The role of small RNAs in susceptibility and tolerance to cassava mosaic disease

A dissertation presented by

Sarah Jane Rogans

to

The Faculty of Science, University of the Witwatersrand, Johannesburg

In fulfilment of the

requirements for the degree of

Doctor of Philosophy

In the School of Molecular and Cell Biology

2016

Supervisor: Prof M.E.C. Rey

Advisor: Prof Rob Veale

ABSTRACT

Cassava (*Manihot esculenta*, Crantz) is considered to be an important food security crop consumed by over a billion peoples globally, many who subsist on it. Cassava mosaic disease (CMD) is one of the main biotic and economically important constraints to cassava cultivation in sub-Saharan Africa. Geminiviruses are the casual agents of CMD and cause disease to many staple food and cash crops of great economic importance worldwide. There are currently 11 species of Begomoviruses that belong to the *Geminiviridae* family. *South African cassava mosaic virus* (SACMV) is a circular ssDNA bipartite (DNA A and DNA B components) begomovirus belonging to the family *Geminiviridae*, and is one of the causal agents of cassava mosaic disease (CMD) endemic to southern Africa. Various strategies to control CMD are currently being investigated, one of which is *cis*-genics, which involves manipulation of endogenous host genes to combat viral pathogens. In order to achieve this, it is imperative to elucidate molecular mechanisms involved in host-virus interactions. Endogenous small RNAs (sRNAs), including microRNAs (miRNAs), have been found associated with gene regulatory mechanisms in response to virus infection. Amongst the non-coding host sRNAs targeting viruses are small interfering RNAs (siRNAs) associated with posttranscriptional gene silencing (PTGS) and transcriptional gene silencing (TGS), which are involved in the host RNA silencing pathway. The RNA silencing pathway is a highly conserved basal immunity pathway involved in host defence against plant viruses. The aim of this study was to identify siRNAs and miRNAs associated with gene regulatory mechanism in response to SACMV infection and to determine if they a play a role in the susceptible or recovery phenotype observed in SACMV tolerant cassava landrace TME3 or T200, respectively. Furthermore, virus-derived siRNA (vsRNA) populations targeting the DNA A and B components of SACMV were also investigated.

MicroRNAs (miRNAs) are an important class of endogenous non-coding single-stranded small RNAs (21-24 nt in length), which serve as post-transcriptional negative regulators of gene expression in plants. Despite the economic importance of *Manihot esculenta* Crantz (cassava) only 153 putative cassava miRNAs (from multiple germplasm) are available to date in miRBase

(V.21). Therefore, both conserved and novel miRNAs needed to be identified in cassava before we could determine what association they had with SACMV infection. In this part of the study, mature sequences of all known plant miRNAs were used as a query for homologous searches against cassava EST and GSS databases, and additional identification of novel and conserved miRNAs were gleaned from next generation sequencing (NGS) of two cassava landraces (T200 from southern Africa and TME3 from West Africa) at three different growth stages post explant transplantation and acclimatization. EST and GSS derived data revealed 259 and 32 conserved miRNAs in cassava, and one of the miRNA families (miR2118) from previous studies has not been reported in cassava. NGS data collectively displayed expression of 289 conserved miRNAs in leaf tissue, of which 230 had not been reported previously. Of the 289 conserved miRNAs identified in T200 and TME3, 208 were isomiRs. Thirty-nine novel cassava-specific miRNAs of low abundance, belonging to 29 families, were identified. Thirty-eight (98.6%) of the putative new miRNAs identified by NGS have not been previously reported in cassava. Several miRNA targets were identified in T200 and TME3, highlighting differential temporal miRNA expression between the two cassava landraces. This study contributes to the expanding knowledge base of the micronome of this important crop.

MicroRNAs play a crucial role in stress response in plants, including biotic stress caused by viral infection. Viruses however can interfere with and exploit the silencing-based regulatory networks, causing the deregulation of miRNAs. This study aimed to understand the regulation of miRNAs in tolerant (TME3) and susceptible (T200) cassava landraces infected with SACMV. Next-generation sequencing was used for analysing small RNA libraries from infected and mock-inoculated cassava leaf tissue collected at 12, 32 and 67 dpi (days post-inoculation). The total number of differentially expressed miRNAs (normalized against mock-inoculated samples) across all three time points was 204 and 209 miRNAs, in TME3 and T200 infected plants, respectively, but the patterns of log_2fold changes in miRNA families over the course of infection differed between the two landraces. A high number were significantly altered at 32 dpi when T200 and TME3 plants showed severe symptoms. Notably, in T200 69% and 28 (100%)

of miRNA families were upregulated at 12 and 32 dpi, respectively. In contrast, TME3 showed an early pre-symptomatic response at 12 dpi where a high number (87%) of miRNAs showed a significant log_2fold downregulation. Endogenous targets were predicted in the cassava genome for many of the identified miRNA families including transcription factors, disease resistance (R)-genes and transposable elements. Interestingly, some of the miRNA families (miR162, miR168 and miR403) that were significantly affected in both T200 and TME3 upon SACMV infection were shown to target proteins (DCL1, AGO1 and AGO2) that play important roles in the RNA silencing pathway. From results, we suggest that the early (12 dpi) miRNA response to SACMV in TME3 appears to involve PTGS-associated AGO1, DCL2 and a cohort of R genes belonging to the miR395 family which may prime the plant for tolerance and recovery downstream, while in T200, SACMV suppresses AGO1, AGO2 (at 32 and 67 dpi), and DCL2 (32 dpi) mediated RNA silencing, leading to severe persistent disease symptoms. This study provides insights into miRNA-mediated SACMV cassava interactions and may provide novel targets for control strategies aimed at developing CMD-resistance cassava varieties

Endogenous small RNAs (sRNAs) associated with gene regulatory mechanisms respond to virus infection, and virus-derived small interfering RNAs (vsRNAs) have been implicated in recovery or symptom remission in some geminivirus-host interactions. Transcriptional gene silencing (TGS) (24 nt vsRNAs) and post transcriptional gene silencing (PTGS) (21-23 nt vsRNAs) have been associated with geminivirus intergenic (IR) and coding regions, respectively. In this Illumina deep sequencing study, we compared for the first time, the small RNA response to *South African cassava mosaic virus* (SACMV) of cassava landrace TME3 which shows a recovery and tolerant phenotype, and T200, a highly susceptible landrace. Interestingly, different patterns in the percentage of SACMV-induced normalized total endogenous sRNA reads were observed between T200 and TME3. Notably, in T200 there was a significant increase in 21 nt sRNAs during the early pre-symptomatic response (12 dpi) to SACMV compared to mock, while in TME3, the 22 nt size class increased significantly at 32 dpi. While vsRNAs of 21 to 24 nt size classes covered the entire SACMV DNA-

A and DNA-B genome components in T200 and TME3, vsRNA population counts were significantly lower at 32 (symptomatic stage) and 67 dpi in tolerant TME3 compared with T200 (non-recovery). It is suggested that the high accumulation of primary vsRNAs, which correlated with high virus titres and severe symptoms in susceptible T200, may be due to failure to target SACMV-derived mRNA. In contrast, in TME3 low vsRNA counts may represent efficient PTGS of viral mRNA, leading to a depletion/sequestration of vsRNA populations, supporting a role for PTGS in tolerance/recovery in TME3. Notably, in TME3 at recovery (67 dpi) the percentage (expressed as a percentage of total vsRNA counts) of redundant and non-redundant (unique) 24 nt vsRNAs increased significantly. Since methylation of the SACMV genome was not detected by bisulfite sequencing, and vsRNA counts targeting the IR (where the promoters reside) were very low in both the tolerant or susceptible landraces, we conclude that 24 nt vsRNA-mediated RNA directed genome methylation does not play a central role in disease phenotype in these landraces, notwithstanding recognition for a possible role in histone modification in TME3.

This work represents an important step toward understanding variable roles of sRNAs in different cassava genotype-geminivirus interactions. Also, by comparing the differences between a tolerant and susceptible host the aim is to achieve better understanding of the effect of pathogens on host sRNAome, an area that is deserving of me attention in plant systems. The expectation is that these findings presented in the PhD will contribute to the long-term goals of devising new methods of disease control against SACMV and understanding the complex interconnected mechanisms involved in virus-host interactome.

DECLARATION

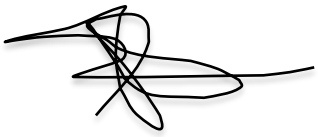
13.2 Formal Declaration (G9.7, G10.3)

A candidate must include in his/her Thesis/Dissertation/Research Report, a formal declaration stating the following:

I, **NAME** (**PERSON NUMBER**), am a student registered for the degree of _____ in the academic year _____.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Thesis/Dissertation/Research Report **HAS/HAS NOT** been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature:  07 day of September **YEAR** 2016

Sarah Rogans

ACKNOWLEDGMENTS

I would firstly like to extend my gratitude to my supervisor Prof Rey for her guidance, patience and support throughout this project.

To all my friends in the School of Molecular and Cell Biology and members of the Plant Biotechnology Laboratory (past and present), thank you for all of your constant support and words and encouragement. A massive thank you to all of my lab mates that have helped me over the years, the wealth of knowledge I have gained from you is priceless.

To my parents, thank you for being my pillars of strength. I would not have been able to accomplish this without your unwavering love, support and patience. Thank you for always believing in me and for your constant encouragement. I love you both very much.

To Jacques, thank you for always being my sholder to cry on and always picking me up when I was down. Thank you for all those much needed pep talks when I was ready to throw in the towel.

A special thank you goes to Hamilton Ganesan from Inqaba Biotech for all the help with the CLC Genomics Workbench.

I would finally like to thank the National Research Foundation (NRF) and the University of the Witwatersrand for their financial support.

RESEARCH OUTPUTS

Publications

Unveiling the Micronome of Cassava (Manihot esculenta Crantz)

Rogans SJ, Rey C (2016) Unveiling the Micronome of Cassava (*Manihot esculenta* Crantz). *PLoS ONE* 11(1):e0147251.

Submitted: Small RNA and Methylation Responses in Susceptible and Tolerant Landraces of Cassava with *South African cassava mosaic virus*. *Virus Research.*

In preparation: Alteration of microRNAs in Susceptible and Tolerant cassava landraces infected with *South African cassava mosaic virus*. *PLoS Pathogens.*

Conference Oral Presentations

ORAL PRESENTATION: Comparison of microRNA populations in SACMV infected tolerant and susceptible cassava cultivars. *South African Society of Microbiology conference* held in Bela Bela, November2013.

ORAL PRESENTATION: Comparison of microRNA populations in SACMV infected tolerant and susceptible cassava cultivars. *ACGT Plant Biotechnology Forum* held in Pretoria at the Agricultural Research Council (ARC), April 2014.

ORAL PRESENTATION: Comparison of microRNA populations in SACMV infected tolerant and susceptible cassava cultivars. *South African Society of Microbiology conference* held in Umhlanga Rocks, January, 2016.

Poster Presentations

INTERNATIONAL POSTER: Biotechnology solutions for engineering virus resistant cassava. *Agricultural Biotechnology International Conference 2011 (ABIC 2011)* held in the Sandton Convention Center. 6-9 September 2011.

NATIONAL POSTER: Artificial miRNAs: A tool for Engineering Resistance to Gemniniviruses. *South African Society of Microbiology 2011 (SASM 2011)* held in the Cape Town, November 2011.

NATIONAL POSTER: Artificial miRNAs: A tool for Engineering Resistance to Gemniniviruses. *University of the Witwatersrand Post-Grad Symposium*. 2012.

INTERNATIONAL POSTER: Identification of putative microRNAs and their targets in cassava (*Manihot esculenta* Crantz.). *International Cassava Conference* held in Uganda, July 2012.

NATIONAL POSTER: Identification of putative microRNAs and their targets in cassava (*Manihot esculenta* Crantz.). *South African Society of Bioinformatics conference* held in Stellenbosch, September, 2012.

INTERNATIONAL POSTER: Identification of conserved microRNAs and their targets in cassava (*Manihot esculenta* Crantz.) using EST, GSS and Deep Sequencing Analysis. *EMBO Symposium* held in Heidelberg Germany, October 2012.

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

TABLE OF CONTENTS

Chapter 1

Literature Review

1. Introduction

1.1 Cassava

Cassava (*Manihot esculenta* Crantz) belongs to the family *Euphorbiaceae* and the Fabid superfamily (also known as eurosids I), which includes several distantly related plants such as rosids, legumes and poplars (Wurdack et al, 2005). Cassava is a woody perennial shrub, which grows from 1 to 5 meters in height (Fig. 1.1A). It is believed to have been cultivated, for 9000 years mainly for its starchy roots (Fig. 1.1B), making it one of agriculture's oldest crops. Although this dicotyledonous plant is reported to have low protein content, it produces clusters of tuberous roots that have high starch content ranging between 24-31% (Cock, 1985; Hillocks, 1997; Pandey et al, 2000).

Figure 1.1 A and B: Cassava plant (A) and roots (B). Photos taken in Uganda.

Due to its roots' high starch content, cassava is a rich source of dietary energy. Its energy per hectare is often very high, and potentially much higher than that of cereals (FAO, 1997; FAO, 2014). Cassava roots' dry mass is very rich in carbohydrates, amounting to about 250-300 kg for every tonne of fresh roots. In

addition, the roots contain significant amounts of vitamin C, thiamine, riboflavin and niacin (FAO, 1997). The leaves are also eaten as a vegetable and have been reported to have excellent nutritional value for both animals and plants, containing up to 25% protein, on a dry weight basis (Chavez et al, 2000; Ceballos et al, 2004). Worldwide, cassava is the second biggest source of starch, after maize, with production estimated at 8 million tonnes of starch a year (FAO, 2013). It is grown in 105 countries and serves as a staple food for nearly one billion people globally, many of who subsist on it (Burns et al, 2010; Nassar et al, 2002; Latif and Muller, 2015). In sub-Saharan Africa, cassava is the cheapest source of calories available and ranks as the second most important source of calories after maize (Nweke et al, 2001).

While some studies indicate that cassava has multiple centres of origin, others suggest that the cultivated species originated on the southern edge of the Brazilian Amazon and brought to West Africa in the 16th century by Portuguese navigators (Nassar, 1978; Hershey, 1987; Olsen and Schaal, 1999; Allem, 2002; Hillocks, 2002). Farming of cassava expanded in the 20th century when is emerged as an important food crop. It is now widely grown in tropical and subtropical countries of Africa, Asia and Latin America (FAO, 2013). Over the past decade, growth in cassava production has accelerated. FAO estimates put the global harvest in 2012 at more than 280 million tonnes, representing a 60% increase since 2000 and an annual growth rate double that of the previous two decades (FAO, 2012). Since 2000, global average yields per hectare have increased by almost 1.8% a year, from 10.4 tonnes per ha in 2000 to 12.8 tonnes in 2011, and the growth rate of cassava output in Africa has been equal to that of maize (FAO, 2013b). The output of cassava has increased most markedly in sub-Saharan Africa, which harvested 140.9 million tonnes, more than half of the global harvest, in 2011 (FAO, 2012).

The world trade in cassava amounted to approximately 8 billion US dollars in 2012 (FAO, 2014), and revolves around the export and import of dried cassava roots or starch. In sub-Saharan Africa it is an important source of employment and income, since most of the processing of this crop into food is done on a small scale

in rural areas. Harvested cassava roots are mainly consumed directly by many farm households or fed to their livestock. However, cassava root starch can also be used in a wide array of industries, including food manufacturing, pharmaceuticals, textiles, plywood, paper and adhesives, and as a feedstock for the production of ethanol biofuel (FAO, 2013; FAO, 2014). It is therefore increasingly being used as an industrial crop (Jansson et al, 2009). African countries have little or no presence in the industrial processing of cassava starch, apart from Nigeria and South Africa.

Relative to other crops, cassava has several agronomic traits that distinguish it as a food security crop that can be counted on to provide a source of nutrition during crop failures. Under the marginal conditions in which cassava is often grown, it produces more energy per unit area than most other crops and with limited human inputs (DAFF, 2010). The resilience of the crop to stress make cassava a major food security crop for subsistence farmers in sub-Saharan Africa (Cock, 1985; Romanoff and Lynam, 1992). Cassava is known to grow in degraded soil where almost nothing else can grow. It is naturally drought tolerant and resilient to climate changes, high temperatures, and poor soils, and in addition, cassava responds extremely well to high CO_2 concentrations, making it a very important crop for 21st century (Kawano et al, 1978; El-Sharkawy, 1993; El-Sharkawy, 2004; Jaramillo et al, 2005). Since it is propagated from stem cuttings, planting material is low-cost and readily available. Cassava is highly tolerant to acid soils, and has formed a symbiotic association with soil fungi that help its roots absorb phosphorous and micronutrients. Thanks to its efficient use of water and soil nutrients, and tolerance to sporadic pest attacks, cassava growers, using few if any inputs, can expect reasonable harvests where other crops would fail. One of the major positive attributes of cassava is that it does not have a specific harvesting period. Harvesting of some varieties can be “as needed”, at any time between six months and two years. During periods of food shortage, they can be harvested whenever needed, often one plant- or even one root- at a time, further demonstrating it as a food security crop and making it an attractive crop that can be used as a famine reserve (DeVries and Toenniessen, 2001 and Nweke et al. 2001). This makes it an excellent food security crop as when all other crops have

been exhausted, cassava roots can still be harvested. It is the reliability of cassava harvests that is most important to cassava farmers. Another factor that favours increased cassava production is the crop's potential to adapt well to climate change. A recent study of the impacts of climate change on major staple crops in Africa found that cassava was the least sensitive to the climatic conditions predicted in 2030, and that its suitability would actually increase in most of the 5.5 million sq. km area surveyed. Conversely, all other major food crops in the region, including maize, sorghum millet, beans, potatoes and bananas, were expected to suffer largely negative impacts (Jarvis et al., 2012).

These "hardy" traits have made cassava highly suitable for low-input, small-scale agriculture, while its inherent potentials have placed it among the crops most suitable for resource-poor farming in the tropics and neotropics under 21st century climate change scenarios (FAO, 2013). Cassava's new status in agriculture is a major step toward the realisation of a Global Cassava Development Strategy, adopted in 2001, after four years of consultation by FAO, the international Fund for Agricultural Development (IFAD), public and private sector partners and 22 cassava-producing countries. The strategy recognises cassava's potential not only to meet food security needs, but also to provide an engine for rural industrial development and a source of higher incomes for producers, processors and traders (FAO/IFAD 2001). In 2003, The New Partnership for Africa's Development (NEPAD) and the International Food Policy and Research Institute (IFPRI) promoted cassava as having the potential of becoming "a powerful poverty fighter in Africa". It was strongly recommended that the expansion of cassava would profit Africa both socially and economically. Following the conference held by NEPAD and IFPRI, the NEPAD PAN African cassava initiative was born. The overall aim of the initiative is to transform cassava production so that it can be exploited for food security as well as for generating capital. Furthermore, the initiative is beneficial, as it has been structured to have a long-term developmental goal where hunger and poverty are envisioned to be reduced by 50%, by the year 2015. African governments and international non-governmental organisations such as the United Nations International Children's Emergency Fund (UNICEF), the International Fund for Agricultural Development

(IFAD), and the international Centre for Tropical Agriculture (CIAT) have become fully involved in the multiplication and distribution of cassava in order to harness the huge potential of this crop for food and non-food applications (Babaleye, 1996).

In South Africa (SA), cassava is grown as a secondary staple food by small-scale farmers in the Mpumalanga, Limpopo and Kwazulu-Natal provinces for local sales or to traders from Swaziland and Mozambique (Daphne, 1980). Industrial processing of cassava roots for starch also presents economic potential for several regions and provinces in SA. Cassava 's food market potential is expanding at a rapid rate due to cassava's drought tolerance and sustainable cropping systems are maintained by small-holder farmers, especially in semi-arid regions of SA (Mathews, 2000). Additionally, cassava is used in the making of ethanol, production is less than maize (1 ton of fresh tuber supplies 180 liters of ethanol) but it yields rawer material (7 – 10 tons) than maize per hectare (Mathews, 2000;

Nuwamanya et al, 2011; Tonukari, 2004). SAB has also started producing cassava beers in Mozambique and Ghana (Fig. 1.2) (www.sabmiller.com).

Figure 1.2: Impala beer produced by SAB in Mozambique. Photo taken in Uganda.

1.2 Cassava mosaic disease (CMD)

Like all major crops, cassava is vulnerable to pests and diseases that can cause heavy yield losses. Diseases such as cassava brown streak virus disease (CBSD), bacterial blight (*Xanthomonas axonopodis* pv. *Manihotis*), and anthrachose (*Colletotrichum gloeosporiodes*) are among the most important diseases. However, the most important constraint limiting cassava production is viral disease. Cassava mosaic disease (CMD) is a geminivirus that is endemic to all cassava-growing regions in Africa. It is the most economically important and the single greatest constraint to cassava production (Herrera-Campo et al, 2011). CMD was first identified in Tanzania in 1894, but it was not evident that a virus was responsible for the disease until Storey (1936) suggested that a virus might be the causal agent as it was shown to be transmissible.

This viral disease is usually transmitted through the use of infected material as cassava is vegetatively propagated. In addition, whiteflies (*Bemisia*

tabaci) are vectors for the 11 species of geminiviruses that cause CMD (Fauquet and Fargette, 1990; Legg and Fauquet, 2004; Patil and Fauquet, 2009; Patil and Fauquet, 2015). Incidences of CMD infection can be as high as 100% of all plants in a given region, with average yield reductions of 30-40% and losses of at least 45 MT of fresh cassava roots each year (Legg and Thresh, 2000; Legg et al, 2006). Common symptoms (Fig. 1.3A-C) include misshapen leaves, chlorosis, mottling and mosaic. Plants suffer stunting and general decline, and the more severe symptoms, the lower the root yield. Symptoms may vary from plant to plant, due to differences in virus species and strains, sensitivity of the host, plant age, environmental factors as well as mixed infections (Legg and Thresh, 2000; Hillocks and Thresh, 2001).

Figure 1.3 A – D: (A –C) Cassava leaves showing typical symptoms of CMD, yellow mosaic patterns, leaf reduction, and distortion compared to (D) healthy leaves. Photos taken in Uganda.

CMD first appeared in 1894 in Tanzania, and several CMD epidemics in Africa have since been reported. The most recent outbreak – and by far the most economically important- began in Uganda in the late 1980s and led farmers to abandon the crop in many parts of the country (Otim-Nape et al, 1997; Otim-Nape and Thresh, 1998; Deng et al, 1997; Zhou et al, 1997). Subsequently the disease has invaded most of East and Central Africa (Legg and Fauquet, 2004). The pandemic of severe CMD has now affected 12 counties and continues to spread (Legg et al, 2006; Legg et al, 2015). The increasing spread of super-abundant

whiteflies raises justifiable fears that CMD will spread further on the African continent and worldwide. This would have major and unanticipated consequences for food security, economic development and social stability in many countries, as much of the world's cassava germplasm is highly susceptible to these viruses. Global warming is likely to also aggravate the situation, as higher temperatures will favour the whitefly vector. This potential additional impact from pests and diseases is all the more significant as cassava is one of the very few crops that may be relatively unscathed by future patterns of climate (Jarvis et al, 2012).

1.3 Geminiviruses

Viruses pose a serious threat to global agriculture, and as revealed by the worldwide crop productivity survey, the yield loss imposed by viral pathogens ranks second next to pathogenic fungi (Oerke and Dehne, 2004). Viruses are primarily grouped into families, of which the family *Geminiviridae* causes devastating diseases in both monocot and dicot crops. Geminiviruses are small circular single-stranded DNA (ssDNA) viruses and are named for their twinned icosahedral particles. They infect and cause severe losses in economically important crops worldwide (Mansoor et al, 2003). The incidence and severity of geminivirus diseases has greatly increased in the past 20 years (Mansoor et al, 2006). Geminivirus genomes can also undergo high levels of mutation, recombination, and reassortment to increase virus diversity (Duffy and Holmes, 2008; Martin et al, 2011).

Geminiviruses have been classified into seven genera namely *Becurtovirus*, *Begomovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus*, and *Turncurtovirus* on the basis of its genome organisation, insect vectors and biological properties (ICTV, 2012, http://ictvonline.org/virusTaxonomy.asp; Brown et al., 2015). These viruses are transmitted through insect and vectors such as whiteflies, leafhoppers, and treehoppers. These non-enveloped viruses possess circular, single-stranded DNA genomes (~2,7 Kb) that are packed into twinned icosahedral capsids. Geminivirus genomes are either monopartite (containing one

ssDNA) or bipartite (containing two ssDNA molecules known as DNA A and DNA B) (Fig. 1.4). Whether monopartite or bipartite, all geminiviruses contain an intergenic region, which holds a stem loops structure, located within the origin of replication, as well as divergent promoter elements responsible for sense and complementary-sense gene expression (Bisaro, 1996). Although the family of geminiviruses is large and includes multiple genera, all geminiviruses in general encode a movement protein, a coat protein and a replication initiator protein, that is required for rolling circle replication (RCR) of the virus. They have few but efficient proteins for their genome replication, movement, encapsidation and host RNAi suppressors. Two additional DNA molecules referred to as alpha- (DNA-α) and beta- (DNA-β) have shown to be associated with a monopartite begomoviruses (Rey et al, 2012). They repeatedly subsist in disease complex due to its high mutation rates, which allows them to acclimatise quickly against unfavourable environments. Furthermore, subgenomic defective interfering (DIs) DNAs, which are from deletions from their monopartite or bipartite helper virus, has also been reported.

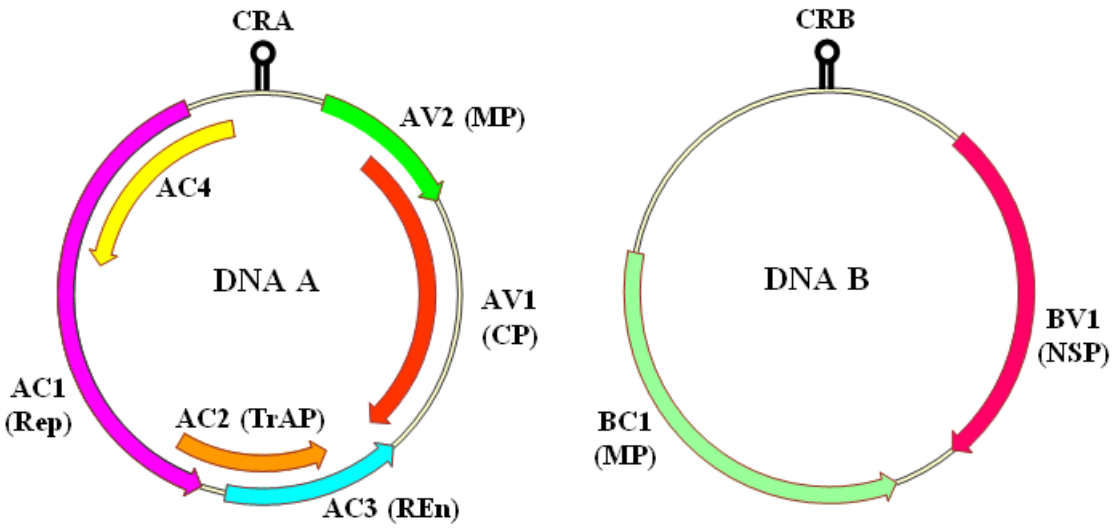


Figure 1.4: Genome organisation of DNA A and DNA B components of cassava-infecting begomoviruses. DNA-A contains six open reading frames (ORFs) and DNA B contains two ORFs (indicated by the coloured arrows). The

direction in which transcription occurs once the virus is within a plant host is depicted by the arrow head (Diagram modified from Berrie et al, 2001).

DNA A has six open reading frames (ORFs) (Fig. 1.4) namely, AV1 (coat protein), AV2 (pre-coat ORF), AC1 (replication initiation protein), AC2 (transcription activator protein), AC3 (replication enhancer), and AC4 (possibly determining symptom expression). The coat protein (CP or AV1) and the pre-coat protein (AV2), which is only found in Old World begomoviruses [such as *South African Cassava Mosaic Virus* (SACMV)] are encoded by the viral strand of DNA A. The complementary strand encodes for four proteins: AC1, AC2, AC3 and AC4 from overlapping ORFs. AC1 is required for initiation of DNA replication and is termed the replication-associated protein (Rep). The Rep protein is essential for the replication of both geminate molecules. It contains conserved functional domains that are responsible for the initiation and termination of the rolling circle replication (RCR) (Lafs et al, 1995; Hanley-Bowdoin et al, 2000). AC2 has been reported to contain a mono-directional promoter and its gene product (TrAP) is responsible for the trans-activation of virion-sense gene expression in both DNA A and DNA B (Gutierrez, 1999; Legg and Fauquet, 2004). TrAP localises to the plant host nucleus where is preferentially binds to ssDNA. AC3 is the DNA replication enhancer (Ren) (Gafni and Epel, 2002; Harrison and Robinson, 2002). The AC4 ORF lies embedded with the coding region of the Rep protein (AC1) and it is the least conserved of all the geminiviral proteins, both is sequence and function (Bisaro, 2006). Agroinfiltration assays conducted by Vanitharani et al. in 2004 and 2005 showed that the AC4 protein of African cassava mosaic virus-Cameroon (ACMV-CM) and Sri Lankan cassava mosaic virus (SLCMV) is involved in RNA silencing, which in turns results in enhanced disease progression. The ORFs on DNA B (Fig.4) encodes the proteins that are necessary and directly involved in the efficient systemic spread of the virus, BV1 (encoding nuclear shuttle protein) and BC1 (encodes for movement protein). BC1 is found on the complementary strand and mediates cell-to-cell movement of the virus. BV1 is the nuclear shuttle protein (NSP), which controls the movement of the viral DNA between the nucleus and the cytoplasm (Gafni and Epel, 2002; Harrison and Robinson, 2002).

Geminivirus DNA replication occurs in the nucleus of the host, requiring two stages for replication, firstly ssDNA conversion to dsDNA, and secondly rolling-circle replication (RCR). In the first stage the viral circular genomic ssDNA (positive strand) is converted into supercoiled covalent dsDNA intermediates through a priming event activating the negative strand origin of DNA replication. These dsDNA intermediates are then amplified through a RCR mechanism (Gutierrez, 1999). The initiation site for RCR has been mapped to the intergenic region, which contains an invariant nine-nucleotide sequence (TAATATTAC) that is common among all geminiviruses. Geminiviruses use a replication at a high conserved stem-loop structure located between two major open reading frames of the genome (Hanley-Bowdoin et al, 1999; Jeske, 2009). Following infection, the virus particle enters the nucleus via a nuclear localisation signal on the coat protein, and after release of the viral ssDNA, host cell DNA polymerase I and components of DNA repair machinery synthesise a complementary strand to generate a dsDNA intermediate. Upon association with nucleosomes to form a "minichromosome" the replication by nicking the virion strand of dsDNA template at a highly conserved nonanucleotide sequence contained within the origin of replication. Then, using the host cell replication machinery, Rep generates multiple copies of virion-sense strand ssDNA using the complementary-sense strand as a template. The virion-sense strand becomes displaced from the template strand, and then is nicked and religated by Rep to be released as multiple copies of circular ssDNAs, which can either undergo RCR or become packaged into mature virions during the late stage of infection (Hanley-Bowdoin et al, 1999; Jeske, 2009). This mechanism of replication is highly effective and can result in expression of tens of thousands of copies of the viral genome per cell. The genomic components have regions and motifs to control the viral gene replication and expression (Orozco and Hanley-Bowdoin, 1996).

In order to complete an infection cycle in a host, geminiviruses disrupt many host processes, which involve transcriptional regulation, DNA replication, cell cycle control, and macromolecular trafficking in plants. Certain pathways such as those involved in plasmodesmata structure and function are altered and

silencing of defence-related mechanisms occurs (Gutierrez, 2002). In addition, geminivirus pathogenicity proteins interfere with host proteins such as NAC transcription factor domains as well as with retinoblastoma-related (RBR) pathways (Gutierrez, 2002). The Rep/AC1 protein functions by nicking the DNA once it has bound to the stem structure at the replication origin and initiates rolling-circle DNA replication. One problem encountered by geminiviruses is that they infect terminally differentiated cells at the resting state (G0), which lack factors required for DNA replication. In order to overcome this problem geminiviruses induce host proliferating cell nuclear antigen (PCNA), which is a DNA polymerase accessory factor, normally found in the S-phase. Rep/AC1 is therefore required to induce PCNA. The process involves binding of Rep/AC1 to the viral replication enhancer (AC3), which then binds to PCNA (Arguello-Astorga *et al.*, 2004). Rep/AC1 in turn, physically interacts with host encoded retinoblastoma-like tumor suppressor proteins (pRbs). In mammals, the pRB protein functions as the G1 checkpoint regulator that prevents completion of G1 and entry into the S-phase. Cyclin-dependent kinases have phosphorylating activity, which stops the cell-cycle inhibition activity of pRb, therefore allowing progression into S-phase. It is believed that geminivirus Rep proteins interact with and either inactivate or divert the pRb-like protein in infected cells, allowing S-phase- specific mRNA production, also providing a pool of factors and enzymes required for viral DNA replication (Hanley-Bowdoin et al., 2004). Plant geminiviruses are therefore analogous to animal DNA tumor-inducing viruses (such as SV40) and adenoviruses. These viruses also encode proteins, which affect cell cycling apparatus (Carrington and Whitham, 1998). In addition, geminiviruses encode multiple silencing suppressors that interfere with plant small interfering RNA (siRNA) production and alter plant DNA methylation and microRNA (miRNA) pathways, often causing developmental abnormalities (Aregger et al, 2012; Bisaro, 2006; Hanley-Bowdoin et al, 2013; Rodriguez-Negrete et al, 2013).

1.4 Plant-Pathogen interactions

The interactions between a plant and its pathogens involve two-way communication. Not only must the plant be able to recognise and defend itself against a potential pathogen landing on its surface, but also the pathogen must be

able to manipulate the biology of the plant to create a suitable environment for its growth and reproduction. Both plants and pathogens have evolved a suite of genes that enables this communication (Jones and Dangl, 2006). According to current plant immunity descriptions, there are two layers of plant immune responses against microbial pathogens. The first line of active plant defences involves the recognition of the pathogen or microbe-associated molecular patterns (PAMPS or MAMPS) by plant pattern recognition receptors (PRR), which triggers the general plant defence responses referred to as PAMP-triggered immunity (PTI) (Jones and Dangl, 2006; Bent and Mackey, 2007; Boller and Felix, 2009; Dodds and Rathjen, 2010; Schwessinger and Ronal, 2012). As a counter-response to plant PTI defence, adapted microbes deliver specific 'effector' proteins into plant cells, which compromise PTI defences and interfere with host defence signalling by Supressing the different components of PTI.

The second line of plant defence involves the recognition of specific effectors [referred to as avirulence (Avr) proteins] by resistance (R) genes encoded by the plant, triggering what is often perceived as a stronger resistance response and referred to as effector-triggered immunity (ETI). This R/Avr-gene recognition has been termed 'gene-for-gene resistance' (Gururani et al, 2012). To further defend the action of the microbial effectors, plants evolved specific surveillance systems involving receptor-like proteins (R proteins) that directly or indirectly recognise the microbial effectors or monitor their activities in the cell to trigger effector-triggered immune (ETI) response. An effector protein can also be the elicitor of ETI defence. Whether the effector or elicitor role of an effector protein prevails is primarily predicted on the presence of the complementary R gene in the plant. The ETI response, and to somewhat lesser extent the PTI responses, are closely associated with or even culminate in HR, thus imparting resistance against invading pathogens (Jones and Dangl, 2006).

Based on current definitions of microbial P/MAMPS and effectors (Jones and Dangl, 2006; Bent and Mackey, 2007; Boller and Felix, 2009; Dodds and Rathjen, 2010; Schwessinger and Roland, 2012) viruses are not generally viewed as encoding P/MAMPs or effectors, and antiviral immune responses triggered via

the R proteins are not typically classified as ETI response, although R protein-mediated responses are triggered by virus-encoded proteins. There is growing evidence that R-gene-mediated immunity and innate immunity i.e. RNA silencing function in concert to defend plants against viruses. To counteract this dual defence and establish infection in susceptible hosts, some viral pathogen proteins have evolved additional functions to supress both PTI/ETI innate responses and RNA silencing.

Plant responses to virus infection

R-gene mediated responses to virus infection

Over the past decade, several R genes that mediate resistance against viruses have been identified (Collier and Moffett, 2009; Gururani et al, 2012). The majority of the cloned dominant R genes encode the conserved nucleotide binding (NB) and LRR family proteins (Collier and Moffett, 2009; Moffett, 2009; Gururani et al, 2012). NB-LRR proteins also contain additional N-terminal domains such as the TIR homology domain, a CC domain, a *Solanaceae* domain, or a predicted BED zinc finger domain. Until recently, the LRR domain was thought to be the major domain critical for R protein function. However, growing evidence indicates that both the LRR and the N terminus domains (TIR and CC) are critical for proper resistance responses. The two domains function through intramolecular interactions and interactions with other proteins (R cofactors) to mediate recognition of pathogen elicitors (Collier and Moffett, 2009; Moffett, 2009).

In viral infections, in addition to the dominant R gene-related resistance responses, another form of recessive resistance exists that is typically derived by a loss of function in host proteins critical for the establishment of disease (Robaglia and Caranta, 2006; Truniger and Aranda, 2009; Gururani et al., 2012). Resistance genes, particularly those encoding the NB-LRR proteins, have well-conserved roles in plants, which are to guard the host cells against diverse viral and nonviral pathogens and to trigger disease resistance. Moreover, the general mechanism of the recognition of R proteins and Avr factors appears to be similar for viral and nonviral pathogens, whereby R cofactors play crucial roles to guide or modulate R/Avr interactions, ultimately activating HR and resistance

responses. While the predominant antiviral resistance responses are mediated by the dominant R genes, other host proteins, such as the elongation initiation factors, TOM proteins, ER chaperones, calreticulins, and lectin proteins, also influence host resistance against diverse viral infections.

Current advances in cassava biology have led to the construction of a genetic map consisting of 1061 genes coding for resistance protein analogs (RPAs) (Soto et al, 2015) and identification of 327 genes encoding for RPAs of NBS-LRR class (Lozano et al, 2015). Developing tolerant or resistant plants requires an initial exploration of immunity-related genes (IRGs) in potential hosts interacting with a putative pathogen. In a study by Louis and Rey, (2015) the transcriptome data of tolerant cassava TME3 (which exhibits a recovery phenotype) and susceptible cassava T200 infected with *South African cassava mosaic virus* (SACMV) were explored for RGAs. Putative resistance protein analogs (RPAs) with amide-like indole-3-acetic acid–Ile-Leu- Arg (IAA-ILR) and leucine-rich repeat (LRR)-kinase conserved domains were unique to TME3. In TME3 and T200 common responsive RPAs were the dirigent-like protein, coil–coil nucleotide-binding site (NBS) and toll- interleukin-resistance, disease resistance zinc finger chromosome condensation-like protein (DZC), and NBS- apoptosis repressor with caspase recruitment (ARC)–LRR domains. Mutation in RPAs in the MHD motif of the NBS- ARC2 subdomain associated with the recovery phase in TME3 was observed. Additionally, a cohort of 25 RGAs mined solely during the recovery process in TME3 was identified. Phylogenetic and expression analyses support that diverse RGAs are differentially expressed during tolerance and recovery. This study revealed that in cassava RGAs participate in tolerance and differentially accumulate during recovery as a complementary defence mechanism to natural occurring RNA silencing to inhibit viral replication.

Viral components, such as dsRNA, single-stranded RNA and DNA are sensed by three classes of receptors: retinoic acid-inducible gene I-like, Toll-like and nucleotide oligomerization domain-like receptors (Takeuchi and Akira, 2009), with the latter being similar to plant NB-LRRs (Bonardi et al., 2012). In plants, there is no evidence for recognition of viral RNA or DNA by immune

receptors and the RNA silencing system have evolved to recognise and target viral nucleic acids. No viral PAMPs have identified so far and PTI-based antiviral response can potentially be elicited by plant DAMPs. Additionally, 'modified self' products of the viral effector activity can be sensed by unconventional CC-NB-LRRs such as ADRs. Viral and non-viral pathogens induce similar immune reactions (Soosaar et al., 2005) thus, local virus infection leads to SAR in uninfected tissues. Little is known about DNA viruses and R gene-mediated immunity. In the case of geminiviruses (family Geminiviridae), the CC-NB-LRR gene CYR1 was recently implicated in resistance to *Mugbean yellow mosaic India virus* (Maiti et al., 2011).

To achieve successful infection, a pathogen must be able to manipulate the cellular environment of the host plant, not only suppressing the natural defence responses of the plant, but also altering the cellular environment to allow it to grow and reproduce. This is achieved by the production of an arsenal of proteins, collectively known as effectors, which target plant defence pathways and metabolism (Koeck et al., 2011). Increasing numbers of pathogen effectors are being identified, and understanding the role of these effectors in suppressing the general PTI plant defence responses will enable the development of new approaches to disease resistance.

Tolerance and recovery in plants

Resistance is defined as the ability to limit parasite burden to non-detectable levels of virus replication (Räberg et al, 2007). Ideally, a fully resistant plant would not be infected by a virus and would not show any symptoms and no detectable viral titre (Bruening et al, 2006). Conversely, tolerance to virus infection leads to reduced crop damage (Fraile and García-Arenal, 2010) and is associated with persistent virus replication at low level (Bruening et al, 2006). Plant recovery from viral-induced symptoms is phenotypically manifested by a progressive reduction in symptom severity or appearance of symptomless leaves at the apices. Recovery in some cases is a key phenotypic indicator of tolerance, characterised by natural RNA silencing defence mechanism (Rodríquez-Negrete, 2009; Nie and Molen, 2015).

1.5 RNA silencing

Plants live in a fluctuant, unpredictable environment and being sessile, they are exposed to a large number of potential stressors. Physiological flexibility is therefore a crucial attribute for plants when coping with biotic and abiotic stresses. Therefore, the regulation of gene expression is a key element in remaining adaptive to variable stresses. RNA silencing, also known as RNA interference (RNAi), is an essential genetic regulatory mechanism conserved in eukaryotic organism. The term RNA silencing refers to the nucleotide-sequence-specific inhibition pathways mediated by small RNAs. RNAi can act at transcriptional (Transcriptional Gene Silencing, TGS) or at post-transcriptional levels (Post-transcriptional Gene Silencing, PTGS) (Ding and Voinnet, 2007; Baulcombe, 2004; Brodersen and Voinnet, 2006; Chapman and Carrington, 2007; Vaucheret, 2006; Voinnet, 2009), and has many diverse roles including developmental regulation, stress response or defence against invading nucleic acids like transposons or viruses (Ding and Voinnet, 2007; Baulcombe, 2004; Brodersen and Voinnet, 2006; Chapman and Carrington, 2007; Vaucheret, 2006; Voinnet, 2009; Pumplin and Voinnet, 2013).

The silencing of RNA relies on host- or virus-derived 21-24 nucleotide long sRNA molecules, which are key mediators of RNA silencing-related pathways in plants and other eukaryotic organisms (Voinnet, 2009; Ruiz-Ferrer and Voinnet, 2009; Llave, 2010). In plants there are two main types of sRNAs, microRNAs (miRNAs) and short interfering RNAs (siRNAs) (Vaucheret, 2006; Brosnan and Voinnet, 2009). MiRNAs are derived from single-stranded (ss) RNAs folded into short imperfect stem-loop structures and siRNAs are derived from dsRNAs, resulting from the folding of long inverted repeats (IR), convergent transcription or the action of RNA-dependent RNA polymerase (RDR) on ssRNA. MiRNAs and siRNAs all associate with Argonaute (AGO) proteins to guide TGS or PTGS on cognate targets based on their homology (Mochizuki et al., 2002). The endogenous small RNA repertoire of wild-type plants grown under standard conditions consists of 10% miRNAs and 90% siRNAs (Kasschau et al., 2007). Among the siRNA category, different types exist, including *trans*-acting siRNAs (tasiRNA),

natural antisense transcript-derived siRNAs (nat-siRNAs), endogenous siRNAs (endo-siRNAs), DNA-dependent RNA Polymerase IV (PollV)/PolV siRNAs (p4/p5-siRNAs) and Needed for RDR2 Independent DNA methylation (NERD) siRNAs.

These sRNAs are produced from double-stranded RNA (dsRNA) or from folded structures by Dicer-like proteins (DCLs), and they guide Argonaute (AGO) proteins to target cognate RNA or DNA sequences (Ruiz-Ferrer and Voinnet, 2009). These endogenous sRNAs play important roles in many aspects of gene regulation in plants, controlling developmental programming or biotic and abiotic stress responses (Ruiz-Ferrer and Voinnet, 2009). Both cellular and antiviral siRNA biogenesis often require RNA-dependent RNA polymerases (RDRs). In the model plant *Arabidopsis* (*Arabidopsis thaliana*), there are four DCLs, ten AGOs and six RDRs (Vaucheret, 2006), which are specialised for different silencing related pathways.

Plants (and other organisms) use silencing for three purposes: creating and maintaining heterochromatin or repetitive DNA and transposons; regulating development, stress response and other endogenous regulatory functions; and defending against viral and bacterial infections (Hohn and Vazquez, 2011). The core of RNA silencing is the formation and recognition of double-stranded (ds) RNA, which otherwise does not play a role in cellular genome replication and expression and which does not elicit an interferon response in plants. When recognised by DICER-LIKE (DCL) and dsRNA binding proteins (DRB), dsRNA is diced into 21-24 nt small RNA (sRNA) duplexes with 2-nt-3'-overhangs (see Ref Vazquez et al, 2010 for more details). These duplexes interact with Argonaute (AGO) and associated proteins to form RNA-induced silencing complexes (RISCs) with one of the sRNA strands in a process that is accompanied by the release/degradation of the other "passenger" strand. RISCs are either involved in chromatin modification or in translation inhibition and cognate RNA degradation (slicing) depending on the AGO effector (Mallory and Vaucheret, 2010) and

associated GW/WG motif containing proteins (El-shami et al, 2007; Azevedo et al, 2010).

Cleavage products of the target RNAs are "aberrant" and can serve as template of RNA-dependent RNA-polymerases (RDR) to form dsRNAs, which again can initiate silencing in an autocatalytic, self-sustained manner. The enzymes involved in the silencing pathways are members of protein families i.e. Arabidopsis has four DCLs, four DRBs, ten AGOs and six RDRs, specifically involved in different silencing pathways, with partially redundant functions. Plant DCLs differ in the size of the sRNAs they produce: DCL3 produces 24 nt long sRNAs, DCL2 22 nt ones and DCL4 21 ones from long perfectly paired RNAs. DCL1 also produces 21 and 22 nucleotides long sRNAs but preferentially form short imperfectly hairpins.

NGS data indicates that specific subsets of AGO proteins are connected to each DCL and that the stabilization of sRNAs into each AGO depends on preferential, hierarchical binding affinity of each AGO for the 5' terminal nucleotide of the sRNAs (Mi et al, 2008). AGO1 favours 21 nt or 22 nt 5'U-, AGO2 21 nt 5'A-, AGO4 24 nt 5'A- and AGO5 21-24 nt 5'C-terminated sRNAs. AGO7 binds specifically to miR390 (Mi et al, 2008; Montgomery et al, 2008; Takeda et al, 2008).

Mechanistically, the RNA silencing process consists of initiation phase, effector phase and the amplification phase**.** During silencing initiation double-stranded RNAs (dsRNAs) of different origins are processed by an RNase III type enzyme Dicer, in plants known as DICER-LIKE proteins (DCL), into short, 21-24 nt long, small RNA (sRNA) duplexes (Hamilton and Baulcombe, 1999; Hutvagner et al, 2001). DICERs require DOUBLE-STRANDED RNA BINDING (DRB) proteins for accurate sRNA production (Eamens et al, 2012a, b; Hiraguri et al, 2005)**.** The sRNAs are stabilised at their 3' end by the HUA Enhancer 1 (HEN1)-dependent methylation (Boutet et al, 2003; Yang et al, 2006) and exported from the nucleus by HASTY (HST) (Park et al, 2005) to be loaded onto Argonaute proteins (Hammond et al, 2001; Liu et al, 2004) the effectors of the RNA-Induced Silencing

Complex (RISC) (Lee et al, 2004; Tomari et at, 2004) or RNA Induced Transcriptional Silencing complex (RITS) (Ekwall, 2004). Guided by the sRNA sequence, RISC induces slicing or translational repression of its target RNAs (during PTGS) in a sequence-specific manner, whereas RITS complex causes histone and/or DNA methylation, resulting in transcriptional gene silencing (TGS) of the homologous gene (Creamer and Partridge, 2011)

In plants, the effector step can result in amplification of silencing response involving RNA-dependent RNA polimerases (RDRs) proteins (Dalmay et al, 2000; Vaistij et al, 2002). Amplification of RNA silencing has been implicated in the spread of an RNA silencing signal, a non-cell-autonomous process (Schwach et al, 2005)

The best-studied plant model, *Arabidopsis thaliana* genome encodes 4 members of DCLs (DCL1-DCL4) (Bologna and Voinnet, 2014), five DRBs (HYL1/DRB1, DRB2, 3,4,5) (Hiraguri et al, 2005), 10 AGOs (AGO1-10) (Mallory and Vaucheret, 2010) and 6 RDRs (Rdr1, 2, 3a, 3b, 3c and 6) (Wassenegger and Krczal, 2006). These proteins have partially redundant roles and combine with each other to result in diverse classes of small RNAs and different effector outputs of the RNA silencing pathways. The small RNA classes identified in plants include microRNAs (miRNAs), trans-acting small interfering RNAs (tasiRNAs), natural-antisense RNAs (nat-siRNAs), repeat-associated siRNAs (ra-siRNAs), viral siRNAs (vsRNAs) and virus-activated siRNAs (vasiRNAs). These classes possess specialised roles during development, stress responses, heterochromatic silencing, viral infection and host-pathogen interplay, respectively (Bologna and Voinnet, 2014).

To understand the molecular mechanism of host defence mechanism during host-virus interaction, it is imperative to study the siRNA generation and their characterisation. Deep sequencing of sRNA pool is powerful tool to identify the consensus and specific siRNA. Simultaneously, the identification of their

targets would further provide insight pathway and uncover the route through which viruses' cause disease.

DICER-LIKE proteins

Dicer RNase-III endonucleases process long dsRNA into sRNA duplexes exhibiting 2-nucleotide (nt) 3' overhangs and 5' monophosphates. Both animal and plant DICER-LIKE (DCL) proteins display DExD-box, Helicase-C, domain of unknown function 283 (DUF283), PIWI/ARGONAUTE/ZWILLE (PAZ), RNase-III, and dsRNA-binding domain (dsRBD) domains (Margis et al, 2006). The distinct dsRNA recognition modes by these domains underlie the accuracy and specificity of sRNA processing from distinct substrates. In the first mode, one end of a near-perfect dsRNA helix is anchored to the PAZ domain, which is connected to the catalytic domain through an α-helix (the "ruler") whose length determines the size of processed sRNAs. In the absence or hindrance of PAZ, a second mode entails nonspecific dsRNA binding via the dsRBD. A third mode involves recognition of imperfect stem-loop substrates (i.e. premiRNA) via the binding of single-stranded loops by the ATPase/helicase domain, acting as the primary RNA sensor for discriminating between premiRNA and long dsRNA substrates, even though the NTPase activity is dispensable for miRNA biogenesis. Dicer helicase likely functions as an ATP-dependent translocase that provides the energy required for multiple cuts along dsRNA substrates. Plant *DCL* genes from a monophyletic group spawned after the plant-animal split but before the monocot-dicot divergence 150 million years ago (Henderson et al, 2006). They share structural similarities with their animal counterparts, suggesting that the biochemical properties mentioned above are also applicable to their varied modes of action.

MicroRNA-generating DCL1

The DCL1 domain architecture comprises a DExD/H-box RNA helicase, DUF283, PAZ, two tandem RNase-III domains, and two tandem dsRBDs (Margis et al, 2006). The involvement of at least two cycles by DCL1 for pri- to pre-miRNA and pre- to mature-miRNA processing likely explains ATP dependency to plant miRNA biogenesis (Bologna et al, 2009, Kurihara and Watanabe 2004). The second C-terminal dsRBD, also found in DCL3 and DCL4 but not in DCL2, promotes

DCL1 localization into the nucleus, where it assembles with other miRNA biogenesis factors in specialised dicing bodies (Burdisso et al, 2012; Fang and Spector, 2007). DCL1 undergoes negative feedback regulation by two of its own miRNA products: miR162 targets the DCL1 mRNA at the junction of exon 12 and 13, whereas the miR838 precursor resides with intron 14, such that its processing results in DCL1 splicing, generating two non-productive mRNA fragments (Rajagopalan et al, 2006).

Small-interfering RNA-generating DCL2, DCL3 and DCL4

Arabidopsis DCL2, DCL3 and DCL4 process long, near-perfect dsRNA substrates into populations of 22, 24 and 21 nt siRNAs, respectively (Henderson et al, 2006). The distance separating the PAZ and catalytic domains of each protein underpins these size specifications. DCL3 action in transcriptional gene silencing (TGS), mediated by 24 nt siRNAs, is generally linked to RDR2 products originating from transposons and repeats undergoing RNA-directed DNA methylation (RdDM) (Pontes et al, 2006). DCL4-dependent posttranscriptional gene silencing (PTGS), mediated 21 nt siRNAs, initiates from endogenous RDR6 and RDR1 products, including *trans*-acting siRNAs (tasiRNAs) precursors. DCL2 and its 22 nt siRNA products act redundantly in TGS and PTGS, downstream, of RDR6 and RDR2 (Jauvion et al, 2012).

Generic substrates for the three DCLs also include intramolecular RNA fold-back transcripts originating from endogenous inverted repeats (IR) loci or their transgenic counterparts; some evolutionary young pre-miRNAs (e.g. miR822) also fold into near-perfect IRs, and, as such, are processed by DCL4 instead of DCL1 (Rajagopalan et al, 2006). Generic exogenous DCL2, DCL3, and DCL4 substrates comprise virus-derived dsRNA produced by the combined action of virus- and host-encoded RDRs, including RDR1/6 (RNA viruses) and RDR2 (DNA viruses) (Pumplin and Voinnet, 2013). The dominant action of a particular DCL on a dsRNA substrate underlies not only the specific size but also the 5'-nucleotide identity of siRNA duplexes, and both of these critically influence their particular ARGONAUTE (AGO) protein and, ultimately, their biological output. Depending on the availability of particular DCLs in specific tissues or cell types, a

given dsRNA may thus be processed into siRNAs with drastically different modes of action.

RNA-dependent RNA polymerases

RDRs, defined by a conserved catalytic domain required for copying single-stranded RNA (ssRNA) into dsRNA, are found in RNA viruses, plants, fungi, protists, and *C. elegans* but are notably absent in *Drosophila* and mammals. One member of each of the three eukaryotic RDR clades – RDRα, RDRβ, and RDRγ- was present in the most recent common ancestor of plants, animals and fungi (Zong et al, 2009). All known functions of plant RDRs are coordinated with the sequential processing of their long dsRNA products by one or several DCLs into secondary siRNAs. These differ from primary siRNAs, which may trigger, directly or indirectly, RDR activity via direct priming or by enabling AGO-directed endonucleolytic cleavage as a starting point for dsRNA synthesis.

Among the six *Arabidopsis* RDRs, RDR1, RDR2, and RDR6 (RDRα clade) share the canonical C-terminal catalytic DLDGD motif of eukaryotic RDRs. They show functional diversification in distinct endogenous silencing pathways by being linked to the action of specific siRNA-processing DCLs. Loss of RDR1, RDR2 and RDR6 function also enhances plant susceptibility to viral infection (Pumplin and Voinnet, 2013). RDRα clade members also have indirect roles in defence against nonviral pathogens (e.g., bacteria, oomycetes, and nematodes) and herbivores by producing endogenous regulatory sRNAs, including tasiRNAs and natural antisense transcript siRNAs (nat-siRNAs) (Katiyar-Agarwal et al, 2006; Pandey and Baldwin, 2007). *Arabidopsis* RDR3, RDR4, and RDR5, defining the RDRγ clade, display an atypical catalytic DFDGD motif and have not yet been assigned any RNA silencing functions (Zong et al, 2009). Nonetheless, all six RDRs

show distinct developmental and stress-responsive expression patterns (Willmann et al, 2011).

AGONAUTE Proteins

AGO and AGO-like proteins are the main RNA silencing effectors across all kingdoms. The *Arabidopsis* genome encodes 10 *AGO* genes, defining three major phylogenetic clades: *AGO1, -5*, and *-10; AGO2, -3,* and *-7*; and *AGO4, -6, -8,* and *-9* (Mallory et al, 2009). Recently, 13 AGOs have been identified in cassava including an AGO9 homologue (Mirzaei et al, 2014). Canonical eukaryotic AGOs contain four main domains; a variable N-terminal domain and the more highly conserved PAZ, MID, and PIWI domains, which together correctly positions sRNAs relative to their targets. PAZ, MID, and PIWI are connected by the L1 and L2 linker regions. AGOs fold into a bilobal structure displaying a central groove for sRNA binding (Wang et al, 2009). A nucleotide-specificity loop lining the sRNA-binding pocket in the MID domain recognises the 5' nucleotide of sRNAs, and the PAZ domain binds the 3' terminal end (Frank et al, 2012). The PIWI domain adopts an RNase-H-like fold and exhibits endonuclease (slicer) activity mediated by an Asp-Asp-His (DDH) catalytic triad (115), although the DDH domain is not always sufficient for slicing (Liu et al, 2004).

Silencing has been experimentally demonstrated for Arabidopsis AGO1, -2, -7 and -10 (mediating PTGS) and AGO4 (mediating TGS) (Zhu et al, 2011). Studies of plant immunoprecipitates have revealed that the sRNA size and 5'-terminal nucleotide bias the loading of these proteins (Mi et al, 2008; Montgomery et al, 2008; Zhu et al, 2011). Thus, AGO4, -6, and -9 associate mostly with 24 nt siRNA, whereas AGO1, -2, -5, -7 and -10 bind 21-22 nt molecules. AGO7 and -10 are associated almost exclusively with miR390 and miR165/166, respectively, whereas AGO1, -2, and -5 preferentially bind sRNAs exhibiting a 5'-end uridine, adenosine, or cytosine, respectively (Mi et al, 2008). In addition, AGO4, -6, and -9 associate primarily with 5'-adenosine sRNAs. Mutational analysis studies have confirmed the importance of the 5'-nucleotide identity in AGO sorting of some sRNAs, further studies also revealed additional requirements (including base-pair mismatches or protein interactions) for sorting of others. For example, most of the

MIR165/166 family members contain a 5'-terminal uridine, normally licensing them for AGO1 loading, but these specifically associate with AGO10 instead (Zhu et al, 2011). Likewise, miR390 selectively loads into AGO7 instead of AGO2 despite having an adenosine at its 5'-terminal (Montgomery et al, 2008).

AGO1, -5, and -10 clade

In addition to its central role in miRNA functioning and tasiRNA production/activity (DCL1- and DCL1/DCL4-dependent processes, respectively), AGO1 also mediates antiviral silencing upon loading with 21- and 22- nt virus-derived siRNAs (vsRNAs) produced by DCL4 and DCL2, respectively (Pumplin and Voinnet, 2013). Verified modes of AGO1 action in some of these pathways include slicing as well as TR, possibly coupled to RNA decay. AGO1 levels are also regulated during its loading with sRNA, which requires HEAT-SHOCK PROTEIN 90 (HSP90) and the Arabidopsis cyclophilin-40 ortholog SQUINT (SQN) (Iki et al, 2010; Iki et al, 2012). Autophagic degradation of unloaded AGO1 and FBW2-mediated control of AGO1 loading/chaperoning are part a homeostatic control mechanism that enables AGO1 steady-state levels to remain relatively constant under adverse environmental or stress conditions, including virus infection. Similar to DCL1 control by miR162, AGO1 homeostasis also entails its regulation by miR168 via (a) miR168-AGO1-dependent slicing of AGO1 mRNA, (b) TR of AGO1 mRNA in miR168-AGO10-dependent manner, and (c) increased miR168 accumulation in response too elevated AGO1 levels (Mallory and Vaucheret, 2010). Homeostatic AGO1 control also entails production of specific 22 nt miR168 isoforms that upon cleavage of the AGO1 mRNA, instigate production of RDR6-dependent secondary siRNAs to further strengthen AGO1 downregulation. This secondary siRNA production, initiated on target 3'-cleavage products, can be triggered by other 22 nt isoforms from other miRNAs that are normally processed by DCL1 as cognate 21 nt species.

AGO10 (also known as ZWILLE or PINHEAD), the closest AGO1 homolog in *Arabidopsis,* regulates shoot apical meristem development by specifically binding members of the *MIR165/166* family. These miRNAs also associate with AGO1 to supress class-III homeodomain-leucine-zipper (HD-ZIP-III) transcription factors

required for shoot apical meristem establishment (Mallory et al, 2009). AGO1 accumulates ubiquitously in all plant tissues; AGO10 is expressed moderately in the shoot apical meristem, in only the adaxial domains of leaf primordial, and at higher levels in vascular precursors. A current model proposes that AGO10 competes with AGO1 for miR165/166 to protect HD-ZIP-III transcripts from repression in the shoot apical meristem and leaf adaxial domains. AGO10 also mediates TR of several endogenous miRNA target genes, including *AGO1* (Mallory et al, 2009). Laser-capture microdissection identified significant *AGO5* expression in and/or around developing megaspores during the transition to megagametogenesis. In male gametophytes, AGO5 localises preferentially in the sperm cell cytoplasm of mature pollen, where, given its analogy to AGO1, it may direct miRNA- and siRNA-mediated functions required for male gametophyte development or cell type specification (Tucker et al, 2012).

The AGO2, -3, and -7 clade

Although it belongs to a different clade, AGO2 displays both additive and overlapping activity with AGO1. For instance, AGO1 and AGO2 redundantly regulate the plantacyanin mRNA via miR408, and both proteins are required for siRNA-mediated silencing of transcribed, nonconserved intergenic regions, pseudogenes, and evolutionary young transposons in Arabidopsis (Maunoury and Vaucheret, 2011). Recent studies also implicate AGO2 in defence against a broad range of viruses. Like AGO1, AGO2 is induced and loaded with DCL4- and DCL2-dependent viRNAs in virus-infected plants. Also like hypomorphic *ago1* mutants, *ago2* plants are hypersusceptible to viruses; a phenotype enhanced in *ago1 ago2* double mutants, indicating the additive and nonoverlapping effects of these proteins. Loss of *AGO2* function was also sufficient to allow systemic infection of viruses not normally hosted by *Arabidopsis*, whereas AGO1 had no effect on this host-range determination (Pumplin and Voinnet, 2013). AGO2 levels are regulated by miR403 in an AGO1-dependent manner (Allen et al, 2005). This

regulatory network may allow AGO2 to take over antiviral defence when AGO1 levels are themselves dampened by viral silencing suppressors.

AGO7 (also known as ZIPPY), is associated almost exclusively with miR390 via mechanisms that require the integrity of the 5'-terminal adenosine and the central region of the miR390:miR390* duplex. MiR390-bound AGO7 triggers biogenesis of *TAS3* family tasiRNAs, which regulate AUXIN-RESPONSE FACTOR 3 (ARF3) and ARF4 (Axtell et al, 2006, Montgomery et al, 2008) to ensure proper juvenile-to-adult phase transition and adaxial-abaxial patterning. No biological role has been ascribed to AGO3 thus far.

The AGO4, -6, -8, and -9 clade

AGO4, -6, and -9 bind DCL3-dependent 24 nt siRNAs displaying a 5'-terminal adenosine bias (Mi et al, 2008). AGO4 is the major effector of RdDM and TGS of transposons and repeats. AGO4-mediated slicing is required for DNA methylation at some loci and dispensable at others (Qi et al, 2006). In addition, AGO4 loads functionally with rare DCL3-dependent 24 nt miRNAs to direct sequence-specific DNA methylation at the *MIRNA* loci of origin and may also functionally substitute AGO1 and -7 in tasiRNA biogenesis initiated by miR172 and miR390, respectively (Montgomery et al, 2008, Qi et al, 2006). AGO4, like AGO1, is widely expressed in most Arabidopsis tissues and associates with DCL3, 24 nt siRNAs and RDR2 into nuclear Cajal bodies (Li et al, 2006, Pontes et al, 2006).

The AGO2-miR168-AGO1-miR403 Loop

An interesting AGO2-miR168-AGO1-miR403 loop has been identified in *Arabidopsis*. It was found that AGO2 has the ability to bind with miR168 (Zhu et al, 2011), while miR403 was found in the database of AGO1 associated small RNAs [NCBI: GSE22252]; furthermore, both *ago1* and *ago2* were considered to play crucial roles in virus defence (Harvey et al, 2011; Morel et al, 2002; Diermann et al, 2010). These observations suggest that AGO1 and AGO2 might cooperate with miR168 and miR403 during virus infection. Loss-of-function of AGO2 did not show obvious defects in development (Harvey et al, 2011), demonstrating that AGO2 and miR403 might act in assistance to AGO1. It has also been suggested that the

overexpression of AGO1 would increase the risk of PTGS in endogenous genes. So the expression of AGO1 should be controlled under a self-check regulator (miR168) to maintain the perfect expression level of AGO1. But viruses have developed a series of molecules to crack this system for example, polerovirus F-box protein P0 degraded AGO1, P21 bound to miRNA/miRNA* and siRNA duplex to inhibit formation of active RISC and P19 had ability to increase level of the endogenous miR168 level to inhibit translational capacity of AGO1 mRNA (Bortolamiol et al, 2007; Chapman et al, 2004; Varallyay et al, 2010). Thus AGO2 could be considered as a secondary defence layer of plants, in case that virus cracked the first defence layer components AGO1. The relationship between miRNAs and AGO proteins is very complicated. MiRNAs can direct AGO proteins to repress the expression of target genes (Mi et al, 2008) and AGO proteins can stabilise the expression of miR168 and miR403 (Vaucheret et al, 2006), but miR168 and miR403 can reduce the protein levels of AGO1 and AGO2. This means that the increase of miR168 or miR403 would down-regulate AGO1 or AGO2, but reduction of AGO1 or AGO2 might decrease the expression of miR168 or miR403, which finally reduces increase of miR168 or miR403. This AGO2-miR168-AGO1-miR403 loop is vulnerable and tends to lose balance, so even slight change of any element in this loop would be amplified constantly. It is believed that transcriptional regulation of AGO1 and AGO2 by miR168 and miR403 and unknown regulatory factors help to keep the balance of this loop.

Transcription gene silencing pathway (TGS)

TGS occurs in the nucleus and functions to initiate and maintain the heterochromatic state of certain DNA regions. Transcripts, which are probably considered aberrant, are produced from heterochromatin or DNA repeats by RNA polymerase IV (Pol IV) in a process that is sometimes preceded by the function of Pol II. These transcripts are transcribed by RDR2 into dsRNA, which are diced by DCL3 into 24 nt long RNA duplexes with 2 nt 3'-overhangs. After methylation of the 2'-OH group by the dsRNA methyltransferase HEN1, the passenger strand is discarded/degraded while the guide strand binds to AGO4 or sometimes AGO6 or AGO9 depending on the loci and tissue (Zilberman et al, 2003; Zheng et al, 2007; Havecker et al, 2010). It then forms together with scaffold transcripts made by Pol

V (Wierzbicki et al, 2008), chromoproteins, histone H3K9 methylases (KYP) and DNA-methylating enzymes (Domain rearranged methylases [DRM2, DRM1] and Chromomethylase [CMT3]) a RNA-induced transcriptional silencing (RITS) complex involved in maintenance of histone- and DNA cytosine methylation (Matzke et al, 2009).

Recently, long-miRNAs (lmiRNAs), 24 nt in length, were identified in Arabidopsis and then in rice and were shown to be produced from pri-miRNAs from exactly the same position as canonical miRNAs, but by the action of DCL3 (Vazquez et al., 2008; Wu et al., 2010). Although the involvement of RDR2 and PollV remains to be explained and the implication of the other proteins of the TGS pathway to be tested, two recent studies show that long miRNAs trigger the methylation of their targets and in some cases of their own genes (Wu et al., 2010; Chellappan et al., 2010). The role of long-miRNAs is still elusive but given their strong expression in inflorescence tissues, one can speculate they act to silence *MIR* genes in meristematic cells and reproductive tissues.

Post-transcriptional gene silencing (PTGS)

Three major silencing pathways are used to control host genes involved in various functions such as development and stress responses, the miRNA pathway, the tasiRNA pathway and the nat-siRNA pathway.

MicroRNAs (miRNAs)

MiRNAs are small endogenous non-coding regulatory RNA sequences that have key roles in regulation of gene expression in most of the eukaryotic cells. In plants, miRNAs regulate gene expression at both transcriptional and post-transcriptional levels (Reinhart et al., 2002; Chapman et al., 2004; Ramachandran and Chen, 2008; Grant-Downton et al., 2009) and are involved in a number of physiological processes, such as growth, development and both biotic and abiotic stress responses (Mathieu et al, 2009; Wang et al, 2010; Grigorova et al, 2011; Thiebaut et al, 2012; Wu et al, 2013). It has been reported that the majority of miRNA sequences are deeply conserved and have near-perfect complementarities with their specific messenger RNA (mRNA) targets (Reinhart et al, 2002). Most

plant miRNAs are ~21 nt long and require DCL1 for their biogenesis and AGO1 their function. Many belong to multigene families that are sometimes conserved over long evolutionary distances (Cuperus et al, 2011, Rogers and Chen, 2013). Plant miRNAs often have narrow sets of target transcripts encoding transcription factors; stress-response proteins; or factors controlling cell identity, development, and growth.

MiRNAs originate from ssRNAs transcribed from *MIRNA* loci. *MIR* genes are usually transcribed by RNA polymerase II (Pol II), and form an imperfect fold-back structure known as primary miRNAs (pri-miRNAs) (Griffiths-Jones et al, 2008; Kim et al, 2011, Wang et al, 2013). Plant *MIRNA* loci are rarely nested within protein-coding genes. TATA boxes and *cis*-regulatory motifs are overrepresented in plant miRNA promoters, allowing their spatiotemporal or stress-responsive regulation by *trans*-acting factors and differential accumulation of individual miRNA isoforms. Pol II synthesizes pri-miRNAs from specific-non-protein coding *MIR* genes. Pri-miRNAs have typical Pol II cap structures at their 5' end and poly(A) tails at their 3' end, and often contain introns (Jones-Rhoades et al., 2006) (this is similar to protein coding transcripts). Due to their intermolecular sequence complementarity, pri-miRNAs adopt a fold-back stem-loop structure and thus miRNA biogenesis does not require an RDR. The pri-miRNA is processed into mature miRNA by Dicer-like 1 (DCL1) in *Arabidopsis* (Kurihara and Watanabe et al., 2004; Park et al., 2002; Reinhart et al., 2002). Accurate maturation and processing of pri-miRNA requires the additional activity of several proteins, including the Cap-binding protein 20 (CBP20) and CBP80/ABH1 (Gregory et al., 2008; Kim et al., 2008; Laubinger et al., 2008), the zinc finger protein Serrate (SE) (Lobbes et al., 2006; Yang et al., 2006), the dsRNA-binding protein/hyponastic leaves 1 (DRB1/HYL1) (Han et al., 2004; Vazquez et al., 2004), the forkhead-associated (FHA) domain containing protein Dawdle (DDL) (Yu et al., 2008), the Tough protein (TGH) (Ren et al., 2012, the proline-rich protein Sickle (SIC) (Zhan

et al., 2012) and the RNA binding protein Modifier of SNC1 2 (MOS2) (Wu et al., 2013).

The stem-loop structure contained within pri-miRNAs defines the pre-miRNA. Pre-miRNA processing is mediated by DCL1 assisted by the dsRNA-binding (DRB) proteins HYPONASTIC LEAVES 1 (HYL1) and SE (Laubinger et al, 2008; Vazquez et al, 2004), which in vitro binds the dsRNA section and ssRNA/dsRNA junctions of pri-miRNAs, respectively. The second dsRBD of HYL1 specifically associates with the DCL1 DUF283 domain, whereas both the N-terminal and zinc-finger domains of SE are required for DCL1-SE interaction. HYL1, SE and DCL1 congregate with pri-miRNA in dicing bodies. HYL1 and SE improve the efficiency and precision of DCL1-mediated cleavage (Dong et al, 2008, Laubinger et al, 2008, Vazquez et al, 2004). The RNA-binding proteins TOUGH (TGH) and MODIFIER OF SNC1 2 (MOS2) both effectively bind pri-miRNAs. TGH associates with DCL1, HYL1 and SE in the dicing bodies and MOS2 is uniformly nuclear. TGH appears to be integral to the complex modulating DCL1 activity, whereas MOS2 might, as an external cofactor, facilitate the recruitment of pri-miRNAs by this complex (Ren et al, 2012 and Wu et al, 2013). The forkhead-associated-domain protein DAWDLE (DDL) also binds pri-miRNAs (Yu et al, 2010). DDL interacts with DCL1 through a protein segment likely phosphorylated in vivo and might thereby promote the access to, or recognition of, pri-miRNAs by DCL1. Without DDL, the portion of pri-miRNA not properly channelled to DCL1 is probably degraded (Machida and Yuan, 2013). Ribosomal protein RECEPTOR FOR ACTIVATED KINASE 1 (RACK1), a direct and specific interactor of SE, localises in nuclear dicing bodies to modulate processing and transcription/stability of only some pri-miRNAs. C-TERMINAL DOMAIN PHOSPHATE-LIKE 1 (CPL1) is another direct interactor of SE. A model was proposed in which its interaction with SE recruits CPL1 to the DCL1 complex, where it might license HYL1 activity via dephosphorylation (Manavella et al, 2012). Once the pre-miRNA is stabilised in the dicing bodies DCL1 generates the mature miRNA/MIRNA* duplex. Structural determinants at least two staggered cleavage sites within the pre-miRNA stem, separated by approximately 21 nt, which releases the miRNA and its opposing fragment (miRNA*). Of key importance is the first cleavage position, which

determines the mature miRNA sequence and therefore its target specificity. The second cut usually proceeds at a fixed distance from the end of the precursor.

Upon miRNA/miRNA* release, the 3' ends of both strands are 2'-O-methylated by the RNA methyltransferase Hua Enhancer 1 (HEN1) (Kankel et al, 2003; Kasschau et al, 2007; Kasschau et al, 2003) and then exported to the cytoplasm by the exportin-5 homologue HASTY (HST) (Katiyar-Agarwal et al, 2006). Loss of HEN1 function incurs 3'-to-5' exonucleolysis (truncation) as well as tailing by HESO1, which adds 3'-oligouridylate tails to unmethylated miRNAs, leading to their degradation via mechanisms genetically distinct from those mediating truncation (Ren et al, 2012; Yu et al, 2005; Zhai et al, 2013, Zhao et al, 2012). Truncation and tailing both require AGO1 but not its slicing activity, and HESO1 colocalises with AGO1, suggesting that 3'modification of unmethylated miRNAs occur after AGO1 loading. Normal miRNA turnover in wild-type plants involves a family of SMALL RNA DEGRADING NUCLEASE (SDN) proteins with 3'-5' exoribonuclease activity capable of degrading 2'-O-methylated ssRNA (Ramachandran and Chen, 2008).

One strand of the miRNA duplex is subsequently incorporated into an RNA-induced silencing complex (RISC), which contains at least, an AGO protein. Plant miRNAs promote the cleavage of their target RNA, to which they bind perfectly or near-perfectly, by employing mostly AGO1 as the RNA slicer. Therefore, cleavage is assumed as the common approach for miRNA-mediated gene regulation in plants (Kim et al, 2011; Kumakura et al, 2009; Kurihara et al, 2004). However, in addition to regulating RNA degradation, miRNAs sometimes direct DNA methylation (Kwak and Tomari et al, 2012) or inhibit translation (Lau et al, 2012; Laubinger et al, 2008; Law et al, 2010; Law et al, 2013; Law et al, 2010; Li et al, 2006). As mentioned, most miRNAs associate the AGO1. However, specific associations with AGO2 (miR408, miR393*), AGO7 (miR390) and AGO 10 (miR165/miR166) have been reported.

Guide-strand selection in plant miRNA/miRNA* duplexes is directed in part by the lower thermodynamic stability of the guide strand 5' end relative to

that of the miRNA*. HYL1 and CPL1 also facilitate this process (Eamens et al, 2009; Manavella et al, 2012). Upon guide strand selection, the miRNA* is generally degraded, although some miRNA* might be stabilised and functional, including under stress conditions and upon their loading into AGO2 (Devers et al, 2011). AGO-miRNA complexes recognise target mRNAs via base complementarity and most plant miRNAs display extended target complementarity. 5'-RACE (rapid amplification of cDNA ends) and degradome analysis show that decreased target levels resulting from plant miRNA action correlates qualitatively with AGO-mediated slicing between paired positions 10-11. The cytoplasmic exosome and 5'-3' exoribonuclease XRN4 degrade the 5'- and 3'-cleavage products, respectively (German, 2008). Despite extended complementarity in most plant miRNA:target pairs, a fraction of AGO-targeted transcripts evade slicing and instead undergo protein-level repression (Brodersen et al, 2008; Brodersen et al, 2012).

Endogenous Inverted repeat-derived small interfering RNAs

Discrete loci scattered in plant genomes are configured as IRs that produce extensively base-paired RNA hairpins of variable lengths. Endogenous hairpins significantly longer than typical pre-miRNAs often generate endogenous siRNAs (Dunoyer et al, 2010; Dunoyer et al, 2007; Kasschau et al, 2007). A detailed study in Arabidopsis showed that all IRs generated active, HEN1-methylated siRNAs upon their coprocessing by DCL4/2 (21-22 nt siRNAs) and DCL3 (24 nt siRNAs). DCL1 also stimulated siRNA accumulation indirectly by facilitating separation of the dsRNA stem from the single-stranded section of IR transcripts, as in pri- to pre-miRNA maturation. IR-derived siRNA production required none of the factors involved in RDR-mediated dsRNA synthesis in the PTGS or TGS pathway, consistent with the intramolecular base pairing of IRs.

EndoIR-siRNAs derive from single-stranded hairpin precursors that are transcribed from different loci found throughout the Arabidopsis genome. These inverted repeats differ from *MIR* genes in terms of hairpin structure and size. Endogenous hairpins are much larger than typical miRNAs precursors and do not fit the criteria for annotation of miRNAs. Like pri-miRNA, EndoIR-siRNA precursors fold back to form molecules with perfect or near-perfect

complementarity, which likely makes them suitable substrates of DCL2, DCL3 and DCL4 instead of DCL1 (Dunoyer et al, 2010). Two such inverted-repeat derived hairpins, *IR71* and *IR2039,* produce 21, 22 and 24 nt through the action of DCL4, DCL2 and DCL3, respectively (Dunoyer et al, 2010). Interestingly, all size classes of small RNAs produced from *IR71* locus move within the plant throughout the vascular tissues. The 24 nt endoIR-siRNAs trigger cognate-specific de novo methylation at distance (Dunoyer et al, 2010). This data suggests that, unlike *MIR* genes, endogenous inverted repeats loci have the potential to trigger local and systemic RNA silencing (Dunoyer et al, 2010). EndoIR-siRNAs function is still unknown but it has been proposed that they could be used in adaption to the environment and also in *trans*-generational memory (Dunoyer et al., 2010). The actual response of EndoIR-siRNAs to environmental factors and contribution to stress-adaption is yet to be demonstrated and more evidence needs to be generated to establish whether these features are typical of a larger class of endogenous inverted-repeat-derived hairpin RNAs and whether they are shared by the potentially numerous shorter-hairpin RNAs.

TasiRNA Pathway

TAS RNAs are transcribed from specific genes, namely *TAS* genes. Three families of these genes are known, the TAS1/2 family made of four members, the TAS 3 family made of three members and TAS 4 family made of a single member (Vazquez et al, 2004; Allen et al, 2005; Peragine et al, 2004; Axtell et al, 2006; Howell et al, 2007). TAS RNAs are originally capped and polyadenylated but become processed by DCL4 to generate secondary siRNAs, termed tasiRNAs, upon cleavage guided by specific miRNAs (Vazquez et al, 2004; Allen et al, 2005; Peragine et al, 2004, Gasciolli et al, 2005; Xie et al, 2005; Montgomery et al, 2008; Hsieh et al, 2009). These miR:AGO complexes are thought to recruit SGS3, which binds dsRNA with 5'-overhang (Elkashef et al, 2009; Fukunaga et al, 2009) and RDR6 to convert TAS RNAs to double-strand form. It is shown that the secondary function of AGO1 depends on a 22 nt miRNA, which can be generated from bulged precursors in contrast to 21 nt one (Cuperus et al, 2010; Chen et al, 2010). It is speculated that the larger miRNA size leads to conformational change of AGO1, enabling it to fulfil this recruiting function (Schwab et al, 2010). For *TAS3* as an

alternative, a very specific AGO protein, AGO7 and targeting at two distant sites are required. The upper site is not sliced in contrast to the lower one, but from there the AGO7:miR390 complex serves to engage RDR6 (Montgomery et al., 2008; Allen et al., 2005). The TAS RNA duplexes are diced by DCL4 and its cofactor DRB4 in phase into 21 nt, and at specific minor sites 22 nt siRNAs (Montgomery et al., 2009; Hiraguri et al., 2005; Adenot et al, 2006; Nakazawa et al, 2007). TasiRNAs are known to trigger themselves a cascade of siRNA biogenesis able to regulate several members of the same gene family (Chen et al, 2007).

TasiRNAs generated from TAS1 and TAS2 mainly regulate the expression of pentatricopepetide mRNAs, those from TAS3 control auxin-response factor mRNAs to regulate abaxial-adaxial leaf polarity and phase change, and those from TAS4 regulate MYB transcription factor mRNAs to regulate anthocyanin biosynthesis in response to stress (Hsieh et al, 2009; Adenot et al, 2006; Chen et al, 2007; Garcia et al, 2006; Marin et al, 2010; Fahlgren et al, 2006; Luo et al, 2011).

NAT-siRNA Pathway

Nat-siRNAs are produced from overlapping dsRNA regions formed by natural antisense transcripts (NAT) and they define two classes: *cis*-nat-siRNAs, which arise from transcripts produced from the same genomic locus, and *trans*-nat-siRNAs, which arise from transcripts produced from physically distant genomic loci. *Cis*-nat-siRNAs are formed under specific stress or developmental conditions (Borsani et al, 2005; Katiyar-Agarwal et al, 2007; Katiyar-Agarwal et al, 2006; Wang et al, 2006; Zhou et al, 2009). Typically, one gene of the NAT pair is always expressed while the second one in induced (Borsani et al, 2005, Katiyar-Agarwal et al, 2006). The long dsRNA region formed by the pairing of the two transcripts is then processed into a single siRNA. The DCLs involved in the biogenesis of each of the nat-siRNA studied so far in *Arabidopsis* are different. The founding nat-siRNA gene pair yields a DCL2-dependent 24-nt siRNA important for tolerance to salt stress; the second one yields a DCL1-dependent 22-nt siRNA with a role in resistance to *P.syringae*; and the third one yields DCL1-DCL4-dependent

39-41 nt siRNAs, called long-siRNAs, important for resistance to bacterial pathogen (Borsani et al. 2005, Katiyar-Agarwal et al, 2006).

Plant viruses and the sRNAome

Antiviral silencing host factors

One of the first discovered and well-studied functions of RNA silencing is the host defence against invading viruses (Baulcombe, 2004). The hallmark of its adaptive antiviral function is the accumulation of virus-derived siRNAs (vsRNAs) at high levels during viral infection (Ruiz-Ferrer and Voinnet, 2009; Hamilton and Baulcombe, 1999; Molnar et al, 2005; Szittya et al, 2010; Pantaleo et al, 2007; Donaire et al, 2009; Qu, 2010). VsRNAs were found to be associated with AGO1, the slicer component of the plant RNA-induced silencing complex (RISC) effecter (Zhang et al, 2006; Csorba et al, 2010). As a counter defensive strategy, many plant viruses have viral suppressors of RNA silencing (VSRs) to counteract antiviral silencing (Ding and Voinnet, 2007; Voinnet et al, 1999; Burgyan, 2008), providing strong evidence for the antiviral nature of RNA silencing. In addition, the lack or inactivation of VSRs leads to the recovery of plants from viral infections, demonstrating the efficient antiviral response of the plant (Baulcombe, 2004, Ratcliff et al, 1997; Szittya et al, 2002).

Mechanism of silencing-based antiviral plant response

The pathway of antiviral silencing can be divided into three major steps:

(1) Sensing and processing viral RNAs to viral vsRNAs.

(2) Amplifying vsRNAs.

(3) Assembling antiviral RISC and targeting viral RNAs.

The silencing-based antiviral plant response starts with the recognition of ds or structured single-stranded (ss) viral RNA by one or more members of plant Dicers (Mlotshwa et al., 2008; Aliyari and Ding, 2009). Dicers then process the recognised viral RNAs into vsRNAs (Ding and Voinnet, 2007; Ruiz-Ferrer and Voinnet, 2009;

Mlotshwa et al, 2008; Molnar et al, 2005; Qu et al, 2010; Wang et al, 2010; Cuperus et al, 2010).

In plants, two distinct classes of vsRNAs have been identified: primary siRNAs, which result from the DCL-mediated cleavage of an initial trigger RNA, and secondary siRNAs, which requires an RDR enzyme for their biogenesis (Ruiz-Ferrer and Voinnet, 2009; Donaire et al, 2008; Qu, 2010; Wang et al, 2010; Wassenegger and Krezal, 2006; Vaistji and Jones, 2009; Garcia-Ruiz et al., 2010). In the *Arabidopsis* model plant, DCL4 and DCL2 are the most important DCLs involved in virus-induced RNA silencing and they can process ds or hairpin viral RNAs into vsRNAs of 21 and 22 nt, respectively (Ruiz-Ferrer and Voinnet, 2009; Wang et al., 2010; Deleris et al., 2006; Fusaro et al., 2006). The amplification and high level of vsRNA accumulation in many but not all virus infections depend on the combined activity of the host-encoded RDRs such as RDR1, RDR2 and RD6. Aberrant viral ssRNAs lacking quality control marks are converted by RDR enzymes to dsRNAs, which serve as a substrate for secondary vsRNA production (Garcia-Ruiz et al, 2010, Diaz-Pendon et al, 2007, Bao et al, 2009). The generated vsRNAs are loaded into distinct AGO-containing effector complexes to guide them to their RNA target molecules (Ding and Voinnet, 2007; Vaucheret, 2008; Hutvagner and Simard, 2008). In plants, the loading of siRNAs into a particular AGO complex is preferentially, but not exclusively, dictated by their 5' terminal nucleotides (Mi et al, 2008). It has been shown that both AGO1 and AGO7 function to ensure the efficient clearance of viral RNAs, and that AGO7 seems to work as a surrogate slicer in the absence of AGO1 (Qu et al, 2008). Moreover, it is probable that AGO1 is capable of targeting viral RNAs with more compact structures, whereas AGO7 favours less structured RNA targets (Qu et al, 2008).

Structure and Function of virus-derived siRNAs (vsRNAs)

The precursor for the stimulation of the RNA silencing machinery against plant viruses is the viral dsRNA, which is cleaved by Dicer to initiate RNA silencing process. There could be three diverse sources of dsRNA: (1) primary siRNA that is generated as an intermediate in genome replication of RNA viruses by the activity of virus-encoded RNA polymerases, or bi-directional transcription (DNA viruses)

or by convergent transcription; (2) structure associated siRNA, which are produced from the transcripts of viral genome having imperfectly based paired secondary structure; and (3) secondary siRNAs which arise from the ssRNA by the action of RDR gene family of hosts (Ahlquist 2006). The dsRNA molecules are formed as an intermediate during RNA virus replication, so they were assumed as a primary vsRNA to trigger silencing (Ahlquist, 2002). Secondary vsRNA production requires DCL processing coupled with RDR-mediated synthesis of longer complementary viral RNA.

The most abundant viral primary siRNAs are 24 nt followed by 21 and 22 nt (Chen, 2009). In this major event, three DCL proteins manage the viral siRNA production process. During the viral infection, DCL4 and DCL2 cleave the dsRNA into 21 nt siRNA, in contrast, DCL3 produces 24 nt viral siRNA which alone can confer viral resistance. The viral primary siRNAs produced are further incorporated in the AGO-containing RISC complex. These AGO proteins can cleave the target viral mRNA or can mediate translation repression or can cause transcriptional silencing. Out of the 10 AGO proteins in *Arabidopsis thaliana,* AGO1 and AGO7 have a role in viral resistance (Qu et al, 2008). Additionally, in infected cells, AGO2 and AGO5 have also been reported to bind the viral siRNA (Takeda et al, 2008). Contrastingly, AGO4 and AGO6 along with DCL3 and 24 nt viral siRNA are responsible for methylation of cytosine and histones of target DNA (Raja et al, 2008). The antiviral immunity involves the production of secondary viral siRNAs, which are responsible for the amplification of the efficacy of RNAi and they represent vast majority of small RNA. It requires RDR1, RDR2 and RDR6, which generate new viral dsRNA (where viral primary siRNA is the primer, and template is targeting viral mRNA). These RDR1/RDR6 processed dsRNA are the target for DCL which convert them into viral secondary siRNA, hence amplifying the viral resistance (Wang et al, 2010). During the characterisation of geminivirus-derived small RNAs, it was evidenced that 24 nt and a significant portion of the 22 nt viral siRNAs were produced by the Dicer-like proteins DCL3 and DCL2, respectively. Thus, dsRNA is processed by Dicer to generate vsRNA of aforementioned sizes. With help of NGS, it was predicted that the hot spots for generation of primary

virus-derived siRNA are the regions of viral genome that can fold into hairpin-like structures and can also act as substrate for Dicer.

Nature of siRNA directs the Defence

Size and population counts of siRNAs govern their specific role during plant-virus interactions. Generally, 21 and 22 nt vsRNAs are involved in PTGS while 24 nt vsRNAs are usually associated with DNA virus genome or associated histone methylation (Axtell, 2013). Geminiviruses, including the bipartite begomoviruses, are the frequent targets of RNA silencing. VsRNAs associated with *Cabbage leaf curl virus* and *African cassava mosaic virus* were elevated and characterised. In these studies, 21, 22 and 24 nt vsRNAs were detected from intergenic and coding regions of both of these geminiviruses genomes (Akbergenov et al, 2006). In a separate study in tomato, it was reported that a higher accumulation of 21 and 24 nt siRNAs corresponding to replication-associated proteins gene (Rep) region of the geminivirus, *Tomato leaf curl New Delhi virus* (ToLCNDV), confers tolerance against the virus (Sahu et al, 2010; Sahu et al, 2012). The 24 nt sized vsRNAs were associated with hypermethylation of the Rep (AC1) ORF (Sahu et al. 2014). Varied size siRNAs (24-26 nt and 26-28 nt) were derived from *Cucurbit leaf crumple virus* (CuLCrV) genome in both symptomatic and recovered leaves of watermelon plants (Hagen et al, 2008). Conflicting results in terms of vsRNA population counts in susceptible and resistant/tolerant plant tissues are reported in the literature. Arabidopsis, which is highly susceptible to CaLCuV, accumulated high counts of vsRNAs (Aregger et al, 2012). For example, in symptomatic watermelon tissues, the levels of the CuLCrV-derived siRNAs were found to be higher in comparison with the recovered tissues. Interestingly, the abundance of CuLCrV-derived small RNA was negatively correlated with recovery. This was dissimilar with the outcomes of earlier studies in several plant species, demonstrating that the higher level of small RNAs was negatively correlated with virus accumulation during recovery from begomovirus infection (Yadav and Chattopadhyay, 2011, Chellappan et al, 2004, Sahu et al, 2010; Sahu et al, 2012). Further they have proposed that the significant reduction in CuLCrV accumulation during recovery was due to decrease in viral-associated RNAs via RISC-mediated RNA degradation, which in turn reduces the levels of viral proteins, leading to a

reduction in viral replication. Thus, in recovered tissues, as the level of the viral-derived RNAs decreases, the CuLCrV-corresponding small RNAs are also simultaneously reduced. Another examination of recovery in pepper from infection of *Pepper golden mosaic virus* (PepGMV) also supported the findings as observed in case of the CuLCrV infection (Carrillo-Tripp et al, 2007). These contrasting results suggest a host-mediated, time-dependent action of recovery or resistance that may vary between plant species/cultivars and different virus interactions.

siRNA-mediated DNA Methylation

siRNA-mediated DNA methylation recently came into the limelight as a major defence pathway against various plant viruses. Several evidences have supported the hypothesis of siRNA-mediated methylation as one of the defence strategies against viruses. In such a study, it was shown that DNA viruses, *Tomato golden mosaic virus* (TGMV) and ACMV, exhibited altered replication when their DNA was methylated prior to transfection, exhibiting the role of siRNA-mediated methylation. Simultaneously, it was demonstrated that specific-sized siRNAs were produced against the two nuclear DNA viruses (geminivirus CaLCuV and pararetrovirus CaMV) and a cytoplasmic RNA tobamovirus involving all the four Dicers (Blevins et al., 2006). There is a considerable body of evidence that demonstrates that plants methylate geminivirus chromatin as an epigenetic defence. Early studies indicated that *in vitro* methylation of geminivirus DNA greatly impaired replication and transcription in protoplasts (Brough et al, 1992; Ermak et al, 1993). More recently, it has been shown that methylation-deficient *Arabidopsis* mutants are hypersusceptible to geminiviruses and that components for the RdDM pathway, e.g. AGO4, are necessary for host recovery from infection (Buchmann et al, 2009). AGO4 specifically plays a role in the production of siRNA that are 24nt long. In addition, geminivirus DNA and associated histones are methylated in infected plants, and viral DNA methylation is reduced in mutants that display enhanced disease. By contrast, the small amount of viral DNA present in recovered tissue is hypermethylated (Raja et al, 2008). These studies clearly

demonstrate that methylation, and likely TGS, acts as a defence against DNA viruses.

Actions of viral suppressors of RNA silencing

Plant viruses are efficient pathogens, which are able to infect and invade distinct plant species. They often cause severe symptoms and damage, which suggests an efficient counter defence strategy against the antiviral silencing response. The most common way to protect viral genome against RNA-silencing mediated inactivation is to encode proteins that act as suppressors of RNA silencing (viral suppressors of RNA silencing, VSRs). Many VSRs have been identified since the discovery of the first VSR more than a decade ago (Anadalakshmi et al, 1998; Brigneti et al, 1998; Kasschau and Carrington, 1998). The fact that most viruses have evolved VSRs underlines the antiviral nature of RNA silencing and reveals a pathogen counter defensive strategy with the active suppression of host surveillance (Ding and Voinnet, 2007; Voinnet et al, 1999; Silhavy and Burgyan, 2004). The VSRs are considered the outcome of recent evolutionary processes and they are surprisingly diverse within and across kingdoms, with no obvious sequence homology.

VSRs were shown to block virtually all steps of RNA silencing such as Dicing, effector assembly, targeting, amplification, transcriptional regulation of endogenous factors that control RNA silencing and its connections with protein-based immunity and hormone signalling. VSRs regulate the multiple layers of the complex defence, counter-defence and counter-counter defence arms race between host and pathogen. Although the study of VSRs was at the frontline of investigations for more than 10 years, many aspects of VSR' molecular behaviours are still elusive. It is becoming more obvious now that VSRs are not just simply blockers of RNA silencing but serve as central hub regulators to dynamically integrate connections between antiviral silencing, protein-based immunity, hormone signalling, RNA metabolism and subcellular organisations (Pumplin and Voinnet, 2013). As most suppressor proteins have parallel functions, the silencing function and the non-silencing activities (e.g. coat protein, movement protein,

replicase, protease etc.) need to be synchronised in order to fulfil these multiple tasks and achieve "optimal" infection.

Driving factors in VSRs' evolution

The high diversity in structure and function, the various position of their gene-code within the viral genome, the alternative expressional strategies like transcriptional read-through, leaky ribosomal scanning, proteolytic maturation and being often encoded by out-of-frame ORFs within conserved viral genes suggests that VSRs are of recent evolutionary origin (Ding and Voinnet, 2007). Therefore, in most cases, the suppressor function of VSRs may have evolved after the ancient role as replicase, coat protein, movement protein, protease, transcriptional regulator etc. or co-evolved with these to combine within the suppressor role and other essential roles important for viral life cycle. The different VSRs can inhibit all steps of the antiviral RNA silencing pathway, including cell-autonomous and non-cell autonomous aspects of it. Using mathematical modelling of dynamics of suppression has been shown that the different strategies employed result in slightly different outcomes regarding suppression of antiviral silencing. Suppressors targeting effector step are more potent at single cell level whereas siRNA binding is more effective at tissue level (Groenenboom and Hogeweg, 2012). Besides this however, an important driving factor in the suppressors' evolution was probably the availability of ancient/original viral protein activities that could be selected from with a minimum number of changes to acquire an additional suppressor features. Many viral proteins have RNA binding capacity (like replicase, coat protein, movement protein). This may explain why an overwhelming number of suppressors act through RNA binding. Silencing functions could have become established in the cases when the trade-off between its positive effect on viral life cycle and negative effects on host were worth it. The suppressors being too weak or too strong were out-selected through evolution.

Blocking initiation of antiviral response

One strategy used by VSRs is to hinder mounting of antiviral silencing by blocking the silencing initiation step. This can be achieved through multiple ways

like dicer protein or co-factor activity inhibition, dsRNA/siRNA-sequestration or AGO protein destabilisation prior of RISC assembly. A widespread suppressor strategy is the ds-siRNA sequestration that is used by several VSRs encoded by diverse virus genera (P19, HC-Pro, p15, p130/p126/p122, γβ, NS3, Pns10, NSs etc.) (Csorba et al, 2007; Harries et al, 2008; Hemmes et al, 2007; Kubota et al, 2003; Lakatos et al, 2006; Merai et al, 2005, 2006; Silhavy et al, 2002). Probably the most characterized siRNA binder is the tombusviral p19 protein (Silhavy et al, 2002). Crystallographic studies have shown that p19 head-to-tail homodimer acts as a molecular caliper to size-select and sequester siRNA duplexes in a sequence-independent manner (Silhavy et al, 2002; Vargason et al, 2003). A consequence of siRNA binding by VSRs is the blocking of HEN1-dependent methylation of sRNAs (Csorba et al, 2007; Lozsa et al, 2008; Vogler et al, 2007), however this also depends on the coexpression of sRNA and the suppressor (Lozsa et al, 2008). The AC4 component of geminiviruses, which is the least conserved protein, does not bind with siRNA or miRNA duplexes, but instead it competes against AGO1 to bind with single-stranded RNA and therefore suppresses the formation of siRNA-RISC assembly (Chellappan et al., 2004).

Begomovirus AC2 and AC4

Little is known about the activities of DNA virus silencing suppressors, but what is understood so far indicates that they are quite unique. The Baulcombe laboratory was the first to demonstrate that the geminivirus AC2 protein could reverse established PTGS (Voinnet et al, 1999). This protein was initially characterised as a transcription factor that stimulates the expression of virus late genes by both activation and depression mechanisms (Sunter and Bisaro, 1992, Sunter and Bisaro, 1997; Sunter and Bisaro, 2003; Lacatus and Sunter, 2008). AC2 has not been shown to bind siRNA or miRNA (Chellappan et al, 2005; Wang et al, 2005). Rather, AC2 ability to reverse PTGS and inhibit systemic spread requires an intact nuclear localisation signal (NLS), the central zinc finger-like domain (CCHC) motif, and the transcription activation domain. This suggests that silencing suppression requires AC2 to stimulation transcription. More specifically, AC2 activates the transcription of cellular genes that negatively regulate silencing pathways (Dong et al, 2003; Wezel et al, 2002; Trinks et al, 2005). This mode of

silencing suppression is termed transcription-dependent, although interactions with the silencing pathway components through the activation domain cannot be formally ruled out (Bisaro, 2006).

The hypothesis that AC2 can alter the host transcriptome comes from studies showing that geminivirus infection can activate the expression of reporter transgenes driven by the viral CP promoter in an AC2-dependent fashion (Hong et al, 1997; Sunter and Bisaro, 1997). Transcription profiling following transient expression of AC2 in Arabidopsis protoplasts identified several genes whose expression was upregulated. One of these genes was *Werner exonucleases-like 1 (WEL1)*, a homologue *of Werner syndrome-like exonucleases (WEX)* (Trinks et al, 2005). Although its role is unclear, WEX is required for PTGS (but not TGS) directed against transgenes (Glazov et al, 2003). It has also been proposed that WEL1 over-expression might compete for factors needed for WEX function (Trinks et al, 2005). Evidence for transcription-independent silencing suppression by AC2 and the related C2 protein from the Curtovirus BCTV has also been gained (Wang et al, 2005). AC2 and C2 share pathogenic functions, and both interact with and inactivate SNF1-related kinase (SnRK1) and adenosine kinase (ADK) (Baliji et al, 2007; Hao et al, 2003; Sunter et al, 2001; Wang et al, 2003; 46,). The SnRK1 interactions inhibit the cellular stress response, which appears to be a component of plant basal defences (Hao et al, 2003).

AC2 and C2 inactivate ADK *in vitro* and when they are co-expressed with ADK in *E.coli* and yeast. Also, ADK activity is reduced in and AC2/C2-dependent manner in geminivirus-infected plant tissue (Wang et al, 2003). A link between ADK activity and silencing comes from the observation that ADK is needed to maintain the methyl cycle responsible for generating SAM, a methyl donor and essential methyltransferase co-factor (Lecoq et al, 2001; Moffatt et al, 2002; Weretilnyk et al, 2001). This is likely the reason why ADK-deficient mutant plants display silencing defects (Moffatt et al, 2002). A connection between methylation and PTGS was established by early observations that post-transcriptional silencing is commonly associated with methylation of coding regions of targeted genes, and particularly reporter transgenes (Ingelbrecht et al, 1994; Jones et al,

1998; Smith et al, 1994). Such methylation can interfere with gene expression (Hohn et al, 1996), possibly by promoting the production of aberrant transcripts that could serve as substrates for dsRNA synthesis and thereby amplify siRNA production. In summary, begomovirus AC2 proteins appear to supress silencing by activating the expression of cellular genes and by inhibiting ADK, whereas Curtovirus C2 is limited to inhibiting ADK (Bisaro et al, 2006).

The AC4 protein encoded by begomoviruses has also been demonstrated to have suppressor activity. The *AC4* gene is embedded within *AC1*, but in a different reading frame. AC4 is among the least conserved geminivirus proteins, despite the highly conserved nature of AC1. The PTGS suppression activities of both AC4 and AC2 from four different cassava-infecting begomoviruses were initially examined in transient assays using *N. benthamiana* leaves (Vanitharani et 2004). The two AC4 proteins from the viruses that elicit recovery-type symptoms (severe disease followed by host recovery), had suppressor activity in the assay. The remaining two from non-recovery-type viruses (which elicit disease from which host plants do not recover) had little or no activity. Conversely, the AC2 proteins of non-recovery viruses were effective suppressors, while the AC2 proteins of recovery-type viruses were not. This study firstly demonstrated that geminiviruses could encode more than suppressor. Secondly it showed that similar proteins from different viruses do not necessarily have equivalent suppressor activities. Thirdly, it suggested that AC2 and AC4 proteins primarily affect different silencing pathways, providing a mechanistic basis for observed severe disease synergy sometimes observed in the field. Severe disease can result when mixed infections, or recombination, combining AC2 and AC4 suppressors. Lastly, the different disease phenotypes induced by this panel of viruses suggests that while the counterdefence effects of AC4 are somewhat transient, an effective AC2 suppressor is associated with the absence of host recovery (Vanitharani et al, 2004, Vanitharani et al, 2005). This is consistent with genetic data linking AC2/C2 function with recovery (Hormuzdi et al, 1995, Raja et al, 2008).

Insight into the AC4 suppression mechanism has been gained from the finding that the silencing-active protein associates with single-stranded, but not

double-stranded, siRNAs and miRNAs *in vitro* and *in vivo* (Chellappan et al, 2005). Thus AC4 is the only known protein that is apparently able to suppress PTGS, and coincidentally the miRNA pathway, by binding small RNAs in single-stranded form. This suggests AC4 acts downstream of small RNA biogenesis and unwinding, and implies that RISC loading and unwinding are not necessarily couples in plants. Presumably single-stranded small RNAs are accessible at some point between these events. Alternatively, small ssRNA may be recycled following release from AGO complexes. In any case, before more can be said about the mechanism it will be important to carefully define the affinity of AC4 for RNAs of varying structures and sizes and to determine whether the protein is able to interfere with RISC loading or disrupt previously formed RISC complexes.

Viral suppressors inhibiting viral RNA sensing and dicing

Inhibition of viral RNA recognition and the subsequent dicing by plant Dicer effectors is not a frequent strategy of known VSRs. Two viral proteins have been identified that were shown to inhibit the processing of dsRNA to siRNA in agroinfiltration assays. P14 of *Pothos latent aureusvirus* and P38 of *Turnip crinkle virus* (TCV). In addition, P38 and P14 have been shown to bind dsRNA in a size-independent way (Merai et al, 2006; Azevedo et al, 2010). P38 has been shown to specifically inhibit DCL4 activity, which has been shown to be the primary antiviral Dicer in the Arabidopsis model plant (Deleris et al, 2006). Recently, it was discovered that the action of the P38 protein occurs through AGO1 binding and that it interferes with the AGO1-dependent homeostatic network, which leads to the inhibition of Arabidopsis DCLs (Azevedo et al, 2010). In addition to P14 and P38, the P6 VSR of the *Cauliflower mosaic virus* (CaMV) (Love et al, 2007) has been shown to interfere with vsRNA processing. A recent discovery showed that one of the nuclear functions of P6 is to suppress RNA silencing by interacting with dsRNA-binding protein 4, which is required for the functioning of DCL4 (Haas et al, 2008).

Viral suppressors preventing RISC assembly

VSRs are able to prevent RISC assembly by targeting one of its essential known or unknown components. The VSRs identified thus far are able to target

siRNAs and miRNAs or AGO proteins in different ways. The most common suppression strategy, evolved by several viral genera, is ds siRNA sequestration (Ding and Voinnet, 2007; Merai et al, 2006; Lakatos et al, 2006; Csorba et al, 2007; Csorba et al, 2009; Wu et al, 2010), which prevents the assembly of the RISC effector. Importantly, these siRNA-binding VSRs are completely unrelated proteins, although they share analogous biochemical properties, suggesting their independent evolution in different viruses. The P19 protein of tombusviruses, probably the best known VSR thus far, prevents RNA silencing by siRNA sequestration through binding dsRNA with a high affinity (Silhavy et al, 2002). Crystallographic studies have shown that P19 forms a tail-to-tail homodimer, which acts like a molecular calliper, measuring the length of siRNA duplexes and binding them in a sequence-independent way, selecting for the 19 bp long dsRNA region of the typical siRNA (Vargason et al, 2003; Ye et al, 2003). Thus, the P19 VSR evolved to bind and inactivate vsRNAs, which are the most conserved key elements of the RNA-silencing pathway. Recent findings have also demonstrated that P19 inhibits the spread of the ds siRNA duplex identified as the signal of RNA silencing (Dunoyer et al., 2010).

Other VSRs, such as the *Tomato aspermy cucumovirus* 2b protein also bind dsRNA in a size-specific manner; however, structural studies have shown that their modes of binding siRNAs do not share any similarity with P19 (Chen et al, 2008; Chao et al, 2005). Two siRNA binding VSRs (HC-Pro and P38) require the RAV2 transcription factor for the suppression of RNA silencing, although the mechanistic role of this plant cofactor is unclear (Endres et al, 2010). The 2'-O methylation step is essential in the biogenesis of miRNAs and siRNAs (Yu et al, 2005), and the siRNA-binding VSRs (*Carnation Italian ringspot virus* P19, *Tobacco etch virus* HC-Pro, *Tobamovirus* P122/P130) also compromise this step by preventing si/miRNA RISC assembly (Csorba et al, 2007, Ebhardt et al, 2005; Vogler et al, 2007; Lozsa et al, 2008; Yu et al, 2006). It is probable that these siRNA-binding VSRs have a higher affinity to siRNA and miRNAs than to HEN1 methyltransferase. However, the inhibition of the methylation step also requires

the temporal and spatial coexpression of the suppressor, endogenous or viral siRNAs and miRNAs (Lozsa et al, 2008).

The VSR of *Potato chlorotic stunt crinivirus* (SPCSV) uses a completely different strategy to prevent RISC assembly. The SPCSV-encoded Rnase3 endonuclease cleaves 21, 22 and 24 vsRNAs into 14 bp products, which are inactive in the RNA-silencing pathways (Cuellar et al, 2009). In the presence of siRNA-binding/targeting VSRs, plants are not able to confine the spread of the viral infection because vsRNAs are sequestered and inactivated before they can be incorporated into the RISC. According to the model suggested previously (Havelda et al, 2003; Havelda et al, 2005) in the absence of siRNA-binding VSRs, virus-specific vsRNAs act as a systemic signal, moving faster than the virus in the infected plant and thereby establishing antiviral silencing in cells ahead of the infection front. Thus, the RISCs already activated by vsRNAs destroy the entering viral RNA, resulting in the fast recovery of the plants (Szittya et al, 2002; Havelda et al, 2003; Havelda et al, 2005). Indeed, siRNA duplexes, as opposed to their precursor molecules, act as mobile silencing signals between plant cells (Dunoyer et al, 2010; Molnar et al, 2010).

Arrest of functional RISC assembly through AGO interaction/ AGO protein targeting VSRs

The arrest in the assembly of a functional RISC can be carried out also through direct binding the protein component of minimal RISC, AGO protein. The prevention of RISC assembly could also occur through direct or indirect interactions between VSRs and the protein components of RISC. The 2b protein of *Cucumber mosaic virus* (CMV) was one of the first described VSRs (Brigneti et al, 1998), and it prevents the spread of the long-range silencing signal facilitating systemic virus infection (Guo and Ding, 2002). The 2b protein of Fny-CMV has been found to physically interact with the PAZ domain and part of the PIWI-domain of AGO1 protein in the nucleus of the cell and also in the cytoplasmic foci (Mayers et al, 2000). A crystallographic study showed that the 2b protein of Tomato aspermy virus (TAV), a cucumovirus related to CMV, binds siRNA duplexes (Chen et al, 2008). The analysis of the crystal structure of TAV-2b-siRNA

showed that 2b adopts an alpha-helix structure to form a homodimer and binds to siRNA by measuring its length. The 2b protein is also known to bind long dsRNA (Goto et al, 2007) and to inhibit the production of viral secondary siRNAs (Diaz-Pendon et al, 2007). Thus, cucumovirus 2b proteins have a dual mode of silencing inhibition, either by sequestering siRNAs or by interacting with AGO1 and preventing RISC assembly.

The P0 protein of the phloem-limited poleroviruses also targets the AGO protein, the core component of the RISC and induces its degradation (Pazhouhandeh et al, 2006; Bortolamiol et al, 2007; Baumberger et al, 2007). P0 has no RNA-binding activity (Zhang et al, 2006; Csorba et al, 2010). Instead, it interacts with the SCF family of E3-ligase S-phase Kinase-related protein-1 components, orthologous to Arabidopsis ASK1 and ASK2, by means of its minimal F-box motif and thereby promotes AGO degradation (Pazhouhandeh et al, 2006; Bortolamiol et al, 2007; Baumberger et al, 2007). It has been found that P0 cannot interfere with the slicer activity of preprogramed siRNA/miRNA containing AGO1, but can prevent the *de novo* formation of siRNA/miRNA-loaded AGO1 (Csorba et al, 2010).

Inactivation of programed antiviral RISC complex

VSRs may mimic cellular protein cofactors to inactivate programmed RISC The conserved GW/WG-motif containing protein family (GW182 family) has been shown to bind to AGOs and to be required for diverse RISC function (Eulalio et al, 2009). Silencing effector complex activity block can be achieved also through targeting holo-RISC's RNA component, the guide RNA. *African cassava mosaic virus* (ACMV) encoded AC4 was shown to bind to the ss-sRNAs but not dsRNA forms *in vitro.* Transgenic expression of AC4 correlated with decreased accumulation of miRNAs and up regulation of target mRNAs.

AC4 acts downstream of the unwinding process: to bind mature miRNAs presumably loaded into AGO protein (Chellappan et al, 2005; Xiong et al, 2009; Zhou et al, 2006). In a study of *Sweet potato mild mottle virus* (SPMMV), the silencing suppressor P1 was shown to interact with AGO1 through its N-terminal

where three GW/WG motifs are present, which hinders RISC activity (Singh et al, 2010). P0 protein of polerovirus constitutes an F-box domain and hence associates with SCF complex of E3 ligase. It further interacts with AGO1 and causes its ubiquitin-dependent proteolysis, hence inhibiting its slicing activity (Bortolamiol et al, 2007). Likewise, AC2 protein of Geminiviruses interferes with the SCF-mediated ubiquitination and usurps with the host cellular machinery to ultimately generate a suitable environment for their function (Lozano-Duran and Bejarano, 2011).

Modulation of AGO1 homeostasis

During tombusviral infection AGO1 transcription is induced as part of the host antiviral arsenal. AGO1 homeostasis in plants depends on the miR168-guided AGO1 mRNA cleavage and translational inhibition (Rhoades et al, 2002). To counteract AGO1-based defence, the virus promotes miR168 transcriptional induction that results in miR168-guided AGO1 down-regulation. The miR168 accumulation spatially correlates with the virus localisation and depends on its p19 VSR (Varllyay et al, 2010). Similarly, to p19 all VSRs, which are very heterogeneous in protein sequence but bind vsRNAs, promote miR168 transcriptional induction and AGO1 down-regulation suggesting that VSR-siRNA complexes are effectors and recognised by the plant surveillance system (Varallyay and Havelda, 2013). A consequence of AGO1 protein deficiency in virus-infected plants can be the misregulation of miRNA targets, resulting in disturbed gene expression, which can lead to the development of viral symptoms.

Plant RDR-based activity suppression

Host RDRs (RDR1, 2 and 6) contribute to amplification of RNA silencing and spread of a systemic signal by synthesis of vsRNAs (Schwach et al, 2005). Interestingly, plant RDR1 itself was suggested to have adverse functions. RDR1 is an antagonist of RDR6-mediated sense-PTGS silencing therefore behaves as an endogenous silencing suppressor (Ying et al, 2010). Suppression of RDR activities may constitute a target point for VSRs since it dampens cell-autonomous silencing

amplification and systemic movement in distant tissues to facilitate the virus replication and spread.

VSR interactions with host factors

There are emerging evidences that besides the "canonical" block of RNA silencing (through ds-, si-, mi-ssRNA-binding e.g. p19, Rnase3 etc. or manipulating silencing-related protein activities via direct/indirect interactions e.g. P0, V2, P1 etc.) some suppressors may target endogenous regulators of the silencing to modulate host defence.

Plants utilise RdDM as a defence against DNA viruses for recovery, and conversely, viruses are equipped with proteins that suppress TGS. The expression of P6 protein in Cauliflower mosaic virus (CaMV) directly affects TGS as it inactivates the nuclear protein DRB4 (double-stranded RNA binding protein), which is critically required for the functioning of DCL4 (Haas et al, 2008). Contrastingly, AC2 protein of Begomoviruses and AL2 of Curtovirus suppress TGS indirectly. AC2 is transcription activator protein (TrAP), which increases the transcription of any silencing suppressor gene, whereas AL2 has been shown to interact and inhibit adenosine kinase (ADK), which is required for the synthesis of S-adenosyl methionine (SAM), a cofactor of all methyltransferase. This inhibition protects viral DNA from methylation (Raja et al, 2008).

VSRs interfering with the epigenetic modification of the viral genome

Suppressors from the Geminiviridae family modulate endogenous biochemical pathways for the benefit of viruses**.** The TGMV-encoded AL2 protein and the closely related *Beet curly top virus* (BCTV) L2 interact with and inactivate adenosine kinase (ADK), a cellular enzyme important for adenosine salvage and the methyl cycle**.** ADK plays a role in sustaining the methyl cycle**.** By inhibiting ADK, the AL2 and L2 proteins indirectly block this cycle and thereby could interfere with the epigenetic modification of the viral genome (Bisaro, 2006, Wang et al, 2005)**.** Evidence for the transcription-dependent activity of *Mugbean yellow mosaic virus* and *African cassava mosaic virus* protein AC2 has also been found**.** This suggests that silencing suppression and transcription activation by AC2 are

functionally connected and that come of the AC2-inducible host genes can code for components of an endogenous network that controls silencing (Trinks et al, 2005).

Viral RNA replication-mediated silencing suppression

Host factors involved in both RNA silencing suppression and viral replication have been proposed as playing roles in RNA silencing suppression during infection by the Red clover necrotic mosaic virus (RCNMV). The putative host factor involved in both processes could be the DCL1 protein because miRNA biogenesis is inhibited by virus replication and dcl1 mutant plants show reduced susceptibility to RCNMV infection (Takeda et al., 2005). In the suggested scenario, DCL1 and its homologues are recruited by the viral replication complex ad are, therefore, depleted from the silencing pathways.

Side effects of VSRs

Many VSRs have been identified as the pathogenic determinants largely responsible for virus-induced symptoms (Voinnet, 2005). It is well established that the antiviral and endogenous silencing pathways share common elements, and VSRs have been shown to interfere with these pathways. SiRNAs-binding VSRs (e.g. HC-Pro and P122) can interact with siRNA and miRNA biogenesis (Csorba et al, 2007; Lozsa et al, 2008; Kasschau et al, 2003; Chapman et al, 2004; Akbergenov et al, 2006) and can compromise these sRNA-regulated plant gene expressions. Similarly, long dsRNA-binding VSRs (e.g. P38 and P14) can compromise the activity of DCLs, and AGO1-targeting VSRs (e.g. 2b, P0, P1 and P38) inhibits RISCS, which in turn can alter expression of an unpredicted number of genes in plant development. A surprising effect of 2b VSR has been demonstrated recently. It has been shown that the 2b protein of CMV facilitates

epigenetic modification through the transport of siRNA to the nucleus (Kanazawa et al, 2011).

Control of pathogen impact on the host

Antiviral and endogenous silencing pathways share common elements. The ability of viruses to block antiviral silencing may have an impact on endogenous silencing pathways that results in alteration in short RNAs expression profile/activity and changes in gene expression both in a direct and in an indirect manner. VsRNA-binding VSRs can bind endogenous si- and miRNAs that could result in alteration of their downstream targets as was previously shown (Chapman et al, 2004; Kasschau et al, 2003; Lozsa et al, 2008). In case of miRNAs that target RNA silencing target components an unpredicted number of genes will be altered indirectly (e.g. miR162-mediated DCL1 negative feedback loop, DCL1-dependent suppression of DCL3 and DCL4, miR168 and AGO1 mRNA-derived siRNA control of AGO1, miR403 control of AGO2) (Allen et al, 2005; Mallory and Vaucheret, 2009; Qu et al, 2008; Rajagopalan et al, 2006; Vaucheret et al, 2006; Xie et al, 2005). This is a similar situation in the case of AGO-targeting VSRs (P0, P1, P38) (Azevedo et al, 2010; Baumberger et al, 2007; Derrien et al, 2012; Giner et al, 2010). VSRs' presence therefore may have a big impact and result in an altered developmental program of host organism and symptom development.

In support of VSRs as contributors to the viral symptoms, VSR-transgenic lines were created and analysed. In many cases the VSR-expressing transgenic plants display phenotypes similar to viral infections. (Dunoyer et al, 2004; Jay et al, 2011; Kasschau et al, 2003; Lewsey et al, 2007; Zhang et al, 2006). However, transgenic expression of VSR does not recapitulate the expression pattern in time and space of an authentic viral infection, therefore conclusions need to be drawn very carefully.

VSRs as links between RNA-based and protein-based immunity

Alteration of silencing pathways (an RNA-based immunity) in the presence of the VSR and/or viral infection triggers the protein-based immunity in host as part of the counter-counter defence response. R genes present in the plant genome

convey disease resistance against pathogens by producing R proteins and their actions are the main component of the protein-based immunity arsenal. Conserved miRNA family controls a plethora of R genes (Li et al., 2012; Shivaprasad et al., 2012; Zhai et al., 2011). It is assumed that R genes are silenced in the absence of the pathogen in order to minimise the cost for the plants and prevent autoimmunity reactions (Tian et al, 2003). It was found that the NBS-LRR genes (the main class of R proteins with nucleotide binding site (NBS) and leucine-rich repeat (LRR) motifs are silenced in a siRNA-regulated cascade similarly to tasiRNA biogenesis scheme: RDR6-dependent secondary siRNAs are produced following the original 22 nt miRNA-mediated cleavage on a R-gene transcript. The secondary siRNA may target other R-genes. When tomato plants were infected with viruses (TCV, CMV, TRV) accumulation of miR482 was reduced. In the absence of miR482 activity the resistance R gene targets get released and consequently R gene products accumulate to enhance immunity of the plants (Shivaprasad et al, 2012).

In summary miRNA-regulated R genes participate in a non-race immunity mechanism where the miRNAs are the sensors of the infection. It is supposed that release of R-gene based defence may be the cause of the inhibitory action of pathogen-encoded suppressors of silencing (VSRs) on miRNA activity during infection, however this assumption needs to be experimentally tested in the future.

Connecting antiviral silencing to hormone signalling

Several studies have shown that antiviral silencing might be connected to signal transduction pathways responsible for induction of SA-mediated resistance (Alamillo et al, 2006; Ji and Ding, 2001). SA is a plant hormone that is involved in local and systemic antiviral defence responses including SAR. SA induces expression of key antiviral silencing factor RDR1 (Liao et al, 2013; Xie et al, 2001). In turn, RDR1 affects many JA-regulated genes (Pandey et al, 2008). JA has been implicated as a defence-related hormone (Lewsey et al, 2010). VSRs seem to interfere with hormone signalling-based responses, although the precise mechanisms are elusive. VSRs therefore emerge as regulators of hormone-based

signalling to create favourable conditions for the virus. Although at the moment the complex interplay between the RNA silencing and SA-mediated defence is elusive VSRs might be important coordinators of this crosstalk during infection.

Various miRNAs and siRNAs have been implicated in innate immunity (Katiyar-Agarwal and Jin, 2010). For example, miR398 targeting superoxide dismutases is downregulated by ROS and plants overexpressing miR398 exhibit enhanced susceptibility to *Pseudomonas syringae* (Li et al, 2010). Emerging evidence implicates components of the nuclear silencing machinery in innate immunity. Most miRNAs are bound to AGO1 and this is a potential target of pathogen effectors. Interestingly, silencing suppressors of some RNA viruses target AGO1. Since AGO1, AGO2 and AGO7 mediate antiviral defence, it is conceivable that, in addition to their interaction with viral siRNAs, these AGOs contribute to anti-viral defence through endogenous miRNAs and siRNAs regulating PTI and ETI. AGO4, in addition to its main function in RdRM, appears to have a distinct function in ETI in *N. benthamiana* (Bhattacharjee et al., 2009). RDR1 is known to be involved in the production of secondary viral siRNAs (Wang et al., 2010) and is induced by SA. This implicates innate immunity signalling in silencing-based antiviral defence. Conversely, RDR6 mediates biogenesis of the endogenous siRNAs induced by bacterial effectors (Katiyar-Agarwal and Jin, 2010). Recent studies reveal that a large proportion of NB-LRRs are associated with RDR6-dependent secondary siRNAs and can be partially suppressed by RNA viruses and virulent P. syringae (Shivaprasad et al., 2012). Thus, plants seem to exploit the pathogen effector activities to achieve inducible expression of NB-LRRs.

1.6 Identification of microRNAs

Several miRNAs from diverse plants have been discovered in recent years In the past 11 years, the total number of registered miRNAs in miRBase has increased from 28 (release 3.0) to 8524 (in the current release version 21) (http://www.mirbase.org/). Accurate prediction and validation of miRNA target genes are important for unravelling the function of specific miRNAs. The three most commonly used methods to identify and validate plant miRNAs include

computational predictions based on conserved sequence and secondary structure without experimental verification, cloning of small RNA libraries and direct capture of miRNAs by high-throughput sequencing. The development of several new advanced and efficient technologies such as high-throughput sequencing has contributed to the marked increase in a number of newly discovered miRNAs. The key features of miRNAs remain unknown and most studies are confined to the model plants or some important plants. It is important to identify and characterise miRNAs to obtain insight into their functions.

Forward genetics is an advanced approach for discovery of miRNAs; however, it is time-consuming and expensive and thus has limited applications. This approach is unlikely to be a main contributor the list of biologically functional miRNAs. In reverse genetics, researchers use known sequences to discover function or phenotypes. The two main reverse genetic strategies used in the identification of miRNAs include bioinformatics and experimental approaches. Discovery of miRNAs through bioinformatic tools has become a widely used method and has been used to predict new miRNAs in both animals and plants. The success is mainly attributed to the low cost, high efficiency, speed and versatility of bioinformatics. The key principle behind miRNA identification using bioinformatics involves the establishment of homologous sequences of known miRNAs both within a single genome and across the genomes of related organism (Lagos-Quintana et al, 2001). The ensuring sequence and structure homologies provide basis for miRNA prediction on the basis of predefined parameters. Computational strategies have provided a reliable and efficient method to predict miRNAs and their target genes and have been used in studies on animals, fungi and higher plants (Rhoades et al, 2002; Bonnet et al, 2004; Adai et al, 2005; Zhang et al, 2005). Cloning and sequencing of small RNA libraries are the current experimental approaches used to identify and characterise miRNAs. Some of the limitations in this approach include tissue and time specificity of miRNA expression as well as the generally low expression levels of miRNAs. MiRNAs are mostly transcribed in response to specific environmental stimuli and they cleave or degrade target mRNAs, thus creating difficulties in cloning and miRNAs and other small RNAs. In this respect, the computational approach offers advantage. Next-generation massive sequencing techniques, such as 454 pyrosequencing and

illumina/Solexa, have been commonly used to identify new miRNAs in plants (Moxon et al, 2008; Song et al, 2010; Wang et al, 2011).

Computational Approaches to miRNA identification

It has been demonstrated that a majority of known miRNAs in the plant kingdom are evolutionary conserved, from mosses and ferns to higher flowering plants (Zhang et al, 2006a), and miRNAs from one species may have homologs or orthologs in other spaces. This observation offers certain practical and powerful strategies to identify novel miRNAs in different plants. Several computational approaches have been developed to identify plants miRNAs (Reinhart et al, 2002; Wang et al, 2004; 2005). Computational approaches have been used to identify miRNAs in: Arabidopsis (Wang et al, 2004), Soybean (Zhang et al., 2008a), Rice (Li et al, 2005), Maize (Zhang et al, 2006b), Tomato (Yin et al, 2008, Zhang et al, 2008b), Grape (Carra et al, 2009), and some other plants (Zhang et al, 2005; 2007; Sunkar and Jagadeeswaran, 2008).

Some miRNA features provide key information to predict novel miRNA sequences. Because of the presence of a characteristic fold-back structure (Berezikov et al, 2006), prediction of miRNAs on the basis of secondary structure of the sequence is applied in almost all approaches. Many approaches also rely on the phylogenetic conservation of both sequence and structure to distinguish between miRNA candidates and irrelevant genomic hairpins (Berezikov et al, 2006). It is well known that miRNAs are conserved in plants; thus, it is possible to computationally search for the homologs or orthologs of miRNAs as well as long hairpin structures in precursors (Wang et al, 2005). Minimum free energy (MFE), a commonly used measure for characterising the secondary structure of different RNAs, is also effective to characterise and/or predict miRNA sequences (Lee et al, 1993; Llave et al, 2002; Reinhart et al, 2002; Thakur et al, 2011). Bioinformatics tools can identify miRNAs using both sequence and secondary structure alignments (Wang et al, 2005). Because conserved miRNA sequences are commonly searched for using algorithms, the major challenge is to find miRNAs that are species specific. The major limitation in most of the bioinformatics techniques is the need to start from a known sequence and the dependence on

conserved secondary structure regions and mature miRNA sequences (Unver et al, 2009).

Analysis of expressed sequence tag databases in miRNA prediction

ESTs, a database of the complementary DNA (cDNA) sequence of the expressed genes, provide a platform for the above-mentioned bioinformatics tools to search for potential mRNA sequences. ESTs are partial sequences of cDNA cloned into plasmid vectors (Adams et al, 1991). Several plant genes have been cloned from sequences in ES databases (Graham et al, 2004). The fact that most miRNAs are deeply conserved from species gives researchers the ability to predict orthologs of previously known miRNAs from EST databases. With the increasing number of plant genome sequences, the number of ESTs in the database has markedly increased. As of 2012, GenBank (http://www.ncbi.nih.gov/Genbank/) had 154130210 EST sequence entries, representing more than 1370 different organisms. Conserved candidate miRNAs and their precursors can be predicted using this resource. EST analysis to predict homologous miRNAs across plant species has been developed using conserved sequence regions from previously known miRNAs. Additional parameters, such as structure prediction filters (secondary structure), have been applied to increase the accuracy (Zhang et al, 2005).

Experimental discovery of candidate miRNAs

The experimental validation of the predicted miRNAs is required to accurately determine their cellular functions. Repertoires of experimental methods are currently available to validate plant miRNAs. These approaches can be divided into two main groups: PCR-based cloning approaches and hybridization-based methods.

PCR-based cloning approaches

Direct cloning and sequencing of small RNA libraries

Direct cloning of small RNAs from plants is one of the earliest basic approaches used for identification of miRNAs. Many plant species have been cloned by this method: Arabidopsis (Llave et al, 2002; Reinhart et al, 2002), Rice

(Sunkar et al, 2005) and Grape (Carra et al, 2009). Direct cloning approach mainly involves synthesis of a cDNA library and includes six key procedures: Total RNA is extracted from the organism of interest; 26 to 28 nt RNAs are selected from the total RNA and excised from the polyacrylamide gel; small RNAs are then ligated with an adapter; subsequently reverse transcribed; amplification of resulting cDNAs with real-time PCR (RT-PCR) using primers specific for the adaptor sites; the RT-PCR product are cloned; and the selected clones are sequences and the sequence data is analysed. This method is not consistent with the prediction and can also identify sequences from different members of the same family.

Reverse transcription-polymerase chain reaction and Real-time analysis

RT-PCR is widely used to detect the expression of mRNA and other RNA molecules. Real-time quantitative RT-PCR (qRT-PCR) offers further advantages and sensitivity in miRNA detection and can circumvent the limitations of northern blotting analysis and cloning; however, it is limited to high cost. qRT-PCR has been successful has been successfully used to detect the expression of miRNAs. Three major qRT-PCR techniques have been used to detect the expression of miRNA: Primer-extension, quantitative PCR (PE-qPCR) (Raymond et al, 2005); Poly(A) tailing assay (Shi and Chiang, 2005); and stem-loop RT-PCR (Chen et al, 2005b). Stem-loop primers are superior to conventional primers in terms of RT efficiency for mature miRNA and can discriminate among related miRNAs that differ as little as one nucleotide. These assays quantify miRNA expression levels with superior performance over existing conventional detection methods, and combine the power of PCR for exquisite sensitivity, real-time monitoring for a large dynamic range and TaqMan assay reporters to increase the specificity.

High-throughput sequencing technologies and miRNA sequencing

Computationally predicted miRNAs have been experimentally validated by PCR-based cloning or hybridization-based methods. PCR-based cloning is challenging when the mature miRNA region is unknown, whereas hybridization-based methods suffer from sensitivity issues to detect less-abundant miRNAs. These methods do not reveal the actual miRNA sequences. Recent advances in technology, including the next-generation high-throughput sequencing

technologies such as Illumina/Solexa, massively parallel and 454 pyrosequencing, can be used to identify plant miRNAs. Species-specific (novel) miRNAs often accumulate at lower levels than conserved miRNAs. Thus, it is often difficult to assess them using traditional sequencing approaches such as Sanger sequencing method, which has been widely used in model plant species with known genome sequences (Song et al, 2010; Ge et al, 2012). The availability of next-generation sequencing (NGS) technologies provides high-throughput tools for new discoveries in plant miRNAs, which have low abundance, tissue specificity and spatiotemporal specificity. NGS has been used successfully to identify miRNAs in multiple plant species including Arabidopsis (Rajagopalan et al, 2006; Fahlgren et al, 2007), Grape (Pantaleo et al, 2010; Wang et al, 2014), Citrus (Song et al, 2010), Strawberry (Ge et al, 2012), and Zea mays (Zhang et al 2009). High-throughput approaches also detect miRNAs by abundance analysis (Fahlgren et al, 2007).

The demand for low-cost sequencing has driven the development of high-throughput (next-generation) technologies that parallelize the sequencing process, resulting in thousands or millions of sequences at once. The application of technologies such as miRNA sequencing (miRNA-seq) has played a considerable role in the discovery of plant miRNAs. MiRNA-seq entails the use of next-generation sequencing or massively parallel high-throughput DNA sequencing technologies to sequence miRNAs. MiRNA-seq often requires specifically enriched small RNAs as input, and this technique allows researchers to discover previously uncharacterised miRNA and to examine tissue- and disease-specific expression patterns, and miRNA isoforms. Similar to other techniques, miRNA-seq offers both advantages (sequence independence and coverage) and disadvantages (high cost, infrastructure requirements, run length and potential artefacts).

Identification of miRNA Targets

With the increase in next-generation sequencing data, new miRNAs are being uncovered in various plant genomes at a rapid pace; one of the major challenges is to determine their function. A crucial step towards functional annotation of miRNA is to identify their targets. Target recognition of plant miRNAs requires near-perfect complementarity base pair matching, and thus,

target prediction is relatively uncomplicated for plant miRNAs. In plants, both experimental and computational techniques are used to identify target miRNAs. Predicting conserved miRNA targets has revealed that homologous mRNAs are targeted by conserved miRNAs within a miRNA family, yet allowing more gaps and mismatches between an individual miRNA and its target (Unver et al, 2009; Sun et al, 2012).

Predicting miRNA targets in plants has been much easier because miRNAs bind to the protein-coding region of target with perfect or near-perfect sequence complementarity. In plants, targets can be identified through perfect base-pair complementarity between miRNAs and mRNA sequences (Rhoades et al, 2002). Based on transcriptome analysis in transgenic Arabidopsis plants overexpressing miRNAs, Schwab et al (2005) designed a set of rules to predict miRNA targets. The criterion allows one mismatch in the region complementary to nucleotides 2-12 of the miRNA, but not in the cleavage site (nucleotides 10 and 11). Three additional mismatches (no more than two continuous mismatches) were permitted between nucleotide positions 12 and 21. Song et al (2010) reported another method that allowed maximum four mismatches, with one mismatch between positions 1 and 9 from the 5'-end of the miRNA, no mismatches between 12 and 21/24 and no gaps at the complementary sites. By applying these rules, miRNA targets have been predicted in plant such as grape (Sun et al, 2012; Wang et al, 2014) and citrus (Song et al, 2010). In plants, miRNAs typically guide cleavage of target RNAs through their high degree of miRNA-target base-pairing matching (Carthew and Sontheimer, 2009). However, some reports have shown that plant miRNAs can also repress target mRNA translation (Aukerman and Sakai, 2003; Chen, 2004; Huntzinger and Izaurralde, 2011).

Microarray expression analysis has been used to analyse the expression of target genes along with specific miRNAs. However, a limitation of this method is that they are observed among a pool of indirect changes in transcript abundance and it detects only miRNA-mRNA interactions that result in cleavage and degradation (Thomson et al, 2011). Among the experimental approaches 5'-RACE [rapid amplification of cDNA ends/RNA ligase-mediated 5' rapid amplification of

cDNA ends (RLM-RACE)], Poly (A) ends (PPM_RACE) and RLM-RACE, and degradome sequencing (Degradome-seq) have been widely utilised to confirm plant miRNA-mRNA target site.

Advances in identification and characterisation of miRNAs and their target genes in plants provide a better understanding of post-transcriptional gene silencing. The information generated from such studies is valuable for RNA research in plants and sheds light on the expression and function of miRNAs. In addition, recently developed high-throughput methods have greatly enhanced the capacity to identify and validate novel miRNAs and their target genes. With the increasing genome sequencing information in plants, it will be intriguing to comprehensively analyse and compare the data across the genome to further broaden the knowledge of small RNA-mediated regulation in plants.

1.7 Plant miRNAs and the adaptive response to viral invasion

The discovery of miRNAs dates back to 1993 (Lee et al, 1993), however it was not until a decade ago that miRNAs were implicated in the host's defence mechanism (Llave, 2004). In plants, miR393 was the first host-derived sRNA recognised to function in antibacterial resistance by modulating the auxin-signalling pathway (Navarro et al, 2006). Plant-derived miRNAs were later reported to be associated with the repression of Plum pox virus (PPV) replication in vivo (Simon-Mateo and Garcia, 2006).

Several studies have demonstrated miRNA-mediated post-transcriptional regulation in response to viral infection. Microarray analysis of tomato plants agroinfected with *Tomato leaf curl New Delhi virus* (ToLCNDV) detected the deregulation of conserved miRNA families including miR319 and miR172 (Naqvi et al, 2010). When *Nicotiana benthamiana* plants were infected with four distinct begomoviruses *African cassava mosaic virus* (ACMV), *Cabbage leaf curl virus* (CbLCuV), *Tomato yellow leaf curl virus* (TYLCV) and *Cotton leaf curl Multan virus/Cotton leaf curl betasatellite* (CLCuV/CLCuMB), miRNAs involved in plant

development processes were found to be upregulated, leading to the suppression of corresponding endogenous targets (Amin et al, 2011). In rice, RNA deep sequencing methods were applied to analyze miRNA profiles during infection with the *Rice dwarf virus* (RDV; dsRNA virus) and *Rice stripe virus* (RSV; negative sense and ambisense RNA virus) (Du et al, 2011). RSV infection triggered the accumulation of miRNA*s rather than the corresponding miRNAs, accompanied by the enhanced expression level of rice *DCL* and *AGO* genes. In contrast, RDV infection resulted in an upregulation of *OsRDR* genes. However, it is not known if the upregulation of *DCL*, *AGO* or *RDR* genes is linked to defence mechanisms. Co-infection of *Nicotiana benthamiana* with Potato virus X, Potato virus Y and the PPV resulted in an altered host miRNA expression profile. Thus the differential modulation of host sRNA metabolism can be observed under the condition of multiple virus infection (Pacheco et al, 2012). In *Brassica*, bra-miR1885 was found induced upon *Turnip mosaic virus* (TuMV) infection (He et al, 2008). Bra-miR1885 targets a TIR-NB-LRR (Toll/inter- leukin-1, nucleotide-binding site leucine-rich repeat) disease resistance gene. Bra-miR1885 probably originates from inverted duplication events of TIR-NB-LRR coding genes. A recent study also investigated the Arabidopsis smRNA profile upon infection with the *Oilseed rape mosaic tobamovirus* (ORMV) (Hu et al, 2011). Thereby, a size-specific enrichment of miRNAs was observed. As the corresponding mRNA targets did not exhibit a corresponding transcriptional change, it has been hypothesized that mature miRNAs only play minor roles during Arabidopsis:ORMV interactions. In a similar study, tomato plants challenged with the *Cucumber mosaic virus* (CMV) and the N5 strain of *Tomato mosaic virus* (ToMV) were subjected to a RNA deep sequencing study (Chen et al, 2012). Over 85% of the analyzed miRNAs were found to be altered; however, the exact role of this phenomenon remains to be elucidated. In grapevine, infection with the *Grapevine vein- clearing virus* also triggers adaptations of the miRNA profile (Singh et al, 2012). MiR169 and miR398 were downregulated in response to viral infection, whereas miR168 and miR3623 were upregulated. However, it remains elusive whether the transcriptional change of these miRNAs has a direct or indirect effect on disease resistance. In summary, virus infections were shown to trigger changes in miRNA transcriptomes of several plant species. Nonetheless, for the majority of novel studies investigating

the role of miRNAs in antiviral defence, the exact contribution to defence mechanisms is still unknown.

Notwithstanding the plethora of plant miRNA regulatory networks that are operational in plants, it is plausible to deduce a common pattern of miRNA regulation due to viral infection. Uncovering these patterns could prove beneficial to the development of biomarkers for the diseased state or towards imparting plant resistance through antiviral strategies. It has also been suggested that miRNA passenger strands (miRNAs*), previously considered degradation products with little role in vivo, are involved in the antiviral defence mechanism of plants (Naqvi et al, 2010). The conserved and abundantly expressed plant miRNA families (miR156, miR159, miR319, miR172, etc.), in general, merit discussion, as they are thought to have a repressive role toward viral invasion. Computational (Pérez-Quintero et al, 2010) and microarray-based experiments have provided evidence that conserved miRNAs generally demonstrate greater antagonism toward viral genomes (Naqvi et al, 2010).

The assumption that those miRNAs, which were able to confer defence against viral invasion, would have survived evolutionary selection and became conserved, provides an explanation for the abundance of conserved miRNAs in the plant small RNAome. It also follows from this assumption that any supplementary functions that miRNAs exhibit would have most likely been acquired by them at later evolutionary stages. The occurrence of ORFs encoding viral suppressors of RNA silencing (VSRs) in the genomes of plant viruses, with their primary function of debilitating the host's sRNA metabolism, stands as evidence in support of this latter hypothesis. The hypothesis is further sup- ported by the duplication and divergence mechanism of miRNA evolution. The mechanism reveals that miRNA families that are conserved across species exhibit copy number variation, followed by qualitative sequence differentiation, which together, are thought to be leading to the evolution of miRNAs with the emergence of novel functions (Ehrenreich and Puruggana, 2008).

It has also been hypothesized that even plant miRNAs that are relatively less abundant could conceivably cater to the host's defence mechanisms during specific host–virus interactions. Of late, miRNAs have also been implicated in the regulation of plant innate immune responses by modulating nucleotide binding site–leucine-rich repeat (NBS– LRR) genes in *Solanaceae* (Li et al, 2012; Shivaprasad et al, 2012). In NBS–LRR-mediated non-specific immunity, when there is absence of infection by a pathogen, only a few miRNAs control the cascade of defence proteins. Conversely, defence proteins under miRNA control are triggered instantly, upon viral invasion, as VSRs depress miRNA- based control of defence proteins. Thus, it appears that miRNA-mediated modulation of plant defence mechanisms functions on the principle of cellular economy. Another perspective on the presence of host-derived miRNAs is that these viral responsive miRNAs, by not targeting all the viral ORFs, might be enabling co-existence of viruses inside the host, and thereby allowing the establishment of a persistent infection (Mahajan et al, 2009). This perspective is plausible considering our deprived understanding of the sRNAome landscape of plants in general, and in particular, of the regulation of virus–plant interactions

Plant miRNAs and viral counter defences

The role of VSRs in debilitating host miRNA pathways also merits its own discussion. The molecular basis behind the manifestation of viral symptoms lies in the ability of VSRs to interfere with host miRNA biogenesis, ultimately affecting host mRNA turnover to the advantage of invading pathogens (Chapman et al, 2004; Chellappan et al, 2005). A p19 VSR of Cymbidium ring spot virus, for instance, induces host-derived conserved miR168 that is involved in restraining AGO-1 accumulation. As AGO-1 accumulation is crucial for the antiviral function of RISC, host miRNA modulations, under the influence of viral infection, lead invariably to an impaired host antiviral response (Bortolamiol et al, 2007; Varallyay et al, 2010).

Interestingly, VSR 2b of CMV has been shown to exhibit miRNA modulating activity and symptom induction, independently of one another, leading to the

conclusion that the RNA suppressor domain acts discretely from the host miRNA inhibitory domain (Lewsey et al, 2009). A report on two unrelated VSRs (Potyvirus HC-Pro and Carmovirus p38) revealed viral activities that were consistent with the notion of distinct domains. In addition, it is known that host TFs are involved in HC-Pro-mediated morphological anomalies but not in their miRNA inhibitory role (Endres et al, 2010). Furthermore, the differential effect of VSRs on siRNA and miRNA AGO-1 loading proposes the presence of two different pools of ARGONAUTE proteins in vivo (Schott et al, 2012). To summarize, because the plant's antiviral defence and endogenous gene regulatory networks share common protein machinery, which would otherwise be involved in maintaining normal cellular processes, leading to the manifestation of disease symptoms.

Resistance (R)-gene mediated immunity is regulated by miRNAs

Beside their role in fine-tuning hormonal pathways during defence responses, miRNAs have been demonstrated to play a pivotal role during the early steps of the plant immune responses. The plant immune system is multilayered and consists of constitutive defence barriers such as cell walls or pre-formed toxins, as well as inducible local and systemic defences (Spoel and Dong, 2012). Upon pathogen and pest perception, plants employ a so-called innate immunity which is mediated by two major receptor classes, namely pattern recognition receptors (PRRs, [Zipfel and Robatzek, 2010]) and resistance (R) proteins which are generally intracellular nucleotide-binding site leucine-rich repeat (NB-LRR) proteins (Elmore et al, 2011). PRRs perceive conserved pathogen- or microbe-associated molecular patterns (PAMPs/MAMPs) such as flagellin or chitin. This triggers the activation of downstream defence pathways including the generation of reactive oxygen species (ROS) that are toxic for invaders, thus leading to a pattern-triggered immunity (PTI). However, successful pathogens are capable of suppressing PTI with the help of (Avr) proteins, pathogen effectors that either interact directly with PRRs or interfere with downstream factors (Boller and He,

2009). These effectors are recognized and attenuated by NB-LRR proteins encoded by plant *R*-genes, thus leading to an effector-triggered immunity (ETI).

Over the past few years, novel findings uncovered a pivotal role of miRNAs during NB-LRR-mediated resistance. In tobacco, two miRNAs (nta-miR6019 and nta- miR6020) were discovered to cleave the mRNA of the tobacco *N*-gene that encodes a NB-LRR receptor (Zhai et al, 2011), *N*-mediated resistance against *Tobacco mosaic virus* (TMV). Overexpression of both miRNAs resulted in an attenuated *N*-mediated resistance to TMV. Intriguingly, cleavage of *N* mRNA resulted in the accumulation of 21-nt siRNAs in phase with the cleavage site of nta-miR6019; the generation of these siRNAs requires RDR6 and DCL4. Hence, miR6019 triggers the generation of secondary siRNAs that act in concert to control innate immunity. In the same study, bioinformatic investigations also led to the identification of miRNAs targeting *R*-genes in tomato and potato. Similarly, the miRNA family miR1507, miR2109 and miR2118 were demonstrated to regulate NB-LRR encoding genes in legumes via phased siRNAs (Zhai et al, 2011). These findings were recently complemented by the demonstration that miR482/2118 targets NB-LRR encoding genes in tomato (Shivaprasad et al, 2012). The miR482/2118 family was found highly abundant in the *Rutaceae, Solanaceae* and *Fabaceae*, suggesting a conserved regulatory role. Expression of miR482 was associated with the synthesis of secondary siRNAs, which also target several NB-LRR encoding genes. Moreover, in bacteria- and virus-infected plants, the miR482-mediated suppression of NB-LRR was alleviated, resulting in enhanced levels of NB-LRR proteins. In summary, miRNAs seem to be heavily implicated in regulating NB-LRR-mediated innate immunity. Normally, NB-LRRs are associated with race-specific immunity, thus specific NB-LRRs recognize race-specific effectors. A pathogen-mediated downregulation of miRNAs targeting NB-LRRs would lead to an overexpression of NB-LRRs in a non-race specific manner, thus resulting in a broader resistance. Therefore, low levels of NB-LRRs under miRNA control might reduce the plant defence costs, as multiple NB- LRRs can be rapidly induced upon pathogen stress.

2. Rationale and Overall Objectives of Study

Food security is one of the most important issues challenging the world today. Any strategy used to solve this problem must include increasing crop yields and quality. Since cassava is robust and can withstand many harsh conditions including low rainfall and nutrient poor soils, it serves as the perfect famine reserve and food security crop for many subsistence farmers. In addition, it is also important source of starch for industrial applications and bioethanol productions. While cassava has been grown for hundreds of years in southern Africa by African tribesmen for subsistence, it is only more recently that the potential value of cassava for both food security and industrial purposes has been realized. Casquip Starch Manufacturing Pty Ltd began the initiative in 1998 in the Limpopo and Mpumalanga Provinces, and also expanded later on into Swaziland. Recently, in 2015, cassava was recognized and acknowledged as an official commodity and the Cassava Industry Association SADC (CIASA) was established through the Department of Trade and Industry (dti). CIASA is registered with the Department of Social Development. National cassava germplasm trials in Limpopo, Mpumalanga and ZwaZulu –Natal provinces have recently been initiated with support from the Agricultural Research Council and Technical Innovation Agency. Government and industry are working with small-scale farmers in these provinces to establish cassava cultivation, and larger commercial cassava farms are planned for the future. T200 is a southern African landrace with high starch and has been identified as one of the appropriate commercial varieties, but other high starch, local environment-adapted varieties are being tested.

However, like many crops, cassava is vulnerable to infection by a number of pathogens. One of the most devastating and economically important constraints to threatening cassava production is infection by at least one of the 11 species of begomoviruses (Family *Geminiviridae*) resulting in cassava mosaic disease (CMD), which causes devastating yield losses. Losses have been reported to be as high as 90% in highly infected cassava fields. *South African cassava mosaic virus* is one of the 11 begomoviruses that infects cassava and was first isolated from a field in South Africa in 1999. These viruses are difficult to control, bringing about the need for effective disease control strategies. One approach for control is to elucidate mechanisms involved in the host defence-responses to an invading pathogen so

that manipulation of endogenous genes or gene networks (*cis*-genics) can be employed to develop virus resistant plants.

Since cassava is considered an "orphan crop", it has been neglected by research in favor of cereals and other economically important crops. Due to the limited information concerning geminivirus-plant host interactions, insufficient information regarding the changes in the small RNAome that occur in CMD-infected cassava is available. Furthermore, while plant pathogen resistance has been well studied, little is known about the molecular mechanisms involved in tolerance and recovery in plants.

Therefore, the main aim of the work presented in this thesis was to investigate the affect SACMV infection has on the small RNA populations in a susceptible cassava landrace (T200) and a SACMV tolerant landrace (TME3) and determine if any of the small RNA populations play a role in TME3 tolerance and recovery to SACMV infection.

2.1 Specific Objectives

The specific objectives of this project are outlined in the chapter format below.

Outline of this thesis

Chapter 2: At the conception of this project, there was very limited information about the microRNAome of cassava. The aim of this study was to identify both conserved and cassava-specific (novel) miRNAs in cassava T200 and TME3 landraces using the freely available cassava EST and GSS databases as well as data generated next-generation sequencing (NGS). T200 was chosen as it is a southern African high starch landrace with industrial-potential, and TME3 is a West African landrace with known tolerance to cassava mosaic disease (CMD). The miRNAs present in miRBase (V.21) were used to identify potential conserved miRNAs in the EST and GSS databases and the NGS data. The miRCat tool that is part of the UEA small RNA workbench was used to identify novel miRNAs in the NGS data.

Targets for both the conserved and novel miRNAs were identified using the psRNATarget webtool.

Chapter 3: The next aim was to determine and compare the changes in expression of miRNA populations in SACMV-infected tolerant (TME3) and susceptible (T200) cassava landraces. To achieve this, first conserved and novel miRNAs were identified in TME3 and T200 in SACMV-infected and mock-inoculated leaf samples collected at 12, 32 and 67 days' post infections (dpi). These three time-points represent the progression of disease; 12dpi represents the early pre-symptomatic stage, 32dpi represent full-systemic infection and 67dpi represents the late infection stage and the recovery stage in TME3.

Chapter 4: The final aim was to determine if methylation and virus-derived vsRNA populations are possible RNA silencing mechanisms involved in the recovery phenotype observed in SACMV-infected cassava cultivar TME3 and the susceptible phenotype in T200.

Chapter 5: This chapter collates all the results obtained from the studies in chapter 2, 3 and 4. The results obtained during the course of this PhD work are discussed in a context of the latest insights regarding plant-virus interactions and defence mechanisms. How the major findings of this research can contribute to elucidating some of the molecular mechanisms that occurs in geminivirus-plant infection systems is discussed. Also, future recommendations and how this data can be used to achieve CMD resistant/tolerant farmer-preferred cassava varieties is put forward.

Chapter 2

Unveiling the Micronome of Cassava (*Manihot esculenta* Crantz)

This Chapter has been published:

Article Source: **Unveiling the Micronome of Cassava (Manihot esculenta Crantz)**

Rogans SJ, Rey C (2016) Unveiling the Micronome of Cassava (Manihot esculenta Crantz). PLoS ONE 11(1): e0147251. doi: 10.1371/journal.pone.0147251

2.1 Abstract

MicroRNAs (miRNAs) are an important class of endogenous non-coding single-stranded small RNAs (21-24 nt in length), which serve as post-transcriptional negative regulators of gene expression in plants. Despite the economic importance of *Manihot esculenta* Crantz (cassava) only 153 putative cassava miRNAs (from multiple germplasm) are available to date in miRBase (Version 21), and identification of a number of miRNAs from the cassava EST database have been limited to comparisons with *Arabidopsis*. In this study, mature sequences of all known plant miRNAs were used as a query for homologous searches against cassava EST and GSS databases, and additional identification of novel and conserved miRNAs were gleaned from next generation sequencing (NGS) of two cassava landraces (T200 from southern Africa and TME3 from West Africa) at three different stages post explant transplantation and acclimatization. EST and GSS derived data revealed 259 and 32 miRNAs in cassava, and one of the miRNA families (miR2118) from previous studies has not been reported in cassava. NGS data collectively displayed expression of 289 conserved miRNAs in leaf tissue, of which 230 had not been reported previously. Of the 289 conserved miRNAs identified in T200 and TME3, 208 were isomiRs. Thirty-nine novel cassava-specific miRNAs of low abundance, belonging to 29 families, were identified. Thirty-eight (98.6%) of the putative new miRNAs identified by NGS have not been previously reported in cassava. Several miRNA targets were identified in T200 and TME3, highlighting differential temporal miRNA expression between the two cassava landraces. This study contributes to the expanding knowledge base of the micronome of this important crop.

2.2 Introduction

MicroRNAs (miRNAs) are an important class of endogenous small RNAs. They are evolutionary conserved, single-stranded, non-coding pieces of RNA that are 21-24 nt in length (Ambros et al, 2003; Bartel, 2004). MiRNAs serve as post-transcriptional negative regulators of gene expression in plants and animals by negatively regulating their target gene expression at post-transcriptional levels through mRNA cleavage or repression of translation, depending on the

complementarity between the miRNAs and their target genes (Carthew and Sontheimer, 2009; Jones-Rhoades et al, 2006; Zhang et al, 2006a) MiRNAs regulate a great number of genes involved in plant growth and development, environmental stress response, signal transduction as well as response to pathogen invasion (Dugas and Bartel, 2004).

The biogenesis of mature miRNAs encompasses a co-ordinated interplay of a few cellular proteins in and outside of the nucleus and is a multi-step process. Like their protein-coding counterparts, miRNAs are also transcribed form their own genes, known as *MIR* genes (Carthew and Sontheimer, 2009; Chen, 2005). *MIR* genes are much longer than their mature sequences and range from several tens to more than 1000 nt. Mature miRNAs are produced from a pathway starting with the *MIR* genes being transcribed to the capped and polyadenylated primary miRNA transcripts (pri-miRNA) by the Pol II enzyme (Lee et al, 2004; Mallory et al, 2008). The pri-miRNA forms an imperfect hairpin-like secondary structure, which undergoes cleavage to form a perfect hairpin precursor called precursor miRNA (pre-miRNA) with the aid of Dicer-like enzyme (DCL1), a plant counterpart of the animal Dicer enzyme (Bartel, 2004; Kurihara and Watanabe, 2004; Xie et al, 2004; Lu et al, 2005). In the dicing process DCL1 interacts with the pri-miRNA with the aid of DWADLE (DDL), which plays a significant role in recruiting DCL1 to the pri-miRNA (Yu et al, 2008). The pre-miRNA is further processed by DCL1 to release the stem portion of the hairpin as a miRNA: miRNA* duplex (miRNA* is the complementary sequence to miRNA on the opposing arm) (Bartel, 2004; Kurihara and Watanabe, 2004). Cellular enzymes like HYL1 (HYPONASTIC LEAVES 1), HEN1 (HUA ENHANCER 1) and HST1 (HASTY 1) are obligatory for the maturation of miRNAs. DCL1 associates with its cohort HYL1, a dsRNA binding protein, and SERRATE (SE), a zinc finger protein, to produce mature miRNAs (Dong et al, 2008; Man et al, 2004). The processed miRNA duplex is 2'-O methylated and polyuridylated at the 3'-terminal nucleotide by a dsRNA methylase HEN1. The methylation protects miRNAs from degradation (Li et al, 2005; Park et al, 2002). The mature miRNA duplex is then exported out of the nucleus by HST1 an EXPORTIN 5 orthologue in plants (Papp et al, 2003; Park et al, 2005; Yi et al, 2003). Out of the two strands of the mature miRNA duplex

(miRNA:miRNA*), the one with the least 5' end thermodynamic stability will function as a mature miRNA, whereas the other strand (miRNA*) termed the passenger strand is specifically degraded (Khvorova et al, 2003). Finally, the single-stranded mature miRNA is incorporated with AGONAUTE (AGO) proteins to form a ribonucleoprotein complex known as RNA-induced silencing complex, where the regulation of target gene expression occurs (Bartel, 2004; Dugas and Bartel, 2004; Voinnet, 2009). The RISC complex along with the mature miRNA negatively regulates gene expression either by inhibiting translation elongation or triggering messenger RNA (mRNA) degradation depending on the degree of complementarity of the miRNA sequence with its target.

A large number of miRNA families are evolutionary conserved in the plant kingdom, which ranges from mosses and ferns to higher flowering plants (Pan et al, 2006). This attribute has been used as a practical indicator for the identification and prediction of miRNAs by homology searches in other species. During recent years the identification and characterisation of miRNA (Pan et al, 2006) and their target genes from plants has been extensively studied (Ghani et al, 2013; Jones-Rhoades and Bartel, 2004; Unver et al, 2009). In the past decade, a large number of miRNAs have been discovered across several plant species; for instance, the miRBase database (Kozomara and Griffiths-Jones, 2014) version 21 contained 8524 mature miRNA sequences for 73 plant species. The majority of these miRNAs have been validated using different computational and experimental approaches including deep sequencing, cloning, northern blots and real-time PCR (Jones-Rhoades et al, 2006; Meyers et al, 2006; Sun, 2012).

Comparison of miRNAs in different plant species by expression sequence tags (EST) analysis had shown that some miRNAs were highly evolutionary conserved among species (Pan et al, 2006). This provided a powerful strategy for identifying miRNAs in a new plant species. Identification of miRNAs using EST analysis has two significant advantages (Frazier and Zhang, 2011): there is no specialized software required and it can be used to identify miRNAs in any species if they have previously registered EST sequences. Since ESTs are derived from transcribed sequences, EST analysis also provides direct evidence for miRNA

expression. In view of these advantages, EST analysis had been used to identify conserved miRNAs in several plants including *Brassica napus* (Xie et al, 2007), *Medicago trunculata* (Zhou et al, 2008), *Lycopersicon esculentum* (Yin et al, 2008), *Glycine max* (Zhang et al, 2008), *Nicotiana tabacum* (Frazier et al, 2010), and *Solanum tuberosum* (Xie et al, 2011). In addition, an *in silico* search of miRNAs in public databases and a bioinformatics approach can greatly assist to identify miRNAs in several plants (Baloch and Muhammed, 2014) especially those whose complete genome sequences are unavailable. It has also been suggested that most of the miRNAs predicted from EST analysis can also be identified by high throughput deep sequencing (Kwak et al, 2009).

There are several miRNAs considered to be recently evolved that show species-specificity and are often expressed at lower levels. Many are tissue specific, and are expressed at certain stages in development or under specific growth conditions, relative to the highly conserved group of miRNAs (Allen et al, 2004; Fahlgren et al, 2007). Next generation sequencing (NGS) technology has great promise to generate an accurate and comprehensive picture of the small RNA transcriptome in different plants, tissues, and at different developmental stages. Using deep sequencing, both species-specific (novel) and conserved miRNAs have been identified in diverse plant species such as *Arabidopsis* (Fahlgren et al, 2007; Rajagopalan et al, 2006), tomato (Moxon et al, 2008; Zuo et al, 2012), cucumber (Martinez et al, 2011), maize (Wang et al, 2011), peanuts (Chi et al, 2011), pepper (Hwang et al, 2013) and rice (Morin et al, 2008). However, homology-based searches in databases are not sufficient for identifying miRNAs; therefore, other additional criteria have been set for distinguishing miRNAs from other types of small RNAs. Predicting the secondary structure of the pre-miRNA and calculating the free energy are necessary for reducing the number of false positive identified miRNAs (Ambros et al, 2003; Bonnet et al, 2004; Meyers et al, 2008; Zhang et al, 2005).

Cassava (*Manihot esculenta* Crantz) is a crop widely grown as a staple food and along with maize, sugarcane and rice is a major source of energy for more than 700 million people in most tropical countries including sub-Saharan Africa (FAO,

2013). Apart from its traditional role as a food crop, there is a growing demand for cassava starch in a diverse set of industries such as animal feed, paper, textile and adhesive as well as an alternative energy resource (El-Sharkawy, 2004). Despite the economic importance of cassava and the potential contribution of miRNAs to cassava improvement, molecular genetic information regarding cassava miRNAs remains sparse. Only recently, 153 conserved miRNAs were made available in miRBase (Version 21 for cassava (Kozomara and Griffiths-Jones, 2014), however other well-studied plant species such as *Arabidopsis thaliana, Glycine max, Populus trichocarpa* and *Oryza sativa* have 427, 639, 401 and 713 reported miRNAs in miRBase, respectively (Kozomara and Griffiths-Jones, 2014). The miRNAs that are available for cassava on miRBase were obtained by Patanun et al., (2013) using a computational prediction method by using homology search based on miRNA conservation among different plant species. In addition, Perez-Quintero et al (2012) analysed small RNA libraries from cassava tissues infected and uninfected with *Xanthomonas axonopodis*, and Zeng et al (2009) studied conserved miRNAs in the Euphorbiaceae family. More recently, Ballen-Taborda et al. (2013) and Xia et al. (2015) both studied cassava miRNAs expressed under abiotic stress conditions.

The identification of a more comprehensive set of miRNAs in cassava is a critical step to facilitate our understanding of regulatory mechanisms or networks, in particular responses to viral pathogens, of particular interest in our laboratory. In this study we employed a combinatorial approach of publicly available cassava EST and GSS data in NCBI, and next-generation sequencing-derived miRNA data collected at 8, 10 and 15 weeks post-planting from two cassava landraces, T200 and TME3, to systematically identify conserved and novel miRNAs in cassava. Our findings revealed 259, 32 and 289 conserved miRNAs using the EST, GSS and NGS data respectively and 39 novel cassava-specific miRNAs of low abundance, belonging to 29 families. In order to understand the function of the newly identified conserved and novel miRNAs in cassava, the targets of these miRNAs were also identified. The knowledge gained from this study contributes to the cassava miRNA database and micronome of this important crop, and unveils differences between landraces, which will be

beneficial in the long term in linking gene regulation, gene targets and germplasm traits.

2.3 Methods and Materials

Identification of miRNAs in cassava using EST and GSS database

Sequence databases

A total of 8524 known plant mature miRNA sequences were downloaded from miRBase database (http://www.mirbase.org/; Release 21: June 2014) (Kozomara and Griffiths-Jones, 2014). The repeated miRNA sequences were removed to avoid redundant miRNAs and the remaining unique sequences were used as the reference set. The ESTs (86 310) and GSSs (77,569) available for cassava (collected from multiple germplasm) were downloaded from NCBI (http://www.ncbi.nlm.nih.gov).

Identification of putative conserved miRNAs in cassava

Two crucial filter conditions in EST and GSS analysis were used to identify conserved miRNAs: the conservation of mature miRNA sequences and the secondary structure of the pre-miRNA (Zhang et al, 2008). The mature sequences of all known plant miRNAs were used as a query for homologous searches against the cassava EST library and GSS database using BLAST search in CLC Main Workbench version 6.6.2. All cassava EST sequences with no more than 3 mismatches against the query sequences were saved. These initial candidate miRNA sequences predicted from the mature reference miRNAs were subjected for protein homology search at NCBI using BLASTx (http://www.ncbi.nlm.nih.gov/BLAST/) with default parameters and the protein coding sequences were removed.

Prediction of stem-loop secondary structures

The precursor sequences of the potential cassava sequences were subjected to hairpin secondary structure prediction using the RNA folding tool using default parameters in CLC Main Workbench version 6.6.2. The following criteria were used for selecting potential cassava pre-miRNAs (Ambros et al, 2003;

Zhang et al, 2005): (1) Pre-miRNA could fold into a typical hairpin secondary structure and the mature miRNA was located in one stem; (2) the length of the pre-miRNA was no less than 50 nt; (3) pre-miRNA had a high minimal folding free energy (MFE) and MFE index (MFEI), which was calculated by MFEI=MFEx100/[length x (G+C%)], where length is the length of the RNA sequence and MFE is a negative folding free energy (-ΔG) [72]; (4) the maximum number of nucleotide mismatches between the mature miRNA and its opposite miRNA* sequence was six; (5) no loops or breaks in miRNA/miRNA* duplex was allowed.

Identification of miRNAs in cassava using Next Generation Sequencing (NGS) data

Micropropagation and acclimatization of cassava

T200 and TME3 cassava landraces were micropropagated by way of nodal culture on Murashige and Skoog (MS) medium (Murashige and Skoog; 1962) supplemented with 20g.L-1 sucrose and 2g.L-1 Phytagel™ (Sigma Aldrich), pH 5.8. Explants for both landraces were grown under identical conditions, and were allowed to grow at 25°C under a 16 h photoperiod. At the appearance of roots (10 days), plantlets were transferred into Jiffy® pellets which were placed on a tray that was covered with plastic film and placed in an insect free, temperature controlled growth chamber (28°C; 16-hour photoperiod). Slits were then gradually made in the plastic film to facilitate acclimatization of explants. Once acclimatised, the plantlets were potted with a 2:1 ratio of potting soil to vermiculite. The potted plants remained in the insect free, temperature-controlled growth chamber (28°C; 16-hour photoperiod). The average light intensity of the growth chamber was 3000 lux. The plants were watered every second day and once a month multifeed fertilizer was added to the plants, following manufacturer's instructions. The newly developing upper most leaves were collected from the T200 and TME3 plants at 8, 10 and 15 weeks after the plantlets had been transferred to the Jiffy® pellets. These time points correlate to early, middle and later growth stages.

RNA extraction, small RNA library preparation and sequencing

Total RNA extraction, using a modified high molecular weight polyethylene glycol (HMWPEG) protocol (Gehrig et al, 2000), was carried out on leaf tissue samples collected from T200 and TME3 at 8, 10 and 15 weeks. Six leaves from each plant in the three biological replicate experiments were pooled to reduce variation. For each sample, 1g pooled leaf tissue was homogenised in liquid nitrogen and added to 5ml preheated (65°C) GHCL buffer (6.5 guanidium hydrochloride, 100mM Tris-HCl pH 8.0, 0.1M sodium acetate pH 5.5, 0.1M β-mercaptoethanol) and 0.1g HMW-PEG (Mr: 20 000, Sigma). The mixture was then pelleted by centrifugation (10000xg) for 10 minutes at 4°C. The supernatant was treated with 0.1ml 1M sodium citrate (pH 4.0), 0.2 ml 2M NaCl and 5 ml phenol:chloroform:isoamyl alcohol (PCI) (25:24:1). The mixture was then vortexed vigorously and again pelleted by centrifugation (10000 x g) for 10 minutes at 4°C. The supernatant was removed and RNA was precipitated by adding 5ml isopropanol (propan-2-ol). The mixture was thoroughly mixed and incubated at -20°C for 60 minutes and pelleted by centrifugation (10000 x g) for 25 minutes at 4°C. RNA pellets were washed with 5ml ice-cold 75% molecular grade ethanol. RNA Pellets were dried at 37°C for 5 minutes. The pellet was resuspended in 100µl preheated (55°C) RNase-free water and 1µl RNase inhibitor (Fermentas). Small RNAs were specifically filtered for using the mirVana™ miRNA isolation kit (Ambion Inc), following the manufacturer's protocol. For cDNA library preparation, approximately 500 ng was used as input for the Illumina TruSeq Small RNA library preparation kit (Illumina, Inc.) and sequencing libraries were created according to the manufacturer's protocol. Briefly, poly-A containing mRNA molecules were purified using poly-T oligo-attached magnetic beads. Following purification, the mRNA was fragmented and copied into first strand cDNA using random primers and reverse transcriptase. Second strand cDNA synthesis was then done using DNA Polymerase I and RNase H. The cDNA was then ligated to adapters and enriched with PCR to create the final cDNA library. The library was then pooled and sequenced on a HiSeq 2000 (Illumina, Inc.) instrument as per manufacturer's instructions. Sequencing was performed up to 2 X 101 cycles. Next generating sequencing (NGS) was done using the Illumina HiSeq2000 platform at LGC Genomics in Berlin, Germany.

Small RNA sequencing analysis

Raw reads generated from the Illumina HiSeq2000 system for the 6 small RNA libraries were cleaned of sequence adapters using the fast-toolkit (htt://hannonlab.cshl.edu/fastx_toolkit/), and low quality tags and small sequences (<15 nt long) were excluded. Quality analysis per cycle was performed for each library. To eliminate all other small non-coding RNAs, high quality trimmed sequences were mapped to rRNA, tRNA and snoRNAs sequences from Rfam (Version 12.0). The sequences that mapped completely and had an E-value <0.06 were removed from the libraries.

Identification of conserved miRNAs, isoforms, and novel miRNAs

In this study, CLC Genomics Workbench version 7.1 was used for data analyses. The FASTQ files containing the Illumina sequencing adapter clipped reads were imported into the CLC Genomics Workbench version 7.1 using the Import NGS option. The first step was to extract and count the small RNAs to create a small RNA sample that could be used for further analysis. The reads had previously had their adapters removed and trimmed. The maximum length of the small RNAs counted was set at 30 nt and the minimum length was 18 nt. The minimum sampling count was left as default, which was 1.

The small RNA samples produced when counting the tags was enriched by CLC Genomic Workbench by comparing the tag sequences with annotation resources such as miRBase. The integrated tool in the workbench was used to download miRBase. The downloaded version was the latest version, release 21, and was downloaded from ftp://mirbase.org/pub/mirbase/CURRENT/miRNA.dat.qz. The downloaded miRBase file contains all precursor sequences from the latest version of miRBase including annotations defining the mature miRNA regions. All plant species were selected from the list of species in miRBase. All settings were left as default except for the maximum mismatches, which was changed to 3.

The reads that mapped to the known miRNAs from miRBase with no more than 3 mismatches were then aligned to the cassava genome. Only the reads that mapped with no mismatches or gaps were considered to be potential miRNAs. In order to select for potential cassava pre-miRNAs, the region 250nt upstream and downstream from where the read mapped to the cassava genome was folded using the RNA folding tool in the CLC genomics workbench and was analysed using the secondary structure identification criteria mentioned previously (Ambros et al, 2003; Zhang et al, 2005).

In order to predict novel miRNAs from cassava, the miRCat program in the UEA small RNA workbench was employed [98]. MiRCat identifies mature miRNAs and their precursors from a sRNA dataset and an input genome, AM5602 available at Phytozome (http://www.phytozomenet/cassava). The sRNA sequences are mapped to the input plant genome using PatMaN (Prüfer et al, 2008) and grouped into loci. In order to enrich for miRNA candidates, a number of criteria for the determination of a bona fide miRNA loci are applied by the software. In brief, the program searches for two-peak alignment patterns of sRNAs on one strand of the locus and evaluates the secondary structures of a series of putative precursor transcripts using the RNAfold (Hofacker et al, 1994) and randfold (Bonnet et al, 2004) programs. According to the recent criteria for annotating novel plant miRNAs, miRNA star (miRNA*) is one of the most important biogenesis proofs for the identification of a novel miRNA (Meyers et al, 2008), and therefore only the identified novel miRNAs that had a corresponding miRNA* sequence identified were considered at potential cassava specific novel miRNAs.

Experimental validation of selected miRNAs using reverse-transcription PCR

For the RT-PCR (reverse transcription) experimental validation, 7 conserved miRNAs and 6 novel miRNAs were randomly chosen from the predicted cassava miRNAs. The primers for the stem-loop sequences of these chosen miRNAs were designed using Integrated DNA technologies Primer Quest tool (www.idtdna.com/Primerquest/Home/index (S9 Table). Total RNA was

extracted from cassava leaves using the Direct-zol™ RNA MiniPrep (Zymo Research), according to the manufacturer's protocol. cDNA was synthesised using the RevertAid™ H minus First Strand cDNA synthesis Kit (Fermentas), according to the supplier's protocol. One hundred ng cDNA was used as template for the PCR. The PCR was programmed as follows: initial denaturation at 95°C for 3 minutes followed by 34 cycles of denaturation at 95°C for 30 seconds, annealing at 59°C for 30 seconds, and extension at 72°C for 30 seconds and final elongation step at 72°C for 10 minutes. The PCR products were separated through 2% (w/v) agarose gel.

Identification of targets and Gene Ontologies

Target genes were identified using psRNATarget server, an automated plant miRNA target prediction server available at plantgrn.noble.org.psRNATarget/ (Dai and Zhou, 2011) using the *Manihot esculenta* (cassava), Unigene, DFCI Gene Index. The analysis parameters were set as default. Briefly, the following criteria were set for predicting the potential cassava miRNA target genes: (1) not more than four mismatches between identified miRNA and target mRNA; (2) no mismatches were allowed between positions 10th, 11th because this site was believed as a cleavage site; (3) one mismatch was allowed between position 2nd and 12th and up to three mismatches between position 12th and 25th; and (4) not more than two consecutive mismatches. To better understand the functions of the newly identified potential targets, proteins were allocated gene ontology (GO) terms using Uniprot (www.uniprot.org).

2.4 Results and Discussion

Small RNA sequencing analysis

In order to identify novel and conserved miRNAs in two cassava landraces, six small RNA-enriched libraries were generated from cassava leaves that were collected from two cassava landraces, T200 and TME3, at 8, 10 and 15 weeks after transferring plantlets from tissue culture to Jiffy® pellets using the Illumina

HiSeq2000 system. The small RNA sequencing yielded a total of 64 827 692 raw reads for the six libraries (Table 1).

After removing low-quality sequences, adapters, and small sequences (<15 nt), 68.8% (44 621 667 reads) of the raw reads remained. A final filtering step, to obtain the sequences that have sizes between 18 and 26 nt, yielded a total 16 302 012 reads for the six libraries (Table 1). The 18-26 nt libraries were normalized per million read counts in order to compare sRNA abundance data. The next filtering step involved the removal of non-coding RNAs such as ribosomal RNA (rRNA), transfer RNA (tRNA) and small nuclear RNA (snoRNA). This filtering step was performed by conducting a BLASTn search of the small RNA libraries against the RNA families database Rfam (Burge et al, 2013) (Table 1). Only sequences with perfect matches and an E-value <0.06 were removed from the libraries.

Table 1. Summary of small RNA sequencing data analysis.

	T200 8 weeks	T200 10 weeks	T200 15 weeks	TME3 8 weeks	TME3 10 weeks	TME3 15 weeks
Raw	6 774 098	14 687 734	23 163 380	2 914 650	4 589 392	12 692 438
Adapter-trimmed Reads	5 921 995	12 914 563	20 603 045	2 668 525	3 625 654	11 414 040
15–50 nt	4 692 091	10 248 438	15 487 598	2 209 332	2 788 194	9 196 014
18–26 nt	1 436 570	4 381 089	6 935 189	634 217	479 048	2 435 899
Normalised 18–26 nt	212 061.83	298 282.158	299 403.153	217 596.28	104 381.58	191 917.35
Total significant Rfam matching sequences*	23 273	34 405	65 997	23 005	15 438	48 811
rRNA	9 032	11 170	22 841	11 129	6 764	15 965
tRNA	5 106	9 873	14 883	2 913	2 604	5 634
snoRNA	2 387	3 190	5 641	1 797	1 286	5 451
Total mapped to miRBase V.20	42 339	57 945	221 841	33 208	30 539	88 200
Exact Matches (Before Hairpin Filtering)	5 150	10 779	63 594	4 294	1 909	9 784
isomiRs (Before Hairpin Filtering)	23 542	28 269	101 274	17 729	22 419	36 304

* Reads that had an E-value of <0.05

doi:10.1371/journal.pone.0147251.t001

The size distribution analysis of small RNA (sRNAs) sequences exhibited a similar pattern of length distribution in all libraries. The small RNA length distribution (18-26 nt) of each library showed that the most abundant and diverse species were those 21 - 24 nt in length (Fig. 2.1), which is typical of Dicer-derived products (Axtel, 2013). In all six libraries, while the 21 nt size class is characteristic of authentic miRNAs (Axtel, 2013), it was most intriguing to note that the 22 nt class was the most abundant, followed by the 23 nt class for T200 at 8 weeks (24.2%) and T200 at 15 weeks (18.1%). The 22 nt miRNA or miRNA* length is important for triggering secondary siRNA biogenesis (Chen et al, 2010;

Dugas and Bartel, 2004). The 22-nt miRNAs or miRNA* are often generated from asymmetric miRNA precursors. The asymmetric miRNA precursors affect the structure of the miRNA/miRNA* duplex, allowing RISC to recruit the RDR6 and SGS3 to trigger the formation of the secondary siRNA (Cuperus et al, 2010; Manavella et al, 2012). The 21 nt class sRNAs was under-represented in T200 at all 3 time points compared to the 22 nt sRNAs, and for TME3, the percentage of 21 nt sRNAs only comprised 17% and 11.5% of the total number of sRNAs in the 10 and 15-week library, respectively. Studies in grapevine (Pantaleo et al, 2010), wheat (Yao et al, 2007); Chinese yew (Qiu et al, 2009) and potato (Lakhotia et al, 2014) also found the 23 nt class to be one of the more abundant size classes in their sRNA libraries. The 24 nt sRNA class was less abundant: for the TME3 8 weeks (15.9%), T200 10 weeks (20.5%) and TME3 15 weeks (19.3%), The presence of the 24 nt small RNAs in our libraries may indicate the complexity of the cassava genome as they are mainly siRNAs that are associated with repeats and heterochromatic modifications (Chen et al 2005, Zhang et al, 2008).

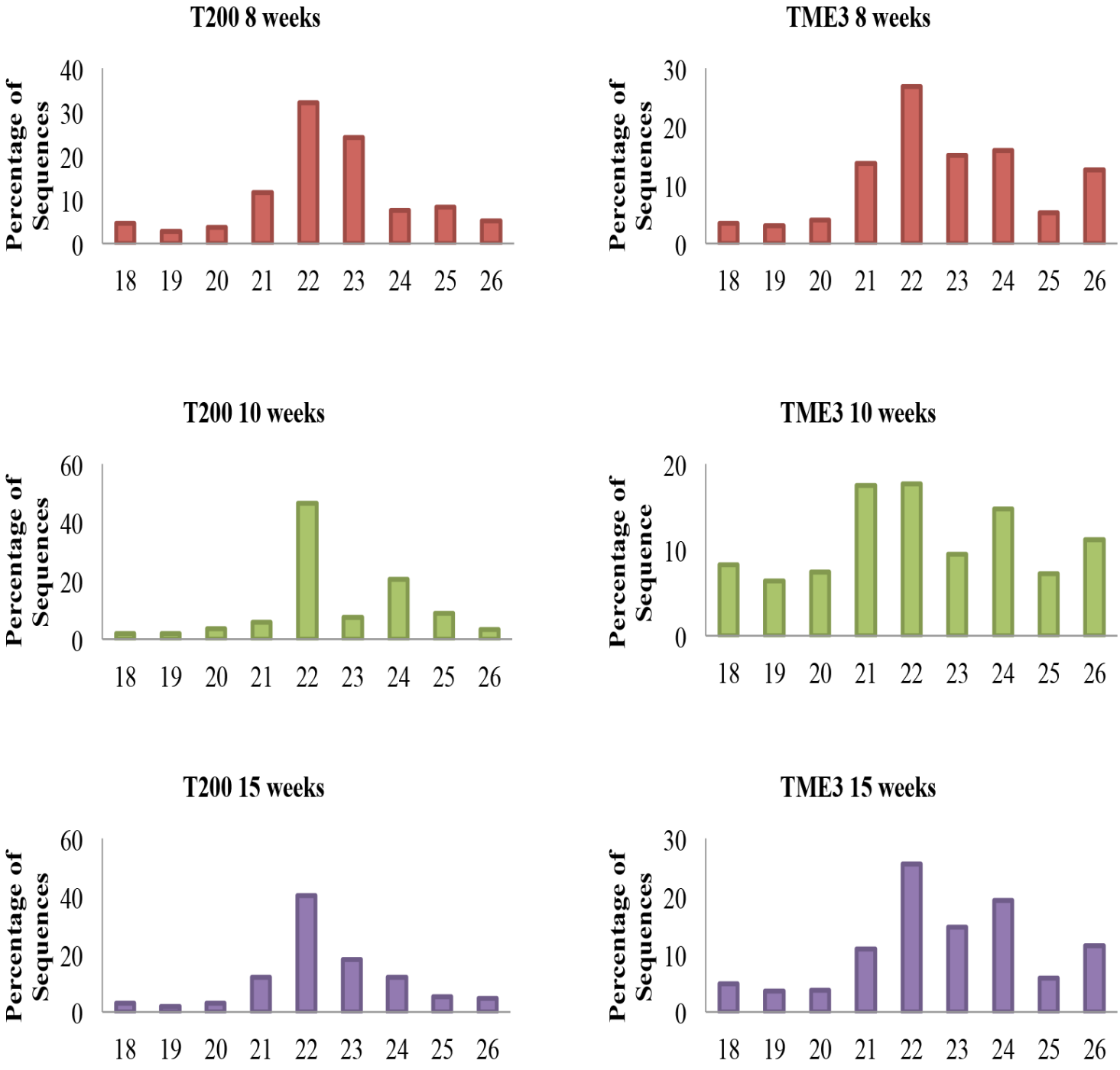


Figure 2.1: Sequence length distribution of cassava small RNAs from T200 and TME3 landraces. Percentage of sequences of 18-26 nt length for each of the six sequenced libraries. The majority of the generated reads were 21 to 24 nt in length.

Identification of conserved miRNAs in cassava

Identification of potential conserved cassava miRNAs from EST and GSS databases

In order to profile and characterize the potential miRNAs in cassava, a comparative genomic approach along with computational and bioinformatics tools was used. In this study, 259 miRNAs in cassava from EST data (S1 Table) and 32 miRNAs from GSS data (S2 Table) were identified. One of the miRNA families, miR2118 identified in this study using the EST database has not been reported in cassava in previous studies (Amiteye et al, 2011; Ballen-Taborda et al, 2013; Pérez-Quintero et al, 2012; Xie et al, 2015; Zeng et al, 2009). The 259 putative cassava miRNAs identified using the EST database belong to 13 families. The largest family was miR408 with 84 individual members and the smallest families were miR170 and miR353 with 1 member each (Fig. 2.2A). The 32 miRNAs identified using the GSS database belong to 7 families. The miR166 family was the largest family with 11 members, while the miR399, miR2275 and miR159 families were the smallest only containing 2 members each (Fig. 2.2B). Also, there were 3 miRNA families that were identified in both EST and GSS databases, miR159, miR166 and miR399. Three miRNAs belonging to the miR166 family were also common to both the EST and GSS databases.

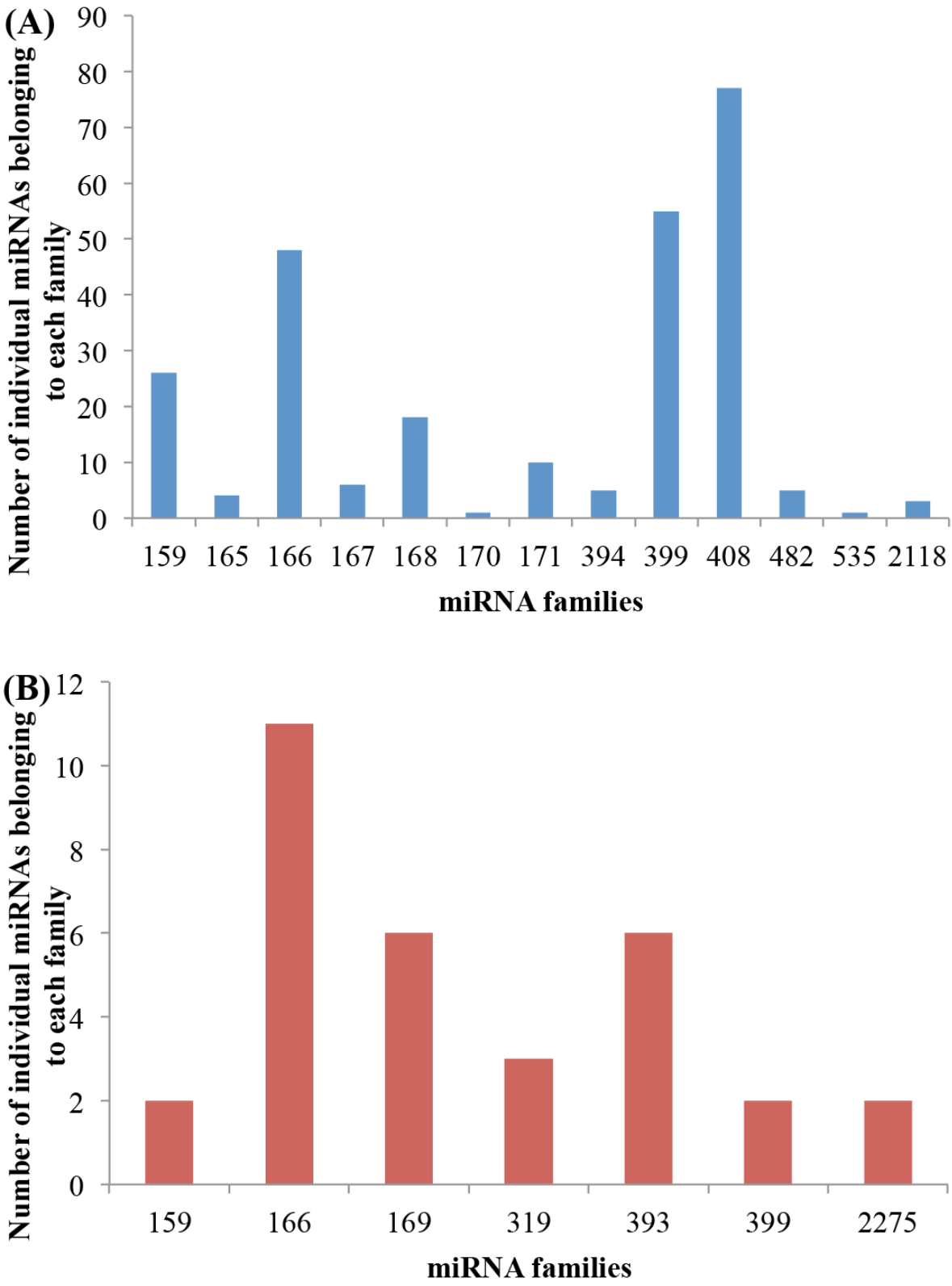


Figure 2.2: The number of individual miRNAs belonging to each miRNA family identified in cassava using (A) EST database and (B) GSS database.

Identification of potential conserved cassava miRNAs from high throughput next generation sequencing (NGS) data

In order to confirm EST and GSS data-derived miRNA results, and identify additional conserved miRNAs in cassava T200 and TME3 landraces, unique sRNA sequences from NGS data at 8, 10 and 15 weeks after transfer from tissue culture to Jiffy® pellets were aligned against the known plant miRNAs deposited in miRBase (Version 21) with a maximum of three mismatches in CLC genomics workbench. A total of 289 potential conserved cassava miRNA sequences belonging to 30 miRNA families were identified from both landraces and developmental stages collectively in this study (S3 Table). Of the 30 miRNA families, the miR166 family was the largest with 33 members. The four families, miR319, miR396, miR482, and miR535 were found to contain 29, 26, 21, and 20 families, respectively. The remaining 25 families contained less than 20 members with 18 of the families containing less than 10 members (Fig. 2.3). It has been previously suggested that most of the miRNAs predicted from EST analysis can be recovered by high throughput NGS (Kwak et al, 2009). In this study 99 (38.2%) of the miRNAs that were identified using the EST database were also identified in the NGS data and are highlighted in green in S1 Table. We were also able to identify 9 (28.1%) of the miRNAs that were identified using the GSS database using the NGS data (highlighted in green in S2 Table).

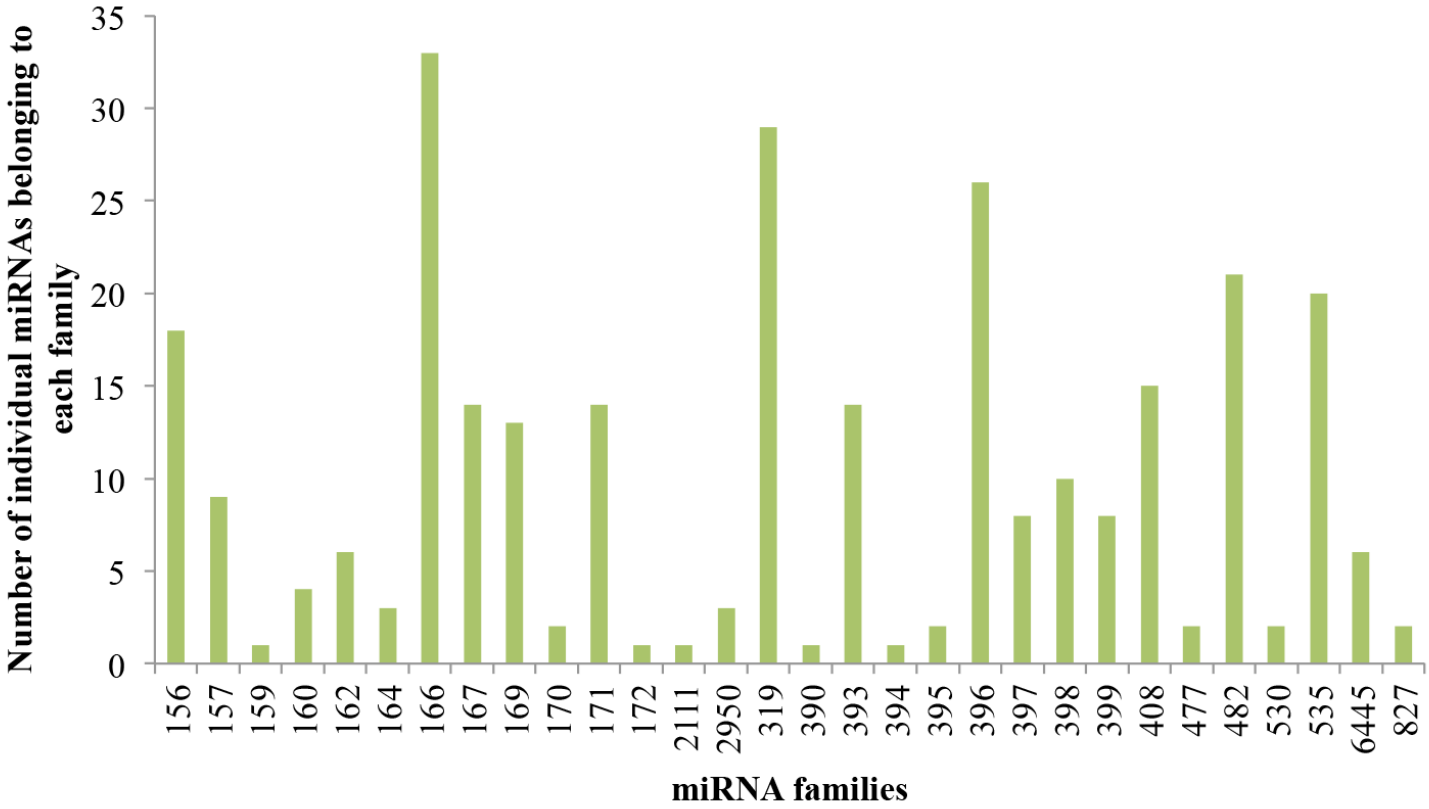


Figure 2.3: The number of conserved individual miRNAs belonging to each miRNA family identified in cassava T200 and TME3 from deep sequencing data.

Characterization of the newly identified conserved cassava miRNAs

EST and GSS data

Characterization of putative candidate miRNAs is a crucial step for their validation as it distinguishes miRNAs from other small RNAs (i.e. tRNAs, rRNAs and mRNAs), as reported earlier (Frazier et al, 2010; Wang et al, 2005). The newly identified potential cassava miRNAs characterized from EST and GSS databases, using accepted criteria/characteristics are summarized in S1 and S2 Tables. The mature miRNA sequences identified from EST database ranged from 18 to 24 nt. The majority (37%) of the miRNAs are 20 nt in length, followed by 21 nt (28%), 19 nt (19%), 18 nt (11%), 22 nt (4%), and 24 nt (1%) (Fig. 2.4 A). The mature miRNAs identified from the GSS database ranged from 18 to 22nt. The majority of the miRNAs either had a length of 19 nt (25%) or 20 nt (25%), followed by 21 nt (22%), 18 nt (19%) and 22 nt (9%) (Fig. 2.4 A). These findings are in agreement

with previously reported studies in other plants species (Amiteye et al, 2011; Frazier et al, 2010; Jones-Rhoades and Bartel, 2004; Panda et al, 2014; Zhang et al, 2008; Wang et al, 2011). The lengths of the potential precursor miRNAs varied from 100 nt to 775 nt for EST-derived (Fig. 2.4B) data and 76 nt to 187 nt for the GSS-derived data (Fig. 2.4B). These results are similar to previous reports in *Arabidopsis*, potato, and rice (Sunker et al, 2005; Zhang et al, 2006a; Zhang et al, 2008). The average A/U% for the pre-miRNAs identified from EST database was 56% and ranged from 42% to 63% and the pre-miRNAs identified from the GSS database also had an average A/U% of 56% and ranged from 47% to 64%. These results are in agreement with the criteria described by Zhang et al (2005), as the A/U% of a potential pre-miRNA should be with 30-70%. In the stem-loop hairpin pre-miRNAs sequences, 66% of the mature miRNAs identified from the EST database were located on the 3' arm, while 34% were located at the 5' arm. The majority (69%) of the mature miRNAs identified from the GSS database were also located at the 3' arm, while 31% were located at the 5' arm.

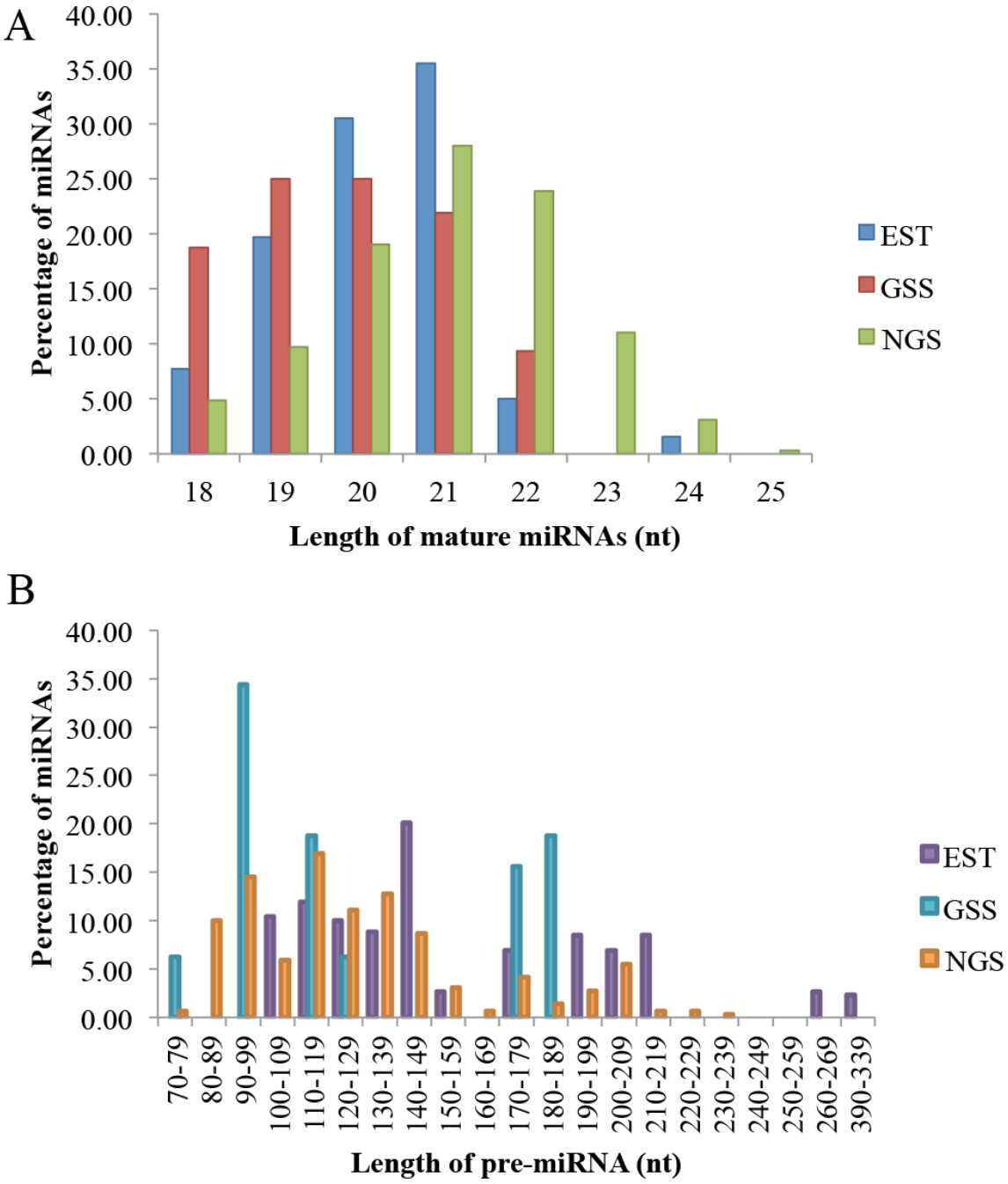


Figure: 2.4 (A) Size distribution of the conserved mature miRNAs and (B) pre-miRNAs identified using the EST database, GSS database and NGS data.

The determination of a hairpin-loop secondary structure of a potential miRNA is not enough for distinguishing miRNAs from other types of non-coding RNAs (Bonnet et al, 2004; Zhang et al, 2005). The minimal folding free energy (MFE) is an important criterion to determine stability of the perfect or near-perfect secondary hairpin structure of pre-miRNAs. The more negative the value of MFE, the higher the thermodynamic stability is of the secondary structure of

the precursor sequence. The MFE of the pre-miRNAs identified from the EST database ranged from -39.9 kcal/mol to -131.3 kcal/mol. The MFE of the pre-miRNAs identified from the GSS database ranged from -26.4 kcal/mol to -95.2 kcal/mol. The minimal folding free index (MFEI) is an important criterion for distinguishing miRNAs from other RNAs. Previous research has suggested that a sequence is more likely to be a potential miRNA if the pre-miRNA had a MFEI more negative than -0.85 kcal/mol (Zhang et al, 2006b). The putative cassava pre-miRNAs identified from the EST database MFEIs ranged from -0.847 kcal/mol to -1.207 kcal/mol. The pre-miRNAs identified from the GSS database MFEIs ranged from -0.964 kcal/mol to – 1.183 kcal/mol. Therefore, the cassava pre-miRNAs identified in this study had more negative MFEIs than other types of RNAs: tRNA (0.64); rRNAs (0.59); mRNAs (0.65) (Zhang et al, 2006b), lending support for their identification as pre-miRNAs.

Next Generation Sequencing Data

A summary of the important characteristics of the miRNAs identified from the NGS data from T200 and TME3 landraces can be found in S3 Table. The identified potential cassava mature miRNA sequences ranged in size between 18 – 25 nt in length. Most of the mature miRNAs were 21 nt in length (28.02%) followed by 22 nt (23.87%), 20 nt (19.03%), 23 nt (7.9%), 19 (9.68%), 18 nt (4.84%), 24 nt (3.11%), and 25 nt (0.34%) (Fig. 2.4 A). Of the 289 identified cassava miRNAs, 170 (58.82%) were found to be located on the 3’ arm of the hairpin secondary structure, while the remaining 119 (41.17%) were located on the 5’ arm. It was also found that miRNAs belonging to the same family miRNA family were not required to be located on the same arm of the pre-miRNA. Most of the identified cassava mature miRNA sequences began with the base uracil (U) (57.78%), which was consistent with previously reported results in other plants [28, 73], due to the high affinity of AGO proteins to bind with U base in the 5’ terminus of mature miRNAs sequences (Mi et al, 2008).

The identified cassava miRNA precursor sequences ranged from 70 – 233 nt in length with an average length of ±127 nt (Fig. 2.4B). The nt composition of these precursor sequences had an average G+C% of 44.35% and A +U% of 55.64% and A +U% ranged from 34.22% - 73.34%, which is consistent with the miRNA

secondary structure filtering criteria by Zhang et al. (2005). The average MFE of the cassava pre-miRNAs was -58.08 kcal/mol. In this study the determined MFEI values of the cassava pre-miRNAs ranged from -0.84 to -1.70 kcal/mol, with an average of -1.03kcal/mol, strongly supporting the validity of these predicted pre-miRNAs in cassava.

RT-PCR validation of data

Seven identified conserved miRNAs: miR169, miR170/171, miR408, miR476 and miR482/2118 were selected for RT-PCR validation studies. All the miRNAs were experimentally validated except for miR482/miR2118 (Fig. 2.5A). This could be due to these miRNAs being present at very low levels in cassava or they could be tissue or developmental stage specific. The experimental validation of these miRNAs provides additional support for the computationally identified miRNAs.

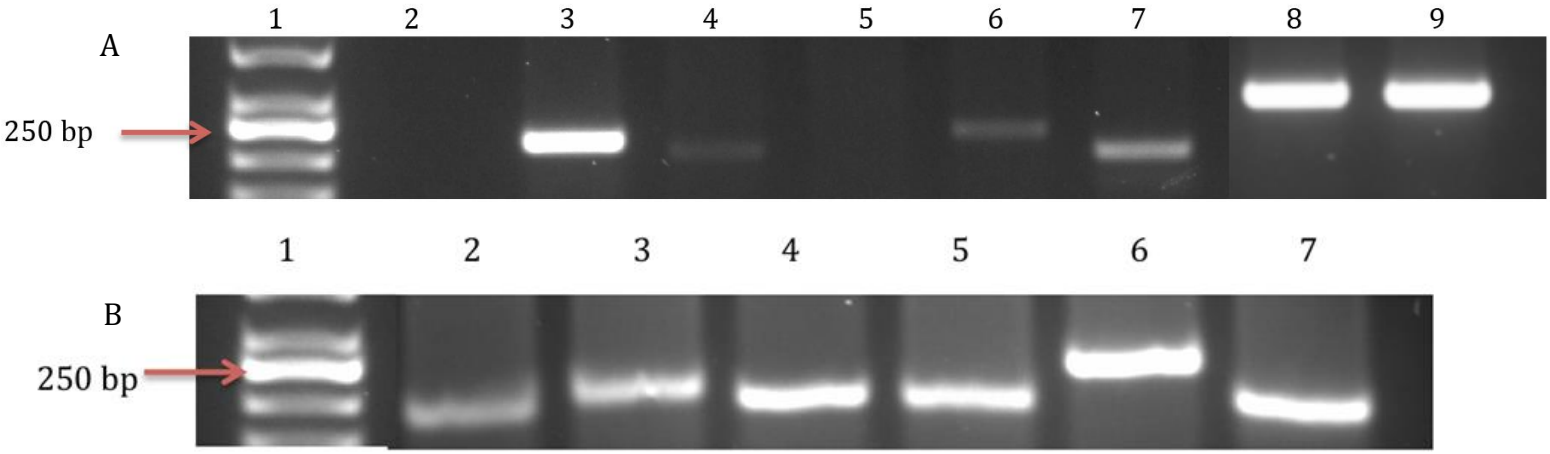


Figure: 2.5 Cassava miRNA RT-PCR expressional validation of the identified (A) conserved cassava miRNAs and (B) novel cassava specific miRNAs. The product of each sample was separated on a 2% agarose gel. (A) Lane 1 = MWM (50bp); Lane 2 = empty; Lane 3 = miR169A; Lane 4 = miR169B; Lane 5 = miR482/miR2118; Lane 6 = miR408; Lane 7 = miR170/miR171; Lane 8 = miR476A; Lane 9 = miR476B. (B) Lane 1 = MWM (50bp); Lane 2 = mes-6; Lane 3= mes-10; Lane 4 = mes-12; Lane 5 = mes-22; Lane 6 = mes-25; Lane 7 = mes-28.

Identification of miRNA isoforms

Small non-coding RNAs such as miRNAs were initially thought to have a specific sequence of a defined length. Identification of more miRNAs from different species has revealed that there is variation in pre-miRNA processing. A single miRNA locus can give rise to multiple distinct isomiRs that differ in their length and sequence composition (Neilsen et al, 2011). In the conventional plant miRNA biogenesis pathway, the 5' and 3' ends are specified by consecutive cleavage events of the primary transcript by the ribonuclease Dicer-like 1 (DCL) (Bartel, 2004). In this study in cassava, of the 289 conserved miRNAs identified in T200 and TME3 using NGS data, 208 were isomiRs. These 208 isomiRs belonged to 27 families. The most frequently observed type of isomiR in both plants and animals is the 3' isomiRs, in terms of both number of miRNAs displaying these variations and their overall abundance (Burroughs et al, 2010; Newman et al, 2011). Seventy-nine (±38%) out of the 208 isomiRs in this study were also found to be 3' isomiRs.

IsomiRs are categorised into three main classes: 5' isomiRs, 3' isomiRs, and polymorphic isomiRs, with 5' and 3' isomiRs subclassified into templated or non-templated modifications (Burroughs et al, 2010). Heterogeneity in length can arise from the imprecise cleavage by DCL, in which case the miRNA sequences will match the parent gene but will vary in length, a situation referred to as 'templated'. Length heterogeneity can also arise by exonucleases 'nibbling' off the end, which produces a shorter templated product, which is referred to as sub-templated. They can also arise from post-transcriptional addition of one or more bases, which is referred to as super. The addition of these bases can result in the end matching the parent gene, templated, or the end may not match the parent gene and is known at non-templated. Polymorphic isomiRs harbour different internal nucleotide sequences, but these are relatively rare (Burroughs et al, 2010). Forty-three of the 3' isomiRs from T200 and TME3 were classified as sub-templated and 36 were classified as super-templated. The super templated 3' isomiRs were further divided into 33 super templated and 3 super non-templated. The second largest class was the 5' isomiRs. There were 45 (±22%) 5' isomiRs divided into 16 sub templated and 29 super templated. Twelve polymorphic isomiRs were identified with either 1 or 2 nucleotides involved in a mismatch between the

isomiR and the reference miRNA. There were also 52 identified isomiRs that had changes in length at both the 5' and 3' ends but contained no sequence differences between itself and the reference miRNA. Twenty isomiRs contained both length differences at both ends as well as polymorphic changes.

IsomiRs were also identified using the EST and GSS databases. For the miRNAs identified from the EST database, 101 of the potential cassava miRNAs were completely identical to their reference miRNA, while 158 were variants of their reference miRNA and are isomiRs (bolded in S1 Table). The largest class of isomiRs was the 3' sub templated (31.6%), followed by the polymorphic class with (24.1%), the 5' sub templated class (22.8%), the 3' sub polymorphic class (12%), 5' sub polymorphic class (7%) and the 5' and 3' sub templated class (2.5%). For the miRNAs identified from the GSS database, 12 of the 32 were identical to their reference miRNAs and 10 were isomiRs (bolded in S2 Table). Again the largest class was the 3' sub templated class (60%) followed by the 5' sub templated class (20%), 3' sub polymorphic class (10%) and the 5' and 3' sub templated and polymorphic classes with 5% each. No super isomiRs were identified in the EST and GSS databases. The main processing steps in the canonical miRNA biogenesis pathway are the sequential cleavage steps catalysed by the Dicer endonucleases (Kim, 2005), which are a source of templated miRNA variation. However, the fact that variability is most commonly associated with the 3' end suggests that other processing activities contribute to the distribution pattern. Insights form Argonaute (AGO) crystallographic studies for example, indicating that the 5' ends of the microRNAs are buried within the MID domain, whereas the 3' ends extend from the PAZ domain and are therefore available to exonucleolytic attack (Schirle and Macrae, 2012), causing shortening. Also, most nucleotidyl transferases that catalyse the addition of nucleotides are 5' – 3' polymerases, thereby causing an abundance of nontemplated nucleotide extensions at 3' rather than 5' ends (Martin and Keller, 2007).

Identification of novel miRNAs

Using the miRCat program in the UEA small RNA workbench (Stocks et al, 2012), 39 novel cassava-specific miRNAs belonging to 29 families were identified

and were named mes-1 to mes-29 (Table 2; S4 Table). The largest family was mes-28 containing 3 members. Seven families contained 2 members and the remaining 22 families only contained a single member (Fig. 2.6). The low abundance (< 3) of miRNAs in each family (Fig. 2.6) strongly suggests that these are cassava-specific miRNAs. Montes et al. (2014) have suggested from their study of miRNAs in vascular plants that the majority (92 to 99%) of species-specific new miRNAs occur in low abundance of less than 10 RPM. The majority of the identified potential novel miRNAs in this study were 22 nt in length (47%) followed by 21 nt (38%), 24 nt (9%), 20 nt (4%), and 19 nt (2%) respectively (Fig. 2.7 A). Most of the novel mature miRNA sequences began with the base uracil (U), which is consistent with previously reported results in other plants (Dhandapani et al, 2011), due to the high affinity of AGO proteins to bind with U base in the 5' terminus of mature miRNAs sequences (Mi et al, 2008). The average length distribution of the predicted novel miRNA precursor sequences was 120.5 nt (Fig. 2.7 B). The mes-16 family exhibited the shortest precursor length of 67 nt, whereas the mes-7 and mes-8 family members exhibited the longest precursor length of 212 nt. The nt composition of the newly identified potential cassava novel miRNA precursor sequences had an average A+U content of 55.1% and G+C content of 44.8%. The average minimal folding free energy (MFE) of the potential cassava novel pre-miRNAs was -58.9 kcal/mol. In this study, the MFEI for the novel miRNA precursors ranged from -0.89 to -1.56 kcal/mol with an average of -1.12 kcal/mol, which agrees with the important rule that plant miRNA precursors should have a MFEI more negative than -0.85 kcal/mol.

Table 2. Novel miRNA families identified in cassava T200 and TME3 and at 3 stages of development post acclimatization from tissue culture explants.

*Novel miRNA families identified		**No. novel miRNA families identified		
T200	TME3	at 3 developmental stages		
		8 wks	10 wks	15 wks
mes-1	*mes-2*	mes-8	mes-7	mes-1
mes-3	mes-3	*mes-12*	*mes-12*	mes-2
mes-5	*mes-4*	*mes-13*	*mes-13*	mes-3
mes-6	mes-5	mes-20	mes-14	mes-4
mes-8	*mes-7*	*mes-22*	mes-15	mes-5
mes-12	mes-8	mes-24	mes-17	mes-6
mes-13	*mes-9*	*mes-25*	*mes-22*	mes-9
mes-14	*mes-10*	*mes-28*	*mes-25*	mes-10
mes-15	*mes-11*	mes-29	*mes-28*	mes-11
mes-17	mes-12			mes-12
mes-19	*mes-16*			mes-13
mes-20	mes-17			mes-16
mes-22	*mes-18*			mes-17
mes-23	mes-19			mes-18
mes-25	*mes-21*			mes-19
mes-27	mes-22			mes-20
mes-28	*mes-24*			mes-21
	mes-25			*mes-22*
	mes-26			mes-24
	mes-28			*mes-25*
	mes-29			mes-26
				mes-27
				mes-28

*miRNA families that are in italics are the miRNA families that are specific to either cassava landrace T200 or TME3

**miRNA families that are in italics are the miRNA families that are common to both cassava landraces and all three developmental stages

doi:10.1371/journal.pone.0147251.t002

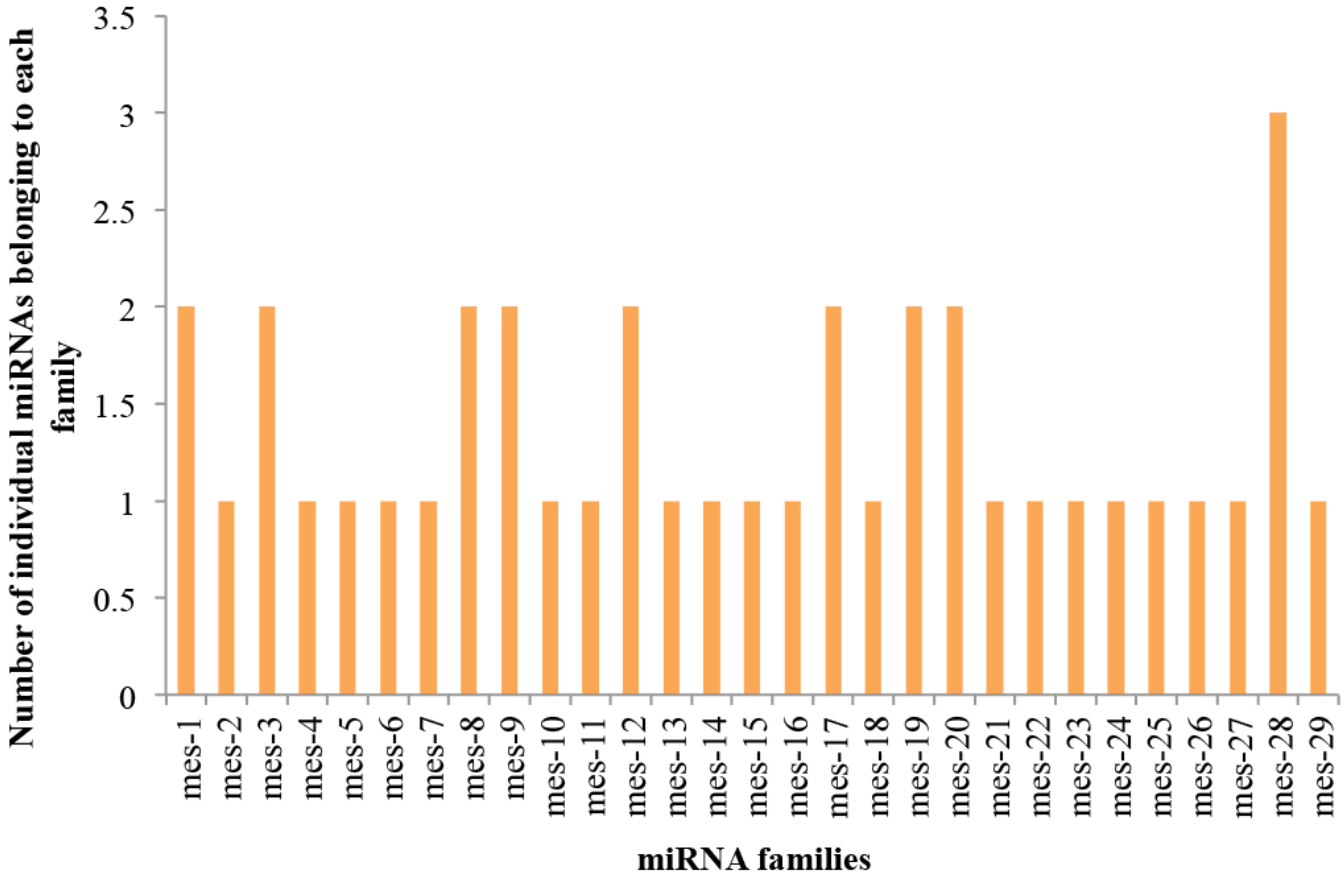


Figure 2.6: The number of individual cassava-specific miRNAs belonging to each novel miRNA family identified in cassava using NGS data.

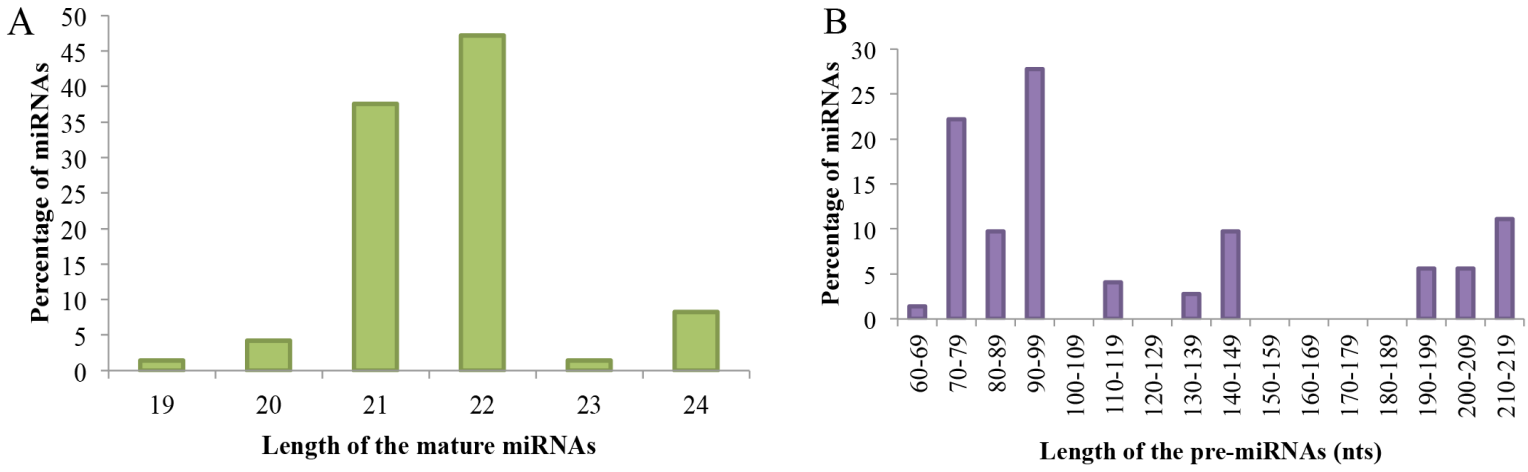


Figure 2.7: Size distribution of the (A) novel mature miRNAs and (B) pre-miRNAs identified in the deep-sequencing data

It was observed that different miRNAs belonging to the same miRNA family were produced from the same scaffold and hairpin. mes-17a and mes-17b were both produced from the same hairpin that is located in scaffold 00631. This was

also observed for mes-19a and mes-19b that were produced from the same hairpin located in scaffold 09876. It was also found that miRNAs belonging to different miRNA families could be produced from the same hairpin. Mes-20a, mes-20b and mes-21 were all produced from the same hairpin located in scaffold 01701. However, mes-6, mes-22 and mes-29 were produced from different hairpins but all the hairpins were located in scaffold 03581. This was also observed for mes-8 and mes-9 that had two cases of members being derived from different hairpins that were located in the same scaffold, scaffold 03429 and scaffold 06557 (S4 Table).

Not all the novel miRNAs identified in this study were identified in both cassava landraces at all three growth stages post transfer of explants from tissue culture (8, 10 and 15 weeks). Only 17 miRNAs were identified in the T200 landrace and 8 were T200 specific. In TME3 21 of the novel miRNA families were identified and 12 of them were TME3 specific. These results are summarized in Table 2. In both the 8 and 10 week samples, 12 novel miRNA families were identified. The 15 week samples had 23 of the novel miRNA families being identified. Only 5 of the identified novel miRNAs were reported for both cassava landraces and in all three developmental stages post explant establishment, namely mes-12, mes-13, mes-22, mes-25 and mes-28 (Table 2). Notwithstanding that different spatial expression patterns of miRNAs may occur, T200 and TME3 were cultivated under the same growth conditions in the growth chamber, and differences in expression between the landraces are therefore likely a consequence of temporal factors and genotype.

A small subset of the newly identified novel cassava miRNAs were experimentally validated using RT-PCR. The randomly selected 6 novel miRNAs: mes-6, mes-10, mes-12, mes-22, mes-25 and mes-28 were used for the RT-PCR validation studies. All 6 novel miRNAs were experimentally validated (Fig. 2.5B). This experimental validation of these miRNAs strengthens the expressed nature for computationally identified miRNAs.

Comparison of current study with previous studies involving cassava miRNA identification

The first homology-based comparative genomics cassava study was by Amiteye et al. (2011). They used 212 previously reported *Arabidopsis thaliana* mature miRNA and precursor sequences that were available in miRBase Version 14 as a reference for a BLASTn search against the publicly available cassava EST database at NCBI (Table 3). This approach resulted in the identification of 35 individual miRNAs belonging to 17 families and their corresponding target genes in cassava that were also conserved in other plant species. However, the ESTs representing 7 of these miRNA families produced foldback structures that showed more than 3 nts not involved in canonical base pairing within a loop or bulge in the mature miRNA:miRNA* duplex. These miRNA families should not have been considered as true miRNAs as they do not follow all the miRNA identification criteria by Ambros et al. (2003) and Zhang et al. (2006b). These families are highlighted in italics in Table 4. Also, there were also 3 miRNA families that had a MFEI less negative than -0.85 kcal/mol. Plant pre-miRNAs should have a MFEI more negative than -0.85 kcal/mol (Zhang et al, 2006b) and these 3 families should also not be considered as true miRNAs (bolded in Table 4).

Table 3. Comparison of four previous studies and this study on miRNA identification in cassava.

	Amiteye *et al*., 2010	Perez-Quintero *et al*., 2012	Patanan *et al*., 2013	Ballen-Taborda *et al*., 2013	Xie *et al*., 2014	This study	T200 & TME3 This Study
Method	EST homology based	Deep-sequencing (Illumina platform)	Genome-wide homology based search	Deep-sequencing (Illumina platform)	Deep-sequencing (Illumina platform)	EST and GSS homology Based search	Deep-sequencing (Illumina platform)
Conditions	N/A	Infected and unifected with *Xanthamonas axonopodis pv*.	N/A	Heat and drought-like conditions	Severe and moderate chilling stresses	N/A	Three developmental stages
Cassava landrace/ cultivar	N/A	MBRA685	AM5602	TAI16	SC124	N/A	T200 and TME3
Part of plant	N/A	Leaves and stem	N/A	Leaves	Leaves and Roots	N/A	Leaves
miRBase version	14	16	16	19	20	20	20
Reference	*Arabidopsis thaliana* miRNAs from miRBase only	All mature Viridiplantae miRNas	All mature Viridiplantae miRNas	All mature Viridiplantae miRNas	All mature Viridiplantae miRNas	All mature Viridiplantae miRNas	All mature Viridiplantae miRNas
Number of conserved miRNAs	35	114	169	60	163	259 and 32	289
Number of conserved miRNA families	17*	56	34	27	32	13 and 7	30
Number of Novel cassava-specific miRNAs	0	12	0	821	17	N/A	39
Number of Novel cassava-specific miRNA families	N/A	N/A	N/A	N/A	N/A	N/A	29

* ESTs representing 7 of the families could not form secondary stem-loop structure and 3 families had MFEI <0.85

doi:10.1371/journal.pone.0147251.t003

Table 4. miRNA families identified from miRBase V.20 in four previous studies compared with the miRNA families identified in this study.

						This Study		
Amiteye *et al*., 2010	Perez-Quintero *et al*., 2012	Balle-Taborda *et al*., 2013	Patanan *et al*., 2013	Xie *et al*., 2013	miR BAse V.20	EST	GSS	Deep-sequencing
miR156	miR1030	miR156	miR1446	MIR1446	miR1446	miR159	miR159	MIR156
miR157	miR156	miR157	miR156	MIR156	miR156	miR165	miR166	MIR157
miR159	miR159	miR159	miR159	MIR159	miR159	miR166	miR169	MIR159
miR160	miR160	miR160	miR160	MIR160	miR160	miR167	miR319	MIR160
miR162	miR164	miR164	miR162	MIR162	miR162	miR168	miR393	MIR162
miR166	miR166	miR166	miR164	MIR164	miR164	miR170	miR399	MIR164
miR167	miR167	miR167	miR166	MIR166	miR166	miR171	miR2275	MIR166
miR168	miR169	miR169	miR167	MIR167	miR167	miR394		MIR167
miR171	miR170	miR171	miR168	MIR168	miR168	miR399		MIR169
miR172	miR172	miR172	miR169	MIR169	miR169	miR408		MIR170
miR390	miR1863	miR2950	miR171	MIR171	miR171	miR482		MIR171
miR394	miR2111	miR319	miR172	MIR172	miR172	miR535		MIR172
miR397	miR390	miR390	miR2111	MIR2111	miR2111	miR2118		MIR2111
miR398	miR393	miR393	miR2275	MIR2275	miR2275			MIR2950
miR399	miR394	miR394	miR2950	MIR2950	miR2950			MIR319
miR408	miR395	miR395	miR319	MIR319	miR319			MIR390
mi414	miR396	miR396	miR390	MIR390	miR390			MIR393
	miR397	miR397	miR393	MIR393	miR393			MIR394
	miR403	miR398	miR394	MIR394	miR394			MIR395
	miR472	miR399	miR395	MIR395	miR395			MIR396
	miR477	miR408	miR396	MIR396	miR396			MIR397
	miR482	miR477	miR397	MIR397	miR397			MIR398
	miR535	miR482	miR399	MIR398	miR399			MIR399
	miR827	miR535	miR403	MIR399	miR403			MIR408
		miR6445	miR414	MIR403	miR408			MIR477
		miR827	miR473	MIR408	miR477			MIR482
		miR828	miR477	MIR477	miR482			MIR530
			miR482	MIR482	miR530			MIR535
			miR529	MIR530	miR535			MIR6445
			miR530	MIR535	miR827			MIR827
			miR535	MIR827	miR828			
			miR827	MIR828				
			miR828					
			miR845					

doi:10.1371/journal.pone.0147251.t004

The study by Perez-Quintero et al. (2012) addressed the role of miRNAs in the *Manihot esculenta-Xanthomonas axonopodis* pv. manihotis (*Xam*) interaction. NGS was used for analysing small RNA libraries from cassava leaf and stem tissue infected and uninfected with *Xam*. A full repertoire of cassava miRNAs was characterized, which included 114 individual conserved miRNAs belonging to 56 families and 12 novel cassava-specific miRNAs. This study used NGS to identify miRNAs in cassava and all available mature Viridiplantae miRNAs obtained from miRBase release (Version 16) were used as the

reference when conducting the BLASTn search against the deep-sequencing reads (Pérez-Quintero et al, 2012) (Table 4). This was also the first study to identify cassava-specific miRNAs. A subsequent report by Patanun et al. (2013), based on homology-based computational prediction, aimed to extend the cassava miRNA knowledge by using all reported plant miRNAs deposited in miRBase (V.16) to search against the cassava genome provided by Phytozome (http://www.phytozomenet/cassava) (Table 4). The cassava genome available at Phytozome was generated from the cassava cultivar AM5602. This study resulted in the identification of 169 individual conserved miRNAs belonging to 34 families in cassava (Table 4).

Ballen-Taborda et al. (2013) used NGS and different bioinformatics methods to identify potential cassava miRNAs expressed in different tissues of the cassava cultivar TAI16 subjected to abiotic stress (heat and drought conditions), and the authors identified 821 novel miRNAs, but these were not submitted to miRBase. In comparison to Ballen-Taborda et al., a NGS study by Xie et al. (2015), profiling miRNAs and target mRNA genes from cassava cv. SC124 plants that experienced severe and moderate chilling stresses, identified 163 individual conserved miRNAs belonging to 32 families and 17 cassava-specific miRNAs (Table 4). Our study combined a homology-based computational prediction approach using the publicly available cassava EST and GSS databases at NCBI as well as a NGS of two different cassava landraces at three developmental stages (8, 10 and 15 weeks). Available Viridiplantae mature miRNAs from the updated miRBase V.21 were used as the reference for a BLASTn search for both approaches. In comparison to the previously mentioned studies above, NGS data from our study unveiled 289 individual miRNAs conserved in other plant species and 39 new previously unreported putative cassava-specific miRNAs (Table 4). Using the EST cassava database, 200 (77.2%) of the identified conserved individual miRNAs had not been reported in the above previous studies. Using the GSS cassava database, 22 (68.8%) of the identified individual conserved miRNAs had not been reported in the above previous studies, while NGS data revealed 230 (79.6%) of the individual conserved miRNAs had not been reported in the above previous

studies. Additionally, we identified a miRNA family (miR2118) using the EST cassava database (underlined in Table 4) that have not been reported in the above previous studies. For the novel cassava-specific miRNAs identified in this study, only one had been reported in a previous study. Therefore, 98.6% of the novel miRNAs identified in this study have not been previously reported. The miRNAs that were reported in the previous studies are bolded in S2-5 Tables. From Tables 3 and 4, it is clear that the results of miRNA discovery studies in cassava varied depending on which miRNA identification method was used, which cultivar/landrace was studied, and not unexpectedly, if the plant underwent any biotic or abiotic stresses.

Evolution of the identified conserved miRNA families in cassava

A small set of miRNAs has been detected in several major lineages of land plants (Axtell and Bowman, 2013). Twenty-one miRNA families (miR156, miR159, miR160, miR162, miR164, miR166-169, miR171, miR172, miR319, miR390, miR393-399, and miR408) seem to be universally expressed among diverse plant species (Fig. 2.8). A subset of these miRNA families is more ancient, because it is also present in gymnosperms, lycopods and bryophytes (Axtell and Bowman, 2013). Eight miRNA families (miR156, miR159/319, miR160, miR166, miR171, miR408, miR390/391, and miR395) have been identified in the common ancestor of all embryophytes. The miR396 family is present in the common ancestor of all tracheophytes (vascular plants). The miR397 and miR398 families were acquired in the common ancestor of all spermatophytes (seed plants). Ten families (miR162, miR164, miR167, miR168, miR169, miR172, miR393, miR399 and miR827) are present in all angiosperm lineages (Fig. 2.8). All of the above miRNA families were identified in cassava T200 and TME3 in this study.

Figure 2.8: Evolutionary conservation of the thirty-four miRNA families identified in cassava within the plant species belonging to the Magnoliophyta Division (also known as Angiosperms) reported in miRBase (v. 21). Closely related species are shown in the same colour.

Populus trichocarpa and *Ricinus communis* (castor bean) had 79.4% and 61.7% of miRNA families in common with cassava (Fig. 2.8). This was expected as both these Euphorbiaceous species are closely related to *Manihot esculenta*, and *Populus trichocarpa* has been well studied in terms of its micronome, with 401 mature miRNAs available on miRBase V.21. However, interestingly, *Populus euphratica* is also closely related to cassava but did not have a single miRNA family on common with cassava. This could be due to the lack of information available about its micronome as there are only 4 mature miRNAs available on miRBase (V. 21) for this plant species. *Glycine max* is a very well-studied plant species with 639 mature miRNAs present on miRBase V.21 but while it is not closely related to cassava, interestingly, it shares 73.5% of the miRNA families with cassava.

At least nine families (miR441, miR444, miR818, miR821, miR1435, miR2118, miR2275 and miR582) likely arose in the monocot lineage (Cuoerus et al, 2011). However, in this study both the miR2118 and miR2275 were identified in cassava. MiR2118 has also been reported in 7 other dicot plant species in miRBase (V.21) (Sun, 2012), including the well-studied *Glycine max* species. MiR2275 has also been reported in cassava by Patanun et al. (2013) and it has been accepted by miRBase as a true cassava miRNA. In rice both of these miRNA families have been implicated in secondary siRNA production. MiR2118 mediates the recruitment of 21 nt secondary siRNA-generating machinery and miR2275-targeted transcripts generate 24 nt siRNAs (Johnson et al, 2009).

Identification of targets

The prediction of miRNA targets is a significant step for validation of the newly identified cassava miRNAs. Most plant miRNAs have perfect or near perfect complementarity with their targets to regulate gene expression at post-transcriptional level (Bartel, 2004; Schwab et al, 2005). Based on this mechanism of miRNAs in plants, a homology search-based method was used for miRNA target prediction in cassava using psRNATarget server. The newly identified conserved and novel miRNAs in cassava were used as queries in the psRNATarget to predict

the potential mRNA targets. Targets for conserved and novel miRNAs are detailed in S5 and S6 Tables, respectively.

Identification of targets for conserved miRNAs

Endogenous miRNAs act as negative regulators of gene expression by facilitating the cleavage of target mRNAs or by repressing their translation. The cleavage of target mRNAs seems to be a prime mode of gene regulation in plants (Sunkar et al, 2005). In this study, 77.5% of the targets identified in cassava were repressed through cleavage while on 22.4% were repressed through translation. A total of 262 targets were identified for 32 of the conserved miRNA families identified in cassava using EST and GSS databases, and NGS data. The miR156 family had the most targets (41) followed by the miR166 family with 29 targets and miR396 with 20 targets (Fig. 2.9). Only 111 of the targets were annotated with a known function. Transcription factors are important components in the transcription process and play an important role in a variety of biological functions, and therefore it was no surprise to observe 4 miRNA families (miR166, miR169, miR319 and miR408) in cassava were associated with 9 transcription factors. Several studies have indicated that miRNAs directly target the transcription factors that regulate plant development as well as specific genes that control various metabolic processes (Fujita and Iba, 2008). The miR169 family targets the CCAAT-binding transcription factor and the nuclear transcription factor Y subunit A-8. The CCAAT-binding transcription factor is a sequence-specific DNA binding transcription factor that is involved in double fertilization forming a zygote and endosperm. The nuclear transcription factor Y subunit A-8 stimulates the transcription of various genes by recognising and binding to a CCAAT motif in promoters (Gusmaroli et al, 2002). The miR408 family targets a probable WRKY transcription factor 71 that interacts specifically with the W box (5'-(T) TGAC [CT]-3'), which is a frequently occurring elicitor-response cis-acting element (Ulker and Somssich, 2004).

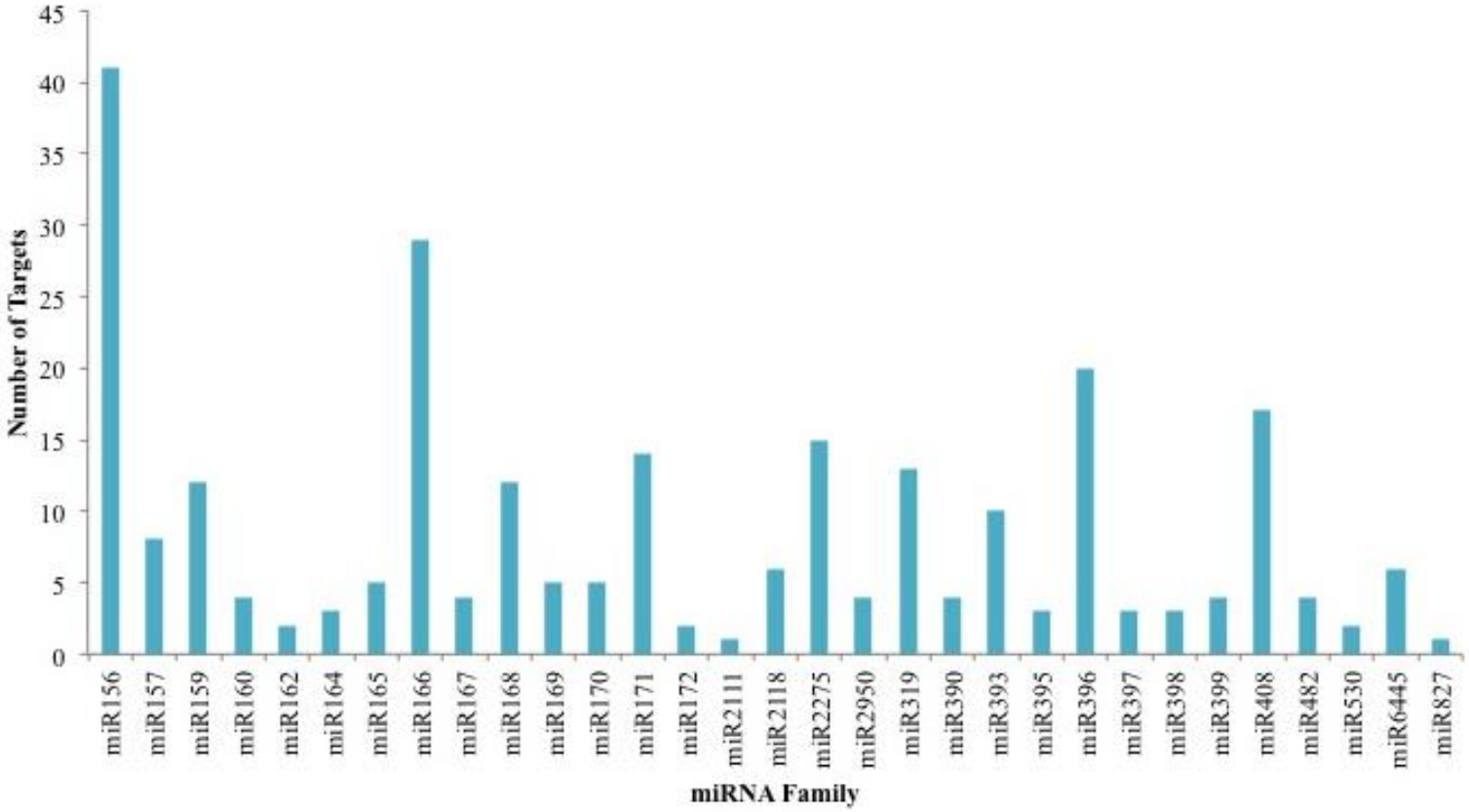


Fig. 2.9 Number of targets predicted for each identified conserved miRNA in cassava

The miR168 family, targeting AGO1, was also identified in cassava in this study. AGO1 is involved in RNA-mediated post-transcriptional gene silencing (PTGS). It is the main component of the RNA-induced silencing complex (RISC) that binds to a short guide RNA such as microRNA (miRNA) or small interfering RNA (siRNA) (Baumberger and Baulcombe, 2005). RISC then uses the mature miRNA or siRNA as a guide for slicer-directed cleavage of homologous mRNAs to repress gene expression. AGO1 mainly associates with miRNAs of 21 nt in length and preferentially recruits small RNAs with a 5' terminal uridine (Mi et al, 2008; Takeda et al, 2008). It also associates with 22 nt miRNAs to trigger RDR6-dependent secondary siRNAs biogenesis (Cuperus et al, 2010). This pathway amplifies silencing by using the target RNA as substrate to generate secondary siRNAs. It also binds to miR168, which targets its own mRNA for repression, establishing a homeostatic regulatory loop. AGO1 is involved in antiviral RNA silencing by contributing to viral RNA clearance (Zhang et al, 2006a). This protein is also essential for multiple processes in development, including proper development of leaves and floral organs, and formation of axillary meristems. Like AGO10, it is required for stem cell function and organ polarity (Bohmert et al, 1998; Unver et al, 2009).

Identification of targets in cassava-specific novel miRNAs

In this study, 37 putative targets were predicted for 17 of the novel miRNA families identified in cassava (Fig. 2.10). Mes-24 had the most targets, 6, followed by mes-20 and mes-6 with 4 targets, and mes-12 and 18 with 3 targets. The remaining novel miRNA families had 1 or 2 targets each.

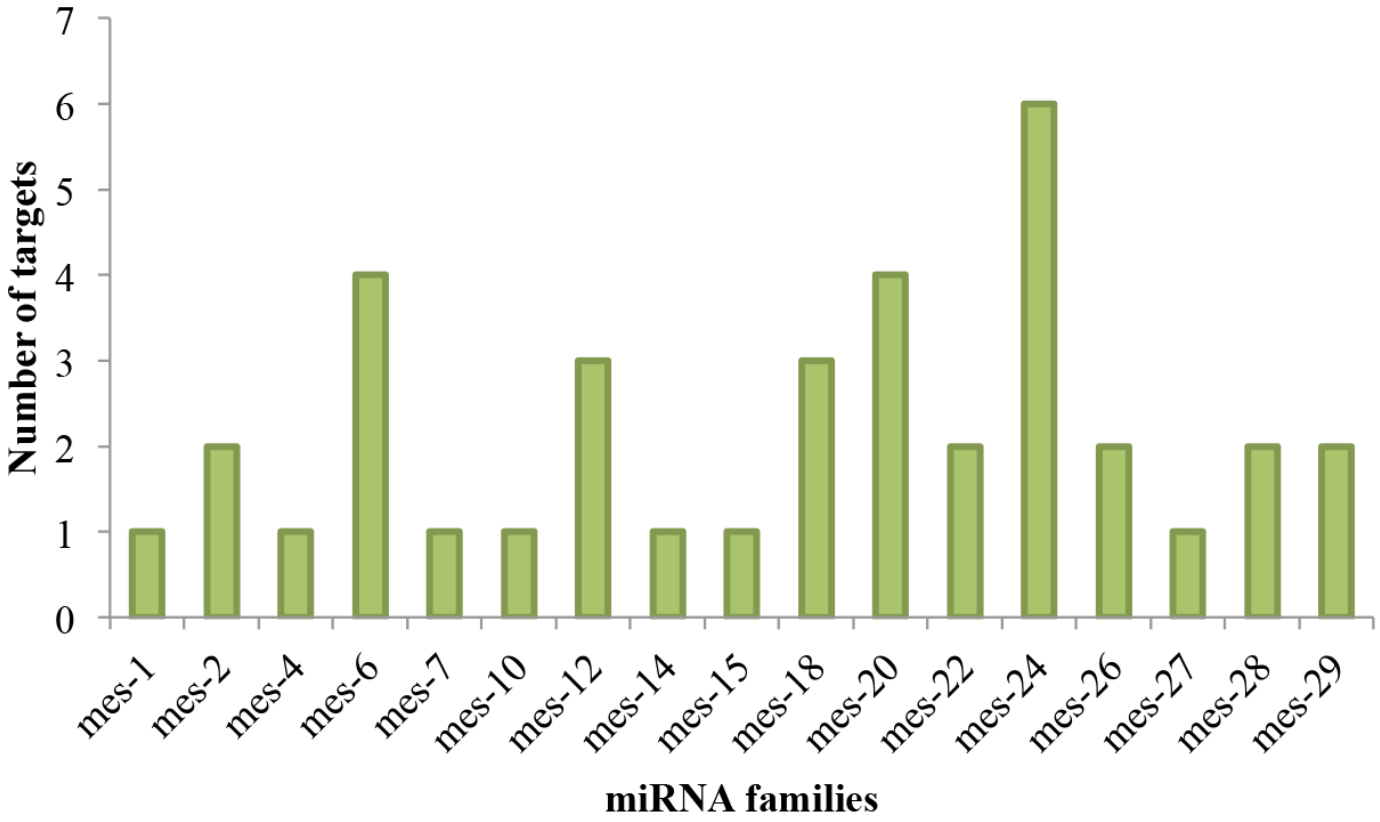


Figure 2.10: The number of targets predicted for the novel miRNA families identified in cassava.

Some of the targets identified for the novel miRNAs were known targets for some conserved miRNA families. An example is the mes-24 family with one of its targets being the Myb1 transcription factor. This transcription factor is a well-documented target for the miR159 family. An interesting target was the NBS-LRR Resistance protein RGH1, which is known to be involved in defence mechanisms in plants against pathogens. The cleavage of target mRNAs seems to be a prime mode of gene regulation in plants (Sunkar et al, 2005). The majority of the novel

miRNAs (70.5%) use cleavage to repress their targets, while 29.5% use translation as their repression mechanism.

GO annotations

To further understand the functions of the identified conserved and novel cassava miRNAs, the identified targets underwent GO term analysis. The results of this analysis are summarised in S7 Table and S8 Table. GO term analysis allows the miRNA-gene regulatory network to be characterised in terms of molecular function, biological process and cellular component. The collective targets in T200 and TME3 from 3 developmental stages for the conserved miRNAs were involved in 101 molecular functions and 192 biological processes. The targets for the novel miRNAs were involved in 26 molecular functions and 37 biological processes. The top ten of the GO terms for each of the three GO categories for the targets identified for the conserved and novel miRNAs are represented in Fig. 2.11 and Fig. 2.12. When comparing these results in the figures, the targets of the conserved and novel miRNAs had six molecular function terms, three biological processes terms and seven cellular components in common. These results suggest, not unexpectedly, that the cassava miRNAs are involved in various biological processes such as oxidation-reduction process, response to biotic and abiotic stresses, regulation of transcription and translation, transport, growth and development, and metabolism.

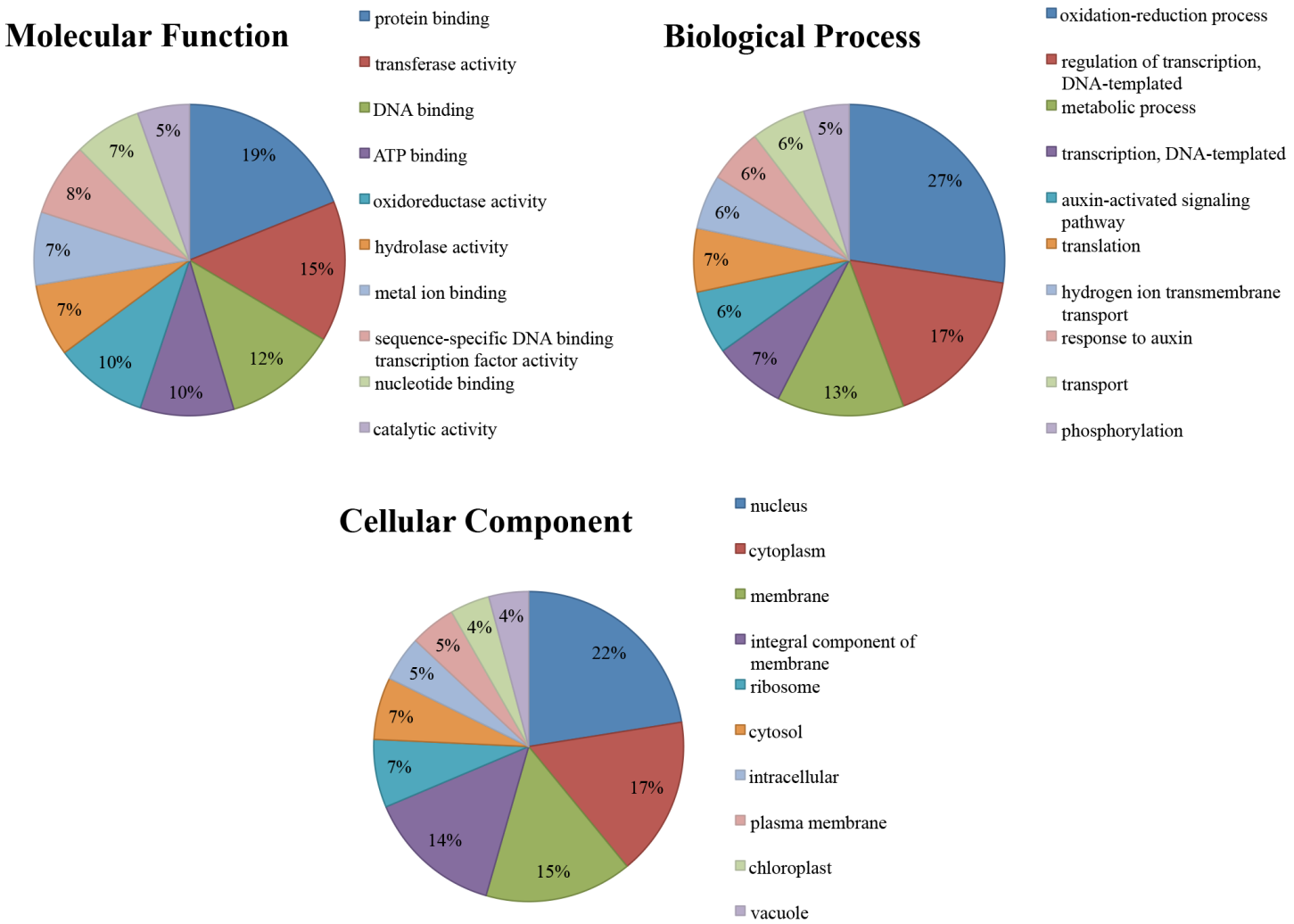


Figure 2.11: The top ten GO terms for the three GO categories for the targets identified for the conserved cassava miRNAs.

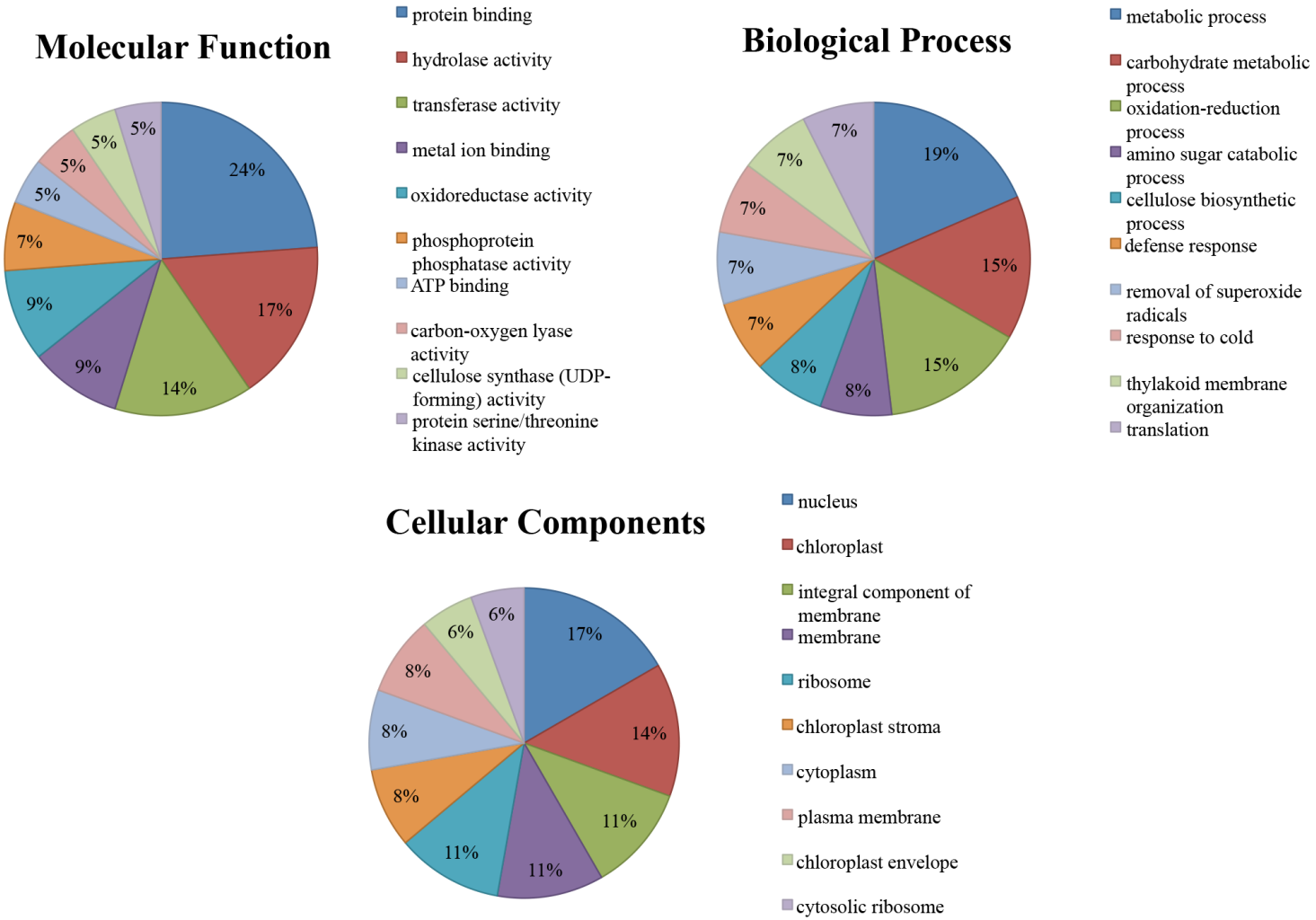


Figure 2.12: The top ten GO terms for the three GO categories for the targets identified for the novel cassava miRNAs.

Secondary siRNA production

Many miRNAs, such as miR390 can direct their targets to generate phased trans-acting siRNAs (tasiRNAs) biogenesis, and the tasiRNAs then regulate other gene expression (Zhang et al, 2005; Allen et al, 2005). MiRNA-dependent transacting siRNAs (tasiRNAs), also known as phased siRNAs, are generated from noncoding TAS loci as well as protein-coding transcripts, and the secondary siRNAs can silence additional genes (Chi et al, 2011; Howell et al, 2007). These miRNAs that trigger tasiRNAs are usually 22 nt in length. Notably, the miR390 identified in this study was not the typical 22 nt tasiRNA triggering 22 nt but was found to be 21 nt. This was also observed in a study by Montes et al, (2014). They found that the miR390 family was the only tasiRNA initiator to be present in all plant species studied and at high abundances but not expressed as a 22 nt. A member of the miR482 family identified in this study was also found to be 22 nt. Studies demonstrate that members of the miR482/2118 superfamily initiate large numbers of phased, secondary tasiRNA accumulation from plant *NB-LRR* class of resistance genes. For example, in tomato, sequence diverse members of the miR482 family target large numbers of *NB-LRR* mRNAs, which in turn produce phased tasiRNAs (Shivaprasad et al, 2012), while in *Medicago truncatula* miR2118 causes large amounts of phased secondary tasiRNAs from *NB-LRR* mRNAs (Zhai et al. 2011).

2.5 Conclusion

In this study, 259 conserved miRNAs belonging to 32 families were identified using EST database, 32 conserved miRNAs belonging to 7 families identified using GSS database and 289 conserved miRNAs belonging to 30 families and 39 novel miRNAs belonging to 29 families were identified in T200 and TME3 landraces in deep-sequencing data. Also, 200 (77.2%) of the miRNAs in the EST library, 22 (68.8%) of the miRNAs identified in GSS, 230 (79.6%) of conserved miRNAs and 38 (98.6%) of the novel miRNAs identified in deep-sequencing data have not been previously reported in cassava. The mR2118 family identified in study has not been previously reported for cassava in other studies. However, we

could not experimentally detect this family using RT-PCR and this could be due to low expression levels or specificity of miRNA. Montes et al. (2014) observed low miRNA abundance and conservation of mR2118. We observed that miRNA abundance increased as the conservation of the sequence increased and that our unique cassava specific miRNAs had an abundance of less than 5 RPM. By comparing miRNA identification in cassava in this study with others, it was demonstrated that the method used for miRNA identification, cultivar/landrace of cassava and environmental conditions can affect the miRNAs that are identified. It is notable that some divergence of miRNAs has taken place since cassava was introduced into West Africa in the 16th century [99]. Differences between TME3 and T200 landraces can be hypothesised to have arisen from geographical separation and adaptation as T200 (history not known) is found in drier regions of southern Africa, while TME3 originates West Africa. Variations could have arisen from hybridizations with local wild *Manihot* species in different locations over the past few hundred years. While this research has unveiled some more important features of the cassava miRNAome, a large number of germplasm-specific cassava miRNAs of low abundance are likely not to have been detected.

Chapter 3

Comparison of microRNA populations in South African cassava mosaic virus infected tolerant and susceptible cassava landraces

3.1 Abstract

South African cassava mosaic virus (SACMV) belongs to the family *Geminiviridae* and is one of the causal agents of cassava mosaic disease (CMD). MicroRNAs (miRNAs) comprise a large group of 21 – 24 nt RNA molecules that play a crucial role in stress response in plants, including biotic stress caused by viral infection. Viruses however can interfere with and exploit the silencing-based regulatory networks, causing the deregulation of miRNAs. This study aimed to understand the regulation of miRNAs in tolerant (TME3) and susceptible (T200) cassava landraces infected with SACMV. Next-generation sequencing was used for analysing small RNA libraries from infected and non-infected cassava leaf tissue collected at 12, 32 and 67dpi (days post-inoculation), and normalized against mock-inoculated samples. A full repertoire of cassava miRNAs was characterized, which included conserved and novel cassava-specific families. The total number of differentially expressed miRNAs across all three time points was 204 and 209 miRNAs, in TME3 and T200 infected plants, respectively, but the patterns of log_2fold changes in miRNA families over the course of infection differed between the two landraces. A high number were significantly altered at 32 dpi when T200 and TME3 plants showed severe symptoms. Notably, in T200 69% and 28 (100%) of miRNA families were upregulated at 12 and 32 dpi, respectively. In contrast, TME3 showed an early pre-symptomatic response at 12 dpi where a high number (87%) of miRNAs showed a significant log_2fold downregulation. Endogenous targets were predicted in the cassava genome for many of the identified miRNA families including RNA silencing associated proteins, transcription factors, resistance (R)-genes and transposable elements. Interestingly, some of the miRNA families (miR162, miR168 and miR403) that were significantly affected in both T200 and TME3 upon SACMV infection were shown to target proteins (DCL1, AGO1 and AGO2) that play important roles in the RNA silencing pathway. Significantly, these miRNA families demonstrated opposite overall expression change patterns between TME3 and T200. From the results, we suggest that the early (12 dpi) downregulated miRNA response to SACMV in TME3 involves upregulation of PTGS-associated AGO1, DCL2 and a cohort of R genes belonging to

the miR395 family which may prime the plant for tolerance and recovery downstream, while in T200, SACMV suppresses AGO1, AGO2 (at 32 and 67 dpi), and DCL2 (32 dpi) mediated RNA silencing, leading to severe persistent disease symptoms. This study provides novel insights into miRNA-mediated SACMV cassava interactions and may provide useful targets for control strategies aimed at developing CMD-resistance cassava varieties.

3.2 Introduction

Due to their sessile lifestyle, plants inevitability suffers all kinds of abiotic and biotic adversity, particularly pathogen-induced plant disease. Annually, typical pathogens including viruses, bacteria and fungi can cause a tremendous loss in cassava production (FAO, 2014). To effectively fight against these pathogens, plants activate precise molecular defences to recognise and resist the intruders (Tsuda and Katagiri, 2010). Simultaneously, successful pathogens in turn evolve diverse counter-defence strategies to avoid or suppress host immunity (Pumplin and Voinnet, 2013). During the past decade, small RNAs (sRNAs) have also been found to be key players in mediating plant-pathogen interactions as well as many other biological processes. There are two main classes of sRNAs, microRNAs (miRNAs) and short interfering RNAs (siRNAs) (Voinnet, 2009), which play a role in post transcriptional gene silencing (PTGS) or transcriptional gene silencing (PTGS).

miRNAs are endogenous short (~21-24 nt) single-stranded non-coding small RNAs that play important roles in development of multicellular organisms and influence the output of many protein-coding genes (Bartel, 2004). miRNA families are highly conserved (Axtell and Bartel, 2005), even though highly specific, recently evolved, miRNA genes have also been observed during the evolutionary process (Voinnet, 2009). Some of the non-conserved miRNAs are generally expressed at low levels, in specific cells, or under specific growth conditions (Rajagopalan et al, 2006). They are transcribed from nuclear *MIRNA* genes by RNA polymerase II (RNA pol II) into primary miRNAs (pri-miRNAs). The pri-miRNAs are then processed in plants by Dicer-like proteins (DCL) into

precursor miRNAs (pre-miRNAs), which form a characteristic hairpin-like structure (Jones-Rhoades et al, 2006). A subsequent processing step by DCL slices the pre-miRNA to form a miRNA:miRNA* duplex (~21-24 nt). The duplex is then methylated and exported from the nucleus to the cytoplasm where it is recognised by an argonaute (AGO) protein and incorporated onto the RNA-induced silencing complex (RISC). Only the mature miRNA strand (usually the one having less stable 5' end pairing) is retained in the complex, while the passenger strand (miRNA*) strand in most cases is degraded (Mateos et al, 2011). Mature miRNAs silence target genes by degrading or repressing the mRNA transcripts at the post-transcriptional level (Brodersen et al, 2008; Kurihara and Watanabe, 2004). An important difference between plant and animal miRNAs is that the regulatory targets of plant miRNAs can be predicted with a fair degree of confidence, simply by identifying mRNAs with near perfect complementarity (Rhoades et al, 2002). Plant miRNAs have a high degree of sequence complementarity to their target mRNAs in the middle (10^{th} and 11^{th} nucleotide) of the complementary regions (Llave et al, 2002; Tang et al, 2003). This has been demonstrated by the detection of 3' cleavage products that have 5' ends that start in the middle of the complementary regions. This is mediated by AGO1 (Baumberger and Baulcombe, 2005; Qi et al, 2005). miRNAs are involved in diverse processes such as development (Jones-Rhoades et al, 2006; Mallory and Vaucheret, 2006), response to nutrients (Chiou, 2007), and environmental stresses (Phillips et al, 2007). They also play critical roles in resistance or susceptibility to pathogens such as plant viruses (Balmer and Mauch-Mani, 2013; Khraiwesh et al, 2012; Pradman et al, 2015; Ramesh et a, 2014; Singh et al, 2012; Wang and Luan, 2015).

The foundation of better understanding the function of disease-associated miRNAs in plants depends on biological experimentation and gene mining (bioinformatic) prediction for discovering miRNAs (de Planell-Saguer and Rodicio, 2011). Among them, high-throughput next-generation sequencing (NGS) is frequently adopted to uncover global disease-related miRNA expression. High-throughput sequencing detects both known (conserved) and novel miRNAs by constructing small RNA libraries, followed by screening of mature miRNA sequence length, pre-miRNA structure and minimal free folding energy. The

advantages of NGS are the large data quantity and high-throughput analysis (Zhang and Wang, 2015), thus, it has been extensively applied in discovering miRNAs in plant disease. For example, via high-throughput sequencing, one novel and 57 conserved miRNAs were discovered to show differential expression levels in *B. cinerea*-infected tomato leaves (Jin and Wu, 2015) and thirty-three *Populus* miRNAs were induced when the host was subjected to canker disease pathogen infection (Chen et al, 2012).

Geminiviruses are an important group of plant viruses with small circular, single-stranded (ss) DNA genomes that replicate as minichromosomes in the nucleus of infected cells (Hanley-Bowdoin et al, 1999). Viruses of the family *Geminiviridae* are divided into four genera based on insect vectors and genome organisation (Brown et al, 2015). Whitefly-transmitted geminiviruses are classified in the genus *Begomovirus* and constitute the largest genus that causes economically important diseases throughout the tropical and sub-tropical regions of the world (Stanley et al, 2005; Varma and Malathi, 2003; Mansoor et al, 2003). *South African cassava mosaic virus* is one such begomovirus first reported in South Africa (Berrie et al, 1998; Berrie et al, 2001) SACMV is bipartite, with the genome consisting of two ssDNA components, known as DNA A and DNA B, of approximately equal size (~2.8kb). SACMV DNA A consists of six open reading frames, AC1-AC4 in the complementary sense and 2 ORFs (AV1 and AV2) expressed in the sense direction. Both AC2 and AC4 of begomoviruses have been shown to be virus suppressors of host RNA immunity (RNA silencing) (Bisaro, 2006).

Plant virus infections can result in disease symptoms that may include chlorosis and/or necrosis, curling, stunting and altered plant stature and morphology, presumably caused by interference of the infection with developmental processes (Zaitlin and Hull, 1987). In recent years, it has been proven experimentally that the short interfering RNAs (siRNAs), and in particular miRNAs, play important roles in plant development and are implicated in host-pathogen interactions (reviewed in Balmer and Mauch-Mani, 2013; Ramesh et al, 2014). Studies have also suggested that viruses can suppress gene expression and

use endogenous RNA-silencing pathways to regulate host gene expression, presumably to benefit virus replication (reviewed in Csorba et al, 2015; Zhang and Qu, 2014). However, the underlying mechanisms that control these activities remain unclear. Several studies have demonstrated that viral suppressors of RNA silencing can interfere with miRNA-mediated regulation of host genes (Chapman et al, 2004, Kasschau et al, 2003). These studies showed that viral proteins interfere with miRNA pathways, although it is unclear whether it is part of the virus replication strategy or a side effect due to the connection between the siRNA and miRNA pathways. For example, transgenic expression in plants of the AC4 protein from *African cassava mosaic virus* (ACMV), a suppressor of post-transcriptional gene silencing (PTGS) (Vanitharani et al, 2004), was correlated with decreased accumulation of host miRNAs and increased developmental abnormalities in *Arabidopsis thaliana* (Chellappan et al, 2005). Furthermore, miRNA-regulation of transcription factors, signalling and hormones, and R (resistance) genes, have been implicated in plant virus disease etiology (Liu et al, 2015; Singh et al, 2012). Recent studies have implicated transcriptome reprogramming (Allie et al, 2014) and R genes (Louis and Rey, 2015) in susceptibility and tolerance/recovery in T200 and TME3 landraces, respectively.

Cassava (*Manihot esculenta*) is a subsistence staple crop whose roots constitute the main source of calories for more than a billion people around the world (Dahniya, 1994; Nassar et al, 2002; Burns et al, 2010; Latif and Muller, 2014). miRNAs have also been proved to be pivotal molecules in plant-pathogen interactions (reviewed in Balmer and Bauch-Mani, 2013). Even though miRNAs have been identified in cassava (Pérez-Quintero et al, 2012; Rogans and Rey, 2016), there have been no reports regarding the role of miRNAs in cassava infected by viruses. The aim of this study was therefore to identify and analyse the change in expression of conserved and novel miRNAs in SACMV-infected landraces. A SACMV-susceptible (T200) and SACMV-tolerant (TME3) cassava landrace were used in this study in order to compare differences in miRNA responses to SACMV. In addition to displaying a disease tolerant phenotype, TME also recovers from initial symptoms (Allie et al, 2014). Recovery from virus pathogens is characterized by initial severe systemic symptoms which

progressively decrease, resulting in a reduction or disappearance of symptoms in newly developed leaves (Bengyella et al, 2015). Symptom recovery is generally accompanied by reduced virus titres and sequence-specific resistance to secondary infection. and has also been linked with the induction of antiviral RNA silencing. Symptom remission or recovery has been reported in several geminivirus-plant interactions for example in *Pepper golden mosaic virus* infected pepper (Rodriquez-Negrete et al, 2009; 2013), and has been associated with TGS and PTGS mechanisms (reviewed in Ghoshal and Sanfaçon, 2015). Tolerance describes the extent to which the host is able to withstand infection without undue damage (Robinson, 1969). The National Academy of Science (Anon, 1968) defines tolerance as "the ability of a host plant to survive and give satisfactory yield at a level of infection that causes economic loss to other varieties of the same host species ". In this study, leaf samples were collected at 12, 32 and 67 days post infection (dpi) as these time points cover the development of disease (early, symptomatic and late/recovery) and allows us to study the changes in miRNA expression as disease develops.

3.3 Materials and Methods

Micropropagation and acclimatization of T200 and TME3 landraces

T200 and TME3 cassava landraces were micropropagated by way of nodal culture on Murashige and Skoog (MS) medium (Murashige and Skoog, 1962) supplemented with 20 g.L-1 sucrose and 2 g.L-1 Phytagel™ (Sigma Aldrich), pH 5.8. The explants for both landraces were grown under identical conditions in the same growth cabinet, and were allowed to grow at 25°C under a 16-hour photoperiod. At the appearance of roots (10 days), plantlets were transferred into Jiffy® pellets which were placed on a tray that was covered with plastic film and placed in an insect free, temperature controlled growth chamber (28°C; 16 h photoperiod). Slits were then gradually made in the plastic film to facilitate acclimatization of explants. Once acclimatised, the plantlets were potted with a 2:1 ratio of potting soil to vermiculite. The potted plants remained in the insect free,

temperature-controlled growth chamber (28°C; 16 h photoperiod). The average light intensity of the growth chamber was 3000 lux. The plants were watered every second day and once a month multifeed fertilizer was added to the plants, following manufacturer's instructions.

Plant growth and SACMV agro-inoculation of cassava T200 and TME3 landraces

Once the T200 and TME3 plantlets were acclimatized and had reached the 4 to 6 leaf stage (approximately 6 weeks), they were either co-inoculated with a total of 60 µl of full length head-to-tail dimers of SACMV DNA-A and DNA-B (Berrie et al, 2001) mobilized in *Agrobacterium tumefaciens* (*Agrobacterium*) AGL1 (OD600 of ±0.8) or were mock inoculated with only *Agrobacterium*. Both mock and infected plants were grown under the same environmental conditions as stated above. *Agrobacterium* containing either SACMV DNA-A or SACMV-DNA-B were cultured independently of each other in Luria broth with working concentrations of 100mg/l carbenicillin and 100mg/l kanamycin. Cultures were incubated at 30°C until an optical density (OD600) of ±0.8 was attained. Each culture was pelleted at 8000 rpm. Pellets were washed in sterile water in a repeated spin at 8000 rpm. Water was removed and pellets were resuspended in 200µl of Luria Broth (i.e. 200ul LB/ml of *Agrobacterium* culture. Each plant was inoculated with 60 µl (20 µl at three different points along the stem below the apical leaves) of SACMV inoculum, using a 1 ml Hamilton syringe. Control plants were mock-inoculated in parallel, with 100 µl of *Agrobacterium* only.

DNA extractions and Quantitative Real-time PCR (qPCR) of SACMV

For each time point (12, 32 and 67 dpi) symptoms were recorded as described in Allie et al, 2014). For each biological experiment, the two leaves closest to the apex were harvested from six plants, for both infected and mock-inoculated plants. This was repeated for two additional biological experiments. A total of 36 apical leaves per time point from the 3 experiments were pooled. Total nucleic acid (TNA) was isolated from these SACMV infected and mock-inoculated leaves using a modified CTAB-based extraction method (Doyle and Doyle, 1987).

From the extracts, fifty milligrams of fresh leaf tissue was homogenized in liquid nitrogen. The resulting tissue powder was suspended in 500 µl of CTAB extraction buffer (2% CTAB, 1.4 M NaCl, 20 mM EDTA, 0.1 M TRIS-HCl, pH 8.0). One µl of 2-mercaptoethanol was added to the suspension, which was incubated at 65 °C for 1 h. The suspension was then purified twice by a chloroform: isoamyl alcohol (24:1) solution and precipitated with isopropanol. The TNA was recovered at 13000 x g at 4 °C for 10 min. Recovered TNA pellets were washed in 70% ice-cold ethanol and later resuspended in TE buffer (10 mM TRIS-HCl, 1 mM EDTA, pH 7.5) as well as treated with 1 µl of RNAse A (10 mg/ml) overnight at 4 °C. The purity of the TNA was assessed using the NanoDrop ND-100 Spectrophotometer (NanoDrop Technologies).

Determination of the viral titre in T200 and TME3 plants was achieved by real time qPCR of TNA extracted from both landraces at time points 12, 32 and 67 dpi. TNA samples were all standardised to a concentration of 100 ng/µl. Duplicates of each pooled sample were prepared as well as a template-free control (NTC) of nuclease-free water. For each sample, a 20 µl reaction was set up in LightCycler capillaries containing 1 µl of 100 ng of leaf tissue TNA added to 4 µl LightCycler ® FastStart DNAMaster Plus SYBR Green I (Roche), 1 µl forward coat protein primer (10µM) 5'ACGTCCGTCGCAAGTACGAT3', 1 µl reverse coat protein primer (10 µM) 5'ATTGTCATGTCGAATAGTACG 3' and 14 µl nuclease-free water. A 150 bp fragment was amplified and quantified using the following amplification conditions: 95°C for 10 min, followed by 35 cycles of 95°C for 10 secs, 60°C for 10 secs, and 72°C for 15 secs. A single fluorescence measurement was taken at the end of each extension step during the PCR amplification cycle. A melting curve (65°C-95°C) with a heating ramp rate of 0.1 °C/sec and a continuous fluorescence measurement was conducted after the amplification and quantification cycle. A 166 bp PCR product of ubiquitin was amplified from 100 ng of the same TNA samples used for viral quantification, which served as an internal loading control. Primers used were previously tested in cassava. Primer sequences used were UBQ10 (fwd): 5' TGCATCTCGTTCTCCGATTG 3' and UBQ10: 5' GCGAAGATCAGTCGTTGTTGG 3' previously described for cassava in Moreno et al (2001).

RNA extraction, small RNA library preparation and sequencing

Total RNA extraction, using a modified high molecular weight polyethylene glycol (HMWPEG) protocol (Gehrig et al, 2000), was carried out on leaf tissue samples collected from T200 and TME3 at 12, 32 and 67 dpi. For each time point and sample, total RNA was extracted from the top two apical leaves from six plants in each of three biological replicates and pooled. For each sample, 1 g pooled leaf tissue was homogenised in liquid nitrogen and added to 5 ml preheated (65°C) GHCL buffer (6.5 guanidium hydrochloride, 100 mM Tris-HCl pH 8.0, 0.1M sodium acetate pH 5.5, 0.1M β-mercaptoethanol) and 0.1g HMW-PEG (Mr: 20 000, Sigma). The mixture was then pelleted by centrifugation (10000 x g) for 10 min at 4°C. The supernatant was treated with 0.1ml 1M sodium citrate (pH 4.0), 0.2 ml 2 M NaCl and 5ml phenol:chloroform:isoamyl alcohol (PCI) (25:24:1). The mixture was then vortexed vigorously and again pelleted by centrifugation (10000 x g) for 10 min at 4°C. The supernatant was removed and RNA was precipitated by adding 5ml isopropanol (propan-2-ol). The mixture was thoroughly mixed and incubated at -20°C for 60 min and pelleted by centrifugation (10000 x g) for 25 min at 4°C. RNA pellets were washed with 5 ml ice-cold 75% molecular grade ethanol. RNA Pellets were dried at 37°C for 5 min. The pellet was resuspended in 100 µl preheated (55°C) RNase-free water and 1µl RNase inhibitor (Fermentas). Enrichment of small RNAs was achieved using the mirVanaTM miRNA isolation kit (Ambion Inc.), following the manufacturer's' protocol. The mirVana™ miRNA Isolation Kit is designed for purification of RNA suitable for studies of both siRNA and miRNA in natural populations and yields highly enriched small RNA species smaller than about 200 bases. For each cDNA library preparation, approximately 500 ng sRNA was used as input for the Illumina TruSeq Small RNA library preparation kit (Illumina, Inc.) and sequencing libraries were created according to the manufacturer's protocol. The libraries were sequenced on a HiSeq 2000 (Illumina, Inc.) instrument as per manufacturer's instructions. Sequencing was performed up to 2 X 101 cycles. Next generating sequencing (NGS) was done using the Illumina HiSeq2000 platform at LGC Genomics in Berlin, Germany.

Small RNA sequencing analysis

Raw reads for the 12 small RNA libraries were cleaned of sequence adapters using the fast-toolkit (htt://hannonlab.cshl.edu/fastx_toolkit/), and low quality tags and small sequences (<15 nt long) were excluded. Reads for each library were filtered for Phred quality scores greater than 20. To eliminate all other small non-coding RNAs, high quality trimmed sequences were mapped onto rRNA, tRNA and snoRNAs sequences from Rfam (Version 12.0). The sequences that mapped completely and had an E-value <0.06 were removed from the libraries and were excluded from further analysis. The sequenced libraries were predicted to have reads within the range ~15-60 nt. Small RNAs in the range of 18 – 26 nt were extracted and the sRNA abundance was calculated. Reads were also normalized per million to take into account different NGS reads from each of the 12 libraries.

Prediction of Conserved miRNAs

The miRProf tool available in the UEA workbench (Stocks et al, 2012) was used to produce the expression profiles of the filtered 18-26 nt sRNAs that matched known miRNAs in miRBase (v.21) (Kozomara and Griffiths-Jones, 2014). The cassava genome, AM5602 available from Phytozome (http://www.phytozomenet/cassava) was used for database search and annotation with 3 mismatches allowed between the miRNA in miRBase and the sRNA. The miRNAs that were predicted by miRProf were aligned to the cassava genome AM5602. Only the miRNAs that mapped with no mismatched or gaps were considered to be potential miRNAs. In order to select for potential pre-miRNAs, the region 250 nt upstream and downstream from where the potential miRNA mapped to the cassava genome was used for secondary structure analysis using the RNA-folding tool in the CLC Genomics Workbench. The following criteria were used for selecting potential cassava pre-miRNAs (Ambros et al, 2003, Zhang et al, 2005): (1) Pre-miRNA could fold into a typical hairpin secondary structure and the mature miRNA was located in one stem; (2) the length of the pre-miRNA was no less than 50 nt; (3) pre-miRNA had a high minimal folding free energy (MFE) and MFE index (MFEI), which was calculated by MFEI=MFEx100/[length x (G+C%)], where length is the length of the RNA

sequence and MFE is a negative folding free energy (-ΔG) (Zhang et al, 2006); (4) the maximum number of nucleotide mismatches between the mature miRNA and its opposite miRNA* sequence was six; (Zhang et al, 2004) no loops or breaks in miRNA/miRNA* duplex was allowed.

Prediction of Novel miRNAs

The miRCat tool in the UEA small RNA workbench (Stocks et al, 2012) was used to identify the novel cassava-specific miRNAs based on the abundance and secondary structure (Moxon et al, 2008). MiRNAs that are not present in the miRBase database or have a very low identity with the known miRNAs are considered to be novel species-specific miRNAs. MiRCat predicts miRNAs from high-throughput sRNA sequencing data without requiring a putative precursor sequence. The sRNA sequences were mapped to the input plant genome, AM5602, using PatMaN (Prüfer et al, 2008) and grouped into loci. In order to enrich for miRNA candidates, the software applies a number of criteria for the determination of a bona fide miRNA loci. In brief, the program searches for two-peak alignment patterns of sRNAs on one strand of the locus and evaluates the secondary structures of a series of putative precursor transcripts using the RNAfold (Hofacker et al, 1994) and Randfold (Bonnet et al, 2004) programs. According to the recent criteria for annotating novel plant miRNAs, miRNA star (miRNA*) is one of the most important biogenesis proofs for the identification of a novel miRNA (Meyers et al, 2008), and therefore only the identified novel miRNAs that had a corresponding miRNA* sequence identified were considered at potential cassava specific novel miRNAs.

Identification of Targets and Gene Ontologies

Target genes were identified using psRNATarget server, an automated plant miRNA target prediction server available at plantgrn.noble.org.psRNATarget/ (Dai and Zhao et al, 2011) using the *Manihot esculenta* (cassava), Unigene, DFCI Gene Index. The analysis parameters were set as default. Briefly, the following criteria were set for predicting the potential cassava miRNA target genes: (1) not more than four mismatches between

identified miRNA and target mRNA; (2) no mismatches were allowed between positions 10th, 11th because this site was believed as a cleavage site; (3) one mismatch was allowed between position 2nd and 12th and up to three mismatches between position 12th and 25th; and (4) not more than two consecutive mismatches. To better understand the functions of the newly identified potential targets, proteins were allocated gene ontology (GO) terms using Uniprot (www.uniprot.org).

Stem-loop RT-PCR for the experimental validation of selected conserved miRNAs

Primer design

The primers for the six individual miRNAs that belonged the miR162, miR168 and miR403 families, that were under investigation, and the two miR166 individuals that were used as the references were designed according to Chen et al, (2005) and are listed in S1 table. The stem-loop reverse transcription (RT) primers have a universal backbone and a specific extension. The universal backbone sequence is designed to form a stem-loop structure because of the complementarity between the nucleotides in the 5'- and 3'- ends. It includes the reverse complement of the universal reverse primer site in the loop region. The specificity of a stem-loop RT primer to an individual miRNA is conferred by a six-nucleotide extension at the 3'-end. This extension is the reverse complement of the last six nucleotides at the 3'-end of the miRNA. Forward primers are specific to the miRNA sequence but exclude the last six nucleotides at the 3'-end of the miRNA. A 5' extension of five to seven nucleotides is added to each forward primer to increase the length and the Tm. These sequences were chosen randomly and are relatively GC rich.

Stem-loop Pulsed Reverse Transcription (cDNA synthesis)

The RevertAid H Minus First Strand cDNA synthesis kit (ThermoFisher Scientific) was used following manufacturer's instructions with some modifications. Stem-loop RT primers (1 μM) were denatured by heating at 65°C for 5 min and then incubated on ice for 2 min. The cDNA synthesis reaction was prepared using 4μl of 5x reaction buffer, 1 μl Ribolock, 2μl of 10mM dNTP mix, 1μl

RevertAid Reverse transcriptase, 10 µl water, 1 µl of the denatured stem-loop RT primers and 1µl of RNA (10ng) extracted from the leave samples that were collected at 32 dpi. Three replicates per RT reaction were prepared. A single "no RNA" control was also prepared. The conditions for the Stem-loop pulsed RT reaction were an initial incubation of 16°C for 30 min followed by pulsed RT of 60 cycles at 30°C for 30 secs, 42°C for 30 sec and 50°C for 1 sec. The reverse transcriptase was then inactivated at 85°C for 5 min.

A qPCR master mix was prepared for each miRNA using the standard reaction of 5µl 2x Maxima SYBR Green/ROX qPCR master mix (ThermoFisher Scientific), 0.3 µl miRNA-specific forward primer, 0.3 µl of universal reverse primer and 3.4 µl water. Nine µl of the master mix was pipetted into each well of a 96-well plate. One µl of the RT product was then added to each well. The qPCR samples were incubated at 95°C for 10 min followed 40 cycles of 95°C for 15 secs, 60°C for 30 sec and 72°C for 30 sec using the LC480 light cycler (Roche). Melting curve analysis was performed by denaturing the sample at 95°C, then cooling to 65°C at 20°C/sec. The florescent signals were collected at 530 nm wavelength continuously from 65 to 95°C at 0.2°C/s. The relative expression software tool (REST) was used to analyse the qRT-PCR data using the crossing points (Pfaffl et al, 2002).

3.4 Results and Discussion

Symptom development of tolerant and susceptible cassava landraces infected with SACMV

T200 and TME3 plantlets were monitored for 67 days for symptom development and SACMC viral titre following agro-inoculation. Leaf samples were collected at three time points (12, 32 and 67 dpi), which represents the progress of disease development, where 12 days post inoculation (dpi) reflects early infection (pre-symptomatic), and 32 dpi represents symptomatic infection and high virus replication. At a later stage of infection, 67 dpi, the recovery phenotype is observed in TME3 (symptom-free newly developing leaves) whereas the susceptible and symptomatic phenotype is observed in T200. No symptoms were

observed in TME3 or T200 plants at 12 dpi. Symptoms (described in detail in Allie et al, 2014) were first observed in both landraces at approximately 15 dpi. At 32 dpi all newly emerging leaves displayed mosaic and leaf curling for both T200 and TME3. TME3 leaves displayed the recovery phenotype at 67dpi where newly emerged leaves had no or reduced symptoms, compared with T200 (typical yellow mosaic on leaves as well as leaf distortion and leaf curling). The mock-inoculated plants did not develop any disease symptoms over the course of the study.

For both susceptible T200 and tolerant TME3, viral load was highest at full systemic infection (32 dpi), but the concentration of SACMV DNA-A was significantly lower (2.5×10^2 and 2.24×10^4) molecules/ng of leaf tissue at 32 and 67 dpi, respectively) in TME3 compared with T200 (1.87×10^3 and 3.19×10^5). The virus titre observations in this study were similar to the infectivity study reported by Allie et al. (2014).

Deep sequencing of small RNAs from mock inoculated and SACMV-infected T200 and TME3 landraces

For each sample, total RNA was extracted from the top two apical leaves from six plants in each of three biological replicates and pooled, and the 12 enriched DNA libraries were sequenced using the Illumina HiSeq 2000 Analyser. All 12 libraries generated had Phred score values of greater than 20. High-throughput sequencing of sRNA populations yielded a total of 186 291 006 raw reads. The SACMV libraries produced a total of 121 469 314 raw reads and the mock libraries produced a total 64 821 692 raw reads. After filtering out adapter sequences and junk reads (length <15 nt) according to the criteria of Illumina's Genome Analyser pipeline software, the total read count dropped down to 157 702 020, with the number of reads in the SACMV and Mock libraries dropping to 100 554 198 and 57 147 822, respectively (S2 Table).

Reads in the size range of 18-26 nt that did not match known plant tRNA, rRNA and snoRNA in the Rfam database, were selected for further analysis. The total number of 18-26 nt reads was 19 412 720, with the SACMV and mock

libraries containing 3 110 708 and 16 302 012 reads, respectively. The 18-26 nt libraries were normalised per million read counts in order to compare sRNA abundance data. The size distribution analysis of the small RNA (sRNA) sequences exhibited a similar pattern of length distribution in all libraries. Small RNA reads of each size class (18 to 26 nt) were calculated as a percentage of total sRNA population counts. The most abundant and diverse species in each library were 21 -24 nt in length (Fig. 3.1), which is typical of Dicer-derived products (Axtel, 2013). Any quantitative changes representing more than 10% of reads were considered significant. In tolerant TME3, the greatest change in the number of sRNA reads (expressed as percentage of total normalized reads) was at the early non-symptomatic 12 dpi phase where there was a significant increase (13 to 40%) in 21 nt sRNA from mock to SACMV infected leaf tissue. However, in T200 the greatest change was at 32 dpi, when symptoms become visible, with the number of sRNAs increasing from 5 to 18%. In mock-inoculated TME3 the 22 nt RNAs were more abundant at 12 and 67 dpi. The most noteworthy difference between SACMV-infected tolerant TME3 and susceptible T200 was that 21 nt (12 dpi) and 22 nt (67 dpi) sRNAs were the most abundant in TME3 (40%) and T200 (33%), respectively. Interestingly, in T200, while 22 nt sRNAs were predominant in SACMV-infected leaf tissue, there was a significant decrease in SACMV-infected compared with mock-inoculated leaves at 12 (31 to 13%) and 32 (46 to 16%) dpi. The 21 nt size class is characteristic of authentic miRNAs (Axtel, 2013). The 22 nt miRNAs or miRNA* are often generated from asymmetric miRNA precursors. The asymmetric miRNA precursors affect the structure of the miRNA/miRNA* duplex, allowing RISC to recruit the RDR6 and SGS3 to trigger the formation of the secondary siRNAs (Manavella et al, 2012). The presence of the 24 nt small RNAs in our libraries represent siRNAs that are associated with genome repeats and heterochromatic modifications (Axtel, 2013).

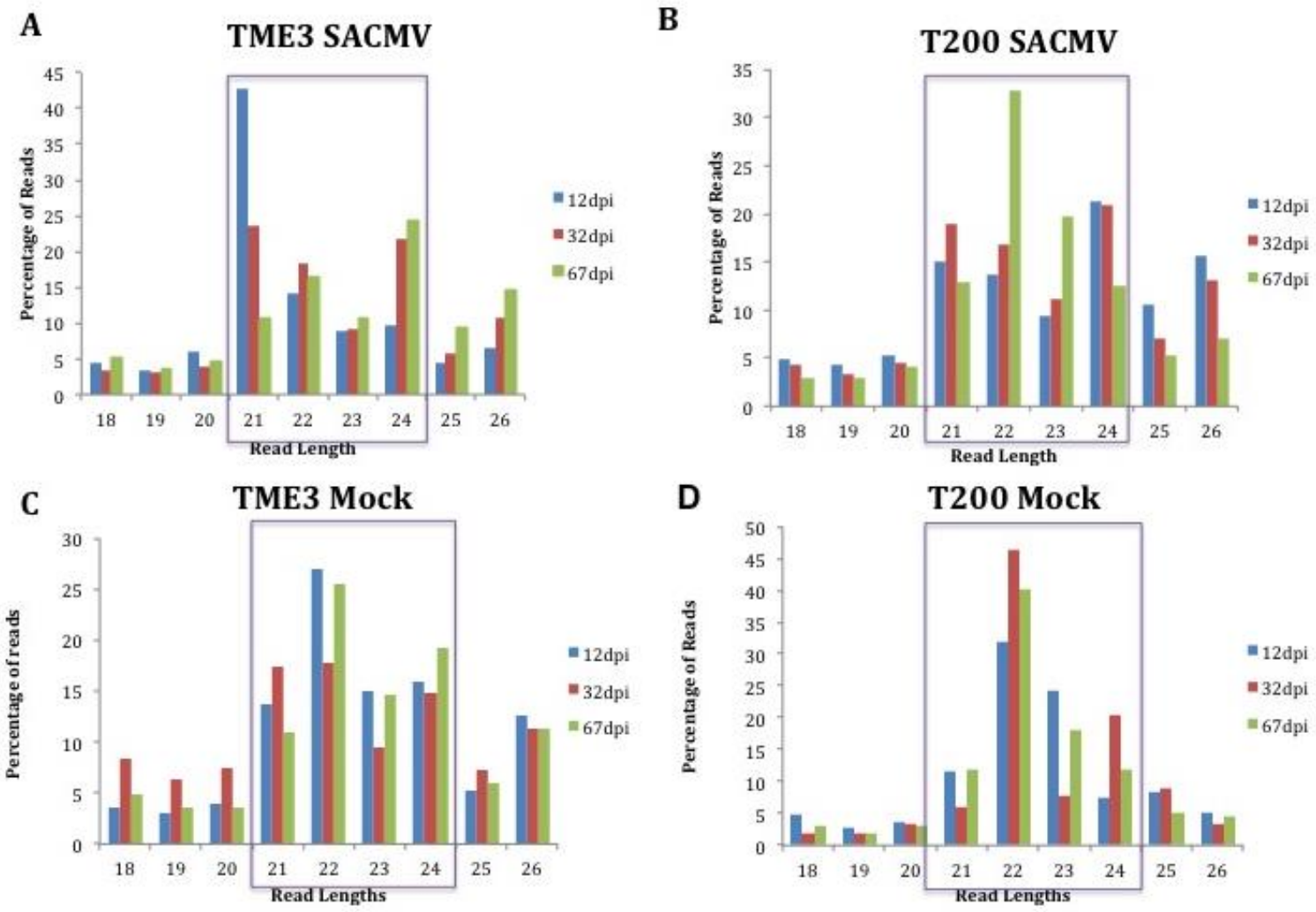


Figure 3.1: Sequence length distribution of cassava sRNAs generated from next generation sequencing. A (SACMV-infected) and C (Mock-inoculated) represent the NGS data generated from TME3 landrace. B (SACMV-infected) and D (Mock-inoculated) represent the NGS data generated for T200. The majority of the generated reads for both landraces mock, and SACMV-infected were 21 – 24 nts in size (indicated by the purple boxes).

Identification of miRNAs in cassava

Identification of conserved miRNAs in cassava

The majority of plant miRNA families are evolutionary conserved from species to species within the plant kingdom (Pan et al, 2006, Denzulian et al, 2006). This conserved nature has greatly enhanced the identification of conserved miRNAs, especially in organisms where DNA sequence information is not yet complete. The 12 filtered sRNA libraries were analysed for potential conserved miRNAs using the miRProf tool available in the UEA sRNA Toolkit (Stocks et al,

2012). This tool is able to determine the normalised expression levels of sRNAs matching known miRNAs in miRBase (V.21) (Kozomara and Griffiths-Jones, 2014) and results for the TME3 and T200 libraries are available in supplementary tables (S3 and S4 tables). Characterisation of putative miRNAs and pre-miRNA candidates is a crucial step for their validations as it distinguishes miRNAs from other small RNAs (i.e. tRNAs, rRNAs and mRNAS) (Frazier et al, 2010; Xie et al, 2010; Wang et al, 2012).

miRNAs are classified into families according to the nucleotide sequence of the mature miRNAs, with identical or very similar sequences grouped into the same family. Members of the same family normally have the same targets. A total of 269 individual miRNA sequences belonging to 34 families were identified in the TME3 libraries (Fig. 3.2A). All 34 of the miRNA families were identified in the SACMV libraries, however only 32 of the families were identified in the mock libraries (miR482 and miR7500 families were not detected). While it is possible that these miRNAs occur in too low an abundance to be detected, and mock library sRNA reads for all 6 samples were consistently lower than the SACMV-infected libraries, we hypothesise that miR482 and miR7500 families are induced by SACMV infection. Induction of certain miRNA families following plant virus infection is well documented (Balmer and Mauch-Mani, 2013; Lui et al, 2015; Ramesh et al, 2014; Wang et al, 2015). It is unlikely that this finding is due to different genetic backgrounds of the landraces as miR482 and miR7500 families were not detected in both T200 and TME3 mock leaf tissue. The miR482 family interestingly target the TIR-NBS-LRR and CC-NBS-LRR classes of disease resistance (R) proteins, but no targets for the miR7500 family could be identified. This family has also been identified in *Gossypium hirsutum* according to miRBase (Kozomara and Griffiths-Jones, 2014), however no targets for miR7500 have been identified in of this plant species. Of the 296 individual miRNAs identified in the TME3 libraries, 193 individual miRNAs were identified in the mock libraries, while 244 individual miRNAs were identified in the SACMV libraries. There were also 25 miRNAs that were absent from the SACMV libraries but present in the mock libraries and 75 miRNAs were absent from the mock libraries but present in the SACMV libraries.

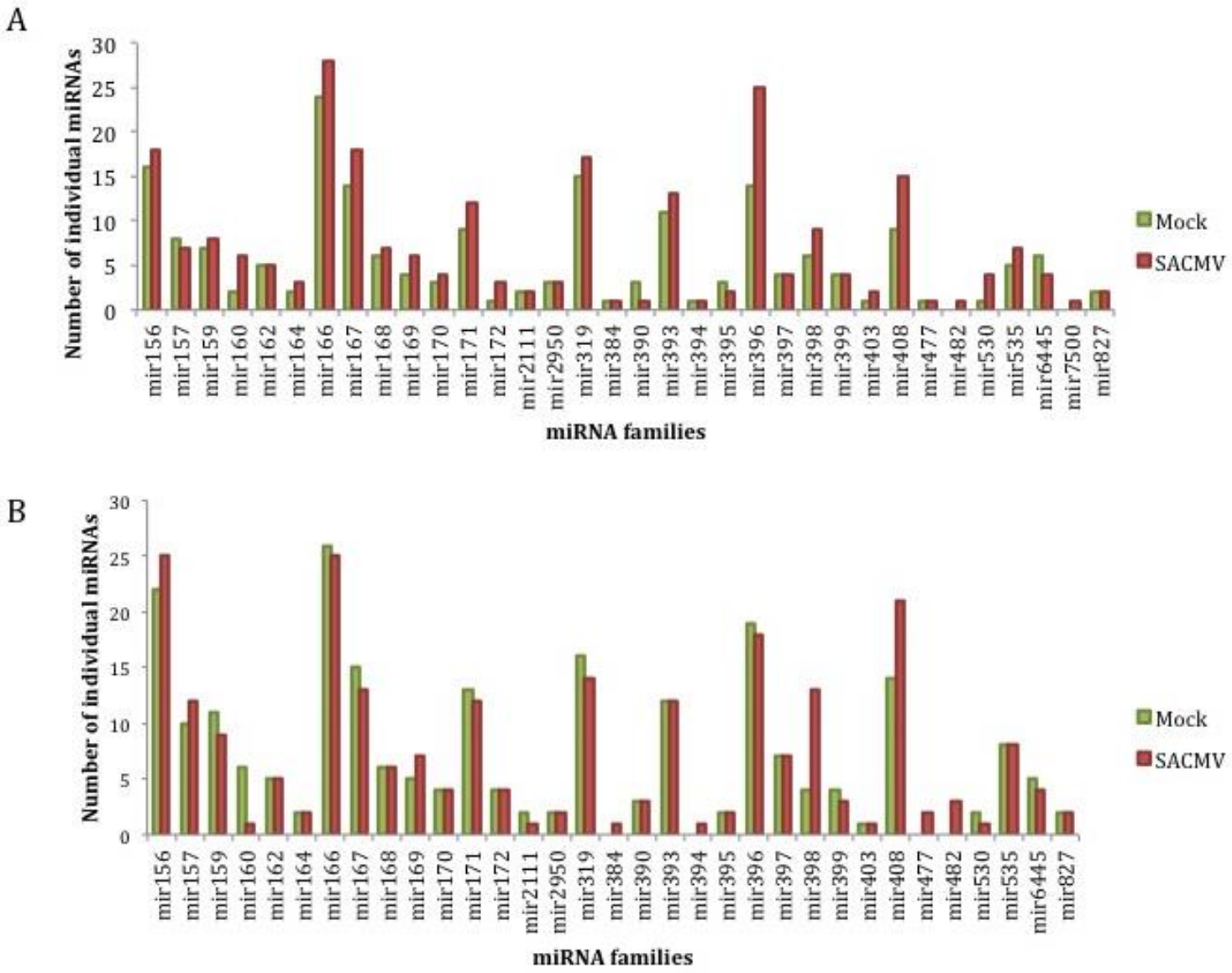


Figure 3.2 A and B: The number of conserved individual miRNAs identified in the TME3 (A) and T200 (B) mock and SACMV-infected libraries. A total of 269 individual miRNA sequences belonging to 34 families were identified in the TME3 libraries and a total of 294 individual miRNA sequences belonging to 33 families were identified in the T200 libraries. The miR7500 family was only identified in TME3 and not in T200. More individual miRNAs were identified in T200 than in TME3

A higher number of individual miRNAs were identified in the T200 libraries compared to the TME3 libraries, with a total of 294 individual miRNA sequences belonging to 33 families identified in T200 libraries (Fig. 3.2B). All 33 families were identified in the SACMV libraries, however only 29 of the families were

identified in the mock samples. As was the case for TME3, the miR482 family was absent from the mock libraries. Additionally, the miR384, miR394 and miR477 families were also absent from the mock libraries. Of the 294 individual miRNAs identified in the T200 libraries, 232 and 243 individual miRNA sequences were identified in the mock and SACMV libraries, respectively. Fifty-one and 62 individual miRNA sequences were identified only in the mock or SACMV libraries, respectively.

Secondary structure analysis of the conserved miRNAs was successful in identifying miRNAs in TME3 and T200 landraces and a summary of the important characteristics of the miRNAs identified from the NGS data can be found in S5 and S6 Tables. Characterisation of putative candidate miRNAs is a crucial step for their validation as it distinguished miRNAs from other small RNAs (i.e. tRNAs, rRNAs and mRNAs), as reported earlier (Frazier et al, 2010; Xie et al, 2010; Wang et al, 2012). The potential cassava mature miRNA sequences identified from the TME3 mock and SACMV-infected libraries ranged in size between 18-24 nt while in the T200 mock and SACMV-infected libraries they ranged from 18 - 23 nt. The largest size class was the 21 nt class for both landraces. The potential pre-miRNAs identified in the TME3 libraries ranged from 73 to 245 nt with an average of 134.34 nt. The pre-miRNAs identified in the T200 libraries had a similar length range, 72 nt to 245 nt with an average of 134.26 nt. These results are similar to previous reports in other plant species (Sunker et al, 2005; Zhang et al, 2006a; Zhang et al, 2008). The determination of a hairpin-loop secondary structure of a potential miRNA is not enough for distinguishing miRNAs from other types of non-coding RNAs (Adai et al, 2005; Bonnet et al, 2004; Zhang et al, 2005). The minimal folding free energy (MFE) is an important criterion to determine stability of the perfect or near-perfect secondary hairpin structure of pre-miRNAs. The more negative the value of MFE, the higher the thermodynamic stability is of the secondary structure of the precursor sequence (Prabua and Mandal, 2010). The MFE of the pre-miRNAs identified in the TME3 libraries ranged from -27.9 kcal/mol to -99.6 kcal/mol with an average of -61.29 kcal/mol and in the T200 libraries –27.9 kcal/mol to -99.6 kcal/mol with an average of -61.56 kcal/mol. The minimal folding free index (MFEI) is an important criterion for distinguishing

miRNAs from other RNAs. Previous research has suggested that a sequence is more likely to be a potential miRNA if the pre-miRNA had a MFEI more negative than -0.85 kcal/mol (Ambros et al, 2003; Zhang et al, 2006). The putative cassava pre-miRNAs identified in the TME3 libraries ranged from -0.83 kcal/mol to -1.49 kcal/mol with an average of -1.06 kcal/mol and in the T200 libraries it ranged from -0.80 kcal/mol to -1.49 kcal/mol with an average of -1.05 kcal/mol. Therefore, the cassava pre-miRNAs identified in this study had more negative MFEIs than other types of RNAs: tRNA (0.64); rRNAs (0.59); mRNAs (0.65) (Zhang et al, 2006), lending support for their identification as pre-miRNAs.

Identification of Novel miRNAs in cassava

Many miRNA sequences are highly conserved within the same kingdom, whereas others are species specific. Species-specific novel miRNAs are also known to have low expression compared to conserved miRNAs (Cuperus et al, 2011) and therefore or are difficult to identify by conventional methods. However, recently established high-throughput sequencing technologies together with powerful bioinformatics tools have allowed efficient identification of not only conserved miRNAs but also low-abundance miRNAs in several plant species (Fahlgren et al, 2007; Sunkar et al, 2008; Pantaleo et al, 2010). Using the miRCat program in the UEA small RNA workbench (Stocks et al, 2011) and based on the primary criteria for annotating novel miRNAs that miRNAs shall form a stem-loop precursor with miRNA* (Meyers et al, 2008), a total of 71 novel miRNAs that can form predicted secondary structures with their corresponding miRNA* sequence were identified in the TME3 and T200 mock and SACMV-infected libraries. The newly identified cassava-specific miRNAs were named mes-1 to mes-58 (S7 Table). Forty-two individual novel miRNAs belonging to 41 families were identified in the TME3 libraries. Twenty and 12 miRNA families were only detected in the SACMV or mock libraries, respectively. Thirty-three individual novel miRNAs belonging to 31 families were identified in the T200 libraries with only 15 miRNA families identified in the SACMV libraries and only 8 families identified in the mock libraries. Interestingly, from these results, it is evident that the number of miRNA families detected in infected leaf tissue in both susceptible and tolerant landraces increased compared with mock infected plants.

A summary of the important characteristics of the novel cassava-specific miRNAs identified in TME3 and T200 libraries can be found in S7 Table. The identified potential novel cassava-specific mature miRNA sequences in the TME3 libraries ranged in size between 18 - 24 nt while in the T200 libraries they ranged from 21 - 24 nt. The largest size classes were the 22 nt and 21 nt class for TME3 and T200 landraces, respectively. The potential pre-miRNAs identified in the TME3 libraries ranged from 67 nt to 212 nt with an average of 129.16 nt. The pre-miRNAs identified in the T200 libraries had a similar length range of 60 nt to 212 nt, with an average of 113.03 nt. The MFE of the pre-miRNAs identified in the TME3 libraries ranged from -41.5 kcal/mol to -114.6 kcal/mol with an average of -54.97 kcal/mol. In the T200 libraries the MFE ranged from -0.86 kcal/mol to -114.6 kcal/mol. As previously mentioned, the MFEI has to be calculated for the potential cassava novel miRNA precursors in order to distinguish a true miRNA from other RNAs precisely (Adai et al, 2005; Bonnet et al, 2004; Zhang et al, 2005). The putative cassava pre-miRNAs identified in the TME3 libraries ranged from -0.83 kcal/mol to -1.49 kcal/mol, with an average of -1.09 kcal/mol, and in the T200 libraries it ranged from -0.86 kcal/mol to -1.57 kcal/mol, with an average of -1.09 kcal/mol. Therefore, the cassava pre-miRNAs identified in this study had more negative MFEIs than other types of RNAs: tRNA (0.64); rRNAs (0.59); mRNAs (0.65) (Zhang et al, 2006), lending support for their identification as pre-miRNAs.

Stem-loop RT-PCR for the experimental validation of selected miRNAs

With increasing interest in the biological functions of small RNAs such as microRNAs (miRNAs) warrant convenient methods for quantification of these small RNA species. Experimental validation of conserved miRNAs was performed based on stem-loop real time RT-PCR. Chen et al, (2005b) developed methods based on reverse transcription (RT) reaction with miRNA specific stem-loop primers followed by qRT-PCR. The expression of selected members of the miR162, miR168 and miR403 families in TME3 and T200 at 32dpi were chosen as these miRNA families targeted DCL1, AGO1 and AGO2 proteins, respectively, and these proteins are essential for the RNA silencing pathway. Using the crossing

points generated from the qRT-PCR and the REST© programme, the six miRNAs' expressions were shown to be all downregulated in TME3 and upregulated in T200 compared to the miR166 reference miRNAs (Fig. 3.3A and 3.3B). This experimentally confirmed that the expression of miR612, miR168 and miR403 was downregulated in the tolerant landrace TME3 and upregulated in T200, and correlated with the NGS miRNA data.

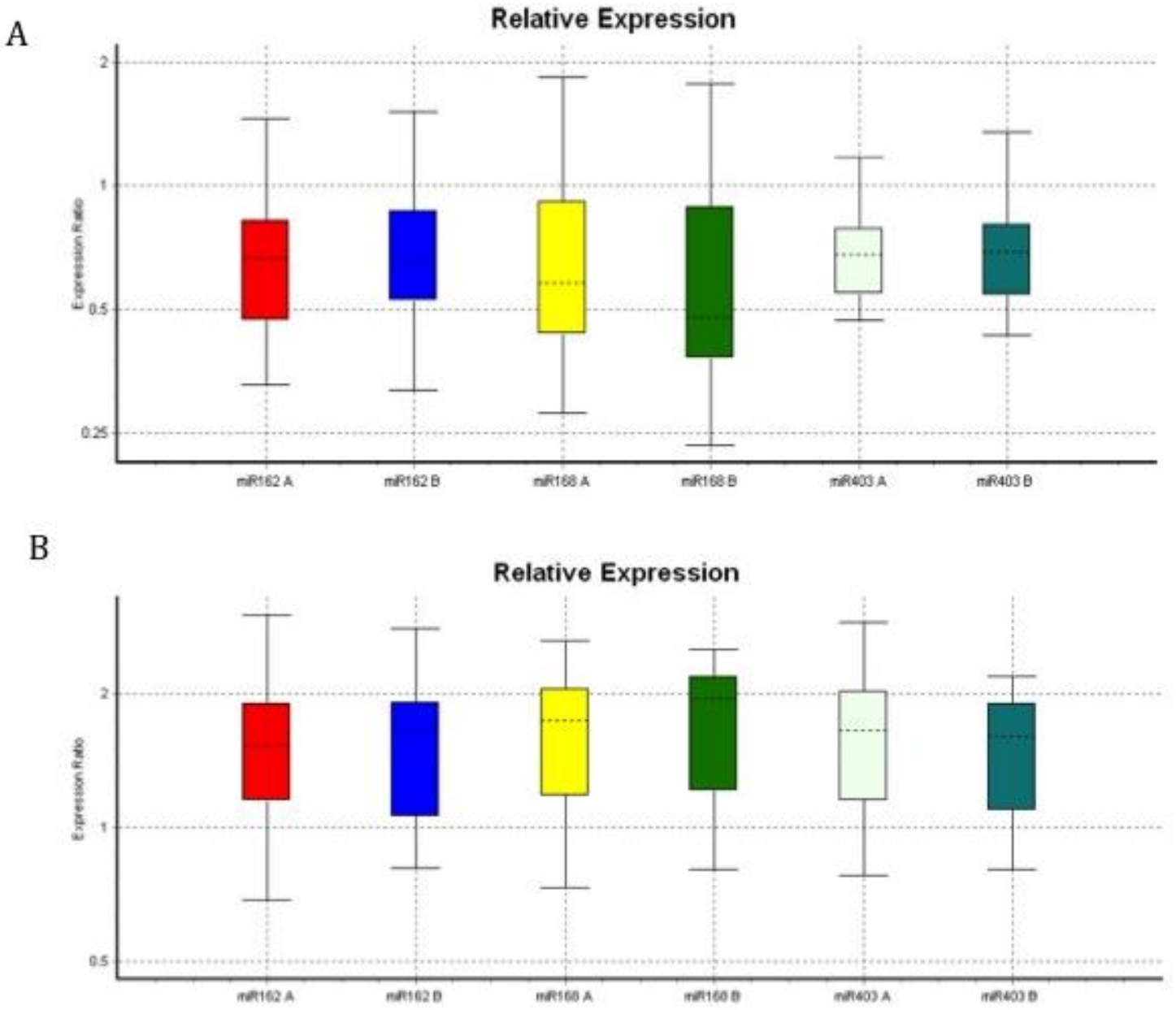


Figure 3.3A and B: Box plots of the relative expression calculated by the Relative Expression Software Tool (REST©) for the individual miRNAs belonging to the moR162, miR18 and miR403 families in (A) TME3 and (B) T200 infected with SACMV compared to mock-inoculated controls. A relative expression below 1 was considered downregulated and above 1 was considered upregulated according to REST. All six individual miRNAs were down regulated in TME3 and upregulated in T200 in the SACMV infected samples relative to the mock-inoculated samples. A combination of two members of the miR166 family was used as the endogenous references. All relative expression values had a p-value <0.05.

SACMV infection-induced log_2 fold changes in conserved miRNAs and their target Gene Ontologies

Insightful information can be drawn from the prediction of the miRNA profiles that are associated with virus infection. In order to study the difference in the miRNA profiles that are associated with SACMV, T200 and TME3 cultivars were infected with SACMV and control plants were mock inoculated with *Agrobacterium.* Newly developing leaves were collected at 12 dpi (early infection stage), 32dpi (full systemic infection) and 67dpi (recovery stage in TME3). It is possible that the changes that occur in host miRNA are related to the life cycle of the virus (Alvaro et al, 2010), which regulates host metabolism and its intracellular environment (Rottiers et al, 2011). Log_2Fold changes were determined for the newly identified individual conserved cassava miRNAs using the equation: Log_2 Fold Change = Log_2 (Normalised count in SACMV library/ Normalised count in the mock library) (S8 Table and S9 Table). We considered a miRNA to be upregulated if the Log_2Fold change was greater than 2 and downregulated if the Log_2Fold change was less than -2. The Log_2Fold Changes in the individual conserved miRNAs in T200 and TME3 were used to construct heat maps (S10 Figure).

Profiles of infection-induced log_2 fold changes in individual conserved miRNA

In the TME3 libraries a total of 204 individual miRNAs belonging to 33 families had their expression altered by SACMV infection, with 93 individuals having a log_2Fold change >2 and <-2 (S8 Table, S10 Figure). Notably, in the T200 libraries, while a similar total number of individual (209) miRNAs and families (32) had their expression altered by SACMV infection compared with TME3, a significantly higher (130) number of individual miRNAs had a log_2Fold change >2 or <-2 (S9 Table, S10 Figure). In TME3 this represents 46% of the total individual miRNAs while in T200, expression of 62% of the miRNAs were significantly altered. This correlates with a greater transcriptome reprogramming response in SACMV-infected T200 compared with TME3, shown in an earlier study (Allie et al, 2014).

At 12dpi in TME3 there was a notable early response of 40 individual miRNAs being downregulated (log_2Fold <-2) and only 6 upregulated (log_2Fold >2). At 32 dpi a larger number (45) of miRNAs were upregulated, while only 5 were downregulated. In contrast to TME3, at 12dpi fewer (10) individual miRNAs were downregulated in T200 while high (39) numbers of miRNAs were upregulated. At 32 dpi, 105 individual miRNAs were upregulated in T200, while none were downregulated. We suggest that the downregulation of a substantially higher percentage (43%) of individual miRNAs at 12 dpi in TME3 is indicative of the induction of an associated cohort of early genes that may be involved in establishing the 'cellular environment' leading to tolerance, a phenotype which persists after recovery for long periods of time in this perennial landrace. While several studies have implicated 21 nt and 24 nt siRNA mediated-PTGS and TGS, respectively, in resistance (Raja et al, 2008; Pooggin, 2013) or at later-stage symptom recovery (Chellappan et al, 2004; Rodriquez et al, 2009) to geminiviruses, no one has yet shown, to date, a definitive role for RNA silencing in tolerance, or examined the very early miRNA responses in plant hosts that develop tolerance, which also persists post-recovery. This invites more in-depth studies ahead on the affected miRNA targeted genes in TME3. In contrast, in susceptible T200, a significantly high percentage (30%) at 12 dpi of > log2 fold upregulated individual miRNA populations were shown. Both T200 and TME3 also exhibited high percentages of upregulated (log_2Fold change >2) miRNAs (80% and 48% for T200 and TME3, respectively), which indicates general suppression of the transcriptome, a pattern which not only correlates with T200/TME3 transcriptome results (Allie et al, 2014) but reflects a similar pattern in several plant virus-host interactions where widespread suppression of transcription occurs at the stages of symptom appearance, and high virus replication and systemic movement (Pierce and Rey, 2013). In TME3, at recovery (67 dpi), 4 individual miRNAs were upregulated and 9 miRNAs were downregulated, while in T200, 28 individual miRNAs were upregulated and 10 were downregulated. While both T200 and TME3 showed a similar lower miRNA response at 67 dpi compared to 12 and 32 dpi, it was apparent that upregulated (>log_2 fold change) individual miRNAs were considerably higher (21%) in T200 than TME3, where

only 4% were upregulated. With the same view that much can be gleaned from studying early responses in tolerance, further investigation of the roles of suppressed target genes in the perennial T200 landrace in disease persistence would prove invaluable.

Gene Ontology classification of conserved log$_2$fold miRNA-targeted genes

The GO annotation classified the target genes into 3 categories, namely molecular function (MF), biological process (BP) and cellular component (CC) since previous research has demonstrated that viral infection can alter the MF, BP and CC of host plants (Laliberte et al, 2013). Gene ontology analysis predicted many target genes associated with the miRNAs that were produced by cassava (T200 and TME3 landraces) in response to SACMV infection**.** The targets identified for the conserved miRNAs in TME3 were associated with 98 CCs, 205 MFs and 348 BPs (S11 and S13 Tables) and in T200 they were associated with 103 CCs, 2111 MFs and 266 BPs (S12 and S14 Tables). In TME3 11 targets were associated with the GO function termed defence response and interestingly 1 target was associated to with the GO functional category, virus induced gene silencing. In T200 only 8 targets were associated with the GO function termed defence response, but T200 also had a target associated with induced gene silencing. Genes with functions related to disease symptoms and resistance/tolerance were of particular interest because they provide clues relating to potential pathogenic mechanisms and resistance genes. The top ten GO terms for each of the GO categories for the targets identified in TME3 and T200 for the conserved miRNAs that had the expression altered by SACMV are represented in Fig. 3.4 A-C and Fig. 3.5 A-C respectively.

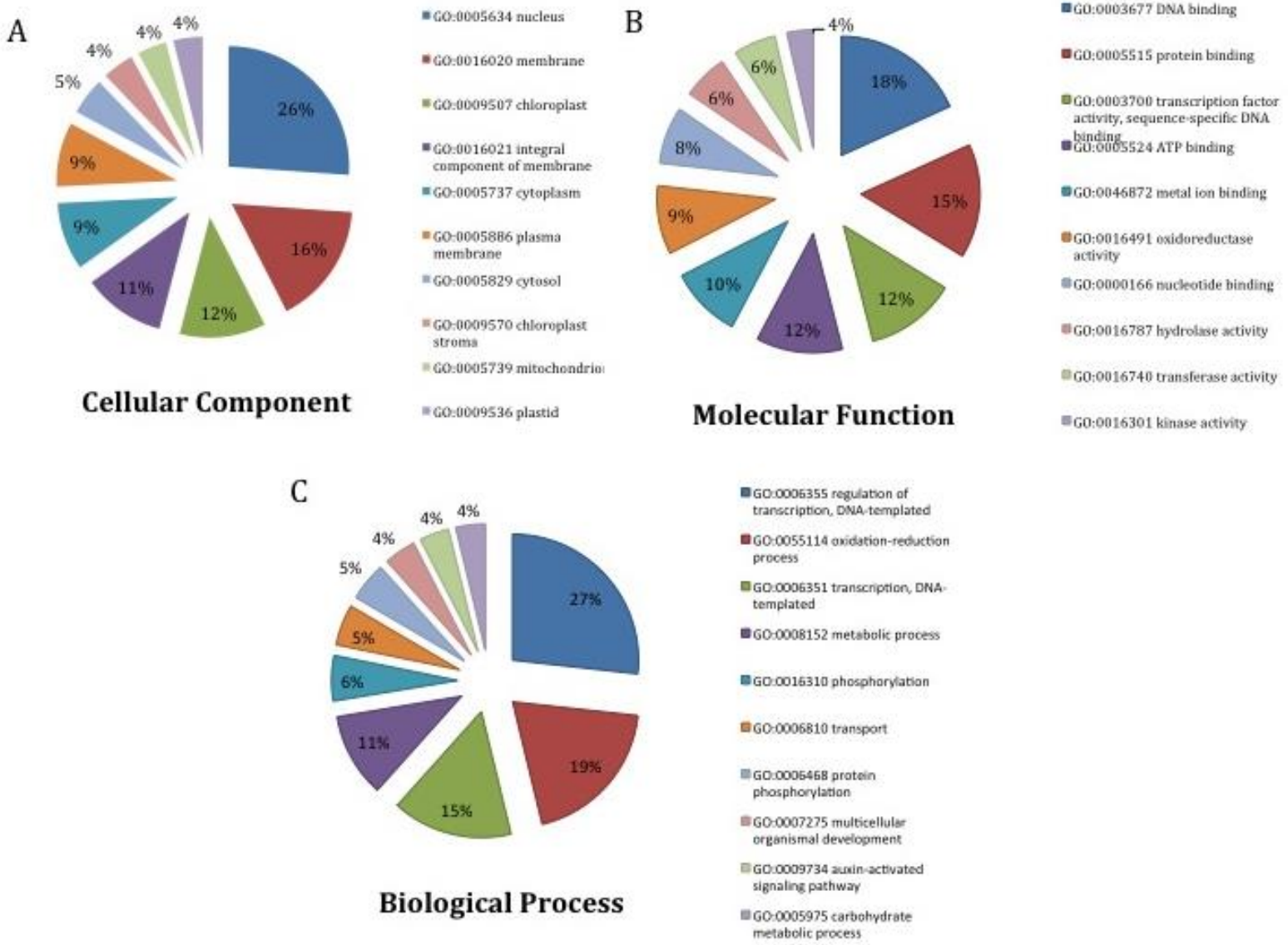


Figure 3.4 A – C: The top ten GO terms for the targets for each GO category (A) Cellular component, (B) Molecular Function, and (C) Biological Process, of the conserved miRNAs whose expression was altered by SACMV in TME3.

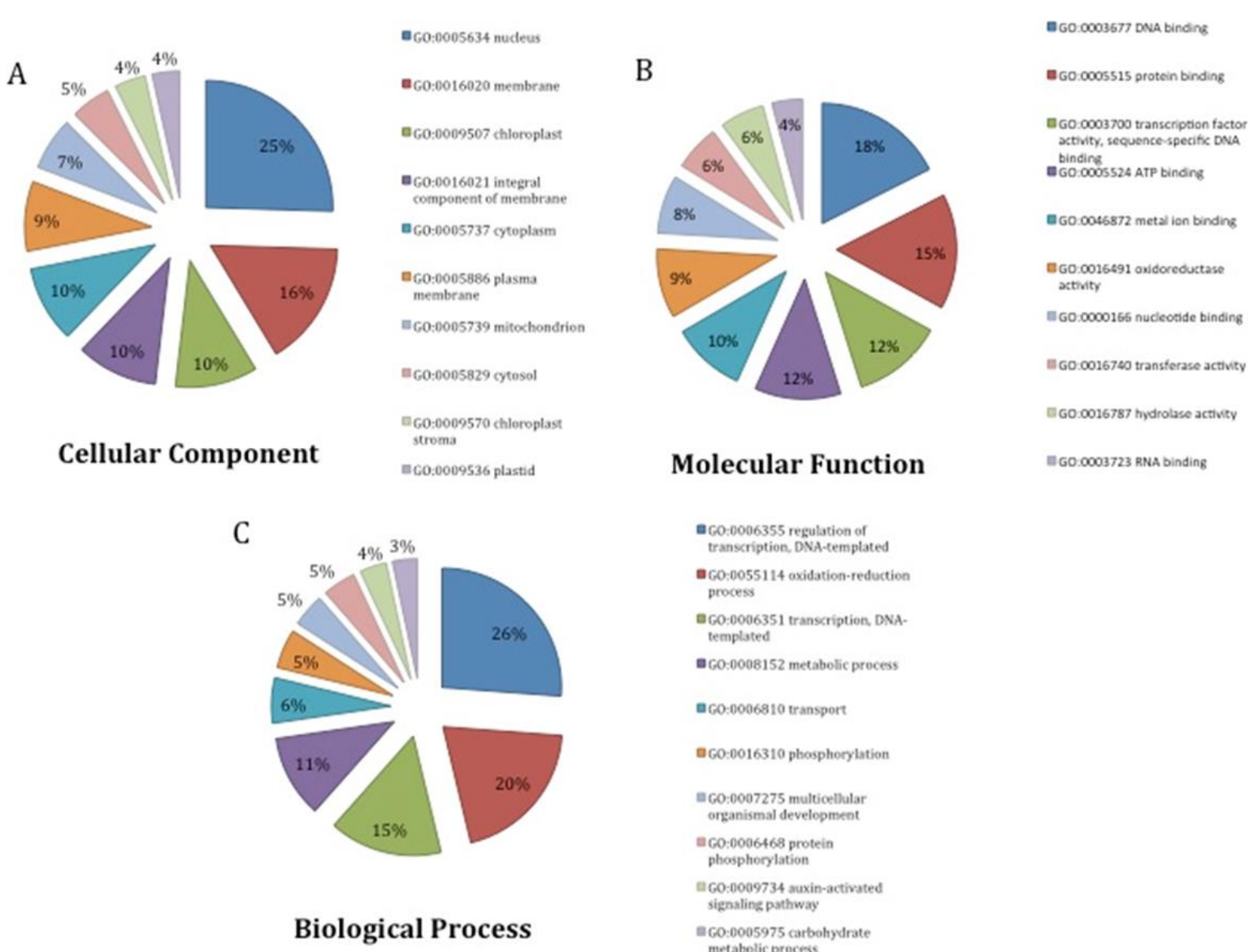


Figure 3.5 A – C: The top ten GO terms for the targets for each GO category (A) Cellular component, (B) Molecular Function, and (C) Biological Process, of the conserved miRNAs whose expression was altered by SACMV in T@00

Roles of log_2fold miRNA-targeted gene functions in disease etiology

The response of cassava to viral stress is complex and involves many genes and interacting molecular mechanisms, operating at both the transcriptional and post-transcriptional level. It is likely that overall changes in miRNA families or in expression of individual miRNAs have important roles in regulating functions of targeted genes in susceptible and resistant/tolerant hosts. The overall log_2Fold changes in conserved miRNA families in SACMV-infected TME3 and T200 at 12, 32 and 67 dpi were used to construct heat maps (Fig. 3.6). The darker the green the more down regulated the miRNA is and the darker the red the more upregulated the miRNA is.

In order to elucidate further the roles these miRNAs (log2fold up- or downregulation <2 or >2) are playing in SACMV infection of a susceptible versus a tolerant cassava landrace it was necessary to identify their targets. To predict the target genes of individual conserved miRNAs that had their expression altered by SACMV infection in TME3 and T200, the web-based plant small RNA target analysis server (psRNATarget) (Dai and Zhao, 2011) was used. Detailed results of target identification by individual miRNAs are presented in S11 and S12 Tables for TME3 and T200, respectively. Due to the large datasets, selected genes of interest, linked to pathogenicity and defence in plant-pathogen interactions, targeted by log_2fold (>2 or < -2) altered conserved miRNA families (Fig. 3.6), are discussed. These predicted selected targets include transcription factors; resistance (R genes); and RNA silencing (Table 3.1)

miRNA Code	TME3 12dpi Log2 Fold Change	TME3 32dpi Log 2 Fold Change	TME3 67dpi Log2 Fold Change	T200 12dpi Log2 Fold Change	T200 32dpi Log2 Fold Change	T200 67dpi Log 2 Fold Change
miR156	-24.53	10.92	-15.53	31.51812295	43.83931432	-7.196139435
miR157	-10.61	7.93	-7.16	11.83542177	22.71994518	-6.431727816
miR159	-15.16	1.35	-3.93	-11.19164324	15.93164084	-6.250332904
miR160	0.56	4.06	1.15	-0.008933125	1.50880623	0.466663188
miR162	-7.59	4.81	0.21	0.635431529	13.50681741	-3.935814752
miR164	-0.05	-3.11	-0.01	0.896674505	1.145850866	-1.121136438
miR166	-16.17	12.41	-3.22	-12.74412659	23.41832502	-2.896182207
miR167	-7.19	23.94	-10.46	0.868777947	31.73162284	3.278059291
miR168	-5.98	0.43	2.06	11.14801098	9.70020729	-2.257159693
miR169	0.00	3.22	0.00	2.538066293	5.179301654	2.507813857
miR170	-2.08	3.62	0.77	-0.368140884	8.095351522	-1.270601514
miR171	-4.89	17.87	0.36	-1.619693503	18.47424763	1.818867193
miR172	0.00	2.32	-1.58	0	5.947125699	5.260402093
miR2111	0.00	2.48	0.00	0	0	0
miR2950	-1.27	4.18	0.00	-1.008933125	4.383520475	-0.194080357
miR319	-0.94	31.24	-3.03	4.913582468	37.91808089	-1.81524312
miR384	1.55	0.00	0.00	0	1.835924074	0
miR390	-1.27	0.97	-2.00	1.156005024	5.340683471	-0.951816466
miR393	-5.96	8.82	1.81	2.147855084	22.46821597	-3.434732558
miR394	1.55	0.00	0.00	0	1.835924074	0
miR395	-3.13	-2.56	1.16	0.687060688	3.992967784	4.518325308
miR396	-8.29	30.62	5.98	-2.1211132	26.90431445	-14.76856407
miR397	-4.02	9.12	-0.42	9.098239335	18.23796327	16.48906532
miR398	2.15	-2.48	3.48	28.97895978	24.47398585	14.89280939
miR399	-0.86	1.51	1.00	-0.219774249	3.582493127	0
miR403	0.00	-2.56	0.00	-0.008933125	3.83491343	2.053637964
miR408	2.32	24.23	-0.67	47.22509356	65.26929247	28.65075153
miR477	0.00	0.00	0.00	0	2.021479727	0
miR482	0.00	0.00	2.70	0	2.514021236	0
miR530	-0.98	1.32	0.00	0	2.835924074	0
miR535	1.89	9.43	-2.17	3.166732511	12.97408644	0.097366094
miR6445	-1.08	5.53	-2.87			
miR7500	1.90	0.00	0.00	-0.330981015	12.0094705	0.979979949
miR827	-2.72	1.94	-1.38	0.687060688	3.005399988	-2.038959029

Figure 3.6: The Log_2Fold Change of the conserved miRNA families identified in TME3 and T200 at 12, 32 and 67dpi. The darker the green the greater the downregulation of the expression of the miRNA family and the darker the red the greater the up regulation of the expression of the miRNA family.

Table 3.1: Summary of the miRNA families that target Transcription Factors, Resistance genes and proteins involved in the RNA-silencing pathway.

miRNA Family	Transcription Factors	Resistance Genes	RNA silencing Protein
miR156	SPL	-	-
miR159	MYB	-	-
miR162	-	Probable disease resistant protein	DCL1
miR164	NAC	-	-
miR167	ARF	-	N/A
miR168	-	-	AGO1
miR172	AP2	-	-
miR2111	-	TIR-NBS-LRR	-
miR396	-	TIR-NBS-LRR	-
miR403	-	-	AGO2
miR482	-	TIR-NBS-LRR class, CC-NBS-LRR class and NB-ARC domain-containing	-

Transcription Factors

Numerous targets of the conserved miRNAs identified in this study were conserved with targets in other plant species, and favoured genes encoding transcription factors (TFs). For example, the miR156 family targets Squamosa promoter binding protein-like (SPL) transcription factors (Wang et al, 2009), miR159 targets Myeloblastosis (MYB)-domain containing transcription factors (Reyes and Chua, 2007), miR167 targets the auxin responsive factor (ARF) gene family (Wang et al, 2005); and miR172 targets AP2-like transcription factors. These miRNAs are classified as highly conserved in plants (Zhang et al, 2006). Transcriptional gene regulation is crucial for host cells to form an efficient defence response (Buscaill and Rivas, 2014). The arsenal of defence-related transcriptional regulators consists of DNA-binding TFs like MYB, NAC and AP2, which have been shown to regulate the expression of defence-related genes (Eulgem, 2005). Complex regulatory networks are formed by transcription factors together with miRNAs in host defence response. Transcription factors regulate miRNAs by binding to the upstream cis-acting elements of *MIRNA* genes. In TME3 and T200 several log_2fold-altered miRNAs were induced by SACMV at different stages of infection (Fig. 3.6). The following differentially expressed miRNA families targeted transcription factors: miR156, miR157, miR159, miR166, miR167, miR169, miR171, miR172, miR2111, miR319, miR393, miR395, miR396 and miR408. In T200 the following miRNA families targeted transcription factors: miR156, miR157, miR159, miR166, miR167, miR169, miR171, miR172, miR319, miR393, miR395, miR396, miR408 and miR482.

Some specific members of the mi156 family identified in this study for both TME3 and T200 were found to target 8 members of the SPL family. Transgenic overexpression of miR156a in tomato led to this disappearance of stem pith (Zhang et al, 2011). Interestingly, the phenotype was similar to that of pathogen-infected tomato. miR156 family was significantly upregulated in TME3 at 32 dpi (log_2fold 10.92) and in T200 at 12 dpi (log_2fold 31.5) and 32 dpi (log_2fold 43.8). This up regulation of miR156 family, which results in the suppression of SPL genes, could be responsible for the leaf symptoms that were observed in both cassava landraces at the 32 dpi stage, which reflects full systemic infection and

symptom development. In contrast, in TME3, symptom recovery at 67 dpi was associated with a notable downregulation (-15.53) of the miR156 family, suggesting a different member set of miR156 may be involved.

miR159 has been previously validated to target MYB TFs (Palatnik et al, 2003), playing crucial regulatory roles in plant development. Members of the miR159 family identified in this study in both TME3 and T200 were found to target MYB transcription factors. In addition to its other functions, it is known that MYB TFs have roles during pathogen infection. Previous studies have shown that *MYB* genes are induced in tobacco plants in response to *Tobacco mosaic virus* (TMV) infection, being important in the hypersensitive response and systemic acquired resistance (Singh et al, 2002). In tomato with the geminivirus, *Tomato leaf curl New Delhi virus* (ToLCNDV), the level of miR159 was increased and symptom development may have been due to the up regulation of miR159 (Naqvi et al, 2010). In both TME3 and T200 the miR159 family was downregulated at 12 and 67 dpi. However, at 32 dpi, the miR159 family was significantly upregulated in T200 (Log_2Fold change of 15.93) but not in TME3. The up-regulation of this family in T200 could contribute to the severe leaf-curling symptom observed in T200 SACMV infected plants.

miR164 has been reported to negatively regulate the expression of *NAM/ATAF/CUC* (controlled by the NAC family of TFs) (Park et al, 2002; Kasschau et al, 2003), which has a large impact on lateral root development (Guo et al, 2005). Transgenic *Arabidopsis* overexpressing miR164 reduced lateral root emergence (Guo et al, 2005), whereas overexpression of *NAC* increased lateral roots (Li et al, 2012). NAC was found to be a target of the miR164 family in both cassava landraces. However, the only significant Log_2Fold change (-3.11) was observed in TME3 at 32 dpi with the miR164 family being downregulated, which would lead to an overexpression of NAC and increase in lateral root development. However, while lateral root development was not measured in TME3, and would prove interesting in future studies, it would not be unreasonable to hypothesise that overexpression of NAC may be a result of TME3 attempting to compensate for SACMV effects.

Previously, miR167 along with its targets, Auxin response factors (ARFs), has been shown to regulate flower fertility and fruit initiation (Nagpal et al, 2005). ARFs are a major class of transcription activators and repressors that facilitate the auxin signal by binding to specific cis-elements in the upstream regions of auxin-inducible genes (Guilfoyle and Hagen, 2001). ARF proteins regulate embryogenesis, root development and floral organ formation (Wang et al, 2005; Mallory et al, 2005). Interestingly ARFs were only identified as targets for the miR167 family in TME3 and not in T200. Also the miR167 targets ARF6 and ARF8. The miR167 family was downregulated at 12dpi and 67 dpi, and upregulated at 32 dpi (Fig. 3.6). We hypothesise that TME3 is able to counteract SACMV by ensuring that specific ARF-associated auxin signals remain functional at 12 dpi in order to stabilize plant development. This would contribute to a tolerance phenotype and recovery at 67 dpi, manifested by the milder symptoms and lower virus load demonstrated in the infectivity assays in this study and also by Allie et al (2014). At 32 dpi, as mentioned before, SACMV symptoms and transcriptome suppression is widespread in both T200 and TME3 when symptoms manifest and virus load is increasing. TME3 recovers whereas T200 does not which may explain why no detectable change in this cohort of ARFs. A next-generation sequencing study by Pérez-Quinter et al (2012) aimed at addressing the role of miRNAs in the *Manihot esculenta* (cassava)-*Xanthomonas axonopodis* pv. manihotis (*Xam*) interaction. They found that 10 conserved miRNA families had highly increased expression (log_2fold >2) in response to *Xam* infection, including miR160, miR167, miR393 and miR390 families which are known to target auxin response factors (ARFs) which results in auxin regulation (Zhang et al, 2012). miR160, miR167, miR390 and miR393 induction was found to be involved in regulating auxin signalling: (Navarro et al, 2006; Zhang et al, 2010). Auxin signalling disruption is an important strategy in all plant-pathogen interactions, including geminiviruses, as an attempt to mount a defence response (Derksen et al, 2013; Pumplin and Voinnet, 2013). Several geminivirus studies have shown similar findings. SACMV (Pierce and Rey, 2013) and *Cabbage leaf curl virus* (CLCuV) (Ascencio-Ibanez et al, 2009) have been shown to disrupt auxin signalling in *Arabidopsis*, and in this study could also play a role in virus infections. In T200 three miRNA families were found

to target ARFs, namely miR160, miR393 and miR482. In TME3 three miRNA families (miR160, miR167 and miR393) were also found to target ARFs. All of these families were upregulated at 32dpi, which would result in auxin signalling disruption that was also observed in cassava when infected with *Xam*, and which would correlate with symptom development. Both host-geminivirus specific and universal auxin signalling factors and their interconnecting networks warrant further investigation.

Disease resistance (R) genes and defence

The plant innate immune system is an ancient and evolutionarily conserved defence strategy (Jones and Dangl, 2006). The adaptive immune system is composed of numerous disease resistance (R) genes that play a vital role in detecting pathogen effectors (avirulence proteins) by recognising effecter-induced modifications to other host proteins (Dodds and Rathjen, 2010; Qi and Innes, 2013). R proteins are generally characterised as having a nucleotide-binding site (NB) domain and a C-terminal leucine-rich repeat (LRR) domain (Elmore et al, 2011). MicroRNAs have been shown to be involved in plant defence against viruses and bacteria (Pelaez and Sanchez, 2013), and miRNAs and siRNAs are involved in the activation of effector-triggered immunity (ETI), often involving R proteins (Weiberg et al, 2015). Small RNA-mediated silencing of R genes regulates host defence against pathogens (Yi and Richards, 2007). MiR482 cleaves mRNA of NBS-LRR at its N terminus, and is often upregulated by plant pathogens. The targeting is accompanied by mRNA degradation and production of secondary siRNAs, which depends on RNA-dependent RNA polymerase 6. These secondary siRNAs further silence other defence-related genes, which forms a miR482-mediated silencing cascade (Shivaprasad et al, 2012; Xiao and Luan 2014). Other recent findings have provided evidence that NB-LRR transcripts are regulated by miRNAs at several conserved motifs in Solanaceae (Shivaprasad et al, 2012; Li et al, 2012) and legumes (Zhai et al, 2011). In *Nicotiana benthamiana*, miR482 has also been shown to be involved in plant immunity (Zhai et al, 2011). In TME3 and T200 three miRNA families were found to target disease resistance proteins, namely miR396 (targeting TIR-NBS-LRR class); miR482 (targeting TIR-NBS-LRR class, CC-NBS-LRR class and NB-ARC domain-containing proteins) and miR162

(Probable disease resistant protein). miR2111 (TIR-NBS-LRR class) was only expressed in TME3 at 32 dpi (Fig. 3.6). In both TME3 and T200 the general trend at 32 dpi was significant log_2fold upregulation of the miRNA families that targeted these R genes, which correlated with symptom development and increasing virus loads. Notably miR396 had a very high log_2fold change at 32 dpi, namely log_2fold 30.62 in TME3 and log_2fold 26.9 in T200 (Fig. 3.6). Additionally, T200 displayed an increased expression of miR162 (13.5 log_2 fold change) at 32 dpi. A recent study has shown that diverse resistance gene analogs (RGAs) are differentially expressed during tolerance and recovery in TME3 (Louis and Rey, 2015). This study revealed that in cassava, a perennial crop, RGAs participate in tolerance and differentially accumulate during recovery at 67 dpi as a complementary defence mechanism to natural occurring RNA silencing to impair viral replication. While this was the first demonstration of the involvement of RGAs at 67 dpi (recovery), there has been no information available to date with regard to early pre-symptomatic responses and miRNA-associated R gene regulation. Interestingly, TME3 in this study exhibited a downregulation of miR162 (-7.59 log_2 fold change) and miR396 (-8.29 log_2fold change) at 12 dpi (Fig. 3.6) which we propose induces multiple NBS-LRR defence proteins which are maintained at a low level leading to tolerance (milder symptoms and lower viral loads later compared with T200) in TME3. This system would also minimize the cost to the plant for defence. Evidence for low-level expression of auto-activated RGAs contributing to tolerance and recovery in TME3 has recently been provided (Louis and Rey, 2015).

miRNAs and pathogen counter-defence

Recent studies of both animal and plant viruses have shown that viruses alter the RNA silencing pathways to regulate host gene expression (Baulcombe, 2005; Voinnet, 2005; Wang et al, 2012). Viral infections can also modify miRNA-derived-*trans*-acting siRNAs or heterochromatic-siRNA production, which also leads to phenotypic changes during virus infection (Raja et al. 2008; Wang et al. 2010). A continuous molecular arms race exists between pathogens and their hosts (Jones and Dangl, 2006). Pathogens have evolved the ability to avoid, supress or even hijack host defence pathways (Pumplin and Voinnet, 2013). To counter RNA silencing viruses encode certain proteins that can block virtually any

step of the RNAi pathway and are referred to as viral suppressors of gene silencing (VSRs) (Voinnet, 2005). CMV 2b is the first and most important identified viral suppressors of RNA silencing (VSRs) as counter-defence against host immunity (Pumplin and Voinnet, 2013). By blocking AGO1 and DCL proteins that are key factors in host gene silencing pathway, 2b assists in virus escaping plant immunity (Guo and Ding, 2002; Ruiz-Ferrer and Voinnet, 2007). It has been demonstrated that viral suppressors of RNA silencing can interact and interfere with the miRNA pathway (Chapman et al, 2004; Kasschau et al, 2003), although it remains unclear whether these interactions are the part of the survival strategy of viruses or just side effects (collateral damage) of their infection cycle. Interestingly, three miRNA families identified in T200 and TME3 in this study were found to target important proteins that are involved in the RNA silencing pathway. The miRNA families miR162, miR168 and miR403 were found to target DCL1, AGO1 and AGO2, respectively. In the model plant *Arabidopsis thaliana*, there are four DCLs and ten AGOs (Vaucheret, 2006), which are specialised for different silencing related pathways, while in cassava 13 AGOs have been recently identified (Mirzaei et al, 2014).

In TME3 the miR162 family was downregulated (-7.59 log_2fold) at 12 dpi whereas in contrast, T200 this family's expression was significantly upregulated (13.5 log_2fold) at 32 dpi (Fig. 3.6). Argonautes are effector proteins that play critical functions in eukaryotic RNA silencing pathways (Carbonell and Carrington, 2015). Plant AGOs associate with sRNAs to target and silence complementary RNA or DNA through PTGS or TGS, respectively, and are targeted by miR168 family. The miR168 family is one of the miRNA families that is closely associated with a range of RNA and DNA viruses. It was found that miR168 was induced significantly in *Nicotiana benthamiana* plants after they were infected with TMV, *Potato virus X* (PVX) and *Tobacco etch virus* (TEV), and also in *Arabidopsis* plants infected with *Ribgrass mosaic virus*, and *Turnip crinkle virus*, in *Medicago truncatula* with TMV, and in *Solanum lycopersicum* with PVX (Varallyay et al, 2010). Expression levels of miRNA168 were also elevated in *N. benthamiana* infected by the geminiviruses CbLCuV and *Tomato yellow leaf curl virus* (TYLCV) (Amin et al, 2011), in *Arabidopsis* plants with ORMV (Hu et al, 2011), and in tomato

plants with *Cucumber mosaic virus* satellite RNAs (Feng et al, 2012). Also in the virus-infected grapevines miRNA168 was found to be upregulated (Singh et al, 2012). Therefore, the induction of miRNA168 is a ubiquitous phenomenon in plant-virus interaction, suggesting that miR168 plays an important role in the host defence and virus's counter defence (Varallyay et al, 2010). It is hypothesised that the virus-induced accumulation of miRNA168 may repress the translation of *AGO1* mRNA and thus inhibits the expression of AGO1 protein. The general induction of miR168 in virus-infected plants may counter the inhibitory activity by AGO1 on the virus replication in susceptible hosts (Varallyay et al, 2010). The high abundance of miRNA168 in the susceptible T200 SACMV-infected libraries is in agreement with previous findings on the virus-induced miR168 expression. Interestingly, in TME3 at 12 dpi the miR168 family was also downregulated (-5.98 log2 fold) while in T200 there was significant up regulation at 12 dpi (11.14 log_2fold) and 32 dpi (9.7 log_2fold) (Fig 3.6). In T200 an accumulation of miR162 and miR168 would result in the suppression of DCL1 and AGO1 activity, respectively, leading to suppression of host RNA silencing and establishment of severe symptoms and high virus titres. Contrary to T200, in TME3 the early downregulation of the miR162 and miR168 families would lead to an early increase in the RNA silencing signal as DCL1 and AGO1 proteins will be expressed, thereby eliciting RNA silencing. AGO1 recruits virus-derived siRNA and then degrades the target viral RNA (Zhang et al, 2006a) which results in a decrease in virus titres. Lower virus levels in TME3 correlate with PTGS-associated AGO1 and DCL. These results indicate a negative or positive role for miRNA-associated PTGS in the symptom phenotypes of the susceptible T200 and tolerant TME3 landraces, respectively. Notably we demonstrate for the first time that both PTGS and R genes are activated at an early pre-symptomatic stage in TME3. Elicitation of R gene-encoded NB-LRR proteins is likely to activate a number of signalling pathways, as NB-LRR-induced defence responses are not necessarily specific to the originating virus or its avirulence protein (Bhattacharjee et al, 2009). Downregulation of miR162 in TME3 was shown to up regulate an unidentified R protein (Probable disease resistant protein). It could be hypothesised that R gene-encoded proteins may activate AGO1 and PTGS in TME3, resulting in tolerance. Clearly NB-LRR proteins and AGOs are linked in virus resistance as, for example,

AGO4-dependent translational control has been shown to be involved in the induction of NB-LRR proteins (Bhattacharjee et al, 2009). Post the symptomatic phase at 32 dpi, neither miR162 or miR168 appeared to be involved in TME3 recovery at 67 dpi as no changes in expression were noted (Fig. 3.6). However, what is very pertinent, is that we have recently demonstrated that a different cohort of resistance gene analogues are specifically associated with recovery in TME3 (Louis and Rey, 2015). Collectively, results from this study, and that of Louis and Rey (2015) provide new insights into molecular mechanisms involved in tolerance and recovery. Additionally, while symptom recovery in infected plants has been correlated with the accumulation of virus-targeting short interfering RNAs (vsRNAs) in other studies (Rodríquez-Negrete et al. 2009), no information on miRNA alterations has been reported. This is the first report to date that shows that early pre-symptomatic host responses may play a role in a tolerance/recovery in a virus-host interaction. Other host-geminivirus interactions need to be investigated to explore further the molecular mechanisms involved in different disease phenotypes.

miR403, which targets AGO2, was found downregulated (-2.56 $\log_2$fold) at 32 dpi in TME3. Interestingly, AGO1 expression is induced at 12 dpi, but at the full systemic infection stage AGO2 is active (miR403 family expression is suppressed) while AGO1 expression is unaltered (miR168 expression 0.43 $\log_2$fold). AGO2 can function in co-operation, and non-redundantly with AGO1 if AGO1 is suppressed by viruses (Harvey et al, 2011). Although it belongs to a different clade, AGO2 displays both additive and overlapping activity with AGO1. It is suggested that as SACMV replication increases in TME3, as was shown by real-time qPCR, a decrease in AGO1 levels occurred, and this may have induced accumulation of AGO2 as a result of reduced AGO2 targeting by AGO1/miR403 complexes (Fig. 3.7. AGO2 levels were reported in an earlier study to be regulated by miR403 in an AGO1-dependent manner (Allen et al, 2005). This regulatory network may allow AGO2 to take over antiviral defence when AGO1 levels are themselves dampened by viral silencing suppressors. Recent studies implicate AGO2 in defence against a broad range of viruses. AGO2 has been found to bind viral siRNA (Takeda et al, 2008) and mutations of AGO2 was linked to hyper-susceptibility to plant virus (Harvey

et al, 2011; Wang et al, 2011). Like AGO1, AGO2 is induced and loaded with DCL4- and DCL2-dependent viRNAs in virus-infected plants. Loss of *AGO2* function is also sufficient to allow systemic infection of viruses not normally hosted by *Arabidopsis*, whereas AGO1 has no effect on this host-range determination (Pumplin and Voinnet, 2013). In contrast to TME3, T200 the miR403 family was found upregulated at 32 (3.83 log_2fold) and 67 (2.05 log_2fold) dpi in T200. This would result in a suppression of AGO2 during the full systemic and the late infectivity stages, which will ultimately aid the repression of the RNA silencing signal in this cassava landrace.

In summary, the miRNA families miR162, miR168 and miR403 which target DCL1, AGO1 and AGO2, respectively, respond differently in susceptible T200 and TME3. This is depicted in a working model (Fig. 3.7). Our hypothesis is that in TME3 during SACMV infection, the transcription of these three miRNA families is decreased, which results in the decrease of their expression. This has a downstream effect on their corresponding targets resulting in an increase in the expression of the *DCL1, AGO1* and *AGO2* genes, which increases the RNA silencing signal. This increase in the RNA silencing signal could play a major role in the recovery phenotype that is observed in TME3. However, in T200 the opposite is observed. The hypothesis is that in T200 during SACMV infection, the transcription of the three miRNA families is increased, resulting in the increase in their expression. The increase in inexpression of miR162, miR168 and miR403 results in an increase in the targeting and inhibition of the expression of DCL1, AGO1 and AGO2 mRNA respectively. These three proteins are known to play major roles in the RNA silencing pathway and by inhibiting the expression of their mRNA, SACMV is inhibiting the RNA silencing pathway. The ability of SACMV infection to increase expression of the miR162, miR168 and miR403 families could be the key to fully understanding the susceptible phenotype that is observed in T200.

Also, an interesting AGO2-miR168-AGO1-miR403 loop has been identified in *Arabidopsis*. Results from previous studies suggest that AGO1 and AGO2 might cooperate with miR168 and miR403 during virus infection (Diermann et al, 2010;

Harvey et al, 2011; Morel et al, 2002; Zhu et al, 2011). It has also been suggested that the overexpression of AGO1 would increase the risk of PTGS in endogenous genes. So the expression of AGO1 should be controlled under a self-check regulator (miR168) to maintain the perfect expression level of AGO1. But viruses have developed a series of molecules to crack this system for example, polerovirus F-box protein P0 degraded AGO1, P21 bound to miRNA/miRNA* and siRNA duplex to inhibit formation of active RISC and P19 had ability to increase level of the endogenous miR168 level to inhibit translational capacity of AGO1 mRNA (Bortolamiol et al, 2007; Chapman et al, 2004; Varallyay et al, 2010). Thus AGO2 could be considered as a secondary defence layer of plants, in case that virus cracked the first defence layer components AGO1. In our study, SACMV was able overcome both AGO1 and AGO2 in the susceptible T200 landrace but not in the tolerant TME3 landrace. This AGO2-miR168-AGO1-miR403 loop is vulnerable and tends to lose balance, so even a slight change of any element in this loop would be amplified constantly. It is believed that transcriptional regulation of AGO1 and AGO2 by miR168 and miR403 and unknown regulatory factors help to keep the balance of this loop.

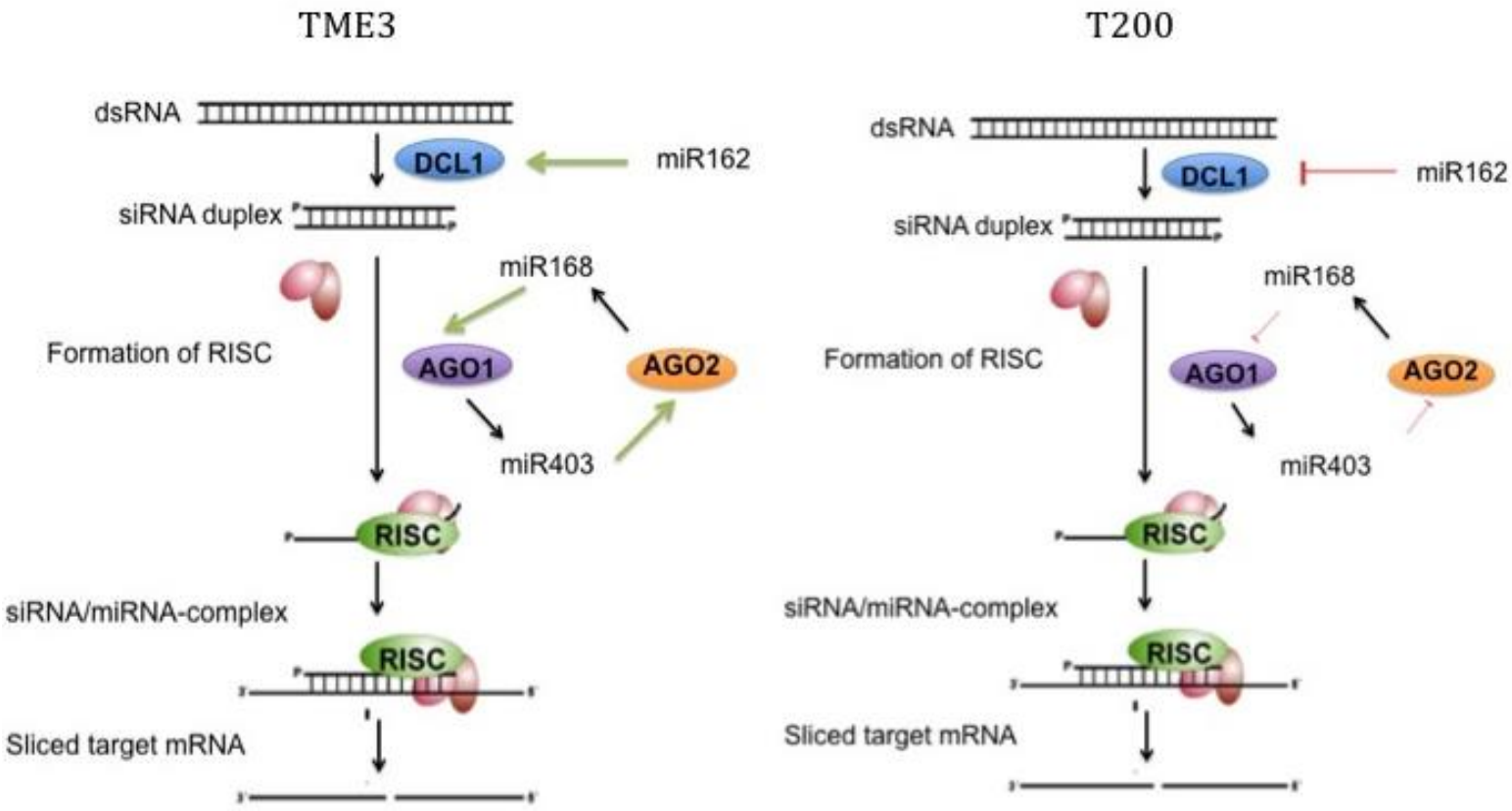


Figure 3.7: Proposed model for the effect SACMV infection has on the expression of the miR162, miR168 and miR403 families and the resulting change in expression of their targets. In TME3 the expression of miR162, miR168 and miR403 decreases during SACMV infection, which results in the increase in expression (indicated by the green arrow) of their targets DCL1, AGO1 and AGO2 respectively. In T200 the opposite occurs. The expression of the miR162, miR168 and miR403 miRNA families increases in expression, which results in the inhibition (indicated by the red inhibition symbol) of their targets DCL1, AGO1 and AG)2 respectively.

DNA-encoded AC2 and AC4 have been shown to be VSRs in several geminivirus infections, and ACMV AC4 was shown to suppress PTGS in cassava (Chellappan et al, 2006). While SACMV AC4 has not been experimentally proven to be a VSR, it is likely that, considering the conserved nature of this protein amongst Old World geminiviruses, this protein may be playing a role in blocking PTGS in T200 and TME3 at 32 dpi. Since VSRs are capable of blocking various steps in PTGS, it is not unreasonable to propose that SACMV AC4 may play a role in downregulation of either one of DCL1, AGO1 and AGO2 in T200 at 32 dpi, leading to severe symptoms.

A comparison between SACMV in cassava in this study and *African cassava mosaic virus* (ACMV) in *N. benthamiana* was performed. Amin et al. (2011) investigated the deregulation of ten developmental miRNAs in *N. benthamiana* infected with four begomoviruses, including ACMV. miR156 and miR160 families had a decrease in expression and miR159, miR164, miR165/miR166, miR167 and miR168 had an increase in expression in *N. benthamiana* infected with ACMV. The miR169 and miR170 families' expression was not altered by ACMV infection. In SACMV-infected T200, different patterns of expression were observed with miRNA families. miR156 and miR168 were upregulated at 12 dpi and 32dpi and downregulated at 67dpi; miR159 and miR166 were downregulated at 12 dpi and 67 dpi and upregulated at 32 dpi; miR167 was upregulated at 32 dpi and 67 dpi; and miR169 was upregulated at all three time-points (Fig. 3.6). The miR160 and mIR164 families' expression was not altered significantly by SACMV infection. These studies show that there was not a common pattern in miRNA regulation for ACMV and SACMV. This is likely due to different hosts and different geminiviral species in the two studies, and illustrates that miRNA-virus responsive transcription patterns are host-virus interaction dependent.

Predicted target genes for identified novel miRNAs

Novel miRNAs were also found to target transcription factors, transposons and R genes similar to the conserved miRNAs, however no novel miRNA was found to target proteins involved in the RNA silencing pathway (S15 Table). Twelve novel miRNAs were found to target transcription factors. Interestingly some of the novel miRNAs targeted similar transcription factors to the conserved miRNAs for example, mes-20, mes-37, mes-40, mes-51 and mes-56 were found to target MYB transcription factors, which are also targeted by the miR159 family. Another transcription factor that was targeted by both novel and conserved miRNAs was the bZIP transcription factor, which was targeted by mes-14, miR395 in TME3 and miR172, miR395 and miR408 in T200. Evolutionary analysis has shown that bZIP transcription factors occur in in all plants as a consequence of sharing a common ancestor, and are important in light and stress signalling, and flower development (Corrëa et al, 2008). In addition to their other functions, it is known that bZIP transcription factors have roles during pathogen infection. Previous studies have shown that bZIP factors, which bind to the G-box of soybean (cv. Glycine max) *Chs15* promoter are also activated during the plant defence response (Alves et al, 2013). Transposable elements (TEs) were targeted by 23 (about 40%) of the novel miRNA families. Global alteration of miRNAs and transposon-derived small RNAs have been shown in cotton during *Cotton leafroll dwarf polerovirus* (CLRDV) infection (Romanel et al, 2012). This large number of novel miRNAs targeting transposable elements is intriguing and future studies should be performed in order to elucidate the exact role that the novel miRNAs and their TE-targets have in virus infection in cassava. In TME3 and T200, 7 and 6 novel miRNAs were found to target R genes respectively. Mes-58 was only found in TME3 and was the most interesting of the novel miRNAs that targeted R genes as it targeted both CC-NBS-LRR and TIR-NBS-LRR classes. Mes-58 expression was also altered significantly at all three time-points, unlike the other novel miRNAs that targeted R genes. It was downregulated at 12dpi and then upregulated at 32dpi and 67dpi. The potential contribution of these novel miRNAs and associated upregulated R genes warrants further investigation.

The targets identified for the novel miRNAs in TME3 were associated with 71 CCs, 150 MFs and 236 BPs (S16 Table) and in T200 they were associated with

78 CCs, 122 MFs and 236 BPs (S17 Table). The top ten GO terms for each of the GO categories for the targets identified in TME3 and T200 for the novel miRNAs that had the expression altered by SACMV are represented in Fig. 3.8A-C and Fig. 3.9A-C respectively. An interesting difference observed between TME3 and T200 was that the GO term defence response is the second highest GO term found in the GO category biological process (BP) in TME3 but in T200 it was only the fourth highest.

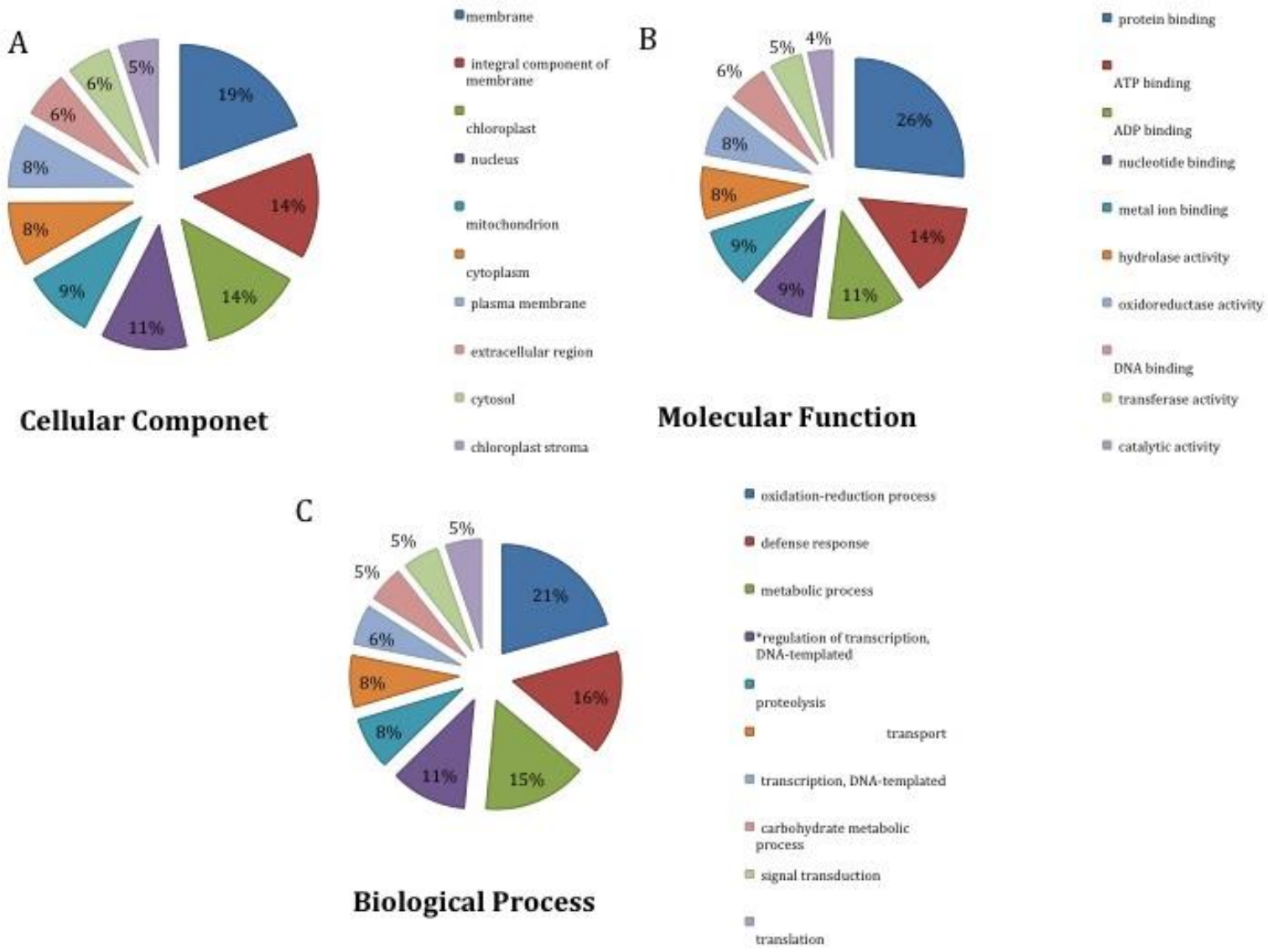


Figure 3.8 A – C: The top ten GO terms for the targets for each GO category (A) Cellular component, (B) Molecular Function, and (C) Biological Process, of the novel miRNAs whose expression was altered by SACMV in TME3.

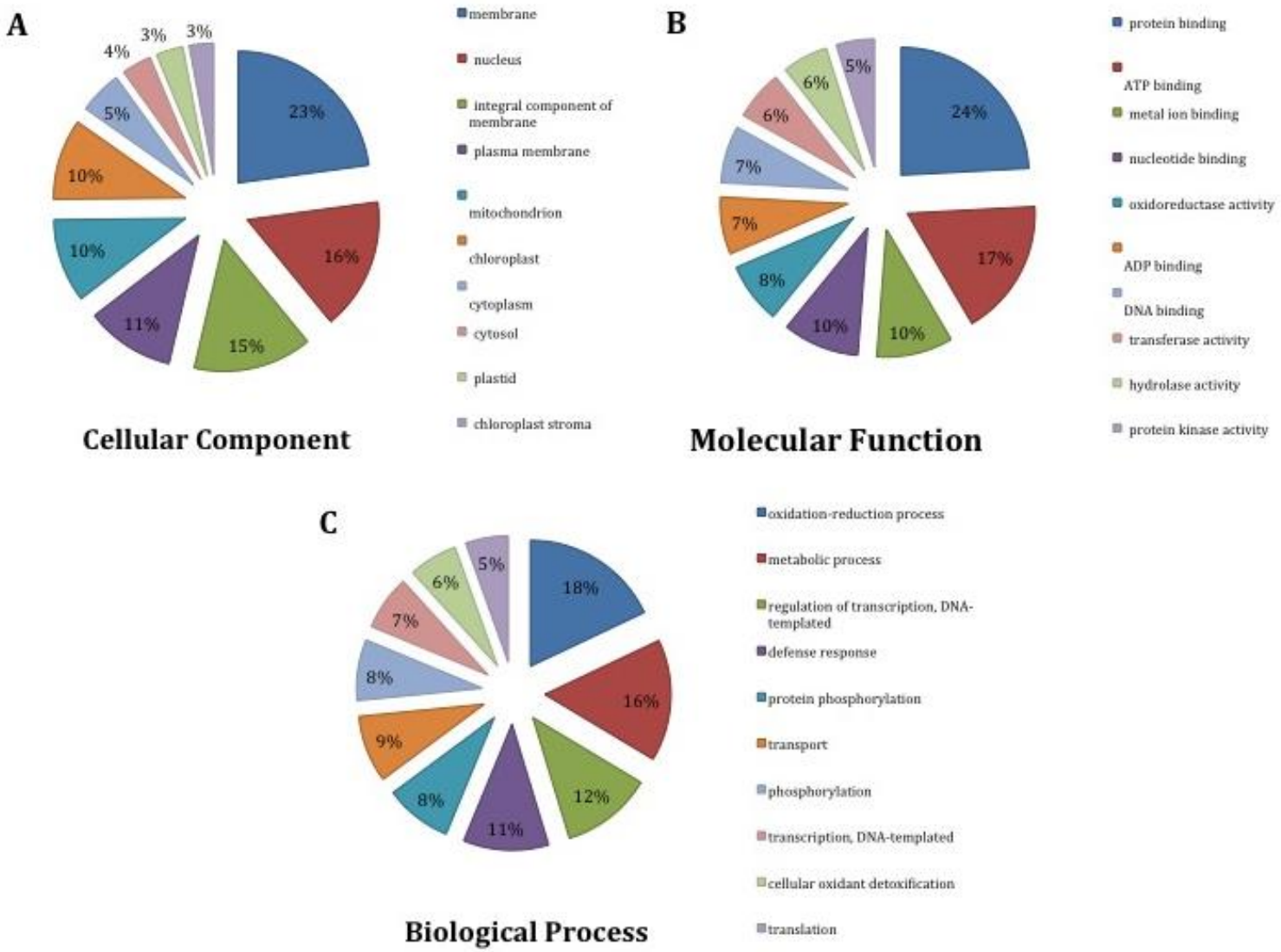


Figure 3.9 A – C: The top ten GO terms for the targets for each GO category (A) Cellular component, (B) Molecular Function, and (C) Biological Process, of the novel miRNAs whose expression was altered by SACMV in T200

3.5 Conclusions

In summary, the current study identified for the first time a broad range of miRNA families (novel and conserved) specifically associated with SACMV infection of T200 and TME3 cassava landraces. Furthermore, the results highlighted differences in both the broad patterns of log_2 fold changes in miRNA families at 12, 32 and 67 dpi between T200 and TME3, as well as differences in specific miRNA-associated target genes between the susceptible T200 and tolerant/recovery TME3 phenotypes. Significantly, it was shown for the first time that an early pre-symptomatic PTGS-R gene response in TME3 may be a contributing factor in tolerance. The discovery of novel and conserved cassava-specific miRNAs in response to a geminivirus brings new perspectives to the understanding of cassava-virus interactions at a molecular level. The changes in miRNA profiles over the period of infection will also provide more interesting data for additional analyses on temporal physiological changes in cassava leaf development. It is envisaged that further investigation of the miRNAs and target genes identified in this study could contribute to disease and tolerance/resistance network models that would lead to the development of disease-resistant cassava plants.

Chapter 4

Small RNA and Methylation Responses in Susceptible and Tolerant Landraces of Cassava Infected with *South African cassava mosaic virus*

4.1 Abstract

Endogenous small RNAs (sRNAs) associated with gene regulatory mechanisms respond to virus infection, and virus-derived small interfering RNAs (vsRNAs) have been implicated in recovery or symptom remission in some geminivirus-host interactions. Transcriptional gene silencing (TGS) (24 nt vsRNAs) and post transcriptional gene silencing (PTGS) (21-23 nt vsRNAs) have been associated with geminivirus intergenic (IR) and coding regions, respectively. In this Illumina deep sequencing study, we compared for the first time, the small RNA response to *South African cassava mosaic virus* (SACMV) of cassava landrace TME3 which shows a recovery and tolerant phenotype, and T200, a highly susceptible landrace. Interestingly, different patterns in the percentage of SACMV-induced normalized total endogenous sRNA reads were observed between T200 and TME3. Notably in T200 there was a significant increase in 21 nt sRNAs during the early pre-symptomatic response (12 dpi) to SACMV compared to mock, while in TME3, the 22 nt size class increased significantly at 32 dpi. While vsRNAs of 21 to 24 nt size classes covered the entire SACMV DNA-A and DNA-B genome components in T200 and TME3, vsRNA population counts were significantly lower at 32 (symptomatic stage) and 67 dpi in tolerant TME3 compared with T200 (non-recovery). It is suggested that the high accumulation of primary vsRNAs, which correlated with high virus titres and severe symptoms in susceptible T200, may be due to failure to target SACMV-derived mRNA. Likewise, in contrast, in TME3 low vsRNA counts may represent efficient PTGS of viral mRNA, leading to a depletion/sequestration of vsRNA populations, supporting a role for PTGS in tolerance/recovery in TME3. Notably, in TME3 at recovery (67 dpi) the percentage (expressed as a percentage of total vsRNA counts) of redundant and non-redundant (unique) 24 nt vsRNAs increased significantly. Since methylation of the SACMV genome was not detected by bisulfite sequencing, and vsRNA counts targeting the IR (where the promoters reside) were very low in both the tolerant or susceptible landraces, we conclude that 24 nt vsRNA-mediated RNA directed genome methylation does not play a central role in disease phenotype in these landraces, notwithstanding recognition for a possible role in histone modification in TME3. This work represents an

important step toward understanding variable roles of sRNAs in different cassava genotype-geminivirus interactions.

4.2 Introduction

Cassava *(Manihot esculenta* Crantz) belongs to the *Euphorbiaceae* family, and is classified as a tuberous perennial crop that is primarily grown in many tropical and sub-tropical regions of the world (FAO, 2008). In South Africa, it is mainly grown by small-scale and subsistence farmers in KwaZulu-Natal, Mpumalanga, and Limpopo provinces. Cassava is considered to be a security crop as it has the ability to grow in nutrient poor soils, and is drought tolerant and resistant to crop pests (Jaramillo et al, 2005). Cassava mosaic disease (CMD) is one of the main biotic and economically important constraints of cassava cultivation in southern Africa. Globally there are 11 recognised species of CMGs (cassava mosaic geminiviruses), of which 9 are reported from Africa and two species from the Indian sub-continent (Adams et al, 2013; Legg et al, 2015; Patil and Fauquet, 2009). *South African cassava mosaic virus* (SACMV) (Berrie et al, 1998; Berrie et al, 2001) is a ss DNA geminivirus belonging to the *Geminiviridae* family. The genome consists of two covalently closed, separately encapsidated, circular ssDNA molecules (DNA-A (2800 nt) and DNA–B (2760 nt). Both DNA molecules in bipartite geminiviruses (GVs) are necessary for infection. DNA-A encodes 2 genes on its virion-sense strand, namely *AV1* (coat protein) and *AV2* [responsible for virus accumulation and symptom development through suppression of host innate RNA silencing mechanism (VSR)] (Bisaro, 2006). The complementary-sense strand of DNA A encodes 4 genes; *AC1-AC3* genes encode the replication-protein (Rep), transcriptional activator protein (TrAP) and the replication enhancer protein (REn), respectively, and *AC4* is involved in virus movement, symptom severity, host range determination and as a VSR (Vanitharani et al, 2004). The DNA-B component is required for inter and intra-cellular movement and encodes 2 genes; *BV1* encodes for a nuclear shuttle protein while *BC1* encodes for proteins required for cell-to-cell movement of the virus (Hehnle et al, 2004; Ward et al, 1997).

The interaction between host plants and virulent pathogens leads to the expression of common sets of defence-related genes. Since constitutive expression of defence pathways in plants can compromise the overall fitness and normal growth of a host plant, plants have evolved intricate mechanisms to exert control over pathogen induced defence pathways. Basal resistance by itself is too weak to protect against virulent pathogens, since it constitutes a residual level of resistance after immune suppression by the pathogen. In addition to PAMP-triggered immunity (PTI) and ETI (Effector-triggered immunity) defence mechanisms (Gomez-Gomez and Boller, 2000; Huffaker et al, 2006; Miya et al, 2007), plants also possess an additional type of adaptive immunity known as RNA silencing which plays a major role specifically in antiviral defence responses (Zvereva and Pooggin, 2012), thereby restricting the accumulation and/spread of viral pathogens. In turn, viruses counteract PTI/ETI-based innate responses and RNA silencing by effectors or suppressor proteins (Burgyan and Havelda, 2011).

RNA silencing also forms part of the conserved silencing mechanism in plants and controls a number of important biological processes including gene expression during development, stress responses, heterochromatin formation and hormone signalling. The plant RNA silencing machinery generates 21 to 24 nt small RNAs (sRNAs) which are broadly classified into transcriptional gene silencing (TGS) and post transcriptional gene silencing (PTGS) associated with microRNAs (miRNAs) and short interfering RNAs (siRNAs) (Khraiwesh et al, 2012). Epigenetic gene regulation is mediated by a highly interactive network of sRNA-directed DNA methylation, and histone and chromatin modifications that control transcription (Matzke et al, 2009). The endogenous siRNAs class have further been categorised into several classes, including *trans*-acting small interfering RNAs (tasiRNAs), natural antisense transcript-derived small interfering RNAs (nat-siRNAs), repeat-associated small interfering RNAs (ra-siRNAs) or heterochromatic small interfering RNAs (hc-siRNAs), and long small interfering RNAs (lsiRNAs) (Chen, 2009; Vazquez et al, 2010). Small RNAs are capable of moving from cell to cell to carry short range signal controlling morphological developmental patterns (Chitwood et al, 2009). The common protein players that participate in all RNA-silencing pathways include the families

of the ribonuclease RNAse III type DICER or DICER-like proteins (DCL), RNA-dependent RNA polymerase (RDRP) and Argonaute (AGO) proteins (Bologna and Voinnet, 2014). Both miRNAs and siRNAs are cleaved or processed from larger double-stranded (ds) dsRNA precursors by DCL with the assistance of DRB proteins. The resulting small RNAs then act as guide molecules for a multi-protein silencing complex that repress genes, in a sequence-specific manner, either post-transcriptionally (PTGS) and/or transcriptionally (TGS).

Growing evidence indicates that innate immunity and RNA silencing are closely linked (Ding and Voinnet, 2007). Small RNAs have recently been shown to be important components of abiotic and biotic stress, and there are many examples of differential siRNA, miRNA or non-coding RNA (ncRNA) expression upon pathogen attack (Khraiwesh et al, 2012; Ruiz-Ferrer and Voinnet, 2007). *Trans*-acting siRNAs (tasiRNAs) for example are a specialized class of siRNAs that are generated by 22 nt miRNA processing of TAS loci transcripts resulting in 21 nt RNAs that are phased with respect to the miRNA cleavage site, a pattern formed by DCL4 (Allen et al, 2005; Zhang et al, 2012), and have been associated with response to cassava blight infection (Quintero et al, 2013). The cleaved products from the miRNA triggering are then processed by a RNA-dependent RNA polymerase (RDR6) and suppressor of gene silencing 3 (SGS3) (Talmor-Neiman et al, 2006) to produce double-stranded RNAs (dsRNAs), which are cleaved by DCL proteins to produce phased 21 nt tasiRNAs.

Virus-infected plants accumulate high levels of virus-derived vsRNAs of three major size-classes: 21 nt, 22 nt and 24 nt (Llave, 2010; Panteleo, 2011). Early in the 1990's (Brough et al, 1992), DNA methylation was already shown to inhibit geminiviruses, *Tomato golden mosaic virus* (TGMV) and *African cassava mosaic virus* (ACMV), in protoplasts. More recently, direct evidence is available that geminivirus-derived vsRNAs (21, 22 and 24 nt) of both polarities are derived from the coding and intergenic regions (reviewed in Raja et al, 2010), and have also been reported from *Manihot esculenta* (cassava) and *N. benthamiana* (Akbergenov et al, 2006), and other geminivirus-plant host interactions (Aregger et al, 2012; Bian et al, 2006; Rodriquez-Negrete et al, 2009). These vsRNAs are

similar to siRNAs derived from dsRNA transgenes, endogenous tasiRNAs and miRNAs, and are phosphorylated at the 5'end and modified at the 3'end, confirming that both TGS and PTGS silencing pathways are involved in plant-geminivirus interactions. Although geminiviruses have DNA genomes and do not replicate using dsRNA intermediates, they have still been shown to be targets of the RNA silencing machinery (Hohn and Vazquez, 2011) since dsRNA intermediates are formed in the process of bidirectional transcription (Bieri et al, 2002; Vanitharani et al, 2005), and precursors of vsRNAs could also form from RdRP activity or from secondary structures of viral RNAs.

RNA-directed DNA methylation (RdDM) is a one of the pathways of plant siRNA silencing machinery and is directed by 24nt siRNAs. RdDM is specifically involved in the regulation of plant gene expression and has been shown in a number of studies to act in plant defence mechanisms against invading and foreign nucleic acids such as transgenes, transposons (Gazzani et al, 2003; Ito, 2013; Michaels et al, 2003) and viruses (Aregger et al, 2012; Raja et al, 2008; Raja et al, 2010; Rodriguez-Negrete et al, 2009). There is a considerable body of evidence that demonstrates that plants methylate geminivirus chromatin as an epigenetic defence. Early studies indicated that *in vitro* methylation of geminivirus DNA greatly impaired replication and transcription in protoplasts (Brough et al, 1992; Ermak et al, 1993). More recently, it has been shown that methylation-deficient *Arabidopsis* mutants are hypersusceptible to geminiviruses and that RdDM pathway components, for example AGO4, are necessary for host recovery from infection (Buchmann et al, 2009). AGO4 specifically plays a role in the production of siRNA that are 24nt long. In addition, geminivirus DNA and associated histones are methylated in infected plants, and viral DNA methylation is reduced in mutants that display enhanced disease. By contrast, the small amount of viral DNA present in recovered tissue is hypermethylated (Raja et al, 2008). These studies clearly demonstrate that methylation, and likely TGS, acts as a defence against DNA viruses. Viruses have consequently evolved diverse counter defence mechanisms to avoid silencing, most notably through the expression of viral suppressors of RNA silencing (Bisaro, 2006).

The most compelling argument for methylation as an anti-geminiviral defence comes from studies associating host recovery and the methylation pathway. Recovery is a phenotype observed in some virus-infected plant hosts characterised by initially severe symptoms which are observed to gradually attenuate until the host appears almost symptomless (Chellappan et al, 2004; Ghoshal and Sanfacon, 2015; Hagen et al, 2008; Sahu et al, 2010). A hallmark of recovery is that virus replication persists at low levels in new tissues in recovered plants. The recovery phenotype is the result of a molecular interplay between the infecting virus and plant host. The phenomenon of recovery has been observed in several plant species, including Cucumis *melo* (cantaloupe) and *Citrullus lanatus* (watermelon) plants infected with *Curcubit leaf crumple virus* (CuLCrV) (Hagen et al, 2008), pepper plants infected with *Pepper golden mosaic virus* (PepGMV) (Rodriguez-Negrete et al, 2009) and tomato infected with *Tomato leaf curl New Delhi virus* (ToLCNDV) (Sahu et al, 2010). In cassava, recovery has been observed when infected with *African cassava mosaic virus* (ACMV) and *Sri Lankan cassava mosaic virus* (SLCMV) (Chellappan et al, 2004). Chellappan et al. (2004) reported a positive correlation between the cassava recovery phenotype post-infection with ACMV and SLCMV and the production of virus-derived small interfering RNAs (siRNAs) through posttranscriptional gene silencing (PTGS). In the case of DNA geminiviruses, transcriptional arrest of viral mini-chromosomes may contribute to recovery, and RNA silencing may also contribute by regulating gene expression (reviewed in Ghoshal and Sanfacon, 2015).

Since the mechanism(s) of tolerance and recovery are not well understood, especially in perennial non-model plants, the objective of this study was to investigate if sRNA responses are associated with host tolerance to and recovery from SACMV infection in the tolerant/recovery TME3 landrace compared with a CMD-susceptible cassava landrace, T200. From this Illumina deep sequencing investigation, there was a clear difference in the pattern and abundance of normalized total sRNA counts and virus-targeted vsRNA populations in leaf tissues between T200 and TME3 in response to SACMV infection during the time course of infection, and between infected and mock inoculated. TGS and PTGS responses in T200 and TME3 differed in some respects to several other studies,

and we conclude that specific sRNA and symptom phenotype responses to geminivirus infection differ between cassava genotypes/landraces. While gene silencing is involved in antiviral defence, this study highlights that RNA silencing is likely to play more complex roles with other mechanisms in both susceptibility and tolerance. A better understanding of the recovery phenotype may lead to advances in breeding programmes, where wild cassava relatives or currently domesticated varieties phenotypically exhibiting the recovery phenotype, may be interbred to create varieties with genotypes resistant to CMD.

4.3 Materials and Methods

Micro-propagation and acclimatization of cassava

T200 and TME3 cassava landraces were micro-propagated by way of nodal culture on Murashige and Skoog (MS) medium (Murashige and Skoog, 1962) supplemented with 20g.L-1 sucrose and 2g.L-1 Phytagel™ (Sigma Aldrich), pH 5.8. Explants for both landraces were grown under identical conditions, and were allowed to grow at 25°C under a 16 h photoperiod. At the appearance of roots (approximately 10 days), plantlets were transferred into Jiffy® pellets which were placed on a tray that was covered with plastic film and placed in a controlled insect-free growth chamber (28°C; 16 h photoperiod). Average light intensity in the cabinet was 300 lux**.** Slits were then gradually made in the plastic film to facilitate acclimatization of explants. Plants were fertilized every 3 weeks with Multifeed fertilizer, following manufactures instructions**.**

Plant growth and virus inoculations of cassava T200 and TME3 landraces

Once T200 and TME3 were acclimatized and the plants had reached the 4 to 6 leaf stage (approximately 6 weeks), they were either infected with SACMV or were mock inoculated with only *Agrobacterium tumefaciens*. Both mock and infected plants were grown under the same environmental conditions as stated above. Leaves were collected from the T200 and TME3 plants at 12, 32 and 67dpi. Eighteen plants, at the 3-4 leaf stage (6 week old plantlets), were co-inoculated with a total of 60 µl of full length head-to-tail dimers of SACMV DNA-A and DNA-B

(Berrie et al, 2001) mobilized in *Agrobacterium tumefaciens* strain AGL1 (OD600 of ±0.8). *Agrobacterium* containing either SACMV DNA-A or SACMV-DNA-B were cultured independently of each other in Luria broth with working concentrations of 100mg/l carbenicillin and 100mg/l kanamycin. Cultures were incubated at 30°C until an optical density (OD600) of ±0.8 was attained. Each culture was pelleted at 8000 rpm. Pellets were washed in sterile water in a repeated spin at 8000 rpm. Water was removed and pellets were resuspended in 200µl of Luria Broth (i.e. 200ul LB/ml of *Agrobacterium* culture. Each plant was inoculated with 60µl (20µl at three different points along the stem below the apical leaves) of SACMV inoculum, using a 1 ml Hamilton syringe. Control plants were inoculated in parallel, and 20 additional plants were mock-inoculated with 100 µl of *Agrobacterium* only.

DNA extractions and Quantitative Real-time PCR (qPCR) of SACMV

For each time point (12, 32 and 67dpi) and each biological experiment, the two leaves closest to the apex were harvested from six plants, for both infected and mock-inoculated plants. This was repeated for two additional biological experiments. A total of 36 apical leaves per time point from the 3 experiments were pooled. Total nucleic acid (TNA) was isolated from these SACMV infected and mock-inoculated leaves using a modified CTAB-based extraction method (Doyle and Doyle, 1987). From the extracts, fifty milligrams of fresh leaf tissue were homogenized in liquid nitrogen. The resulting tissue powder was suspended in 500 µl of CTAB extraction buffer (2% CTAB, 1.4 M NaCl, 20 mM EDTA, 0.1 M TRIS-HCl, pH 8.0). One µl of 2-mercaptoethanol was added to the suspension, which was incubated at 65 °C for 1 h. The suspension was then purified twice by a chloroform: isoamyl alcohol (24:1) solution and precipitated with isopropanol. The TNA was recovered at 13000 x g at 4 °C for 10 min. Recovered TNA pellets were washed in 70% ice-cold ethanol and later resuspended in TE buffer (10 mM TRIS-HCl, 1 mM EDTA, pH 7.5) as well as treated with 1 µl of RNAse A (10 mg/ml) overnight at 4 °C. The purity of the TNA was assessed using the NanoDrop ND-100 Spectrophotometer (NanoDrop Technologies).

Determination of the viral titre in T200 and TME3 plants was achieved by qPCR of TNA extracted from both cultivars at time points 12, 32 and 67 dpi. TNA samples were all standardised to a concentration of 100ng/µl. Duplicates of each pooled sample were prepared as well as a no template control (NTC) of nuclease-free water. For each sample, a 20 µl reaction was set up in LightCycler capillaries containing 1µl of 100ng of leaf tissue TNA added to 4 µl LightCycler ® FastStart DNAMaster Plus SYBR Green I (Roche), 1 µl forward coat protein primer (10µM) 5'ACGTCCGTCGCAAGTACGAT3', 1 µl reverse coat protein primer (10 µM) 5'ATTGTCATGTCGAATAGTACG 3' and 14 µl nuclease-free water. A 150 bp fragment was amplified and quantified using the following amplification conditions: 95°C for 10 min, followed by 35 cycles of 95°C for 10 secs, 60°C for 10 secs, and 72°C for 15 sec. A single fluorescence measurement was taken at the end of each extension step during the PCR amplification cycle. A melting curve (65°C-95°C) with a heating ramp rate of 0.1 °C/s and a continuous fluorescence measurement was conducted after the amplification and quantification cycle. A 166 bp PCR product of ubiquitin was amplified from 100 ng of the same TNA samples used for viral quantification, which served as an internal loading control. Primers used were previously tested in cassava. Primer sequences used were UBQ10 (fwd): 5' TGCATCTCGTTCTCCGATTG 3' and UBQ10: 5' GCGAAGATCAGTCGTTGTTGG 3' previously described in Moreno et al. (2001).

Bisulfite sequencing

Bisulfite-sequencing analysis was carried out with the EZ DNA Methylation Kit (Zymo Research). TNA of SACMV-infected 32 and 67 dpi leaf tissue samples (the time points as which virus was detected in the viral titre qPCR assay) were collected. Ten µg of TNA was treated with sodium bisulfite for 16 h. The treated DNA was purified in accordance with the manufacturer's instructions and used for subsequent PCR reactions using Zymo Taq, which is specific for bisulphite treated templates. Nine primer sets each were used to amplify 200-300 bp regions on DNA-A and DNA-B of SACMV. Primers targeting the intergenic regions and ORFs of DNA-A and DNA-B of SACMV were designed using Invitrogen Methyl software package. Following PCR, the products were cloned in pJet1.2 (Thermo Scientific), and individual clones were sequenced with M13F or T7 primers by Inqaba biotech

(Pretoria, South Africa). The Sanger sequences were then aligned with Clustal X and compared with the DNA sequence from sequences of DNA-A (AF155806.1) and DNA-B (AF155807.2) available in NCBI. All primers used for bisulfite-sequencing analysis are listed in Supplementary Table S1.

RNA extraction, small RNA library preparation and sequencing

Total RNA extraction, using a modified high molecular weight polyethylene glycol (HMWPEG) protocol (Gehrig et al, 2000), was carried out on leaf tissue samples collected from T200 and TME3 at 12, 32 and 67 dpi. For each time point and sample, total RNA was extracted from the top two apical leaves from six plants in each of three biological replicates and pooled. For each sample, 1g pooled leaf tissue was homogenised in liquid nitrogen and added to 5 ml preheated (65°C) GHCL buffer (6.5 guanidium hydrochloride, 100 mM Tris-HCl pH 8.0, 0.1M sodium acetate pH 5.5, 0.1M β-mercaptoethanol) and 0.1g HMW-PEG (Mr: 20 000, Sigma). The mixture was then pelleted by centrifugation (10000 x g) for 10 min at 4°C. The supernatant was treated with 0.1ml 1M sodium citrate (pH 4.0), 0.2ml 2M NaCl and 5ml phenol:chloroform:isoamyl alcohol (PCI) (25:24:1). The mixture was then vortexed vigorously and again pelleted by centrifugation (10000 x g) for 10 min at 4°C. The supernatant was removed and RNA was precipitated by adding 5ml isopropanol (propan-2-ol). The mixture was thoroughly mixed and incubated at -20°C for 60 min and pelleted by centrifugation (10000 x g) for 25 min at 4°C. RNA pellets were washed with 5ml ice-cold 75% molecular grade ethanol. RNA Pellets were dried at 37°C for 5 min. The pellet was resuspended in 100 µl preheated (55°C) RNase-free water and 1µl RNase inhibitor (Fermentas). Enrichment of small RNAs was achieved using the mirVanaTM miRNA isolation kit (Ambion Inc.), following the manufacturer's' protocol. The mirVana™ miRNA Isolation Kit is designed for purification of RNA suitable for studies of both siRNA and miRNA in natural populations and yields highly enriched small RNA species smaller than about 200 bases. For each cDNA library preparation, approximately 500 ng sRNA was used as input for the Illumina TruSeq Small RNA library preparation kit (Illumina, Inc.) and sequencing libraries were created according to the manufacturer's protocol. The libraries were sequenced on a HiSeq 2000 (Illumina, Inc.) instrument as per manufacturer's instructions. Sequencing was

performed up to 2 X 101 cycles. Next generating sequencing (NGS) was done using the Illumina HiSeq2000 platform at LGC Genomics in Berlin, Germany.

Small RNA sequencing analysis

Raw reads for the 12 small RNA libraries were cleaned of sequence adapters using the fast-toolkit (htt://hannonlab.cshl.edu/fastx_toolkit/), and low quality tags and small sequences (<15 nt long) were excluded. Reads for each library were filtered for Phred quality scores greater than 20. To eliminate all other small non-coding RNAs, high quality trimmed sequences were mapped onto rRNA, tRNA and snoRNAs sequences from Rfam (Version 12.0). The sequences that mapped completely and had an E-value <0.06 were removed from the libraries and were excluded from further analysis. The sequenced libraries were predicted to have reads within the range ~15-60 nt. Reads were also normalized per million to take into account different NGS reads from each of the 12 libraries, and data is presented in Table 1 and Supplementary Table S2.

Analysis of viral-derived vsRNA

The small RNA analysis tool in the CLC Genomics workbench was used to facilitate trimming of sequencing reads, counting and annotating resulting tags. The NGS data was imported using the NGS import tool. The minimum length of small RNAs was set to 21 and the maximum length was set to 24 nt. All other parameters were left as default. The resulting small RNA samples were then used to map against the SACMV Genome allowing no mismatches. To check that no host miRNAs targeted SACMV, miRNAs from T200 and TME3 were aligned against SACMV DNA A and B did not show any matches (data not shown). The SACMV DNA A (AF155806.1) and DNA B (AF155807.2) FASTA sequences were downloaded from NCBI (www.ncbi.nlm.nih.gov). The Map Reads to Reference tool in the CLC Genomics Workbench was used to map the sRNA samples generated for each time point for both mock and SACMV infected in both TME3 and T200. MISIS (Seguin et al, 2013) was used to visualise, analyse and compare maps of small RNAs (sRNAs). Extracting the sRNA counts from the Sam files generated from the CLC Genomics Workbench Mapping tool generated an Input Table. The generated Input Table for each time point was then used by MISIS to draw the

histograms corresponding to the selected Input Table. Since a low number of vsRNAs were detected in mock, these were subtracted from the total vsRNA counts in infected samples. The frequency of viral-derived vsRNA populations targeting DNA A or B components and individual ORFs were then calculated as either total (redundant) or unique (non-redundant) vsRNA counts or as percentages of the total vsRNA reads. Additionally, the frequencies of vsRNAs targeting the ORFs were also calculated per 100 nt length to take into account the different lengths of the ORFs. Differences in counts targeting the SACMV genome or individual ORFs were considered significant if differences were greater than 10%.

4.4 Results and Discussion

Symptom severity correlates with SACMV titres

Following agro-inoculation of T200 and TME3, plantlets were monitored over a 67-day period for symptom development (Fig. 4.1) and SACMV viral titre. Based on numerous infectivity assay trials, this time frame covers different phases of infection, where 12 days post inoculation (dpi) represents early infection (pre-symptomatic), 32 dpi represents symptomatic infection and high virus replication and 67 dpi represents a later infection stage (persistently symptomatic in T200 and recovery in TME3). No symptoms were observed in T200 and TME3 plants at 12 dpi. When compared to mock-inoculated plants (Fig. 4.1A), symptoms were first observed in both T200 and TME3 at approximately 15 dpi. At 32 dpi all newly emerging leaves displayed mosaic for both T200 and TME3 (Fig. 4.1B and C). Mock-inoculated plants did not develop any disease symptoms over the course of this study. Interestingly, leaf tissue in TME3 displayed the recovery phenotype at 67 dpi, compared to T200 (typical yellow mosaic on leaves as well as leaf distortion and leaf curling), where newly emerged leaves had no or reduced symptoms (Fig. 4.1E). For both susceptible T200 and tolerant TME3, viral load was highest at full systemic infection (32 dpi), but the concentration of SACMV DNA-A was significantly lower (2.5×10^2 and 2.24×10^4) molecules/ng of leaf tissue at 32 and 67 dpi, respectively) in TME3 compared with T200 (1.87×10^3

and 3.19 X10^5) (Fig. 4.1f). The observations in this study were similar to the infectivity study reported by Allie et al. (2014).

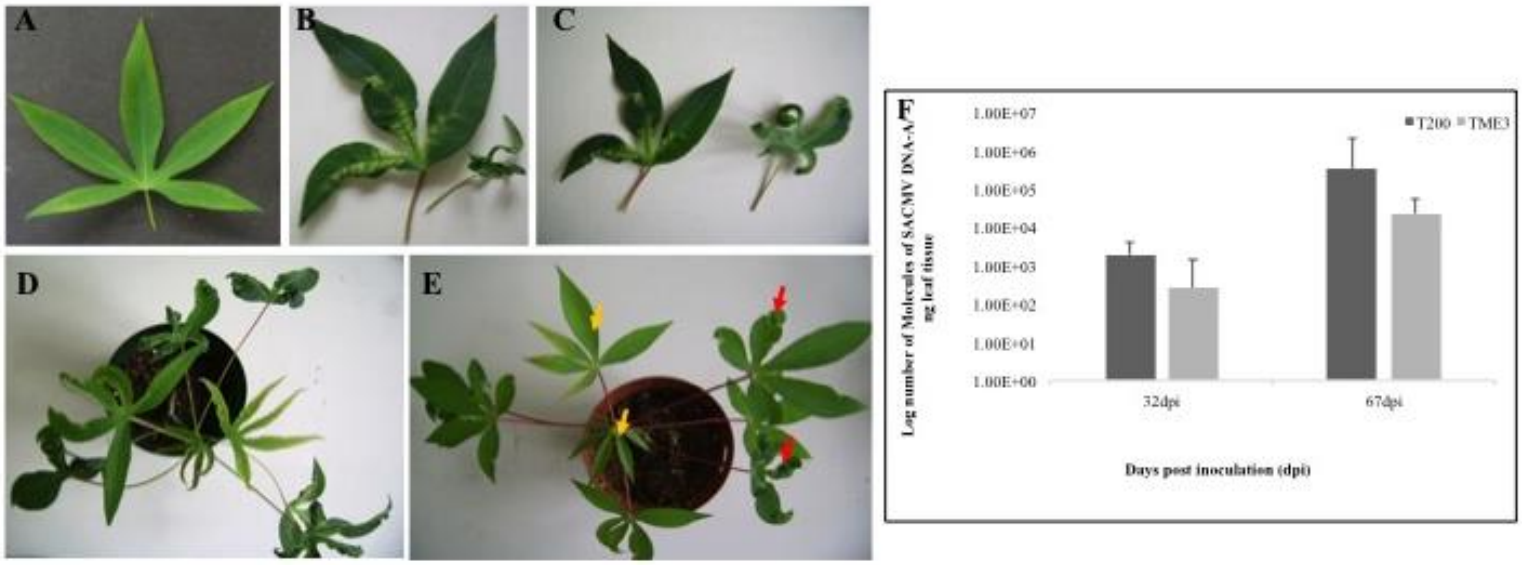


Figure 4.1 Symptoms observed in T200 and TME3 landraces of cassava infected with *South African cassava mosaic virus* (SACMV) at 32 and 67 days post-infection (dpi), and corresponding virus titres. Mock-inoculated plants (A) were symptomless at 32 dpi, while T200 (B) and TME3 (C) displayed typical mosaic, leaf curling and distortion. At 67 dpi, all newly emerged leaved in T200 (D) continued to be severely infected whereas the recovery phenotype was observed in TME3 (E) in newer emerging leaves (indicated with yellow arrows) compared to older infected leaves (indicated with red arrows). (F) The concentrations of SACMV DNA-A were measured in infected and mock-inoculated T200 and TME3 plants at 12, 32 and 67 dpi for 6 biological replicates in triplicate. Viral titre is represented as the Mean Log concentration of DNA-A molecules/ng TNA. SACMV was not detected in both T200 and TME3 at 12 dpi but while viral load increased over time in susceptible T200, it declined significantly in TME3 at 67 dpi (recovery).

Quantitative profiling of endogenous small nuclear RNA in SACMV-infected T200 and TME3 cassava landraces

Small RNA biogenesis factors are involved in plant immunity or susceptibility, and the production of small nuclear sRNAs and their targets altered by pathogen infection can influence the outcome of host-virus interactions. To analyse and compare SACMV interactions with small RNA (sRNA) silencing pathways, we deep-sequenced populations from mock-inoculated and SACMV-infected cassava T200 and TME3 at 12, 32 and 67 dpi. For each sample, total RNA was extracted from the top two apical leaves from six plants in each of three biological replicates and pooled, and the 12 enriched DNA libraries were sequenced using the Illumina HiSeq Analyzer. The Ambion method was chosen to create libraries of size-selected sRNAs less than ~ 200 nt since sRNAs involved in gene silencing are mostly considered to fall between 18- 26 nt size range (Bologna and Voinnet, 2014). All 12 libraries generated had Phred score values of greater than 20. The next generation sequencing (NGS) data is presented in Supplementary Table S2 Table and Table 4.1. Raw reads, actual and normalized counts for the enriched libraries, in addition to 18-26 nt counts, and counts for rRNA, tRNA and snoRNA are presented in Supplementary Table S2. The enriched libraries generated between 2.6 to 24 million high quality adapter-trimmed reads per sample (Table 4.1). In a study in apple, a narrow range library (NRL) was created where 97% of all reads were 17-26 nt, and the libraries generated between 7 to 14 million high quality reads per sample depending on cultivar and *Apple stem grooving virus* (ASGV) isolate (Visser et al, 2014).

Table 1: Next generation sequencing reads (< 200 nt) for the enriched cDNA libraries from mock and SACMV-infected cassava T200 and TME3.

Samples	Phred (Q) values	Adapter trimmed Reads			18-26 nt Reads		
		Actual Reads	Size of Data files (MB)	% of Raw Reads	Actual Reads	Size of Data files (MB)	% of Raw Reads
T200 12dpi SACMV	>24	**24 139 388**	685	70.54	**377 302**	1.10	**1.56**
T200 12dpi Mock	>28	5 921 995	166	87.42	**1 436 570**	21.21	**24.26**
T200 32dpi SACMV	>26	**22 498 560**	666	85.39	**771 382**	2.93	**3.43**
T200 32dpi Mock	>36	12 914 563	358	87.93	**4 381 089**	29.83	**33.92**
T200 67dpi SACMV	>28	**6 352 055**	173	90.89	**190 884**	2.73	**3.01**
T200 67dpi Mock	>24	20 603 045	556	88.95	**6 935 189**	29.94	**33.66**
TME3 12dpi SACMV	>24	**21 163 150**	617	88.86	**373 239**	1.57	**1.76**
TME3 12dpi Mock	>28	2 668 525	75	91.56	**634 217**	21.76	**23.77**
TME3 32dpi SACMV	>24	**15 233 143**	429	88.60	**890 603**	5.18	**5.85**
TME3 32dpi Mock	>20	3 625 654	95	79.00	**479 048**	10.44	**13.21**
TME3 67dpi SACMV	>22	**11 167 902**	306	86.56	**507 298**	3.93	**4.54**
TME3 67dpi Mock	>26	11 414 040	310	89.93	**2 435 899***	19.19	**21.34**

Notably, a consistent pattern in library reads was observed, where the total adapter trimmed reads (<200 nt) increased in both mock-inoculated T200 and TME3 over time (12 to 67 dpi), but decreased from 12 dpi to 32 and 67 dpi in both SACMV-infected T200 and TME3. In contrast, while the 18- 26 nt sized sRNA reads showed the same pattern as adapter-trimmed reads in mock T200 i.e. increased over time, in TME3 counts declined at 32 dpi and increased significantly at 67 dpi. In SACMV-infected T200 and TME3 samples a different pattern in 18-26 nt reads compared with total adapter-trimmed counts was observed, where 18-26 nt counts increased at 32 dpi and declined at 67 dpi. When mock is compared to infected apical leaf tissue in T200, interestingly, the total 18-26 nt sRNA reads were consistently lower in SACMV-infected T200 at all time points, whereas in TME3, while there was a significant decrease in infected vs mock samples at 67 dpi, at 32 dpi 18-26 nt sRNA counts increased. Different sRNA reads in mock-inoculated leaves between the two landraces likely represent different genetic backgrounds, but the highly significant reduction in percentage of expressed 18-24 nt sRNA in infected T200 and in TME3 represents the effect of SACMV infection.

Small endogenous RNA reads of each size class were calculated as a percentage of total sRNA population counts. These results are discussed in chapter 3. The 24 nt siRNAs were also represented in SACMV-infected TME3 and T200, but fluctuated depending on the age of the host post-inoculation (Fig. 3.1). Notably, there was no significant (<10%) change (in terms of the percentage of total sRNA reads) in prevalence of the TGS-associated 24 nt sRNAs in mock (Fig. 3.1C) vs SACMV infected (Fig. 3.1A) TME3 at each time point, but in T200 there was a significant increase during the early pre-symptomatic response (12 dpi) to SACMV (Fig. 3.1B) compared to mock (Fig. 3.1D). In the *Arabidopsis* study, leaves were harvested 30 days post CaLCuV infection, which represents the stage where symptoms are highly visible in the host and replication levels high (Aregger et al, 2012). They found a 7% decrease in 24 nt total sRNAs in CaLCuV infected compared to mock *Arabidopsis*, while in our study at 32 dpi, where virus replication is high and symptoms severe (Fig. 4.1), there was a 6% increase in TME3, and in T200 there was no change. There was also a 6% and 5% increase in

24 nt sRNAs at 32 dpi and recovery (67 dpi), respectively in TME3. There was no significant change (20% and 67% of total reads were represented at both 32 and 67 dpi, respectively) in total endogenous 24 nt sRNA percentages in susceptible T200 from mock-inoculated to SACMV infection. In conclusion, it is clear that SACMV alters levels of endogenous sRNAs. Notably we demonstrate that these changes fluctuate as infection in cassava progresses, but most importantly, we show that there are different patterns of sRNA responses to infection between the susceptible and tolerant genotypes/phenotypes, T200 and TME3, despite their different genetic backgrounds. In contrast to these cassava landraces, a study by Sahu et al. (2014), total snRNAs did not change in expression between mock and ToLCNDV inoculated tomato. More in depth studies on contrasting sRNA responses in different virus-host interactions may provide further clues as to their roles in susceptibility and tolerance in cassava.

vsRNAs map along the entire viral DNA components but accumulate at higher levels in susceptible T200 compared to tolerant TME3

Sequencing reads from cDNA libraries of 21-24 nt vsRNAs mapping to the SACMV DNA A and B genome are depicted in Supplementary Table S3. Notably the highest vsRNA reads targeting (100% match) SACMV DNA A and B components (~2800 nt and 2760 nt, respectively) were the highest in infected T200 at 32 and 67 dpi. Normalized redundant (total) counts were highest at 32 dpi targeting DNA A (20,582) and DNA B (13,196) in T200. In ASGV infected apple, the total (redundant) and non-redundant (unique) vsRNAs targeting ASGV genomes (~6400 nt) ranged from 1659 to 27069 and 1659 to 5897 counts, respectively, and genome coverage was 52-98% (Visser et al, 2014). In mock inoculated leaf tissue low numbers of virus-matched sRNAs were detected, suggesting that the cassava landraces harbour endogenous sRNAs that share homology to SACMV. Integration of geminivirus sequences have been reported (Chu et al, 2014), and a recent study identified Rep-like and capsid protein-like sequences in *Populus trichocarpa* and *Nicotiana tabacum*, respectively (Liu et al, 2011). Host transcripts targeted by vsRNAs were identified in grapevine infected

with two viruses (Miozzi et al, 2013), however relatively few vsRNAs were involved in RNA silencing of host mRNAs. Studies propose that these endogenous genetic elements may be involved in viral resistance/susceptibility (Bertsch et al, 2009; Maredza et al, 2015), but the effect, if any of the vsRNA host gene targeting is poorly understood. In virus-infected plants two classes of vsRNAs have been identified, namely primary vsRNAs which result from the initial Dicer-mediated cleavage of the initial viral trigger RNA and secondary vsRNAs generated by RDR (Ding and Voinnet, 2007; Ruiz-Ferrer and Voinnet, 2009). It is likely that the SACMV-triggered vsRNAs are primary siRNAs derived from dsRNA generated by bidirectional read through transcription of viral DNA by RNA polymerase II. Geminiviral mRNAs appear to be poor templates for RDR-dependent production of secondary siRNAs (Aregger et al, 2012).

Several patterns emerged from the vsRNA deep sequence data, and differences were considered significant if there was an alteration in normalized vsRNAs greater than 10%. Virus-derived vsRNAs were detected at very low levels in either T200 or TME3 at the early pre-symptomatic stage of infection (12 dpi) (Fig. 4.2), but there was a increase in counts from 12 to 32 dpi targeting DNA A and B in both T200 and TME3 landraces (Fig. 4.2) as the plants responded to infection. However, deep sequencing results showed that total numbers of vsRNAs and unique vsRNAs were higher in T200 compared with TME3 at 32 and 67 dpi, and total vsRNA counts targeting DNA A were higher compared with DNA B in both landraces (Fig. 4.2). In contrast, in a different study with ToLCNDV susceptible and tolerant cultivars of tomato no considerable difference was observed between the vsRNAs derived from DNA B (Sahu et al, 2014). Since T200 is highly susceptible and virus loads are high, we conclude that high accumulation of vsRNAs observed in susceptible T200, especially at 32 dpi when symptoms are severe, is due to failure to target SACMV-derived mRNA resulting in replication and symptom persistence. Similarly, high levels of 21, 22 and 24 nt vsRNAs were reported in *Arabidopsis* infected with CaLCuV, yet plants remained highly susceptible (Aregger et al, 2012). In contrast, in tolerant TME3, low vsRNA numbers correlated positively with virus titres, symptoms and recovery, and may represent efficient PTGS of viral mRNA, leading to a depletion/sequestration of

vsRNA populations, which in turn reduces the levels of viral proteins, and subsequent virus replication. Thus, in recovered tissues, as the levels of viral mRNAs decrease, the corresponding small RNAs are also simultaneously reduced. Hagen et al. (2008) also demonstrated that the abundance of CuLCrV-derived small RNAs was negatively correlated with recovery in watermelon and cantaloupe. Recovery in pepper from infection with PepGMV (Carrillo-Tripp et al, 2007) showed similar findings as observed in case of the CuLCrV infection (Hagen et al, 2008). These results differ from earlier studies (Chellappan et al, 2004; Sahu et al, 2012; Sahu et al, 2014; Yadav and Chattopadhyay, 2011) where high numbers of vsRNAs correlated with reduced virus replication during recovery from begomovirus infection. Notably, in contrast to TME3, the study with ACMV in cassava cv. 60444 showed an increase in vsRNAs in tolerance and recovery (Chellappan et al, 2004). Contrasting results suggest that recovery is both host specific and geminivirus species dependent. Viral RNA silencing suppressor proteins (VSRs) (Lewsey et al, 2009) typically counteract RNA-mediated defence by (1) preventing the generation of siRNAs, (2) by inhibiting the incorporation of siRNA molecules into effector complexes or (3) by interfering the RISC effector complex (Lakatos et al, 2004). Since high numbers of vsRNA accumulate in T200, we conclude that SACMV counteracts defence by either or both the latter two mechanisms.

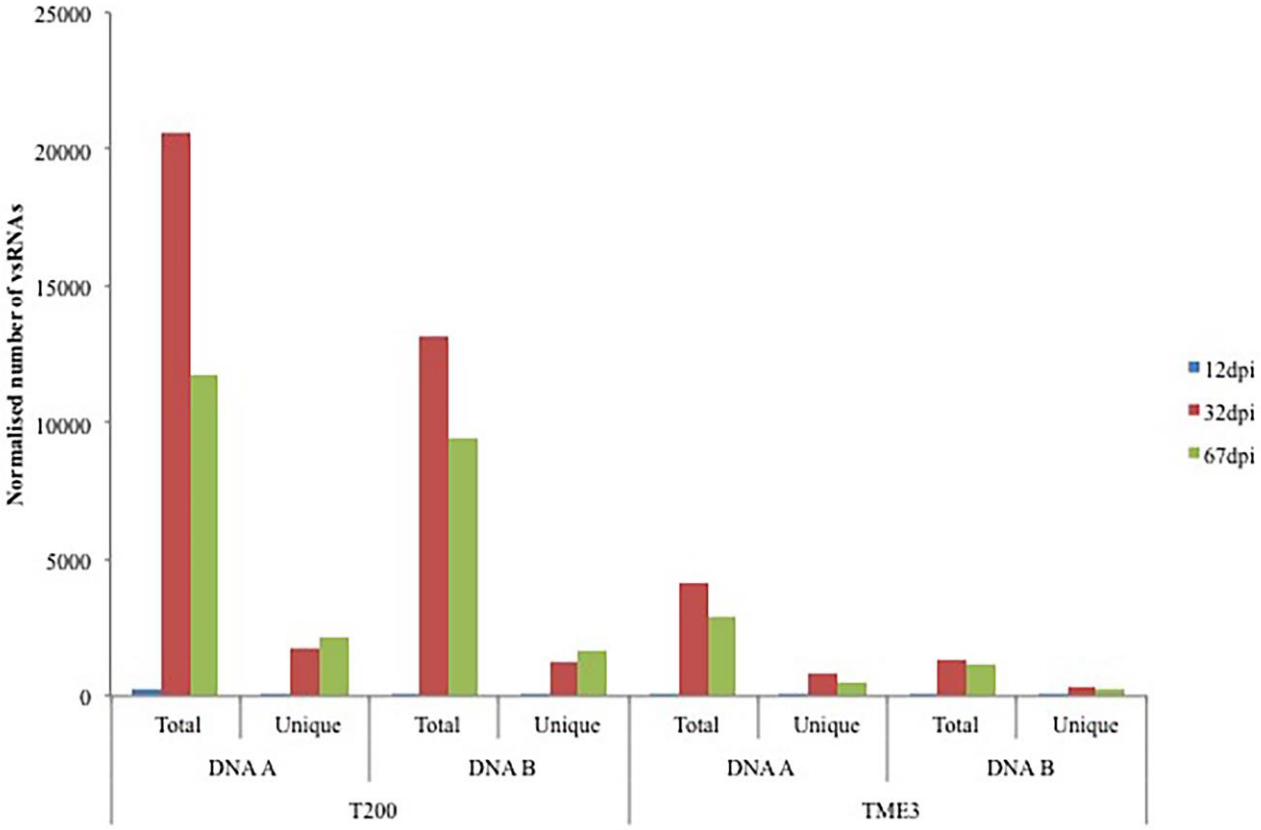


Figure 4.2 Total and unique number of vsRNAs mapping to SACMV DNA A and DNA B in SACMV-infected T200 and TME3 at 12, 32 and 67 dpi. There was a very low level of vsRNAs at the early pre-symptomatic 12 dpi stage. Total numbers of vsRNAs and unique vsRNAs are significantly higher in T200 compared with TME3 at 32 and 67 dpi. vsRNA counts targeting DNA A are higher compared with DNA B for both T200 and TME3 at all three time points.

Both antisense and sense total (redundant) vsRNAs counts were represented at 12, 32 and 67 dpi in T200 and TME3, but dropped as infection progressed (67 dpi) (Fig. 4.3A), with the exception of the 24 nt vsRNA targeting A and B in TME3, demonstrating that both the genomic sense ssDNA and complementary templates formed during rolling circle replication are targeted. Redundant antisense vsRNA reads expressed as total counts (Fig. 4.3A) or as a percentage of total vsRNA at each time point (Fig. 4.3C) were higher than sense in T200, suggesting a decline in targeting of the antisense ssDNA strands which form the template for the sense genomic strands during replication at this time point, which positively correlates with increased virus titres. Interestingly, while

redundant antisense vsRNA reads in T200 are highest targeting DNA A at 32 dpi, compared to 12 and 67 dpi and all time points in TME3, a different pattern emerges when the percentage of vsRNA is calculated as a percentage of total vsRNA reads at each time point. In this case, while numbers of vsRNAs are low, the percentage of sense and antisense vsRNA is high at (12 dpi), in particular targeting DNA B in T200 (Fig. 4.3C).

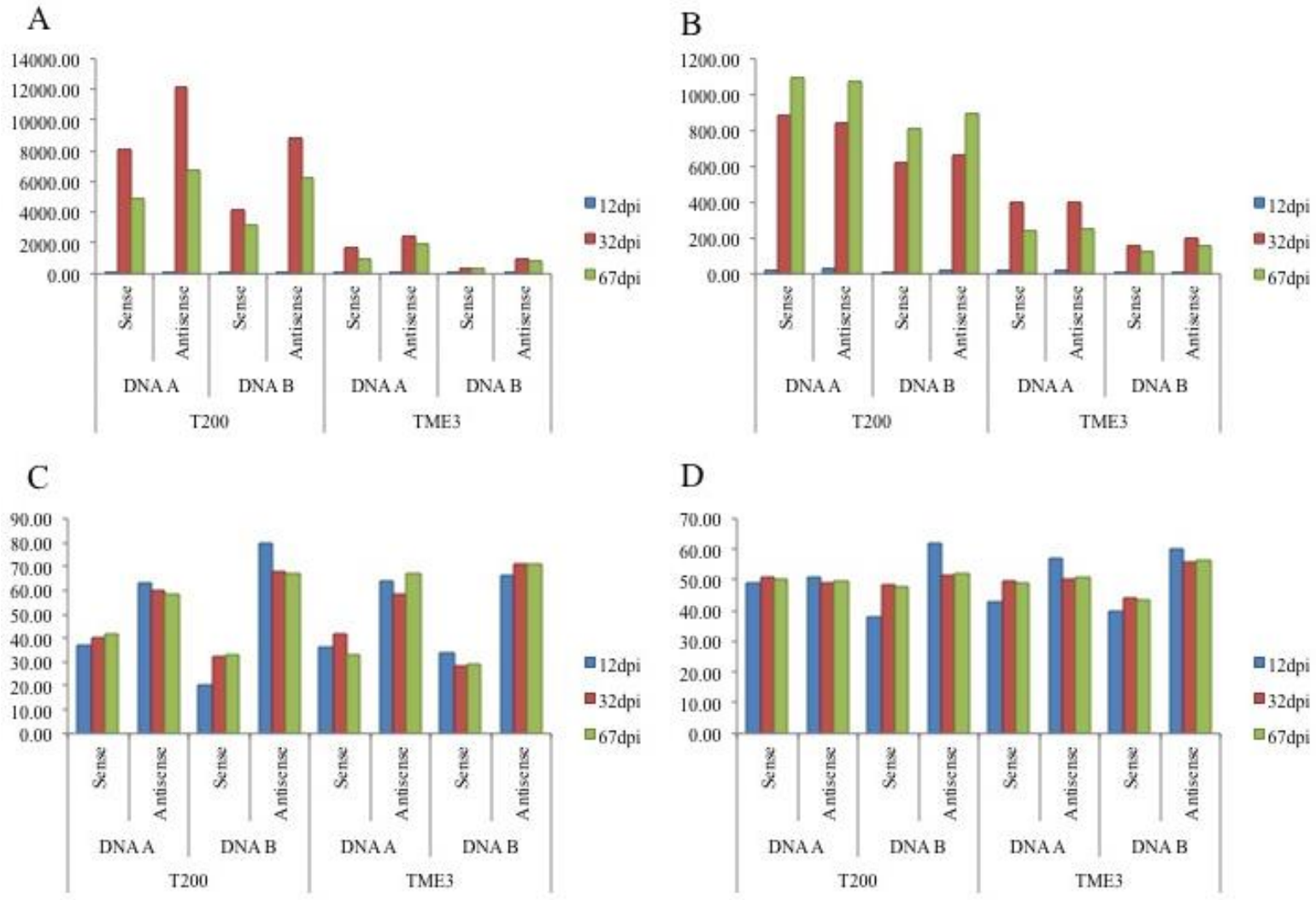


Figure 4.3 A - D: The number of total (A) and unique (B) sense and antisense vsRNA sequences that map to SACMV DNA A and DNA B in TME3 and T200 cassava landraces at 12, 32 and 67 dpi. Number of total (C) or unique (D) vsRNAs expressed as a percentage of total mapped vsRNAs. Sense and antisense vsRNAs are not equally represented at each time point in T200 and TME3 and generally total counts decline from 32 to 67 dpi.

The number of unique (non-redundant) sense and antisense vsRNA reads were also higher in T200 compared to TME3 (Fig. 4.3B), as expected. However, while a bias towards antisense is noted at the respective time points for total counts, no bias between sense and antisense polarities could be discerned for unique vsRNAs targeting DNA A and B components in T200 and TME3 (Fig. 4.3D). Strand biases are usually attributed to preferential processing of highly structured single-stranded genomic viral RNAs by DCLs (Ding and Voinnet, 2007) and different viruses have been shown to produce, in the same host plant, virus-derived small RNAs with different ratios of sense to antisense polarity (Pantaleo, 2011).

While vsRNA populations remained consistently higher in T200 compared with TME3, in T200 there was a significant decrease in total vsRNAs targeting DNA A (43%) and B (29%) from the systemic infection stage (32 dpi) to the 67 dpi time point. A similar pattern was noted in TME3 where there was a 30% and 15% decrease in vsRNAs targeting DNA A and B, respectively. Interestingly, while the total counts declined at 67 dpi for both T200 and TME3 from 32 to 67 dpi, the number of unique (non-redundant) vsRNAs increased for T200 but the pattern remained the same (declined at 67 dpi) in TME3. We conclude from this data that the large increase in virus load from 12 dpi to 32 dpi for T200 and TME3 (Fig.4.1) implies that siRNAs are not successfully targeting virus-derived mRNA hence suppressing virus replication, but a decline in total sense and antisense vsRNA counts at 67 dpi is indicative of the host attempting to counterattack SACMV by vsRNA-mediated PTGS. Whether plants are tolerant, resistant or susceptible, basal innate immunity is always detected, but the outcome is a result of multiple complex interacting factors. Cassava is perennial, and T200 continues to grow despite symptom persistence and considerable virus load for extended periods of time (we have monitored this over 18 months; data not shown), and we believe that there is a persistent but low innate defence response which we have termed non-recovery accommodation (Bengyella et al, 2015) where T200 and SACMV may co-exist without death of the host, a hallmark perhaps of some other virus-infected perennial crops. In contrast, TME3 exhibits recovery and tolerance where, contrary to non-recovery accommodation, virus replication is detectable

at lower levels compared to T200 and mild to no symptoms are observed. While tolerance and recovery in TME3 was shown from this study to be associated with adaptive antiviral mechanisms mediated by PTGS, other molecular mechanisms are involved. A more recent study revealed that structurally specific resistance gene analogs (RGAs) participate in tolerance in TME3, and differentially accumulate during recovery as a complementary defence mechanism to natural occurring RNA silencing to impair viral replication (Louis and Rey, 2015).

Size class abundance of vsRNA populations differ between T200 and TME3 at different time points post infection and between DNA A and B components

Virus-infected plants accumulate vsRNAs of three major size-classes, namely 21, 22 and 23 nt (Pantaleo, 2011). Additionally, in geminivirus infections TGS-associated 24 nt vsRNAs have also been demonstrated (Aregger et al, 2012). All total and unique 21-24 nt vsRNA size classes were found to target DNA A and B in T200 and TME3 at 32 and 67 dpi, but the distribution patterns and population counts differed between the two landraces (Fig. 4.4). In *Arabidopsis*, DCL4 and DCL2 act redundantly to produce 21 or 22 nt siRNAs responsible for antiviral silencing (Bouche et al, 2006; Deleris et al, 2006) whereas DCL3 gives rise to 24 nt siRNAs that are not active in directing RNA cleavage (Deleris et al, 2006; Fusaro et al, 2006). Since all size classes were detected in T200 and TME3 we conclude that DCL2, 3 and 4 cassava homologues were responsible for generation of vsRNA populations. The presence of 21-23 nt size vsRNA classes is evidence that a PTGS response is initiated by both T200 and TME landraces in response to SACMV. Notably all redundant size vsRNA classes targeting DNA A and B were more highly represented in terms of total counts in T200 at 32 dpi than 67 dpi (Fig. 4.4A) but the number of non-redundant (unique) 21-24 nt vsRNA counts increased at 67 dpi (Fig. 4.4B). In contrast, in TME3, the non-redundant 21-24 nt vsRNAs declined at 67 dpi (Fig. 4.4B), with the exception of DNA B-targeting 24 nt vsRNAs.

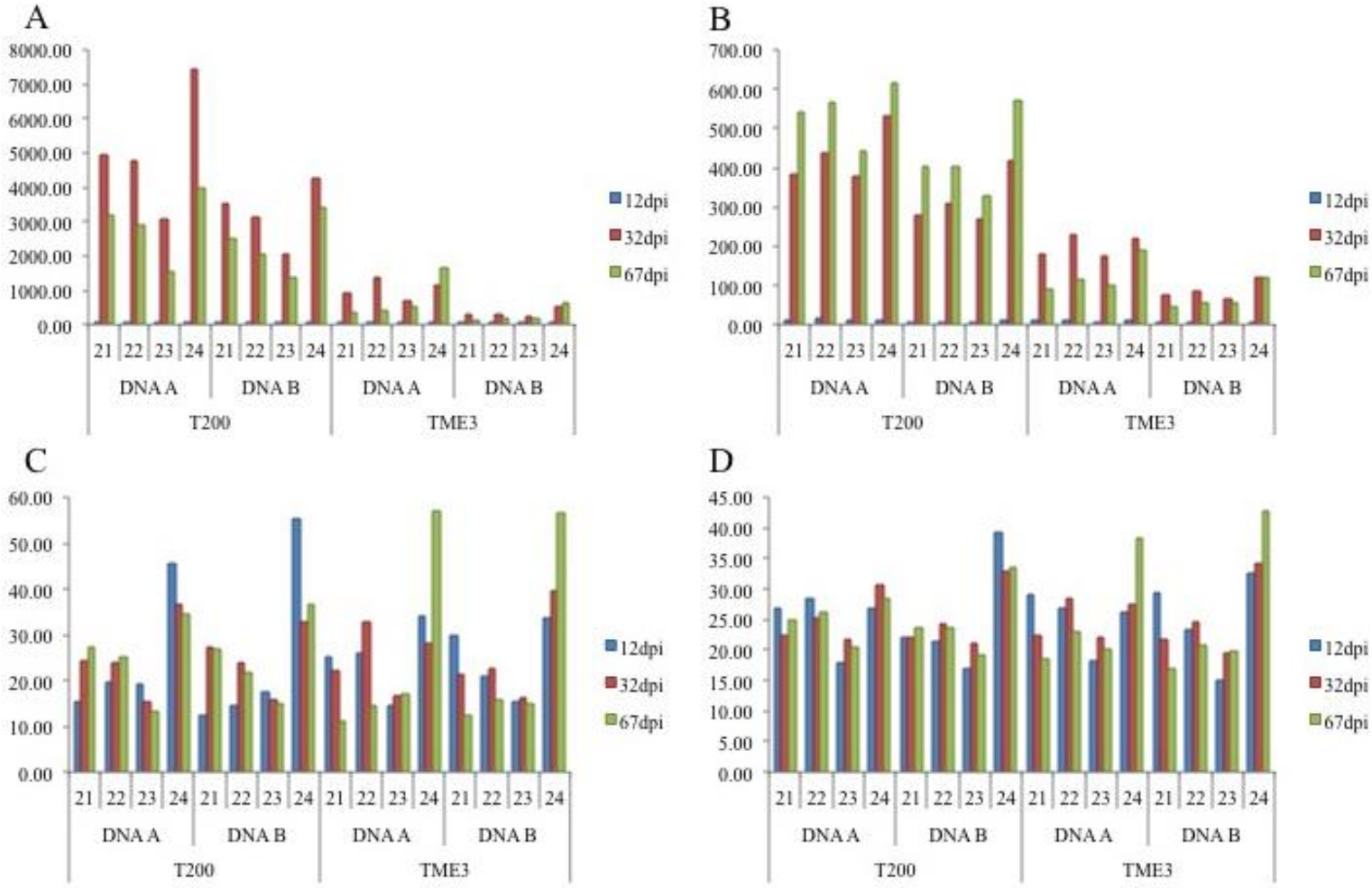


Figure 4.4 Size categories of total (A) and unique (B) vsRNAs counts, and total (C) and unique (D) counts expressed as a percentage of total vsRNAs, mapping to SACMV DNA A and DNA B in T200 and TME3 at 12, 32 and 67 dpi. All categories are less abundant in TME3 compared to T200. The redundant 24 nt category counts targeting DNA A and B are more highly represented in T200 at 32 and 67 dpi. In TME3 both the redundant 22 and 24 nt vsRNA counts targeting DNA A are the most abundant at 32 dpi compared to the other classes. Notably the percentage (as a percentage of total 21-24 nt vsRNAs) of TME3 redundant and non-redundant 24 nt vsRNAs targeting DNA A and B increase significantly at 67 dpi (recovery).

The 24 nt vsRNAs were predominant in T200 at 32 dpi, with 38% and 32% of total 21-24 nt vsRNAs targeting DNA A and B, respectively (Fig. 4.4A), and 31% and 33% of total unique vsRNAs counts (Fig. 4.4B) targeting DNA A and B, respectively, and were comparatively higher than the 21 and 22 nt size classes,

implicating increased activity of cassava homologues of DCL3 (Blevins et al, 2006) in response to virus infection. Interestingly, while actual redundant and non-redundant 21-24 nt counts were very low early in infection at 12 dpi in T200 and TME3 (Fig. 4.4A and B), when expressed as a percentage of total vsRNA classes, in T200 the redundant 24 nt vsRNAs targeting DNA A and B represented 45 % and 55%, respectively (Fig. 4.4C), and the non-redundant 27% and 38% targeting DNA A and B, respectively (Fig. 4.4D). In TME3, while overall redundant and non-redundant counts were significantly lower compared to T200, at 32 dpi, when counts were represented as a percentage of total vsRNAs at each time point a different picture emerged. It was shown that a sizable percentage of total 24 nt RNAs was represented in TME3, with 28% and 39% of total 24 nt vsRNAs targeting DNA A and B, respectively (Fig. 4.4C), while 27% and 34% of the unique 24 nt vsRNAs targeted DNA A and B, respectively (Fig. 4.4D). At 67 dpi, 38% and 43% of total 24 nt vsRNAs targeted DNA A and B in TME3, respectively, demonstrating a 11% and 9% increase in 24 nt vsRNAs targeting DNA A and B, respectively, from 32 to 67 dpi. Populations of 24 nt vsRNAs, produced by DCL3, have been reported to be the most abundant in DNA virus infected tissues (Blevins et al, 2006). In the dsDNA *Cauliflower mosaic virus* (CaMV) over-accumulation of the 24-nt siRNAs was observed in infected Arabidopsis (Blevins et al, 2006). While TME3 showed a late stage (67dpi; recovery) increased response in 24 nt vsRNA generation, T200 exhibited an early response at 12 dpi which was undetectable at 12 dpi when total or unique counts were scrutinized (Fig. 4.4A and B) but were noticeable when figures were calculated as a percentage of total vsRNAs at each time point. The percentage of 24 nt vsRNAs targeting both DNA A and B components at 12 dpi was higher in T200 compared to TME3, illustrating generation of this class of siRNAs as an early response by this susceptible genotype. What was also interesting was that while total vsRNA counts are more highly represented in T200 at 32 dpi, the numbers of unique vsRNAs are higher at 67 dpi in T200. We speculate that this may be due to a small degree of secondary vsRNA amplification of overlapping RNA fragments by RNA-dependent RNA polymerase 6 (Ruiz-Ferrer and Voinnet, 2009).

DNA A and B genome methylation does not play a role during recovery in TME3

From the 24 nt vsRNA results in TME3, it was speculated that the recovery phenotype may be the result of TGS of genomic SACMV DNA. Symptom recovery is a phenomenon reported in several plant studies, including ACMV-[CM] and SLCMV infected cassava and *N. benthamiana*, and pepper infected with the geminivirus, *Pepper golden mosaic virus* (PepGMV) (Rodríguez-Negrete et al, 2009), and has been associated with 21-24 nt siRNAs. Furthermore, cassava landrace TME7 when inoculated with EACMV-Ug alone or in combination with *Cassava brown streak virus* (CBSV), developed typical mosaic symptoms on the first emerging leaves followed by a recovery phenotype where CMD disease symptoms were reduced (Vanderschuren et al, 2012). Since methylation of geminivirus genomes has previously been associated with recovery for geminiviruses such as *Beet curly top virus* (BCTV) (Akbergenov et al, 2006), *Mungbean yellow mosaic India virus* (Yadav and Chattopadhyay, 2011), and ACMV in cassava (Akbergenov et al, 2006; Chellappan et al, 2004; Ermak et al, 1993), we predicted that methylation may play a role in SACMV-infected tolerant TME3, but not susceptible T200. TGS has two major roles, one of which involves defending the host plant against invasive DNA, such as ssDNA geminiviruses. (reviewed in Hohn and Vazquez, 2011; Raja et al, 2010). To analyse whether SACMV DNA A and B were extensively methylated via TGS, we performed bisulfite sequencing on viral DNA extracted from infected T200 (susceptible) and TME3 (tolerant) leaf tissue at 12, 32 and 67 dpi. TME3 is a West African landrace that has in earlier years been described as resistant to CMD (Akano et al, 2002; Dixon et al, 2001; Fregene et al, 2004), but in fact shows a tolerant/recovery phenotype (Allie et al, 2014). Bisulfite sequencing (Frommer et al, 1992) did not reveal any specific patterns of methylation of any of the ORFs and IR's on SACMV DNA-A or DNA-B that were amplified from T200 and from TME3 leaf tissue. Sequence analysis showed that all cytosine residues in SACMV on either DNA-A or DNA-B were successfully converted to thymine, which is an indication that no cytosines were modified by host methyltransferases (Supplementary Table S4). This result would be expected in T200 as this is highly susceptible to SACMV and shows severe persistent symptoms over 67 days, suggesting suppression of host TGS, but in

TME3 it was anticipated that genome methylation would have occurred at 67 dpi (recovery). It is possible that the absence of genome methylation detection could be attributed to a number of factors. Since methylation is a reversible process, SACMV may have been able to reverse or reduce DNA methylation to low-to-undetectable levels. Additionally, methylation of geminivirus DNA is conformation selective (Paprotka et al, 2010), and this may play a role at different stages of infection. It has also recently been proposed that RNA silencing induction is triggered by a threshold of virus accumulation in the leaves (Santovito et al, 2014), and low levels of SACMV in TME3 (Fig. 4.1F) may have been below the threshold of TGS induction. Correlation between high numbers of 24 nt siRNAs and RNA-directed DNA-methylation associated with recovery from geminiviruses has been reported in the literature (Aregger et al, 2012; Bian et al, 2006; Brough et al, 1992; Raja et al, 2010; Rodríguez-Negrete et al, 2009). However, while the percentage of 24 nt vsRNAs increased from 32 to 67 dpi in TME3, vsRNA counts targeting the IR and CR (Fig. 4.5), associated with TGS, were highly under represented compared to T200. It is possible that since there is no evidence for genome methylation by sequencing, that methylation of histones associated with mini-chromosomes in the nucleus, which has been shown in geminivirus infections (Pilartz and Jeske, 1992; Pilartz and Jeske, 2003), may be linked to tolerance and recovery in TME3. A decrease in the expression of some host methyltransferase genes in Arabidopsis infected with *Tomato yellow leaf curl Sardinia virus* has been demonstrated (Rodríguez-Negrete et al, 2013).

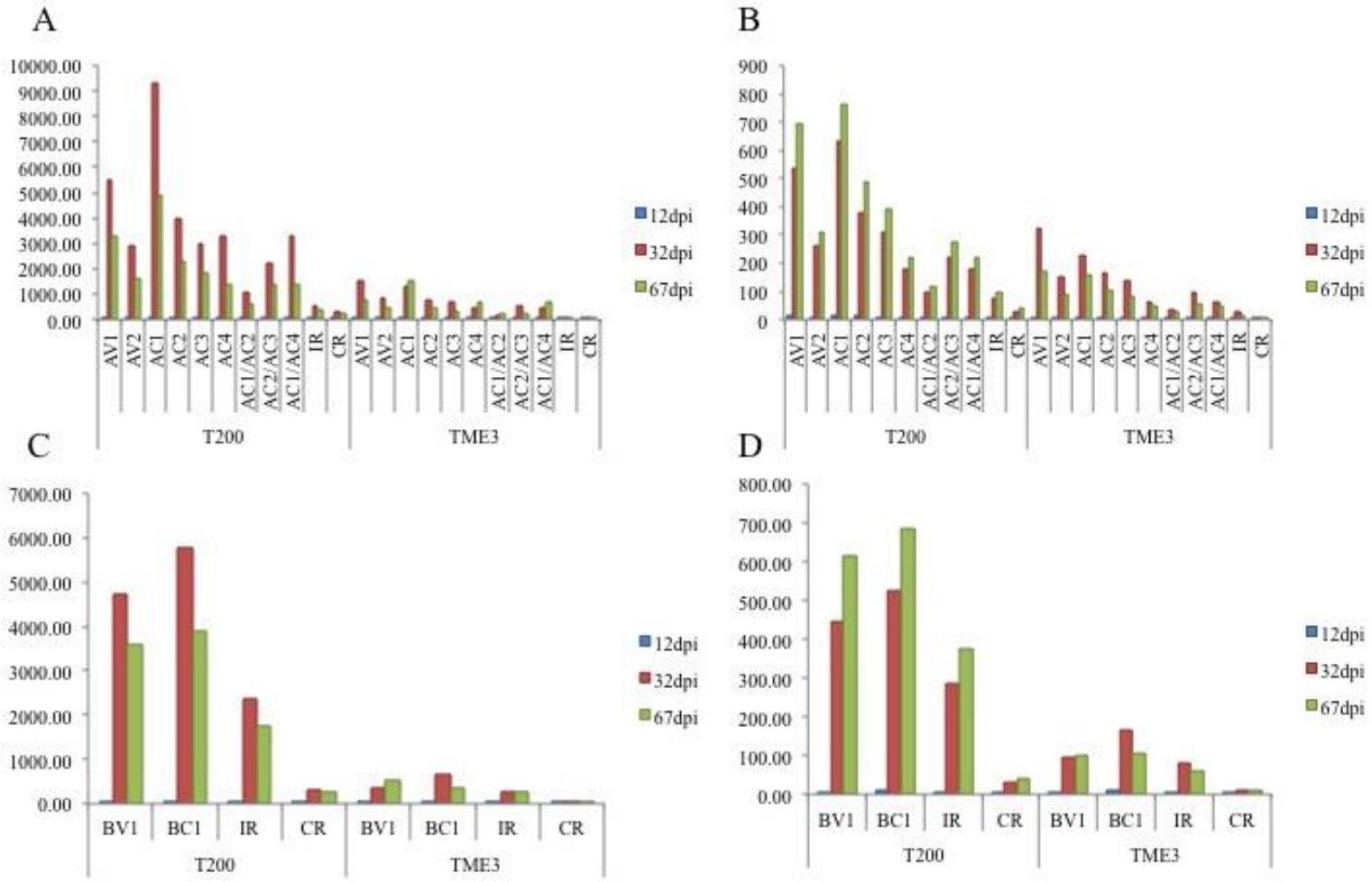


Figure 4.5: The total (redundant) and unique (non-redundant) vsRNAs mapping to each ORF of SACMV DNA A and DNA B in T200 and TME3 at 12, 32 and 67 dpi. (A) Total counts DNA A; (B) Unique counts DNA A; (C) Total counts DNA B; (D) Unique counts DNA B.

However, in a transcriptome profiling study by Allie et al (2014), there was no clear evidence for extensive histone methylation reprogramming at 12, 32 dpi or 67 dpi, or for significant alterations in any DNA-methylation associated enzymes or proteins throughout infection, with the exception in TME3 at 67 dpi (recovery), where there was a significant down-regulation (-3.175 log2 fold) of histone acetyltransferase of the MYST family 1 (cassava4.1_029570m.g; AT5G64610.1) in recovered leaves (Allie et al, 2014) and in susceptible T200 where a methyltransferase (cassava4.1_022835m.g; AT5G10620.1) was down-regulated at 32 dpi. In the absence of any genome or histone methylation evidence, the extent of the role of TGS-associated 24 nt RNAs in the outcome of disease phenotype in T200 or TME3 remains unclear.

DNA viruses encounter both TGS and PTGS, but the disease resistance/susceptibility outcome is complex, depending not only on the strength of virus suppressors of RNAi, but also on host plant age, lifespan of host (annual or perennial), other genetic co-factors and aspects of the agro-ecosystems. It is interesting to note that DNA virus promoters used to drive expression of transgenes can be targeted by repressive methylation, but this may depend on the context of the host chromatin. Since geminivirus proteins, e.g. AC2 and AC4 actively suppress methylation (Vanitharani et al, 2005), it is possible that host genome instability due to transposon activation may be a component of cassava mosaic disease pathogenesis. Recently, involvement of transposon-like elements in CMD modulation in cassava has been demonstrated (Maredza et al, 2015).

vsRNA hotspots targeting ORFs differed between susceptible T200 and tolerant TME3

The relative abundance of total (Fig. 4.5A and C) and unique (Fig. 4.5B and D) vsRNAs varied between the ORFs (hotspots) across DNA A and B of T200 and TME3 (Fig. 4.5A - D), but a notable difference was that numbers were significantly lower in TME3 compared to the susceptible landrace T200 at 32 and 67 dpi. As was the case for total and sense and antisense vsRNA reads, prior to symptom appearance, vsRNAs were almost undetectable at 12 dpi. Total and unique vsRNAs targeting the intergenic region (IR)/common region (CR) were low (less than 100) for both T200 and TME3 for DNA A (Fig. 6A and B) and DNA B (Fig. 4.5C and D), with the notable exception of the IR in DNA B of T200. As was the pattern with sense/antisense and size class counts, the trend was the same where for T200 the total counts for all ORFs was highest at 32 dpi for DNA A and B components (Fig. 4.5A and C), but the number of unique vsRNAs increased for all ORFs as infection progressed (at 67 dpi) (Fig. 4.5B and D). Interestingly the trend was opposite for TME3 where numbers declined from 32 to 67 dpi for DNA A and DNA B (with the exception of *BV1*). For both T200 and TME3, vsRNAs were over-represented in *AC1*, *AV1, BV1* and *BC1* regions (Table 4.2), in particular *AC1* and *AV1* in T200 and TME3, respectively at 32 dpi. One major difference was that in TME3, the percentage of unique vsRNAs (based on the total number of vsRNAs;

Fig. 4.2) targeting the coat protein (*AV1*) was the highest (40% and 36% at 32 and 67 dpi, respectively), while in T200 *AC1* (Rep) had the highest vsRNA targets (37% and 35% at 32 and 67 dpi, respectively). For T200 and TME3, the percentage of unique vsRNAs targeting *BC1* (cell-to-cell movement) was the most prevalent for T200 (41% and 40% at 32 and 67 dpi, respectively) and TME3 (48% and 39% at 32 and 67 dpi, respectively).

Table 4.2 Total (redundant) and Unique (non-redundant) reads, and percentages of the total vsRNA reads 21-24 nt vsRNAs targeting SACMV AV1, AC1, BV1 and BC1 at 32 and 67 days post infection.

	32 dpi															
	AV1				AC1				BV1				BC1			
	Total		Unique		Total		Unique		Total		Unique		Total		Unique	
Landraces	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
T200	5464	27	535	30	9279	46	635	37	4744	36	446	35	5759	44	524	41
TME3	1550	37	320	39	1299	31	227	28	356	27	96	28	673	51	167	48
	67 dpi															
	AV1				AC1				BV1				BC1			
	Total		Unique		Total		Unique		Total		Unique		Total		Unique	
Landraces	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
T200	3303	28	697	32	4889	42	762	35	3611	38	614	36	3902	41	683	40
TME3	736	25	175	35	1495	51	162	33	524	45	102	37	344	30	107	38

Interestingly, if the number of vsRNAs were normalised (per 100 nt) to the length of the ORFs, *AV1* was more highly targeted than *AC1* in T200, and the prevalence of vsRNAs per 100 nt in the *AC2/AC3* overlapping region increased significantly (Supplementary Figure S5). Furthermore, *BV1* was more highly targeted compared with *BC1* in T200. When comparing the 5'-ter, 3'-ter and central domains of the targeted ORFs of SACMV, the general patterns were the same in T200 and TME3 (Supplementary Figure S6). Not all ORFs on DNA A and B showed the same pattern. For example, *AV2* was targeted more highly at the 3'-ter compared to the 5'-ter and central region, while in contrast in *AC2* both the 5' and 3' termini were highly targeted, especially at the early stage (12 dpi) of infection. In a study by Patil and Fauquet (Patil and Fauquet, 2015) investigating the infection dynamics of several species of cassava mosaic geminiviruses (CMGs) and their isolates in *Nicotiana benthamiana*, it was found that in most CMGs the regions corresponding to the 3'-ter of *AC1* and *BC1* had higher densities of siRNAs compared to the other ORFs. Notably, in TME3 *AC4*-targeting vsRNAs were highly represented at 32 dpi prior to recovery (Supplementary Figure S5). *AC4* and *AC2* are associated with suppression of host RNA silencing (Gupta et al, 2014; Latham et al, 1997; Vanitharani et al, 2004). For *BV1*, all regions were highly targeted at all time points post infection, while for BC1 there was a spike in numbers at the 3'-ter at 12 dpi, but in T200 the numbers decreased thereafter but remained high in TME3. In cassava infected with ACMV-[CM], a recovery-type virus, the 3'-ter of *AC1* was the primary target, while the 3'-ter of *BC1* was targeted by *East African cassava mosaic Cameroon virus* (EACMCV), a non-recovery type geminivirus (Chellappan et al, 2004; Pita et al, 2001). Interestingly, in this study T200, a non-recovery host to SACMV, both *AC1* and *BC1* were highly targeted. SLCMV, also a recovery-type geminivirus, behaved differently in two hosts, cassava and *N. benthamiana*. In cassava, plants recovered at a later stage of the infection cycle with high siRNA accumulation similar to ACMV-[CM], but in *N. benthamiana* the plants were highly susceptible and died at 3 weeks post infection, with low siRNA levels detected (Chellappan et al, 2004). In *Arabidopsis* susceptible to CaLCuV, the highest abundance of reads targeted the *AV1* ORF (Aregger et al, 2012). This study,

concomitant with previous studies, illustrates differences in vsRNA hot spot targets on CMG genomes are host-virus interaction specific.

Comparisons were performed between T200 and TME3 in terms of 24 nt unique vsRNA abundance targeting *AV1*, *AC1*, *BC1*, and *BV1* (Fig. 4.6). For all four ORFs the 24 nt size class had the highest number of vsRNAs targeting the ORF for both T200 and TME3 at 32 and 67 dpi. If one examines the number of vsRNAs as a percentage of the total unique vsRNAs, at 32 and 67 dpi there appeared to be a small difference (4% and 8%, respectively) in the percentage of 24 nt vsRNAs targeting *AV1* between T200 and TME 3, in contrast, T200 and TME3 had 61% and 29% 24 nt vsRNAs (32% difference) targeting *AC1*, respectively, at 32 dpi (Fig. 4.6A). At 67 dpi there was a 12% difference in 24 nt vsRNAs targeting *AC1* between T200 (30%) and TME3 (42%) (Fig. 4.6B). In both T200 and TME3 DNA B *BC1* had the highest vsRNA targets, but T200 vsRNA populations were significantly more abundant than TME3 (Fig. 4.6C). A similar pattern was noted for *BV1* where in T200 vsRNAs were significantly higher compared to TME3 (Fig. 7d). As in the case of *AV1* and *AC1* in T200, the numbers of vsRNAs increased from 32 to 67 dpi, while in TME they declined (Fig. 4.6C).

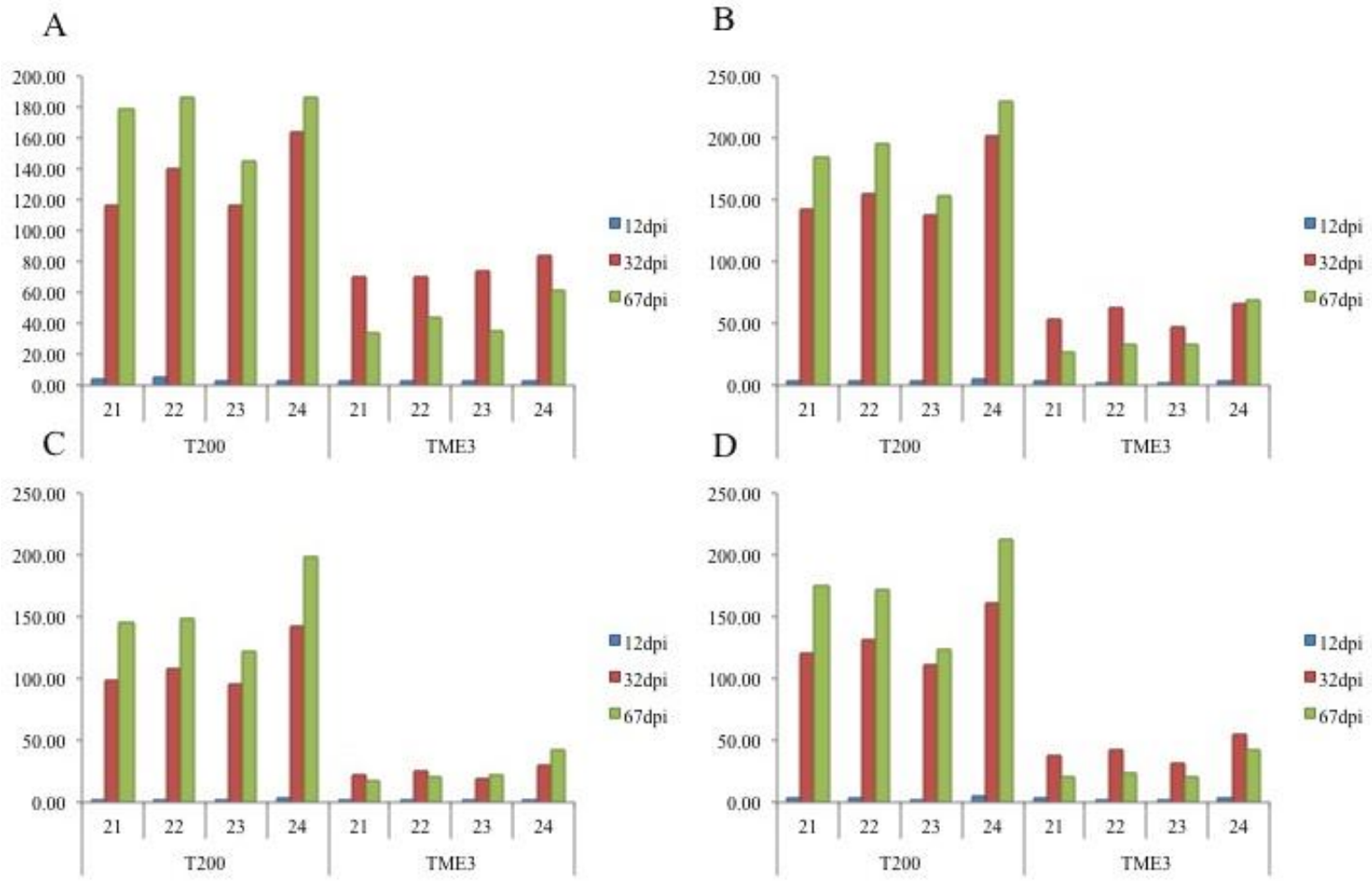


Figure 4.6 Number of unique vsRNA size categories mapping to SACMV *AV1* (A), *AC1* (B), *BV1* (C) and *BC1* (D) ORFs in T200 and TME3 at 12, 32 and 67 dpi. For all four ORFs the 24 nt size class had the highest number of vsRNAs targeting the ORF for both T200 and TME3 at 32 and 67 dpi.

In summary, our vsRNA results clearly demonstrate differences between the susceptible T200 and tolerant TME3 landraces in response to SACMV infection with regard to RNA silencing. Results herein suggest, that while 21-24 nt vsRNAs are generated by both susceptible and tolerant cassava landraces, the expected vsRNA numbers are contrary to other reported geminivirus studies (Chellappan et al, 2004; Rodriguez-Negrete et al, 2009). TME3 demonstrated a significantly lower vsRNA response compared to T200, while symptoms and virus load were lower, hallmarks of tolerance. In contrast, induction of 24 nt vsRNAs have been reported from recovered leaves in pepper and cassava (Chellappan et al, 2004; Rodriguez-Negrete et al, 2009). It is not unreasonable to suggest, as mentioned earlier, that in fact lower 24 vsRNAs in TME3 are a result of a steady targeting of

SACMV mRNA and subsequent depletion, manifested by attenuated symptoms and virus load (Fig. 4.1), which are hallmarks of tolerance. Overall, results suggest that RNA silencing is not effective in counteracting SACMV in T200 as vsRNAs accumulate in abundance. In TME3, low 21-23 nt vsRNA counts suggest a role for PTGS of SACMV DNA A and B resulting in significantly reduced viral loads compared with T200, contributing to the tolerance and recovery phenotype. RNA silencing is not the only factor in determining disease outcome in T200 and TME3 phenotypes, as transcriptome reprogramming was shown to differ between these two landraces (Allie et al, 2014). Different hosts respond variably to different geminiviruses since TGS has been shown to be associated with a cassava recovery phenotype with ACMV-[CM] (Chellappan et al, 2004), but not in tomato infected with *Tomato leaf curl New Delhi virus* (ToLCNDV). Both susceptible and tolerant tomato cultivars respond to ToLCNDV by producing 21-23 nt Rep targets, but in contrast to TME3, the tolerant cultivar H-88-78-1 was reported to produce 90-fold higher vsRNAs at 21 dpi compared to 7 dpi (26-fold) (Sahu et al, 2010). While siRNAs increased over the 3-week period in tolerant cultivar H-88-78-1, in contrast in TME3, at recovery (67 dpi) siRNA populations declined. Other intrinsic features of the viral genome, and its molecular interaction with host, are likely to influence the efficacy of virus-induced PTGS and play a role in natural resistance. The molecular mechanisms of tolerance and recovery are not well-studied, and further research on more plant-virus interactions is required.

Predictive interaction of vsRNAs with Argonaute complexes directed by first 5' nucleotide

It has been shown in *Arabidopsis*, that preferential sorting of small RNAs into RNA silencing associated Argonaute (AGO) complexes are directed by the first 5' nucleotide (Mi et al, 2008). Specifically, AGO1 has been shown to have preference for U, AGO2 and AGO4 have preference for A or U, while AGO5 prefers C at the first 5'-end of the siRNA (Brough et al, 1992; Mlotshwa et al, 2008). Similarly, virus-derived small RNAs, dictated by their first 5'-end nucleotides, are preferentially sorted and loaded into multiple AGO complexes (Hohn and Vazquez, 2011). To predict cassava AGO interactions with SACMV, 5'-end nucleotides of vsRNAs targeting DNA A and B in T200 and TME3 at 12, 32 and 67 dpi were

investigated. Our results demonstrated that vsRNAs (as a percentage of total vsRNAs) and actual counts (data not shown) with A as the first 5' nucleotide was most prevalent in targeting DNA A and B in both T200 and TME3, and there was a peak at 32 dpi (Fig. 4.7) when virus replication is high and symptoms visible. As a percentage of total vsRNA populations with A as the first 5' nucleotide there was no significant difference (greater than 5%) between T200 and TME3, but actual counts were higher targeting DNA A in T200 (462, 15,200 and 4690 at 12, 32 and 67 dpi, respectively) compared to TME3 (212, 4147 and 2012 at 12, 32 and 67 dpi, respectively) (data not shown). The only notable difference was a peak (6%) in 5' C nt-vsRNAs at 12 dpi targeting DNA A in TME3 compared to T200 (Fig. 4.7). A similar trend in general was noted with the DNA B component. In contrast to SACMV in cassava, a few cases of preferential use of C as the first 5'-terminal nucleotide has been reported, for example in tomato plants infected with the geminivirus *Tomato yellow leaf curl Sardinia virus* (TYLCSV) (Mlotshwa et al, 2008). While vsRNA-5'A populations were the highest, those with C and U nucleotides were also highly abundant in targeting DNA A and B in T00 and TME3 (Fig. 4.7). A tendency to avoid vsRNAs with G residues at the first 5'-end has been reported (Mi et al, 2008; Molnár et al, 2005), and this was also observed in T200 and TME3, where vsRNAs with a G as the first 5'-nucleodide were significantly under-represented. These results suggest that SACMV vsRNAs preferentially interact with AGO2 or AGO4. However, high abundance of U and C implies involvement of multiple AGOs in sorting vsRNAs in cassava T200 and TME3. Abundance and functionality of virus-derived small RNAs depend on many factors including the secondary structure of each gene (Molnár et al 2005). In SACMV infected cassava, *AC1*, *AV1* and *BC1* were preferentially targeted. The extensive secondary structures in these genomic areas may influence accessibility, affinity, or enzymatic activity leading to the biogenesis of small RNA by one or more components of the RNAi machinery (Molnár et al, 2005).

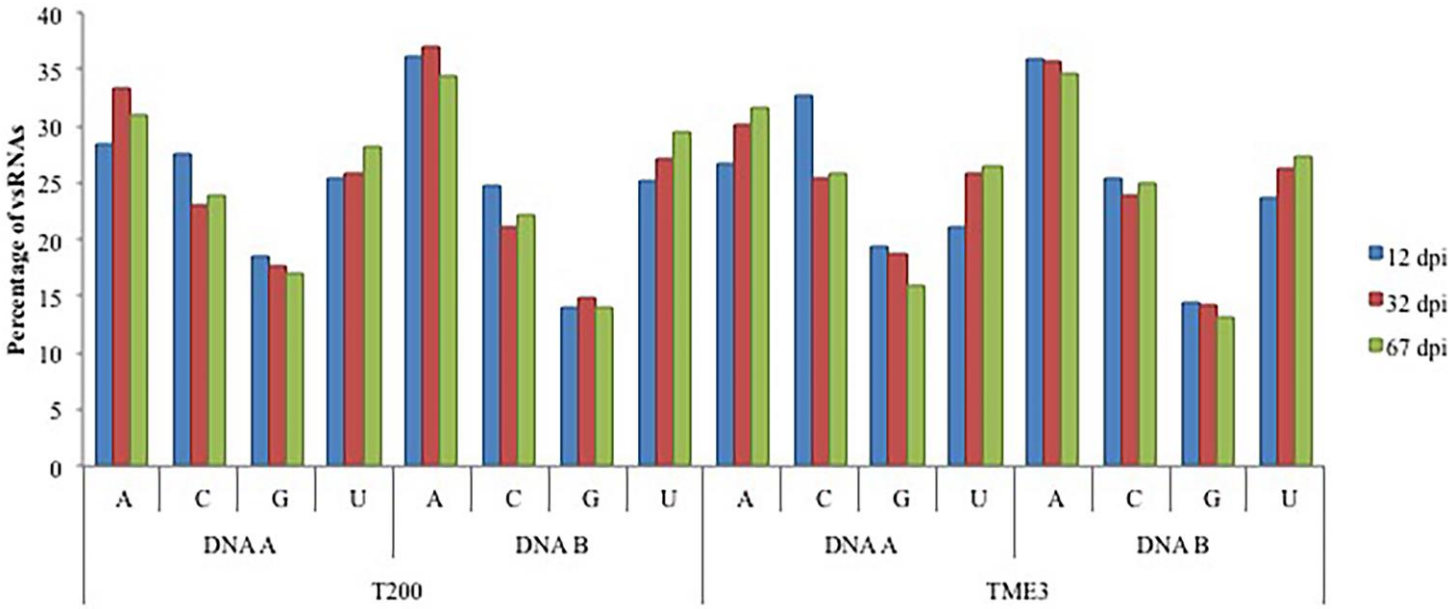


Figure 4.7: 5'-end nucleotides of SACMV-derived vsRNAs targeting DNA A and B in T200 and TME3 at 12, 32 and 67 dpi. The majority of vsRNAs targeting DNA A and DNA B in T200 and TM3 at all three time points start with 'A' at the 5'-end, except at 12 dpi in TME3 where a greater number the vsRNAs start with 'C' at the 5'-end.

4.5 Conclusions

Evidence for PTGS in susceptibility and tolerance in T200 and TME3 phenotypes, respectively, was demonstrated. The patterns of sRNA and vsRNA expression differed between the susceptible and tolerant landraces in response to SCMV infection, and fluctuated over the period of infection (up to 67 dpi). Differences were noted between this study and other CMG-cassava genotype interaction studies. Distribution and frequency of vsRNA has also been shown recently to differ in resistant NASE 3 and susceptible genotypes TME 204 and 60444 infected with ssRNA viruses, *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) (Ogwok et al, 2016).

We propose that high populations of vsRNAs in T200 represent accumulation and failed host defence leading to a susceptible phenotype, and in TME3, low populations represent depletion due to efficient viral mRNA targeting, leading to a tolerant phenotype. This is contradictory to some other studies where they suggest high levels of siRNAs indicate resistance or tolerance/recovery. According to a model suggested previously (Havelda et al, 2005; Szittya et al,

2002), in TME3 where virus titres are lower compared to T200 and decline further in recovered leaves, SACMV VSRs (AC2 or 4) may not bind, and virus-specific vsRNAs act as a systemic signal, moving faster than the virus and thereby establishing antiviral silencing in cells ahead of the infection front. Thus, the RISCs, already activated by vsRNAs, target the entering virus by TGS or PTGS of transcribed mRNA, resulting in recovery of the plants. In contrast, a study by Chellappan et al. (2004) suggested that the AC4 component (VSR) of ACMV competed against AGO1 to bind with single-stranded RNA and therefore suppressed the formation of siRNA-RISC assembly, leading to lower vsRNAs in the susceptible cultivar. In the presence of siRNA-binding/targeting VSRs, plants are not able to confine the spread of the viral infection because vsiRNAs are sequestered and inactivated before they can be incorporated into the RISC. However, in contrast, a high accumulation of vsRNAs in T200 correlates with severe symptoms and a high viral load and suggests a different model where SACMV AC4 may act later post RISC assembly and may prevent AGO-directed vsRNA targeting of mRNA.

Genome methylation was not detected in either SACMV infected T200 or TME3 at any time points in contrast to some other geminivirus studies where high numbers of 24 nt vsRNAs correlating to geminivirus genome intergenic regions. Extremely low numbers of 24 vsRNAs targeting the CR and IR of DNA A and B in both T200 and TME3 were observed. Notwithstanding the possibility that methylation levels fluctuate and may be difficult to measure spatially and temporally in cassava, a role for genome methylation in disease outcome in perennial hosts such as cassava T200 and TME3 landraces remains unclear. In another report, the extent of methylation in geminivirus DNA has been disputed (Paprotka et al, 2010), although an increase in DNA methylation of *Tomato yellow leaf curl Sardinia virus* (TYLCSV) and *Abutilon mosaic virus* geminivirus has been shown in recovered leaves of *Arabidopsis*. Interestingly, in these recovered leaves, sub-populations of highly methylated (associated with histone 3 lysine 9 dimethylation) and hypomethylated (associated with active acetylated histone 3) viral minichromosomes were reported (Raja et al, 2014). Notably, in susceptible T200, significant up-regulation of several histone superfamily transcripts,

including histone H4, and H2A-7, 8 and 10, were present throughout the infection period (12-67 dpi), and down-regulation (log2 fold -1.84) of a methyltransferase (cassava4.1_022835m.g; AT5G10620.1) at 32 dpi correlated with severe systemic symptoms, indicative of active replicative minichromosomes (Allie et al, 2014). Increased histones, known to be associated with geminivirus replication (Hanley-Bowdoin et al, 2013), and high levels of 21-24 nt vsRNAs provide strong evidence that suppression of host RNA and histone silencing mechanisms by SACMV in T200 contributes to susceptibility.

Although recovery has, in most cases documented to date, been associated with RNA silencing, demonstration that recovery from a VSR-deficient BCTV was not prevented in *dcl2* of *dcl2 Arabidopsis* mutants (Raja et al, 2014) certainly suggest other mechanisms besides PTGS that contribute to recovery. Recent studies in cassava have shown a role for transcriptome reprogramming (Allie et al, 2014; Bengyella et al, 2015; Louis and Rey, 2015) and R genes in TME3 (Louis and Rey, 2015) recovery. Recovery may be a hallmark of a complex balance between virus-host defence and virus anti-defence mechanisms, which leads to virus threshold fluctuations during the duration of infection in a plant. This would be more likely in perennial hosts such as cassava. It is more probable in the case of geminivirus infection in cassava that a complex ongoing fluctuating interaction of RNA silencing with plant gene expression shapes symptom phenotype during the course of infection, and in recovery a delicate equilibrium between plant and virus responses may contribute to maintain reduced symptoms and virus titres. This work represents a significant step toward understanding the roles of sRNAs in the recovery response of cassava.

Chapter 5

Summary and Recommendations

5.1 Overall summary and Discussion

Like all major crops, cassava is vulnerable to pests and diseases that can cause heavy yield losses. Cassava mosaic disease (CMD) is the most economically important and single greatest constraint to cassava production (Herrera-Campo et al, 2011). CMD is triggered by the emergence and spread of 11 species of Begomoviruses (Fauquet and Fargette, 1990; Legg and Fauquet, 2004; Patil and Fauquet, 2009), one of which is South African cassava mosaic virus (SACMV). This virus has been identified in regions of South Africa and some neighbouring countries including Zimbabwe, Mozambique and Swaziland. Understanding the mechanism underlying CMD could facilitate control strategies to combat this virus. The work demonstrated in this thesis therefore presents the analysis of microRNA and vsiRNA expression changes associated with SACMV infection in a tolerant (TME3) and susceptible (T200) cassava landraces. This work aimed at addressing a number of questions with regards to SACMV infection and the plant response in this two different cassava landraces. These included:

1. How does SACMV infection affect the small RNA, including miRNAs and vsiRNAs, populations in a susceptible cassava landrace (T200) compared to a SACMV tolerant landrace (TME3)?
2. Do any of these small RNA populations play a role in TME3 recovery or T200 susceptibility to SACMV infection?
3. Does methylation play a role in TME3 recovery to SACMV infection?
4. Does the plant target any "hotspots" on the SACMV genome with virus-derived siRNAs?

Before we could answer the questions about the roles that miRNAs play in SACMV infection in a SACMV tolerant (TME3) and susceptible (T200) cassava landraces, we needed to first identify and characterise the microRNA population in cassava. The main objective of study was to update the available cassava micronome. The identification of a more comprehensive set of miRNAs in cassava is a critical step to facilitate our understanding of regulatory mechanisms or

networks, in particular responses to viral pathogens, of particular interest in our laboratory.

Despite the economic importance of cassava and the potential contribution of miRNAs to cassava improvement, only 153 putative cassava miRNAs are available to date in miRBase (v21). However other well-studied plant species such as *Arabidopsis thaliana, Glycine max, Populus trichocarpa* and *Oryza sativa* have 427, 639, 401 and 713 reported miRNAs in miRBase, respectively (Kozomara and Griffiths-Jones, 2014). The miRNAs that are available for cassava on miRBase were obtained by Patanun et al, (2013) using a computational prediction method by using homology search based on miRNA conservation among different plant species. In addition, Perez-Quintero et al, (2012) analysed small RNA libraries from cassava tissues infected and uninfected with *Xanthomonas axonopodis*, and Zeng et al, (2009) studied conserved miRNAs in the *Euphorbiaceae* family. More recently, Ballen-Taborda et al, (2013) and Xia et al, (2015) both studied cassava miRNAs expressed under abiotic stress conditions. The advantage of this study is that the miRNA discovery was able to capture information from two landraces, that had not been studied previously, and also at three different time points post mock-inoculation, representing different physiological conditions in leaf tissue.

In this study, mature sequences of all known plant miRNAs were used as a query for homologous searches against the publicly available cassava EST and GSS databases (NCBI), and additional identification of novel and conserved miRNAs from next generation sequencing (NGS) of two cassava landraces (T200 from southern Africa and TME3 from West Africa) at three different stages post explant transplantation and acclimatization. miRNAs are classified into families according to sequence similarity and members of the same family usually have the same targets. In this study, 259 conserved miRNAs belonging to 32 families were identified using EST database, 32 conserved miRNAs belonging to 7 families identified using GSS database and 289 conserved miRNAs belonging to 30 families and 39 novel miRNAs belonging to 29 families were identified in T200 and TME3 landraces in deep-sequencing data. Also, 200 (77.2%) of the miRNAs in the EST library, 22 (68.8%) of the miRNAs identified in GSS, 230 (79.6%) of conserved

miRNAs and 38 (98.6%) of the novel miRNAs identified in deep-sequencing data have not been previously reported in cassava. The mR2118 family identified in study has not been previously reported for cassava in other studies. However, we could not experimentally detect this family using RT-PCR and this could be due to low expression levels or specificity of miRNA. We were also able to identify 39 novel cassava specific miRNAs using the NGS data. In order to understand the function of the newly identified conserved and novel miRNAs in cassava, the targets of these miRNAs were also identified. Targets for these newly identified miRNAs were predicted using the psRNATarget web tool. We were able to predict 262 targets for 32 of the conserved miRNAs and 37 targets for 17 of the novel miRNAs.

We also compared the resulting cassava miRNAs identified in this study to the cassava miRNAs that were identified in previous studies, which entailed different germplasm or environmental conditions. This comparison demonstrated that the method/criteria used for miRNA identification, cultivar/landrace of cassava and environmental conditions can affect the miRNAs that are identified and should all be carefully considered when designing a miRNA identification study. Interestingly the miRNA populations identified in mock T200 and TME3 had some differences, which was not unexpected as these have different genetic backgrounds. The differences between TME3 and T200 landraces can be hypothesised to have arisen from geographical separation and adaptation as T200 (history not known) is found in drier regions of southern Africa, while TME3 originates West Africa. Variations could have arisen from hybridizations with local wild *Manihot* species in different locations over the past few hundred years. While this research has unveiled some more important features of the cassava miRNAome, a large number of germplasm-specific cassava miRNAs of low abundance are likely not to have been detected. The knowledge gained from this study contributes to the cassava miRNA database and micronome of this important crop, and unveils differences between landraces, which will be beneficial in the long term in linking gene regulation, gene targets and germplasm traits.

The knowledge gathered over the past few years corroborate the fact that miRNAs exert a broad impact on regulatory mechanisms during plant-microbe interactions. Recent bioinformatic studies uncovered large networks of miRNA families that show altered expression patterns upon infections by bacteria, viruses, fungi and oomycetes (Dunoyer et al, 2006; Fahlgren et al, 2007; He et al, 2008; Kulcheski et al, 2012; Li et al, 2012; Singh et al, 2012). These studies demonstrate that miRNAs are part of regulatory networks orchestrating a multi-layered plant defence. The vast majority of studies demonstrate a stress-specific transcriptional change of miRNA transcriptomes; however, the exact role of given miRNAs during defence responses remains highly elusive. Even though miRNAs have been proved to be pivotal molecules in plant-pathogen interactions, there have been no reports regarding the role of miRNAs in cassava infected by viruses. The aim of this part of the study was to identify and analyse the change in expression of conserved and novel miRNAs in SACMV-infected landraces compared to mock-inoculated controls. We chose the two landraces T200 and TME3, as T200 is a SACMV-susceptible landrace and TME3 is a SACMV-tolerant (TME3). Tolerance describes the extent to which the host is able to withstand infection without undue damage (Robinson, 1969). Not only did we want to compare the changes in expression of microRNAs in a susceptible and tolerant host infected with SACMV, but we also wanted to compare the expression changes of the miRNAs at three different time points: 12, 32 and 67-day post infection (dpi). These time points were chosen as they represent the progress of disease development, where 12 days post inoculation (dpi) represents early infection (pre-symptomatic), and 32 dpi represents symptomatic infection and high virus replication. At a later stage of infection, 67 dpi the recovery phenotype is observed in TME3 (symptom free newly developing leaves) whereas the susceptible and symptomatic phenotype is observed in T200. Symptom remission or "recovery" is a phenomenon reported in several plant studies and has been associated with TGS and PTGS mechanisms (Rodriquez-Negrete et al, 2009). No symptoms were observed in TME3 or T200 plants at 12dpi. Symptoms were first observed in both landraces at approximately 15 dpi. At 32 dpi all newly emerging leaves displayed mosaic and leaf curling for both T200 and TME3. The TME3 leaf tissue displayed the recovery phenotype at 67dpi compared with T200 (typical yellow mosaic on

leaves as well as leaf distortion and leaf curling), where newly emerged leaves have no or reduced symptoms. Log_2Fold changes were determined for the newly identified conserved and novel cassava miRNAs using the equation: Log_2 Fold Change = Log_2 (Normalised count in SACMV library/ Normalised count in the mock library). We considered a miRNA to be upregulated if the Log_2Fold change was greater than 2 and downregulated if the Log_2Fold change was less than -2.

The response of cassava to viral stress is complex and involves many genes and molecular mechanisms, operating at both the transcriptional and post-transcriptional level. It is possible that the changes that miRNAs have important roles in regulating functions of such target genes. In order to elucidate further the roles these miRNAs are playing in SACMV infection it was necessary to identify their targets. Both conserved and novel miRNAs associated gene targets included transcription factors, transposable elements and *R*-genes. Multiple miRNA families in T200 and TME3 were found to target transposable elements and this will prove interesting in further studies. Transcription factors are proteins involved in the process of converting, or transcribing DNA to RNA. Transcriptional gene regulation is crucial for host cells to form an efficient defence response. The arsenal of defence-related transcriptional regulators consists of DNA-binding transcription factors like MYB, NAC and AP2, which have been shown to regulate the expression of defence-related genes, and were all identified as targets of the miRNAs identified in this study. The adaptive immune system of plants is composed of the numerous resistance (R) genes that play a vital role in detecting pathogen effectors (avirulence proteins) by recognising effecter-induced modifications to other host proteins. It has also been shown that sRNA-mediated silencing of R genes regulates host defence against pathogens. A good example of this is the miR482 family. miR482 cleaves mRNA of NBS-LRR at its N terminus. The targeting is accompanied by mRNA degradation and production of secondary siRNAs, which depends on RNA-dependent RNA polymerase 6. These secondary siRNAs further silence other defence-related genes, which forms a miR482-mediated silencing cascade. The miR482 was identified in both T200 and TME3 and interestingly only in the SACMV libraries, which suggests that the presence of SACMV triggers the transcription of the *MIR482* gene. We were also able to

identify multiple miRNA families in TME3 and T200 that targeted R-genes. The molecular mechanisms of tolerance are poorly understood. Notably, while a role for R genes has recently been demonstrated at 67 dpi (recovery) (Louis and Rey, 2015), we show for the first time in this study that an early R gene response in TME3 may also play a role in tolerance.

The most interesting and important result from this study was the identification of miRNAs that targeted important proteins that are necessary for the RNA silencing pathway. The miRNA families miR162, miR168 and miR403 were found to target DCL1, AGO1 and AGO2 respectively. DCLs process long dsRNA into sRNA duplexes and are involved in the processing of the pri-miRNA and pre-miRNA during the miRNA biogenesis pathway. AGO1 has many functions including miRNA processing, and recruitment of virus-derived sRNA (vsRNA) and then degrades the target viral RNA, it is therefore a central node of the RNA silencing pathway in host defence against viruses. AGO2 displays both additive and overlapping activity with AGO1. The induction of this miRNAs in virus-infected plants may counter the inhibitory activity of these proteins and the RNA silencing pathway. In TME3 the miR162 family was downregulated at 12dpi whereas in T200 this family's expression was altered at 12dpi but significantly upregulated at 32dpi. This means that in T200 there will be an accumulation of miR162 in T200, which results in the suppression of DCL1 activity, whereas in TME3 the early downregulation of the miR162 family will result in an early increase in the RNA silencing signal as the DCL1 protein will be expressed. In TME3 at 12dpi the miRNA168 family is downregulated. We believe that this early decrease in miR168 expression will result in an increase in AGO1 expression that should result in an increase in the RNA silencing signal. This could be contributing to the susceptible phenotype that is observed in T200 and the tolerant phenotype that is observed in TME3. In TME3 the miR403 family is downregulated at 32dpi but in T200 this family it upregulated at 32dpi and 67dpi. This will result in a suppression of AGO2 in T200 during the full systemic infection stage and the late infectivity stage, which will ultimately add the repression of the RNA silencing signal in this cassava landrace. However, the SACMV tolerant TME3 cassava landrace will have an increase in AGO2 expression as the miR403 family is

downregulated. We confirmed these results with stem-loop qRT-PCR using RNA samples from both landraces collected at 32dpi. Members of the miR162, miR168 and miR403 were analysed and were found to be upregulated in T200 (susceptible) and downregulated in TME3 (Tolerant). We believe the expression changes of these miRNAs, along with R genes, could be playing large roles in the susceptible and recovery phenotypes we observe in T200 and TME3 respectively.

To counter RNA silencing viruses encode certain proteins that can block the RNAi pathway and are referred to as suppressor of gene silencing (VSR). Silencing suppressors in other viruses haven shown to alter miR168 in order to target AGO1 mRNA. SACMV contains two VSR AC2 and AC4. One or both of these suppressors could be interfering with the expression of these miRNA families in T200 and inhibiting the RNA silencing pathway resulting in suitability. TME3 may have developed a way to inhibit with the SACMV VSRs therefore these miRNA families expression is not interrupted and therefore the RNA silencing pathway is not interrupted and TME3 is able tolerate SACMV infection. Recently a fifth ORF (AC5) has been described in the DNA-A component of many bipartite and monopartite begomoviruses. The AC5 ORF is located downstream, of AC3 in the complementary strand of DNA-A, and overlaps a portion of the CP ORF. Li et al, (2015) demonstrated several important functions of the AC5 protein of *Mugbean yellow mosaic India virus* (MYMIV) for example, AC5 was found to effectively suppress PTGC induced single-stranded but not double stranded RNA. Their results demonstrated that MYMIV AC5 is a pathogenicity determinant and a potent RNA silencing suppressor that employs novel mechanism to supress antiviral defences. They also suggested that the AC5 function may be conserved among Old World begomoviruses, however, while AC5 has been identified by BLAST in SACMV, its putative functions not been studied. Future research should focus on whether SACMV contains AC5 and if it plays a role in the expression changes of important miRNA families during SACMV infection.

In summary, this part of the study identified a broad range of miRNAs (conserved and novel) associated with SACMV infection of T200 and TME3 cassava landraces at three different stages of viral disease stages. By identifying

different miRNAs at different time points, it was hoped that such data would increase our understanding of the host-virus interactions with regard to both pathogenic mechanism of the virus and the resistance response of the host. This discovery and analysis of virus infection-associated miRNA and cassava-specific miRNAs brings new perspectives on the understanding of cassava-virus interaction at molecular levels. Bioinformatic analysis was used to predict the functions of the targets of the identified miRNAs, which indicated that many were involved in or mediated the regulation of physiological mechanism in cassava including pathogenesis-related genes associated with the symptom and characteristic growth and development of cassava infected with SACMV. It is hoped that further investigation of the miRNAs and target genes implicated in this study could lead to the development of disease-resistant cassava plants. The gene targets identified in tolerant TME3 in particular will form the basis for further studies into the molecular networks associated with the tolerance phenotype in plants, and will provide clues for future strategic plans to manipulate virus resistance, not only in cassava but other crops.

The role of siRNA and virus-derived vsRNAs in defence against different viruses has been demonstrated in various plants (Chellappan et al., 2004; Akbergenov et al., 2006; Sahu et al., 2010; Yadav and Chattopadhyay 2011; Sahu et al., 2012b; Sharma et al., 2012). Thus the final aim of this study was to identify vsRNAs in the tolerant (TME3) and susceptible (T200) cassava landraces and see if they play a role the recovery and susceptible phenotypes were observed in these landraces respectively. We also anticipated potential "hotspots" on the SACMV genome that were being targeted by vsiRNAs. Hotspots are specific regions of the viral genome were various vsRNAs are clustered. The NGS data that was generated for the microRNA study previously mentioned was also used to achieve this aim. The sRNA sequences generated by NGS that were 21-24 nt in size, did not map to any sequences in Rfam, and did not match previously identified miRNAs were mapped to SACMV DNA A and DNA B. The mapping generated two very interesting results. Firstly, the deep sequencing results showed that total numbers of vsRNAs and unique vsRNAs were significantly higher in T200 compared with TME3 and total vsRNA counts targeting DNA A were significantly higher compared with DNA

B in both landraces. Since T200 is highly susceptible and virus loads were found to high, we concluded that high accumulation of vsRNAs observed in susceptible T200, especially at 32 dpi when symptoms are severe, is due to failure to target SACMV-derived mRNA resulting in replication and symptom persistence. In contrast, in tolerant TME3, low vsRNA numbers correlated positively with virus titres, symptoms and recovery, and could represent efficient PTGS of viral mRNA, leading to a depletion/sequestration of vsRNA populations, which in turn reduces the levels of viral proteins, and subsequent virus replication. Therefore, in recovered tissues, as the levels of target viral mRNAs decrease, the corresponding vsRNAs are also simultaneously reduced. Secondly, the relative abundance of vsRNAs mapping varied between the ORFs across DNA A and DNA B for both TME3 and T200. When the data was normalised to the raw read count the highest amount of vsRNA clustering occurred in the *AC1, AV1, BV1* and *BC1* for both T200 and TME3 and were considered hotspots. This result was not surprising at AV1 and AC1 are the largest ORFs on DNA A. Therefore, we decided to normalise the number of vsRNAs to the length of the size of the ORF. The number of vsRNAs mapping to the over-lapping regions of the ORFs, especially *AC2/AC3*, significantly increased. The number vsRNAs targeting the intergenic region (IR)/common region (CR) were low (less than 100) for both T200 and TME3 for DNA A and DNA B. The common region includes a stem-loop structure containing the nonanucleotide TAATATTAC and the origin of replication is the last A in the nonanucleotide sequence. The identification of these potential "hotspots" is important for generating pathogen-derived constructs that are designed for genetic engineering experiments aimed at produced a CMD resistant farmer preferred cassava cultivar.

The 24 nt size class was identified as the largest class for the vsRNAs mapping to SACMV DNA A and DNA B for both T200 and TME3. Also the number of vsRNAs that were 24 nt in length increased at 67 dpi (recovery stage) in TME3. From the 24 nt vsRNA results in TME3, it was speculated that the recovery phenotype that we observed in TME3 may be the result of TGS of genomic SACMV DNA. Methylation of virus DNA by plant hosts has already been demonstrated as an epigenetic defence against geminivirus (Raja et al, 2008). Raja et al.

demonstrated that *Arabidopsis* mutants defective in a number of genes that are key players in the RdDM pathway (e.g. *drm1, drm2, kyp2, ago4* and others) results in hypersusceptibility to infection with the geminiviruses *Cabbage leaf curl virus* (CaLCuV) and *Beef curly top virus* (BCTV). Since methylation of geminivirus genomes has previously been associated with recovery for geminiviruses such as *Beet curly top virus* (BCTV) (Akbergenov et al, 2006), *Mungbean yellow mosaic India virus* (Yadav and Chattopadhyay, 2011), and ACMV in cassava (Akbergenov et al, 2006; Chellappan et al, 2004; Ermak et al, 1993), we predicted that methylation may play a role in SACMV-infected tolerant TME3, but not susceptible T200. To analyse whether SACMV DNA A and B were extensively methylated via TGS, we performed bisulfite sequencing on viral DNA extracted from infected T200 (susceptible) and TME3 (tolerant) leaf tissue at 12, 32 and 67 dpi. However, Bisulfite sequencing (Frommer et al, 1992) did not reveal any specific patterns of methylation of any of the ORFs and IR's on SACMV DNA-A or DNA-B that were amplified from T200 and from TME3 leaf tissue. It is possible that since there is no evidence for genome methylation by sequencing, that methylation of histones associated with mini-chromosomes in the nucleus, which has been shown in geminivirus infections (Pilartz and Jeske, 1992; Pilartz and Jeske, 2003), may be linked to tolerance and recovery in TME3. Since methylation of the SACMV genome was not detected by bisulfite sequencing, and vsRNA counts targeting the IR (where the promoters reside) were very low in both the tolerant or susceptible landraces, we conclude that 24 nt vsRNA-mediated RNA directed genome methylation does not play a central role in disease phenotype in these landraces, notwithstanding recognition for a possible role in histone modification in TME3. This work represents an important step toward understanding variable roles of vsRNAs in different cassava genotype-geminivirus interactions.

In summary, this study has contributed significantly to our knowledge of the mechanisms of tolerance and recovery, which are poorly understood, especially in perennial non-model plants. Furthermore, results demonstrate that sRNA responses are associated with host tolerance to SACMV infection in the tolerant TME3 landrace compared with a CMD-susceptible cassava landrace, T200. From this Illumina deep sequencing investigation, there was a clear

difference in the pattern and abundance of normalized total sRNA counts and virus-targeted vsRNA populations in leaf tissues between T200 and TME3 in response to SACMV infection during the time course of infection, and between infected and mock inoculated. While gene silencing is involved in antiviral defence, this study highlights that RNA silencing is likely to play more complex roles with other mechanisms, such as effector and non-effector elicitation of R proteins, in both susceptibility and tolerance. A better understanding of the recovery phenotype may lead to advances in breeding programmes, where wild cassava relatives or currently domesticated varieties phenotypically exhibiting the recovery phenotype, may be interbred to create varieties with genotypes resistant to CMD.

5.2 Future recommendations

Food security is one of the most important issues challenging the world today**.** Any strategies to solve this problem must include increasing crop yields and quality**.** Increasing food production, especially in developing countries, is essential to solve the problem of food security**.** MicroRNA-based genetic modification technology (miRNA-based GM tech) can be one of the most promising solutions that contribute to agricultural productivity directly by developing superior crop cultivars with enhanced biotic and abiotic stress tolerance and increased biomass/yields**.** Indirectly, the technology may also increase the usage of marginal soils and decrease pesticide use, among other benefits**.** MiRNAs and their targets not only provide an invaluable source of novel transgenes, but also inspire the development of several new GM strategies, allowing advances in breeding novel crop cultivars with agronomically useful characteristics. Further advances in small RNA sequencing and their efficient control in plant tissues will certainly provide the necessary tools to better understand the concrete molecular and chemical role of miRNAs during plant-microbe interactions, ultimately leading to miRNA-based improvements of biotic stress responses in important crops. Also, miRNAs likely provide an additional layer of flexibility to cope with diverse biotic stresses. As maintaining a defensive

state is accompanied with high physiological costs, this flexibility provides an intricate cost minimising system to plants, as proposed for miRNA-regulated NB-LRR genes (Shivaprasad et al, 2012). In addition, the fact that pathogens employ suppressors that counteract miRNA pathways (Navarro et al, 2008) implicates promising new ways for pathogens resistance programs in crops

In the near future, exploiting miRNA pathways in regard to pathogen resistance might provide a valuable option to control a broad range of pathogens and pests. Such an advance has been recently proposed in generating virus resistant plants (Qu et al, 2012). It has been demonstrated that expression of modified miRNAs triggering the synthesis of artificial miRNAs (amiRNAs) that target viral RNA sequences can efficiently induce virus resistance. AmiRNAs is engineered by replacing mature miRNA/miRNA* sequence with the designed complementary sequences on host miRNA precursors (Alvarez et al, 2006; Schwab et al, 2006). The advantage of this approach is the fact that artificial miRNAs can be generated in such a way that they are not targeting plant genes (no off target effects)

Based on the precursor backbone of miR171a, amiRNA that targeted 2b gene of CMV was introduced to tobacco, and transgenic tobacco plants acquired defence against CMV (Qu et al, 2007). Transgenic tomato plants of two amiRNAs that targeted VSRs of turnip yellow mosaic virus (TYMV) and turnip mosaic virus (TuMV gained 100% resistance against TYMV TuMV (Nu et al, 2006). By using the precursor of miR159, two amiRNAs were constructed, one targeted the overlapping sequence between CMV 2a and 2b, and the other targeted the 3' untranslated region (UTR). Transgenic tomato plants developed good resistance against CMV, TMV and TYLCN (Zhang et al, 2011). In addition, Vu et al (2013) constructed two amiRNAs, which targeted the AV1 gene of ToLCNDV and the overlapping gene region of AV1 and AV2. Transgenic tomato plants displayed high tolerance to ToLCNDV. Recently, amiRNAs were generated targeting conserved sequences within the genomes of *Cassava brown streak virus* (CBSV) and Ugandan cassava brown steak virus (UCBSV). Transgenic plants challenged with CBSV and UCBSV isolates showed resistance levels that ranged from 20 to 60% against the

viruses. These results indicate a potential application of amiRNAs for engineering resistance to CBSD-causing viruses in cassava. AmiRNAs that target virus genes can process and function like natural miRNAs in plant cells, enhancing disease resistance. All of these findings above provide a new way to plant protection and disease control and further investigation is needed to determine if this technology could be applied to acquiring a SACMV resistant farmer preferred cassava landrace.

Therefore, we were successfully able to address the questions that set out the framework of this PhD. This work contributes significantly to the existing limited knowledge specifically underlying small RNA, including miRNA and vsRNA, expression changes in CMD causing geminiviruses and their host cassava. By comparing the differences between a tolerant and susceptible host there is now a better understanding of the effect of pathogens on host sRNAome, an area that is deserving of more attention in plant systems. The expectation is that these findings presented in the PhD will contribute to the long-term goals of devising new methods of disease control against SACMV and understanding the complex interconnected mechanisms involved in virus-host interactome.

Chapter 6

References

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

Adams M.D, Kelley J.M, Gocayne J.D, Dubnick M, Polymeropoulos M.H, Xiao H, Merril C.R, Wu A, Olde B, Moreno R.F. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252: 1651-1656.

Adams M.J, King A.M.Q, Carstens E.B. (2013) Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses. *Arch. Virol*. 158: 2023–2030.

Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M. (1997) Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. *Plant Cell*, *9*: 841-857.

Akano A, Dixon A, Mba C, Barrera E, Fregene M. (2002) Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor. Appl. Genet*, 105: 521–525.

Akbergenov R, Si-Ammour A, Blevins T, Amin I, Kutter C, Vanderschuren H, Zhang P, Gruissem W, Meins F. Jr., Hohn, T, Pooggin M.M. (2006) Molecular characterization of geminivirus-derived small RNAs in different plant species. *Nucleic Acids Res*, 34: 462–471.

Albert M, Jehle A.K, Lipschis M, Mueller K, Zeng Y, Felix G. (2010) Regulation of cell behaviour by plant receptor kinases: Pattern recognition receptors as prototypical models. *European journal of cell biology*, 89: 200-207.

Allem, A.C. (2002) "The Origins and Taxonomy of Cassava", in R. J. Hillocks, J. M. Thresh and A. C. Bellotti (eds.), Cassava: Biology, Production and Utilization, CAB International, Wallingford, UK. pp. 1–16.

Allen E, Xie Z, Gustafson A.M, Sung G.H, Spatafora J.W, Carrington JC. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana. Nat. Genet*, 36: 1282-1290. doi:10.1038/ng1478.

Allen E, Xie Z, Gustafson A.M, Carrington J.C. (2005) microRNA-Directed Phasing during *Trans*-Acting siRNA Biogenesis in Plants. *Cell*, 121: 207-221. doi:10.1016/j.cell.2005.04.004.

Allie F, Pierce E.J, Okoniewski M.J, Rey C. (2014) Transcriptional analysis of *South African cassava mosaic virus*-infected susceptible and tolerant landraces of cassava highlights differences in resistance, basal defence and cell wall associated genes during infection. *BMC Genom***.** 15, 1006.

Alves M. S, Dadalto S. P, Gonçalves A. B, De Souza G. B, Barros V. A, Fietto L. G. (2013). Plant bZIP transcription factors responsive to pathogens: a review. *International J. Mol. Sciences*, 14: 7815-7828.

Ambros V, Bartel B, Bartel D.P., Burge C.B., Carrington JC, Chen X, Dreyfuss G, Eddy S.R, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T. (2003) A uniform system for microRNA annotation. *RNA,* 9: 277-279. doi:10.1261/rna.2183803.

Amin I, Patil B. L, Briddon R. W, Mansoor S, Fauquet C. M. (2011). A common set of developmental miRNAs are upregulated in Nicotiana benthamiana by diverse begomoviruses. *Virology J.,* 8: 1.

Amiteye S, Corral J.M, Sharbel T.F. (2011) Overview of the potential of microRNAs and their target gene detection for cassava (*Manihot esculenta*) improvement. Afr. J. Biotechnol., 10: 2562-2573. doi:10.5897/AJB10.017.

Arguello-Astorga G, Lopez-Ochoa L, Kong L.J, Orozco B.M, Settlage S.B, Hanley-Bowdoin L. (2004) A novel motif in geminivirus replication proteins interacts with the plant retinoblastoma-related protein. *J. Virol.,* 78: 4817–4826

Aregger M, Borah B.K, Seguin J, Rajeswaran R, Gubaeva E.G, Zvereva AS, Windels D, Vazquez F, Blevins T, Farinelli L, Pooggin M.M. (2012) Primary and secondary siRNAs in geminivirus-induced gene silencing. *PLoS Pathog.* 8:e1002941. Doi:10.137/journal.ppat.1002941.

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

Axtell M.J, Bowman J.L. (2008) Evolution of plant microRNAs and their targets. *Trends Plant Sci*, 13: 343-349. doi:10.1016/j.tplants.2008.03.009.

Axtel M.J. (2013) Classification and Comparison of Small RNAs from Plants. *Annu. Rev. of Plant Biol.,* 64: 137-159. doi: 10.1146/annurev-arplant-050312-120043.

Azevedo J, Garcia D, Pontier D, Ohnesorge S, Yu A, Garcia S, Braun L, Bergdoll M, Hakimi M.A, Lagrange T, Voinnet O. (2010) Argonaute quenching and global changes in Dicer homeostasis caused by a pathogen-encoded GW repeat protein. *Genes & Devel.*, 24: 904-915.

Babaleye T. (1996) Cassava, Africa's food security crop. *Consultative Group on International Agricultural Research (CGIAR) Newsletter* 3(1).

Ballen-Taborda C, Plata G, Ayling S, Rodriguez-Zapata F, Lopez-Lavalle L.A.B, Duitama J, Tohme J. (2013) Identification of Cassava MicroRNAs under Abiotic Stress. *Int. J. Genomics*, 2013: 1-10. doi.org/10.1155/2013/857986.

Baloch I.A, Muhammed D. (2014) Bioinformatic Hunting of MicroRNAs. Pure Appl. Biol., 3: 72-80.

Bartel D.P. (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell Press*, 116: 281-297. doi:10.1016/S0092-8674(04)00045-5.

Baulcombe D. (2004) RNA silencing in plants. *Nature*, 431: 356-363.

Baulcombe D. (2005). RNA silencing. *Trends Biochemical Sciences*, 30: 290-293.

Baumberger N, Baulcombe D.C. (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits micro RNAs and short interfering RNAs. *PNAS*, 102: 11928–11933. doi:10.1073/pnas.0505461102.

Bengyella L, Waikhom S. D, Allie F, Rey C. (2015) Virus tolerance and recovery from viral induced symptoms in plants are associated with transcriptome reprogramming. *Plant Mol. Biol*, 89: 243–252. doi: 10.1007/s11103-105-0362-6.

Bent A.F, Mackey D. (2007) Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annu. Rev. Phytopathol.*, 45: 399-436.

Berezikov E, Cuppen E, Plasterk R. H. (2006). Approaches to microRNA discovery. *Nature genetics*, *38*: S2-S7.

Berrie L.C, Palmer K.E, Rybicki E.P, Rey M.E.C. (1998) Molecular characterisation of a distinct South African cassava infecting geminivirus. *Arch. Virol*., 143: 2253–2260.

Berrie L.C, Rybicki E.P, Rey M.E.C. (2001) Complete nucleotide sequence of *South African cassava mosaic virus*: further evidence for recombination among begomoviruses. *J. Gen. Virol*., 82: 53-58.

Bertsch C, Beuve M, Dolja V.V, Wirth M, Pelsey F, Herbach E, Lemaire O. (2009) Retention of virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance. *Biol. Direct*., 4, 21.

Bian X.Y, Rasheed M.S, Seemanpillai M.J, Rezaian M.A. (2006) Analysis of silencing escape of *Tomato leaf curl virus*: An evaluation of the role of DNA methylation. *Mol. Plant-Microbe Interact*., 19: 614–624.

Bieri S, Potrykus I, Fűtterer, J. (2002) Geminivirus sequences as bidirectional transcription termination/polyadenylation signals for economic construction of stably expressed transgenes*. Mol. Breeding*, 10: 107-117.

Bisaro D.M. (2006) Silencing suppression by geminivirus proteins. *Virology*, 344: 158-168.

Blevins T, Rajeswaran R, Shivaprasad P.V, Beknazariants D, Si-Ammour A, Park H.S, Vazquez F, Robertson D, Meins F, Hohn T, Pooggin M.M. (2006) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res., 34: 6233–6246.

Bhattacharjee S., Zamora A, Azhar M. T, Sacco M. A, Lambert L. H, Moffett P. (2009). Virus resistance induced by NB–LRR proteins involves Argonaute4-dependent translational control. P*lant J.*, 58: 940-951.

Bohmert K, Camus I, Bellini C, Bouchez D, Caboche M, Benning C. (1998) AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J.*, 17: 170-180. doi: 10.1093/emboj/17.1.170.

Boller T, Felix G. (2009) A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual review of plant biology*, 60: 379-406.

Boller T, He S. Y. (2009). Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. *Science*, *324*: 742.

Bologna N. G, Mateos J. L, Bresso E. G, Palatnik J. F. (2009). A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *EMBO L*, 28: 3646-3656.

Bologna N.G, Voinnet O. (2014) The diversity, biogenesis, and activities of endogenous silencing small RNAs in *Arabidopsis*. *Annu. Rev. Plant Biol*., 65: 473–503. doi: 10.1146/annurev-arplant-050213-35728.

Bonardi V, Dangl J.L. (2012) How complex are intracellular immune receptor signaling complexes?. *Frontiers in plant science*, 3: 237.

Bonardi V, Cherkis K, Nishimura M.T, Dang, J.L (2012) A new eye on NLR proteins: focused on clarity or diffused by complexity?. *Curr. Opin. Immunol.,*, 24: 41-50.

Bonnet E, Wuyts J, Rouzé P, Van de Peer Y. (2004) Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences*.* *Bioinformatics*, 20: 2911-2917. doi:10.1093/bioinformatics/bth374.

Bortolamiol D, Pazhouhandeh M, Marrocco K, Genschik P, Ziegler-Graff V. (2007). The Polerovirus F box protein P0 targets ARGONAUTE1 to suppress RNA silencing. *Current Biology*, 17: 1615-1621.

Boutet S, Vazquez F, Liu J, Béclin C, Fagard M, Gratias A, Morel J.B, Crété P, Chen X, Vaucheret H. (2003) Arabidopsis HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Current Biology*, 13: 843-848.

Bouche N, Lauressergues D, Gasciolli V, Vaucheret H. (2006) An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J*., 25: 3347.

Brodersen P, Voinnet O. (2006) The diversity of RNA silencing pathways in plants. *TRENDS in Genetics*, 22: 268-280.

Brosnan C.A, and Voinnet O. (2009) The long and the short of noncoding RNAs. *Curr. Opin. Cell Biol.*, 21: 416-425.

Brough C.L, Gardiner W.E, Inamdar N.M, Zhang X.Y, Ehrlich M, Bisaro D.M. (1992) DNA methylation inhibits propagation of *Tomato golden mosaic virus* DNA in transfected protoplasts. *Plant Mol. Biol.,* 18**:** 703–712.

Brown J.K, Zerbini F.M, Navas-Castillo J, Moriones E, Ramos-Sobrinho R, Silva J.C, Fiallo-Olive E, Briddon R.W, Hernandez-Zepeda C, Idris A, Malathi V.G, Martin D.P, Rivera-Bustamante R, Ueda S, Varsani A. (2015) Revision of Begomovirus taxonomy based on pairwise sequence comparisons. *Arch. Virol*., 160: 1593 – 1619.

Bruening G. (2006) Resistance to infection. In: Loebenstein G, Carr J.P. (eds.). Natural Resistance Mechanisms of Plants to Viruses, pp. 211-240. Springer, Dordrecht, The Netherlands.

Buchmann R.C, Asad S, Wolf J.N, Mohannath G, Bisaro D.M. (2009) Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. *J. Virol*., 83: 5005–5013.

Burdisso P, Suarez I. P, Bologna N. G, Palatnik J. F, Bersch B, Rasia R. M. (2012). Second double-stranded RNA binding domain of dicer-like ribonuclease 1: structural and biochemical characterization. *Biochemistry*, 51: 10159-10166.

Burge S.W, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki E.P, Eddy S.R, Gardner P.P, Bateman A. (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res..* 2012: 1-7. doi: 10.1093/nar/gks1005.

Burgyan J, Havelda Z. (2011) Viral suppressors of RNA silencing. *Trends Plant Sci.* 16: 265–272.

Burns A, Gleadow R, Cliff J, Zacarias A, Cavagnaro T. (2010) Cassava: The Drought, War and Famine Crop in a Changing World*. Sustainability*, 2: 3572-3607.

Burroughs A.M, Ando Y, de Hoon M.J.L, Tomaru Y, Nishibu T, Ukekawa R, Funakoshi T, Kurokawa T, Suzuki H, Hayashizaki Y, Daub C.O. (2010) A comprehensive survey of 3′ animal miRNA modification events and a possible role for 3′ adenylation in modulating miRNA targeting effectiveness. *Genome Res.,* 20: 1398–1410. doi:10.1101/gr.106054.110.

Buscaill P, Rivas S. (2014). Transcriptional control of plant defence responses. *Current opinion in plant biology*, *20*, 35-46.

Carbonell A, Carrington J. C. (2015). Antiviral roles of plant ARGONAUTES. *Curr. Opin. Plant Biol.*, *27*: 111-117.

Carra A, Mica E, Gambino G, Pindo M, Moser C, Pè M. E, Schubert A. (2009). Cloning and characterization of small non-coding RNAs from grape. *Plant J.*, *59*: 750-763.

Carrillo-Tripp J, Lozoya-Gloria E, Rivera-Bustamante R.F. (2007) Symptom remission and specific resistance of pepper plants after infection by *Pepper golden mosaic virus*. *Phytopathology,* 97: 51–59.

Carrington J.C, Whitham S.A. (1998) Viral invasion and host defense: strategies and counter-strategies. *Curr Opin Plant Biol,* 1: 336-341.

Carter S.E, Jones P.G. (1993) A model of the distribution of cassava in Africa. *Appl. Geogr.,* 13: 353-371. doi:10.1016/0143-6228(93)90037-2.

Carthew R.W, Sontheimer E.J. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 126: 642-655. doi:10.1016/j.cell.2009.01.035.

Ceballos H, Iglesias C.A, Perez J. C, Dixon A.G.O. (2004) Cassava breeding; opportunities and challenges. *Plant Mol. Biol*., 56: 503 – 516.

Chapman E. J, Prokhnevsky A. I, Gopinath K, Dolja V. V, Carrington J. C. (2004). Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. *Genes & development*, 18:1179-1186.

Chapman E.J, Carrington J.C. (2007) Specialization and evolution of endogenous small RNA pathways. *Nature Reviews Genetics*, 8: 884-896.

Chavez A.L, Bedoya J.M, Sanchez T, Iglesias C, Ceballos H, Roca W. (2000) Iron, carotene, and ascorbic acid in cassava roots and leaves. *Food Nutr Bull*, 21:410–3.

Chellappan P, Vanitharani R, Fauquet C.M. (2004) Short interfering RNA accumulation correlates with host recovery in DNA virus-infected hosts and gene silencing targets specific viral sequences. *J. Virol*, 78, 7465-7477.

Chellappan P, Vanitharani R, Fauquet C. M. (2005). MicroRNA-binding viral protein interferes with Arabidopsis development. *PNAS,* 102: 10381-10386.

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

Chen X. (2005a) MicroRNA biogenesis and function in plants. *FEBS Letters*, 579: 5923-5931. doi:10.1016/j.febslet.2005.07.071.

Chen C, Ridzon D.A, Broomer A.J, Zhou Z, Lee D.H, Nguyen J.T, Barbisin M, Xu, N.L, Mahuvakar V.R, Andersen M.R, Lao K.Q. (2005b) Real-time quantification of microRNAs by stem–loop RT–PCR. *Nucleic Acids Res.*, 33: e179-e179.

Chen X. (2009) Small RNAs and their roles in plant development. *Annu. Rev. Cell. Dev. Biol.,* 25: 21-44.

Chen H.M, Chen L.T, Patel K, Lia Y.H, Baulcombe D.C, Wu SH. (2010) 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. *PNAS,* 107: 15269–15274. doi:10.1073/pnas.1001738107.

Chi X, Yang Q, Chen X, Wang J, Pan L, Chen M, Yang Z, He Y, Liang X, Yu S. (2011) Identification and characterization of microRNAs from Peanut (*Arachis hypogaea* L.) by High-Throughput Sequencing. *PLOSone,* 6: e27530. doi:10.1371/journal.pone.0027530.

Chitwood D, Nogueira F, Howell M, Montgomery T, Carrington J, Timmermans M. (2009). Pattern formation via small RNA mobility. *Genes Dev.,* 23: 549–554.

Chu H, Jo Y, Cho W.K. (2014) Evolution of endogenous non-retroviral genes integratd into plant genomes. *Curr. Plant Biol*., 1: 55-59.

Chuck G, Candela H, Hake S. (2009). Big impacts by small RNAs in plant development. *Cur. Opin. Plant Biol.*, 12: 81-86.

Cock J. (1985) Cassava New potential for a neglected crop. *Westview*, 191 p.

Coll N.S, Epple P, Dangl J.L. (2011) Programmed cell death in the plant immune system. *Cell Death Differ.* 18: 1247–1256

Collier S.M, Moffett P. (2009) NB-LRRs work a "bait and switch" on pathogens. *Trends in Plant Sci.*, 14: 521-529.

Corrêa L.G.G, Riaño-Pachón D. M, Schrago C. G, dos Santos R. V, Mueller-Roeber B, Vincentz M. (2008). The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. *PLoS One*, 3: e2944.

Creamer K.M, Partridge J.F. (2011) RITS—connecting transcription, RNA interference, and heterochromatin assembly in fission yeast. *Wiley Interdisciplinary Reviews: RNA*, *2*: 632-646.

Cuperus J.T, Carbonell A, Fahlgren N, Garcia-Ruiz H, Burke R.T, Takeda A. (2010) Unique functionality of 22-nt miRNAs in triggering RDR6-dependet siRNA biogenesis from target transcripts in Arabidopsis. *Nature Struct. Mol. Biol.,* 17: 997-1003. doi:10.1038/nsmb.1866.

Cuperus J.T, Fahlgren N, Carrington J.C. (2011) Evolution and Functional Diversification of MIRNA Genes. *Plant Cell,* 23: 431-442. doi:http://dx.doi.org/10.1105/tpc.110.082784.

Culver J.N, Padmanabhan M.S. (2007) Virus-induced disease: altering host physiology one interaction at a time. *Annu. Rev. Phytopathol.*, 45: 221-243.

Dai X, Zhao P.X. (2011) psRNATarget: a plant small RNA Target analysis server. *Nucleic Acids Res.,* 39: W155-W159. doi:10.1093/nar/gkr319.

Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe D.C. (2000) An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell*, 101: 543-553.

Dangl J.L, Jones J.D. (2001) Plant pathogens and integrated defence responses to infection. *nature*, 411*:* 826-833.

Daphne P. (1980) Cassava: a South African venture. *Optima,* 29: 61-68.

De Lorenzo G, Brutus A, Savatin D.V, Sicilia F, Cervone F. (2011) Engineering plant resistance by constructing chimeric receptors that recognize damage-associated molecular patterns (DAMPs). *FEBS letters*, 585: 1521-1528.

Deng D, Otim-Nape G.W, Sangare A, Ogwal S, Beachy RN, Fauquet C.M. (1997) Presence of a new virus closely related to East African cassava mosaic geminivirus associated with cassava mosaic outbreak in Uganda. *Afr. J. Root and Tuber Crops*, 2:23-28.

DeVries J, Toenniessen G. (2001) Securing the harvest: biotechnology, breeding and seed systems for African crops. *CABI publishing*, Oxon, UK.

Deleris A, Gallego-Bartolome J, Bao J, Kasschau K.D, Carrington J.C, Voinnet O. (2006) Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defence. *Science,* 313: 68.

Dhandapani V, Ramchiary N, Paul P, Kim J, Choi SH, Lee J, Hur Y, Lim Y.P. (2011) Identification of potential microRNAs and their targets in *Brassica rapa* L. *Mol. Cells*, 32:21–37. doi:10.1007.s10059-011-2313-7.

Diermann N, Matoušek J, Junge M, Riesner D, Steger G. (2010) Characterization of plant miRNAs and small RNAs derived from potato spindle tuber viroid (PSTVd) in infected tomato. *Biol. Chem.*, *391*:1379-1390.

Ding S.W, Voinnet O. (2007). Antiviral immunity directed by small RNAs. *Cell*, 130: 413–426.

Dixon A.G.O, Whyte J.B.A, Mahungu N.M, Ng S.Y.C. (2001) Tackling the cassava mosaic disease (CMD) challenge in sub-saharan Africa: The role of host-plant resistance and germplasm deployment. In: Taylor, N.J., Ogbe, F., Fauquet, C.M., (eds), Cassava, an ancient crop for modern times: Food, health, culture. Donald Danforth Plant sciences Center, St. Louis, USA.

Dodds P.N, Rathjen J.P. (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics*, 11: 539-548.

Dong Z, Han M, Fedaroff N. (2008) The RNA-binding proteins HYL1 and SE promote accurate *in vitro* processing of pri-miRNA by DCL1. *PNAS,* 105: 9970-9975. doi:10.1073/pnas.0803356105.

Doyle J.J, Doyle J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*, 19: 11-15.

Du Z, Xiao D, Wu J, Jia D, Yuan Z, Liu Y, Hu L, Han Z, Wei T, Lin Q, Wu Z. (2011) p2 of rice stripe virus (RSV) interacts with OsSGS3 and is a silencing suppressor. *Mol. Plant Path.*, 12:.808-814.

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. (2010). MYB transcription factors in Arabidopsis. *Trends Plant Scie.,15*: 573-581.

Dugas D.V, Bartel B. (2004) MicroRNA regulation of gene expression in plants. *Curr. Opin. Plant Biol.,* 7: 512-520. Doi:10.1016/j.pbi.2004.07.011.

Duffy S, Holmes E.C. (2007) Phylogenetic Evidence for Rapid Rates of Molecular Evolution in the Single-Stranded DNA Begomovirus Tomato Yellow Leaf Curl Virus*, J. Virol*., 82: 957 – 965.

Ehrenreich I. M, Puruggana M. D. (2008). Sequence variation of MicroRNAs and their binding sites in Arabidopsis. *Plant Physiol.*, 146: 1974-1982.

Ekwall K. (2004) The RITS complex—a direct link between small RNA and heterochromatin. *Mol. Cell*, 13: 304-305.

Elmore J. M, Lin Z. J. D, Coaker G. (2011). Plant NB-LRR signaling: upstreams and downstreams. *Curr. Opin. Plant Biol.*, 14: 365-371.

El-Shami M, Pontier D, Lahmy S, Braun L, Picart C, Vega D, Hakimi M.A, Jacobsen S.E, Cooke R, Lagrange T. (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes & Devel.*, 21: 2539-2544.

El-Sharkawy M.A. (1993) Drought-tolerant cassava for Africa, Asia, and Latin America: breeding projects work to stabilize productivity without increasing pressures on limited natural resources. Bioscience, 43: 441 – 451.

El-Sharkawy M.A. (2004) Cassava biology and physiology. *Plant Mol. Biol.,* 56: 481-501. doi:10.1016/j.cell.2012.05.017.

Endres M.W, Gregory B.D, Gao Z, Foreman A.W, Mlotshwa S, Ge X, Pruss G.J, Ecker J.R, Bowman L.H, Vance, V. (2010) Two plant viral suppressors of silencing require the ethylene-inducible host transcription factor RAV2 to block RNA silencing. *PLoS Pathog*., 6: p.e1000729.

Ermak G, Paszkowski U, Wohlmuth M, Mittelsten Scheid, O, Paszkowski J. (1993) Cytosine methylation inhibits replication of *African cassava mosaic virus* by two distinct mechanisms. *Nucleic Acids Res*., 21: 3445–3450.

Fahlgren N, Howell M.D, Kasschau K.D, Chapman E.J, Sullivan C.M, Cumble J.S, Givan S.A, Law T.F, Grant S.R, Dangl J.L, Carrington J.C. (2007) High-Throughput Sequencing of Arabidopsis microRNAs: Evidence for Frequent Birth and Death of MIRNA Genes. *PLOSone.,* 2: e219. doi:10.1371/journal.pone.0000219.

FAO (1997) Human nutrition in the developing world. Food and Nutrition series No. 29. Rome.

FAO. (2008) Cassava. In: http://www.fao.org/ag/agp/agpc/gcds/index_en.html.

FAO. (2012) Food outlook. Global market analysis – November 2012. Rome.

FAO. (2013) Save and Grow: Cassava A guide to sustainable production intensification. www.fao.org/ag/save-and-grow/cassava/index_en.html.

FAO. (2014) Cassava, A guide to sustainable production intensification, Rome.

Fang Y, Spector D. L. (2007). Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr. Biol.*, *17*: 818-823.

Fauquet C, Fargette D (1990). African cassava mosaic virus: etiology, epidemiology and control. *Plant Dis.* 74: 404-411.

Feng J, Lai L, Lin R, Jin C, Chen J. (2012). Differential effects of Cucumber mosaic virus satellite RNAs in the perturbation of microRNA-regulated gene expression in tomato. *Molecular biology reports*, *39*(1), 775-784.

Fraile A, García-Arenal F. (2010) The coevolution of plants and viruses: resistance and pathogenicity. *Adv. Virus Res.*, 76: 1-32.

Frank F, Hauver J, Sonenberg N, Nagar B. (2012) Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. *EMBO J.* 31: 3588-3595.

Frazier T.P, Xie F, Freistaedler A, Burklew C.E, Zhang B. (2010) Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum). Planta*., 232: 1289-1308. doi:10.1007/s00425.010.1255.1.

Frazier T.P, Zhang B. (2011) Identification of Plant microRNAs using Expressed Sequence Tag Analysis. *Methods Mol. Biol.,* 678: 13-25.

Fregene M, Matsumura H, Akano A, Dixon A, Terauchi R. (2004) Serial analysis of gene expression (SAGE) of host-plant resistance to the cassava mosaic disease. *Plant Mol. Biol*. 56, 563-571.

Frommer M, McDonald L.E, Millar D.S, Collis C.M, Watt F, Grigg G.W, Molloy P.L, Paul C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *PNAS,* 89: 1827–1831.

Fujita S, Iba H. (2008) Putative promoter regions of miRNA genes involved in evolutionarily conserved regulatory systems among vertebrates. *Bioinformatics*, 24: 303-308. doi: 10.1093/bioinformatics/btm589.

Fusaro A.F, Matthew L, Smith N.A, Curtin S.J, Dedic-Hagan J, Ellacott G.A, Watson J.M, Wang M.B, Brosnan C, Carroll B.J, Waterhouse P.M. (2006) RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. *EMBO Rep*., 7: 1168–1175.

Gafni Y, Epel B.L. (200). The role of host and viral proteins in intra-and inter-cellular trafficking of geminiviruses. *Physiol. Mol. Plant Path.*, 60: 231-241.

Gazzani S, Gendall A.R, Lister C, Dean C. (2003) Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiol.,* 132: 1107–1114.

Ge A, Shangguan L, Zhang X, Dong Q, Han J, Liu H, Wang X, Fang J. (2013) Deep sequencing discovery of novel and conserved microRNAs in strawberry (Fragaria× ananassa). *Physiologia plantarum*,*148*: 387-396.

Gehrig H.H, Winter K., Cushman J, Borland A, Taybi T. (2000) An improved RNA isolation method for succulent plant species rich in polyphenols and polysaccharides. *Plant Mol. Biol. Rep*., 18: 369-376.

Ghani A, Din M, Baloch I.A, Barozai M.Y.K. (2013) Identification of microRNAs in 12 plant species of *Fabaceae*. *Pure Appl. Biol.,* 2: 104-115.

Ghoshal B., Sanfacon H. (2015) Symptom recovery in virus-infected plants: Revisiting the role of RNA silencing mechanisms. *Virol.,* 479: 167-179. doi:10.1016/j.virol.2015.01.008.

Glazebrook J. (2005) Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.*, *43*: 205-227.

Gomez-Gomez L, Boller T. (2000) FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis. Mol. Cell*. 5: 1003–1011.

Graham J, Smith K, MacKenzie K, Jorgenson L, Hackett C, Powell, W. (2004). The construction of a genetic linkage map of red raspberry (Rubus idaeus subsp. idaeus) based on AFLPs, genomic-SSR and EST-SSR markers. *Theoretical and Applied Genetics*, 109: 740-749.

Guilfoyle T. J, Hagen G. (2001). Auxin response factors. *J. Plant Growth Regulation*, 20: 281-291.

Guo H. S, Ding S. W. (2002). A viral protein inhibits the long range signaling activity of the gene silencing signal. *EMBO J*., 21: 398-407.

Guo H. S, Xie Q, Fei J. F, Chua N. H. (2005). MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for Arabidopsis lateral root development. *Plant Cell*, 17: 1376-1386.

Gupta S, Ganguli S, Datta A. (2014) The Silent Assassins: Informatics of Plant Viral Silencing Suppressors. In: Kishor, P.B.K., Bandopadhyay, R., Suravajhala, P., (eds). *Agricultural Bioinformatics*. pp. 21 – 32.

Gururani M.A, Venkatesh J, Upadhyaya C.P, Nookaraju A, Pandey S.K, Park S.W. (2012) Plant disease resistance genes: current status and future directions. *Physiol. Mol. Plant Pathol.*, 78: 51-65.

Gusmaroli G, Tonelli C, Mantovani R. (2002) Regulation of novel members of the Arabidopsis thaliana CCAAT-binding nuclear factor Y subunits. *Gene,* 283: 41–48. doi:10.1016/S0378-1119(01)00833-2.

Gutierrez C. (1999) Geminivirus DNA replication. *Cellular and Molecular Life Sciences CMLS*, 56: 313-329.

Gutierrez C. (2002) Strategies for geminivirus DNA replication and cell cycle interference. *Physiol. Mol. Plant Pathol.*, 60: 219-230.

Hagen C, Rojas M.R, Kon T, Gilbertson R. (2008) Recovery from *Cucurbit leaf crumple virus* (family *Geminiviridae*, genus *Begomovirus*) infection is an adaptive antiviral response associated with changes in viral small RNAs. *Phytopathology*. 98: 1029–1037.

Hamilton A.J, Baulcombe D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, 286: 950-952.

Hammond S.M, Boettcher S, Caudy A.A, Kobayashi R, Hannon G.J. (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science, 293*: 1146-1150.

Han M.H, Goud S, Song L, Fedaroff N. (2004) The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. PNAS, 101: 1093-1098. doi:10.1073/pnas.0307969100.

Hanley-Bowdoin L, Settlage S.B, Orozco B.M, Nagar S, Robertson D. (1999) Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. *Critical Reviews in Plant Sciences*, 18: 71-106.

Hanley-Bowdoin L, Bejarano E.R, Robertson D, Mansoor S. (2013) Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat. Rev. Microbiol.* 11: 777–788.

Harrison B.D, Robinson D.J. (2002) Green shoots of geminivirology. *Physiol. Mol. Plant Path.*, 60: 215-218.

Harvey J. J, Lewsey M. G, Patel K, Westwood J, Heimstädt S, Carr J. P, Baulcombe, D. C. (2011). An antiviral defense role of AGO2 in plants.*PloS one*, *6*(1), e14639.

Havelda Z, Hornyik C, Valoczi A, Burgyan J. (2005) Defective interfering RNA hinders the activity of a tombusvirus-encoded posttranscriptional gene silencing suppressor. *J. Virol.* 79: 450–457.

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

Henderson I. R, Zhang X, Lu C, Johnson L, Meyers B. C, Green P. J, Jacobsen S. E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nature genetics*, 38: 721-725.

Herrera-Campo B.V, Hyman G, Belloti A. (2011) Threats to cassava production: known and potential geographic distribution of four key biotic constraints. *Food Sec.* 3: 329–345.

Hershey C.H. (1987) Cassava germplasm resources. In C.H. Hershey, ed. Cassava Breeding: A Multidisciplinary Review. Proc. of a Workshop, held in the Philippines. March 4-7, 1985. pp. 1-24.

Hehnle S, Wege C, Jeske H. (2004) Interaction of DNA with the movement proteins of geminiviruses revisited. *J. Virol*. 78, 7698–7706.

Hillocks R.J, Thresh J.M. (2001) Cassava Mosaic and Cassava Brown Streak Virus Diseases in Africa: A comparative guide to symptoms and aetiologies. *Roots* 7: 36-50.

Hillocks R.J. (2002). Cassava in Africa. In: Hillocks RJ, Thresh JM, Bellotti AC, editors. Cassava: Biology, Production and Utilization. CABI Publishing, Wallingford, UK. pp. 41–54.

Hofacker I.L, Fontana W, Stadler P.F, Bonhoeffer L.S, Tacker M, Schuster P. (1994) Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie / Chemical Monthly.* 125: 167-188.

Hohn T, Vazquez F. (2011) RNA silencing pathways of plants: silencing and its suppression by plant DNA viruses*. BBA* 1809: 88–600. 10.1016/j.bbagrm.2011.06.00.

Howell M.D, Fahlgren N, Chapman E.J, Cumbie J.S, Sullivan C.M, Givan S.A, Kasschau K.D, Carrington J.C. (2007) Genome-Wide Analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 Pathway in *Arabidopsis* Reveals Dependency on miRNA- and tasiRNA-Directed Targeting. *Plant Cell*, 19:926-942. doi: http://dx.doi.org/10.1105/tpc.107.050062.

Hu Q, Hollunder J, Niehl A, Kørner C.J, Gereige D, Windels D, Arnold A, Kuiper M, Vazquez F, Pooggin M, Heinlein M. (2011) Specific impact of tobamovirus infection on the Arabidopsis small RNA profile. *PLoS One*, *6*: p.e19549.

Huffaker A, Pearce G, Ryan C.A. (2006) An endogenous peptide signal in *Arabidopsis* activates components of the innate immune response. *PNAS,* 103: 10098–10103.

Huntzinger E, Izaurralde E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews Genetics*, 12: 99-110.

Hutvágner G, McLachlan J, Pasquinelli A.E, Bálint É, Tuschl T, Zamore P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, *293:* 834-838.

Hwang D.G, Park J.H, Lim J.Y, Kim D, Choi Y, Kim S, Choi I.Y, Choi D, Shin C. (2013). The Hot Pepper (*Capsicum annuum*) MicroRNA Transcriptome Reveals Novel and Conserved Targets: A Foundation for Understanding MicroRNA Functional Roles in Hot Pepper. *PLoS ONE.* 2013; 8: e64238. doi:10.1371/journal.pone.0064238.

Iki T, Yoshikawa M, Nishikiori M, Jaudal M.C, Matsumoto-Yokoyama E, Mitsuhara I, Meshi T, Ishikawa M. (2010) In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol. Cell*, 39: 282-291.

Iki T, Yoshikawa M, Meshi T, Ishikawa M. (2012). Cyclophilin 40 facilitates HSP90-mediated RISC assembly in plants. *EMBO J.,* 31: 267-278.

Ito, H. (2013) Small RNAs and regulation of transposons in plants. *Genes Genet. Syst.* 88: 3-7.

Jansson C, Westerbergh A, Zhang J, Xinwen H.S.C. (2009) Cassava, a potential biofuel crop in (the) People's Republic of China. Appl. Energy 86, S95–S99 10.1016/j.apenergy.2009.05.011

Jaramillo G, Morante N, Perez J.C, Calle F, Ceballos H, Aria B, Bellotti A.A. (2005) Diallele analysis in cassava adapted to the mid-altitude valleys environment. *Crop Sci.* 45: 1058-1063.

Jarvis A, Ramirez-Villegas J, Campo B.V.H, Navarro-Racines C. (2012) Is Cassava the Answer to African Climate change Adaption? Tropical Plant Biol., 5: 9-29.

Jauvion V, Rivard M, Bouteiller N, Elmayan T, Vaucheret H. (2012) RDR2 partially antagonizes the production of RDR6-dependent siRNA in sense transgene-mediated PTGS. *PLoS One*, *7*(1), p.e29785.

Jeske H. (2009) Geminiviruses. In TT Viruses (pp. 185-226). Springer Berlin Heidelberg.

Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan GL, Walbot V, Sundaresan V, Vance V, Bowman L.H. (2009) Clusters and superclusters of

phased small RNAs in the developing inflorescence of rice. Genome Res., 19: 1429–1440. doi:10.1101/gr.089854.108.

Jin H, Martin, C. (1999). Multifunctionality and diversity within the plant MYB-gene family. *Plant Mol. Biol.*, *41*: 577-585.

Jones J.D, Dangl J.L. (2006) The plant immune system. *Nature*, 444: 323-329.

Jones-Rhoades M.W, Bartel D.P. (2004) Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. *Mol. Cell.,* 14: 787-799. doi:10.1016/j.molcel.2004.05.027.

Jones-Rhoades M.W, Bartel D.P, Bartel B. (2006) MicroRNAs and their regulatory Roles in Plants. Annu. Rev. Plant Biol., 57: 19-53. doi:10.1146/annurev.arplant.57.032905.105218.

Kasschau K. D, Xie Z, Allen E, Llave C, Chapman E. J, Krizan K. A, Carrington J. C. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA function. *Devel. Cell*, 4: 205-217.

Kasschau K.D, Fahlgren N, Chapman E.J, Sullivan C.M, Cumbie J.S, Givan S.A, Carrington J.C. (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol*, *5*, p.e57.

Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas A, Zhu J.K, Staskawicz B.J, Jin H. (2006) A pathogen-inducible endogenous siRNA in plant immunity. *PNAS*, 103: 18002-18007.

Kawano K. (1978) Genetic improvement of cassava (Manihot esculenta Ctrantz) fro productivity. *Tropical Agricultural Research Series,* 11: 9-21.

Khraiwesh B, Zhu J.K, Zhu J.H. (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim. Biophys. Acta.* 1819: 137-148.

Khvorova A, Reynolds A, Jayasens S.D. (2003) Functional siRNAs and miRNAs Exhibit Strand Bias. *Cell*, 115: 209-216. doi:10.1016/S0092-8674(03)00801.8.

Kim V.N. (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.,* 6: 376–385.

Koeck M, Hardham A.R, Dodds P.N. (2011) The role of effectors of biotrophic and hemibiotrophic fungi in infection. *Cell. Microbiol*., 13: 1849-1857.

Kozomara A, Griffiths-Jones S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, 42: 68-73.

Kurihara Y, Watanabe Y. (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *PNAS,* 101: 12753-12758.

Kwak P.B, Wang Q.Q, Chen X.S, Qiu C.X, Yang Z.M. (2009) Enrichment of a set of microRNAs during the cotton fiber development. *BMC Genomics*. 10: 1-11.

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

Lakatos L, Szittya G, Silhavy D, Burgyan J. (2004) Molecular mechanism of RNA silencing suppression mediated by the p19 protein of tombusviruses. *EMBO J*. 23: 876–884.

Lakhotia N, Joshi G, Bhardwaj A.R, Katiyar-Agarwal S, Agarwal M, Jagannath A, Goel S, Kumar A. (2014) Identification and characterization of miRNAome in root, stem, leaf and tuber developmental stages of potato (Solanum tuberosum L.) by high-throughput sequencing. *BMC Plant Biology*. 14: 1-16 doi:10.1186/1471-2229-14-6.

Lamb C, Dixon R.A. (1997) The oxidative burst in plant disease resistance. *Annual review of plant biology*, 48: 251-275.

Latif A, Muller J. (2015) Potential of cassava leaves in human nutrition: A review. *Trends Food Sci. Tech.,* 44: 147 – 158.

Latham J.R, Saunders K, Pinner M.S, Stanley J. (1997) Induction of plant cell division by *Beet curly top virus* gene C4. *The Plant J*. 11: 1273-1283.

Law J.A, Ausin I, Johnson L.M, Vashisht A.A, Zhu J.K, Wohlschlegel J.A, Jacobsen S.E. (2010) A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in *Arabidopsis*. *Curr. Biol.* 20: 951–956.

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

Lee Y, Kim M, Han J, Yeom K, Lee S, Baek S.H, Kim V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal*, 23: 3909-4125.

Lee D. Y, An, G. (2012). Two AP2 family genes, supernumerary bract (SNB) and Osindeterminate spikelet 1 (OsIDS1), synergistically control inflorescence architecture and floral meristem establishment in rice. *Plant J.*, 69: 445-461.

Legg J.P, Thresh J.M. (2000) Cassava mosaic virus disease in East Africa: a dynamic disease in a changing environment. *Virus Res.,* 71: 135-149.

Legg J.P, Fauquet C.M. (2004) Cassava mosaic geminiviruses in Africa. *Plant Mol. Biol.,* 56: 585 – 599.

Legg J.P, Owor B, Sseruwagi P, Ndunguru J. (2006) Cassava Mosaic Virus Disease in East and Central Africa: Epidemiology and Management of A Regional Pandemic. *Adv. Virus Res.* 67: 355 -418.

Legg J.P, Kumar P.L, Makeshkumar T, Tripathi L, Ferguson M, Kanju E, Ntaruhunga, P, Cuellar W. (2015) Cassava virus diseases: biology, epidemiology, and management. *Adv. Virus. Res*. 91: 85–142.

Lewsey M.G, Surette M, Robertson F.C, Ziebell H, Choi S.H, Ryu K.H, Carr J.P. (2009) The role of the *Cucumber mosaic virus* 2b protein in viral movement and symptom induction. *MPMI*, 22: 642–654.

Li J, Yang Z, Yu B, Liu J, Chen X. (2005) Methylation protects miRNAs and siRNAs from a 3'-End Uridylation Activity in Arabidopsis. *Curr. Biol.,* 16: 1501-1507.

Li C.F, Pontes O, El-Shami M, Henderson I.R, Bernatavichute Y, Chan S.W.L, Lagrange T, Pikaard C.S, Jacobsen S.E. (2006) An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. *Cell*, *126*:93-106.

Li F, Pignatta D, Bendix C, Brunkard J.O, Cohn M.M, Tung J, Sun H, Kumar P, Baker B. (2012) MicroRNA regulation of plant innate immune receptors. *PNAS,* 109: 1790-1795.

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

Liu H, Fu Y, Li B, Yu X, Xie J, Cheng S.A, Ghabrial S.A, Li G, Yi X, Jiang D. (2011) Widespread horizontal gene transfer from circular ssDNA viruses to eukaryotic genomes. *BMC Evol. Biol.* 11: 276.

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

Llave C. (2004). MicroRNAs: more than a role in plant development?. *Mol. Plant Path.*, 5: 361-366.

Llave C. (2010) Virus-derived small interfering RNAs at the core of plant-virus interactions. *Trends Plant Sci.* 15: 701–707.

Lozano R, Hamblin M. T, Prochnik S, Jannink J.L (2015). Identification and distribution of the NBS-LRR gene family in the Cassava genome. *BMC Genomics* 16:360. 10.1186/s12864-015-1554-9

Louis B, Rey C. (2015) Resistance gene analogs involved in tolerant cassava geminivirus interaction that shows a recovery phenotype. *Virus Genes*. 51:393-407. Doi:10.1007/s1262-015-1246-1.

Lu S, Sun Y, Shi R, Clark C, Li L, Chiang VL. (2005) Novel and Mechanical Stress-Responsive MicroRNAs in Populus trichocarpa that are absent From *Arabidopsis*. *ASPB,* 17: 2186-2203.

Mallory A. C, Bartel D. P, Bartel B. (2005). MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell*, 17: 1360-1375.

Mallory A.C, Elmayan T, Vaucheret H. (2008). MicroRNA maturation and action –The expanding roles of ARGONAUTES. *Curr. Opin. Plant Biol.,* 11: 560-566.

Mallory A, Vaucheret H, (2010) Form, function, and regulation of ARGONAUTE proteins. *Plant Cell*, 22: 3879-3889.

Mahajan V. S, Drake A, Chen J. (2009). Virus-specific host miRNAs: antiviral defenses or promoters of persistent infection?. *Trends Immunol.*, 30: 1-7.

Maiti S, Basak J, Kundagrami S, Kundu A, Pal A. (2011) Molecular marker-assisted genotyping of mungbean yellow mosaic India virus resistant germplasms of mungbean and urdbean. *Mol. Biotechnol. 47*: 95-104.

Manavella P.A, Koenig D, Weigel D. (2012) Plant secondary siRNA production determined by microRNA-duplex structure. *PNAS*, 109: 2461–2466. doi: 10.1073/pnas.1200169109.

Mandadi K.K, Scholthof K.B.G. (2012) Characterization of a viral synergism in the monocot Brachypodium distachyon reveals distinctly altered host molecular processes associated with disease. *Plant physiol.*, 160: 1432-1452.

Mansoor S, Briddon R.W, Zafar Y. Stanley J. (2003) Geminivirus disease complexes: an emerging threat. Trends Plant Sci., 8: 128 – 134.

Mansoor S, Zafar Y, Briddon R.W. (2006) Geminivirus disease complexes: the threat is spreading. *Trends Plant Sci.*, 11: 209 – 212.

Maredza A.T, Allie F, Plata G, Rey M.E.C. (2015) Sequences enhancing cassava mosaic disease symptoms occur in the cassava genome and are associated with South African cassava mosaic virus infection. *Mol. Genet. Genomics*. 1-19. DOI 10.1007/s00438-015-1049-z.

Margis R, Fusaro A. F, Smith N. A, Curtin S. J, Watson J. M, Finnegan E. J, Waterhouse P. M. (2006). The evolution and diversification of Dicers in plants. *FEBS letters*, 580: 2442-2450.

Martin G, Keller W. (2007) RNA-specific ribonucleotidyl transferases. *RNA*,13: 1834–1849.

Martin D.P, Biagini P, Lefeuvre P, Golden M, Roumagnac P, Varsani A. (2011) Recombination in eukaryotic single stranded DNA viruses. *Viruses,* 3: 1699–1738.

Martinez G, Forment J, Llave C, Pallos V, Gomez G. (2011) High-Throughput Sequencing, Characterisation and Detection of New and Conserved cucumber miRNAs. *PLOSone*. 6: e19523. doi:10.1371/journal.pone.0019523.

Mathews C. (2000) Cassava production by small holer farmers inMpumalanga: A case study. Lowveld Reaserch unit, Department of Agriculture Conservation and Environment, P/ Bag x 11318, Nelspruit-1200.

Matzke M, Kanno T, Daxinger L, Huettel B, Matzke A.J.M. (2009). RNA-mediated chromatin-based silencing in plants. Targets of RNA-directed DNA methylation. *Curr. Opin. Cell Biol*. 21: 367–376.

Maunoury N, Vaucheret H. (2011). AGO1 and AGO2 act redundantly in miR408-mediated Plantacyanin regulation. *PLoS One*, *6*(12), e28729.

Meyers B.C, Souret F.F, Lu C, Green P.J. (2006) Sweating the small stuff: microRNA discovery in plants. *Curr. Opin. Biotech.*, 17: 139-146.

Meyers B.C, Axtell M.J, Bartel B, Bartel D.P, Baulcombe D, Bowman J.L, Cao X, Carrington J.C, Chen X, Green P.J, Griffiths-Jones S, Jacobsen S.E, Mallory A.C, Martienssen R.A, Poethig R.S, Qi Y, Vaucheret H, Voinnet O, Watanabe Y, Weigel D, Zhu J.K. (2008) Criteria for Annotation of Plant MicroRNAs. *ASPB*, 2008; 20: 3186-3190.

Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, Chen S, Hannon G.J, Qi Y. (2008) Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5' terminal nucleotide. *Cell*, 133: 116–127. doi:10.1016/j.cell.2008.02.034.

Michaels S.D, Ditta G, Gustafson-Brown C, Pelaz S, Yanofsky M, Amasino R.M. (2003) AGL24 acts as a promoter

of flowering in *Arabidopsis* and is positively regulated by vernalization. *Plant J.* 33: 867–874.

Miozzi L, Pantaleo V, Burgyán J, Accotto G.P, Noris E. (2013) Analysis of small RNAs derived from Tomato yellow leaf curl Sardinia virus reveals a cross reaction between the major viral hotspot and the plant host genome. *Virus Res*., 178: 287-296.

Mirzaei K, Bahramnejad B, Shamsifard M. H, Zamani W. (2014). In silico identification, phylogenetic and bioinformatic analysis of argonaute genes in plants. *International journal of genomics*, *2014*.

Miya A, Albert P, Shinya T, Desaki Y, Ichimura K, Shirasu K, Narusaka Y, Kawakami N, Kaku H, Shibuya N. (2007) CERK1, a LysM receptor kinase, is essential for chitin elicitor signalling in *Arabidopsis. PNAS*, 104: 19613–19618.

Mlotshwa S, Pruss G.J, Peragine A, Endres M.W, Li J, Chen X, Poethig R.S, Bowman L.H, Vance V. (2008) DICER-LIKE2 plays a primary role in transitive silencing of transgenes in *Arabidopsis*. *PLoS One.* 3: e1755.

Mochizuki K, Fine N.A, Fujisawa T, Gorovsky M.A. (2002) Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell*, 110: 689–699

Moeder W, Yoshioka K. (2008) Lesion mimic mutants: a classical, yet still fundamental approach to study programmed cell death. *Plant Signal. Behav.*, 3: 764-767.

Moffett P. (2009) Mechanisms of recognition in dominant R gene mediated resistance. *Adv. Virus Res.*, 75: 1-229.

Molnár A, Csorba T, Lakatos L, Várallyay É, Lacomme C, Burgyán J. (2005) Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs. *J. Virol.* 79: 7812-7818.

Monaghan J, Zipfel C. (2012) Plant pattern recognition receptor complexes at the plasma membrane. *Curr. Opin. Plant Biol.*, 15: 349-357.

Montes R.A.C, Rosas-Cárdenas D.F, Paoli E.D, Accerbi M, Rymarquis L.A, Mahalingam G, Marsch-Martinez N, Meyers B.C, Green P.J, de Folter S. (2014) Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs*. Nat. Commun.,* 5: doi:10.1038/ncomms4722.

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

Morel J.B, Godon C, Mourrain P, Béclin C, Boutet S, Feuerbach F, Proux F, Vaucheret H. (2002) Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. *The Plant Cell*, 14: 629-639.

Moreno I, Gruissem W, Vanderschuren H. (2001) Reference genes for reliable potyvirus quantitation in cassava and analysis of *Cassava brown streak virus* load in host varieties. *J. Virol. Methods*. 177: 49-54.

Morin R.D, Aksay G, Dolgosheina E, Ebhardt H.A, Magrini E.R, Sahinalp S.C, Unrau P.J. (2008) Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa.* Genome Res*.,* 18: 571-584.

Moxon S, Jing R, Szittya G, Schwach F, Pilcher R.R, Moulton V, Dalmay T. (2008) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res.,* 18: 1602-1609.

Mur L.A, Kenton P, Lloyd A.J, Ougham H, Prats E. (2008) The hypersensitive response; the centenary is upon us but how much do we know?. *J. exp. Bot.*, 59: 501-520.

Murashige T, Skoog F. (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiol. Plant.,* 1962; 15: 473-497. doi: 10.1111/j.1399-3054.1962.tb08052.

Nagpal P, Ellis C.M, Weber H., Ploense S.E, Barkawi L.S, Guilfoyle T.J, Hagen G, Alonso J.M, Cohen J.D, Farmer E.E, Ecker J.R. (2005) Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development*, 132*:* 4107-4118.

Naqvi A. R, Haq Q. M, Mukherjee S. K. (2010). MicroRNA profiling of *tomato leaf curl new delhi virus* (tolcndv) infected tomato leaves indicates that deregulation of mir159/319 and mir172 might be linked with leaf curl disease. *Virology J.*, *7*: 1.

Nassar, N.M.A. (1978). Conservation of the genetic resources of cassava (Manihot esculenta Crantz). Determination of wild species localities with emphasis on probable origin. *Econ. Bot.,* 32: 311-320

Nassar N.M.A. (2002) Cassava, Manihot esculenta Crantz, genetic resources: origin of the crop, its evolution and relationships with wild relatives. *Genet. Mol. Res.*, 1: 298 – 305.

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

Neilsen C.T, Goodall G.J, Bracken C.P. (2011) IsomiRs – the overlooked repertoire in the dynamic microRNAome. *TIG*, 28: 544-549. doi:10.1016/j.tig.2012.07.005.

Newman M.A, Mani V, Hammond S.M. (2011) Deep sequencing of microRNA precursors reveals extensive 3′ end modification. *RNA,* 17: 1795–1803. doi:10.1261/rna.2713611.

Nie X, Molen T.A. (2015) Host recovery and reduced virus level in the upper leaves after Potato virus y infection occur in tobacco and tomato but not in potato plants. *Viruses*, 7: 680-698.

Nishimura M.T, Dangl J.L. (2010) Arabidopsis and the plant immune system. *The Plant Journal*, 61: 1053-1066.

Nuwamanya E, Chiwona-Karltun L, Kawauki R.S, Baguma Y. (2012) Bio-Ethanol Production from Non-Food Parts of Cassava (Manihot esculenta Crantz). *Ambio*. 41: 262-270.

Nweke F.I, Spencer D.D.C, Lynam J.K. (2001) The cassava Transformation: Africa's best kept secret. *Michigan State University Press*, East Lansing.

Oerke E.C, Dehne H.W. (2004) Safeguarding production – losses in major crops and the role of crop production. *Crop Prot*., 23: 275 – 285.

Ogwok E, Ilyas M, Alicai T, Rey M.E.C, Taylor N.J. (2016) Comparative analysis of virus-derived small RNAs within cassava (*Manihot esculenta* Crantz) infected with cassava brown streak viruses. *Virus Res.* 215: 1-11.

Olen K.M, Schaal B.A. (1999) Evidence on the origin of cassava: Phylogeography of Manihot esculenta. *PNAS*, 96: 5586 – 5591.

Orozco B.M, Hanley-Bowdoin L. (1996) A DNA structure is required for geminivirus replication origin function. *J. Virology*, 70: 148-158.

Otim-Nape G.W, Thresh J.M, Shaw M.W. (1997) The effects of cassava mosaic virus disease on yield and compensation in mixed stands of healthy and infected cassava. Annual Appl. Biol., 130: 503 – 521.

Otim-Nape G.W, Thresh J.M. (1998). The current pandemic of cassava mosaic virus disease in Uganda. In: The Epidemiology of Plant Diseases. Ed: D Gareth Jones, Kluwer Academic, London (pp 423-443).

Pacheco R, García-Marcos A, Barajas D, Martiáñez J, Tenllado F. (2012). PVX–potyvirus synergistic infections differentially alter microRNA accumulation in Nicotiana benthamiana. *Virus Research*, *165*: 231-235.

Palatnik J. F, Allen E, Wu X, Schommer C, Schwab R, Carrington J. C, Weigel D. (2003). Control of leaf morphogenesis by microRNAs. *Nature*, *425*: 257-263.

Pallas V, García J.A. (2011) How do plant viruses induce disease? Interactions and interference with host components. *J. Gen. Virol.*, *92*(12), pp.2691-2705.

Panda D, Dehury B, Sahu J, Barooah M, Sen P, Modi MK. (2014) Computational identification and characterisation of conserved miRNAs and their target genes in garlic (*Allium sativum* L.) expressed sequence tags. *Gene*, 537: 333-342. doi:10.1016/j.gene.2014.01.010.

Pandey, S. P, Baldwin I. T. (2007). RNA-directed RNA polymerase 1 (RdR1) mediates the resistance of Nicotiana attenuata to herbivore attack in nature. *The Plant Journal*, *50*(1), 40-53.

Paprotka T, Venkatachari N.J, Chaipan C, Burdick R, Delviks-Frankenberry K.A, Hu W.S, Pathak, V.K. (2010) Inhibition of xenotropic murine leukemia virus-related virus by APOBEC3 proteins and antiviral drugs**.** *J. Virol.,* 84: 5719–5729.

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

Park W, Li J, Song R, Messing J, Chen X (2002) CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel protein, Act in microRNA Metabolism in *Arabidopsis thaliana. Curr. Biol*., 12: 1484-1495.

Park M.Y, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig R.S. (2005) Nuclear processing and export of microRNAs in Arabidopsis. PNAS, 102: 3691-3696.

Papp I, Mette M.F, Aufsatz W, Daxinger L, Schauer S.E, Ray A, van der Winden J, Matzke M, Matzke A.J.M. (2003) Evidence for Nuclear Processing of Plant MicroRNA and Short Interfering RNA Precursors. *Plant Physiol*., 132: 1382-1390.

Pandey A, Soccol C.R, Nigam P, Soccol V. T, Vandenberghe L.P.S, Mohan R. (2000) Biotechnological potential of agro-industrial residues II: cassava bagasse. *Bioresource Technology*, 74: 81-87.

Pantaleo V, Szittya G, Moxon S, Miozzi1 L, Moulton V, Dalmay T. (2010) Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *Plant J.,* 62: 960-976. doi: 10.1111/j.1365-313X.2010.04208.x.

Patanun O, Lertpanyasampatha M, Sojikul P, Viboonjun U, Narangajavana J. (2013) Computational Identification of MicroRNAs and Their Targets in Cassava (*Manihot esculenta* Crantz.). *Mol. Biotechnol*., 53: 257-269.

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

Patil B.L, Fauquet C.M. (2015) Studies on differential behaviour of cassava mosaic geminivirus DNA components, symptom recovery patterns, and their siRNA profiles. *Virus Genes*, 50: 474-86. doi: 10.1007/s11262-015-1184-y.

Peláez P, Sanchez, F. (2013). Small RNAs in plant defense responses during viral and bacterial interactions: similarities and differences. *Front Plant Sci*, 4: 343.

Pérez-Quintero Á. L, Neme R, Zapata A, López C. (2010). Plant microRNAs and their role in defense against viruses: a bioinformatics approach. *BMC plant Biol.*, 10: 138.

Pérez-Quintero A.L, Quintero A, **Urrego O**, **Vanegas P, López C.** (2012) Bioinformatic identification of cassava miRNAs differentially expressed in response to infection by *Xanthomonas* axonopodis pv. Manihotis. BMC Plant Biol., 12: 1-11. doi:10.1186/1471-2229-12-29.

Pilartz M, Jeske H. (1992) Abutilon mosaic geminivirus double-stranded DNA is packed into minichromosomes. *Virol.*, 189**:** 800-802.

Pilartz M, Jeske H. (2003) Mapping of Abutilon mosaic geminivirus minichromosomes. *J. Virol.* 77: 10808–10818.

Pita J.S, Fondong V.N, Sangare A, Otim-Nape G.W, Ogwal S, Fauquet C.M. (2001) Recombination, pseudorecombination, and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J Gen. Virol.* 82**:** 655-665.

Pontes O, Li C. F, Nunes P. C, Haag J, Ream T, Vitins A, Jacobsen S.E, Pikaard, C. S. (2006). The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell*, 126: 79-92.

Pooggin M. M. (2013). How can plant DNA viruses evade siRNA-directed DNA methylation and silencing?. *International journal of molecular sciences*,*14*: 15233-15259.

Prüfer K, Stenzel U, Dannemann M, Green R.E, Lachmann M, Kels J. (2008) PatMaN: rapid alignment of short sequences to large databases. *Bioinformatics*, 24: 1530-1531. doi: 10.1093/bioinformatics/btn223.

Pumplin N, Voinnet O. (2013) RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nature Reviews Microbiol.*, 11: 745-760.

Qi Y, He X, Wang X. J, Kohany O, Jurka J, Hannon G. J. (2006). Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature*, 443: 1008-1012.

Qiu C.X, Xie F.L, Zhu Y.Y, Guo K, Huang S.Q, Nie L, Yang Z.M. (2009) Computational identification of microRNAs and their targets in *Gossypium hirsutum* expressed sequence tags. *Gene,* 395: 49-61. doi:10.1016/j.gene.2007.01.034.

Quintero A, Pérez-Quintero A.L, López C. (2013) Identification of ta-siRNAs and Cis-nat-siRNAs in Cassava and Their Roles in Response to Cassava Bacterial Blight**.** *GPB,* 11: 172-181.

Råberg L, Sim D, Read A.F. (2007) Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science,* 318: *812*-814.

Rabergain S, Vailleau F, Balagué C, Roby D. (2003) Lesion mimic mutants: keys for deciphering cell death and defense pathways in plants?.*Trends Plant Sci.*, 8: 263-271.

Raymond C. K, Roberts B. S, Garrett-Engele P, Lim L.P, Johnson J. M. (2005). Simple, quantitative primer-extension PCR assay for direct monitoring of microRNAs and short-interfering RNAs. *RNA*, 11: 1737-1744.

Raja P, Sanville B.C, Buchmann R.C, Bisaro D.M. (2008) Viral genome methylation as an epigenetic defence against geminiviruses. *J. Virol*. 82: 8997–9007.

Raja P, Wolf J.M, Bisaro D.M. (2010) RNA silencing directed against geminiviruses: post-transcriptional and epigenetic components. *Biochim. Biophys Acta.* 1799: 337–351.

Raja P, Jackel J.N, Li S, Heard I.M, Bisaro D.M. (2014) *Arabidopsis* double stranded RNA binding protein DRB3 participates in methylation-mediated defence against geminiviruses. *J. Virol*. 88: 2611–2622.

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP. (2006) A diverse and evolutionary fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev.,* 20: 3407-3425.

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

Rey M.E, Ndunguru J, Berrie L.C, Paximadis M, Berry S, Cossa N, Nuaila V.N, Mabasa K.G, Abraham N, Rybicki E.P, Martin D. (2012) Diversity of dicotyledenous-infecting geminiviruses and their associated DNA molecules in Southern Africa, including the south-west Indian Ocean islands. *Viruses*, *4*: 1753-1791.

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

Robaglia C, Caranta C. (2006) Translation initiation factors: a weak link in plant RNA virus infection. *Trends Plant Sci.*, 11: 40-45.

Rodríguez-Negrete E.A, Carrillo-Tripp J, Rivera-Bustamante R.F. (2009) RNA silencing against geminivirus: Complementary action of posttranscriptional gene silencing and transcriptional gene silencing in host recovery. *J. Virol*. 83: 1332–1340.

Rodríguez-Negrete E, Lozano-Durán R, Piedra-Aguilera A, Cruzado L, Bejarano E.R, Castillo A.G. (2013) Geminivirus Rep protein interferes with the plant DNA methylation machinery and suppresses transcriptional gene silencing. *New Phytol.* 199: 464–475.

Rogans S. J, Rey, C. (2015). Unveiling the Micronome of Cassava (Manihot esculenta Crantz). *PloS one*, *11*: e0147251-e0147251.

Romanel E, Silva T. F, Corrêa R. L, Farinelli L, Hawkins J. S, Schrago C. E, Vaslin M. F. (2012). Global alteration of microRNAs and transposon-derived small RNAs in cotton (Gossypium hirsutum) during Cotton leafroll dwarf polerovirus (CLRDV) infection. *Plant molecular biology*, *80*(4-5), 443-460.

Romanoff S, Lynam J. (1992) Commentary: Cassava and African food security: Some ethnographic examples. Ecol. Food Nutr., 27: 29-41.

Rottiers V, Najafi-Shoushtari S.H, Kristo F, Gurumurthy S, Zhong L, Li Y, Cohen D.E, Gerszten R.E, Bardeesy N, Mostoslavsky R, Näär A.M. (2011) MicroRNAs in metabolism and metabolic diseases. In *Cold Spring Harbor symposia on quantitative biology* (Vol. 76, pp. 225-233). Cold Spring Harbor Laboratory Press.

Ruiz-Ferrer V, Voinnet O. (2009) Roles of plant small RNAs in biotic stress responses. *Annu. Rev. Plant Biol.*, 60: 485–510.

Sahu P.P, Rai N.K, Chakraborty S, Singh M, Chandrappa P.H, Ramesh B, Chattopadhyay D, Prasad M. (2010) Tomato cultivar tolerant to Tomato leaf curl New Delhi virus infection induces virus-specific short interfering RNA accumulation and defence-associated host gene expression. *Mol. Plant Pathol*. 11: 531–544.

Sahu P.P, Rai N.K, Puranik S, Roy A, Khan M, Prasad M. (2012) Dynamics of defence related components in two contrasting genotypes of tomato upon infection with *Tomato Leaf Curl New Delhi Virus*. *Mol. Biotech.,* 52: 140–150.

Sahu P.P, Sharma N, Puranik S, Prasad M. (2014) Post-transcriptional and Epigenetic Arms of RNA Silencing: A Defence Macinery of Naturally Tolerant Tomato Plant against *Tomato Leaf Curl New Delhi Virus*. *Plant Mol. Biol. Rep*., 32: 1015-1029.

Santovito E, Mascia T, Siddiqui S.A, Minutillo S.A, Valkonen J.P, Gallitelli D. (2014) Infection cycle of artichoke Italian latent virus in tobacco plants: meristem invasion and recovery from disease symptoms. *PLoS One,* 9, e99446.

Schirle N.T, Macrae I.J. (2012) The crystal structure of human Argonaute2. *Science*, 336: 1037–1040.

Schott G, Mari-Ordonez A, Himber C, Alioua A, Voinnet O, Dunoyer P. (2012). Differential effects of viral silencing suppressors on siRNA and miRNA loading support the existence of two distinct cellular pools of ARGONAUTE1. *EMBO J*, *31*: 2553-2565.

Schwab R, Palatnik J.F, Riester M, Schommer C, Schmid M, Weigel D. (2005) Specific Effects of MicroRNAs on the Plant Transcriptome. *Dev. Cell*, 8: 517-527. doi:10.1016/j.devcel.2005.01.018.

Schwach F, Vaistij F. E, Jones L, Baulcombe D. C. (2005). An RNA-dependent RNA polymerase prevents meristem invasion by potato virus X and is required for

the activity but not the production of a systemic silencing signal. *Plant physiology*, 138: 1842-1852.

Schwessinger B, Ronald P.C. (2012) Plant innate immunity: perception of conserved microbial signatures. *Plant Biology*, *63*.

Seguin J, Otten P, Baerlocher L, Farinelli L, Pooggin M.M. (2013) MISIS: a bioinformatics tool to view and analyze maps of small RNAs derived from viruses and genomic loci generating multiple small RNAs. *J. Virol. Methods*, 195: 120-2. doi: 1016/j.viromet.2013.10.013.

Shivaprasad P.V, Chen H.M, Patel K, Bond D.M, Santos B.A, Baulcombe D.C. (2012) A microRNA superfamily regulates nucleotide binding site leucine-rich repeats and other mRNAs. *Plant Cell*. 24:859–874. doi:http://dx.doi.org/10.1105/tpc.111.095380.

Simón-Mateo C, García J. A. (2006). MicroRNA-guided processing impairs Plum pox virus replication, but the virus readily evolves to escape this silencing mechanism. *J. Virology*, 80: 2429-2436.

Singh K. B, Foley R. C, Oñate-Sánchez L. (2002). Transcription factors in plant defense and stress responses. *Curr. Opin. Plant Biol.*, *5*: 430-436.

Singh K, Talla A, Qiu W. (2012). Small RNA profiling of virus-infected grapevines: evidences for virus infection-associated and variety-specific miRNAs. *Functional & integrative genomics*, 12: 659-669.

Spoel S.H, Dong X. (2012) How do plants achieve immunity? Defence without specialized immune cells. *Nat. Rev. Immunol.*, 12: 89-100.

Song C, Wang C, Zhang C, Korir N. K, Yu H, Ma Z, Fang J. (2010). Deep sequencing discovery of novel and conserved microRNAs in trifoliate orange (Citrus trifoliata). *BMC genomics*, 11: 431.

Soosaar J.L, Burch-Smith T.M, Dinesh-Kumar S.P. (2005) Mechanisms of plant resistance to viruses. *Nature Reviews Microbiol.*, 3: 789-798.

Soto J, Ortiz J.F, Perlaza-Jiménez L, Vásquez A.X, Lopez-Lavalle L.A, Mathew B, Leon J, Bernal A.J, Ballvora A, Lopez C.E. (2015) A genetic map of cassava (Manihot esculenta Crantz) with integrated physical mapping of immunity-related genes. *BMC Genomics*, 16:190. Doi: 10.1186/s12864-015-1397-4

Stocks M.B, Moxon S, Mapleson D, Woolfenden H.C, Mohorianu I, Folkes L, Schwach F, Dalmay T, Moulton V. (2012) The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics.* 28 :2059-2061. doi:10.1093/bioinformatics/bts311.

Storey H.H. (1936) Virus diseases on East African plants. VI. A progress report on studies of diseases of cassava. *East African Agricultural Journal* 2: 34–39.

Sun G. (2012) MicroRNAs and their diverse functions in plants. Plant Mol. Biol., 80: 17-36.

Sunkar R, Girke T, Jain P.K, Zhu J.K. (2005) Cloning and characterization of microRNAs from rice. Plant Cell, 17:1397-1411. doi:http://dx.doi.org/10.1105/tpc.105.031682.

Szittya G, Molnar A, Silhavy D, Hornyik C, Burgyan J. (2002) Short defective interfering RNAs of tombusviruses are not targeted but trigger post-transcriptional gene silencing against their helper virus. *Plant Cell*, 14: 359–372.

Talmor-Neiman M, Stav R, Frank W, Voss B, Arazi T. (2006) Novel micro-RNAs and intermediates of micro-RNA biogenesis from moss. *Plant J.* 47, 25–37.

Takada S, Hibara K. I, Ishida T, Tasaka M. (2001). The CUP-SHAPED COTYLEDON1 gene of Arabidopsis regulates shoot apical meristem formation. *Development*, 128: 1127-1135.

Takeda A, Iwasaki S, Watanabe T, Utsumi M, Watanabe Y. (2008) The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. Plant Cell Physiol., 49: 493-500. doi:10.1093/pcp/pcn043.

Takeuchi O, Akira S. (2009) Innate immunity to virus infection. *Immunological reviews*, 227: 75-86.

Taylor D.D, Taylor C.G. (2013) The origin, function, and diagnostic potential of RNA within extracellular vesiv=cles present in

Thakur V, Wanchana S, Xu M, Bruskiewich R, Quick W. P, Mosig A, Zhu X. G. (2011). Characterization of statistical features for plant microRNA prediction. *BMC Genomics*, 12: 108.

Thomson D. W, Bracken C. P, Goodall G. J. (2011). Experimental strategies for microRNA target identification. *Nucleic Acids Res.*, 39: 6845-6853.

Tomari Y, Matranga C, Haley B, Martinez,N, Zamore P.D. (2004) A protein sensor for siRNA asymmetry. *Science*, 306: 1377-1380.

Tonukari N. J. (2004) Cassava and the future of starch. *Electro. J. Biotechnol.,* 7.

Truniger V, Aranda M.A. (2009) Recessive resistance to plant viruses. .*Adv. Virus Res.*, 75: 119-231.

Tsuda K, Katagiri F. (2010). Comparing signalling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Curr. Opin. Plant Biol.*, 13: 459-465.

Tucker M. R, Okada T, Hu Y, Scholefield A, Taylor J. M, Koltunow A. M. (2012). Somatic small RNA pathways promote the mitotic events of megagametogenesis during female reproductive development in Arabidopsis.*Development*, 139: 1399-1404.

Ulker B, Somssich I.E. (2004) WRKY transcription factors: from DNA binding towards biological function. *Curr. Opin. Plant Biol.* 7:491-498. doi:10.1016/j.pbi.2004.07.012.

Unver T, Namath-Covert DM, Budak H. (2009) Review of Current Methodological Approaches for Characterising MicroRNAs in Plants. *Int. J. Plant Genomics*, 2009: 1-11.

Vaistij F. E, Jones L, Baulcombe D. C. (2002) Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell*, 14: 857-867.

Vanderschuren H, Moreno I, Anjanappa R. B, Zainuddin I. M, Gruissem W. (2012) Exploiting the combination of natural and genetically engineered resistance to cassava mosaic and cassava brown streak viruses impacting cassava production in Africa. *PLoS ONE,* 7:e45277 10.1371/journal.pone.0045277.

Vanitharani R, Chellappan P, Pita J.S, Fauquet C.M. (2004) Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and posttranscriptional gene silencing suppression. *J. Virol.* 78: 9487–9498.

Vanitharani R, Chellappan P, Fauquet C.M. (2005) Geminiviruses and RNA silencing. *Trends Plant Sci*., 10**:** 144–151.

Várallyay É, Válóczi A, Ágyi Á, Burgyán J, Havelda Z. (2010). Plant virus-mediated induction of miR168 is associated with repression of ARGONAUTE1 accumulation. *EMBO J.*, 29: 3507-3519.

Vaucheret H. (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes & Devel.*, 20: 759-771.

Vazquez F, Legrand S, Windels D. (2010) The biosynthetic pathways and biological scopes of plant small RNAs. *Trends Plant Sci.,* 15: 337–345.

Visser M, Maree H.J, Rees D.J.G, Burger J.T. (2014) High throughput sequencing reveals small RNAs involved in ASGV infection. *BMC Genom*., 15: 568.

Voinnet O. (2005). Induction and suppression of RNA silencing: insights from viral infections. *Nature Reviews Genetics*, 6: 206-220.

Voinnet O. (2009) Origin, biogenesis, and activity of plant microRNAs. *Cell,* 136:669-687.

Vlot A.C, Klessig D.F, Park S.W. (2008) Systemic acquired resistance: the elusive signal (s). *Curr. Opin. Plant Biol.*, 11: 436-442.

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

Wang J. W, Czech B, Weigel D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. *Cell*, *138*: 738-749.

Wang X.B, Wu Q, Ito T, Cillo F, L, W.X, Chen X, Yu J.L, Ding S.W. (2010) RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. *PNAS*, 107: 484-489.

Wang L, Liu H, Li D, Chen H. (2011) Identification and characterization of maize miRNAs involved in very early stage of seed germination. *BMC Genom*. 12:154 doi: 10.1186/1471-2164-12-154.

Wang J, Yang X, Xu H, Chi X, Zhang M, Hou X. (2012) Identification and characterisation of microRNAs and their target genes in *Brassica oleracea*. *Gene*, 505: 300-308. doi:10.1016/j.gene.2012.06.002.

Wang C, Leng X, Zhang Y, Kayesh E, Zhang Y, Sun X, Fang,J. (2014). Transcriptome-wide analysis of dynamic variations in regulation modes of grapevine microRNAs on their target genes during grapevine development. *Plant Mol. Biol.*, 84: 269-285.

Ward, B.M, Medville R, Lazarowitz S.G, Turgeon R. (1997) The geminivirus BL1 movement protein is associated with endoplasmic reticulum-derived tubules in developing phloem cells. *J. Virol*., 71: 3726–3733.

Wassenegger M, Krczal G. (2006). Nomenclature and functions of RNA-directed RNA polymerases. *Trends in plant science*, 11:142-151.

Wei S, Gruber M.Y, Yu B, Gao M.J, Khachatourians G.G, Hegedus D.D, Parkin I.A, Hannoufa A. (2012) Arabidopsis mutant sk156 reveals complex regulation of SPL15 in a miR156-controlled gene network. *BMC Plant Biol.*, *12*: 169.

Weiberg A, Bellinger M, Jin H. (2015). Conversations between kingdoms: small RNAs. *Curr. Opin. Biotech.*, 32: 207-215.

Willmann M.R, Endres M.W, Cook R.T, Gregory B.D. (2011) The Functions of RNA-Dependent RNA Polymerases in *Arabidopsis. Arabidopsis* Book, 9: e0146.

Wu G, Poethig R. S. (2006). Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3. *Development*,133: 3539-3547.

Wurdack K.J, Hoffmann P, Chase M.W. (2005) Molecular phylogenetic analysis of univulate Euphorbiaceae (Euphorbiaceae sensu stricto) using plastid RBCL and TRNL-F DNA sequences. Am. J. Bot. 92: 1397 – 1420.

Xiao Y, Luan Y. S. (2014). MiR482 mediates the biotic stress response of plant. *Plant Physiol J*, *50*: 741-748.

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Ziberman D, Jacobson S.E, Carrington J.C. (2004) Genetic and Functional Diversification of small RNA Pathways in plants. *PLOS Biol.,* 2: 642-652.

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

Xie F, Frazier TP, Zhang B. (2011). Identification, characterization and expression analysis of MicroRNAs and their targets in the potato *(Solanum tuberosum). Gene*, 473: 8-22.

Xie J, Zeng C, Chen Z, Zhang K, Chen X, Zhou Y, Song S, Lu C, Yang Z, Zhou J, Peng H, Wang W, Peng M, Zhang W. (2015) Endogenous small-noncoding RNAs and their roles in chilling response and stress acclimation in Cassava. *BMC Genom*., 15: 1-19. doi:10.1186/1471-2164-15-634.

Xu P, Roossinck M.J. (2000) Cucumber Mosaic Virus D Satellite RNA–Induced Programmed Cell Death in Tomato. *Plant Cell*, 12: 1079-1092.

Yadav R.K, Chattopadhyay D. (2011) Enhanced viral intergenic region–specific short interfering RNA accumulation and DNA methylation correlates with resistance against a geminivirus. *MPMI*, 24: 1189–1197.

Yang L, Liu Z, Lu F, Dong A, Huang H. (2006) SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J.* 47: 841-850.

Yao Y, Guo G, Ni Z, Sunkar R, Du J, Zhu JK, Sun Q. (2007) Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.). *Genome Biol.,* 8 :R96. doi:10.1186/gb-2007-8-6-r96.

Yi R, Qin Y, Macara IG, Cullen BR. (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes and Dev.* 17: 3011-3016.

Yi H, Richards E. J. (2007). A cluster of disease resistance genes in Arabidopsis is coordinately regulated by transcriptional activation and RNA silencing. *Plant Cell*, 19: 2929-2939.

Yin Z, Li C, Han X, Shen F. (2008) Idnetification of conserved microRNAs and their target genes in tomato (*Lycopersicon esculentum*). *Gene*, 414: 60-66.

Yu B, Bi L, Zheng B, Ji L, Chevalier D, Agarwal M, Ramachandran V, Li W, Lagrange T, Walker J.C, Chen X. (2008) The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *PNAS,* 105: 10073-10078.

Yu S, Galvão V.C, Zhang Y.C, Horrer D, Zhang T.Q, Hao Y.H, Feng Y.Q, Wang S, Schmid M, Wang J.W. (2012) Gibberellin regulates the Arabidopsis floral transition through miR156-targeted SQUAMOSA PROMOTER BINDING–LIKE transcription factors. *Plant Cell*, *24*: 3320-3332.

Zeng C, Wang W, Zheng Y, Chen X, Bo W, Song S, Zhang W, Peng M. (2009) Conservation and divergence of microRNAs and their functions in Euphorbiaceous plants. *Nucl. Acids Res*., 2009: 1-15. doi: 10.1093/nar/gkp1035.

Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD, Li Y, Gonzalez A.J, Yan Z, Kitto S.L, Grusak M.A., Jackson S.A, Stacey G, Cook D.R, Green P. J, Sherrier J, Meyers B.C. (2011). MicroRNAs as master regulators of the plant NB-LRR defence gene family via the production of phased, trans-acting siRNAs. *Genes & Dev.,* 25: 2540-2553. doi:10.1101/gad.177527.111.

Zhang B.H, Pan X.P, Wang Q.L, Cobb G.P, Anderson T.A. (2005) Identification and characterization of new plant microRNAs using EST analysis. *Cell Research*, 15: 336–360. doi:10.1038/sj.cr.7290302.

Zhang B, Pan X, Cobb G.P, Anderson T.A. (2006a) Plant microRNA: A small regulatory molecule with big impact. *Dev. Biol.,* 289: 3-16.

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

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

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

Zhang L, Chia J.M, Kumari S, Stein J.C, Liu Z, Narechania A, Maher C.A, Guill K, McMullen M.D, Ware, D (2009) A genome-wide characterization of microRNA genes in maize. *PLoS Genet*, *5*: p.e1000716.

Z**hang Z, Yu J, Li D, Zhang Z, Liu F, Zhou X, Wang T, Ling Y, Su Z.** (2010) PMRD: plant microRNA database. *Nucleic acids research*, 38: D806-D813.

Zhang X, Zou Z, Zhang J, Zhang Y, Han Q, Hu T, Xu X, Liu H, Li H, Ye Z. (2011) Over-expression of sly-miR156a in tomato results in multiple vegetative and reproductive trait alterations and partial phenocopy of the sft mutant. *FEBS letters*, *585*: 435-439.

Zhang X, Xia J, Lii Y.E, Barrera-Figueroa B.E, Zhou X, Gao S, Lu L, Niu D, Chen Z, Leung C, Wong T, Zhang H, Guo J, Li Y, Liu R, Liang W, Zhu J.K, Zhang W, Jin H. (2012). Genome-wide analysis of plant nat-siRNAs reveals insights into their

distribution, biogenesis and function. *Genome Biol.,* 13: R20. Doi:10.1186/gb-2012-13-3-r20.

Zhou X, Liu Y, Calvert L, Munoz C, Otim-Nape G.W, Robinson D.J, Harrison B.D. (1997) Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *J. Gen. Virol.,* 78: 2101–11

Zhou Z.S, Huang S.Q, Yang Z.M. (2008) Bioinformatic identification and expression analysis of new microRNAs from *Medicago trunculata. Biochem. Biophys. Res. Commun*., 374: 538-542.

Zhu H, Hu F, Wang R, Zhou X, Sze S.H, Liou L.W. Barefoot A, Dickman M, Zhang X. (2011) Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. *Cell*, 145: 242-256.

Zipfel C, Robatzek S. (2010) Pathogen-associated molecular pattern-triggered immunity: veni, vidi…?. *Plant Physiology*, 154: 551-554.

Zong J, Yao X, Yin J, Zhang D, Ma H. (2009) Evolution of the RNA-dependent RNA polymerase (RdRP) genes: Duplications and possible losses before and after the divergence of major eukaryotic groups. *Gene,* 447: 29 – 39.

Zuo J, Zhu B, Fu D, Zhu Y, Ma Y, Chi L Ju Z, Wang Y, Zhai B, Luo Y. (2012) Sculpting the maturation, softening and ethylene pathway: The influences of microRNAs on tomato fruits. *BMC Genom*., 13: 1-12. doi:10.1186/1471-2164-13-7.

Zvereva A.S, Pooggin M.M. (2012) Silencing and Innate Immunity in Plant Defence against Viral and Non-Viral Pathogens. *Viruses*, 4: 2578-2597.