364
           ||||| ||||||||||||
Sbjct: 19  ggcaattcatccttggcta 1
```

Query= 600092884T1 9738 view annotation for est 600092884T1. Length = 555

>osa-miR169f MIMAT0001051, length = 21

Score = 40.1 bits (20), Expect = 2e-006

Identities = 20/20 (100%)

Strand = Plus / Minus

```
Query: 347 aggcaagtcatccttggcta 366
           ||||||||||||||||||||
Sbjct: 20  aggcaagtcatccttggcta 1
```

>osa-miR169q MIMAT0001062, length = 21

Score = 38.2 bits (19), Expect = 8e-006

Identities = 19/19 (100%)

Strand = Plus / Plus

```
Query: 269 tagccaaggagactgccca 287
           ||||||||||||||||||
Sbjct: 1   tagccaaggagactgccca 19
```

>osa-miR169e MIMAT0001050, length = 21

Score = 38.2 bits (19), Expect = 8e-006

Identities = 19/19 (100%)

Strand = Plus / Minus

```
Query: 348 ggcaagtcatccttggcta 366
           ||||||||||||||||||
Sbjct: 19  ggcaagtcatccttggcta 1
```

>osa-miR169n MIMAT0001059, length = 21

Score = 32.2 bits (16), Expect = 5e-004

Identities = 19/20 (95%)

Strand = Plus / Minus

```
Query: 347 aggcaagtcatccttggcta 366
           |||||||||| ||||||||
Sbjct: 20  aggcaagtcattcttggcta 1
```

>osa-miR169d MIMAT0001049, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Minus

```
Query: 348 ggcaagtcatccttggcta 366
           ||||| ||||||||||||
Sbjct: 19  ggcaattcatccttggcta 1
```

Query= 600129804T1 19373 view annotation for est 600129804T1. Length = 241

>osa-miR156k MIMAT0001020, length = 21

Score = 38.2 bits (19), Expect = 3e-006

Identities = 19/19 (100%)

Strand = Plus / Plus

```
Query: 105 tgacagaagagagagagca 123
           ||||||||||||||||||
Sbjct: 1   tgacagaagagagagagca 19
```

>bna-miR156a MIMAT0004445, length = 21

Score = 30.2 bits (15), Expect = 8e-004

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 105 tgacagaagagagagagca 123
           |||||||||||||| |||||
Sbjct: 1   tgacagaagagagtgagca 19
```

>osa-miR156l MIMAT0001021, length = 21

Score = 30.2 bits (15), Expect = 8e-004

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 106 gacagaagagagagagcat 124
           |||||||||||| ||||||
Sbjct: 2   gacagaagagagtgagcat 20
```

Query= 600133517T1 20491 view annotation for est 600133517T1. Length = 466

>osa-miR156k MIMAT0001020, length = 21

Score = 38.2 bits (19), Expect = 7e-006

Identities = 19/19 (100%)

Strand = Plus / Plus

```
Query: 195 tgacagaagagagagagca 213
           ||||||||||||||||||
Sbjct: 1   tgacagaagagagagagca 19
```

>bna-miR156a MIMAT0004445, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 195 tgacagaagagagagagca 213
           ||||||||||||| |||||
Sbjct: 1   tgacagaagagagtgagca 19
```

>osa-miR156l MIMAT0001021, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 196 gacagaagagagagagcat 214
           ||||||||||||| ||||||
Sbjct: 2   gacagaagagagtgagcat 20
```

Query= 600156376T1 27348 view annotation for est 600156376T1. Length = 496

>osa-miR156k MIMAT0001020, length = 21

Score = 38.2 bits (19), Expect = 7e-006

Identities = 19/19 (100%)

Strand = Plus / Plus

```
Query: 135 tgacagaagagagagagca 153
           ||||||||||||||||||
Sbjct: 1   tgacagaagagagagagca 19
```

>bna-miR156a MIMAT0004445, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 135 tgacagaagagagagagca 153
           ||||||||||||| |||||
Sbjct: 1   tgacagaagagagtgagca 19
```

>osa-miR156l MIMAT0001021, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 136 gacagaagagagagagcat 154
           ||||||||||| ||||||
Sbjct: 2   gacagaagagagtgagcat 20
```

Query= 600160669T1 28781 view annotation for est 600160669T1. Length = 474

>osa-miR156k MIMAT0001020, length = 21

Score = 38.2 bits (19), Expect = 7e-006

Identities = 19/19 (100%)

Strand = Plus / Plus

```
Query: 146 tgacagaagagagagagca 164
           ||||||||||||||||||
Sbjct: 1   tgacagaagagagagagca 19
```

>bna-miR156a MIMAT0004445, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 146 tgacagaagagagagagca 164
           ||||||||||||| |||||
Sbjct: 1   tgacagaagagagtgagca 19
```

>osa-miR156l MIMAT0001021, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 147 gacagaagagagagagcat 165
           ||||||||||||| ||||||
Sbjct: 2   gacagaagagagtgagcat 20
```

Query= 600081801T1 4333 view annotation for est 600081801T1. Length = 601

>osa-miR156k MIMAT0001020, length = 21

Score = 40.1 bits (20), Expect = 2e-006

Identities = 20/20 (100%)

Strand = Plus / Plus

```
Query: 518 tgacagaagagagagagcac 537
           ||||||||||||||||||||
Sbjct: 1   tgacagaagagagagagcac 20
```

>bna-miR156a MIMAT0004445, length = 21

Score = 32.2 bits (16), Expect = 5e-004

Identities = 19/20 (95%)

Strand = Plus / Plus

```
Query: 518 tgacagaagagagagagcac 537
           ||||||||||||| ||||||
Sbjct: 1   tgacagaagagagtgagcac 20
```

Query= 600145360T1 24058 view annotation for est 600145360T1. Length = 594

>osa-miR156k MIMAT0001020, length = 21

Score = 40.1 bits (20), Expect = 2e-006

Identities = 20/20 (100%)

Strand = Plus / Plus

```
Query: 550 tgacagaagagagagagcac 569
           ||||||||||||||||||||
Sbjct: 1   tgacagaagagagagagcac 20
```

>bna-miR156a MIMAT0004445, length = 21

Score = 32.2 bits (16), Expect = 5e-004

Identities = 19/20 (95%)

Strand = Plus / Plus

```
Query: 550 tgacagaagagagagagcac 569
           |||||||||||||| ||||||
Sbjct: 1   tgacagaagagagtgagcac 20
```

Query= 600104506T1 16427 view annotation for est 600104506T1. Length = 523

>osa-miR156k MIMAT0001020, length = 21

Score = 42.1 bits (21), Expect = 5e-007

Identities = 21/21 (100%)

Strand = Plus / Plus

```
Query: 355 tgacagaagagagagagcaca 375
           |||||||||||||||||||||
Sbjct: 1   tgacagaagagagagagcaca 21
```

>bna-miR156a MIMAT0004445, length = 21

Score = 34.2 bits (17), Expect = 1e-004

Identities = 20/21 (95%)

Strand = Plus / Plus

```
Query: 355 tgacagaagagagagagcaca 375
           ||||||||||||| |||||||
Sbjct: 1   tgacagaagagagtgagcaca 21
```

Query= 600145442T1 24105 view annotation for est 600145442T1. Length = 616

>osa-miR156k MIMAT0001020, length = 21

Score = 42.1 bits (21), Expect = 6e-007

Identities = 21/21 (100%)

Strand = Plus / Plus

```
Query: 575 tgacagaagagagagagcaca 595
           |||||||||||||||||||||
Sbjct: 1   tgacagaagagagagagcaca 21
```

>bna-miR156a MIMAT0004445, length = 21

Score = 34.2 bits (17), Expect = 1e-004

Identities = 20/21 (95%)

Strand = Plus / Plus

```
Query: 575 tgacagaagagagagagcaca 595
           ||||||||||||| |||||||
Sbjct: 1   tgacagaagagagtgagcaca 21
```

Query= 600079015T1 2976 view annotation for est 600079015T1. Length = 581

>osa-miR172c MIMAT0001071, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 253 gaatcctgatgatgctgca 271
           ||||| ||||||||||||
Sbjct: 2   gaatcttgatgatgctgca 20
```

Query= 600080802T1 3881 view annotation for est 600080802T1. Length = 615

>osa-miR169n MIMAT0001059, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 177 agccaagaatgaattgcct 195
           ||||||||||||| ||||||
Sbjct: 2   agccaagaatgacttgcct 20
```

Query= 600080875T1 3945 view annotation for est 600080875T1. Length = 583

>osa-miR169n MIMAT0001059, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 208 agccaagaatgatttgcct 226
           ||||||||||||| ||||||
Sbjct: 2   agccaagaatgacttgcct 20
```

Query= 600084823T1 4751 view annotation for est 600084823T1. Length = 497

>osa-miR172c MIMAT0001071, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 423 gaatcctgatgatgctgca 441
           ||||| ||||||||||||
Sbjct: 2   gaatcttgatgatgctgca 20
```

Query= 600085062T1 4903 view annotation for est 600085062T1. Length = 600

>osa-miR169n MIMAT0001059, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 193 agccaagaatgaattgcct 211
           ||||||||||||| ||||||
Sbjct: 2   agccaagaatgacttgcct 20
```

Query= 600087026T1 6095 view annotation for est 600087026T1. Length = 545

>osa-miR169n MIMAT0001059, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 212 agccaagaatgaattgcct 230
           ||||||||||||| ||||||
Sbjct: 2   agccaagaatgacttgcct 20
```

Query= 600087712T1 6490 view annotation for est 600087712T1. Length = 572

>osa-miR169n MIMAT0001059, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 198 agccaagaatgatttgcct 216
           ||||||||||||| ||||||
Sbjct: 2   agccaagaatgacttgcct 20
```

Query= 600087966T1 6639 view annotation for est 600087966T1. Length = 421

>osa-miR172c MIMAT0001071, length = 21

Score = 30.2 bits (15), Expect = 0.001

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 317 gaatcctgatgatgctgca 335
           ||||| ||||||||||||
Sbjct: 2   gaatcttgatgatgctgca 20
```

Query= 600093195T1 9955 view annotation for est 600093195T1. Length = 490

>osa-miR172c MIMAT0001071, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 440 gaatcctgatgatgctgca 458
           ||||| ||||||||||||
Sbjct: 2   gaatcttgatgatgctgca 20
```

Query= 600100194T1 13842 view annotation for est 600100194T1. Length = 581

>osa-miR172c MIMAT0001071, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 424 gaatcctgatgatgctgca 442
           ||||| ||||||||||||
Sbjct: 2   gaatcttgatgatgctgca 20
```

Query= 600103944T1 16162 view annotation for est 600103944T1. Length = 567

>osa-miR172c MIMAT0001071, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 353 gaatcctgatgatgctgca 371
           ||||| ||||||||||||
Sbjct: 2   gaatcttgatgatgctgca 20
```

Query= 600106033T1 17205 view annotation for est 600106033T1. Length = 565

>osa-miR172c MIMAT0001071, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 358 gaatcctgatgatgctgca 376
           ||||| ||||||||||||
Sbjct: 2   gaatcttgatgatgctgca 20
```

Query= 600124872T1 18512 view annotation for est 600124872T1. Length = 456

>osa-miR172c MIMAT0001071, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 187 gaatcctgatgatgctgca 205
           ||||| ||||||||||||
Sbjct: 2   gaatcttgatgatgctgca 20
```

Query= 600142076T1 23069 view annotation for est 600142076T1. Length = 480

>osa-miR172c MIMAT0001071, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 33 gaatcctgatgatgctgca 51
           ||||| ||||||||||||
Sbjct: 2  gaatcttgatgatgctgca 20
```

Query= 600153776T1 26699 view annotation for est 600153776T1. Length = 440

>osa-miR169n MIMAT0001059, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 224 agccaagaatgaattgcct 242
           ||||||||||| ||||||
Sbjct: 2   agccaagaatgacttgcct 20
```

Query= 600160442T1 28683 view annotation for est 600160442T1. Length = 397

>osa-miR172c MIMAT0001071, length = 21

Score = 30.2 bits (15), Expect = 0.001

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 153 gaatcctgatgatgctgca 171
           ||||| ||||||||||||
Sbjct: 2   gaatcttgatgatgctgca 20
```

Query= 600173221T1 32142 view annotation for est 600173221T1. Length = 480

>osa-miR164d MIMAT0001034, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 233 gagaagcagggcacctgct 251
           |||||||||||||| ||||
Sbjct: 3   gagaagcagggcacgtgct 21
```

Query= 600177165T1 32938 view annotation for est 600177165T1. Length = 506

>osa-miR169n MIMAT0001059, length = 21

Score = 30.2 bits (15), Expect = 0.002

Identities = 18/19 (94%)

Strand = Plus / Plus

```
Query: 228 agccaagaatgaattgcct 246
           |||||||||||| ||||||
Sbjct: 2   agccaagaatgacttgcct 20
```

Query= 600136111T1 21376 view annotation for est 600136111T1. Length = 461

>osa-miR156k MIMAT0001020, length = 21

Score = 36.2 bits (18), Expect = 3e-005

Identities = 20/21 (95%)

Strand = Plus / Plus

```
Query: 400 tgacagangagagagagcaca 420
           ||||||| ||||||||||||
Sbjct: 1   tgacagaagagagagagcaca 21
```

Appendix 2: Solution preparation for RNA extraction using CTAB-NETS.

i. CTAB extraction buffer of 500 ml

- 2 g of 2% CTAB
- 2 g of 2% PVP40
- 5 ml of 0.5 M EDTA

Treat with 0.1 % DEPC, shake vigorously and left overnight before autoclave at 121 ºC for 45 min.

- 40 ml of 5 M NaCl
- 10 ml of 1 M Tris-Cl, pH 8.0
- DEPC-water adjust the final volume to 500 ml

ii. 0.5 M of EDTA (100 ml)

- About 18.61 g of disodium ethylene diamine tetraacetate powder was added to 80 ml of water. The solution was stirred vigorously with a magnetic stirrer. Approximately 2 g of NaOH pellet was added into the solution and was adjusted to pH 8.0. The solution was then sterilized by autoclave.

iii. 5 M of NaCl (100 ml)

- About 29.22 g of NaCl powder was dissolved in 80 ml of water. The volume is adjusted to 100 ml and sterilized by autoclave.

iv. 1 M of Tris, pH 8.0 (100 ml)

- Before preparing Tris solution, spatula and container were sprayed with RNase AWAY^{TM}, distilled water and bottle were treated with DEPC. About 12.11 g of Tris base powder was dissolved in 80 ml of water. The pH was adjusted to the 8.0 by adding 4.2 ml of concentrated HCl. The solution was allowed to cool in

room temperature before making final adjustment to the pH. It was sterilized by autoclave.

v. chloroform-isoamylalcohol, CI (250 ml)

- With the ratio 24:1 of CI, 240 ml of chloroform were added to 10 ml of isoamylalcohol.

vi. 3 M of sodium acetate, NaOAC with pH 5.2 (100 ml)

- About 40.81 g of sodium acetate powder was dissolved into 80 ml of water. The pH was adjusted to 5.2 with glacial acid. The solution was then sterilized by autoclave.

vii. 10 X Tris-Borate EDTA, TBE buffer (500 ml)

- 54 g of Tris base
- 27.5 g of boric acid
- 2 ml of 0.05 M EDTA, pH 8.0
- Distilled water adjust to final volume of 500 ml
- Sterilized with autoclave at 121 ºC for 45 min

Appendix 3: Sequencing results of the RT-PCR.

5'-AGAGAACTTCCTCCCCCAAACCCTATAATTTGCAGGCAGATGAGAGAGAGGAGGAAGGGATG
GGATCTCGAGAGGAGGAGATTTCATGCGAGGGGGGTTGACAGAAGAGAGTGAGCACACAGCG
GGGGAAAATGGCATGCGACGGTGCTGTTTCCGTTGCGTGCTCACTTCTCTTCCTGTCAGCCTTT
TGTTGTTCCTTCCTTTTTTGAGCCGTGCCTGGAGATGCGATCCCGTTCCCCTTCCGTGCTTATCC
GTCGCCGATCTCCTCTCGGGAGCTTCGTCTCGATCCGTGCTTCGCTTCGCTTTGCTGTGCTGTCG
GTCTTATCTTATCTGCCTGTCCGCCCGTCGGATTTCTCCGGGCGAAGAGAGACCGATGCTTGAT
GTCGTTAAATCTCTTGGCGTTGCGTGCCTAAATGGCTTCTTTTCTTAATTATTTGCCTTTGTTCA
AGAAATTGTTGGTCAGATTTTTCGGATAATTAAGGAAATTGTGCAGGAGGAAGC-3'

Figure A.1: Sequencing results for RT-PCR amplified cDNAs from *Musa acuminata* var. *Berangan* with miR156a.

5'-ATCTGCTTCCGTCTGTGCTTGATGCAGTTGAACTCCTCCTTACAGCATAGTGTTCTGCAAGGGT
GAACACATGCGATTGCATGGCTGCATGCATTAGCGCAGTGCAATACATCAGGAAGATGAAACC
CTTGTCTTTGTATCAGTTGCGAGCTCACGAGCTGCTACCGGAGAAGACTCTGAGGTCTAGCGTG
AGGGATGAGAGCCACTCAGACTAGCCAAGGAGACTGCCTGTGTAGCCAACCCACTGAGAGTTT
TTAGGGCACTGAAACCTGAACCGAGAGGGCATATGGCACAGGCAAGTCATCCTTGGCTTGCCA
GAGAGACTCTTCGTCTCCATGCAAGCCCTTTCT-3'

Figure A.2: Sequencing results for RT-PCR amplified cDNAs from *Musa acuminata* var. *Berangan* with miR169f/n.

5'-ATCCTTCCAAAGTCCTGCAAATGGGGAAGAGATAGTGGTCCAGCATGGGGGACAAGAGTCCCT
CTGATTAGCCAAGGAGACTGCCCAGTGACCCACATGGAGCTTCTGAAACCCTCATGCAGCCAC
CAGGACCTCCATGAGGACACAGGCAAGTCATCCTTGGCTACCAAAGACAGACTCTCATCTCCA
TGCAAGACCTCTCTTCCACACCTCTCCTCCTTGGCTGCTTGCTCAGACCTGCACTGCTTGCCCAT
GGCAGCCATTTATAGCGTGGCCATGGAT-3'

Figure A.3: Sequencing results for RT-PCR amplified cDNAs from *Musa acuminata* var. *Berangan* with miR172c.

REFERENCES

References:

1. Adai, A., Johnson, C., Mlotshwa, S., Archer-Evans, S., Manocha, V., Vance, V. and Sundaresan, V. (2005). Computational prediction of miRNAs in *Arabidopsis thaliana*. *Genome Research*, 15 (1): 78-91.

2. Aert, R., Sagi, L., Volckaert, G. (2004). Gene content and density in banana (*Musa acuminata*) as revealed by genomic sequencing of BAC clones. *Theoretical and Applied Genetics*, 109: 129-139.

3. Agot, A. (1968). Etat actuel de nos connaisances sur la biochimic, la physiologie et la valeur nutritive de la banane (Biochemistry, physiology and nutritive value of the banana). *Bulletin de lu Sociétés Scienifique d'Hygiene Alimentaire et d'Alimentation Rationnelle de I'Homme*, 56: 27-41.

4. Allen, E., Xie, Z. X., Gustafson, A. M. and Carrington, J. C. (2005). mRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, 121: 207-221.

5. Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X.M., Dreyfuss, G., Eddy, S.R., Jones, S.G., Marshall, M., Matzke, M., Ruvkun, G. and Tuschl, T. (2003). A uniform system for MicroRNA annotation. *RNA*, 9 (1-2): 277-279.

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

7. Arora, A. and Simpson, D. A. C. (2008). Individual mRNA expression profiles reveal the effects of specific microRNAs. *Genome Biology*, 9 (5): R82.

8. Arteaga-Vázquez M., Caballero-Pérez J. and Vielle-Calzada J. (2006). A family of microRNAs present in plants and animals. *The Plant Cell*, 18: 3355-3369.

9. Australian Government, Department of Health and Ageing Office of the Gene Technology Regulator. (2008). The biology of *Musa* L. (banana). Accessed at http://www.health.gov.au/internet/ogtr/publishing.nsf/Content/banana-3/$FILE/biologybanana.pdf on May 2009.

10. Axtell, M. J. and Bartel, D. P. (2005). Antiquity of MicroRNAs and their Targets in Land Plants. *The Plant Cell*, 17: 1658-1673.

11. Axtell, M. J., Snyder, J. A. and Bartel, D. P. (2007). Common Functions for Diverse Small RNAs of Land Plants. *The Plant Cell*, 19: 1750-1769.

12. Barker, W. G., Steward, F. C. (1962). The growing regions of the vegetative shoot. *Annals of Botany*, 26: 389-411.

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

14. Berezikov, E., Cuppen, E., and Plasterk, R. A. (2006). Approaches to microRNA discovery. *Nature Genetics Supplement*, 38: 2-6.

15. Billoud, B., Paepe, R. D., Baulcombe, D. and Boccara, M. (2005). Identification of new small non-coding RNAs from tobacco and *Arabidopsis. Biochimie*, 87: 905-910.

16. Blanco, E., Guigo, R. and Messeguer, X. (2007). Multiple non-collinear TF-map alignments of promoter regions. *BMC Bioinformatics*, 8: 138.

17. Boning, C. R. (2006). Florida's best fruiting plants. Pineapple Press Inc., Florida.

18. Bonneti, E., Wuytsl, J., Rouze, P. and Van de Peerl, Y. (2004). Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics*, 20 (17): 2911-2917.

19. Chen, X. (2008). MicroRNA Metabolism in Plants. *Current Topics in Microbiology and Immunology*, 320: 117-136.

20. Chen, X. M., Zhao, Li., Kim, Y. J. (2006). miR172 modulates the output of the AGAMOUS / APETALA2 antagonistic pair in floral patterning. *Developmental Biology*, 295: 324-326.

21. Cheung, F., Town, C.D. (2007). A BAC end view of the *Musa acuminata* genome. *BMC Plant Biology (online)*, 7: 29.

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

23. Daniells, J., Jenny, C., Karamura, D., Tomekpe, K. (2001). Musalogue, a catalogue of *Musa* germplasm: Diversity in the Genus *Musa*. International Network for the Improvement of Banana and Plantain, Montpellier, France. Accessed at http://bananas.bioversityinternational.org/files/files/pdf/publications/musalogue2.pdf on May 2009.

24. De Langhe, E. (1995). Banana and Plantain: the Earliest Fruit Crops? International Network for the Improvement of Banana and Plantain, Montpellier, France.

25. De Peer, Y. V. (2005). 'Horizontal plant biology on the rise. *Genome Biology*, 6 (1): 302.

26. Dezulian, T., Palatnik, J. F., Huson, D. and Weigel, D. (2005). Conservation and divergence of microRNA families in plants. *Genome Biology*, 6: P13.

27. Dezulian, T., Remmert, M., Palatnik, J. F., Weigel, D. and Huson, D. H. (2006). Identification of plant microRNA homologs. *Bioinformatics*, 22 (3): 359-360.

28. Dugas, D. V. and Bartel, B. (2004). MicroRNA regulation of gene expression in plants. *Plant Biol.*, 7: 512-520.

29. Dunoyer, P., Lecellier, C-H., Parizotto, E.A., Himber, C. and Voinnet, O. (2004) Probing the MicroRNA and Small Interfering Pathways with Virus-Encoded Suppressors of RNA Silencing. *The Plant Cell*, 16: 1235-1250.

30. Glazinska, P., Zienkiewicz, A., Wojciechowski, W. and Kopcewicz J. (2009). The putative miR172 target gene in APETALA2-like is involved in the photoperiodic flower induction of *Ipomoea nil*. *Journal of Plant Physiology*, 166: 1801-1813.

31. Gollapudi, R., Revanna, K.V., Hemmerich, C., Schaack, S. and Dong, Q. (2008). BOV – A Web-based BLAST Output Visualization Tool. *BMC Genomics*, 9: 414.

32. Grad Y., Aach J., Hayes G. D., Reinhart B. J., Church G. M., Ruvkun G. and Kim J. (2003). Computational and experimental identification of *C.elegans* microRNAs. *Molecular Cell*, 11(5): 1253-1263.

33. Griffiths-Jones S. (2004). The microRNA Registry. *Nucleic Acids Research*, 32: D109- D111.

34. Griffiths-Jones S., Grocock R.J. van Dongen S., Bateman A. and Enright J. (2005). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, 34: 140-144.

35. Heslop-Harrison, J.S., Schwarzacher, T. (2007). Domestication, genomics and the future of the banana. *Annals of Botany*, 100:1073-1084.

36. Horigome, T., Sakaguchi, E., Kishimoto, C. (1992). Hypocholesterolaemic effect of banana (*Musa sapientum* L. var. Cavendishii) pulp in the rat fed on a cholesterol-containing diet. *British Journal of Nutrition*, 68: 231-244.

37. Horry, J.-P., Arnaud, O.E., Crouch, J.H., Ferris, R.S.B., Jones, D.R., Mateo, N., Picq, C., Vuylsteke, D. (1997). Banana and Plantain. Chapter 6. In: D Fuccillo, L Sears, P Stapleton, eds. Biodiversity in Trust. Conservation and Use of Plant Genetic Resources in CGIAR Centres. Cambridge University Press, Cambridge, pp. 67-81.

38. Hyam, R., Pankhurst, R. (1995). Plants and Their Names: A Concise Dictionary. Oxford University Press, Oxford.

39. INIBAP (2000). Bananas. International Network for the Improvement of Banana and Plantain, International Plant Genetic Resources Institute. Accessed at http://bananas.bioversityinternational.org/files/files/pdf/publications/an00_en.pdf on May 2009.

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

41. Kim, V. N. (2004) MicroRNA precursors in motion: exportin-5 mediates their nuclear export. *Trends in Cell Biology*, 14: 156-159.

42. Kiple, K. F. and Ornelas, K. C. (2000), The Cambridge World History of Food Vol. I, Cambridge University Press, New York, pp. 180.

43. Lee R. C. and Ambros V., 2001, An extensive class of small RNAs in Caenorhabditis elegans, *Science*, 294(5543): 862–864.

44. Li, S. C. T. (2008), Vegetables and Fruits – Nutritional and Therapeutic Values, CRC Press, United States of America, pp. 124.

45. Li Y., Li W. and Jin Y. X. (2005). Computational identification of novel family members of microRNA genes in *Arabidopsis thaliana* and *Oryza sativa*. *Acta Biochim Biophys Sin*, 37: 75-87.

46. Lim L. P., Lau N. C. and Weinstein E. G. (2003) The microRNAs of Caenorhabditis elegans. *Genes and Development*, 17(8): 991–1008.

47. Lin, S-L., Miller, J. D. and Ying, S-Y. (2006). Intronic MicroRNA (miRNA). *Journal of Biomedicine and Biotechnology*, pp. 1-13.

48. Lingel, A. and Sattler, M. (2005). Novel modes of protein-RNA recognition in the RNAi pathway. *Current Opinion in Structural Biology*, 15 (1), 107-115.

49. Low, E-T, Alias, H., Boon, S-H., Shariff, E.M., Tan, C-YA., Ooi, C.L.L., Cheah, S-C., Abdul, R. R., Wan, K-L. and Singh, R. (2008). Oil palm (*Elaeis guineensis* Jacq.) tissue culture ESTs: Identifying genes associated with callogenesis and embryogenesis. *BMC Plant Biology*, 8: 62.

50. Mallory, A. C. and Vaucheret, H. (2006). Functions of microRNAs and related small RNAs in plants. *Nat Genetics*, 38: S31-S37.

51. Matthews D.H., Sabina J., Zuker M. and Turner D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*, 288: 911-940.

52. Matukumalli L. K., Grefenstette J. J., Sonstegard T. S. and van Tassell C. P. (2004). EST-PAGE - managing and analyzing EST data. *Bioinformatics*, 20: 286-288.

53. Mendes, N. D., Freitas, A. T., Sagot, M. F. (2009). Current tools for the identification of miRNA genes and their targets. *Nucleic Acids Research*, 37 (8), 2419-2433.

54. Miller, K. and Storts, D. R. (1995) Promega Notes 53, pp. 2–5.

55. Nasaruddin, N., Harikrishnam K., Othman, R.Y., Lim, S. H., and Harikrishna, J. A. (2007). Computational prediction of microRNAs from oil palm (*Elaeis guineensis*) expressed sequence tags. *Asia Pacific Journal of Molecular Biology and Biotechnology*, 15 (3), 107-113.

56. Nelson, S. C., Ploetz, R. C. and Kepler, A. K. (2006), Musa species (Banana and Plantain). *Species Profiles for Pacific Island Agroforestry*, pp. 22.

57. Pillay, M., Tripathi, L. (2007). Banana. Chapter 15. In: C Kole, ed. *Genome Mapping and Molecular Breeding in Plants, Volume 4 Fruits and Nuts.* Springer-Verlag, Berlin. pp 281-301.

58. Pillay, M.A., Tenkouano, A., Ude, G., Ortiz, R. (2004). Molecular characterization of genomes in Musa and its applications. Chapter 23. In:SM Jain, R Swennen, eds. *Banana Improvement: Cellular, Molecular Biology and Induced Mutations*. Science Publishers Inc., Enfield (NH), USA.

59. Qiu C. X., Xie F. L., Zhu Y. Y., Guo K., Huang S. Q., Nie L. and Yang Z. M. (2007). Computational identification of microRNAs and their targets in *Gossypium hirsutum* expressed sequence tags. *Gene*, 395: 49-61.

60. Qu, J., Ye, J., and Fang, R. (2007). Artificial MicroRNA-Mediated Virus Resistance in Plants. *Journal of Virology*, 81(12): 6690-6699.

61. Reinhart B. J., Weinstein E. G., Rhoades M. W., Bartel B. and Bartel D. P. (2002). MicroRNA in plants. *Genes & Development*, 16: 1616-1626.

62. Resourses Council (1982). Standard Tables of Food Composition in Japan, 4th revised ed. Tokyo: Science and Technology Agency, Japan.

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

64. Ronemus, M., Vaughn, M. W. and Martienssen, R. A. (2006) MicroRNA-Targeted and Small Interfering RNA–Mediated mRNA Degradation Is Regulated by Argonaute, Dicer, and RNA-Dependent RNA Polymerase in *Arabidopsis*. *Plant Cell*, 18(7): 1559-1574.

65. Ruby, G. J., Stark, A., Johnston, W. K., Kellis, M., Bartel, D. P. and Lai, E. C. (2007). Evolution, Biogenesis, Expression and Target Predictions of Substantially Expanded Set of Drosophila microRNAs. *Genome Research*, 17(12): 1850-1864.

66. Rudd, S. (2005). openSputnik—a database to ESTablish comparative plant genomics using unsaturated sequence collections. *Nucleic Acids Research*, 33: D622-D627.

67. S. Sharrock. (1997). INIBAP annual report 1996. Uses of Musa, pp. 42-44.

68. Segers, G. C., van Wezel, R., Zhang, X., Hong, Y., and Nuss, D. L. (2006), Hypovirus Papain-like Protease p29 Suppresses RNA Silencing in the Natural Fungal Host and in a Heterologous Plant System. *Eukaryotic Cell*, 5(6): 896-904.

69. Sharaf, A., Sharaf, O. A., Hegazi, S. M. & Sedky, K. (1979). Chemical and biological studies on banana fruit. *Zeitschrift für Ernahrungswissenschaft*, 18: 8-15.

70. Simmonds, N.W. (1962). The Evolution of Bananas. Longmans, Green and Co. Ltd, London.

71. Simmonds, N. W., Shepherd, K. (1955). The taxonomy and origins of the cultivated bananas. *Journal of the Linnean Society of London (Botany)* 55:302-312.

72. Stover R. H. and N. W. Simmonds. (1987). Bananas. Longman. UK.

73. Sunkar, R., Girke, T., Jain, P. K., and Zhu, J-K. (2005). Cloning and Characterization of MicroRNAs from Rice. *Plant Cell*, 17(5): 1397-1411.

74. Sunkar, R. and Zhu, J. K. (2007). Mirco RNAs and short-interfering RNAs in plants. *Journal of Integrative Plant Biology*, 49 (6): 817-826.

75. Tang, G., Reinhart, B. J., Bartel, D. P., and Zamore, P. D. (2003), A biochemical framework for RNA silencing in plants. *Genes & Development*, 17: 49-63.

76. Thomassen, G. O. S., Rosok, O., and Rognes, T. (2006). Computational Prediction of MicroRNAs Encoded in Viral and Other Genomes. *Journal of Biomedicine and Biotechnology*, pp. 1-10.

77. Thompson A.K. (1995). Banana Processing. In *Bananas and Plantains*. (S. Gowen, ed.). Chapman and Hall, UK, pp. 481-492.

78. Valmayor, R. V., Jamaluddin, S. H., Silayoi, B., Kusumo, S., Danh, L. D., Pascua, O. C., Espino, R. R. C. (2000). Banana Cultivar Names and Synonyms in Southeast Asia. International Network for the Improvement of Banana and Plantain – Asia and the Pacific Office, Laguna, Philipines. Accessed at http://bananas.bioversityinternational.org/files/files/pdf/publications/synonyms.pdf on May 2009.

79. Vazquez, F. (2006). *Arabidopsis* endogenous small RNAs: highways and byways. *Trends Plant Science*, 11: 460 -468.

80. Wang, J. F., Zhou, H., Chen, Y. Q., Luo, Q. J. and Qu, L. H. (2004). Identification of 20 microRNA from *Oryza sativa*. *Nucleic Acid Research*, 32 (5): 1688-1695.

81. Weng, J-K., Tanurdzic, M. and Chapple, C. (2005). Functional analysis and comparative genomics of expressed sequence tags from the lycophyte *Selaginella moellendorffii*. *BMC Genomics*, 6: 85.

82. Willmann, M. R. and Poethig, R. S. (2008). Conservation and evolution of miRNA regulatory programs in plant development, *Current Opinion in Plant Biology*, 10 (5): 503-511.

83. Xie, F. L., Huang, S. Q., Guo, K., Xiang, A. L., Zhu, Y. Y., Nie, L., and Yang, Z. M. (2007). Computational identification of novel microRNA and targets in *Brassica napus*. FEBS Letter 581 (7), 1464-1474.

84. Yang, T. W., Xue, L. G., and An, L. Z. (2007). Functional diversity of miRNAs in plants. *Plant Science*, 172: 423-432.

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

86. Zhang, B. H., Pan, X. P., Cox, S. B., Cobb, G. P. and Anderson, T. A. (2005b). Evidence that miRNAs are Different from other RNAs. *Cellular and Molecular Life Sciences*, 63: 246-254.

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

88. Zhang, B. H., Pan, X. P., Cobb, G. P. and Anderson, T. A. (2006b). Plant microRNA: a small regulatory molecule with big impact. *Development Biology*, 289: 3-16.

89. Zhang, B. H., Wang, Q. L., Pan, X. P. (2006c). MicroRNAs and their regulatory roles in animals and plants. *Journal of Cell Physiology*, 210: 279-289.

90. Zhang, B. H., Pan, X. P., Wang, Q. L., Cobb, G. P. and Anderson, T. A., (2006d). Computational identification of microRNAs and their targets. *Computational Biology and Chemistry*, 30: 395-407.

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

92. Zhu, Q-H., Spriggs, A., Matthew, L., Fan, L., Kennedy, G., Gubler, F., and Helliwell, C. (2008). A Diverse Set of microRNAs and microRNA-like Small RNAs in Developing Rice Grains. *Genome Research*, 18 (9): 1456-1465.

93. Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31 (13), 3406-15.