

Gene Expression studies towards the elucidation of host responses to *South African cassava mosaic virus*

A dissertation presented by

Farhahna Allie

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
2013



Project Supervisor: Prof. M.E.C Rey

Advisor: Prof Colin Straker

Abstract

Cassava Mosaic Disease (CMD) is one of the main biotic and economically important constraints of cassava (*Manihot esculenta* Crantz) cultivation in southern Africa. *South African cassava mosaic virus* (SACMV-[Za:99]) in particular is endemic to South Africa. Compatible virus infection induces and suppresses host gene expression at the global level. These gene-expression changes are the molecular basis of symptom development and general stress and defence-like responses of the host. Currently, only limited amounts of expression data are available for geminivirus infection in both model and natural host systems. The research presented in this thesis demonstrates that geminivirus-plant interactions can successfully be investigated using high-throughput technologies such as microarrays and RNA-seq Next-generation sequencing.

The first gene expression study undertaken for study made use of a heterologous cross-species hybridization (CSH) approach to elucidate host gene responses to SACMV-[ZA:99] in a susceptible host, *Nicotiana benthamiana*, at four different time points (7, 14, 21 and 35 days post inoculation (dpi). A high degree (>70%) of nucleotide conservation in CSH between several crop species, including *Arabidopsis* and *N. benthamiana*, has been demonstrated, and facilitated a CSH cDNA microarray study in the absence of a cassava or *N.benthamiana* EST chip. Evaluation of symptoms during the infectivity study revealed that symptoms appeared at 14 dpi, and plants were fully symptomatic at 35 dpi (1.79×10^3 molecules SACMV DNA A/ng TNA using quantitative real-time PCR), with the appearance of characteristic disease symptoms such as leaf distortion, chlorosis and stunting. Two biological replicates were conducted for microarray hybridizations. Following hybridization and scanning, data was analysed using the web-based Limma package, Carmaweb. By using stringent selection criteria (≥ 1.5 or ≤ -1.5 -fold change, $p < 0.05$) transcripts were identified to be up- and down-regulated. Seven genes were selected and five of these were independently validated using semi-quantitative Real-time PCR. The expression levels of many transcripts encoding proteins involved in the regulation of transcription and translation, metabolism, cytoskeleton organization, chloroplast functions and photosynthesis were repressed at 14 dpi and were associated with the increasing levels of SACMV-[ZA:99] titre in host cells. Comparisons with gene expression changes induced by geminiviruses (TYLCV and CaLCuV), and other RNA viruses in solanaceous species formed the basis of the discussion for this study.

With the availability of a more recently developed commercial *Nicotiana benthamiana* array (Nb-array) from Nimblegen, we embarked on conducting a homologous hybridization between systemically SACMV-[ZA:99]-infected *N.benthamiana* leaf tissue cDNA and 13 014 unique gene oligo

probes printed on the Nb-array, to compliment the previously conducted heterologous array. Infectivity assays conducted on SACMV-[ZA:99] infected *N.benthamiana* exhibited leaves that were clearly symptomatic at 21 days post inoculation (6.40×10^2 molecules SACMV-[ZA:99] DNA-A/ng TNA), with the appearance of characteristic disease symptoms such as leaf distortion, chlorosis and stunting. As a result, the 21 dpi time point was chosen to conduct the microarray study in order to dissect host gene responses in *N. benthamiana* using the commercially available custom-made 60-mer oligo NimbleGen platform. Three biological replicates were conducted for microarray hybridizations. The expression levels of many transcripts were significantly ($p < 0.05$) altered in response to SACMV-[ZA:99], including encoded proteins involved in transcription networks, defence responses and plant hormone signalling. Differentially expressed transcripts with putative functions in virus movement, and geminivirus-activated metabolic pathways related to starch and sucrose are discussed.

For the final aim of this study, we performed the first next-generation sequencing (NGS) study using the ABI SOLiD platform to identify gene expression changes associated with disease development and recovery in two cassava (*Manihot esculenta*) landraces, susceptible T200 and tolerant TME3, induced by SACMV- [ZA:99] infection. Analysis of the SACMV-[ZA:99] infection in T200 and TME3 at 12, 32 and 67 days post inoculation revealed that 4181 and 1008 transcripts, respectively, were differentially expressed (>2 -fold changes at $p < 0.05$) in these landrace across all three timepoints. Comparative analysis between SACMV-[ZA:99]-infected T200 and TME3 , across 12, 32 and 67dpi, revealed that there was a positive correlation between the number of regulated genes at each timepoint during infection, the increase in viral titre and the severity of symptoms between these two landraces. This was particularly interesting for TME3, where the increase in virus titre between 32 and 67dpi was found to be 10-fold lower than that observed in SACMV-infected T200 between the same two timepoints. Additionally, an attenuation of symptoms, characteristic of the “recovery phenotype”, was observed in SACMV-infected TME3 leaves from about 50dpi but was not observed in T200. From the SOLiD data, alterations observed in the expression of transcription factors, R genes, genes involved in the DNA methylation pathway and those involved in the putative movement of SACMV-[ZA:99] were selected for further discussion. This is the first study reporting NGS of a susceptible and tolerant landrace of cassava infected with geminivirus.

The activation and suppression of suites of genes identified in model hosts *N.benthamiana* and natural hosts cassava during this study, further demonstrates the complex relationship that exist between host plants and geminiviruses. In general, the data presented in this thesis highlights that there are core genes that partake in defence and are commonly expressed in different plant hosts

(*Nicotiana* and cassava) in response to SACMV-[ZA:99], such as WRKY, HSP70, ethylene responsive element binding factor (ERF) , and PR-genes.

In conclusion, the data collectively presented in this thesis identify a large number of candidate genes that could be further dissected and interrogated in downstream functional genomics studies *in planta*, such as using a Virus-induced gene silencing (VIGS) reverse genetics approach, or by exploiting *Arabidopsis* mutants. Also, since “omics” approaches are becoming increasingly popular, protein-protein interactions between SACMV-[ZA:99] and host proteins can be assessed using two-yeast hybrid assays or protein arrays. The combination of our gene expression data and other complementing functional techniques can therefore be used to build a predictive model of geminivirus-host infection which can be used as a long term resource for dissecting virus-plant interactions. This will be particularly beneficial for the identification of candidate host genes that could be used in the future to improve control strategies against SACMV-[ZA:99].

Declaration

I declare that this thesis titled "*Gene Expression studies towards the elucidation of host responses to South African cassava mosaic virus*", is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy in the Faculty of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



Farhahna Allie

Johannesburg, 27 September 2013

Acknowledgements



(In the name of Allah, the Beneficent, the most Merciful)

To my Supervisor, Prof Chrissie Rey, I am eternally grateful to you for giving me the opportunity to do this project and to be a part of your lab. You were the eternal optimist when things did not always go according to plan with this project and you kept me motivated particularly on the darkest of days. Your insightful guidance and advice throughout this PhD study has been invaluable.

To all my friends in the School of Molecular and Cell Biology and my fellow peers (both past and present) from the Plant Biotechnology Laboratory, thank you for creating such a stimulating working environment. I have gained a wealth of knowledge from each of you!

To my husband, Zaid, thank you for your unwavering love, support and patience on this journey. I am truly grateful for your loving shoulders on the turbulent days and the 'pats-on-the-back' on my victorious ones. I am blessed to have you as my lifelong partner!

To my mom, dad and sister...your constant faith and love in me is humbling. Thank you for believing in me even when I did not. Your love and support has not only carried me through my University career, but also through everyday life.

To my in-laws, thank you for all of your support and for reminding me that hard work and sacrifice always pays off.

A special thank you to the National Research Foundation (NRF), the Department of Science and Technology (DST), the National Bioinformatics Network (NBN) and the University of the Witwatersrand, for their financial support.

Thank you to all the members of the Microarray Facility, at the University of Pretoria, for all of their advice and technical assistance.

Last but not least, to my son, Ayaan, thank you for giving me the motivation and final 'kick' towards the end...

Research Outputs

*Research outputs for the duration of PhD term 2008-2012 for Farhahna Allie (*nee* van Schalk)

Conference outputs

1. **Van Schalk, F., and Rey, M.E.C.** (2008). Elucidation of host responses to South African cassava mosaic virus using cDNA microarrays. BIO 08 Conference Grahamstown, 21-25th January 2008.
2. **Rey, M.E.C., Van Schalk, F.** (2008) Gene expression profile of *Nicotiana benthamiana* infected with SACMV using cDNA microarrays. 9th International Congress of Plant Pathology (ICPP 2008), 24-29 August, Turin, Italy.
3. **van Schalk, F and M.E.C. Rey** (2009) Gene Expression profile of *Nicotiana benthamiana* infected with SACMV using microarrays. Proceedings of the 46th Congress of the Southern African Society for Plant Pathology, Gordons Bay, 26-28 January 2009. p74
4. **Rey, M.E.C., E. Pierce and F. van Schalk** (2010) Global transcriptome profiling of *Arabidopsis thaliana* infected with South African cassava mosaic virus. 4th International ssDNA Comparative Virology Workshop, 7-12 November, Guanajuato, Mexico
5. **van Schalk, F and M.E.C. Rey** (2010) Gene expression profile of *Nicotiana benthamiana* infected with *South African cassava mosaic virus* using a cross-species cDNA microarray approach. 4th International ssDNA Comparative Virology Workshop, 7-12 November, Guanajuato, Mexico.
6. **Rey, M.E.C., Pierce, E.J. and van Schalk, F.** (2011) Transcriptome profiling of *South African cassava mosaic virus*-responsive genes in *Arabidopsis* and *Nicotiana benthamiana*. Abstracts of the American Society for Virology 30th Annual Meeting, Minnesota, US, pg.79
7. **C. Chetty; M. Moralo; R. Mundembe; E.J. Pierce; S. Rogans, F. van Schalk, H. Vanderschuren and M.E.C. Rey** (2011) Biotechnological solutions for engineering virus resistant cassava. Agricultural Biotechnology for economic Development: moving towards a Bioeconomy. 6-9 September, Sandton Convention Centre, Johannesburg, South Africa.

Original Publications

1. **Allie, F and Rey, M.E.C** (2013). Transcriptional alterations in model host, *Nicotiana benthamiana*, in response to infection by *South African cassava mosaic virus*. *European Journal of Plant Pathology*, (DOI) 10.1007/s10658-013-0286-4
2. **Allie, F., Pierce, E.J., Okoniewski, M.J. and Rey, M.E.C.** (2013). Monitoring transcriptional changes in a susceptible and tolerant cultivar of cassava infected with *South African cassava mosaic virus*-[South Africa:99] using next-generation sequencing. *BMC Genomics*. [Manuscript in preparation]

Original Publications (not part of this PhD)

3. **Djami-Tchatchou, A.T., Straker, C.J., and Allie, F.** (2012). 454 Sequencing for the identification of genes differentially expressed in avocado fruit (cv. Fuerte) infected by *Colletotrichum gloeosporioides*. *Journal of Phytopathology*, Volume 160, Issue 9, pg 449–460
4. **Djami–Tchatchou A.T., Allie, F. Straker, C.J.** (2013). Expression of defence–related genes in avocado fruit (cv. Fuerte) infected with *Colletotrichum gloeosporioides*, *South Africa Journal of Botany*, 86, pg 92-100

List of Figures

Chapter 1

Figure 1.1	Symptoms development of ACMV and SACMV-infected susceptible cassava.....	4
Figure 1.2	Genome organization of DNA-A and DNA-B of <i>South African cassava mosaic virus</i> (SACMV)-[ZA:99].....	6
Figure 1.3	mi-RNA (A) and si-RNA (B) mediated pathways in plants.....	13

Chapter 2

Figure 2.1	Symptom development and confirmation of SACMV-[ZA:99] infection in <i>N.benthamiana</i> leave.....	38
Figure 2.2	Conventional PCR carried out on twenty SACMV-[ZA:99] infected plants at 14dpi to confirm viral infection.....	38
Figure 2.3	Standard Curve for SACMV-[ZA:99] –DNA-A and viral quantification of SACMV in <i>N. benthamiana</i> leaf tissue	39
Figure 2.4	Microarray scan of “yellow” test hybridization between <i>Arabidopsis</i> cDNA probes and <i>N. benthamiana</i> cDNA from healthy leaf tissue.....	39
Figure 2.5	MA plots for biological replicate 1 at 7, 14, 21 and 35dpi.....	41
Figure 2.6	RT-qPCR Validations of CSH microarray study.....	43

Chapter 3

Figure 3.1	Symptom development and PCR confirmation of SACMV-[ZA:99] –infection in <i>N. benthamiana</i> leaf tissue over a 35 day period.....	61
Figure 3.2	Viral quantification and 155bp DNA-A PCR product obtained from qPCR from SACMV-[ZA:99] -infected <i>N.benthamiana</i> leaf tissue.....	62
Figure 3.3	Heat map and distribution of 611 differentially expressed genes (DEGs) showing the upregulated (red) and downregulated (green) genes identified as differentially expressed at 21dpi in SACMV-[ZA:99] -infected <i>N.benthamiana</i> leaf tissue.....	63
Figure 3.4	Gene Ontology annotations of DEGs. Pie chart showing the functional categorization of genes differentially expressed at 21dpi.....	64
Figure 3.5	Log ₂ ratios of qPCR validations versus log ₂ ratios obtained from the microarray data	65
Figure 3.6	KEGG pathways for Starch and Sucrose metabolism in <i>Arabidospis</i>	66

Chapter 4

Figure 4.1	T200 and TME3 infected leave tissue evaluated for the development of symptoms over a 67 day period.....	88
-------------------	---	----

Figure 4.2	Viral quantification from SACMV-[ZA:99] –infected T200 and TME3 leaf tissue ...	89
Figure 4.3	Venn diagrams showing the differential distribution of upregulated (>2.0-fold) and downregulated (<2.0-fold) transcripts in SACMV-[ZA:99] infected T200 and TME3 leaf tissues at three different time points	94
Figure 4.4	GOSlim functional characterization for DEGs in T200 and TME3	95
Figure 4.5	RT-qPCR validations for SOLiD data	97

List of Tables

Chapter 2

Table 2.1	Primer sequences designed and used for the independent validation of microarray results.....	35
------------------	--	----

Chapter 3

Table 3.1	Primers used for qPCR validation of NimbleGen microarray data.....	59
Table 3.2	Symptom severity index scores used to assess the development of symptoms in SACMV-infected leaf tissue.....	60

Chapter 4

Table 4.1	Primers used for RT-qPCR validations of SOLiD data for T200 and TME3.....	86
Table 4.2	Pairing statistics for the total number of bases and forward and reverse tagged sequences that were mapped to the cassava draft genome (version 4.1) for T200 and TME3 libraries.....	91

List of Abbreviations

ACMV	<i>African cassava mosaic virus</i>
AGO	Argonaute protein
BCTV	<i>Beet Curly Top Virus</i>
bp	base pair
CCP	Core coat protein
CP	coat protein
CSH	Cross-species hybridization
CTAB	Cetyltrimethylammonium bromide
DCL	DICER or DICER-like protein
DEG	Differentially expressed gene
DNA	deoxyribonucleic acid
dpi	days post inoculation
ds	double-stranded
EDTA	ethylene diamine tetra-acetate
FC	fold change
g	gram
GV	geminivirus
LB	Luria Bertani
LRR	Leucine rich repeat
MAMP	microbe-associated molecular Patterns
min	minute
ml	millilitre
NBS	nucleotide-binding site
ng	nanograms
NGS	Next Generation Sequencing
nt	nucleotide
NTP	nucleotide triphosphate
ORF	open reading frame

PAMP	pathogen-associated molecular patterns
PCR	Polymerase chain reaction
PD	plasmodesmata
PRR	Pathogen recognition receptors
PVX	<i>Potato virus X</i>
<i>R</i>	resistance
RMA	Robust Multi-Array Average
RCR	Rolling circle replication
RdDM	RNA-mediated DNA Methylation Pathway
RDRP	RNA-dependent RNA polymerase
RISC	<i>RNA-induced silencing complex</i>
<i>RNA</i>	ribonucleic acid
RSS	RNA silencing suppressors
RT	reverse transcriptase
SDS	sodium dodecyl sulphate
SEL	size exclusion limit
sec	seconds
ss	single-stranded
SSC	Saline sodium citrate
TCV	<i>Turnip crinkle virus</i>
TIR	Toll-interleukin-1 receptor
TMV	<i>Tobacco mosaic virus</i>
TNA	Total nucleic acid
µg	microgram
µM	micromolar

Contents

Gene Expression studies towards the elucidation of host responses to South African cassava mosaic virus

Abstract	
Declaration	i
Acknowledgements	ii
Research Outputs	iii
List of Figures	v
List of Tables	vii
List of Abbreviations	viii
CHAPTER 1 – Literature review	1
1. Introduction	2
1.1 Cassava, the crop	2
1.2 Cassava disease and pests	3
1.2.1. Cassava mosaic disease	3
1.3 Geminiviruses	4
1.3.1 Classification of <i>Geminiviridae</i> and genome organization.....	4
1.3.2 Geminivirus-host interactions.....	6
1.4 Plant-pathogen interactions.....	7
1.4.1 Incompatible interactions (Resistance).....	8
1.4.2 Compatible Interactions (Susceptibility/Pathogenecity)	9
1.5 Plant defence response	10
1.5.1 Basal Innate Immunity in plants	10
1.5.2 RNA silencing pathways and anti-viral immunity.....	10
1.5.3 Viral counter defence – Viral RNA silencing suppressors	13
1.6 Gene expression profiling	15
1.6.1 Microarray technology	15
1.6.2 Next-generation sequencing	17
1.6.2.1 The Applied Biosystem (ABI) SOLiD™ sequencing system	18
1.6.2.1.1 Paired-end sequencing	18
1.7 Model and natural host systems in host-pathogen interactions studies	19
1.7.1 Arabidopsis and <i>N.benthamiana</i> as model host plants	19
1.7.2 Cassava: EST's, Genomics and Transcriptomics	20
2. Rationale and Overall Objectives of study	21
2.1 Specific Aims.....	22

CHAPTER 2 – Heterologous microarray approach to identify host responses involved in SACMV-[ZA:99] Infection	24
2.1 Abstract.....	25
2.2 Introduction	26
2.3 Materials and Methods	28
2.4 Results	36
2.5 Discussion	43
2.6 Conclusions	47
 CHAPTER 3 – Transcriptional alterations in model host, <i>Nicotiana benthamiana</i>, in response to infection by <i>South African cassava mosaic virus</i>.....	 49
3.1 Abstract.....	50
3.2 Introduction	51
3.3 Materials and Methods	53
3.4 Results	59
3.5 Discussion	67
3.6 Conclusions	74
 CHAPTER 4 – Monitoring transcriptional changes in a susceptible and tolerant cultivar of cassava infected with <i>South African cassava mosaic virus</i> using next-generation sequencing technology.....	 76
3.1 Abstract.....	77
3.2 Introduction	78
3.3 Materials and Methods	81
3.5 Results.....	87
3.5 Discussion	98
3.6 Conclusions	105
 CHAPTER 5 – Summary and Future Recommendations.....	 107
5.1 Overall summary and Discussion.....	108
5.2 Future recommendations	113
 CHAPTER 6 – References	
References	115

Chapter 1

Literature Review

1. Introduction

1.1 Cassava, the crop

Cassava, (*Manihot esculenta* Crantz, Family: *Euphorbiaceae*), is a tuberous crop that is widely grown in tropical Africa, Asia and South America. This crop serves as a food source to almost one billion people worldwide (Nassar et al. 2002). Although this dicotyledonous plant is reported to have a low protein content, it produces clusters of tuberous roots that have a high starch content ranging between 24-31% (Hillocks, 1997 and Cock, 1985). The leaves are eaten as a vegetable and have been reported to have excellent nutritional value for both animals and plants (Ceballos et al. 2004). This crop ranks as the second most important source of calories in sub-Saharan Africa (Nweke, 2001). The global production of cassava in 2009 was reported as being more than 19 million hectares of land used with the global production being over 240 million metric tonnes (FAOSTAT, 2008).

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.

Cassava is considered to be a security crop as it has the ability to grow in nutrient poor soils; it is drought tolerant and is resistant to crop pests such as locusts (Kawano et al. 1978; Jaramillo et al. 2005). Furthermore, the tuberous roots of cassava can be stored in the ground and harvested when needed, 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. 2002). The major cassava growing areas in South Africa include; northern Kwa-Zulu Natal, Limpopo and Mpumalanga provinces. Cultivation has also been reported in as well as neighbouring countries Swaziland and Mozambique. These areas are particularly favourable for cultivation as they are well known for their extreme temperatures, erratic rainfall, and a sandy infertile soil which lends itself to the cultivation of this recalcitrant crop.

In addition to being a staple food crop, cassava is economically important as it can be used as raw material for the production of industrial starch and bioethanol. Commercially produced cassava starch in South Africa alone, in 2010, was reported to be around 20 000 tons (Department of

Agriculture, Forestry and Fisheries, Republic of South Africa, 2010). In November 2011, one of the world's largest brewers, SABMiller, launched a cassava beer in the hope of targeting low-income Africans, and the subsequent switch from home brew to commercial beer (<http://mg.co.za/article/2011-11-02-sabmillers-cassava-beer-to-compete-with-moonshine/>).

1.2 Cassava Diseases and pests

Cassava like other crops is significantly affected by a number of diseases. These include but are not limited to, Cassava Brown Streak Disease (CBSD), Cassava Bacterial Blight (CBB), Cassava Anthracnose Disease (CAD) and Cassava Mosaic Disease (CMD). In addition to bacterial and viral attack, cassava is also challenged by insect pests, including mealybugs (*Phenacoccus manihoti* Mat. Ferr.) and green spider mites (*Mononychellus tanajoa* Bondar), where green mites in particular have been reported to cause losses of up to 80% (Dahniya, 1994).

1.2.1 Cassava Mosaic Disease

Cassava mosaic disease (CMD) has been identified as the major biotic constraint to cassava production in 17 out of 38 main cassava growing sub-Saharan African countries surveyed (Legg and Thresh, 2003). CMD affects the foliar parts of the plant and reductions in yield have been shown to range between 20 - 90%. Cassava losses in 2002 were of the order of 19 million tonnes from a total continent wide production of 101 million tonnes (FAOSTAT, 2004). Cassava is a crop that is vegetatively propagated through stem cuttings. This mode of propagation results in the spread of CMD. The incidence of CMD has been reported in eastern parts of South Africa as well as Swaziland. CMD is recognized by the development and appearance of a characteristic mosaic pattern on the leaves (Fig 1.1). Chlorotic regions are scattered on the leaf surface and appear as pale yellow areas surrounded by normal, darker green islands. Severe infection results in an overall stunting of the plant as well as distortion of leaves (Agrios, 2005). The most well studied and characterized casual agent of CMD is *African cassava mosaic virus* (ACMV). There are currently nine species that are recognized (Fauquet et al. 2003; Legg 2008), of which eight have been reported from Africa.

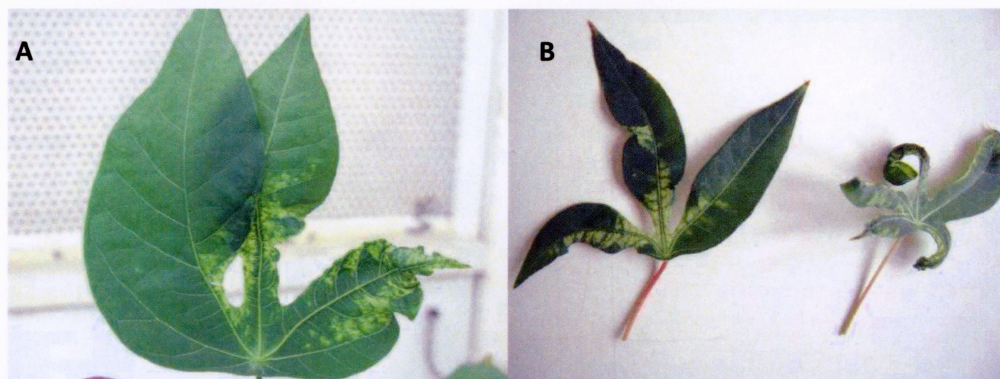


Figure 1.1 Symptoms development of ACMV (A) and SACMV-infected (B) susceptible cassava. Symptoms typically include, leaf margin distortion, green (A) and yellow (B) mosaic on leaves and reductions in leaf size.

1.3. Geminiviruses

1.3.1 Classification of *Geminiviridae* and genome organization

Geminiviruses (GV) are plant-infecting viruses which have been reported to cause major economical losses to many agricultural crops worldwide. The family *Geminiviridae* is divided into four genera, (*Mastrevirus*, *Curtovirus*, *Topocuvirus*, and *Begomovirus*) based on their genome organization, host range and insect vectors (Fauquet and Stanley, 2003). Begomoviruses, curtoviruses and topocuviruses have been found to naturally infect dicotyledonous plants whereas the mastreviruses have unique genomic features and have been found to infect both monocotyledonous and dicotyledonous plants (Stanley et al. 2005). These viruses possess circular, single-stranded DNA genomes that are packaged into twinned icosahedral capsids. GV genomes are either monopartite (containing one ssDNA molecule known as DNA-A) or bipartite (containing two ssDNA molecules known as DNA-A and DNA-B). Additionally, it has been reported that many monopartite begomoviruses are associated with additional subviral ssDNA satellites which are referred to as either alphasatellites (DNA- α) or betasatellites (DNA- β). Furthermore, subgenomic defective interfering (DIs) DNAs, which arise from deletions from their monopartite or bipartite helper virus, has also been reported.

The distinguishing attributes of the begomoviruses in particular, is that they contain a bipartite DNA genome, they have a specific dicotyledonous host range and that they are transmitted by the whitefly, *Bemisia tabaci* (Shahid et al. 2003). *South Africa cassava mosaic virus*-[South Africa:99] (SACMV-[ZA:99]) is one such begomovirus first reported in South Africa, (Berrie et al. 1998; Berrie et al. 2001), and has now been reported in Zimbabwe and Madagascar. SACMV-[ZA:99] has a bipartite genome which contains two covalently closed, separately encapsidated, circular ssDNA molecules

(DNA-A (2800nt) and DNA-B (2760nt) (Fig. 1.2). Both DNA molecules in bipartite GV's are necessary for viral infection. Replication is achieved by means of the rolling circle replication (RCR) mechanism where the ssDNA employs the replication machinery of the host and dsDNA replicative intermediates form mini chromosomes and associate with histones (Jeske 2009; Pilartz and Jeske, 2003). These dsDNA intermediates serve as templates for DNA replication and transcription which takes place in the plant nucleus (Hanley-Bowdoin et al. 2000; Gutierrez et al. 2004).

Sequence and phylogenetic analysis have shown that the coat protein (CP) of SACMV-[ZA:99] DNA-A is most closely related to EACMV isolates from Malawi, whereas DNA-B is mostly closely related to EACMV isolates from Uganda. Furthermore, the common region (CR) is most closely related to the monopartite virus *Tomato yellow leaf curl virus* (TYLCV), (Berrie *et al.*, 2001). DNA-A and DNA-B differ in their sequences except for the CR. This region ranges between 200-250bp in size and shares approximately 90% homology (Shahid et al. 2002). The CR contains important cis-elements that facilitate DNA transcription and replication *via* rolling circle mechanism. Six genes are encoded by DNA-A whereas two genes are encoded by DNA-B. DNA-A encodes for the coat protein (CP) as well as a set of four proteins. These viral proteins are necessary for viral replication and the encapsidation of both DNA molecules. The CP is a multifunctional protein involved in virus encapsidation and insect transmission. This was established in a study creating a chimeric/hybrid of the DNA-A molecule of ACMV by replacing its CP with the CP of the curtovirus; *Beet Curly Top Virus* (BCTV) (Briddon et al. 1990). Subsequent infection of the ACMV hybrid into *N. benthamiana* seedlings was still able to produce symptoms typical of ACMV infections as well as encapsidate ssDNA molecules of the hybrid ACMV molecule. In addition, the hybrid ACMV was able to be transmitted to the insect vector, *Circulifera tenellus* (beet leafhopper), of BCTV but the original ACMV molecule was not, indicating that the CP also plays an important role in vector specificity (Briddon et al. 1990).

The minus strand of DNA-A has two open reading frames (ORFs) that encode for the Rep/AC1 protein and a transcriptional activator (TrAP/AC2). 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 RCR (Laufs 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 (Legg and Fauquet, 2004). TrAP localizes to the plant host nucleus where it preferentially binds to ssDNA.

AC4 and (AC3) REn are proteins that are encoded from the positive strand of DNA-A. The AC4 ORF lies entirely embedded within the coding region of the Rep protein and it is the least conserved of all

the geminiviral proteins, both in sequence and in 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 turn results in enhanced disease progression.

The ORFs on DNA-B on the other hand encodes the proteins that are necessary and directly involved in the efficient systemic spread of the virus. These two proteins are encoded by BV1 from the positive strand and BC1 which is located on the minus strand.

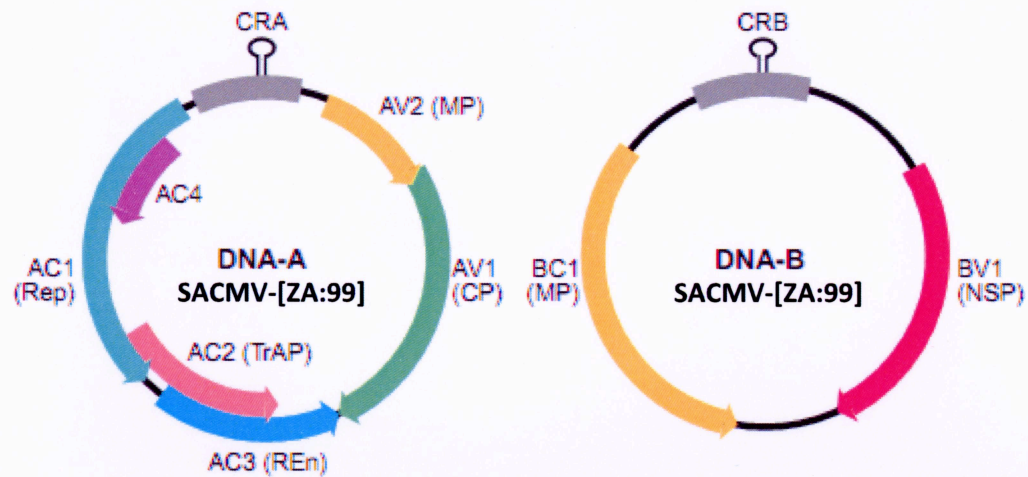


Figure 1.2 Genome organization of *South African cassava mosaic virus* (SACMV). DNA-A contains six open reading frames (ORFs) and DNA-B contains two open reading frames. ORFs are indicated by coloured arrows. The arrow head represents the direction in which transcription occurs once the virus is within it plant host (adapted from Berrie *et al*, 1998).

1.3.2 Geminivirus-host protein associations

Plant viruses, in particular geminiviruses (GVs) which have small ssDNA genomes, are limited in the number of molecular processes they can carry out. It is for this reason that GV manipulates host gene expression and associate with a number of host proteins in order to successfully replicate, move and overcome host defences. Since geminiviruses are unable to infect actively dividing meristematic tissue of the host; it needs to employ host proteins that will facilitate the transition and progression of infected G_0 cells into active S-phase cells. In order to overcome this, geminivirus use the Rep (AC1) protein to interact with host retinoblastoma protein (Kong et al. 2000, Nash et al., 2011). The retinoblastoma (Rb) gene family contains critical regulators of the G_1/S transition in the cell cycle. This protein is a substrate for cyclin-dependent kinases and has multifunctional and critical roles in DNA initiation, by binding to the origin and inducing the replication machinery of the host plant. Rb binds to members of the E2F-family, which are transcriptional activators. In doing this, it

camouflages the E2F activation domain without disturbing DNA binding at the cell-cycle regulated promoters (Gutierrez et al. 2004). Viral dsDNA replicative forms are most abundant in the S-phase nuclei.

AC1 is involved in protein-protein interactions in the cell. A study carried out by Settlege et al. 2005, illustrated that AC1 also interacts and binds to viral replication enhancer protein AC3 which in turn binds to the proliferating cell nuclear antigen (PCNA) and the C-complex replication apparatus early in replication. PCNA is a DNA accessory factor that is normally found in the S-phase (Castillo et al. 2003). In past studies it has been shown that AC3 protein greatly enhances the accumulation of viral DNA in infected cells and therefore enhances symptoms in the diseased plant tissue (Elmer et al. 1988). Before AC3 can interact and bind to PCNA and pRBR, it needs to undergo homo-oligomerization (Settlege et al. 2005). Another example of host proteins implicated in protein-protein interactions with geminiviruses are the host protein kinases known as GRIK (Geminivirus Rep-interacting kinase). Also, the *Arabidopsis* Synaptotagmins (SYTs), are a large family of Ca^{2+} /lipid binding proteins. SYTA, specifically, has been shown to localize to membrane-derived endosomes in plant cells and directly binds to the cell-to-cell movement proteins of CaLCuV and *Tobacco mosaic virus* (TMV) (Lewis and Lazarowitz, 2010). Several other proteins that have been shown to interact with geminiviruses are depicted in Table 1 (Jeske, 2009).

1.4. Plant-pathogen Interactions

Plants are continuously challenged by biotic and abiotic stresses which results in coordinated changes in the cellular and molecular control of the plant cell. In the case of a virus, replication and movement within the susceptible host is achieved by efficiently recruiting and exploiting host factors. This is true for both RNA and DNA viruses (Miguel et al. 1996; Gutierrez 2002; Ahlquist et al. 2003 and Gutierrez 2004). The identification of host factors and their contribution in viral replication is an area that has long been under scrutiny. Virus infection typically disrupts a host's normal physiology and this in turn is an indication of the biochemical and metabolic processes implicated in the disruption. Furthermore, the disruption in plant physiology is dependent on the compatibility and recognition that occurs between the host and invading pathogen. In general there are two broad pathways that a host-pathogen interaction could follow: either a resistant interaction or a susceptible one. This interaction between host and pathogen was first proposed by Flor in 1971 in the gene-for-gene model. The gene-for-gene model proposes that the interaction between host and pathogen is governed by the interaction and specificity between a single gene product encoded by the pathogen which is recognized by a specific plant encoded protein. There has been a wealth of advances made in the field of plant-pathogen interactions since Flor's hypothesis which have lead to

alternative hypothesis for plant pathogen interactions. These include the Guard hypothesis (van der Biezen and Jones, 1998), the Decoy model proposed by Van der Hoorn and Kamoun (2008) and more recently, the Bait and Switch model (Collier and Moffett, 2009).

1.4.1 Incompatible Interactions (Resistance)

An incompatible reaction involves the direct recognition and interaction of the dominant plant host genes and pathogen derived genes. R-and-Avr models are based on the “gene-for-gene” theory where it has been recognized that gene products encoded by pathogen genes (Avr genes or effectors) will interact with a receptor-like protein encoded by genes of a host (*R*) (Keen 1992, reviewed in Martin et al. 2003; Nimchuk et al. 2003). This interaction will subsequently result in the culmination of downstream cascading events that will lead a resistant phenotype. The major class of *R* gene proteins are well characterized and form the recognition component of the plant immune system. These *R* proteins belong to a small number of conserved structural classes where their recognition domain is either present in an extracellular (membrane bound) or intracellular environment. Structurally, *R*-genes consist of a nucleotide-binding site (NBS) and leucine rich repeat (LRR), which is connected by a homologous region called ARC (adaptor shared by APAF-1, certain R gene products and CED-4). Furthermore, *R* genes proteins are divided into further sub categories based on the presence of a N-terminal coil-coil (CC) domain or an amino terminal Toll-interleukin-1 receptor (TIR) domain (Bent, 1996). Interestingly, LRR proteins have shown to have significant sequence variation among *R* gene families which could directly correlate to pathogen recognition specificity (Ellis et al. 1999).

There are numerous *avr* and *R* gene models that have been reported in fungi, bacteria, oomycetes and viruses. Many of these models have been demonstrated in *Arabidopsis* or members of the *Solanaceae* family (Bendahmane et al. 1999; Cooley et al. 2000; Whitham et al. 1994). The *RPS4* *R*-gene for example belongs to the TIR-NBS-LRR family and has been reported to mediate resistance in *Arabidopsis* plants infected with bacterial pathogen *Pseudomonas syringae* (Gassman et al. 1999; Zhang et al. 2003). For viral pathogens, a number of avirulence determinants have been identified. These include the RNA replicase, coat protein, or movement protein. This has been studied in great detail in *tobacco mosaic virus* (TMV) infections where the *R* gene, *N*, has shown to confer resistance in transgenic tomato plants and in turn functions to limit replication and movement of TMV (Erickson et al. 1999). *N* encodes either full-length or truncated proteins and belongs to the TIR-NBS-LRR class of plant disease resistance genes (Li et al. 1999). This incompatible interaction between host plant and TMV is triggered by the presence of TMV 126 kDa replicase protein.

The recognition of Avr proteins by the matching host R-protein will often lead to the rapid activation of various defence mechanisms. The most common and rapid mechanism associated with active plant defence, following recognition of the pathogen, is manifested by the hypersensitive response (HR) (Greenberg 1997, 2004). The HR is characterised by a rapid increase in ion fluxes, extracellular oxidative burst (O_2^- and H_2O_2), transcriptional responses near the infection sites, and a localized cell death at the primary site of infection. Together, this is thought to limit the spread of the pathogen from the infection site (Chisholm *et al.* 2006; Lamb and Dixon 1997; Nemchinov *et al.* 2008).

The HR is followed by the culmination of a more generalized form of resistance known as Systemic Acquired Resistance (SAR). This mechanism is activated by the plant to combat any possible secondary infection from a broader range of pathogens. The development of SAR depends on pathogen recognition, translocation of the recognition signal, and the expression and regulation of defence genes upon pathogen infection. SAR requires an accumulation of signalling molecule Salicylic acid (SA) as well as downstream signalling molecule NON EXPRESSOR OF PR1 (NPR1), (Durrant and Dong, 2004). Early work on *Arabidopsis* mutant *npr1* showed that these mutants have accumulated levels of wild-type SA in response to infection with avirulent pathogens, however, they were unable to activate pathogen-related genes or develop biologically or chemically induced SAR (Cao *et al.* 1994; Delaney *et al.* 1995). This, therefore demonstrated that NPR1 is likely a key regulator of SAR. More recently, studies have revealed that the establishment of SAR requires a number of additional signals, which precede the systemic accumulation of salicylic acid. Some of these include jasmonates, and indole-derived metabolites (Truman *et al.* 2007; Truman *et al.* 2010).

1.4.2 Compatible Interaction (Susceptibility/Pathogenecity)

Compatible interactions between a susceptible host and a virulent pathogen lead to the expression of common sets of defence-related and heat shock genes (reviewed in Whitman *et al.* 2006). Unlike resistance mechanism, there is no recognition between a plant and a pathogen in a susceptible reaction which leads to a susceptible host developing symptoms as an indication of successful pathogen replication. Although, susceptible hosts do not have rapid and direct defence responses such as the HR, they do possess defence responses that have been reported to limit but not stop the progress of pathogen growth. The induced signalling intermediates that have been reported to facilitate defence in a susceptible host include: ethylene (ET), jasmonic acid (JA) and salicylic acid (SA) (Schenk *et al.* 2000). Each of these hormones can activate a specific defence signalling pathway which can act individually, synergistically, or antagonistically and activation is dependent upon the invading pathogen (Thomma *et al.* 2001).

1.5 Plant Defence Response

1.5.1 Basal Innate Immunity in plants

Since constitutive expression of defence pathways in plants can compromise the overall fitness and normal growth of a host plant, it has been suggested that plants have evolved intricate mechanisms to exert control over pathogen induced defence pathways. Extracellular surface receptors in plants referred to as pattern recognition receptors (PRRs), are able to recognize conserved microbe-associated molecular patterns (MAMPs) or pathogen-associated molecular patterns (PAMPs) which results in the triggering of a general broad defence response described as PAMP-triggered immunity (PTI) (Gomez-Gomez and Boller 2000; Huffaker et al. 2006; Miya et al. 2007) . PTI is therefore considered to constitute the first layer of nonspecific plant immunity that restricts pathogen proliferation.

Some pathogens have evolved the ability to evade PTI by secreting specific effector molecules that suppress or interfere with the induction of PTI. Effector-triggered immunity (ETI) is brought about by plant disease resistance (R) proteins that activate highly efficient defence responses upon specific recognition of pathogen effectors (Jones and Dangl 2006). ETI is particularly effective against many biotrophic pathogens and activated immune responses in ETI are reported to occur faster and are more prolonged and more robust than those in PTI (Jones and Dangl, 2006; Tao et al. 2003; Tsuda and Katagiri 2010).

1.5.2 RNA silencing pathways and anti viral immunity

In addition to PTI and ETI defence mechanisms, plants also possess an additional type of adaptive immunity known as RNA silencing which plays a major role specifically in antiviral defence response. RNA silencing 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. In addition it serves as an antiviral defence strategy against invading viral pathogens thereby restricting the accumulation and/spread of viral pathogens (Waterhouse et al. 2001) The RNA silencing pathway is triggered by double stranded RNA structures. Since the majority of known plant viruses have RNA genomes, it is not at all surprising that dsRNA intermediates formed during RNA virus replication or secondary structures formed from the self-complementary folding of ss viral genomes, are strong inducers of RNA silencing and sequence-specific degradation against viruses. Although GVs have DNA genomes and do not replicate using dsRNA intermediates, they have been still been shown to be targets of RNA

the RNA silencing machinery. One likely reason is that dsRNA intermediates are formed in the process of bidirectional transcription for GV (Bieri et al. 2002; Vanitharani et al. 2005).

Although complex, in general, RNA silencing pathways can be divided into two basic groups: Post-Transcriptional Gene Silencing (PTGS) and Transcriptional Gene Silencing (TGS). This is dependent on the activation of one of three RNA silencing pathways – small-interfering (si) RNA, micro (mi) RNA and RNA-mediated DNA methylation as described by Baulcombe, (2004) and others (Carrington and Ambros 2003; Vaucheret, 2006; Vazquez, 2006). 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.

The RNA silencing process involves several different pathways and target molecules, and each of these is mediated by specific enzyme catalysis. The main enzymatic steps in the silencing process is the conversion of ssRNAs into double-stranded form by RdRP, the cleavage of the dsRNAs by DCL, association of AGO to siRNA and the degradation of the target RNAs by the RNA-induced silencing complex (RISC) (Fig 1.3). DCL plays a crucial role in this process and will determine the length of siRNA produced. In *Arabidopsis*, DCL2, DCL3, and DCL4 produce virus-derived siRNAs of distinct size classes (Borsani et al. 2005; Park et al. 2002; Xie et al. 2004; Xie et al. 2005) 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)

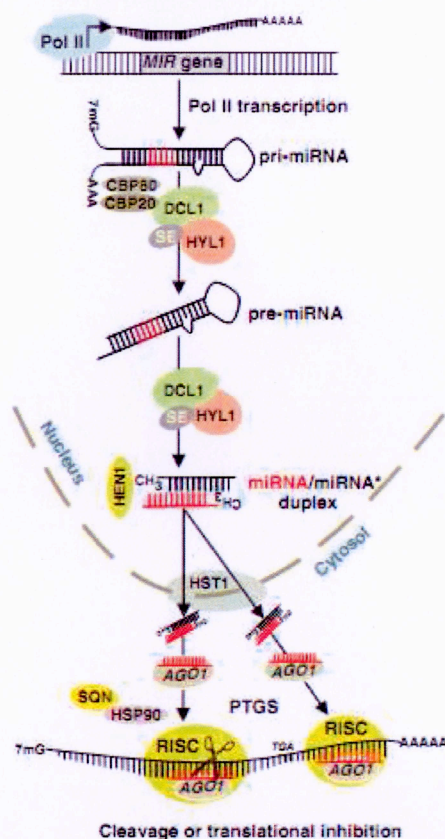
The siRNA pathway is mediated by siRNAs that are entirely complimentary to their targets and are made from both endogenous and exogenous mRNA. siRNAs can be induced by over-expressed transgenes, transposons and viral pathogens, and suppress the expression of the corresponding RNAs, usually by sequence-specific cleavage of the target.

The miRNA pathway, in contrast to siRNAs, is mediated by miRNAs, which are endogenous small RNAs that allow for mismatches and are produced by genes other than those that they are found to regulate. miRNA-mediated regulation is therefore referred to as heterosilencing (Bartel 2004). miRNA is transcribed in the nucleus by RNA polymerase II, into long primary miRNA (pri-miRNA). Pri-miRNAs are then cleaved by DCL1 (Park et al. 2002) into precursor (pre)-miRNA of approximately 65nt in length. A single *pri-miRNA* may contain from one to six miRNA precursors. This is subsequently translocated to the cytoplasm (Zhu 2008). In the nucleus, DCL1 has been shown to interact strongly with other proteins, HYPONASTIC LEAVES (HYL1), and C2H2 Zn-finger SERRATE (SE) (Dong et al. 2008). Once cleaved, the mature miRNA is methylated by the methyltransferase HEN1,

which adds a methyl group at the 2' OH of the 3' end of mature miRNA (Yang et al. 2006). The mature miRNA then translocates into the cytoplasm, where it is recruited in the RISC.

The RNA-mediated DNA Methylation pathway (RdDM) was first reported in the early 1990's in transgenic tobacco expressing the potato spindle tuber viroid (PSTVd) where PSTVd showed evidence of viroid methylation after autonomous viroid RNA-RNA replication (Wassenegger et al. 1994). Soon after, Park et al (2002), demonstrated that the 35S promoter of the hygromycin phosphotransferase (hpt) gene showed heritable promoter methylation which was directed by dsRNA cleaved into 23nt siRNAs in tobacco. These studies indicated that small RNAs were involved in transcriptional gene silencing and that it may be similar to siRNA-induced post-transcriptional gene silencing observed in plants therefore suggesting a link between RNAi and DNA methylation (Hamilton and Baulcombe 1999). There is a considerable body of evidence that demonstrates how 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, GV 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. Silencing pathways leading to DNA methylation is mediated by RNA Polymerase IV (Pol IV), RDRP2 and DCL3. In this pathway Pol IV and RDRP2 synthesize dsRNA which is then cleaved by DCL3 to generate siRNA. In one other pathway, Pol IV acts together with AGO4 and one of the RDRPs to produce specific siRNAs (Herr et al, 2005).

A) miRNA biogenesis and function



B) siRNA biogenesis and function

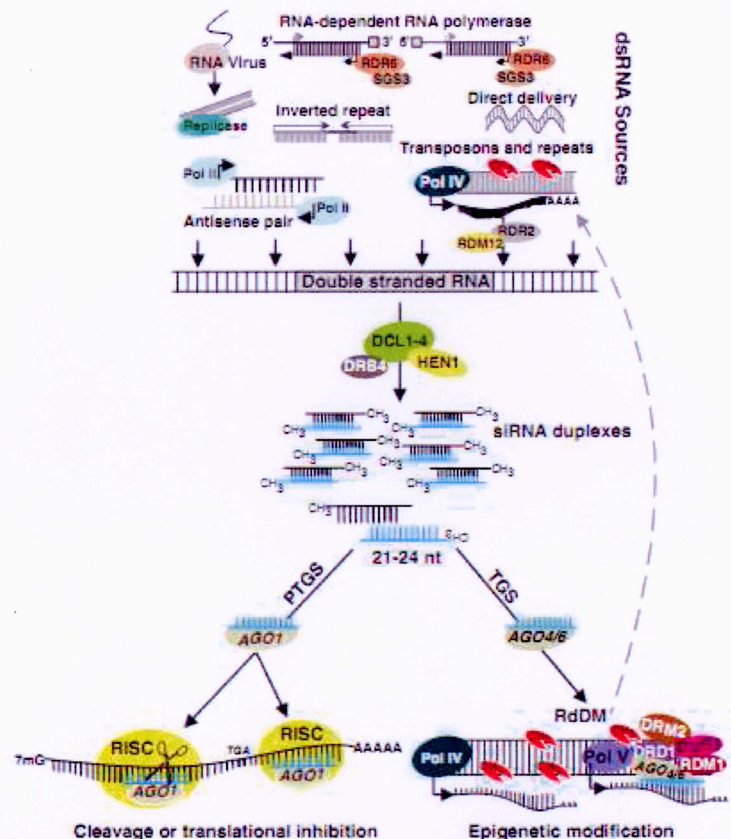


Figure 1.3 mi-RNA (A) and si-RNA (B) mediated pathways in plants. (Obtained from Khraiweh et al. 2012)

1.5.3 Viral Counterdefence – Viral RNA Silencing Suppressors

The examples above clearly demonstrate that host plants have developed a number of different defence strategies in order to reduce the fitness cost of a pathogen therefore avoiding or limiting pathogen infection (Agnew et al. 2000). In short, host plants will activate a number of cascading events which results in resistance or tolerance to invading pathogens and pathogens in return will attempt avoiding these defence strategies. Tolerance in plants is explained by the ability of a host plant to adapt to disease. Tolerant plants are usually not associated with any disease symptoms even though pathogen growth is maintained (O'Donnell et al. 2003). Due to the complex interplay and dynamics underlying plant-pathogen interactions, compounded by the strong selection pressures mutually exerted between the host plant and invading pathogen, it is not surprising that plants and pathogens are thought to co-evolve (Frank, 1992; Sacristán and García-Arenal, 2008; Burdon and Thrall, 2009). Viruses, in particular, have acquired a counter defence mechanism against RNA silencing pathways by means of viral suppressor proteins termed Viral RNA silencing suppressors (RSS) (reviewed in Lewsey et al. 2009). RSS proteins typically function by binding to

either siRNA or dsRNA therefore decreasing the availability of these intermediates for the silencing machinery (Csorba et al. 2009). Therefore RSS are likely to counteract RNA-mediated defence at one of three possible steps, 1) preventing the generation of siRNAs (2) inhibiting the incorporation of siRNA molecules into effector complexes or (3) interfering the RISC effector complex (Lakatos 2004).

One of the first RNA suppressor proteins characterized and described has been for the 2b protein of for positive-sense RNA virus, Cucumber Mosaic Virus (CMV) (Brigneti et al. 1998). The 2b protein is a small 12kDa multifunctional polypeptide that binds siRNAs thereby inhibiting the anti-viral silencing of CMV. In addition, 2b also disrupts mi-RNA regulated gene expression pathways through its interaction with host AGO1 or disruption of DNA methylation through its interaction with AGO4 (González et al. 2010; Goto et al. 2007; Lewsey et al. 2009 ; Ye et al. 2009 ; Zhang et al. 2006). Studies investigating RSS in other virus families have shown that these suppressor proteins behave differently for different viruses. For example, the CP of Turnip crinkle virus (TCV) inhibit the generation of siRNAs that are derived from the early initiation step of the siRNA pathway where dsRNA is processed into siRNA molecules by DICER (Qu et al. 2003). Also, the 19kDa (p19) RSS encoded by tombusvirus *Tomato bushy stunt virus*, forms homodimers and preferentially binds to 21-nt siRNAs thereby preventing a downstream step involving cleavage of cognate RNA by RISC (Lakatos et al. 2004; Lakatos et al. 2006; Liu et al. 2012; Omarov et al. 2006). The P25 movement protein of Potato virus X prevents systemic spread of silencing through the vascular system (Bayne et al. 2005) and more recently has been demonstrated to interact with AGO1 from *Nicotiana benthamiana* to mediate degradation through the ubiquitin-proteasome pathway (Chiu et al. 2010).

Our group however, is particularly interested in viral suppressors identified in the *Geminiviridae* family. For these dsDNA viruses, at least four main ORFs have been demonstrated to encode viral suppressor proteins. These include AC2 and AC4 which are present on DNA-A of bipartite GV genomes, C2 present of monopartite GV genomes and β C1, encoded on betasatellite DNA β . Studies carried out on the small protein AC2 (also known as TrAP) of the Kenyan strain of *African cassava mosaic virus (ACMV-[KE])* and the homologous C2 of the related monopartite virus, *Tomato yellow leaf curl China virus (TYLCCNV)*, have implicated these proteins in suppression of RNA silencing (van Wezel et al. 2002; Voinnet et al. 1999). AC4, which is completely embedded in the AC1, is the least characterized ORF of the GVs and the function of AC4 is not completely defined, but it is believed to be membrane bound and involved in symptom development. Additionally, it has been shown to act similarly to AC2/TrAP and has PTGS silencing activity (Latham et al. 1997; Patil and Dasgupta 2006, Vanitharani et al. 2004). β C1 is a positionally conserved ORF which is known as major symptom modulating component and has been demonstrated to function as an RSS (Cui et al. 2005). Nuclear

localized β C1 protein appears to affect the silencing pathway similar to that of AC2 at the initiation step (Cui et al. 2005).

1.6. Gene Expression Profiling

In order to understand the molecular mechanisms underlying host-pathogen interactions, it is necessary to identify gene transcripts that are differentially expressed in infected host vs. a healthy host. At present, there is a variety of well established methods that have been used for quantifying the mRNA abundance in plant tissues. These include a number of techniques such as *In situ* hybridization, massively parallel signature sequencing (MPSS), RNA-blot, Differential display, and Serial analysis of gene expression (SAGE), macroarrays, microarray technology and more recently, next generation sequencing (NGS).

1.6.1 Microarray technology

Molecular biology is entering the area of functional genomics, where DNA microarrays are being used to establish which genes are overexpressed or repressed following a given treatment. Although a number of techniques such as *in situ* hybridization, massively parallel signature sequencing (MPSS), RNA-blot, Differential display, Serial analysis of gene expression (SAGE) and macroarrays have been exploited over the past decade for identifying genes expression changes that occur as a consequence on pathogen attack, DNA microarray technology allows for multiparallel analysis of thousands to tens-of-thousands of genes simultaneously (Schena *et al.*, 1995, Whitman *et al.*, 2006).

DNA Microarray technology involves the hybridization of nucleic acid to a 'gene chip' thereby allowing one to systematically monitor the expression of hundreds to thousands of genes on a global scale simultaneously (Schena et al, 2000; Reinsink and Buell, 2005; Kuella et al. 2001). An additional favourable feature of microarray analysis is that the arrays produce sensitive and accurate expression data. DNA microarrays have a wide range of applications including, SNP genotyping, gene discovery, gene expression profiling and diagnostics (Schena *et al.*, 1998). There are two basic types of microarrays: the oligonucleotide (oligo)- based arrays and cDNA microarrays. The two different types are based on differences in the chemical state of the probes that has been immobilized on the solid surface. cDNA microarray in particular are designed from the known sequences of expressed sequenced tags (ESTs) which represent a single gene. cDNA is amplified from these ESTs, followed by purification and high-density robotic spotting onto positively charged poly-lysine or aminosilane glass microscope slides (Harrington et al. 2002). The cDNA probes, which are immobilized onto the microarray slide, provide a platform which allows one to monitor the steady-state RNA levels in the

samples under investigation. The basis of the method is the specificity and affinity of nucleotide complementary base-pairing. In order to allow a side-to-side comparison of the changes in gene expression, two target samples need to be labelled with two distinct fluorochromes– the control and an experimental sample. The two fluorescent dyes that have been predominately used are the cyanine dyes (Cy3 and Cy5). Differential expression analysis utilizing microarrays is an advantageous technique as these two differently labelled samples can be hybridized together on a single gene chip.

Microarray analysis has already been used in studies investigating host-pathogen interactions in plants (Babu et al. 2008; Dowd et al., 2004; De Vos et al. 2005, Dardick, 2007; Okinaka et al. 2002; Whitham et al, 2003;). Virus associated host-pathogen interactions in particular, have primarily been conducted in the plant model system *Arabidopsis* (Whitham et al. 2003; Golem and Culver, 2003). *Arabidopsis* is a well developed experimental model system that has been exploited in a number of diverse molecular studies, including those involving microarrays. Some of these studies include customised cDNA microarrays for analysis of gene expression during programmed cell death in *Arabidopsis* cell suspension culture (Swidzinski et al. 2002) and a custom array containing defence responsive genes used to investigate the response of a lipopolysaccharide in *Arabidopsis* (Zeidler et al. 2004) *Arabidopsis* is a favourable plant system as the genome has been fully sequenced. There have also been extensive generations of ESTs for this model system which have been used as probes on microarray slides. Whole genome microarrays have been developed for *Arabidopsis* including the *Affymetrix Arabidopsis* ATH1 Genome Array and the *Agilent Arabidopsis* (V4) 4x44K Gene Expression Microarray.

Microarrays have proven to be an extremely useful tool and it is almost indispensable in the field of functional genomics and molecular biology. For this reason, great efforts have been made in order to easily compare and interpret the results of individual microarray experiments which can be independently verified if required (Brazma et al. 2001; Spellman et al. 2002). In order to confidently assess thousands of genes from microarray experiments, it is essential to rigorously plan experiments, ensure uniform and precise sampling/harvesting, ensure high quality processing of RNA samples and to use appropriate data analysis tools for downstream data processing. For this reason, it is now mandatory for all microarray experiments to meet all standards and comply with the rules set out in MIAME (minimum information about microarray experiment) prior to publication (Brazma et al. 2001 and 2006).

1.6.2 Next-generation sequencing technologies

There has been a huge leap in recent years in the “omics” field with the introduction of Next-generation sequencing (NGS) platforms (Shendure and Ji, 2008). Next-generation sequencing offers a broad range of applications including Whole genome sequencing, SNP-genotyping, RNA-seq, Chip-seq, Exome sequencing, Microbiome sequencing, Ribonomics and Epigenomics. Measuring changes at the transcriptome level (RNA-seq) is an ideal application of NGS. Traditionally global transcriptome profiling of this nature would involve a number of steps including the construction of complementary DNA (cDNA) libraries followed by Sanger sequencing of these ESTs and possibly microarray analysis. Unlike microarray or other hybridization-based technologies, RNA-Seq is not limited to the detection of known transcripts; therefore allowing the identification, characterization and quantification of new splice isoforms. RNA-seq is a therefore favourable and practical application for non-model species as it requires little to no genetic tools and can still be employed even when sequence data and resources are limited. Given that the primary focus of RNA-seq is restricted to the coding regions of genes, which typically have fewer repetitive elements and higher GC content, it is much easier to execute than whole genome sequencing. To date, a number of studies have used RNA-seq transcriptomic approaches in non-model plant systems, for example, some of these non-model systems include soybean (Severin et al. 2010), pea (Franssen et al. 2011), olive (Alagna et al. 2009), eucalyptus (Novaes et al. 2008), chickpea (Garga et al. 2011) and poplar (Qiu et al. 2011).

Today, there are a number of NGS machines available but the three main platforms used for the generation of high-throughput sequence data include, the 454 GS FLX Titanium system from Roche, the Genome Analyzer (GA) IIx system from Illumina and the SOLiD system from Applied Biosystems. These systems differ in their chemistry, complexity and time required for the generation of libraries, but they are fundamentally similar in that they follow four major steps: (1) random shearing of DNA either through sonication or nebulisation, (2) the ligation of universal adapters to 5' and 3' ends of the sheared DNA fragments (3) immobilization of adapter-flanked DNA fragments and (4) amplification of immobilized fragments that serve as templates for the sequencing reaction (Shendure and Ji, 2008). Although 454, Illumina and SOLiD platforms have dominated the market over the past few years, it should also be mentioned that newer NGS platforms have recently been commercialized and introduced onto the market. Other NGS technologies include the Ion torrent system (Life technologies) and Heliscope (Helicos BioScope). Unlike microarray technology, where hundreds to tens of thousands of genes can be assessed at a time; NGS can generate sequence data in the order of hundreds of thousands to hundreds of millions of sequence reads. Although it's been reported that error rates of NGS is on average ten times higher than that of Sanger sequencing

(Shendure and Ji , 2008), the massive throughput resulting from the sequencing results in a high base coverage, allows for the screening of sequencing errors variations (Hillier et al. 2008).

1.6.2.1 The Applied Biosystems (ABI) SOLiD™ Sequencing system

The SOLiD sequencing technology was developed by Life Technologies and first become commercially available in 2008. In comparison to Roche 454 and Illumina sequencing platforms that make use of “*sequencing-by-synthesis*” technology, the SOLiD system uses the “Sequencing by Oligo Ligation Detection” technology. Library preparation is the first crucial step of the SOLiD sequencing work flow and is similar that of Roche 454 and Illumina where adaptor-flanked DNA fragments are produced (Shendure & Ji, 2008). This is subsequently followed by the amplification of DNA libraries. Like 454 sequencing, SOLiD technology requires DNA amplification where DNA fragments are immobilized to the surface of paramagnetic beads and carried out using emulsion PCR (Dressman et al. 2003). Following amplification, the emulsion is broken and the resulting PCR amplicons attached to the beads are then covalently bound to a specially treated glass slide (Mardis, 2008). Sequencing of amplified DNA is achieved by annealing of a sequencing primer (P1) followed by the synthesis of DNA which is driven by a ligase enzyme and not DNA polymerase (Shendure et al. 2005).

One of the unique features of SOLiD sequencing is that use of 2-base coding probes which binds fluorescent primers to DNA two nucleotides at a time therefore resulting in a sequencing reaction based on “colourspace”. For any given sequence in colorspace, there are four possible mappings into basespace. The conversion from colourspace to basespace requires knowledge of at least one base in the sequence that the colourspace mapping was derived from. The main advantages of colourspace measurements are that it improves the accuracy of the sequencing reaction and enables the detection of errors. After sequencing, the end output is sequence files in SOLiDs colourspace format .csfasta (Life Technologies 2008). However, these can be converted into regular sequence files in FASTA format using dynamic programming (Li & Durbin. 2009).

1.6.2.1.1 Paired-end sequencing

The Roche 454 GS FLX Titanium system is able to generate read lengths of up to 700 bp with accuracy of approximately 99.9%. The Illumina and SOLiD platforms however are short-read sequencing technologies that generate between 85bp and 150bp read-lengths respectively (Liu et al, 2012). Short, however, this is read lengths are often considered to be a limitation but this is partially overcome by the development of paired-end sequencing. Pair-end sequencing can be performed on

all three platforms and makes use of paired-end tags (PETs) are shorter sequences originating from the both the 5' and 3' ends of target DNA (Fullwood et al. 2009). Although short, PETs contain enough sequence information which allows for uniquely mapped of tags to the genome which in turn represents the entire DNA fragment of interest.

1.7. Model and natural plant host systems in host-pathogen interaction studies

1.7.1 *Arabidopsis and Nicotiana*

Arabidopsis belongs to the mustard family (*Brassicaceae*). *Arabidopsis* has proven to offer several advantages for laboratory use, for example, it has a small plant size, it has a short lifespan and it is highly fertile and produces large amounts of seed (Anderson and Wilson, 2000, Hays 2002). It has one of the smallest plant genomes (approximately 125Mb, Mysore et al., 2001) which is one of the main factors that it was chosen to be fully sequenced by The Arabidopsis Genome Initiative in 2000. The sequences generated from this project were subsequently annotated by The Arabidopsis Genome initiative. This has therefore made a large number of *Arabidopsis* expressed sequence tags (ESTs), gene expression databases and cDNA microarrays readily available. Furthermore, *Arabidopsis* also serves as an invaluable model system in plant pathology. It is host to a large number of microbes belonging to different taxonomic classes, including oomycetes, fungi, viruses and bacteria (reviewed in Schlaich 2011). Its susceptibility, in particular, to a number of RNA viruses was demonstrated by Whitham et al. in 2003 where the gene expression changes in *Arabidopsis* in response to five positive sense RNA viruses including : *turnip vein clearing tobamovirus* (TVCV), *oilseed rape tobamovirus* (ORMV), *potato virus X potexvirus* (PVX), *cucumber mosaic cucumovirus* (CMV), and *turnip mosaic potyvirus* (TuMV) was monitored and compared. In addition, a number of companies such as Affymetrix and Agilent have exploited the availability of the *Arabidopsis* genome and have designed and created oligo microarray genechips that represent the full genome of this model host. These are available commercially as the *Arabidopsis AH1* Genome array and 4x44k *Arabidopsis Expression* microarray respectively. These slides have already successfully been used in various studies including host-pathogen interactions (Ascencio-Ibáñez et al. 2008)

Although *Arabidopsis* is considered a model plant system for many host-pathogen interactions (Schlaich 2011), *N. benthamiana* Domin is the plant system of choice for virus infection (Goodin et al. 2008). *N. benthamiana* Domin belongs to the nightshade family (*Solanaceae*). It has been described as “universally susceptible” to many plant viruses, which may in part be attributed to the presence of a mutation which occurs naturally in the *Nicotiana* RNA-dependent-RNA polymerase gene (NbRdRP1m) (Yang et al. 2004; Goodin et al. 2008). In addition, since *N. benthamiana* serves as

a host for multiple virus species, it allows one to comparatively analyse host responses to different viral pathogens. The commercial availability of a 13 000 oligo *N.benthamiana* array through NimbleGen from 2008 (Feulner, 2008) and the recent release of the draft genome sequence for *N. benthamiana* in 2012 (Bombarely et al. 2012) supports its use as a model for plant–pathogen interactions.

Arabidopsis and *Nicotiana* are therefore perfectly well suited for gene expression studies elucidating the mechanisms underlying host-pathogen interactions.

1.7.2 Cassava: EST's, Genomics and Transcriptomics

In the last decade, great attempts have been made in generating sequence data for natural hosts, such as cassava, in an effort to develop genetic mapping tools and markers for crop improvement. The availability of EST, genomics or transcriptome data for cassava has begun to make a great contribution to the improvement of knowledge and information of this once neglected crop. Suarez et al. (2000) were the first group to generate approximately 500 cassava ESTs from a cross between Tropical Manihot Series 30572 (TMS30572) and Colombia 2177-2 (CM2177-2) using cDNA-AFLP. TMS30572 x CM2177-2 are both elite cassava cultivars from Nigeria and Colombia, respectively (Fragene et al. 1997). TMS30572, in particular has good food quality characteristics and is disease resistant (FAO). Subsequently, various laboratories started constructing a number of EST and cDNA libraries in order to identify genes involved in the various developmental stages of cassava tissues, and genes involved in starch biosynthesis (Anderson et al. 2004; Lokko et al. 2007; Sakurai et al. 2007). An EST search in January 2013 using the search term “*Manihot esculenta*” revealed that there are currently 80 652 cassava ESTs deposited into Genbank which have come from at least 13 different submissions.

More recently, with the introduction of high-throughput “omics” tools, cDNA and oligo microarrays have been developed using publically available EST sequence data from NCBI, TIGR and RIKEN (Ayling et al. 2012). The first of these, an array printed with 5700 cDNAs, was used to monitor gene expression changes in an incompatible interaction between *Xanthomonas axonopodis* pv. *manihotis* (*Xam*) and the resistant cassava variety MBra685 (Lopez et al. 2005). More recently though, a number of studies have employed 60-mer oligo microarray platforms, consisting of just over 20 000 oligo probes, which has been designed and printed through Agilent technologies (An et al. 2012, Utsami et al. 2012, Yang et al. 2011) More promising advances are still being made with the introduction of next-generation sequencing (NGS). In 2009, collaboration between 454 Life Sciences (Roche) and the U.S. Department of Energy -Joint Genome Institute (DOE-JGI) resulted in the whole

genome shotgun sequencing of cassava using the 454 GS FLX Titanium platform (Prochnik et al. 2012). Sequencing was carried out on cassava cultivar AM560-2, which is a S3 line selfed from cultivar MCOL1505. The draft genome sequence was released in November 2009 but was later updated to version Cassava 4.1 in October 2010 which included additional sequence data. The *de novo* assembly of cassava has resulted in 12,977 scaffolds spanning approximately 70% of the genome (533Mb of the predicted 760Mb). This data is publically accessible through phytozome (<http://www.phytozome.net/cassava#D1>).

2. RATIONALE AND OVERALL OBJECTIVES OF STUDY

Cassava is cultivated In Southern Africa in the Northern Province, Mpumalanga, KwaZulu Natal, Swaziland, Zimbabwe and Mozambique. Since this crop is robust and can withstand many harsh conditions including nutrient poor soils and low rainfall, it serves as the perfect famine reserve and food security crop for many subsistence farmers. In addition, it is an important source of starch for industrial applications and biofuels. The Department of Agriculture, Forestry and Fisheries, Republic of South Africa reported that in South Africa alone in 2005, that gross income generated from cassava starch production was estimated in the order of R27million. Cassava, like many crops however is vulnerable to infection by a number of pathogens in the field. One of the main and most economically important constraints threatening cassava production is infection by at least one of eight 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. Among the eight species infecting cassava, is *South African cassava mosaic virus* (SACMV) which was isolated from field-infected cassava in SA in 1999.

Since cassava has long been neglected by research in favour of cereals (i.e. maize, rice, wheat, and barley, millets), compounded with the limited information concerning Geminivirus-plant host interactions, means that there is insufficient information regarding gene expression changes that occur in CMD-infected cassava. With the recent availability of advanced high-throughput gene expression technologies and the cassava genome sequences, we hypothesised that a number of these tools could be used in order to elucidate some of the mechanisms underlying disease in model hosts *Nicotiana benthamiana* and natural host cassava caused by *South African cassava mosaic virus*-[ZA:99] (SACMV-[ZA:99]). Improved disease control strategies against CMD will depend on the outcome of the candidate genes identified from these studies.

Therefore, the main objective of work presented in this thesis was to indentify genes involved in the replication, progression and maintenance of SACMV-[ZA:99] infection using a combination of

gene expression approaches including microarray technology and SOLiD next generation sequencing technology (NGS).

2.1 Specific Aims

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

Outline of this thesis

Chapter 2: At the conception of this project, there were only a limited number of microarray studies investigating virus-host interactions. Furthermore, the numbers of species-specific arrays commercially available for non-model host were limited. The main aim of this study was to use a customised *Arabidopsis* SSH array (representing a pathogen-inducible cDNA subset) in order to identify genes that were activated in response to SACMV infection in *N. benthamiana*. We therefore took a heterologous /Cross-species approach in order to elucidate some of the genes involved in a SACMV-[ZA:99]-*Nicotiana benthamiana* interaction over the progression of disease. Microarray analysis was carried out on infected leave tissue sampled at four timepoints post inoculation (7, 14, 21 and 35dpi) and monitored for temporal changes in gene expression. Results indicate that many of the genes identified are involved in the regulation of transcription and translation, defence response, metabolism, cytoskeleton and plasmodesmata organization, chloroplast functions and photosynthesis.

Chapter 3: With the availability and access to a commercially constructed oligo *Nicotiana benthamiana* array (NB-array) through Roche Nimblegen, our main aim was to conduct a self-species microarray hybridization assay of systemically SACMV-[ZA:99]-infected *N. benthamiana* leaves tissue at 21days post inoculation (dpi). By using the Nb-array, it was successfully demonstrated that many of the transcripts associated with this geminivirus-plant interaction, in systemically infected leaf tissue, were consistent with those observed in similar RNA virus studies dissecting the molecular interaction of a pathogen in a plant host.

Chapter 4: The main aim of this study was to use a RNA-seq next- generation sequencing approach in order to identify host responses to SACMV-[ZA:99] in a susceptible(T200) and tolerant cultivar (TME3) of cassava from early infection (12 dpi) to systemic infection (T200) and recovery(TME3) (67dpi). A time course (12, 32 and 67dpi) RNA-seq study was carried out using the ABI SOLiD platform. To our knowledge, this study provides the first time course study of a cassava-geminivirus infection, using high-throughput analysis. In brief, transcriptome alterations induced by SACMV-[ZA:99] in TME3 were significantly lower compared to T200. This finding strongly suggests that TME

is not a classical “resistant” landrace, associated with CMD2, but rather is tolerant due to failure of SACMV-[ZA:99] to elicit wide transcriptome changes. TME3 rather appears to display a recovery phenotype, which may involve persistent RNA silencing maintenance.

Chapter 5: This chapter collates all the results obtained from research chapters 2, 3 and 4. The results obtained during the course of PhD work are discussed in a context of the latest insights regarding plant-virus interactions and defence mechanisms. The similarities and differences between these studies are discussed and how the major findings of this research can contribute to elucidating some of the mechanisms that occurs in geminivirus-plant infection systems. Furthermore, future recommendations and how this data can be used to formulate new hypothesis for follow-on experiments has been put forward.

Chapter 2

A heterologous microarray approach to identify host responses in
Nicotiana benthamiana to SACMV-[ZA:99] infection

2.1 Abstract

Studies elucidating the mechanisms underlying host responses to a pathogen attack, although well studied in model host *Arabidopsis*, is poorly understood in other model and non-model plants, such as *Nicotiana benthamiana* and cassava, respectively. In order to analyse changes in gene transcript levels in *Nicotiana benthamiana* leaf tissue infected by geminivirus, *South African cassava mosaic virus*-[ZA:99], a heterologous cDNA microarray approach was used. cDNA microarray slides contained 5000 ESTs obtained from a suppressive subtractive hybridization (SSH) library of *Arabidopsis thaliana* Col-0 leaves in response to inoculation with the bacterial pathogen *Ralstonia solanacearum*. Evaluation of symptoms during the infectivity study revealed that plants started to show symptoms at 14 days post inoculation (dpi), and by 35 dpi were fully symptomatic, with the appearance of characteristic disease symptoms such as leaf distortion, chlorosis and stunting. Infected leaf tissue was harvested at 4 different timepoints and used for DNA extractions (to confirm infection and viral titre at the various timepoints) as well as RNA extractions (used in subsequent microarray experiments). Real-Time qPCR revealed that plants showed an increase in viral titre between 7dpi (0 molecules of SACMV-[ZA:99]-DNA-A/ng of TNA) and 35 dpi (1.79×10^3 molecules of SACMV-[ZA:99]-DNA-A/ng of TNA). A test of cross-hybridization percentage was performed between *Arabidopsis* and *N. benthamiana*, and was shown to be high (70%). Following normalization, a cut-off criteria of 1.5-fold change at $p < 0.05$ was used to filter data, a total of 73 differentially expressed genes were identified across all timepoints in SACMV-[ZA:99]-infected leaf tissue. In general, the total number of genes differentially expressed during the infection increased over time. Suites of genes that exhibited altered transcript abundance were identified and found to encode proteins involved in the regulation of transcription and translation, defence response, metabolism, cytoskeleton and plasmodesmata organization, chloroplast functions and photosynthesis.

2.2 Introduction

Successful virus infection in host plants is a complex event that often leads to drastic changes in cell physiology and metabolism. Since viruses possess small genomes, they often recruit and interact with a number of host proteins and metabolites in order to facilitate the replication and movement of their genomes, which in turn results in the disruption of normal host processes (Whitham et al. 2003). Current plant-virus models demonstrate that successful RNA and DNA viral infection in a number of model and non-model host plant species affects the expression of many plant genes both temporally and spatially (reviewed in Culver and Padmanabhan, 2007). In addition to changes in cellular processes, basal host defences to these pathogens is also elicited by susceptible plants (Postnikova and Nimchinov, 2012).

Monitoring gene expression changes as a result of pathogen attack can be successfully monitored using microarray technology, where hundreds to thousands of genes can be assessed simultaneously. Gene expression patterns monitored at this high-throughput level can provide useful information about genes that are co-expressed or co-regulated. Since the early 2000's, microarray analysis has been routinely used for transcriptomic analysis in plants, yet the number of species for which microarray slides is commercially available is still limited. To date, a number of gene expression studies conducted for compatible plant-virus interactions have already been undertaken, primarily monitoring differential gene expression that result from infection by a number of RNA viruses) (Golem and Culver, 2003, Postnikova and Nimchinov, 2012; Whitham et al. 2003; Whitham et al. 2006; Yang et al. 2007). These studies have shown groups of genes whose biological functions are common not only for viruses, but also fungal and bacterial pathogens. The microarray slides used for this study were custom printed from an enriched subset of a cDNA library of bacterial induced genes. Although we are primarily interested in virus-induced changes, we still made use of this bacterial subset for our study as it has been established that cross talk does occur between signalling and defence responses between a diverse number of pathogens (Reviewed in Kunkel and Brooks 2002).

Although *Arabidopsis thaliana* is considered a model plant system, *Nicotiana benthamiana* Domin is the plant system of choice for virus infection (Goodin et al., 2008). It has been described as "universally susceptible" to many plant viruses, which may in part be attributed to the presence of a mutation which occurs naturally in the *Nicotiana* RNA-dependent-RNA polymerase gene (NbRdRP1m) (Goodin et al. 2008). In addition, since *N. benthamiana* serves as a host for multiple virus species, it allows one to comparatively analysis host responses to different viral pathogens.

Since there is limited information regarding global gene expression changes as a result of DNA geminivirus (GV) infections, we hoped to carry out our GV infections this model host.

South African cassava mosaic virus-[ZA:99] is a circular ssDNA geminivirus belonging to the *Geminiviridae* family. Geminiviruses behave like most RNA virus, where in order for successful infection to occur; a few crucial steps are necessary, which include the expression of viral gene proteins, replication of their viral genome and movement from cell-to-cell as well as systemically throughout a host. Many studies have shown that geminivirus infection relies extensively on the recruitment of host proteins which facilitates basic compatibility between virus and host, as well as assists the virus in the diversion of active or induced defence responses.

A number of studies investigating heterologous or cross-species hybridization (CSH) across plant species have provided further support that this type of microarray application is a good alternative to species-specific hybridizations in the absence of a species-specific gene chip. A study conducted by Horvath et al., in 2003 made use of a 11 522 cDNA *Arabidopsis* microarray in order to test the feasibility of a cross-species hybridization approach between *Arabidopsis* and cDNA's obtained from the mature leaves and apices of three distantly related plant species including leafy spurge, wild oat and poplar. Results indicated that the expression of between 23 and 47% genes were detected following hybridization and scanning, therefore indicating successful use of an *Arabidopsis* cDNA microarray for monitoring changes in distantly related non-model plant systems. Furthermore, in the absence of a commercially available *Nicotiana benthamiana* microarray, a successful heterologous array study between enveloped virus-infected *N.benthamiana* leaf tissue and 5000 *S.tuberosum* cDNA clones further demonstrated that a comparative heterologous approach is feasible in plant gene expression studies (Senthil et al. 2005). Bar-Or et al. further highlights the many favourable features of using a cross-species microarray approach.

The number of gene expression studies elucidating changes as a result of geminivirus infection is few (Ascencio-Ibáñez et al. 2008; Trinks et al. 2005). Since CSH is a feasible microarray application that can be used to assess gene expression changes in non-model systems, we hypothesised that an *Arabidopsis* heterologous cDNA microarray, based on a bacterial pathogen-inducible cDNA subset, could successfully be exploited in this study, in the absence of the natural host cassava, in order to detect some transcriptional changes in leaf tissue infected by geminivirus SACMV-[ZA:99] over a 35 day period. Genes identified at four timepoints (7, 14, 21 and 35 days post inoculation) were compared in order to identify changes that occur from the onset of symptoms to full systemic infection. Through this early preliminary study, we aimed to better understand some of the complexities underlying host-geminivirus interactions.

2.3 Materials and Methods

2.3.1 Plant growth and virus inoculations

Nicotiana benthamiana seeds were germinated in jiffy pellets and allowed to grow to the 4-6 leaf (6 weeks) stage in a controlled growth chamber at 25 °C with a photoperiod of 16 h light/8 h dark . Twenty 6 week old plantlets, were co-inoculated with a total of 60 µl of full length head-to-tail dimers of SACMV-[ZA:99] DNA-A and DNA-B mobilized in *Agrobacterium tumefaciens* strain C58C1 (OD₆₀₀ of ±0.8). *Agrobacterium* containing infectious clones of SACMV DNA-A or SACMV-DNA-B (partial dimers cloned into pBIN19) were cultured independently of each other in Luria broth with working concentrations of 50 mg/l rifampicin and 100 mg/l kanamycin. *Agrobacterium* containing the pBIN19 plasmid alone (control) was also cultured. Cultures were incubated at 30°C until an optical density (OD₆₀₀) of ±0.8 was attained. Each culture was pelleted at 8000 x g. Pellets were washed in sterile water in a repeated spin at 8000 x g. Water was removed and pellets were resuspended in 200 µl 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) of SACMV-[ZA:99] inoculum, using a Hamilton syringe. Control plants were inoculated in parallel, and eighteen additional plants were mock-inoculated with 60 µl of *Agrobacterium* containing the plant transformation vector pBIN19.

2.3.2 Sample Collection and DNA Extraction

Twenty plants were randomly divided into two biological replicates, each containing 10 plants. For each time point (7, 14, 21 and 35 dpi), the two newly emerged but fully expanded leaves closest to the apex were harvested from the 10 plants representing each group, and these leaves were later pooled to form each of the biological replicas. This was carried out separately for SACMV-[ZA:99]-infected leaf tissue as well as *Agrobacterium* mock-inoculated leaf-tissue. Although pooling is not a replacement for experimental variation, it served to normalize plant-to-plant variation and variability introduced by differences in infection efficiencies or differences in sample collection procedures. The leaves harvested at each timepoint were immediately frozen in liquid nitrogen and stored at -80°C until required for downstream DNA and RNA extractions.

Total nucleic acid (TNA) was isolated from SACMV-[ZA:99]- infected and mock-inoculated leaves using a modified CTAB-based extraction method (Doyle and Doyle, 1987). Approximately 50 milligrams of 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 hour. 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 10min. Recovered TNA pellets were washed in 70% ice-cold ethanol and later resuspended in 1x 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 by determining the ratio of the absorbance at 260 and 280nm (A260/280) using the NanoDrop ND-100 Spectrophotometer (NanoDrop Technologies, Thermo Scientific, USA).

2.3.3 Conventional PCR used to confirm SACMV-[ZA:99] infection in harvested leaf tissue

SACMV-[ZA:99] infection of leaf tissue was confirmed by conventional PCR. A 550bp fragment of the core coat protein (CCP) on DNA-A was amplified using degenerate forward primer: 5' GCCHATRTAYAGRAAGCCMAGRAT 3' and reverse primer 5' GGRTTDGARGCATGHGTACANGCC 3'. Each PCR reaction mixture contained 0.4 µM of each primer, 200 µM dNTPs, 2 units DreamTaq DNA polymerase (Fermentas, Thermo Scientific, USA), 1x DreamTaq Buffer (Fermentas Thermo Scientific, USA), and nuclease-free water to a final volume of 50 µl. Approximately 500 ng of the genomic DNA template was added to the reaction mixture. Reactions were cycled in a MyCycler™ Thermal Cycler (Bio-Rad, California, USA) at 95 °C for 5 min to activate the Taq DNA polymerase, followed by 30 cycles of denaturation at 95°C for 30 sec, annealing at 55 °C for 30 sec, primer extension at 72 °C for 45 sec, and a final extension step of 72 °C for 5 minutes. As a negative control, genomic DNA extracted from a mock-inoculated leaf tissue was used. A “no template control” or “NTC” was also set up as a second negative control. As a positive control, 50 ng of SACMV-[ZA:99] DNA-A cloned into pBLUESCRIPT was used. Amplification products were examined by electrophoresis on a 1.2% (w/v) agarose TAE (Tris-acetate-EDTA) gel containing 10 µg/ml ethidium bromide

2.3.4 Absolute Quantification Real-Time PCR of SACMV-[ZA:99] DNA-A molecules

In order to quantify levels of SACMV-[ZA:99] in infected leaf tissue, a standard curve was generated on the LightCycler 1.5 using known 1:10 serial dilution concentrations of SACMV-[ZA:99] DNA-A cloned into plasmid pBluescript. The concentrations ranged from 100 ng to 0.1 pg. The standard curve constructed represents a linear relationship between the cycle threshold and the initial concentration of DNA added. The copy number (molecules/ng of TNA) present for each concentration was calculated using Avagadro's constant. The equation incorporating Avagadro's

constant $[Xg/\mu l \text{ DNA}/(\text{plasmid length} + \text{insert length in basepairs} \times 660) \times \text{Avagadro's constant} \times 10^{-12}]$ was used for each 10-fold dilution. This equation assumes that one plasmid (DNA-A in pBluescript) has one copy or molecule of DNA-A therefore if 100 ng plasmid DNA = 0.026pmol plasmid then, using Avagadro's number, 1pg plasmid DNA = 1.56×10^5 molecules. Therefore, the standard curve generated had a range from 1.54×10^8 molecules of DNA-A /ng of TNA to 1.54×10^4 molecules of DNA-A/ng of TNA .

$$\text{Molecules of SACMV-[ZA:99]DNA-A/ng of leaf tissue TNA} = \frac{[\text{calculated concentration (pg)} \times 1.56 \times 10^5 \text{ molecules}]}{\text{Concentration of DNA added to qPCR reaction (ng)}}$$

Viral titer was quantified using the TNA extracted from 7, 14, 21 and 35 dpi leaf tissue as well as 100 pg Standard of DNA-A in pBluescript on the LightCycler (Roche Applied Science). Quantitative PCR was carried out using the LightCycler® FastStart DNA Master^{PLUS} SYBR Green kit (Roche Applied Science). Twenty μl LightCycler PCR master mix was prepared according to manufacturer's instructions to contain, 1 of 10 μM forward primer and 1 μl of 10 μM reverse primer, 2 μl of LightCycler FastStart DNA Master^{PLUS} SYBR Green I and 2 μl of sample DNA and 13 μl sterile water . Primers designed to amplify a 155 bp fragment of the SACMV coat protein (CP) were used in Real-Time PCR reactions. Primer sequences were as follows: Forward primer -5' GCACAAACAAGCGTCGAT 3' and Reverse primer-5' CTGCCAGTATGCTTAACGTCA 3'. Cycling conditions of an initial activation cycle of 95°C for 10min, followed by thirty amplification cycles consisting of denaturation at 95°C for 10 sec, annealing at 60°C for 10 sec and elongation at 72°C for 10 sec. The intensity of fluorescent emission of the SYBR Green dye present in each LightCycler capillary was measured at 520 nm at the end of each elongation phase.

2.3.5 Microarray Study

2.3.5.1 *Arabidopsis thaliana* and *N. benthamiana* test hybridization

One of the pre-requisites for carrying out a heterologous microarray study is to confirm sufficient sequence homology to facilitate hybridization between the target sample and the probe set on the array. To our knowledge, there have been no previous microarray studies carried out between *Arabidopsis thaliana* and *N. benthamiana*. Therefore, a "yellow experiment" which is a proof-of-concept experiment, was conducted in order to test the hybridization efficiency of healthy wild-type *N.benthamiana* leave tissue cDNA against the spotted *Arabidopsis* cDNA probes. Therefore, the assumption made is that although there are sequence mismatches between the target and probe,

the target RNA will still hybridize at a sufficient level to the probes therefore allowing reliable detection of the majority of gene expression changes in a given treatment. In the case of a two-colour cDNA microarray, similar to the one in this study, two independent RNA samples were obtained from 6-week old healthy wild-type *N.benthamiana* leaf tissue and reverse –transcribed into cDNA and subsequently labelled separately with Cy3 and Cy5 fluorescent dyes. Labelled cDNA was then added in equimolar concentrations and hybridised onto the *Arabidopsis* SSH array. Since there is no differential expression expected to occur between the two healthy leaf tissue samples, spots appear yellow after scanning (The full methodology for this test hybridization was carried out in the same way outlined in steps 2.3.8 – 2.3.10.2.)

2.3.6 Array design

The cDNA microarray glass slides used for this study were obtained from the ACGT Microarray Facility based at the University of Pretoria (<http://www.microarray.up.ac.za/index.html>). cDNA clones printed on these slides were obtained from a suppression subtractive hybridization (SSH) library enriched with pathogen-inducible genes in an *Arabidopsis* Col-0 –*Ralstonia Solanacearum* interaction. cDNA clones were spotted onto glass slides using the GenIII microarray spotter system (Amersham Bioscience) in 24 blocks (8 print tip groups).

2.3.7 Microarray experimental design

A direct comparison design (reviewed in Naidoo et al. 2005) was used for this microarray where SACMV-[ZA:99] infection was compared to a mock-inoculated control. Furthermore, the experiment aimed at monitoring temporal gene expression changes over four timepoints (7, 14, 21 and 35 dpi). Two biological replicates were conducted for each timepoint and a technical replicate (dye swap) was included within each biological replicate. A total of 16 slides were used for this study.

2.3.8 RNA extractions and RNA purification

One hundred milligrams of leaf tissue for both infected and healthy *N.benthamiana*, was homogenized in liquid nitrogen. One millilitre TRI® Reagent (Sigma-Aldrich) was added to each sample, incubated for 10 min at room temperature and centrifuged at 13000 x g for 10 min at 4°C. The supernatant was then treated with 200 µl of chloroform, vortexed for 15 sec and incubated at

room temperature for 15 min. The chloroform-TRI reagent mixture was then centrifuged at 13,400 rpm at 4°C for 10 min. The aqueous layer was carefully removed and placed in a new centrifuge tube. RNA was precipitated by adding 500 µl of isopropanol to the aqueous layer and gently inverting the tube. The mixture was incubated at 28°C for 10 min followed by an additional centrifugation step at 13,000 x g at 4°C for 10 min. The RNA pellet was washed with 75% ethanol, centrifuged at 13,000 x g at 4°C for 10 min. The pellet was briefly air dried for 15 min and resuspended in 100 µl RNase-water.

RNA obtained from SACMV-[ZA:99]-infected and mock-inoculated leave tissue samples were purified using the Qiagen RNeasy Plant MiniKit (Qiagen, Valencia, California) by following manufacturer's instructions.

2.3.9 cDNA synthesis and cyanine dye coupling

Probes were prepared for microarray analysis from RNA templates by the incorporation of modified nucleotide, amino allyl dUTP (AA-dUTP) during first strand synthesis by Reverse transcriptase. Two µl of Oligo dT₁₈ primer (0.5 µg/µl), 3 µl of random nonamers (6 µg/µl) and 0.5 µl RNase inhibitor were added to 15 µg of purified leaf tissue RNA. RNA leaf tissue samples were incubated at 70°C for 5 min, immediately cooled on ice and used as a template for first-strand cDNA synthesis. To each RNA leaf tissue sample, 2 µl of 20X aa-dNTP (10 mM dATP, dCTP, dGTP, 4 mM dTTP, 6 mM aa-dUTP), 8 µl 5x strand buffer, 4 µl 0.1M dithiothreitol [DTT], and 2 µl Reverse transcriptase (Superscript III, Invitrogen) was added. The reverse transcription reactions were incubated overnight (approximately 16 hours) at 42°C. After reverse transcription, the sample RNA was hydrolysed by treating the cDNA reaction cocktails with 20 µl of stop solution (1 M NaOH/0.5 M EDTA) and incubating at 65°C for 10 min. Following incubation, 25 µl of 1 M HEPES (pH 7.0) was added to each reaction in order for neutralization to occur. Unincorporated aa-dUTP was removed using the Qiagen RNeasy®MinElute™ Cleanup kit according to manufacturer's instructions with one modification. β-mercaptoethanol was not used in RLT buffer in any of the purification steps.

Purified cDNA was then used to in a dye-coupling procedure involving the chemical labelling of Cy3 and Cy5 NHS-esters to the aa-dUTP synthesized cDNA. This was achieved by concentrating cDNA samples to a total of 5µl. cDNA was re-dissolved in 5 µl of NaHCO₃ buffer (pH 9.0) and incubated for 10 min at 37°C. 5 µl of either Cy3 or Cy5 dyes (Amersham CyDye Post-Labeling Reactive Dye Packs) were added to their respective samples and mixed thoroughly. For the biological replicates, the SACMV-infected samples were labelled with Cy5 and the mock-inoculated controls with Cy3. For the technical replicates, SACMV-infected samples were labelled with Cy3 and the mock-inoculated

controls with Cy5. Labelling reactions were allowed to occur by incubating cDNA samples at room temperature in the dark for 2 hrs. 35 μ l NaOAc (pH 5.2) was added to samples after incubation. This was followed by the removal of unincorporated dye using the Qiagen PCR purification kit according to manufacturer's instructions (Qiagen, Valencia, California).

The total amount of dye incorporation (measured in picomoles of dye per cDNA probe) and the ratio of unlabeled to fluorescent labelled nucleotide in the probe were assessed by measuring probe absorbance on the NanodropTM spectrophotometer at 260, 550, and 650 nm to assess cDNA, Cy3, and Cy5 concentration, respectively. A total of 300 pmoles of incorporated dye (i.e. 150 pmoles / labelled sample) were used for downstream hybridizations.

2.3.10 Hybridizations

2.3.10.1 Pre- hybridization slide treatment

Before hybridization, microarray slides were pre-treated in a solution of 1% BSA, 0.2% SDS and 3.5 $\times$ SSC at 60 °C for 20 min. Slides were rinsed several times in sterile distilled water (Sigma) and dried using a slide spinner at 1000g for 2 min.

2.3.10.2 Cy3-Cy5 hybridizations and washes

Hybridization solution, consisting of 50% formamide, 25% Amersham hybridisation buffer (GE Health-care Ltd.), 25% sterile distilled water and the Cy3 and Cy5 labelled targets were used for downstream hybridization to Arabidopsis SSH slides. Labelled samples were heated to 95 °C and then cooled on ice. The hybridisation solution was added to the slides under a clean coverslip (24mm $\times$ 60mm) and then allowed to hybridise in a HybUP hybridisation chamber (NB Engineering, Pretoria, South Africa) overnight in a waterbath set at 42 °C. Following overnight hybridization, the slides were washed in a number of stringency washes in order to remove all excess probe solution.

Following overnight hybridization, slides were washed two times (separately) in a wash solution containing 1.0 $\times$ SSC, 0.2% SDS for 4min at 42 °C. The first wash was done for 1 minute in order to remove the cover slip and the second wash was carried out in fresh buffer for 4min. This was followed by two washes in 0.1 $\times$ SSC, 0.2% SDS for 4 min at 42 °C, and three washes in 0.1 $\times$ SSC for 1min at room temperature. The slides were dipped in distilled water a 5 times before being dried in slide spinner , and scanned using the GenepixTM 4000B scanner (Axon Instruments, Foster City, CA, USA).

2.3.11 Image acquisition and data analysis

In order to normalize raw fluorescence intensities, the web-based data-analysis microarray package CARMAweb (Comprehensive R-based Microarray Analysis web service) was used. CARMAweb (<https://carmaweb.genome.tugraz.at/carma/>) is a web-based interface to the 'limma' package of Bioconductor (Rainer et al. 2006). For statistical analysis, moderated t statistics (Bioconductor "limma" package) were used and p-values calculated using Benjamini-Hochberg correction for multiple testing. Before uploading .gpr files into CARMAweb for normalization, the GenePix Array List file (.gal) containing the spot coordinates and clone ID names for the SSH array was uploaded into Genepix Pro 6.0 software and the foreground and background intensities of all the cDNA probe spots on the microarray were calculated. Also, all poorly hybridized spots and those that were saturated and had raw fluorescence intensities >65000 in either one or both channels were flagged as "bad" spots before subsequent analysis in CARMAweb.

The settings of 'minimum', 'robust spline', and 'quantile' were used for background correction, within-array normalization, and between-array normalization, respectively. Between-array normalization was done using dye-swap hybridizations.

2.3.12. Independent expression validation using semi-quantitative Real-time PCR (RT-qPCR)

RT-qPCR was performed on selected genes (N = 7) in order to validate data obtained from this microarray study. Genes selected for validation were chosen based on whether they were detected in at least 2 or more timepoints during SACMV-[ZA:99]-infection, and were of interest for discussion in terms of possible involvement in SACMV-[ZA:99] systemic infection. Primers used for RT-qPCR were designed using the real-time PCR primer design tool available through Integrated DNA technologies, Inc (<http://www.idtdna.com/Primerquest/Home/Index>). Primers were designed for: Heat shock protein 70 (HSP70), Heat shock protein 20.2 (HSP 20.2), tubulin, actin, Vesicle-associated movement protein (VAMP), Lipxygenase 2 (LOX2) and Nitric oxide synthase (NOS1) (Table 2.1). Furthermore, primers were designed for 4 Housekeeping genes (Elongation factor 1 (EF1), 18S, cyclophilin and Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) in order to test which was most stably expressed across all 4 timepoints for SACMV-[ZA:99] infected *N. benthamiana* as well as in mock-inoculated leaf tissue samples. From these, the most stably expressed housekeeping endogenous control was used for downstream qPCR calculations.

Table 2.1 Primer sequences designed and used for the independent validation of microarray results.

Gene Name	Accession number	Forward primer (5'-3')	Reverse primer (5'-3')
HSP70	ABW76421.1	AGCAGCTGAAGGAACTTGGAGACA	AGCCATCGCATCCTTCATGGTTTG
HSP20.2	BQ511516	TGACTCCACCCTTGGCATCTTTGA	TGTGTTTACCGGCACCTCCTCTGT
Tubulin	ABY54099.1	GCATCTTTGCGTACACTTTGCT	ACATAAGCCCCAAAAGTAGCTGGA
Actin	X55749	GACAATGGAACAGGAATGGTTAAG	CTCAAGAGTCCGAGCAACATAG
VAMP	CN747483.1	ACTGTCTCAACCTGTTGCCGGG	GGCTGCAGCAGAAGCTGACAAAGAA
LOX2	AY254348.1	TGCAGAACTGGTCGCCCAAGG	TGCAGAACTGGTCGCCCAAGG
EF1	X53043	TCAGGCTGACTGTGCTGTTCTCAT	GACACCAAGGGTGAAAGCAAGCAA
18S	AY552528	GGCAAATAGGAGCCAATGAA	GGCAAATAGGAGCCAATGAA
Cyclophilin	AY972080.1	GCGCCAAATTCAAGGACGAGAACT	ACAGCCTCGGCCTTCTTAATCACA
GAPDH	U97257.1	ATGGCCTTCAGAGTACCAACTGCT	GCTTGACCTGCTGTACCAACAAA
NOS1	AB303300.1	GCGTTGCAACTTCATATGGTGCCT	TTCCTGTGCGCGGTGTCAATAGAA

In brief, 1 µg of DNase-treated total RNA was reverse transcribed using the Improm-II-reverse transcriptase kit (Promega, Madison, WI) according to manufacturer's instruction. RNA, dNTPs and Oligo dT₁₈ primer were denatured for 10 min at 70°C; then kept at 25 °C for 5 min before the reverse transcription master mix was added. Reverse transcription was performed at 42°C for 1 h followed by a 10 min incubation step at 70°C. Control reactions were set up without the addition of reverse transcriptase and used as negative controls in the real-time PCR study. RT-qPCR experiments were conducted on the Lightcycler 1.5 for all genes using the appropriate primer pair for each reaction (Table 1). Relative quantification standard curve method (Larionov et al. 2005) was chosen for each of the seven genes. Standard curves were generated for each gene using a 10-fold serial dilution of cDNA reverse transcribed from RNA extracted from healthy *N.benthamiana* leaf tissue. All reactions were based on the following recommended protocol using 0.5µl of each primer and 1µl of template per reaction. In brief, all qPCR reactions were performed in LightCycler® capillaries using the LightCycler 1.5 using LightCycler® FastStart DNA Master^{Plus} SYBR Green I kit (Roche). Three biological replicates and two technical replicate were run for SACMV-[ZA:99]-

infected and mock-inoculated leave cDNA samples. One μ l of undiluted cDNA was used for each reaction. The cycling conditions used were as follows: initial denaturation for 10 min at 95°C (hot start) followed by an amplification and quantification cycle repeated 35 times, each consisting of 10 sec denaturing at 95°C, 10 sec annealing at primer specific temperatures, 15 sec primer extension at 72°C with a single fluorescence measurement. Melting curve cycle was obtained by heating to 65°C for 15 s with a heating rate of 0.1°C per second with a continuous fluorescence measurement. GAPDH was the gene found to be most stably expressed across all qPCR runs and was using as an endogenous control for normalization. Statistical analysis was carried out in Microsoft Excel.

2.4 Results

Gene expression changes in SACMV-infected *N.benthamiana* leave tissue for two biological replicates were found to be statistically similar where correlations coefficients between biological replicates at each time point ranged between 0.69 and 0.78 and these are the results which are presented in this study. In order to increase the statistical power between our microarray data sets, we conducted a third biological replicate for infected leave tissue at 7,14, 21 and 35dpi. These hybridizations, however, were not completely successful in that, when we used the same normalization calculations for the third biological replicate (after hybridization and scanning) as we did for biological replicate one and two, we found that the genes were not significantly expressed at a 1.5-fold cut-off ($p < 0.05$). Furthermore, the initial raw fluorescence signal values observed post hybridization, prior to normalization was found to be lower than that observed for the original two biological replicates conducted. Furthermore, this resulted in a reduction of the overall ratios of Cy3 to Cy5 after background correction and normalization for this third biological replicate compared bio-replicate 1 and 2. This could be attributed to various factors. Firstly, re-amplification of SSH cDNA clones from freezer stocks proved to be problematic for the technical team at the ACGT microarray facility, so this could have affected downstream hybridization of *N.benthamiana* cDNA. Also, the hybridization and washing steps were all carried out manually which can introduce variation as a result of the user. Furthermore, since this replicate was carried out at a completely different time to the first two replicates, it is inherently subjected to different variation factors at different step of the study from RNA extraction to hybridization to scanning and data analysis. This is primarily why replicates should ideally be conducted at the same time and on the slides printed from the same batches. Since our study was the first to carry out an *Arabidopsis*-*N.benthamiana* cross-species array, we could have tested various optimization procedures such as hybridization temperature, washing steps or customized normalization procedures on an initial trial set of

microarray slides in order to optimize the best combination of methodologies for our study before carrying out our experimental SACMV-[ZA:99] microarray hybridizations. In terms of normalization, using the command-line based version of limma as opposed to the web graphical interface version, (GUI), allows more flexibility with normalization algorithms and statistical testing. This is also a factor to consider as it has been shown that “over normalizing” data can lead to differences in data outputs.

2.4.1 Infectivity study

2.4.1.1 Symptom development and viral titre

Infection progress and onset of symptoms was monitored in SACMV-[ZA:99]-infected leaf tissue. Symptoms characteristic of SACMV-[ZA:99] infection, i.e. mosaic, chlorosis, leaf distortion and stunting, were not evident at early infection at 7dpi. Plants became visibly infected when newly emerged leaves at 14 dpi (Fig 2.1 A) exhibited slight leaf margin distortion. By 21 dpi, *Nicotiana benthamiana* leaves were clearly symptomatic with the appearance of mild mosaic, leaf distortion and a reduction in leaf size (Fig 2.1 B). Thirty-five dpi represented fully symptomatic leaf tissue characterised by the appearance of mosaic, chlorosis, and overall plant stunting compared to mock-inoculated controls (Fig 2.1 C and D). Conventional PCR using primers targeting a 550bp of the CCP of SACMV-[ZA:99] was found to be positive for leaf tissue harvested from twenty infected plants at 14 dpi. Eighty-five percent (17/20) of plants were positive for SACMV-[ZA:99] infection (Fig 2.2 A and B) where two had weak amplification (plants 12 and 13). PCR was carried out again (at 21dpi) on plants 8, 12, 13, 15 and 18 which showed weak or no amplification at 14dpi and was confirmed for infection (Fig 2.2 C). The appearance of symptoms was directly correlated with an increase of viral titre over a 35 day period. We compared the replication levels of SACMV-[ZA:99] in leaf tissue samples from each timepoint post inoculation, using absolute quantitative PCR using known concentration of DNA-A from a standard curve (Fig 2.3 A), in order to monitor temporal changes in virus titre during the course of SACMV-[ZA:99] infection. SACMV-[ZA:99] DNA-A was not detected in any infected plants at early infection at 7 dpi. Viral titre increased rapidly by 17.2-fold between 14dpi (2.83×10^1 SACMV-[ZA:99]-DNA-A/ng of TNA) and 21dpi (1.89×10^1 SACMV-[ZA:99]-DNA-A/ng of TNA), but was significantly lower with a 3.6 fold increase between 21 and 35 dpi (1.79×10^3 SACMV-[ZA:99]-DNA-A/ng of TNA). This pattern of virus titer increase with symptom severity increase over time is typical of most plant virus infections. For example, the same pattern was observed in cassava plants infected with a potyvirus, *Cassava brown streak virus* (CBSV), where the presence of symptoms on leaves correlated with an increase in CBSV RNA titers (Moreno *et al.*, 2011).

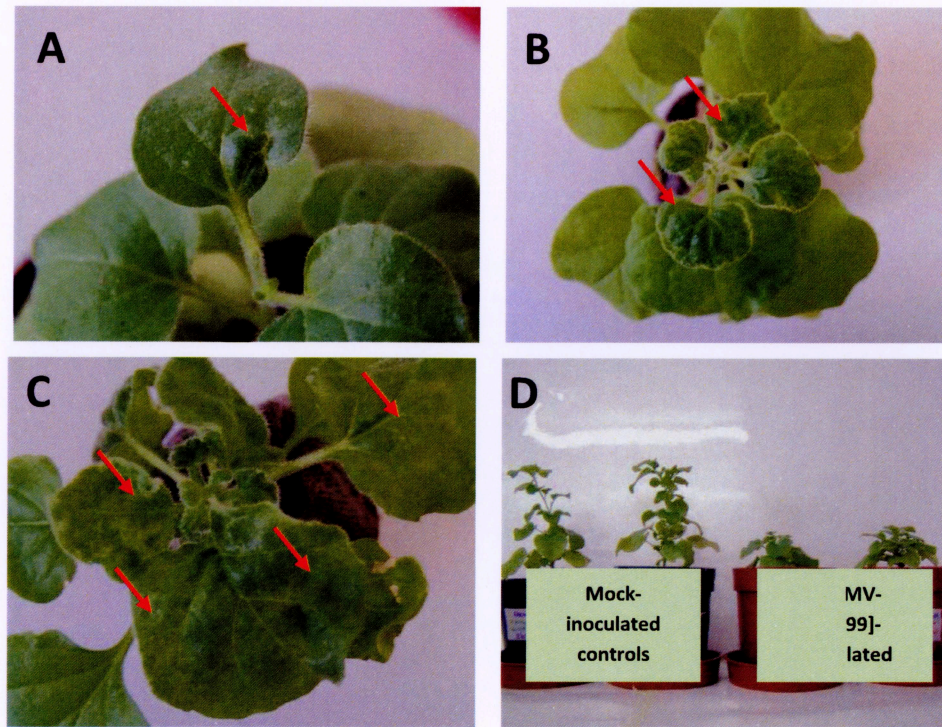


Figure 2.1 Symptom development of SACMV-[ZA:99] infected leaf tissue of a 35 day period. (A) A newly emerged leaf shows slight leaf distortion at 14dpi. (B) Development of leaf curling and a reduction in leaf size was observed at 21dpi. (C) Leaves are fully symptomatic at 35dpi with the appearance of mosaic, chlorosis and leaf distortion. (D) Overall stunting of SACMV-[ZA:99] infected plants are observed at 35dpi as compared to mock-inoculated control plants.

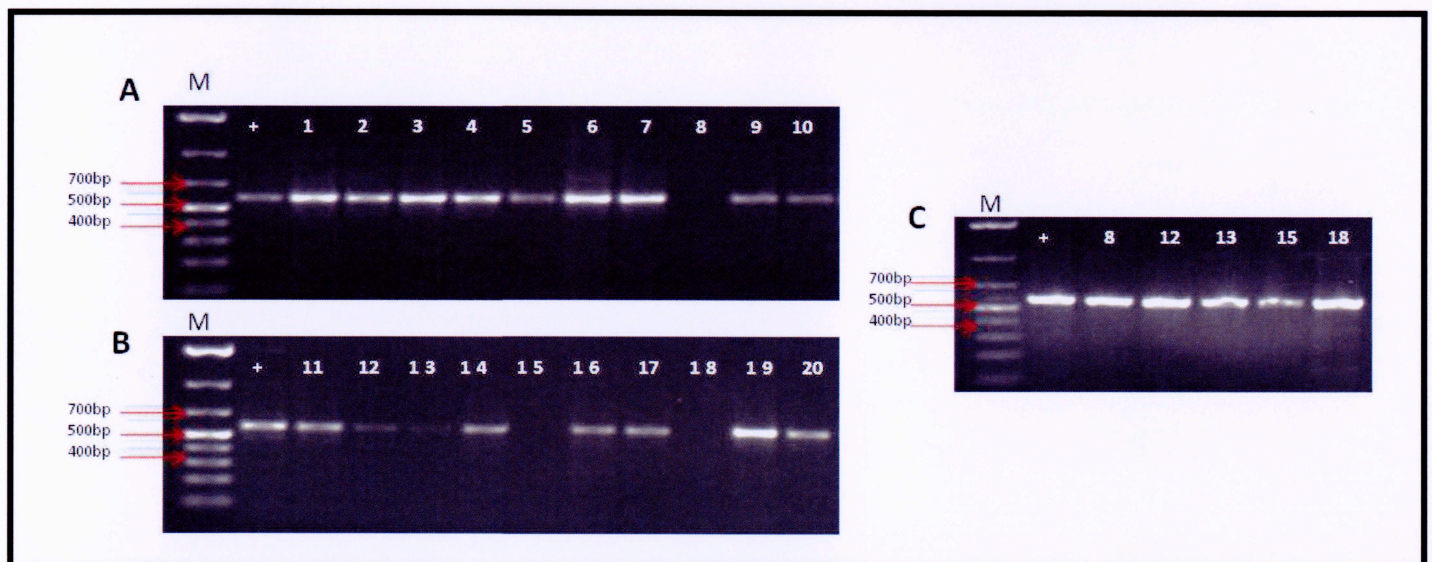


Figure 2.2 Conventional PCR carried out on twenty SACMV-[ZA:99] infected plants at 14dpi (**A and B**) confirming infection in 17/20 plants. PCR was once again carried out in plants 8, 12, 13, 15 and 18 which showed no or weak amplification at 14dpi. Infection was confirmed in these 5 plants at 21dpi (**C**). A 550bp fragment of the coat protein of SACMV-[ZA:99] was targeted using degenerate primers. M = Thermoscientific O'GeneRuler 1kb plus, "+" = positive PCR control (DNA-A cloned into pBluescript plasmid).

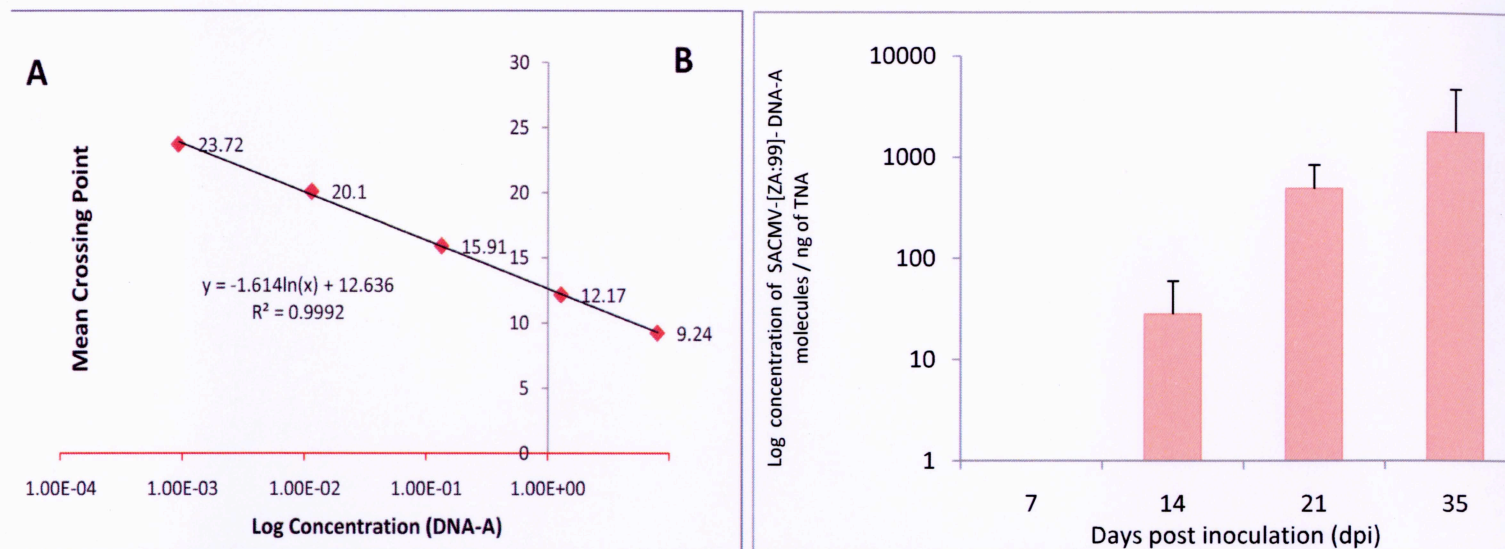


Figure 2.3 (A) Standard Curve (SC) set up using Roche FastStart Master^{PLUS} SYBR green^I and primers designed to amplify a 155bp region of the coat protein of DNA-A. The SC was generated using the mean crossing points vs. Log concentration of DNA-A (B) Viral titre calculated in molecules of SACMV-[ZA:99]-DNA-A/ ng of Total nucleic acid (TNA) at 7, 14, 21 and 35dpi. SACMV-[ZA:99] DNA-A increases over the progression of infection. SC was used to estimate the concentration of SACMV-[ZA:99] in qPCR.

2.4.2 Microarray Study

2.4.2.1 *Arabidopsis thaliana* and *N. benthamiana* test hybridization

Following hybridization and scanning, results indicated that there was successfully hybridization between target and probe where more than 70% of cDNAs (3659 spots of 5000) (Fig 2.3) showed good hybridization signal for probe-target matches. Furthermore, duplicate spots printed on SSH slides were consistent in hybridization signals.

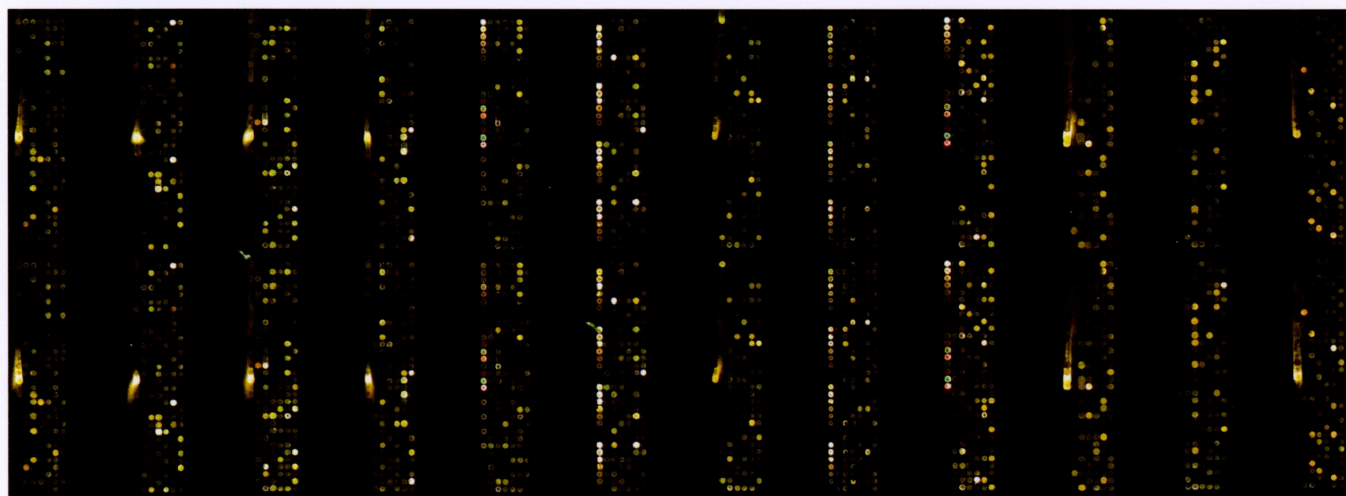


Figure 2.4 Microarray scan for “yellow” experiment showing successful hybridization between *Arabidopsis thaliana* cDNA probes and *N. benthamiana* targets.

2.4.2.2 Normalization and Identification of differentially expressed genes

Normalization algorithms used for this study successfully centred the intensities values for green and red signal around zero which can be observed in MA plots (Fig 2.5 A-D). This indicated that intensity and spatially-dependent biases were accounted for and corrected accordingly. Following normalization, cut-off criteria of 1.5-fold change at $p < 0.05$ was used to filter data. Since this was a heterologous array, we expected hybridization signals to be slightly lower than those expected from self-self-hybridizations. Therefore, in order to detect even the smallest changes in gene expression, we used a 1.5 fold cut-off as opposed to the standard 2-fold cut-off as observed in most gene expression studies. Using this cut-off, 163 genes were found to be differentially expressed but this number was later reduced to 73 genes (discussed in section 2.4.2.3). Of the 75 genes, 15 genes were found to be differentially expressed at 7 dpi, 32 genes at 14 dpi, 50 genes at 21dpi and 44 genes at 35 dpi. One of the most noticeable observations in this study was an over representation of downregulation across timepoints. For 7 dpi, 9/15 (60%) genes were found to be downregulated, for 14 dpi, 22/33 (69%) genes, for 21dpi, 38/50 (76%) genes and for 35 dpi 38/45 (86%) genes. Furthermore, not all genes were consistently expressed across all four timepoints, where only 2 genes (cytochrome P450 and HSP70-2D) were detected across all four timepoints. For upregulation of genes, the highest numbers of genes were detected at late infection at 21 dpi (12 of 38 genes)

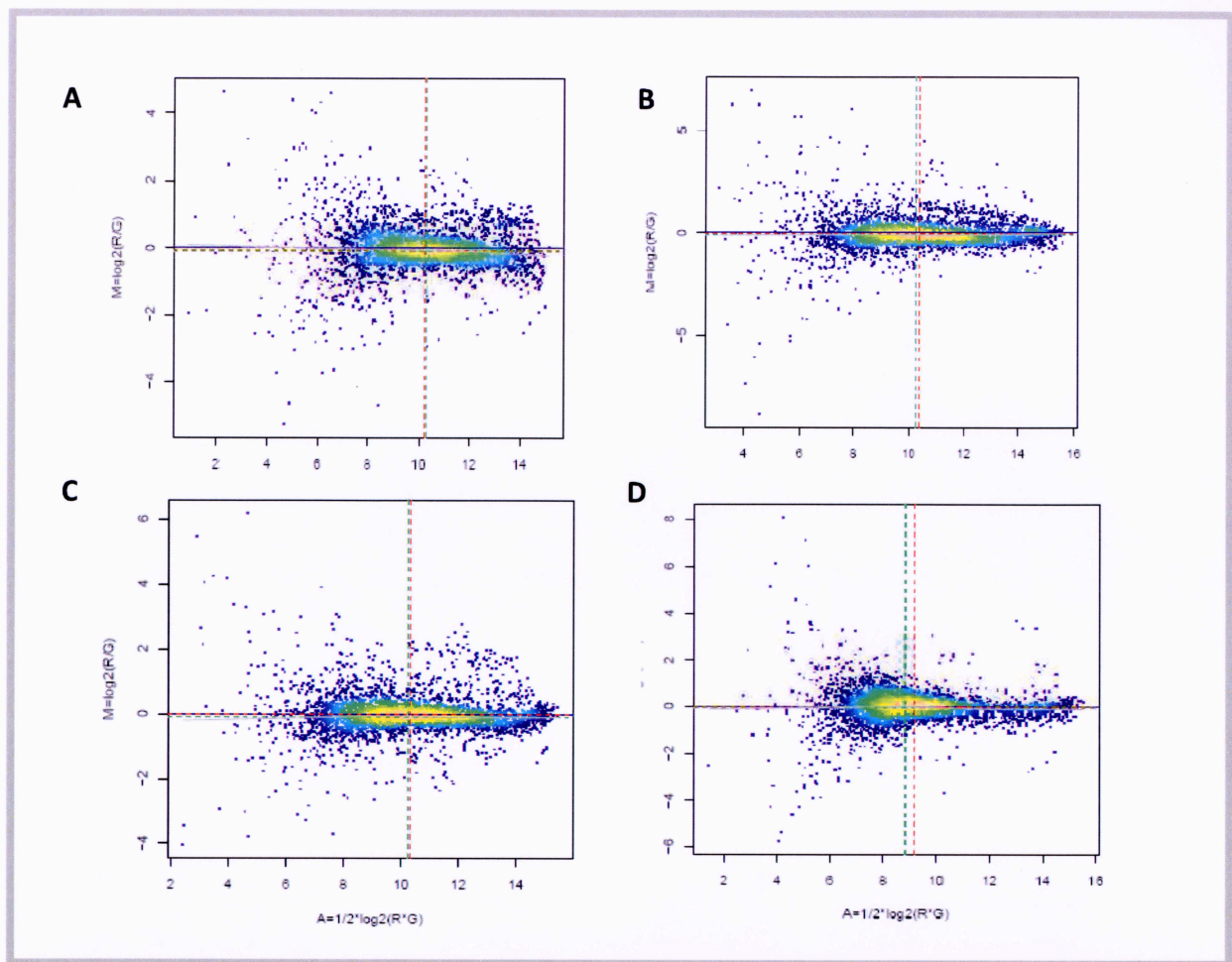


Figure 2.5 MA plots for normalized microarray data shows centering around zero for 7 dpi (A), 14 dpi (B), 21 dpi (C) and 35dpi (D) . MA plots only shown for biological replicate 1 for each timepoint.

2.4.2.3 Sequencing and annotation of *Arabidopsis* Clones

Of the 5000 cDNA clones printed on the *Arabidopsis* SSH slide used for hybridizations, only 278 cDNA had been sequenced and annotated. Therefore, of the total 163 genes (across all timepoints) identified as differentially expressed in this study, none had assigned annotation. For this reason, we harvested bacterial clones (obtained from the Microarray Facility at University of Pretoria) and carried out plasmid isolation of 163 plasmid clones containing the *Arabidopsis* gene sequences representing the genes that were identified as differentially expressed. Plasmid clones were sequenced by Inqaba biotech (Sunnyside, Pretoria, South Africa) using M13 primers and standard Sanger sequencing. All sequences obtained were edited in Finch TV and BLASTed against the *Arabidopsis* genome in TAIR. Accessions and annotations were assigned to each clone. Furthermore, of the 163 clones sequenced, several were eliminated from downstream analysis for various reasons. Twelve genes could not be successfully sequenced by Inqaba biotech, eleven genes had annotations with unknown function and as many as sixty-seven genes were found to have aligned to duplicate *Arabidopsis* accessions therefore these were characterized as redundant clones. A total of

71 (across all timepoints) were consequently verified, and were considered SACMV-[ZA:99] responsive genes (Supplementary Data S2.1).

2.4.3 Microarray Validations using Semi-quantitative Real-time PCR (RT-qPCR)

Semi-quantitative Real-time PCR validations were conducted on SACMV-infected and mock-inoculated cDNA samples from timepoints 14, 21 and 35dpi. No validations were carried out on leaf tissue cDNA samples at 7dpi. The expression patterns observed for HSP70, VAMP, actin, tubulin and LOX were consistent with the expression patterns observed for the microarray study where all the genes were shown to be downregulated at the selected timepoints when normalized against GAPDH. However, the expression ratios for AtNOS1 and HSP20.2 were not consistent with those obtained from microarray results. Results obtained from the microarray data suggested that AtNOS1 (AT3G47450) was upregulated at 21 and 35dpi. However, our qPCR data indicated this gene was downregulated at both timepoints. Also, the primers used for the amplification of HSP20.2 were taken from Nicot et al. (2005). Amplification using these primers was unsuccessful. We then downloaded the FASTA sequence using the HSP20.2 accession (BQ511516) provided in Nicot et al. and designed our own primers, but amplification was still unsuccessful even under different MgCl₂ concentrations and under different amplifications temperatures. Based on the fact that two of the seven genes chosen for validation did not work consistently, it would have been beneficial to choose more genes for microarray validations.

Furthermore, the expression levels (expressed as Log₂ fold change) of the genes successfully validated, while consistent, appeared higher by qPCR compared with microarray hybridization Log₂ fold change ratios. Based on observations made from many microarray studies, this is a phenomenon that frequently occurs and can be attributed to several factors, one main factor being the differential sensitivity between the two techniques used. Another factor is likely to be that microarray measures the ratio in gene expression where “absolute amounts” of gene transcripts of target genes is being measured in qPCR experiments which is later expressed relative to the “absolute” amount an endogenous gene. The most critical factor when doing microarray validation is to make sure that the pattern of expression observed for genes validated is consistent with those of the microarray experiment regardless of the fold-change differences between the two techniques.

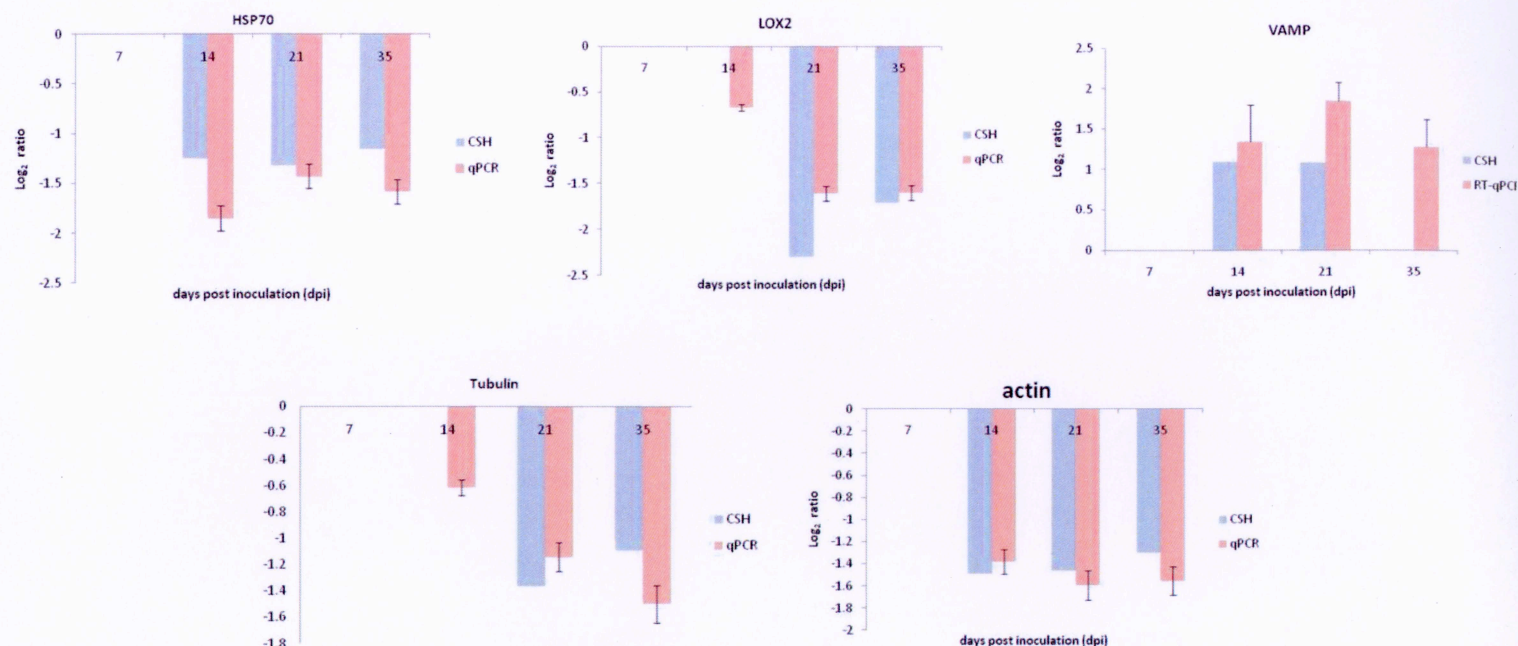


Figure 2.6 Log₂ fold-change RT-qPCR (pink bars) and Log₂ fold change of microarray data (blue bars) for genes LOX, VAMP, HSP70, tubulin and actin.

2.5 Discussion

This study reports, for the first time, a comparative approach using a heterologous hybridization between an *Arabidopsis* customized microarray and *N. benthamiana* leaf tissue cDNA to interrogate global gene expression changes induced in *N. benthamiana* by a ssDNA geminivirus over a 35 dpi period. Despite differences in the probe target sets, valuable information on altered expression of a cohort of common and unique genes belonging to functional gene categories putatively involved in geminivirus movement and replication, general biotic stress responses, photosynthesis and cytoskeleton organization was obtained. Furthermore, since the slides used for this study were created from an SSH library, i.e. enriched for pathogen responsive genes to bacteria, successful hybridization results suggest that expression of many biotic or defence-related genes in susceptible host plants might share components of signalling pathways involved in virus host-pathogen interactions.

Supporting evidence by a number of studies have already shown that geminiviruses have the ability to affect a number of host cell pathways in order to establish successful infection (Andleeb et al., 2010, Ascencio-Ibanez et al. 2008, Hanley-Bowdoin et al. 2004; Morilla et al. 2006). A previously

conducted CSH study has been successful in revealing differential changes in response to virus infection in leaf tissue in the absence of homologous gene probe microarrays (Senthil et al. 2005). In the Senthil study, cDNA microarrays derived from *Solanum tuberosum* expressed sequence tags were used to monitor the gene expression changes of distantly related enveloped viruses, *Sonchus yellow net virus* (SYNV) and *Impatiens necrotic spot virus* (INSV), elicited in systemically infected *Nicotiana benthamiana* leaves. Results obtained showed that this *Solanum-Nicotiana* heterologous microarray study was successful and that although gene expression fold-changes were lower in SYNV-infected leaf tissue as compared to INSV-infected leaf tissue, overall the number of genes detected increased over the course of virus infection (measured across three timepoints). In another example, Horvarth et al. (2003) reported the detection of 23-47% of genes in mature leaf tissue and shoot apices of wild oat, leafy spurge and poplar by utilizing an *Arabidopsis* cDNA microarray.

In this study, by using cut-off criteria (≥ 1.5 or ≤ -1.5 -fold at $p < 0.05$), 163 differentially expressed genes were detected in the two biological replicate experiments. Following sequencing and annotation of these genes, more than half i.e. 90, were eliminated from downstream analysis, while 73 common transcripts from across the two biological replicates were identified to be up- or down-regulated in SACMV-challenged *N. benthamiana* from 5000 enriched pathogen-inducible cDNAs on the *Arabidopsis* SSH array during the 35 dpi. The highest number of genes (12) were up-regulated at 21dpi, while the highest number were down-regulated between 21 and 35dpi (Supplementary Data S2.1). Only 21% (15/73) and 44% (32/73) of selected genes were SACMV-responsive at 7 dpi (plants symptomless; low virus titer) and 14 dpi (symptoms appear and increased virus titers), respectively. From these results, it would appear that there is a correlation between repression of gene transcription, increasing symptom severity and increasing levels of virus titre in the host cells. A similar pattern was reported in a study of *Tobacco mosaic virus* (TMV)-inoculated *A. thaliana* ecotype Shahdara, where a low average (3 biological replicates) of 51 responsive genes from 8-10,000 *Arabidopsis* cDNAs was reported at 4 dpi (low virus accumulation and symptomless), compared with 129 genes at a later (14 dpi) timepoint, when systemic symptoms and significant TMV accumulation occurred (Golem and Culver, 2003). This pattern of general gene alteration is not uncommon in general. In a comparison of 11 virus-plant host interactions in *Arabidopsis*, late stages of pathogenesis were characterized by repression of a large number of genes, including those involved in the stress-responsive pathways (Postnikova and Nemchinov, 2012). Furthermore, it was also noted in this comparative study, that many of the altered genes were not the same at different stages of virus infection.

2.5.1. Genes involved in photosynthesis

The 1.5-fold cut-off used for expression analysis, as well as cluster analysis, revealed that photosynthesis was one of the most significantly altered pathways in term of gene expression. Ribulose biphosphate carboxylase small chain 1B (AT5G38430), ribulose / RuBisCO small subunit 3B (RBCS-3B, AT5G38410), ribulose biphosphate carboxylase small chain 1A (AT1G67090) and LHCB5, Light Harvesting Complex of Photosystem II (AT3G5489 and AT4G10340), were found to be down-regulated at 21 and 35 dpi. The combination of down-regulation in RuBisCo and the altering in chloroplast structure, suggests that SACMV-[ZA:99] may be significantly affecting the photosynthesis apparatus of these *N. benthamiana* leaf cells. This is consistent with the appearance of yellow mosaic and chlorosis observed at these timepoints. Furthermore, these observations are in line with those reported in other virus-host interactions (Hodgson et al. 1989; Pompe-Novak et al. 2006).

2.5.2 Defence-related genes

Another gene of interest, lipoxygenase (LOX2) enzyme was found to be repressed at 21 (Log_2 -2.30) and 35 dpi (Log_2 = -1.17). LOX2 (AT3G45140) is a dioxygenase that catalyses the hydroperoxidation of polyunsaturated fatty acids (Sanz et al. 1992). Induction of LOX activity has been consistently considered as a pathogen-induced defence response, such as in potato infected by a fungal pathogen (Kolomiets et al. 2000). In virus-plant interactions, Aranda et al. (1996) showed that the level of LOX1 transcript was depleted significantly in pea embryo cells undergoing pea seed-borne mosaic virus replication, and recovered at a shorter distance behind the infection front when virus replication declined. Based on the above study, LOX is probably associated with virus infection. However, past microarray studies, for example on soybean, have shown that LOX transcripts change dramatically under diverse abiotic stresses as well by infection with oomycetes (Moy et al. 2004), and therefore, the expression changes of this gene family is likely a general response to biotic stress.

In addition to repression of LOX2, several other host defence-related genes were downregulated in *N. benthamiana*. Notably, HSP20 (AT4G27670), HSP70-2D (AT5G02490) and HSP70-2B (AT3G12580) were repressed at 14, 21 and 35 dpi. The downregulation of several heat shock gene families have been shown to be involved in a number of abiotic and biotic stress responses in plants (Wang et al, 2004). While upregulation of members of heat shock protein families are often induced early in pathogen infection (Babu et al. 2008, Whitham et al. 2003) due to a basal stress or defence response, repression in later symptomatic stages is usually associated with specific virus-directed changes in order to support a cellular environment conducive for replication and movement. Interestingly, an ATP-binding cassette transporter (AT1G59870), involved with systemic acquired

resistance (SAR) in resistant or incompatible reactions as well as a disease R family gene containing LRR domains (AT5G23400), were both suppressed at 21 and 35 dpi. Collectively, down regulation observed for these genes suggest that there is a general downregulation of host defence responses in model host, *N. benthamiana* during systemic infection stages which are associated with both cellular processes redirected by viruses for their demands, and defensive reactions to pathogenesis.

2.5.3 Genes associated with the cytoskeleton and plasmodesmata arrangement

The other interesting change observed in this study was the significant repression of cytoskeleton-associated genes. There have been several different structural proteins and cytoskeleton-related genes identified that are known to be major determinants of cell morphogenesis in plants (reviewed in Mathur and Hulskamp, 2002). In SACMV-[ZA:99]-infected leaves, 4 genes associated with the cytoskeleton, were downregulated at 14, 21 and 35dpi. These included tubulin (TUA6, AT4G14960) and three actin (AT35G09810, AT3G46010 and AT2G37620) encoding genes. Both actin microfilaments and tubulin microtubules play an important role in determining cell shape and growth (Smith and Oppenheimer 2005). This is an interesting observation because previous studies have shown that proteins such as actin, tubulin and heat shock factors remain unaffected during *Pea seedborne mosaic virus* infection in pea seeds (Aranda et al. 1998; Escaler et al. 2000). More interestingly, was the upregulation of two RAB-GTPases (AT2G21880 and AT4G20360) and vesicle-associated movement protein (VAMP714). Unlike other GTPase (e.g. Rac-GTPases, Tao et al. 2002) which are associated with oxidative stress in plants in response to pathogen invasion and hormone responses, the Rab-GTPase family has been intimately linked to the intercellular cell-to-cell trafficking of virus in plants (Takai et al. 2001). This trafficking of virus is mediated by the interaction of Rabs with viral movement proteins and/ the plasmodesmata (PD). They are known to participate in all aspects of intracellular vesicle trafficking, including vesicle budding, targeting, docking, and fusion (Takai et al., 2001). It could be speculated that the two Rab-GTPase found to be upregulated in this system could be hijacked by the movement proteins of SACMV-[ZA:99] to facilitate movement of their movement to the PD. Similarly, VAMPs are involved in similar protein docking function as observed with the Rabs. VAMPs are required for subcellular sorting and vesicle targeting of proteins that are destined delivery and movement through the PD (Uemura et al. 2004 and 2005). Therefore, SACMV-[ZA:99] may be associated with these as an alternative to Rab-GTPases. Therefore, taken together, the downregulation of actin and tubulin could be play a role a synergistic role with Rab- GTPase or VAMP in this movement process as downregulation could lead to remodelling of the cytoskeleton which may become more flexible to facilitate SACMV-[ZA:99] movement. Alternatively, downregulation of cytoskeleton-related transcripts like actin and tubulin

could have dual function this SACMV infection and may be associated with a reduction in cell wall/plasmembrane cross-linking that contributes to disease symptom development.

2.5.4 Limitations of CSH and improvement strategies

Although useful as a preliminary study, and also performed in light of the absence of cassava genomic data, the cross-species approach was met with some limitations. The first of which was the limitation of cDNA probes printed on the slide. Although cross-talk between defence responses from different pathogens have been reported, one can safely assume that not all genes activated or suppressed during the *N.benthamiana*-SACMV interaction were present on the slide and therefore could not be detected. Furthermore, a whole *Arabidopsis* microarray CSH would have preferable, allowing the detection of a higher number of differentially expressed transcripts, but was not possible due to limited funds. This was the first time this type of study was conducted by our group in 2007/2008, and required technology capacity development and training in our laboratory. Additionally, due to the availability of this enriched pathogen-responsive array along with associated expertise and usage of this array by several students in the Microarray Facility at the University of Pretoria (<http://www.microarray.up.ac.za/>), it was a sensible array to begin our geminivirus-host interaction research. A CSH study conducted by Senthil et al. (2005) showed high correlations efficiencies (R^2 between 0.92 and 0.98) in healthy control samples at each timepoint as well as correlation efficiencies between dye swap experiments which also encouraged us to pursue this course of interrogation in *N. benthamiana*. The correlations obtained from our study were much lower though. We did not test the correlation efficiencies between our healthy control samples, but we did test the correlation between SACMV-infected biological replicates as well as between the technical dye swap experiments. Correlation efficiencies between biological replicates at 7, 14, 21 and 35 dpi ranged between 0.69 and 0.78 indicating some variation between hybridizations. To our knowledge, there has been no value suggested as the “golden standard” for correlation coefficients. The values from our study may be an indication of high variability due to sub-optimal hybridization efficiencies or the use of incorrect normalization calculation resulting in possible over normalization of data (as discussed in section 2.4).

2.6 Conclusions

In conclusion, this SSH heterologous microarray study served as an adequate starting point for the identification of genes identified in *N. benthamiana* leaf tissue infected with SACMV-[ZA:99] ,

particularly in the absence of a species-specific *N. benthamiana* array. The most noticeable observation made from this study was the increase in the number of genes found to be differentially expressed as SACMV-[ZA:99] infection progressed over 35 days. Also, it was observed that there was an overall overrepresentation of downregulation/repression of genes across timepoints 7, 14, 21 and 35dpi. Given that the genes primarily repressed were identified to be involved in general defence response, cytoskeleton organization and photosynthesis, this downregulation may outwardly be due to the virus overtaking host defence-related pathways and causing physiological changes associated with the disease symptoms. This ultimately results in successful SACMV-[ZA:99] infection in *N. benthamiana* leaf tissue.

While this initial study proved both useful and interesting, in uncovering some genes involved in a geminivirus-plant interaction, more recent advances in sequencing technology, such as RNAseq [i.e. next-generation sequencing (NGS)], and the availability of a larger homologous EST *N. benthamiana* microarray (Nimblegen) and cassava genomic sequence data (www.phytozome.org), allow for more comprehensive host-responsive studies to be undertaken. A homologous microarray study of fully symptomatic *N. benthamiana* infected with SACMV-[ZA:99] at 21 dpi, and a RNAseq study at three timepoints in cassava are described in chapter 3 and 4, respectively.

Chapter 3

Transcriptional alterations in model host, *Nicotiana benthamiana*, in response
to infection by *South African cassava mosaic virus*

3.1 Abstract

Successful systemic infection by plant viruses is the outcome of complex molecular interactions between host and viral pathogen, leading to spatial and temporal alterations in plant gene expression. Microarray technology has predominantly been used to profile global gene expression of model host species infected with several positive-sense RNA viruses, but only one comprehensive Affymetrix ATH1 GeneChip study has been reported on a DNA geminivirus, *Cabbage leaf curl virus*, at symptom appearance (12 days post-infection). *South African cassava mosaic virus*-[South Africa:99] (SACMV-[ZA:99]), a bipartite circular ssDNA geminivirus, was used to agroinoculate the susceptible model host *Nicotiana benthamiana*. Plants were clearly symptomatic at 21 days post inoculation (6.40×10^2 molecules SACMV-[ZA:99] DNA-A/ng TNA), with the appearance of characteristic disease symptoms such as leaf distortion, chlorosis and stunting. Consequently, the 21 dpi time point was chosen to conduct a microarray study in order to dissect host gene responses in *N. benthamiana* using the commercially available custom-made 60-mer oligo NimbleGen platform (Roche). Three biological replicates were conducted for microarray hybridizations. The expression levels of many transcripts were altered in response to SACMV-[ZA:99], including encoded proteins involved in transcription networks, defence responses and plant hormone signalling. Differentially expressed transcripts with putative functions in virus movement, and geminivirus-activated metabolic pathways related to starch and sucrose are discussed.

3.2 Introduction

The last decade has seen an exponential increase in the number of plant virus-host studies investigating gene function, gene discovery and expression analysis in plants using high-throughput techniques such as microarray technology and whole genome sequencing (Alba et al., 2004; Ascencio-Ibáñez et al., 2008; García-Marcos et al., 2009; Hanssen et al., 2011; Postnikova and Nemchinov 2012; Rensink and Buell 2005; Whitham et al., 2003 and Whitham et al., 2006). In plants, cDNA and oligo microarray analyses have been used extensively to monitor global transcription profile changes that occur in host-virus interactions (Ascencio-Ibáñez et al., 2008; Dardick 2007; De Vos et al., 2005; Dowd et al., 2004; Okinaka et al., 2002 and Wan et al., 2002). Genomic tools and high-throughput sequencing technologies generate large sets of gene data which require a genomic systems biology (GSB) approach (Elena et al., 2011), encompassing bioinformatic tools, for dissection and interpretation of these interactomes.

Systemic infection can be considered to be a sum of virus replication and movement, and host defence and viral counter-defence reactions, and the relationship has been previously demonstrated to be a complex and intimate co-existence (Reviewed in Culver and Padmanabhan 2007; Pallas and Garcia 2011; Wan et al., 2002 and Whitham et al., 2006). During the onset and progression of viral disease, a number of spatial and temporal changes take place in the levels of gene expression within a host plant (Yang et al., 2007), and replication and movement within the susceptible host is achieved by efficiently recruiting and exploiting host factors, this being true for both RNA and DNA viruses (Ascencio-Ibáñez et al., 2008; Gutierrez 2004; and Whitham et al., 2003). Geminiviruses, in particular, do not encode their own DNA polymerases and rely on host nuclear DNA replication machinery in order to successfully replicate their genome. A large number of plant-virus interaction studies have focused on RNA viruses, in particular *Tobacco mosaic virus* (TMV), demonstrating that RNA viral infection is accompanied by vast changes in the plant transcriptome (Marathe et al., 2004; Whitham et al., 2006; Yang et al., 2007). While genome-wide virus associated plant-microbe studies to date have been primarily conducted in the model plant, *Arabidopsis*, infected with positive-sense ssRNA viruses, and more recently a dsRNA fijivirus of maize (Jia et al., 2012), very few have focused on the DNA geminiviruses. To our knowledge, only one comprehensive microarray study has been performed with *Cabbage leaf curl virus* (CaLCuV) in *Arabidopsis* (Ascencio-Ibáñez et al., 2008). Differential display reverse transcription-PCR and serial analysis of gene expression (SAGE) studies have revealed differential gene expression in geminivirus-resistant pepper (Anaya-Lopez et al., 2005) and cassava (Fregene et al., 2004), respectively. In another geminivirus study by Eybishtz et al., (2009), a reverse genetics approach was applied to identify genes involved in *Tomato yellow leaf curl virus* resistance. Approximately 70 different cDNAs, representing genes preferentially expressed in a

resistant tomato line compared to a susceptible line from the same breeding program, were identified.

Specifically, geminivirus proteins have been shown to interact with a diverse set of host factors in *Arabidopsis thaliana*, *Solanum lycopersicum* and *Nicotiana benthamiana* (Jeske, 2009), which both support viral replication and movement, but also can affect several metabolic pathways such as carbon metabolism (Shaitlin and Wolf 2000; Tecsí et al., 1996). Sugars, such as sucrose, are recognized as signalling molecules in plants, and are important in plant responses to abiotic and biotic stress (Bolouri-Moghaddam and van den Ende 2011). The co-ordination of sugar and hormonal pathways, in addition to their roles as carbon and energy sources, are likely to play a role in pathogen innate immunity. With regard to cytoplasmic or plasmadesmatal virus movement, different membrane-bound vesicles accumulate at cellular sites of viral replication, and plasma membrane domains, such as syntaxins and ATP-binding cassette (ABC) transporters, appear at the site of pathogen attack (Hückelhoven 2007). The *Arabidopsis* Synaptotagmins (SYTs), are a large family of Ca^{2+} /lipid binding proteins. SYTA, specifically, has been shown to localize to membrane-derived endosomes in plant cells and directly binds to the cell-to-cell movement proteins of CaLCuV and *Tobacco mosaic virus* (TMV) (Lewis and Lazarowitz 2010). The actin cytoskeleton or actin-ER network complex has been implicated in intra- and intercellular movement of plant viruses (Harries et al., 2009; Harries et al., 2010) and it has been demonstrated that RNA viruses have evolved differently in their requirement for microfilaments. Associated with the cytoskeleton are motor proteins (ATPases), such as kinesins and myosins, that derive energy from ATP hydrolysis to move along microtubules and actin microfilaments (Lee and Liu, 2004), and facilitate cargo trafficking. In plants, one of the key regulatory protein families involved in endosome organization, membrane targeting and trafficking, and recruitment of molecular motors to membranes, are the Rab GTPases (Woollard and Moore 2008), and therefore their potential role in virus movement is highly probable.

Cassava mosaic disease (CMD), an important disease of cassava (*Manihot esculenta* Crantz), in sub-Saharan Africa, is caused by whitefly-transmitted viruses of the genus *Begomovirus* (family *Geminiviridae*), including *South African cassava mosaic virus*-[South Africa:99] (AF155806) (SACMV-[ZA:99]) (Berrie et al., 2001). SACMV-[ZA:99] possesses two circular DNA molecules, designated DNA-A and DNA-B, of approximately 2.8 kb (Berrie et al., 2001), both of which are required for systemic infection of plants. *Nicotiana benthamiana* Domin has been extensively used as an experimental host (Goodin et al., 2008), to study the molecular characterization of geminivirus replication and protein function. Notably, two microarray studies profiling transcript expression changes induced by an enveloped RNA rhabdovirus and bunyavirus (Senthil et al., 2005) and three fruit tree RNA viruses

(potyvirus, nepovirus and ilarvirus) (Dardick, 2007) have been undertaken to study host-virus interactions. In the absence of the cassava genome and access to cassava oligoarray (Utsumi et al., 2012) at the start of this study, we used a 60-mer 13k *N. benthamiana* NimbleGen microarray (Nb-array) to monitor changes in the gene expression profile of model host, *N. benthamiana* systemically infected with SACMV-[ZA:99], and from the dataset obtained, candidate cassava homologs through protein sequence BLAST (www.phytozome.org) searches were identified. Several host candidate gene responses to SACMV-[ZA:99] were observed including differentially expressed transcription factors, and cytoplasmic and cell wall-associated proteins, which may play a putative role in SACMV-[ZA:99] infection and movement. In this report, we focus on genes that may play a role in cytoplasmic movement, or are involved in several metabolic pathways, where in particular, significant alterations in sucrose and starch metabolism were identified. This is the first study, to our knowledge, that explores a geminivirus-*N.benthamiana* microarray study as a platform to get a better understanding of the processes underlying geminivirus-host interactions at the molecular level, which could ultimately lead to strategies to control these virus pathogens.

3.3 Materials and Methods

3.3.1 Infectivity Study

3.3.1.1 Plant growth and virus inoculations

Nicotiana benthamiana plants were grown at in controlled growth chambers at 25 °C with a photoperiod of 16 h light/8 h dark for 6 weeks. 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 mobilized in *Agrobacterium tumefaciens* strain C58C1 (OD₆₀₀ of ±0.8) . *Agrobacterium* containing SACMV DNA-A as well as *Agrobacterium* containing SACMV-DNA-B were cultured independently of each other in Luria broth with working concentrations of 50mg/l rifampicin and 100mg/l kanamycin. *Agrobacterium* containing the pBIN19 plasmid (containing no dimeric insert) was also cultured. 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) of SACMV-[ZA:99] inoculum, using a 1 ml Hamilton syringe. Control plants were inoculated in parallel, and eighteen additional plants were mock-inoculated with 60µl of *Agrobacterium* containing the plant transformation vector pBIN19.

3.3.1.2 Total Nucleic Acid Extractions

In order to reduce variation, eighteen plants were randomly divided into three independent groups representing three biological replicates, each containing 6 plants. For each time point (7, 14, 21 and 35 dpi), the leaves closest to the apex were harvested from the six plants representing each group, and these leaves were pooled to form each of the biological replicates. Total nucleic acid (TNA) was isolated from these SACMV-[ZA:99] infected and mock-inoculated leaves using a modified CTAB-based extraction method (Doyle and Doyle 1987). 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 g at 4 °C for 10min. 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).

3.3.1.3 Confirmation of SACMV-[ZA:99] infection using conventional PCR

Systemic infection of *N. benthamiana* leaf tissue was confirmed by conventional PCR. A 550 bp fragment of the core coat protein (CCP) on DNA-A was amplified using degenerate forward primer: 5' GCCHATRTAYAGRAAGCCMAGRAT 3' and reverse primer 5' GGRTTDGARGCATGHGTACANGCC 3'. Each 50 µl PCR reaction mixture contained 0.4 µM of each primer, 200 µM dNTPs, 2.5mM MgCl₂, 2 units DreamTaq DNA polymerase (Fermentas), 1x DreamTaq Buffer (Fermentas), and nuclease-free water to a final volume of 50 µl. Approximately 500 ng of the genomic DNA template was added to the reaction mixture. Reactions were cycled in a MyCycler™ Thermal Cycler (Bio-Rad) at 95 °C for 2 minutes to activate the Taq DNA polymerase, followed by 30 cycles of denaturation at 95°C for 15 seconds, annealing at 57 °C for 15 seconds, primer extension at 72 °C for 15 seconds, and a final extension step of 72 °C for 5 minutes. As a negative control, genomic DNA extracted from a mock-inoculated leaf tissue was used. A “no template control or NTC” was also set up as a second negative control. As a positive control, 50 ng of DNA-A cloned into pBLUESCRIPT was used. Amplification products were examined by electrophoresis on a 1.2% agarose TAE gel containing 10 µg/ml ethidium bromide

3.3.1.4 Quantification of SACMV-[ZA:99] using quantitative Real-time PCR

Quantitative real-time PCR experiments were carried out according to the MIQE guidelines for qPCR experiments where the PCR amplification efficiency of SACMV-[ZA:99]-CP primers used for viral load experiments was calculated using the equation: Efficiency % = $(10^{[-1/\text{slope}]} - 1) \times 100\%$ (Bustin et al., 2009). Quantification of SACMV-[ZA:99] DNA-A in infected and mock-inoculated leaf tissue was carried out on the LightCycler 1.5 (Roche). A standard curve was set up for five known concentrations of DNA-A (10 ng, 1 ng, 100 pg, 10 pg and 1 pg). Each of these standards were spiked with 100 ng TNA extracted and purified from healthy *N. benthamiana* leaf tissue, in order to set up a standard curve that closely mimics the infected leaf TNA being used for quantification. All qPCR reactions were performed in LightCycler® Capillaries using the LightCycler 1.5 using LightCycler® FastStart DNA Master^{Plus} SYBR Green I kit (Roche). TNA extracted from infected and mock-inoculated leaf tissue was standardised to a concentration of 100 ng/ul. Quantification of SACMV-[ZA:99] DNA-A was carried out on all 3 biological replicates set up in duplicate. A 150 bp region of the coat protein was targeted during amplification using the primers: SACMV-[ZA:99] CP (fwd) : 5'ACGTCCGTCGCAAGTACGAT3' and SACMV-[ZA:99] CP (rev) 5'ATTGTCATGTCGAATAGTACG 3'. The amplification conditions used were as follows: 95°C for 10 min, followed by 35 cycles of 95°C for 10 sec, 60°C for 10 sec, and 72°C for 15 sec. A melting curve (65°C-95°C with a heating rate of 0.1 °C/s and a continuous fluorescence measurement) was run after the PCR cycles. 18S rDNA was used as the internal loading control using 18S (fwd) : 5' GGCAAATAGGAGCCAATGAA3' and 18S(rev) : 5' GGGGTGAACCAAAAGCTGTA 3' primers. Data were exported to Microsoft Excel for further analysis and statistical data testing using the Student's T-test.

3.3.2 NimbleGen *Nicotiana benthamiana* Microarray Study

2.3.2.1 Experimental design and sample preparation

A reference design was used for this microarray study (Yang and Speed, 2003). In order to obtain robust and statistically accurate data, three independent sets of plants representing three biological replicates were used for the virus-infected leaf tissue for all virus treatments and mock-inoculated controls. As with the infectivity assay, eighteen plants were used in total, where each biological replicate contained 6 plants. Two symptomatic leaves, closest to the apex, was harvested from each of the six plants in each group at each timepoint. Leaf tissue was immediately frozen in liquid nitrogen and stored at -80°C until used for downstream RNA extractions.

3.3.2.2 RNA extractions and Quality Control of RNA leaf tissue samples

Total RNA was extracted from infected and mock-inoculated leaf tissue at 21dpi using TRI-reagent (Sigma Aldrich,) according to the manufacturer's specifications. Briefly, the two upper-most infected and symptomatic (approximately 100mg) leaf tissue was ground to a fine powder in liquid nitrogen using a pestle and mortar. One millilitre TRI-reagent was added to each homogenized leaf tissue sample and incubated at room temperature for 5 min. RNA was separated from DNA and protein by the addition of 200 µl chloroform, followed by vigorous shaking, incubation at room temperature for 15 min, and centrifugation at 13000 g for 10 min at 4°C. RNA was precipitated from the aqueous phase by addition of 500 µl isopropanol, and incubation for 10 min at room temperature. RNA pellets were washed with 75 % ethanol to remove residual chloroform. The pellets were briefly air dried, resuspended in 100 µl RNase-free water, and quantified on a Nanodrop 1000 spectrophotometer (Nanodrop). The purity of the extracts was assessed based on the A260/ A280 and A260/A230 ratios. After quantification, 1µl Ribolock RNase inhibitor (Fermentas) was added to each sample.

3.3.2.3 Generation of double-stranded cDNA for NimbleGen arrays.

RNA extracted from the infected and mock –inoculated leaves from 21dpi leaf tissue (as above) was converted to double-stranded cDNA by using the SuperScript Choice System for cDNA Synthesis (Invitrogen). Briefly, 10 µg of total RNA was incubated with 1X first strand buffer, 10 mM DTT, 500 µM dNTPs, 5 pM primer for 60 min at room temperature. Second-strand synthesis was accomplished by incubation with 200 µM dNTPs, 0.1 units DNA ligase, 0.3 units DNA polymerase I, 0.02 units RNase, 1X second-strand buffer and 10 units T4 DNA polymerase for 2 hours. Double-stranded cDNA was purified by using phenol-chloroform extraction and precipitated with ethanol. This was washed with 80% ethanol, and resuspended in 30 µl of water. Double-stranded cDNA was sent off to Roche NimbleGen Iceland (<http://www.nimblegen.com/>), where all hybridization and scanning was conducted.

Hybridizations were made against a custom 1-plex, 385K *N. benthamiana* 60-mer oligonucleotide microarray (NimbleGene™ Nb-array) slide. Double-stranded cDNA was sent to NimbleGen Systems, Inc. (Reykjavik, Iceland), where all hybridizations were performed. Briefly, 2.5 µg of high-quality ds cDNA for each biological replicate and reference control sample was labelled with Cyanine-3 using a random priming method. Cy3-labelled ds cDNA was hybridized to the *Nb-array* expression array according to the manufacturer's instructions (<http://www.nimblegen.com/>).

The probes, that were designed and printed onto the Nb-array, corresponded to 13,014 unique *N. benthamiana* ESTs [collaboration between Prof S. Lommel's group (North Carolina State University) and NimbleGen (Madison, WI, U.S.A.)]. Of the original 13,719 submitted ESTs, 13,415 were included on the array, where 12,738 of 13,415 genes are *N. benthamiana* genes corresponding to approximately 38% *N. benthamiana* transcriptome, and 677 of the 13,415 genes are non-*N. benthamiana* (*Drosophila* EST) orthologs. (Feulner, 2007, Supplementary Data S3.1). There are 9 different (60-mer) probes representing each EST and each of these probes has been replicated 3 times, resulting in 27 probes (NBENT sequence IDs) per EST.

3.3.2.4 Data analysis: Cluster analysis, Gene Ontology and pathway analysis

DNASTAR ArrayStar 4 software was used to calculate the expression intensities from scanned images and normalized using the Robust Multichip Average (RMA) method (Irizarry et al, 2003). Genes were considered differentially expressed when a difference in fold change (FC) was observed at ($p < 0.05$). FC was calculated as normalized signal intensity of SACMV-inoculated sample/normalized signal of the mock-inoculated reference sample. The FC was calculated as the mean of three replicates. Data was visualized and hierarchical cluster analysis was performed using Euclidean distance metric and average linkage clustering with TIGR MultiExperiment Viewer (MEV) to identify patterns of gene expression and differentially expressed genes, available at <http://www.tigr.org/software/>. Functional category classification using the GO of the Arabidopsis Information Resource (TAIR10, <http://www.arabidopsis.org/tools/bulk/go/index.jsp>) for three components ('biological process', 'molecular function' and 'cellular component') was performed on the differentially expressed genes. To determine whether the SACMV-[ZA:99] responsive transcripts were involved in specific pathways, 423 *Arabidopsis* AGI TAIR loci representing the 611 differentially expressed genes were used to search against Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in *Arabidopsis thaliana* using the online tool DAVID (Bioinformatics Resources 6.7 : <http://david.abcc.ncifcrf.gov/tools.jsp>).

3.3.2.5 Data retrieval

MIAME information on this SACMV-*N. benthamiana* microarray study has been deposited in the Gene Expression Omnibus (GEO) of NCBI (www.ncbi.nlm.nih.gov/geo/) (Edgar et al., 2002 ; Barrett et al., 2007). In order to retrieve data presented in this paper, the following accession number can be used: GSE43584

3.3.2.6 Microarray Validations using RT-qPCR

Semi-quantitative real-time PCR was performed on selected genes (Table 3.1), in order to validate data obtained from this microarray study. In brief, 1 µg of DNase-treated total RNA was reverse transcribed using the Improm-II-reverse transcriptase kit (Promega, Madison, WI) according to manufacturer's instruction. RNA, dNTPs and Oligo dT primer were denatured for 10 min at 70°C; then kept at 25 °C for 5 min before the reverse transcription master mix (1 µl of 10 mM dNTP mix, 0.5 µl of RNase inhibitor , 2.4 µl of 25 mM MgCl₂, 4 µl of 5×RT buffer) was added in a final volume of 20µl. Reverse transcription was performed at 42°C for 1 hour followed by a 10 min incubation step at 70°C. Control reactions were set up without the addition of reverse transcriptase and used as negative controls in the real-time PCR study. RT-qPCR experiments were conducted on the Lightcycler 1.5 for all genes using the appropriate primer pair (Table 3.1) for each reaction. In brief, all qPCR reactions were performed in LightCycler® Capillaries using the LightCycler 1.5 using LightCycler ® FastStart DNA Master^{Plus} SYBR Green I kit (Roche). Three biological replicates and a technical replicate were run for SACMV-[ZA:99]-infected and mock-inoculated leave cDNA samples at 21dpi . A volume of 1 µl of undiluted cDNA was used for each reaction. The cycling conditions used were as follows: initial denaturation for 10 min at 95°C (hot start) followed by an amplification and quantification cycle repeated 35 times, each consisting of 10 sec denaturing at 95°C, 10 sec annealing at primer specific temperatures, 15 sec primer extension at 72°C with a single fluorescence measurement. Melting curve cycle was obtained by heating to 65°C for 15 sec with a heating rate of 0.1°C per second with a continuous fluorescence measurement. GAPDH was the gene found to be most stably expressed across all qPCR runs and was used as an endogenous control for normalization. Statistical analysis was carried out in excel using the data analysis plug-in.

Table 3.1 Primers used for qPCR validation of NimbleGen microarray data

Gene Name	Accession number	Forward primer (5'-3')	Reverse primer (5'-3')
Glutaredoxin	AB679830.1	TGCGGGTATTGCACGAGGGT	TGCTGTCGCAGCCACCAACAT
ERF4	NM_001247384.1	AATCTCCGCTCCGATTTCCTGAA	TGGATCTCTTATCTCCGCCGCAAA
WRKY30	AY547496.1	GGTGCGCACACAAGTACGATCA	ACAGAGCTGCAAGTGTGCTGGC
β -tubulin	U91564.2	GCATCTTTGCGTACACTTTGCT	ACATAAGCCCCAAACTAGCTGGA
CYP450	EH371009.1	ACACGTTTATACCCTTCGTGGGAG	TTTGGCACAGGAGAAGCATCGAAC
CP12	AF359459.2	ACCTTAAGGCTGTAAGTGCAACACC	CTGCTGCGCTCGCCTCTTCA
Histone H3	AB015760.1	TGCCCGTAAATCTACTGGAGGCAA	ACGAAGAGCAACAGTTCAGGTCT
1,3 Glucanase	X53600.1	ACGAGTATCCAGATCCGCACATGA	TCTCTGGGTGCAAACAATGGAGGT
Chitinase	S44869.1	AACACCAATGACTACTGTGGCCCT	CCTTGGCATGCATTATCGTTGCGA
Knotted 3	AF544054.1	ACGAGCAGAAGATCGGGAAGTAA	ACTCCGAAGGATATGGCCACTTGT
GAPDH (Reference)	AF5297779.1	ATGGCCTTCAGAGTACCAACTGCT	GCTTGACCTGCTGTCACCAACAAA

3.4 Results

3.4.1 Infectivity assay

Cassava mosaic disease (CMD) symptom development and severity on new emerging *N. benthamiana* leaves was assessed weekly over a 35 day period (Figs. 3.1 A –D). A general yellowing associated with agroinfection disappeared within 2 days post inoculation (dpi). Phenotypes of SACMV-[ZA:99] infected leaf tissue was scored at each of the time points (i.e. 7, 14, 21 and 35 dpi) using the symptom severity index outlined by Fauquet and Fargette (1988, Table 3.2). This symptom severity index assigns the minimum score of 0 to plants that are asymptomatic and the maximum score of 5 to plants that exhibit CMD-like symptoms such as severe mosaic as well the severe reduction in leaf size. CMD-like symptoms appeared at 12 to 14 dpi (Fig. 1B), with a mean severity score of 1. Newly emerged leaves at 12dpi started displaying signs of leaf margin distortion. All plants were clearly symptomatic by 21 dpi, with a mean severity score of 4.5, and by 35 dpi were

fully symptomatic. Symptoms of leaves assessed at 21-35 dpi included chlorosis, reduction in leaf size and blistering when compared with the leaves of the healthy control (Figs. 3.1C and D). Temporal changes in the replication (virus titer) levels of SACMV-[ZA:99] were monitored, using absolute quantitative real time PCR (qPCR), in young apical leaf tissue samples from each of the 4 timepoints (7, 14, 21 and 35 dpi) which corresponded to early (pre-symptomatic) infection, onset of virus replication, onset of symptoms and fully symptomatic, respectively. qPCR results (Fig. 3.2B) showed a clear trend of increasing virus titer (DNA-A/ng TNA) from 7 dpi over 28 days (total of 35 dpi). Virus titer was largest with a 20-fold increase between 14dpi (3.23×10^1 molecules of SACMV NAA/ng TNA) and 21dpi (6.40×10^2 molecules SACMV DNA A/ng TNA). Virus titer between 21dpi and 35 dpi (2.89×10^3 molecules SACMV DNA A/ng TNA) increased by 4.5-fold. Virus was not detected in any of the mock-inoculated leaf tissue sample at any of the timepoints. From these results, timepoint 21 dpi was selected for the microarray study, as this reflects a phase where SACMV-[ZA:99] is actively replicating and moving, and symptoms are well established, thereby representing a period of intense host responses to infection.

Table 3.2 Symptom severity indexing scores used to assess and describe symptom development in infected *N. benthamiana* leaf tissue.

Symptom Score	Symptom description
0	No visible symptoms
1	Mild chlorosis without deformation
2	Clear mosaic with or without slight leaf deformation
3	Strong mosaic all over the leaflets with some leaf deformation
4	Same as score 3 except with severe leaf deformation
5	Severe mosaic and severe reduction of leaf size

3.4.2 Bioinformatics and identification of differentially expressed genes

3.4.2.1 Updating Annotations for Nb-array

The sequences of the ESTs used for the probes designed for the Nb- array were last annotated in June 2009. More recent available draft sequence data of *Ricinus communis* (castor bean), *Nicotiana benthamiana*, *Solanum lycopersicum* (tomato), *Vitis vinifera* (grapevine), *Solanum tuberosum* (potato), and *Eucalyptus grandis* (eucalyptus) (Chan et al., 2010; Grattapaglia et al., 2011; Bombarely et al., 2012; The Tomato Genome Consortium, 2012; Jaillon et al., 2007; Xu et al., 2011) and sequence information available through the *The Arabidopsis Information Resource Database* (TAIR),

was used to update the annotations for the sequences used for the microarray probe set, particularly our 611 differentially expressed gene set (Supplementary Data S3.2).

3.4.2.2 Cassava homologs identified using differentially expressed genes

In November 2009, the first draft of the cassava genome was made freely available at Phytozome (<http://www.phytozome.net/cassava.php>). Potential homologs for all 611 genes were evaluated. Protein sequences corresponding to the probes spotted on the *Nb*-array were downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/protein/>) and queried against the *M. esculenta* protein database available through phytozome using BLASTX. A cut-off expectation (E)-value of 10^{-5} and sequence identity greater than 50%, was used in this analysis. Seventy one percent (434/611) of the genes found to be differentially expressed had homolog matches in cassava (Supplementary Data S3. 3).

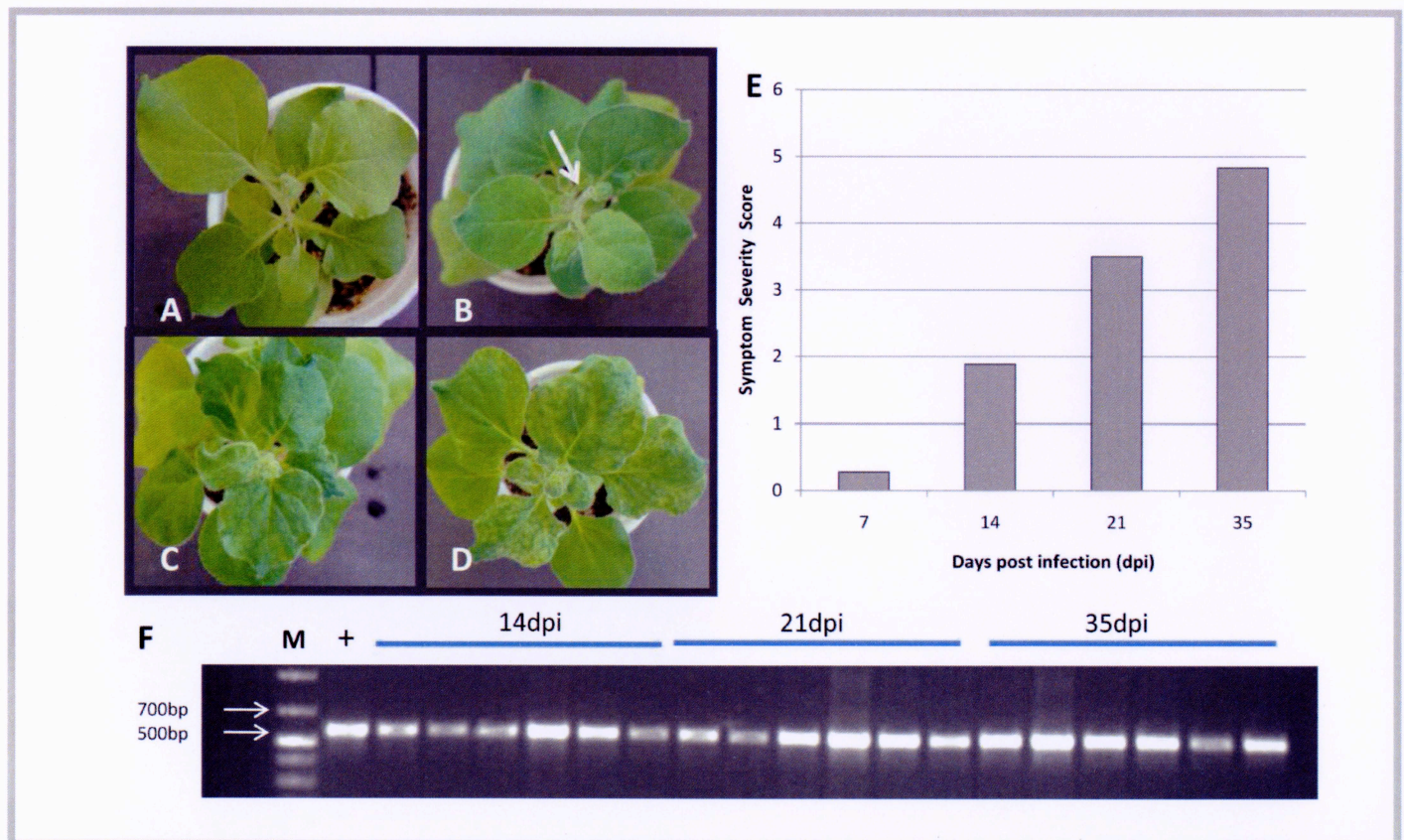


Figure 3.1 Symptom development in SACMV-[ZA:99] -infected *N. benthamiana* over a 35 day period. **(A)** Healthy plants with no visible symptoms. **(B)** First appearance of symptoms observed at 14dpi. **(C)** Arrow-head indicates the presence of newly emerging leaf which is reduced in size and distorted. **(D)** Fully symptomatic plants by 21 dpi and by 35 dpi; newly developed and older existing leaves were severely distorted, reduced in size and displaying characteristic mosaic (yellowing). **(E):** Infected plants assigned a score based on the symptom severity score index (Fauquet & Fargette, 1988). **(F)** Confirmation of SACMV-[ZA:99] infection using PCR with primers targeting the core coat protein (CCP) of SACMV-[ZA:99]. SACMV-[ZA:99] cloned into pBluescript was used as a positive control (+) for PCR (50ng).

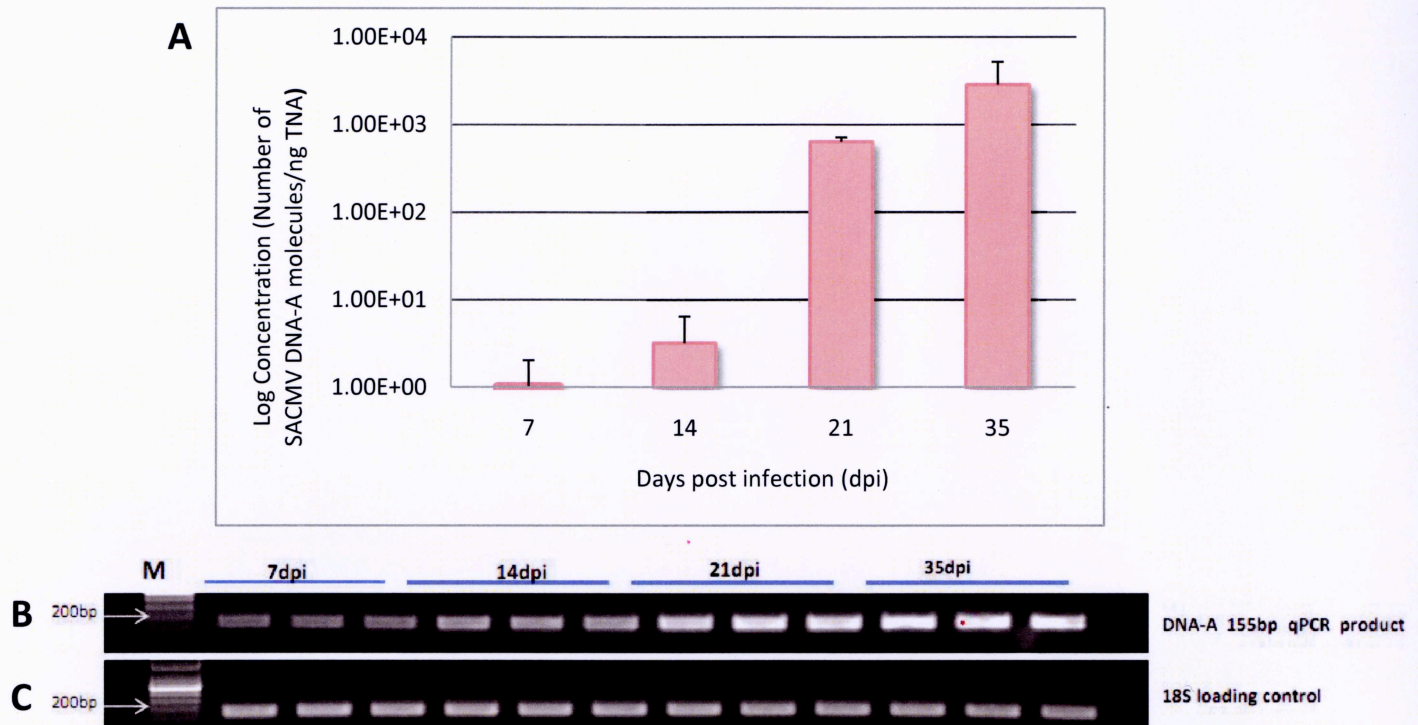


Figure 3.2 (A) Viral quantification from SACMV-[ZA:99]-infected *N.benthamiana* leaf tissue (n=6). The molecules of DNA-A/ng TNA for 7-35dpi in infected leaf tissue samples were calculated using Avagadro's constant **(B)** 155bp DNA-A PCR product obtained from qPCR **(C)** 18S rDNA 150bp PCR product obtained from qPCR. 18S rDNA served as a loading control in qPCR experiments

3.4.3 NimbleGen Microarray study and qPCR validations

Following quantile (RMA) normalization and employing a stringent cut-off (2-fold, $p < 0.05$, 5% FDR), a total of 611 genes were found to be differentially expressed in systemically infected leaf tissue at 21 dpi. Genes were considered differentially expressed when a difference in fold change (FC) was observed at ($p < 0.05$). FC was calculated as normalized signal intensity of SACMV-inoculated sample/normalized signal of the mock-inoculated reference sample. The FC was calculated as the mean of three replicates. This represents approximately 4.7% (611 of 13 014) of the transcripts that were significantly changed in response to SACMV-[ZA:99] infection. Pair-wise comparisons between each of the three biological replicate and the reference slide were all highly correlated and reproducible ($R^2 > 0.96$). Of these, 483 genes were found to be upregulated and 128 genes were downregulated (Figs. 3.3 A and B). Figure 3.3A clearly demonstrates a landscape snapshot of over-

representation of upregulated genes identified, which correlates to dense clustering on the positive axis of the $\log_2$ graph (Fig. 3.3 B), where the negative axis was more sparsely clustered (Fig. 3.3B). Additionally, using the P-value <0.05 as the threshold value, we were able to functionally categorize 585 of the 611 differentially expressed transcripts in the library. GOSlim terms were counted and grouped in a hierarchical manner into the major GO functional categories of Cellular Component, Biological Process and Molecular Function (Fig. 3.4), which included fifteen cellular components, fourteen molecular functions and fourteen biological processes.

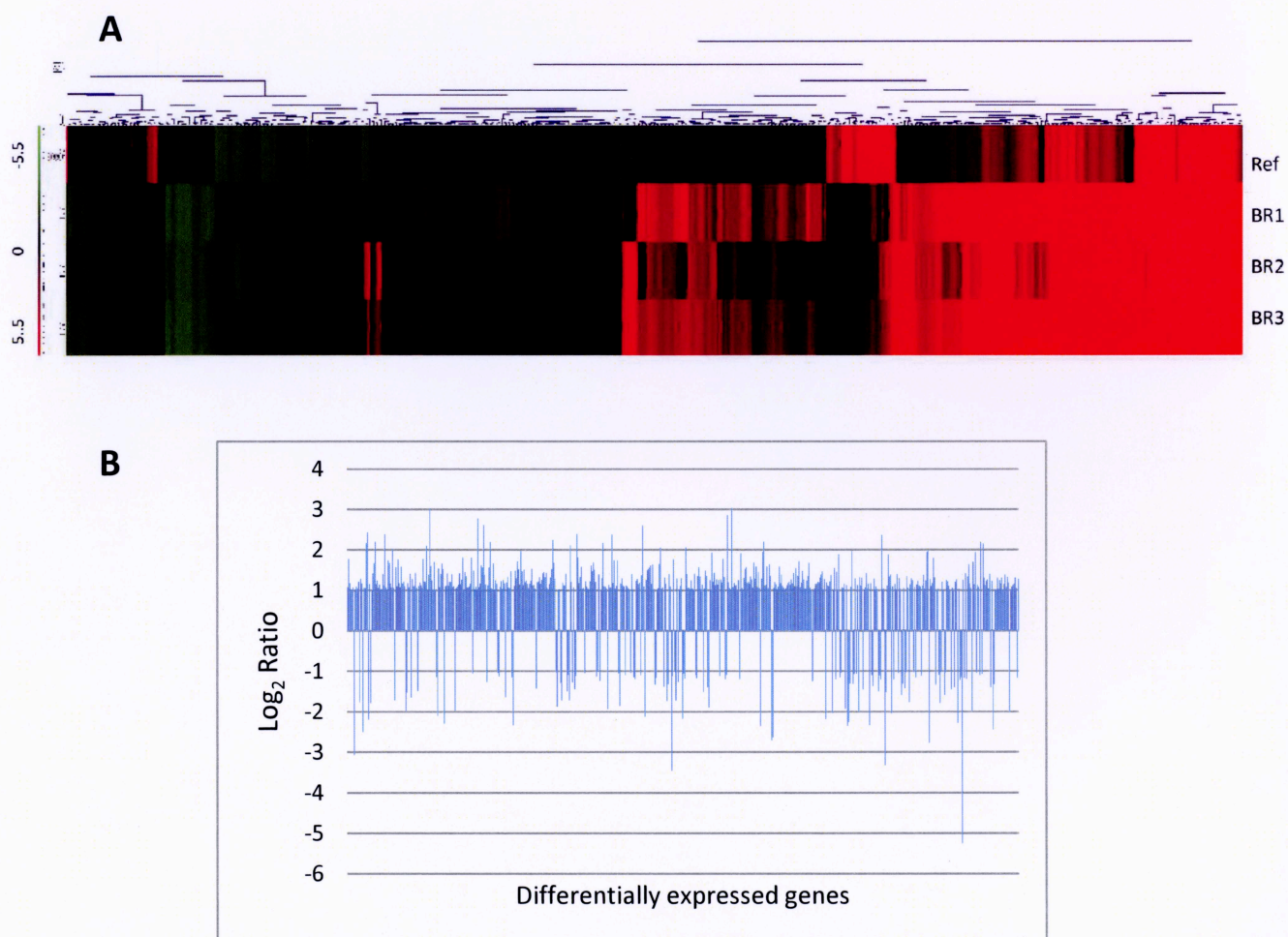


Figure 3.3 (A) Heat map showing the distribution of upregulated (red) and downregulated (green) genes (x-axis) identified as differentially expressed at 21dpi. **(B)** Distribution of 611 differentially expressed genes (DEGs), where 483 genes were upregulated and 128 were downregulated.

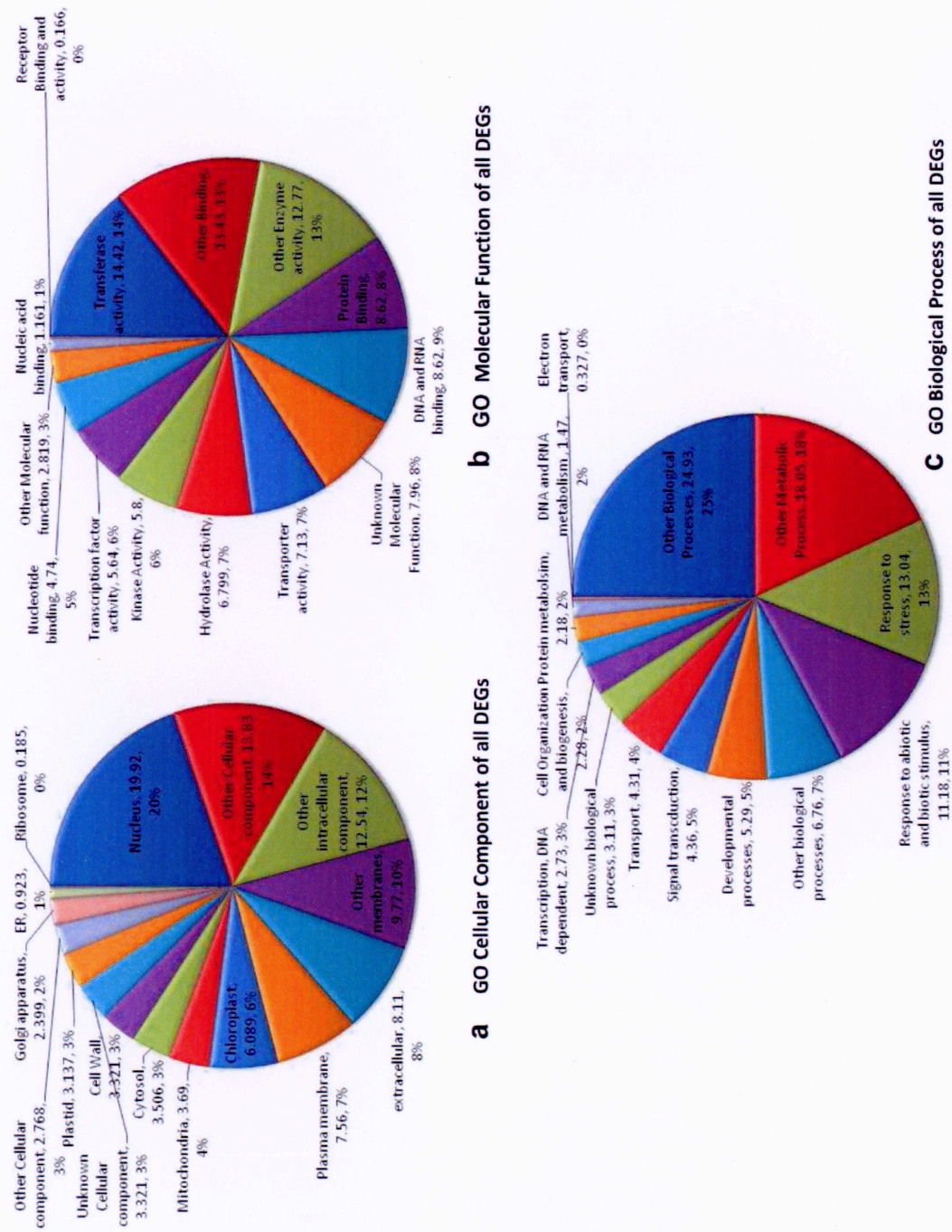


Figure 3.4 (a-c) Gene Ontology annotation of DEGs. Pie charts showing the functional categorization of genes differentially expressed at 21dpi, when plants are clearly symptomatic. The numbers are the percentage of genes in each category differentially expressed during infection. Significantly induced and repressed *N. benthamiana* transcripts were grouped into functional groups using TAIR 10 (<http://www.arabidopsis.org/tools/bulk/go/index.jsp>)

In order to independently validate the expression profiles obtained by microarray analysis, we also performed qPCR analysis. Ten genes were selected from the genes shown in Supplementary Data S3.2. The genes chosen included Chitinase, 1,3 Glucanase, Histone H3, Knotted 3, WRKY3, ERF4, Glutaredoxin, β -tubulin, CYP450 and CP12 (Table 3.1). GAPDH was used as a control because its expression level remained constitutively expressed between mock-inoculated and SACMV-[ZA:99] infected leaf tissue samples. The Log_2 expression results from the qPCR analysis were calculated using the relative standard curve method (Larionov et al. 2005). Standard curves were set up for each gene using five dilutions of cDNA (ie undiluted, 1:10, 1:100, 1:1000 and 1:10000. Results obtained for each gene were consistent in their expression pattern compared with those of the microarray analysis, where the R^2 value between the two experiments for the eleven expression values was 0.9403 (Fig. 3.5).

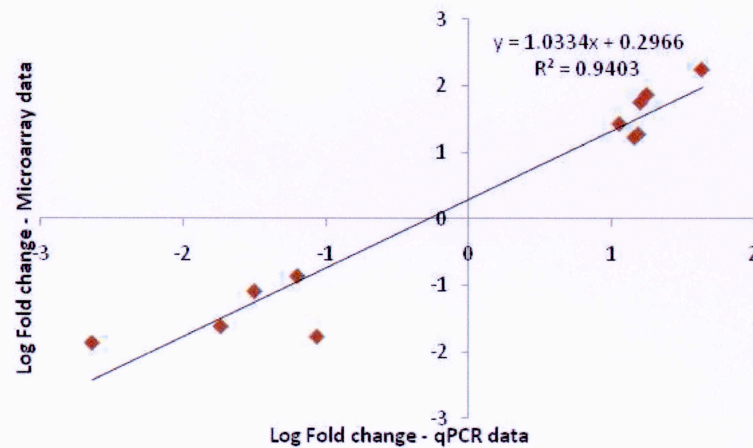


Figure 3.5 Log ratios of qPCR validations were found to correlate to the log ratios obtained from the microarray data ($R^2 = 0.94$)

From the identified altered gene transcript data, proteins that may be implicated in the response of *N. benthamiana* to infection, or may directly interact with SACMV-[ZA:99] proteins to promote disease-associated differentially expressed transcripts (Supplementary Data S3. 5), were targeted for further discussion. Furthermore, in order to determine whether the SACMV-[ZA:99] responsive transcripts were involved in specific metabolic pathways, 423 *Arabidopsis* AGI TAIR loci representing the 611 differentially expressed genes were used to search against Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in *Arabidopsis thaliana* using the online tool DAVID (Bioinformatics

Resources 6.7 : <http://david.abcc.ncifcrf.gov/>) (Huang et al., 2009a and Huang et al., 2009b). The relationship between microarray probe sequences and *Arabidopsis* was established by downloading the microarray full probe sequences for 611 differentially expressed genes and running a BLAST search against *Arabidopsis* sequences using AmiGO (http://amigo.geneontology.org/cgi-bin/amigo/blast.cgi?session_id=4054amigo1358853163) (E-value of $\leq 10^{-5}$ and sequence identity greater than 50%, was used in this analysis). Of the 423 AGI loci used, only 16 transcripts mapped to 3 different pathways: 8 genes mapped to 'Starch and sucrose metabolism' (KEGG: ath00500), 5 genes mapped to "Amino acid and nucleotide sugar" metabolism' (KEGG: ath00520) and 3 genes mapped to 'alpha-Linolenic acid metabolism' (KEGG: ath00592), and these are discussed below.

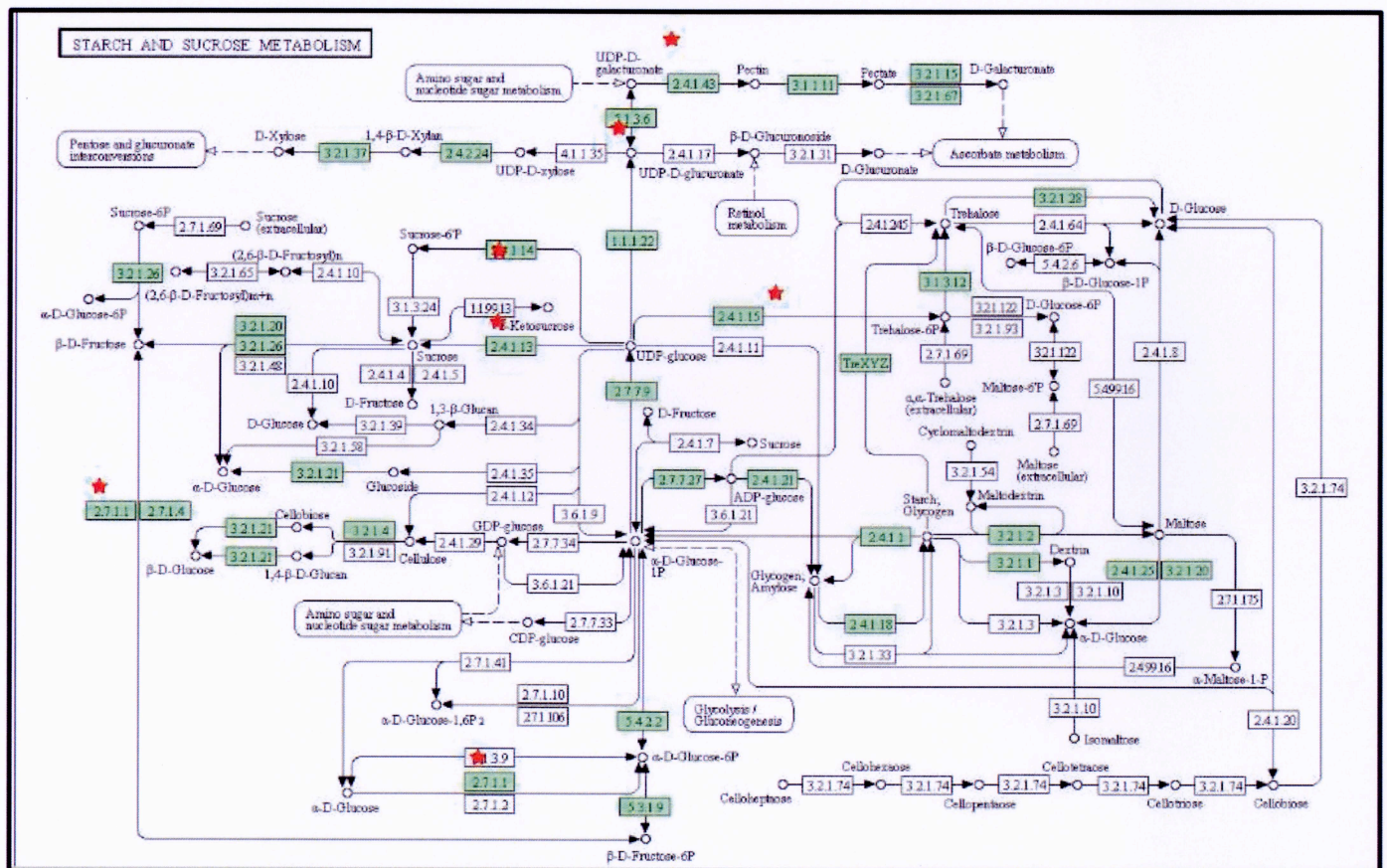


Figure 3.6 KEGG pathways for Starch and Sucrose metabolism in *Arabidopsis*. Red stars represent points at which differentially expressed genes participate in the pathway

3.5 Discussion

It has been reported in a number of studies that viruses alter global transcriptional networks of their hosts (Golem and Culver 2003; Itaya et al., 2002; Whitman et al., 2003), as a result of both the direct and indirect effects that viral replication can have on the host during the development of disease. Gene expression in host plants can change for two reasons; the first is an active defence or basal immunity response, be it successful or not, against the invading pathogen and the second, particularly in the case of geminiviruses, is the reprogramming of plant gene expression to favour viral replication and movement. Therefore, understanding virus-induced changes in host gene expression provides a valuable means of identifying the cellular processes of the host that are involved in the replicative cycle of a plant geminivirus. This study, reports for the first time, the use of the NimbleGene™ Nb-array, representing 13 014 ESTs, to interrogate gene expression changes induced in systemically infected *N.benthamiana* by a ssDNA geminivirus at 21 dpi. This timepoint was selected since, from a previous cross-species microarray experiment [using hybridization custom slides printed with an enriched library of 5500 defence-related cDNA library created from an *Arabidopsis-Ralstonia solanacearum* plant-pathogen interaction study (<http://www.microarray.up.ac.za/>) (unpublished)], it was determined that the greatest number of genes that were putatively involved in geminivirus movement and replication, or general biotic stress responses host responses to SACMV-[ZA:99], were expressed at 21 dpi. Furthermore, several homologs were identified in cassava and putative annotations assigned despite the current estimated annotation status of only 67% (20 641 of the 30 666 transcripts have Pfam domain annotation and KOG orthology) (Prochnick et al., 2012).

3.5.1 Infectivity assay

Geminivirus concentrations in single infections have previously been determined using Southern blot assays (Méndez-Lozano et al., 2003). In this study, the concentration of SACMV-[ZA:99] DNA-A was determined by real-time PCR, a procedure that allows a more direct and accurate quantification of DNA concentration. SACMV-[ZA:99] replication titers increased between early timepoints (7-14 dpi) followed by a significant 20-fold increase later on in infection (14-21dpi). Concentrations of 18S (loading control) remained constant between the different timepoints (7-35dpi) regardless of the increase SACMV-[ZA:99]. Greater virus replication levels are expected to correlate with the onset of visible symptoms, and once the plant is fully systemically infected, one would predict a high virus load, which indeed was observed (Fig.3.2 A). A similar observation has previously been reported in

tomato infected with TYLCV, where viral DNA accumulation was positively correlated with symptom severity (Rom et al, 1993). Similarly, the same pattern was observed in cassava plants infected with a potyvirus, *Cassava brown streak virus* (CBSV), where the presence of symptoms on leaves correlated with an increase in CBSV RNA titers (Moreno et al., 2011)

3.5.2 Global gene expression profiles in *N. benthamiana* response to SACMV-[ZA:99] at 21dpi

Supporting evidence by a number of studies has already shown that geminiviruses have the ability to affect a number of host cell pathways in order to establish successful infection. Some of these include DNA replication, changes of cell cycle control, signal transduction, transcriptional regulation, and trafficking and movement of macromolecules (Ascencio-Ibáñez et al., 2008; Gutierrez, 2002; Castillo et al., 2004; Sánchez-Durán et al., 2011), and in this study, global changes in transcription were also noted. However, no direct comparisons can be made between our and their data sets, especially with the geminivirus CaLCuV-*Arabidopsis* global transcriptome (Affymetrix ATH1 GeneChip) study (Ascencio- Ibáñez et al., 2008). This is due to differences in probes represented on the two microarrays, even though the CaLCuV study has proven useful in supporting evidence for putative geminivirus-responsive candidate genes in *N. benthamiana*.

A number of individual experimental studies have demonstrated host proteins that are involved in interacting with geminivirus proteins supporting replication, movement, symptom development, nuclear import and other geminivirus-related functions (Table 2; Jeske 2009). Additionally, several microarray and yeast two-hybrid studies with a range of RNA viruses have revealed virus-host protein interactions (Ahlquist et al., 2003; Lellis et al., 2002). Collectively, these can inform putative roles in SACMV-[ZA:99] infection in *N. benthamiana* and its natural host cassava, where homologs have been identified. From the Gene Ontology results discussed above, and the altered gene profiles identified in response to SACMV-[ZA: 99] (Supplementary Data S3. 3, Supplementary Data S3. 5), candidate genes with a putative role in infection, were selected for further discussion.

3.5.3 Induction of defence-related transcripts

Some of the transcripts involved in the induction of general plant defense response included lipoxygenase (LOX), chitinase, heat shock proteins (HSP), glutathione-S-transferase and neutral PR-5. These findings suggest that SACMV-induced expression changes are similar and consistent with those gene expression changes already observed and reported in other stress treatments and virus

infections (Babu et al., 2008; Golem and Culver, 2003; Pompe-Novak et al., 2006; Whitman et al., 2003). Of these transcripts, glutathione-S-transferase (GST) was found to have the largest fold change of 5.144 ($\log_2 = 2.362$). In a gene expression study conducted by Whitham et al., in 2003, it was observed that *Arabidopsis* leaves infected with five different RNA viruses resulted in the induction of four GST homologues. Glutathione S-transferases (GSTs) are a ubiquitous class of enzymes that display a number of antioxidant properties. The induction of reactive oxygen species (ROS) has been previously linked with plant immune responses to virus infection (Love et al., 2005; Riedle-Bauer 1998). Similarly, as observed in other plant and animal virus infections, SAMCV-[ZA:99] infection resulted in the upregulation of three heat-shock proteins/factors (HSPs) (Supplementary Data S3.1 and Supplementary Data S3.2) (Aparicio et al., 2005; Aranda and Maule 1998, Ventelon-Debout et al., 2004; Whitham et al., 2003). HSPs are a major class of protein chaperones that are involved in responses to abiotic stress and/or infection by pathogens and in a microarray analysis, Whitham and colleagues confirmed that the upregulation of HSP in response to viral infection was a common occurrence in *Arabidopsis* leaves infected with one of five different positive stranded RNA. In addition, the upregulation of pathogenesis related (PR) proteins such as chitinase and neutral PR-5 was also observed in this susceptible *N. benthamiana* study. The upregulation of these genes represent the host attempting to mount a broad non-specific defence response against SAMCV-[ZA :99], although the increase in virus titer and appearance of symptom development is an indication that this active defence response is unsuccessful to halt SAMCV-[ZA:99] disease development. It has been shown in many compatible plant virus-host studies, that despite progression of disease symptoms, defence-related responses persist throughout the infection but have no effect on viral infection (Babu et al., 2008 ; Senthil et al., 2005). This data supports a delayed host defence response, where basal immune responses in the susceptible host *N.benthamiana* persist even at full symptomatic infection (21dpi).

3.5.3 Plant hormone responses

Hormones, such as ethylene, jasmonic acid, abscisic acid, gibberellins and salicylic acid are present in plants in basal amounts, yet act in a well-balanced and regulative manner during plant growth and development (Culver and Padmanabhan, 2007). Any change from normal levels of phytohormones such as those caused by infection with virus pathogens could significantly alter physiological processes and morphology, resulting in symptoms such as stunting and leaf deformation, as was observed in our study. Transcripts encoding proteins involved in the synthesis and signal transduction of hormones such as ethylene, auxin, and jasmonate were found to be induced by 2-

fold or more during systemic infection by SACMV-[ZA:99]. These include Nt-IAA4.3, Jasmonate ZIM domain protein, and six ethylene-responsive transcription factors. A report in 2005, by Love et al., showed that ethylene-signalling mutants reduced virus titers of *Cauliflower mosaic virus* and hindered long-distance movement of the virus. The upregulation, in particular, of several ethylene-related transcripts by SACMV- [ZA:99] may indicate that ethylene is favouring the replication and movement of the virus during systemic infection at 21dpi. It is also possible that upregulation of some signalling pathways may be linked to an attempt by *N. benthamiana* to mount a defence response, albeit unsuccessfully. For example, in susceptible *Arabidopsis* infected with *Cucumber mosaic virus* (CMV) and *Oilseed rape mosaic virus* (OSRV), expression of many defence-related genes share components of signalling pathways involved in incompatible interactions, but their increased expression have no effect on virus infection (Huang et al., 2005).

3.5.4 Transcription factors

From the GO data for this study, it was observed that 17.44% (Fig. 3.5) of genes associated with the nucleus were differentially expressed during SACMV [ZA:99] infection at 21dpi, highly indicative of a consequence of replication of DNA-A and DNA-B components by host machinery. Furthermore, the activation of several WRKY transcription factors and suppression of histone-related transcripts were also observed. The same observation was made by Senthil and colleagues in 2005, where they observed the downregulation of histone H1e, which is required for cell biogenesis, and the upregulation of transcriptional factor WRKY in *N. benthamiana* infected with either a rhabdovirus (*Impatiens necrotic spot virus*) which replicates in the nucleus, or a plant bunyavirus (*Sonchus yellow net virus*). WRKY factors belong to a superfamily of DNA-binding transcriptional regulators which are crucial for plant growth and development, and mediate responses to biotic and abiotic stress and innate immunity (Rushton et al., 2010). WRKY proteins function through interactions with a diverse number of partners including MAP kinases, calmodulin and histone deacetylases. There have been at least 49 WRKY transcription factors identified in *Arabidopsis* that are specifically associated with pathogen infection (Dong et al., 2003 and Eulgem and Somssich, 2007) or salicylic acid (SA) treatment (Li et al., 2004). The expression of a number of WRKY genes is known to be induced by infection by various pathogens, including viruses (Yang et al., 1999) and bacteria (Chen & Chen, 2002). Changes in the expression of WRKY genes specifically in virus-plant interactions, has been observed in *Arabidopsis* infected with either *Tobacco mosaic virus* (TMV), *potato virus X* (PVX), ORMV, CMV, *Turnip mosaic virus* (TuMV) and *Turnip vein clearing virus* (TVCV) (Golem and Culver 2003 ; Marathe et al., 2004; Whitham et al., 2003). Similarly, the same observation has been made in

Plum pox virus (PPV)-infected protoplasts (Babu et al., 2008). The upregulation of WRKY in this system could be to support SACMV-[ZA:99] replication and transcription in the nucleus by inducing increased host factors required for these processes. It is also important to mention that besides being involved in defence responses, WRKYs also participate in various plant physiology responses such as the development of trichomes, seed coat (Johnson et al. 2002) and senescence (Robatzek and Somssich, 2001). Previous reports also indicated that WRKY's are associated with plant sensitivity in response to phosphorus (Chen et al., 2009) and boron (Kasajima et al., 2010) deficiency. Therefore, it can be hypothesised that the downstream genes that are altered or regulated by the WRKY gene family may play an additional regulatory role in geminivirus infection.

3.5.5 Cytoskeleton, endosome, ER or membrane associated fractions

Our results show a large number of host genes that are differentially expressed as a consequence of SACMV-[ZA:99] infection, but of prime interest are those associated proteins which may directly interact with SACMV-[ZA:99] proteins and facilitate intracellular and cell-to-cell movement. Viral movement requires the action of virus-encoded movement proteins to coordinate the replication of the viral genome with its cell-to-cell transport (Lazarowitz and Beachy 1999; Tzfira et al., 2000). SACMV- [ZA:99] encodes the cell-to-cell movement protein (DNA-B encoded BC1-encoded MP) which must interact with a number of host proteins in order to facilitate cell-to-cell and systemic movement through the vascular system. Virus movement and trafficking through the plasmodesmata of host cells have interested many scientists in recent years (Reviewed in Fernández-Calvino et al., 2011; Lucas 2006; Taliansky et al., 2008). It is also clear that plant RNA virus proteins and replication or movement complexes associate with or traffic along the host cytoskeleton and membranes (Harries et al., 2010). While it is evident that bipartite geminivirus BC1 mediates intercellular and cell to-cell transport via the plasmodesmata (Jeske, 2009), how host proteins co-operate in viral DNA-MP or DNA-MP-NSP complexes is still unclear.

3.5.5.1 Myosin and Tubulin

Notably, from our data, we were able to detect the upregulation of two cytoskeleton-associated genes, namely myosin heavy chain and beta-tubulin. The 30K movement protein of TMV has been reported to bind to components of the cytoskeleton (namely actin, myosin and tubulin) to effect transport of viral replication complexes to the plasmodesmata (Harries et al., 2009; Harries et al., 2010). A number of immunolocalization studies have suggested that myosin (and actin) localise at

the plasmodesmata (PD) (Overall et al., 2000; Radford and White 1998; Volkmann et al., 2003), and furthermore, myosin is likely to be involved PD targeting of a number of viral proteins (Avisar et al., 2008; Hofmann et al., 2009; Liu et al., 2005). While the direct interaction between geminiviral proteins and the cytoskeleton-associated proteins has not been demonstrated, it is interesting to note that geminivirus BC1 MPs are classified within the '30K' superfamily involved in many RNA viruses (including TMV) movement (Melcher 2000).

3.5.5.2 Rab-GTPases

In addition to the upregulation of myosin and tubulin, the upregulation of Ras-GTPase (Rab6A) by SACMV-[ZA:99] was an interesting observation, reported for the first time in this study. Rabs are a family of small guanosine triphosphatases (GTPases), and their regulator proteins, the Rab-GTPase-activating proteins are well-characterized regulators that form endomembrane vesicles. They are primarily involved in all aspects of intracellular vesicle trafficking, including vesicle budding, targeting, docking, and fusion (reviewed in Takai et al., 2001). Alternatively, it is possible that SACMV-[ZA: 99] MP targets Rab6A in order to facilitate endosome/vesicle-assisted trans-cytoplasmic movement to the plasmamembrane and release into the PD. BC1 movement proteins from SACMV-[ZA:99] could either be associating with Rab6A either by binding directly to a Rab protein that trafficks to PD, or alternatively, by becoming attached to one or more vesicles that are being delivered to PD by Rab6A (reviewed in Oparka 2004). Carvalho et al., (2008a and 2008b) also have implicated the binding of a cytosolic NSP-interacting GTPase as a cofactor in redirecting geminiviral NSP-DNA complexes to the cell surface from the nucleus and trans-cytoplasmic movement.

3.5.5.3 β -1,3-glucanase and HSP

β -1,3-glucanase was also found to be upregulated in our study by 2.3 fold ($\text{Log}_2 = 1.202$). In a study monitoring *Arabidopsis* leaves infected with five different viruses, Whitham et al., 2003 also observed the induction of the β -1,3-glucanase gene, although their observations were made at early infection at 2, 4 or 5 dpi. It has been reported that an increase in β -1,3-glucanase expression is particularly important for viruses that make use of the movement protein for spreading through the plant (Iglesias and Meins 2000). Furthermore, Levy et al., (2007) and Epel (2009) have presented evidence that viruses may exploit the activity of β -1,3-glucanases to remove callose barriers and increase the size exclusion limit of plasmadesmata. It is important to mention that although HSP70 and heat shock cognate 70 have been grouped under defence response transcripts (Supplementary

Data S3.2), these transcripts have also been implicated in cell-to-cell transport of macromolecules to the plasmodesmata (Aoki et al., 2002). HSP's have been shown to be recruited during infection by a number of viruses to assist genome replication or virion assembly (Aranda et al., 1996 and Glotzer et al., 2000). The homologues of HSP70 of *Beet yellow virus* has been shown to assist movement of the virions through plasmodesmata (Prokhnevsky et al., 2002). It could be hypothesised that HSP is functioning in two diverse roles in this study. Firstly, the plant attempts to mount a defence response against SACMV-[ZA:99] as mentioned earlier, or alternatively, SACMV-[ZA:99] is able to exploit the increasing levels of HSP and use it for its own movement.

3.5.5.4 *Pectin methyltransferase*

Lastly, the downregulation of pectate lyase ($\text{Log}_2 = -1.438$) and 2.2 fold upregulation of pectinesterase ($\text{Log}_2 = 1.178$) was observed in this microarray study. In higher plants, *pectinesterase* or *pectin methylesterase (PME)* play a vital role in cell wall metabolism, plant growth and plant development (Li et al., 2002). In earlier studies, PME has been shown to interact with TMV movement protein which assists the virus with cell-to-cell movement during infection (Chen et al., 2000; Dorokhov et al., 1999; Heinlein 2002). This interaction was confirmed by Chen and Citovsky in 2003, when their work using a reverse genetics antisense approach to reduce PME in tobacco plants successfully demonstrated that reduced levels of PME lead to a considerably delayed systemic movement of TMV. Additionally, the enzymatic activity of PME has been shown to regulate the pH and ion balance of the cell wall which in turn results in a change of cell wall porosity (Nairn et al., 1998). It could be hypothesised that the BC1 movement protein of SACMV-[ZA:99], which mediates cell-to-cell movement of the virus, may be interacting with the increased levels of pectinesterase and in turn is modulating or altering the ion balance of the cell wall. This change in ion balance may induce changes in the permeability of the plasmodesmata therefore assisting SACMV-[ZA:99] in cell-to-cell movement. Therefore, the upregulation of β -1,3-glucanase, HSP70, Rab6A and pectinesterase in SACMV-[ZA:99]-infected leaf tissue may act synergistically to aid movement through the PD, and further work will prove useful in developing a holistic movement model for geminiviruses

3.5.6 *Transcripts primarily map to sucrose and starch metabolism pathway*

Although we have not measured sucrose or starch levels in this study, the assignment of eight genes including sucrose-phosphate synthase, sucrose synthesis 4, α -trehalose-phosphate synthase, hexokinase-2, pectinesterase, UDP-glucuronate-4-epimerase-3 and UDP-glucuronate-4-epimerase-6,

to the sucrose and starch metabolism pathway (P-value = 0.00042) (Fig. 3.6) is significant, and indicates that there may be an accumulation of sucrose in infected leaf tissue. All eight genes were found to be upregulated with $\log_2$ values ranging between 1.087 and 1.474 (Supplementary Data S3.2) An increase in sucrose levels, in particular, in plant tissues as a consequence plant RNA virus infection has already been demonstrated in a number of studies (Shalitin and Wolf, 2000). Starch and sucrose are known to be carbon and energy sources for the host, but have also been implicated in plant defence where it has been recognised as an endogenous signalling molecule used to induce defence responses against pathogens (Bolouri-Moghaddam and Van den Ende 2012). Babu et al., (2008) reported a similar finding where they reported that PPV-infected leave tissue resulted in changes of several metabolic pathways where 23% of genes were associated with metabolism of starch, soluble sugars and amino acids. The increase in sucrose-phosphate synthase is of particular interest as it a key enzyme in plant sucrose biosynthesis (Smith 1999; Subbaiah et al., 2006). In a study by Eybishtz et al., (2009), a reverse genetics approach was applied to identify the role of hexose transport gene (*LeHT1*) in *Tomato yellow leaf curl virus* resistance. Results from this study indicated that TYLCV-infected *LeHT1*-silenced resistant plants became susceptible to TYLCV infection. Furthermore, TYLCV-infected plants became necrotic and silencing of this gene appeared to be connected with TYLCV long-distance movement because vascular bundles of *LeHT1*-silenced resistant plants became necrotic following TYLCV inoculation. We could also hypothesis that the change in sucrose or starch metabolism as a consequence of infection or could be linked to the movement of SACMV-[ZA: 99] as it has already been demonstrated that long-distance movement of viruses is known to occur through the phloem, following the stream of sugar transport (Maule 1991; Gilbertson and Lucas, 1996). Interestingly enough, in a recent study in melon, it has been demonstrated that the constitutive expression of CMV movement protein in transgenic plants, results in a considerable change in carbon metabolism processes and carbon resource distribution in plants (Shalitin et al., 2000 and Shalitin et al., 2002). Since plasmodesmata and sucrose transport to sink organs is intimately linked, it would be interesting to test whether the changes in sucrose and starch metabolism are a non-specific and indirect result of SACMV-[ZA:99] infection or a directed virus-induced alteration to the plasmodesmata.

3.6 Conclusions

By using the Nb-array, we have successfully demonstrated that many of the transcripts associated with a geminivirus-plant interaction are consistent with those observed in similar RNA virus studies dissecting the molecular interaction of a pathogen in a plant host. However, many unique

differentially expressed transcripts, such as Rab6A, and others (Supplementary Data S3.2 and 3.3) from *N. benthamiana* and cassava (homologs), are differentially regulated by our geminivirus, SACMV-[ZA:99]. In order to understand the functional roles that each of these genes, and others not discussed here, play in the SACMV-[ZA:99]-*N.benthamiana* system, it will be necessary to explore infection in *Arabidopsis* knockout mutants or make use of a virus-induced gene silencing (VIGS) and systems biology approaches. The recent release of a draft *N. benthamiana* genome sequence on the Sol Genomics Network (<http://solgenomics.net/>) (Bombarely et al., 2012) will prove invaluable in expanding molecular plant-microbe biology studies in this model host. Understanding a geminivirus-host interactome can inform designing improved broad-based strategies to combat both RNA and DNA virus diseases through identification of unique or conserved host responses.

Chapter 4

Monitoring transcriptional changes in a susceptible and tolerant landrace of
cassava infected with *South African cassava mosaic virus*
using next- generation sequencing technology.

4.1 Abstract

Cassava mosaic disease (CMD) is caused by several *cassava mosaic geminiviruses* and has been the major constraint on cassava production in sub-Saharan Africa and Asia. The identification of genes that are differentially expressed in cassava upon geminivirus infection represents an important step towards understanding the plants defence or susceptibility responses and would assist in designing appropriate control strategies. In this study, a time course (12, 32 and 67dpi) RNA-seq study monitoring *South African cassava mosaic virus*-[South Africa:99] infection in a susceptible (T200) and tolerant (TME3) cassava landrace, was performed using the Applied Biosystems (ABI) SOLiD platform. The multiplexed sequencing run produced a total of 523MB and 693MB of paired-end reads for SACMV-[ZA:99] infected susceptible and resistant cDNA libraries respectively. Of these, approximately 50.7 % of the T200 reads and 55.06% of TME3 reads mapped to the cassava reference genome available at phytozome. Using a 2 fold cut-off ($p < 0.05$), comparative analysis between these six cDNA libraries showed that T200 and TME3 had 4181 and 1008 transcripts, respectively, that were differentially expressed in total across 12, 32 and 67 dpi. GOslim functional groups illustrated that differentially expressed genes in T200 and TME3 were overrepresented in the cellular component category for plasma membrane and nucleus. Alterations in the expression of transcription factors, R genes, genes involved in the DNA methylation pathway and those involved in the putative movement of SACMV-[ZA:99] were selected for further discussion. To our knowledge, this study provides the first time course study of a cassava-geminivirus infection, using high-throughput analysis.

4.2 Introduction

Cassava, *Manihot esculenta* Crantz, is a tropical crop that is important for food security and income generation for poor farmers in several Asian and African countries. Fresh tubers of cassava are suitable for consumption by both humans and animals and provides the most important dietary source of calories for more than a billion people in about 105 countries, providing an estimated one third of calorie intake (FAO, 2008). Cassava's tolerance to unfavourable conditions and abiotic stress make it an excellent crop for small-scale farmers with limited resources, in comparison with other cereals such as wheat, rice and maize (Blagbrough et al., 2010; El-Sharkawy, 2003). Cassava mosaic disease (CMD), is the most important biotic constraint of cassava production in sub-Saharan Africa and is caused by whitefly-transmitted viruses of the genus Begomovirus (family *Geminiviridae*), including *South African cassava mosaic virus*-[South Africa:99] (AF155806) (SACMV-[ZA:99]) (Berrie et al., 2001). SACMV-[ZA:99] has two circular DNA molecules, designated DNA-A and DNA-B, of approximately 2.8 kb (Berrie et al., 2001), both of which are required for systemic infection of plants. CMD has been identified as the major biotic constraint to cassava production in 17 out of 38 main cassava-growing sub-Saharan African countries surveyed (Legg et al., 2006).

In past years, high levels of resistance/tolerance to CMD have been identified in several West African cassava landraces, including TME3 (Akano et al., 2002; Dixon et al., 2001; Fregene et al., 2004). By using classical genetic techniques such as genetic mapping, resistance in several resistant cassava cultivars has been attributed to the presence of a major dominant gene, CMD2 (Akano et al., 2002; Fregene et al., 1997). Furthermore, several molecular markers have been associated with CMD2, including SSRY28, NS158 and RME1 (Akano et al., 2002). Currently, further efforts are being made in order to dissect the genetic architecture of cassava resistance and other economically important traits using an EST-derived SNP and SSR genetic linkage map approach (Rabbi et al. 2012). T200 is a South African landrace, which has been grown recently for its high starch content, but which is highly susceptible to cassava mosaic-infecting geminiviruses (Laminski et al., 1997).

Gene expression technologies such as microarrays represent a well-established technology and have been widely exploited in the last years leading to a vast amount of gene expression information particularly in the area of host-pathogen interactions (Babu et al., 2008; Dardick 2007, Senthil et al., 2005; Whitham et al., 2003). Notwithstanding the economical importance of cassava, particularly in developing countries, it has received comparatively little attention in the scientific community in contrast to the model species *Arabidopsis*, *Nicotiana benthamiana* or crops such as rice, potato and tomato. The only comprehensive global transcriptome profile for a plant geminivirus-host interaction has, to date, been that of *Cabbage leaf curl virus* (CaLCuV) in *Arabidopsis* (Ascencio-

Ibanez et al., 2008). There are only a handful of gene expression studies that have been carried out in cassava. These include studies on storage root formation in cassava (de Souza et al., 2004) a microarray gene expression profile of *Xanthomonas axonopodis* pv. *manihotis* (Xam) strain CIO151 infection (Lopez et al., 2005), and the effect of cold stress on the apical shoots of cassava (An et al., 2012). More recently, a 60-mer oligonucleotide Agilent microarray representing $\pm 20,000$ cassava genes was used to profile three cassava genotypes under abiotic (drought) stress (Utsumi et al., 2012).

The introduction of next-generation sequencing (NGS) in the last few years has provided new and innovative ways to speed up the identification of genes in many plant and animal species, particularly those under biotic and abiotic stresses (An et al., 2013; Shendure and Ji, 2008; Mortazavi et al., 2008; Sharma et al., 2010; Wang et al., 2010; Wang et al., 2012; Zhu et al., 2013). NGS has become the new method of choice for gene expression experiments because of it is an extremely sensitive technique and results in high-throughput data outputs. Furthermore, NGS is a better alternative to microarray technology as it allows one to quantify all transcripts in a given system without the limitation of probes printed on a slide. Several NGS platforms have emerged, including Roche 454, Illumina GA, and Applied Biosystems (ABI) SOLiD. The Roche 454 sequencing system uses a “sequencing-by-synthesis” (SBS) approach, in which a huge number of adapter flanked DNA fragments are bound to the surface of specifically designed beads and amplified by means of emulsion PCR (Dressman et al., 2003). The Illumina GAIIx system generates a large number of clustered amplicons through “bridge amplification”, this occurs by capturing adapter flanked DNA fragments to a flow cell coated by a thick lawn of covalently attached primers, whose sequences are complementary to those of the adapters flanking the DNA fragments (Fedurco et al., 2006). Cluster formation and bridge amplification is also followed by a SBS approach. SBS in both these systems is primed by DNA polymerase. Unlike Roche and Illumina, the SOLiD platform developed by Applied Biosystems, differs in that it uses a “sequencing-by-ligation” approach where adapter-flanked DNA fragments are fixed to the surface of paramagnetic beads and amplified by means of emulsion PCR using DNA ligase (Dressman et al., 2003). Each SOLiD sequencing run takes seven days and generates around 4 terabytes (TB) of raw data which increases substantially after bioinformatics analysis. The operating system used by most researchers is GNU/LINUX (Liu et al. 2012). One of the main advantages of the SOLiD system is that the reads generated are highly accurate, however, the generation of short read lengths (50bp+35bp) or (50bp+50bp) are considered to be a disadvantage when compared to the 454 GS FLX system which generates reads up to 700bp (Liu et al. 2012).

In 2009, the first draft genome assembly of cassava was made publically available through phytozome (<http://www.phytozome.net/>). Whole genome shotgun (WGS) strategy in combination with 454-sequencing technology was used to sequence South American cassava cultivar AM560-2, which is a partially S3-derived inbred line selfed from cultivar MCOL1505, that was generated at the International Centre for tropical Agriculture (CIAT, Prochnik et al., 2012). In addition to the cassava draft genome, there are also a number of additional cassava EST resources available through different databases (reviewed in Ayling et al., 2012). Some of these include the availability of more than 80 000's ETS through Genbank as well as two further large EST libraries containing between 20 000 and 30 000 Sanger reads which was generated as a collaborative effort between RIKEN (Rikagaku Kenkyusho—Institute of Physical and Chemical Research, Japan), and CIAT (Centro Internacional de Agricultura Tropical). The availability of these resources and easy accessibility to this sequence information will facilitate the rapid data analysis of many future gene expression studies carried out in cassava as well as accelerate many marker-assisted breeding programs therefore permitting improvements in disease-resistance and tolerance in many cassava cultivars. Genetic information will also help provide a better understanding of the genetic basis underlying disease resistance.

To our knowledge, there has been no global gene expression studies reported for geminivirus (GV) infection in cassava. With the availability of the cassava genome, compounded by the limited information regarding GV infection in cassava, the main aim of this study was to elucidate the gene expression responses of susceptible (T200) and tolerant (TME3) cassava to SACMV-[ZA:99] infection. The data presented in this chapter is the first comparative report of global expression changes in cassava infected with a CMD-infecting GV. SOLiD transcriptomic data from six different SACMV-[ZA:99] infected cDNA libraries, was successfully generated by SOLiD sequencing and used to compare expression patterns between a susceptible and tolerant cassava cultivar at three timepoints post inoculation. RNA-Seq data generated from the SOLiD platform was assembled and reference-based mapping against the cassava genome was carried out. Since the cassava genome is publically accessible and available through phytozome, this facilitated easier identification of the paired-end reads generated during this study. The data obtained demonstrates that differential responses occur to SACMV-[ZA: 99] infection between the susceptible T200 and tolerant TME3, and that TME3 had a significantly lower number of altered transcripts compared with T200.

4.3 Materials and Methods

4.3.1 Micropropagation and acclimatization of cassava

T200 and TME3 cassava cultivars were micropropagated by nodal cutting culture on Murashige and Skoog (MS) medium (Murashige and Skoog, 1962) supplemented with 20 g/L sucrose and 7.8g/L plant agar (Sigma Aldrich), pH 5.8. Cassava explants were allowed to grow at 25°C under a 16 hour 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 growth chamber (28°C; 16 hour photoperiod). Plantlets were gradually acclimatized by adding slits to plastic film. Acclimatized plantlets were allowed to grow until they reached a 4-6 leaf stage.

4.3.2 Agroinoculations of T200 and TME3 plantlets

Agroinoculations of T200 and TME3 cassava plantlets were achieved by an agroinoculation protocol adapted from Hayes *et al.*, (1988). Infectious, head-to-tail, dimers of SACMV-[ZA:99] DNA-A and DNA-B were previously cloned separately into binary vector pBIN19 (Berrie *et al.*, 2001) and transformed into *Agrobacterium tumefaciens* Ag/1. The two transformed cultures containing DNA-A and DNA-B were cultured separately in Luria Bertani (LB) Broth supplemented with carbenicillin (100µg.ml⁻¹) and kanamycin (100µg.ml⁻¹). Wild-type *Agrobacterium* Ag/1 cultures served as a negative control for inoculations and was inoculated into LB broth supplemented with carbenicillin (100µg.ml⁻¹). Cultures were grown overnight at 30°C until optical densities of 1.8-2.0 (OD₆₀₀) were reached. From each of the three cultures, 5ml was sub-inoculated into 30ml fresh LB Broth, containing the correct combination of antibiotics as previously described. Cultures were once again grown overnight at 30°C until cultures reached optical densities of 1.8-2.0 (OD₆₀₀). For each culture, 25ml aliquots were pelleted by centrifugation at 13000xg, washed in sterile distilled water and subsequently resuspended in 5 ml LB Broth. Ag/1 -SACMV-[ZA:99]-DNA-A and Ag/1 -SACMV-[ZA:99]-DNA-B were resuspended and combined to form a homogenous mixture of Ag/1 - SACMV-[ZA:99]-DNA-A and Ag/1 - SACMV-[ZA:99]-DNA-B cells. T200 and TME3 plantlets were wounded along the stem with a hypodermic needle and each plantlet was inoculated with 100 µl the Ag/1 DNA-A/DNA-B suspension using a 1 ml Hamilton syringe. Control plants were mock-inoculated with 100 µl wild-type untransformed *Agrobacterium* Ag/1 inoculum.

4.3.3 Experimental procedure and sample collection

SACMV-[ZA:99]-infected and mock-inoculated plants were monitored over a 67 day period. Two newly developed and fully expanded symptomatic leaves were collected below the apex from each plant (n=6) at each timepoint i.e 12, 32 and 67dpi, and pooled. Time points were however kept separate and therefore a total of six SACMV-infected samples were used in down stream sequencing (i.e 12, 32 and 67 dpi for T200 and 12, 32 and 67 dpi for TME3). The same procedure was carried out on mock-inoculated leaf tissue at the same timepoints therefore resulting in six samples of mock-inoculated controls. Leaf tissue was immediately frozen in liquid and stored at -80°C until further use for DNA and RNA extractions.

4.3.4 DNA extractions from leaf tissue

For each time point (12, 32 and 67dpi), the leaves closest to the apex were harvested from six plants, representing 6 biological replicates, for both infected and mock-inoculated plants. Total nucleic acid (TNA) was isolated from these SACMV-[ZA:99] infected and mock-inoculated leaves using a modified CTAB-based extraction method (Doyle and Doyle, 1990). 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 hour. The suspension was then purified twice by a chloroform: isoamyl alcohol (24:1) solution and precipitated with isopropanol. The TNA was recovered at 13000xg at 4 °C for 10min. 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).

4.3.5 Confirmation of SACMV-[ZA:99] infection using conventional PCR

Systemic infection in cassava leaf tissue for T200 and TME3 at 12, 32 and 67dpi was confirmed by conventional PCR. 50 µl PCR reactions were set up and contained 0.4 µM of each primer, 200 µM dNTPs, 2 units DreamTaq DNA polymerase (Fermentas), 1x DreamTaq Buffer (Fermentas), and nuclease-free water to a final volume of 50 µl. A 550 bp fragment of the core coat protein (CCP) on SACMV-[ZA:99] DNA-A was amplified using degenerate forward primer: (V524) 5' GCCHATRTAYAGRAAGCCMAGRAT 3' and reverse primer: (C1048) 5'

GGRTTDGARGCATGHGTACANGCC 3'. Approximately 500 ng of the total nucleic acid (TNA) template was added to the reaction mixture. Reactions were cycled in a MyCycler™ Thermal Cycler (Bio-Rad) at 95 °C for 5 minutes to activate the Taq DNA polymerase, followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at 55 °C for 30 seconds, primer extension at 72 °C for 45 seconds, and a final extension step of 72 °C for 5 minutes. DNA-A of SACMV-[ZA:99] cloned into pBluescript vector (50ng) was used as positive control for PCR reactions. Amplification products were examined by electrophoresis on a 1.2% agarose TAE gel containing 10 µg/ml ethidium bromide.

4.3.6 Quantitative Real-time PCR (qPCR) of SACMV-[ZA:99] DNA-A

Determination of the viral titre in T200 and TME3 plants was achieved by use of qPCR on TNA extracted from both cultivars at time points 12, 32 and 67dpi. TNA samples were all standardised to a concentration of 100ng/µl. Duplicates of each sample were prepared as well as a no template control (NTC) of nuclease-free water. For each sample, a 20µl reaction was set up in LightCycler capillaries containing 1µl of 100ng of leaf tissue TNA added to 4µl LightCycler® FastStart DNA Master^{Plus} SYBR Green I (Roche), 1µl forward coat protein primer (10µM) 5'ACGTCCGTCGCAAGTACGAT3', 1µl reverse coat protein primer (10µM) 5'ATTGTCATGTGAATAGTACG 3' and 14µl nuclease-free water. A 150bp fragment was amplified and quantified using the following amplification conditions: 95°C for 10 min, followed by 35 cycles of 95°C for 10 sec, 60°C for 10 sec, 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 166bp PCR product of ubiquitin was amplified from 100ng 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, (2011). Data was exported to Microsoft Excel for analysis statistical using the Students T-test.

4.3.7 RNA extractions

Total RNA was extracted from SACMV-[ZA:99]-infected and mock-inoculated leaf tissue using a modified high molecular weight polyethylene glycol (HMW-PEG) protocol (Gehrig et al., 2000). In order to minimise variability, the leaves were harvested at each time point from mock inoculated and infected plants at the same time of day on collection days as circadian rhythms have been

shown play a role in regulating many defence response genes. In brief, 1 g leaf tissue, for each biological replicate, was homogenised in liquid nitrogen and added to 5ml preheated (65°C) GHCL buffer (6.5M guanidium hydrochloride, 100 mM Tris-HCl pH 8.0, 0.1 M sodium acetate pH 5.5, 0.1 M 2-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 (10000xg) for 10 minutes at 4°C. The supernatant was removed and RNA was precipitated by adding 5ml isopropanol (Sigma). The mixture was thoroughly mixed and incubated at -20°C for 60 minutes and pelleted by centrifugation (10000xg) for 25 minutes at 4°C. RNA pellets were washed with 5ml ice-cold 75% 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). Concentrations were determined using the NanoDrop and RNA integrity was assessed using an Agilent 2100 Bioanalyzer.

4.3.8 cDNA library preparation and sequencing

cDNA libraries were generated at the Functional Genomics Center UNI ETH Zurich, Switzerland. Briefly, 12µg of total RNA for each sample was used to generate cDNA libraries. RNA was fragmented and subjected to hybridization and ligation using the SOLiD Total RNA-Seq Kit (Applied Biosystems) according to the manufacturer's instructions. cDNAs were selected by size on a polyacrylamide gel before and after the library amplification. A total of 12 libraries were multiplexed using the SOLiD RNA Barcoding Kit (Applied Biosystems) and pooled in an equimolar ratio. The samples were then diluted and used for emulsion PCR. Beads containing a multiplex of 12 samples were deposited onto a single flow cell. Libraries were sequenced operating on 50bp forward and 35 bp reverse paired-end sequencing chemistry on the ABI SOLiD V4 system.

4.3.9 Bioinformatics: assembly, mapping and annotation

The SOLiD v4 sequencer was used for the generation of sequence reads and was run in paired-end mode (50+35bp). This resulted in two csfasta and two quality files per sample. The reads generated for each library were mapped to the genome assembly (Phytozome, *Manihot esculenta* 147) using the Lifescope software from LifeTech. As a result, SAM/BAM alignment files were prepared, sorted and indexed using samtools (<http://samtools.sourceforge.net/>). In the secondary data analysis phase, the BAM data were matched with the genome annotations available in Phytozome as a GTF/GFF3 file, which describes genes, transcripts and their exons with the genomes coordinates. The

alignments were then transformed to counts using rnaSeqMap library of Bioconductor (Leśniewska and Okoniewski, 2011). The count table for all genes from the annotation were analyzed using DESeq (Anders and Huber, 2010). In order to identify and quantify the number of differentially expressed genes common between timepoints 12, 32 and 67dpi in each landrace, data was imported into SQL 2012 where 'inner join' and 'left join' queries were executed using the cassava transcript ID number as the unique feature used to identify all of the genes common between timepoints.

4.3.10 RT-qPCR validations for genes differentially expressed in T200 and TME3

Eight genes (5 from T200 and 3 from TME3) that were found to be differentially expressed were selected based on the SOLiD RNA-seq results (i.e., >2- fold change, $p < 0.05$) and analyzed using real-time quantitative RT-PCR. One of the criteria used to select genes, was the differential expression observed in at least 2 of the 3 timepoints in T200 and TME3 SACMV-infected leaf tissue. Primers for each gene were designed using software available online through Integrated DNA technologies (IDT, <http://www.idtdna.com/Primerquest/Home/Index>, Table 4.1). In brief, 1 µg of DNase-treated total RNA was reverse transcribed using the Improm-II-reverse transcriptase kit (Promega, Madison, WI) according to manufacturer's instruction. RNA, dNTPs and Oligo dT₁₈ primer were denatured for 10 min at 70°C; then kept at 25 °C for 5 min before the reverse transcription master mix was added. Reverse transcription was performed at 42°C for 1 hour followed by a 10 min incubation step at 70°C. Control reactions were set up without the addition of reverse transcriptase and used as negative controls in the real-time PCR study. RT-qPCR experiments were conducted on the Lightcycler 1.5 for all genes using the appropriate primer pair for each reaction (Table 4.1). Relative quantification standard curve method (Larionov et al., 2005) was used to calculate the relative expression changes in each of the 8 genes assessed. Standard curves were generated for each gene using a 10-fold serial dilution of cDNA reverse transcribed from RNA extracted from either healthy T200 or TME3 leaf tissue. All reactions were based on the following recommended protocol using 0.5µl of each primer and 1µl of template per reaction. In brief, all qPCR reactions were performed in LightCycler® capillaries using the LightCycler 1.5 using LightCycler® FastStart DNA Master^{Plus} SYBR Green I kit (Roche). Three biological replicates and two technical replicate were run for SACMV-[ZA:99]-infected and mock-inoculated leave cDNA samples for T200 or TME3 at 12, 32 and 67 dpi. One µl of undiluted cDNA was used for each reaction. The cycling conditions used were as follows: initial denaturation for 10 min at 95°C (hot start) followed by an amplification and quantification cycle repeated 35 times, each consisting of 10 sec denaturing at 95°C, 10 sec annealing at primer

specific temperatures, 15 sec primer extension at 72°C with a single fluorescence measurement. Melting curve cycle was obtained by heating to 65°C for 15 s with a heating rate of 0.1°C per second with a continuous fluorescence measurement. UBQ10 (Moreno et al, 2011) was the gene used as an endogenous control for normalization. Statistical analysis was carried out in Microsoft Excel using the Student's T-test.

Table 4.1 Primers used for RT-qPCR validations of SOLiD data for T200 and TME3

Gene Name	Phytozome Accession number	Forward primer (5'-3')	Reverse primer (5'-3')
Cellulose synthase	cassava4.1_001280m.g	AGCAGCTGAAGGAACCTGGAGACA	AGCCATCGCATCCTTCATGGTTTG
Thaumatococin-like protein	cassava4.1_011211m.g	TGACTCCACCCTTGGCATCTTTGA	TGTGTTTACCGGCACCTCCTCTGT
PHE-ammonia lyase(PAL)	cassava4.1_002591m.g	GCATCTTTGCGTACACTTTGCT	ACATAAGCCCCAAAAGCTAGCTGGA
Cyclin P4	cassava4.1_016519m.g	GACAATGGAACAGGAATGGTTAAG	CTCAAGAGTCCGAGCAACATAG
Plant Invertase	cassava4.1_034263m.g	ACTGTCTCAACCTGTTGCCGGG	GGCTGCAGCAGAAGCTGACAAAGAA
JAZ protein 10	cassava4.1_016821m.g	TGCAGAACTGGTCGCCCCAAGG	TGCAGAACTGGTCGCCCCAAGG
Cytochrome P450	cassava4.1_029565m.g	TCAGGCTGACTGTGCTGTTCTCAT	GACACCAAGGGTGAAAGCAAGCAA
Rubisco methyltransferase	cassava4.1_006869m.g	GGCAAATAGGAGCCAATGAA	GGCAAATAGGAGCCAATGAA
UBQ10 * (endogenous control)	DV441403	GCGCCAAATTCAAGGACGAGAAGT	ACAGCCTCGGCCTTCTTAATCACA

4.4 Results

4.4.1 Comparison of CMD disease infectivity assays in T200 and TME3

Following agroinoculation of T200 and TME3, plantlets were monitored over a 67 day period for symptom development (Fig. 4.1) and concentrations of virus (Fig. 4.2). Based on trial SACMV-[ZA:99] infections, timepoints chosen for this study represent different phases of infection where 12 dpi represents early infection (pre-symptomatic), 32 dpi represents active replication (symptomatic) and 67 dpi represents a later infection stage (persistently symptomatic in T200 and recovery in TME3). The symptom severity score index (Fauquet and Fargette, 1988) was used as a guideline for the assessment of symptom development in cassava plants. The mean symptom severity (n=6) scores calculated for T200 at 12, 32 and 67 dpi showed increasing development of symptoms over time. At 12 dpi, the mean symptom severity score indicated that plants were asymptomatic (score of 1). By 32 dpi, symptoms developed uniformly on SACMV-[ZA:99] infected plants which displayed typical mosaic and mild leaf distortion (Fig 4.1 B) associated with CMD infection and the mean symptom severity score of 3.5 indicated that plants were showing moderate symptoms. At 67 dpi, plants were fully symptomatic (Fig 4.1 C) and the mean symptom severity score of 4. Although the West African landrace TME3, possessing a marker-linked CMD2 resistant gene (Akano et al., 2002), is reported to exhibit resistance to CMD, we observed that TME3 does in fact become infected by SACMV-[ZA:99] in the laboratory using agroinoculation with SACMV-[ZA:99] infectious clones. Furthermore, we observed that late infection (approximately 50 dpi onwards) results in leaf tissue becoming asymptomatic and displaying a recovery phenotype, suggesting that TME3 is more appropriately defined as a tolerant landrace rather than resistant. This is confirmed by both symptom severity indexing of infected TME3 as well as viral titre described later. The mean (n= 6) symptom severity scores calculated for TME3 at 12, 32 and 67 dpi showed a different trend to that observed in T200 plants, where the mean symptom severity score peaked at 32 dpi with leaves visibly showing mosaic and distortion of leaf margins (Fig 4.1E). At 32 dpi and 67dpi, the mean symptom severity score of 3.0 and 3.5, respectively, indicated that plants were displaying milder symptoms as compared to T200 at the same timepoints. Interestingly, newly emerging leaves on some plants (every 2 out of six plants) showed an attenuation of symptoms and had lower symptom severity scores (between 0 and 1) at 67 dpi (Fig 4.1 G).

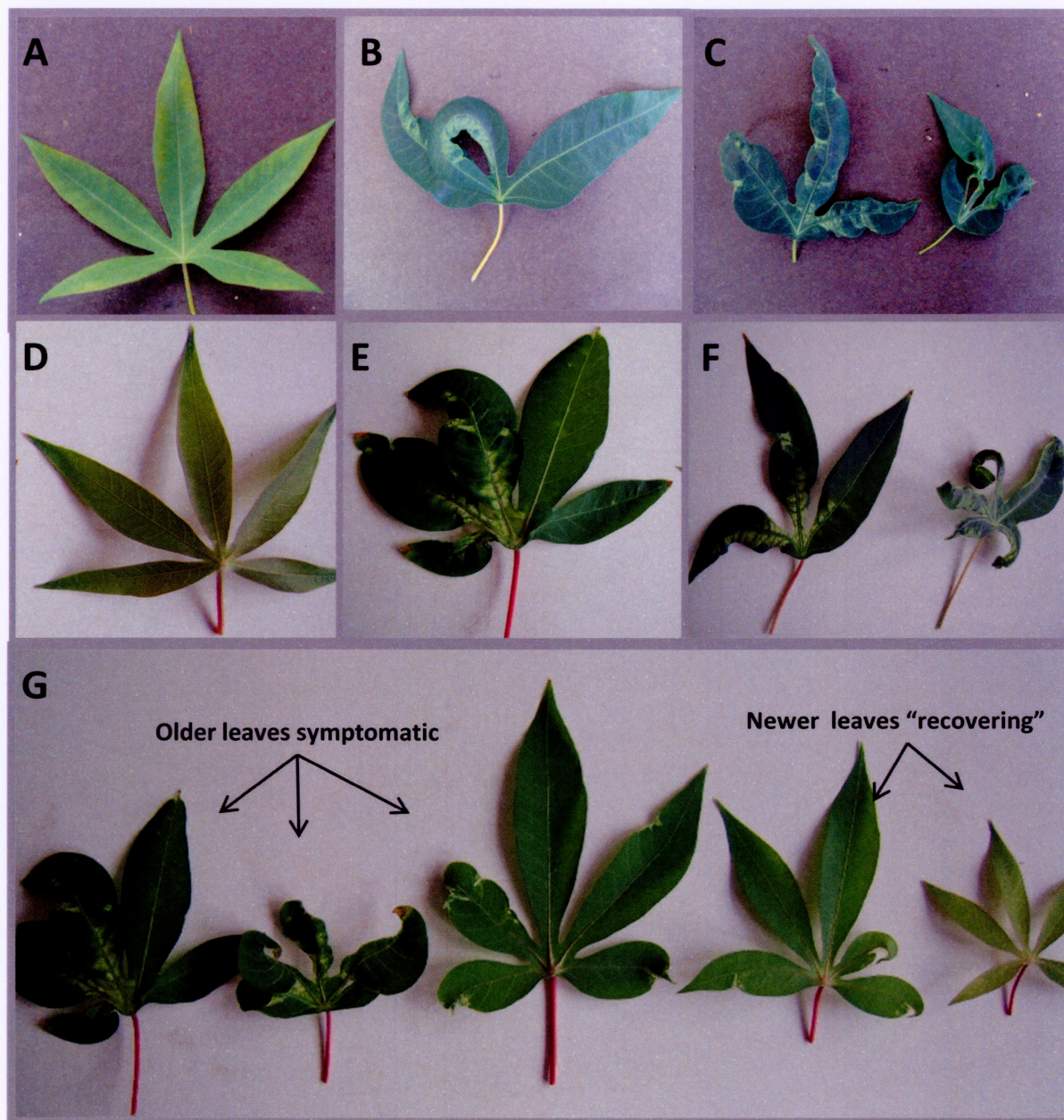


Figure 4.1 T200 and TME3 infected leaf tissue was evaluated for the development of symptoms over a 67 day period. Leaves for both T200 (A) and TME3 (D) at 12dpi (early infection) appear symptomless. Leaves at 32dpi for T200 (B) and TME3 (E) display characteristic mosaic, distortion of leaf margins and leaf curling. Leaves at T200 (C) and TME3 (F) are severely reduced in size. Newer emerged leaves on some cassava TME3 (G) plants (on average, 2 out of every 6), displayed an attenuation in symptoms and almost appear symptomless.

4.4.2 SACMV viral titre in T200 and TME3

The concentration of SACMV-[ZA:99] DNA was accurately measured in infected and mock-inoculated T200 and TME3 plants at 12, 32 and 67 dpi (n=6) using real time qPCR (Fig. 4.2). A technical replicate was included for each biological replicate. For susceptible T200, the concentrations of DNA-A in 12 dpi were extremely low and almost undetectable at a concentration

of 0.14×10^1 , and at 32 dpi 2.19×10^3 SACMV-[ZA:99] molecules of DNA-A were present, whereas 4.43×10^5 SACMV-[ZA:99] molecules of DNA-A were detected at 67 dpi. For tolerant TME3, concentration of DNA-A were significantly lower (t-test, $p < 0.05$) than those detected in T200 where 1.79×10^2 SACMV-[ZA:99] molecules of DNA-A were present at 32dpi and 3.23×10^3 SACMV-[ZA:99] molecules of DNA-A were detected at 67 dpi (Fig 4.2). Overall, viral titre in T200 between 32 and 67 dpi was 10-fold higher than that observed in TME3 at the same timepoints. These concentrations correlated well with the mean symptom severity score recorded for both landraces.

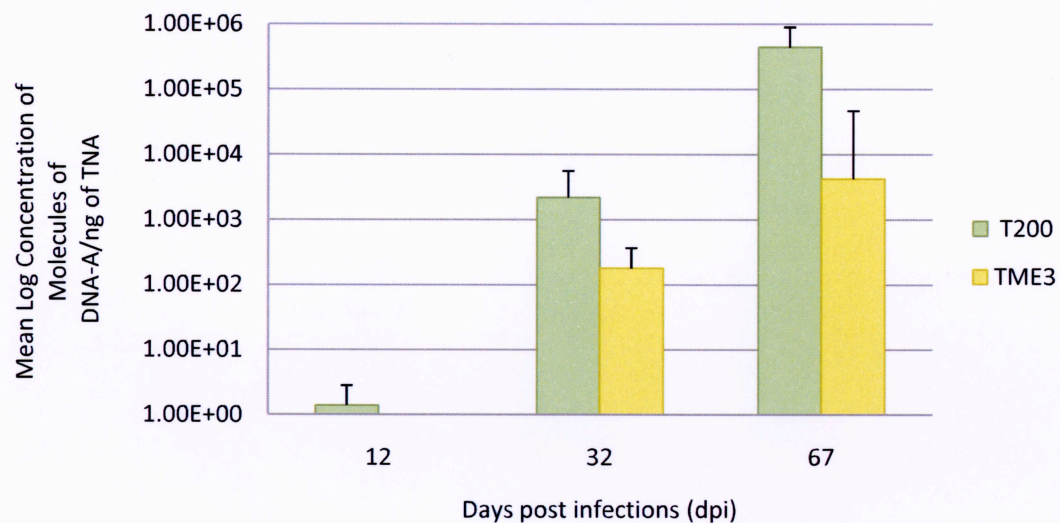


Figure 4.2 Viral quantification from SACMV-[ZA:99] –infected T200 and TME3 leaf tissue. The mean Log concentration of DNA-A molecules/ng TNA for 12, 32 and 67dpi in infected leaf tissue samples were calculated using Avagadro’s constant (calculation shown is Chapter 2).

4.4.3 Transcriptome Data

Sequence reads were generated using the SOLiD v4 sequencing platform in order to generate a gene expression profile of T200 and TME3 infected with SACMV-[ZA: 99]. The sequencer was run in the paired end mode with 50bp forward (F3 tag) and 35bp reverse (F5 tag). Forward and reverse pairs were mapped to reference genome *Manihot esculenta* 147 available through phytozome (www.phytozome.net/cassava) and percentages were calculated for each F3 and F5 mapping combination (Table 4.2). In general, for tolerant TME3 library, an average of 23.41% of both the forward and reverse reads mapped to the reference, 22.74% of the forward read mapped only but the F5 did not, 6.50% of the reverse read mapped, but the F3 did not and finally 47.19% of F3+F5 reads did not map at all. Similarly, for T200, an average of 23.79% of both the forward and reverse reads mapped to the reference, 22.19% of the forward read mapped only but the F5 did not, 5.91%

of the reverse read mapped, but the F3 did not and finally 48.11% of F3+F5 reads did not map at all. Normalization was carried out as an averaging geometric mean of replicates for each library. Normalized data was then imported into DESeq R software package where the counts for differentially expressed genes were calculated using the negative binomial distribution estimated from the complete dataset. Cassava transcripts indentified as differentially expressed were annotated using the *Mesculenta* 147 annotation info file available from phytozome.

Table 4.2 Pairing statistics for the total number of bases and forward and reverse tagged sequences that were mapped to the cassava draft genome (version 4.1) for T200 and TME3 libraries.

	T200 infected 12dpi	T200 Infected 32dpi	T200 infected 67dpi	T200 mock-inoculated 12dpi	T200 Mock-inoculated 32dpi	T200 Mock-inoculated 67dpi	TME3 infected 12dpi	TME3 Infected 32dpi	TME3 infected 67dpi	TME3 Mock-inoculated 12dpi	TME3 Mock-inoculated 32dpi	TME3 Mock-inoculated 67dpi
Total number of bases in paired reads	748,005,833	312,918,869	484,858,536	607,626,022	534,428,823	633,895,179	634,953,645	567,510,203	878,731,651	766,608,396	651,492,969	787,945,631
F3 and F5 mapped (%)	32.16	11.28	20.32	33.44	18.8	26.79	22.06	22.91	31.5	18.4	21.28	25.28
F3 mapped, F5 unmapped (%)	22.87	22.99	22.57	21	21.66	22.02	23.23	22.87	22.84	21.65	22.86	22.99
F3 unmapped, F5 mapped (%)	5.82	6.9	7.19	4.19	5.15	6.23	7.27	6.32	6.18	6.5	6	6.71
F3 and F5 unmapped (%)	39.15	58.83	49.93	41.38	54.39	44.96	47.44	47.89	39.48	53.43	49.86	45.01
Total count of mapped reads	200,801,05	163,420,84	167,942,02	150,030,93	184,254,97	183,472,80	217,820,32	182,493,03	238,740,00	270,717,82	215,499,20	249,149,44
Total reads mapped (%)	60.85	41.17	50.08	58.63	45.61	55.04	52.56	52.1	60.52	46.65	50.14	54.89

4.4.4 Global expression profiles of T200 and TME3 after SACMV-[ZA:99] inoculation

The background genetics of a host plant and the virulence of the invading pathogen, as well as the compatibility between the two, will determine the combination of physiological, metabolic and phenotypic changes that occur as a result of infection. We used a SACMV-[ZA:99]–susceptible and -tolerant landrace in a transcriptomics study to investigate and compare different gene expression alterations following infection over time. In order to quantify the differential expression of genes at 12, 32 and 67 dpi, we computed the tag count for all genes at 12, 32 and 67 dpi versus the tag counts the same timepoint in mock-inoculated samples. This allowed us to calculate the change in expression between SACMV-[ZA:99]-infected and mock-inoculated leaf tissue samples at all three timepoints for both cultivars. After statistical filtering of the data (2-fold cut-off, $p < 0.05$), the total number of differentially expressed genes (DEGs) were identified as SACMV-[ZA:99] responsive genes for T200 (Supplementary data S4.1 –S4.4) and TME3 (Supplementary data S4.5 –S4.8). Overall, the number of differentially expressed genes in TME3 infected with SACMV-[ZA:99] were less than that observed for T200 plants. Of these, 632 DEGs were detected at 12 dpi in T200 where 417 genes were upregulated and 215 genes were downregulated. At 32 dpi, this number increased to 1763 where 742 genes were upregulated and 1021 genes were downregulated and at 67 dpi, a total of 1786 DEGs were detected where 991 genes were upregulated and 795 were downregulated. For TME3, a considerably lower number of DEGs were identified as compared to the susceptible T200 cultivar. For early response at 12 dpi, only 251 DEGs were found where 63 were upregulated and 188 were downregulated. At 32dpi 461 DEGs occurred where 294 genes were upregulated and 167 were downregulated, and at 67 dpi, 290 genes were altered where 88 genes were upregulated and 202 genes were downregulated. Genes that were consistently expressed across timepoints (Fig 4.4 A and B) in each cultivar, and comparisons between timepoints (Fig 4.3 C-E) across cultivars, are represented in Venn diagrams. For T200 (Fig 4.4 A), the largest overlap between timepoints was observed where 252 genes were common between 12 and 32 dpi, 281 genes were common between 12 and 67 dpi and 812 genes were common between 32 and 67dpi, and 209 genes between all three timepoints. For TME3 (Fig 4.4 B), the overlap was considerably smaller, where only 30 genes were common between 12 and 32dpi, 18 genes between 12 and 67dpi, 30 genes between 32 and 67dpi and only 5 between all three timepoints.

To evaluate the potential functions of genes altered at the transcription level in T200 and TME3, Gene Ontology (GO) analysis was conducted using the Arabidopsis Information Resource (TAIR10, <http://www.arabidopsis.org/tools/bulk/go/index.jsp>) for three components ('biological process', 'molecular function' and 'cellular component'). Arabidopsis accessions, corresponding to cassava

transcripts, were used for this analysis. Data has been shown in Fig 4.4. For the cellular component category (Fig 4.4 A), genes were found to be primarily assigned to the plasma membrane ($\pm 20\%$ of T200 genes, $\pm 15\%$ of TME3 genes) and the nucleus ($\pm 40\%$ of T200 genes, $\pm 35\%$ of TME3 genes). For biological processes category (fig 4.4 B), cell organization and biogenesis was over-represented ($\pm 18\%$ for T200, $\pm 10\%$ for TME3), and in addition, noticeable changes in transcription, signal transduction and response to stress and, response to abiotic and biotic stimulus was evident. Lastly kinase activity and hydrolase activity seemed to be overrepresented at 32 and 67 dpi , respectively, in TME3 plants but not in T200 at the same timepoints.

4.4.6 KEGG pathway analysis

Virus infection has been shown to disrupt the highly ordered primary metabolism of the host plant. KEGG pathway analysis was carried out for T200 and TME3 for commonly regulated transcripts using DAVID (<http://david.abcc.ncifcrf.gov/>). Arabidopsis accession numbers for the annotation of cassava transcripts in phytozome were used for this analysis. The small set of genes commonly expressed for TME3 across timepoints did not map to any significant pathways, but T200 had 11% of transcripts mapping to three main pathways. These included: flavanoid biosynthesis (10 genes, $P = 1.2E-9$), biosynthesis of phenylpropanoids (18 genes, $P = 0.01$), phenylpropanoid biosynthesis (9 genes, $P = 0.014$), and stilbenoid, diarylheptanoid and gingerol biosynthesis (6 genes, $P = 0.051$).

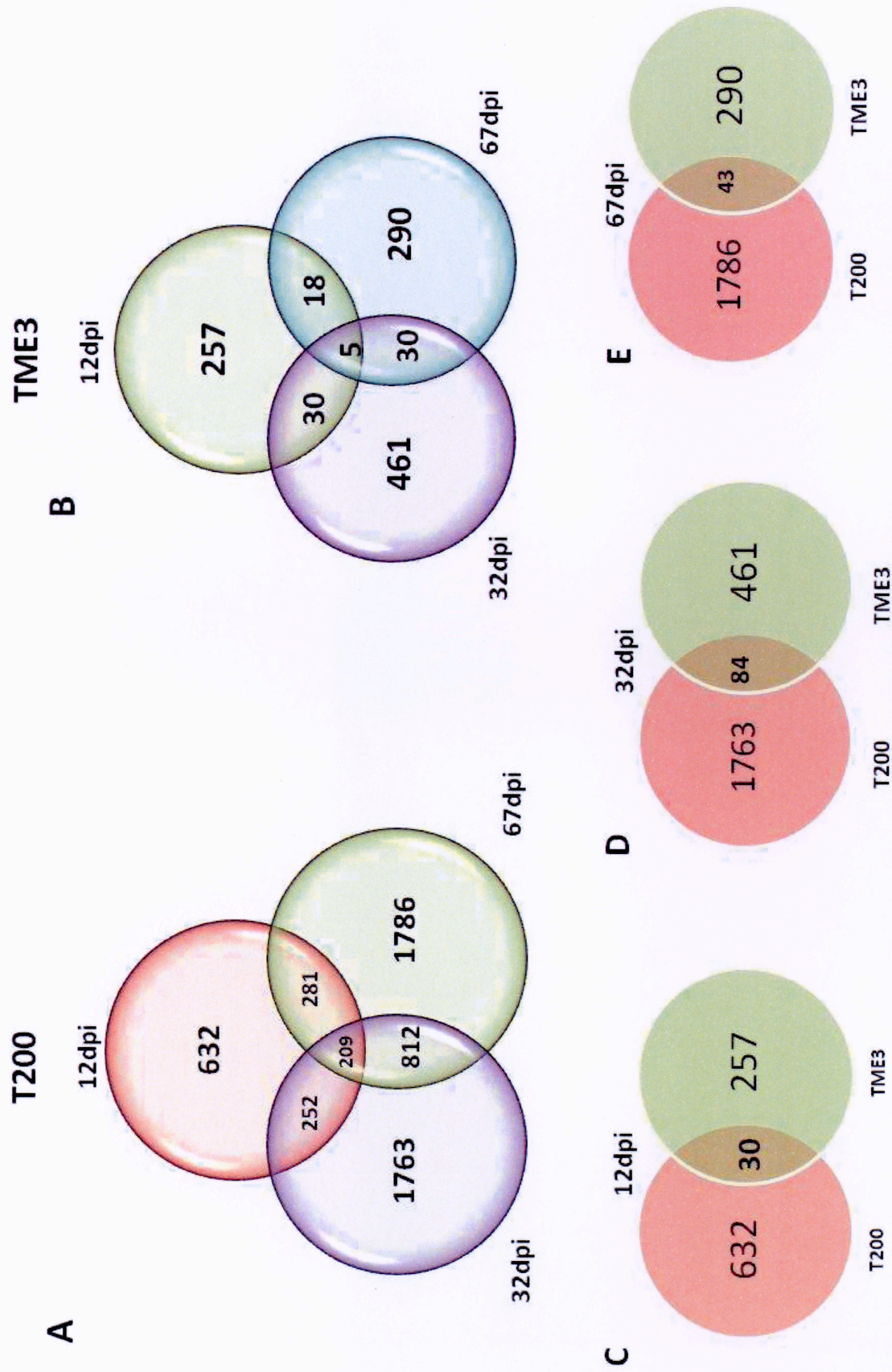


Figure 4.3 Venn diagrams showing the differential distribution of differentially regulated transcripts (2 fold up or down regulated) in SACMV-[ZA:99] infected T200 (**A**) and TME3 (**B**) leaf tissues at three different time points. A statistical cut-off of $p < 0.05$ was used to determine differentially expressed genes (DEGs). Number of DEGs shown in the non overlapping sectors represents unique significant DEGs at each time point, whereas DEGs at an intersection of sectors represent the number of shared genes between the given timepoints. Comparisons between the number of overlapping DEGs at timepoints 12dpi (**C**), 32dpi (**D**) and 67dpi (**E**) between T200 and TME3 were also shown

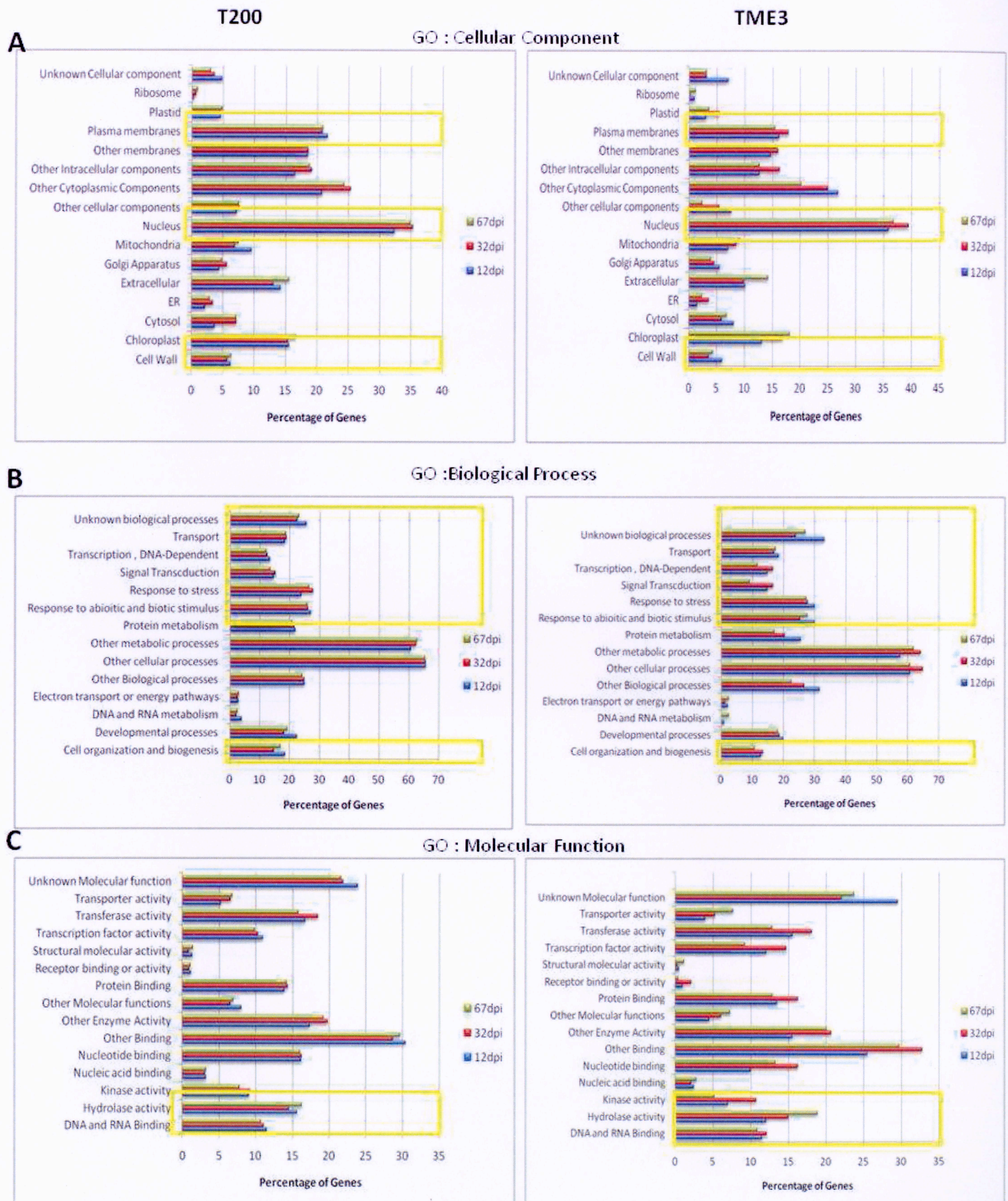


Figure 4.4 : GOSlim Functional Characterization of T200 and TME3 DEGs at 12, 32 and 67dpi for cellular component (A), biological process (B) and molecular function (C). Yellow demarcated areas indicate the most noticeable changes in the percentage of DEGs categories between landraces.

4.4.5 Validation of data by real time quantitative RT-qPCR

To validate the SOLiD RNA-seq data, RT-qPCR was performed on eight (5 from T200 and 3 from TME3) genes that were significantly changed upon SACMV-[ZA:99] infection (>2 -fold, $p < 0.05$). The expression levels for cellulose synthase (Fig. 4.5 A), cyclin p4 (Fig. 4.5 B), PHE-ammonia lyase (Fig. 4.5 C), plant invertase (Fig. 4.5 D), thaumatin PR protein (Fig. 4.5 E), cytochrome P450 (Fig. 4.5 F), JAZ protein 10 (Fig. 4.5 G) and Rubisco methyltransferase (Fig. 4.5 H) were independently validated on cDNA samples of 12, 32 and 67 dpi previously used for the SOLiD RNA-seq study. The standard curve method (Larionov et al., 2005) was used to determine expression values for each target gene from SACMV-[ZA:99] infected leaf tissue at each timepoint in relation to the expression of the same target in mock-inoculated leaf tissue. Relative expression values for each target gene were then expressed as a Log_2 ratio of target gene expression level to *UBQ10* expression level measured in the same cDNA sample. Therefore, expression levels are presented as the relative Log_2 ratio of the infected cassava leaf tissue sample compared with the control mock-inoculated sample at each timepoint. All eight genes showed similar Log_2 gene expression patterns in agreement to those observed in SOLiD sequencing data (i.e. upregulated genes were confirmed to be upregulated and down regulated genes were confirmed to be downregulated). Since such a small number of genes were validated, particularly from the TME3 landrace, it would be beneficial to do additional assays on more genes especially since the SOLiD RNA-seq study produced such high-throughput data outputs.

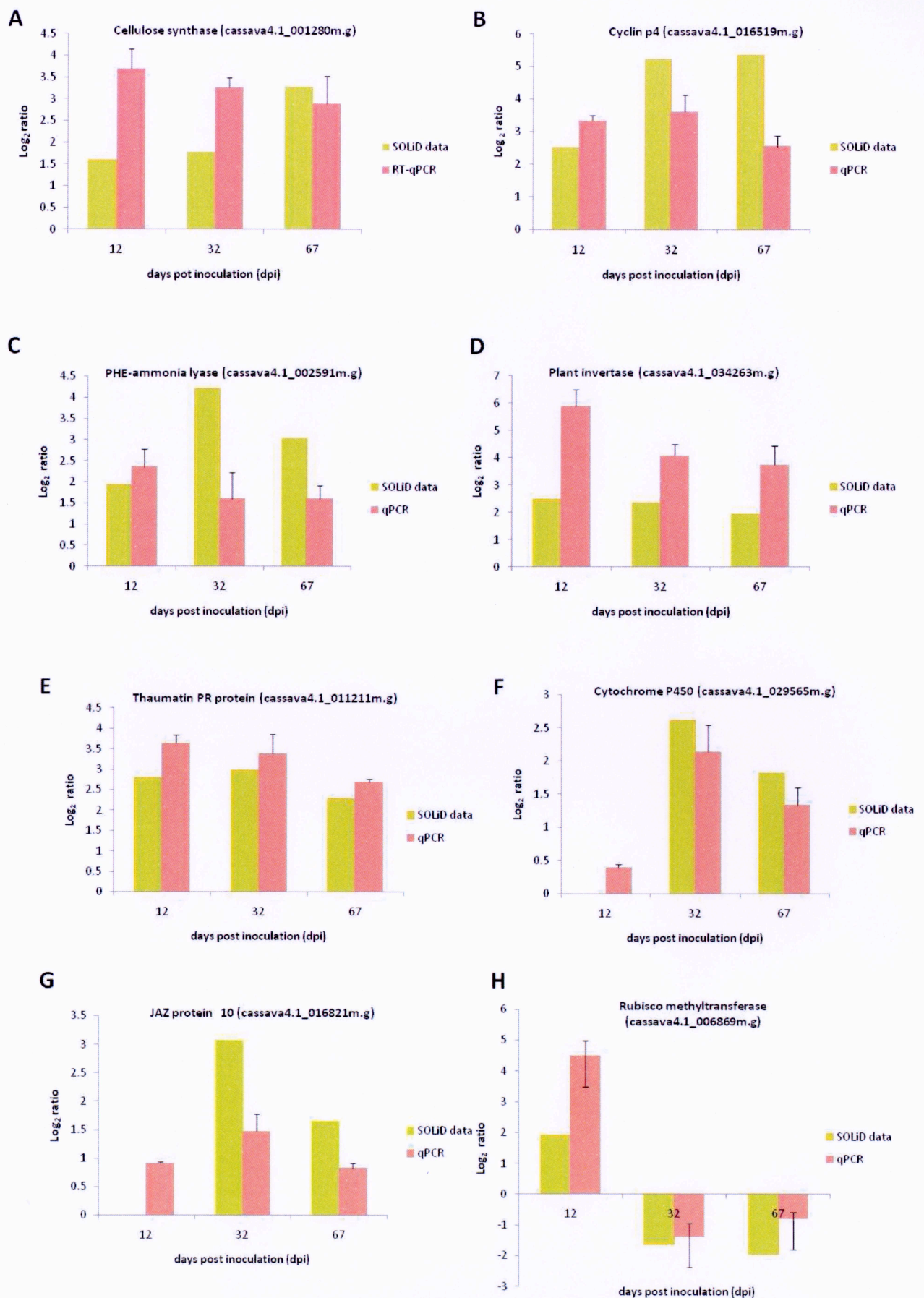


Figure 4.5 RT-qPCR vs. SOLiD $\log_2$ gene expression ratios of eight genes measured from SACMV-[ZA:99] leave tissue at 12, 32 and 67 dpi. Expression values were calculated by normalizing SACMV-infected expression ratios by mock-inoculated expression ratios for each gene. Five genes were chosen for T200 (A-E) and 3 for TME3 (F-H). The expression of each gene was normalized to endogenous UBQ10.

4.5 Discussion

Here we present for the first time, transcriptomic data that has been generated through the ABI SOLiD 4 platform, for an interaction between the natural host cassava and SACMV-[ZA:99]. In this study, we used NGS to detect and quantify the transcripts present in mock-inoculated apical cassava leaves compared with leaf expression after infection with SACMV-[ZA:99]. Data was generated for 12 libraries in total, six for susceptible cultivar T200 (3 infected, 3 mock-inoculated) and six for tolerant landrace TME3 (3 infected, 3 mock-inoculated). Furthermore, infection was followed over a period of 67 days post infection, with sampling at three timepoints, namely 12, 32 and 67 dpi. This study clearly demonstrated that the interaction of cassava with SACMV-[ZA:99] involves transcriptional activation or repression of a large number of genes. These levels and patterns of expression are particularly interesting as they differ between T200 and TME3. Despite different genetic backgrounds between the susceptible and tolerant landraces, it is noted that all altered genes in each landrace are compared at each time post-virus inoculation with the corresponding mock-inoculated counterpart. Consequently the study depicts a global transcriptome snapshot, representing host response to a geminivirus, at three times within a 67 day infection process in two landraces, one known to be tolerant and the other highly susceptible to SACMV-[ZA:99]. All annotated gene functions for both T200 and TME3 are determined from the reference genome in phytozome blasted against *Arabidopsis*, and some altered genes in both these landraces may not have been detected/identified due to genetic differences with the sequenced reference cultivar (AM560-2) and unknown functions ($\pm$ 67% of the reference cassava genome is annotated). Furthermore, it is difficult to do true justice to this wealth of data in a single report. Consequently, we identified SACMV-[ZA:99]-responsive transcripts that were common in both but showed differing expression patterns at various time points post infection. Some of these groups include transcription factors, R genes, genes involved in the DNA methylation pathway and those involved in the putative movement of SACMV-[ZA:99]. This work provides the foundation for further investigation of molecular mechanisms underlying GV-host interactomes (disease vs. tolerance) in the agriculturally important natural host, cassava.

4.5.1 Downregulation of TIR-NBS-LRR, WRKY and MAPK

Transcript quantification showed that T200 had a far greater change in the number of differentially expressed genes as well as the magnitude of expression changes across time points (Supplementary data S4.1- S4.4). For T200, 209 genes (Fig 4.3 A) were found to be common between 12, 32 and 67dpi and $\log_2$ expression values ranged between $\log_2$ -6.53 to 4.89 across timepoints indicating strong transcriptional changes in this susceptible cultivar. A comparison between timepoints showed

that T200 had the greatest overlap of genes between 32 and 67dpi (812, Fig 4.3 A). Many of the cassava host genes found to be down-regulated are key players in the defence response signalling pathways reported for many R-mediated resistance phenotypes. Of these, the most noticeable observation with regard to the T200 transcript data, was the consistent downregulation of ten TIR-NBS-LRR disease resistant genes and ten different WRKY transcription factors. In addition, mitogen-activated protein kinase 3, mitogen-activated protein kinase kinase kinase 15 and mitogen-activated protein kinase kinase 9 were also found to be consistently downregulated. Interrogation of the TME3 data at the same timepoints did not show any of the same patterns with regards to the expression of TIR-NBS-LRR, WRKY and MAPK genes. One comparable observation however was made between early timepoint 12dpi and 32dpi where only one WRKY40 (cassava4.1_011696m.g) and one MAPKKK19 (cassava4.1_020998m.g) were found to be upregulated. The identification and characterization of R-genes has long been under scrutiny, and several of these have been cloned from bacteria, fungi and oomycetes. To date, research has been focused on three dominant viral R genes. These include *Rx* against *Potato virus X* (Bendahmane et al. 1995 and 1999), *RT4-4* against *Cucumber mosaic virus* and *N* resistance against *Tobacco mosaic virus*. The identification and sequencing of TIR-type NBS-LRR and absence of CC-NBS-LRR genes in this study is interesting in itself as it compares with previous cloned *Rx*, *RT4-4* and *N* resistance genes, which also contain TIR domains. Our data indicates that several R genes (or splice variants thereof), with TIR-domains, are actively downregulated by the presence of SACMV-[ZA:99]. Although cassava resistance genes *CMD1* and *CMD2* have been located on linkage maps of cassava (Fregene et al. 1997; Akano et al. 2002), these genes have not yet been identified and mapped to any scaffolds of version 4.1 of the cassava draft genome presently available on phytozome, and the identification of many TIR-NBS-LRR genes in T200 suggests that dominant resistance in different cassava landraces/cultivars is not monogenic, and may differ between genotypes. *CMD1* and *CMD2* may contribute to resistance, but resistance to geminiviruses is likely to be due to several genes acting epistatically or additionally (Garcia-Neria and Rivera-Bustamante. 2011; Vidavski et al., 2008). Studies have shown that virus-host interactions can directly contribute to the establishment of infection, via direct interactions with host proteins and hijacking cellular pathways (Nelson and Citovsky, 2005; Pallas et al. 2011), or by suppression of host defences (Hevelde et al, 2008). In this study, the downregulation of TIR-NBS-LRR transcripts in susceptible T200 is strong evidence for defence repression, and susceptibility being a consequence of failure of host recognition of viral proteins. SACMV-[ZA:99] may be avoiding detection through inhibition of R-gene mediated plant defence responses, therefore promoting virus replication and movement. Furthermore, suppression of TIR-NBS-LRR could negatively affect other signalling pathways downstream of TIR-activation such as the mitogen-activated protein kinase pathway.

A MAPK signalling cascade is sequentially activated by three protein kinases, a MAP kinase kinase Kinase (MAPKKK or MEKK), a MAP kinase kinase (MAPKK or MKK) and a MAP kinase (MPK). Activation of this multi-tiered cascade is phosphorylation-dependent (Mészáros et al. 2006). Twenty MPKs have been identified in *Arabidopsis* (Innes 2001; Pitzschke et al., 2009) where MPK3, MPK4 and MPK6 in particular are stress/pathogen-responsive and have been the most comprehensively studied (Asai et al., 2002; Beckers et al., 2009; Menke et al., 2004; Yang et al., 2001).

Studies in *Arabidopsis* have provided direct lines of evidence that some WRKY proteins are involved in plant defence response. *Arabidopsis* leaves transiently transformed with bacterial flg22 was shown to induce transcription factors WRKY22 and WRKY29 which was mediated by a MAPK pathway involved in both bacterial and fungal pathogens which in turn resulted in the attenuation of disease symptoms (Asai et al. 2002). Furthermore, *RRS1*, a gene that confers resistance to bacterial pathogen *Ralstonia solanacearum* encodes a TIR-NBB-LRR protein with a C-terminal WRKY motif (WRKY52). This additional WRKY structural feature of *RRS1* could indicate a direct relationship between Avr-recognition and the downstream transcriptional activation of defence genes (Deslandes et al. 2003). Similarly, a study by Liu et al., (2004) demonstrated that virus-induced gene silencing of three WRKY genes (NtWRKY1, NtWRKY2 and NtWRKY3) in *Nicotiana tabacum* resulted in compromised N-gene-mediated resistance to *Tobacco mosaic virus*. Suppression of WRKY and MAPK transcription, as observed in our study for T200, in conjunction with TIR-NBS-LRR, led to enhanced susceptibility, and the disease phenotype is maintained in T200 throughout the 67 day period with the avoidance of the R-mediated resistant mechanism. This correlates well with viral quantification data showing increase in SACMV-[ZA:99] titre over the sixty-seven day period, as well as the appearance of symptoms over time. Furthermore, although the effect of MAPK-mediated phosphorylation on the function of WRKY remains to be defined, we also speculate that due to the downregulation of MPK3 (cassava4.1_010219m.g) and reduced levels of MAPK3 (cassava4.1_010219m.g), phosphorylation of transcription factors such as WRKY is reduced, which may contribute to the down regulation of defence-related genes.

4.5.2 Induction of defence-related genes

As stated earlier, the observation that many transcription factor genes e.g. WRKY were found to be downregulated, we did not expect to find a large number of defence –related genes induced during infection with SACMV-[ZA:99]. However, the transcripts levels for two pathogenesis-related (PR) protein genes were shown to be increased upon infection by SACMV-[ZA: 99] primarily at 32 and 67 dpi. These PR-proteins included peroxidase (cassava4.1_011768m.g, cassava4.1_012124m.g) and

thaumatin superfamily protein (cassava4.1_014480m.g, cassava4.1_014683m.g, cassava4.1_011211m.g) which may be induced by the activation of transcription factors other than WRKY. Log₂ expression ratios ranged between 1.76 and 2.05 for peroxidase and between 2.28 and 3.59 for thaumatin. The induction of pathogenesis-related genes has been reported in other stress treatments and virus infections using gene expression tools (Babu et al. 2008; Golem and Culver, 2003; Pompe-Novak et al. 2006; Whitman et al. 2003). Unlike these studies, where a whole range of PR-related genes were expressed, the induction of only two PR genes here, hints at the mounting of a weaker immune response. The induction of these two genes at 32 and 67dpi in compatible T200 therefore supports a delayed host defence response, where basal immune responses persist even at full symptomatic infection indicating a general yet unsuccessful defence response.

4.5.3 Role of flavonoids and phenylpropanoid in general stress response

One of the most interesting findings, which have not extensively been reported in cassava and other compatible virus-host interactions, is the role of genes involved in the flavonoids and phenylpropanoid pathway. KEGG pathway analysis showed that several genes involved in flavanoid and phenylpropanoid biosynthesis were induced in T200-SACMV-[ZA:99] infection which were prominently altered at 32 dpi and maintained at 67dpi. Genes mapping to these pathways included flavonol synthase (cassava4.1_011509m.g), UDP-glycosyltransferase (cassava4.1_005848m.g), chalcone synthase (cassava4.1_009206m.g, cassava4.1_009295m.g, cassava4.1_009402m.g) and phenylalanine ammonia lyase (cassava4.1_002591m.g, cassava4.1_002709m.g, cassava4.1_034377m.g). Furthermore, these genes were all found to be highly induced with expression ratios in the range of Log₂ 1.95 – Log₂ 4.45. Phenylalanine ammonia lyase (PAL) is an enzyme that catalyzes the first and most important step in the phenylpropanoid pathway. Several lines of evidence indicate that PAL may participate in depending host plants against invading pathogens. This has been shown in a very early study conducted by Pallas *et al* (1996), where PAL-suppressed tobacco leaves did not result in the induction of downstream PR proteins in systemic leaves which therefore impaired an active defence response against TMV. Also, more recently, Hoa *et al* (2011), demonstrated that PAL was highly induced (5.8-fold) in a resistant rice variety early hours after infection with *Rice stripe virus*, but not in a susceptible variety suggesting that PAL plays a defence response. Similarly, the silencing of a pathogen-inducible UDP-glycosyltransferase in tobacco resulted in the depletion of UDP-glycosyltransferase in tobacco which enhanced oxidative stress and weakened resistance of silenced tobacco plants to TMV infection (Chong et al., 2002).

We however observed the activation of PAL, CHS and UDP-glycosyltransferase at late stages of infection in T200 (32 and 67 dpi). The induction of PAL and CHS in particular may be the host plant actively trying to mount a defence response against SACMV-[ZA:99] in order limit SACMV-[ZA:99] albeit unsuccessfully, or could be a general stress response or consequence of infection.

4.5.4 Ethylene and Jasmonic acid phytohormone signalling

Hormones, such as ethylene, jasmonic acid, abscisic acid, gibberellins and salicylic acid are present in plants in basal amounts, and act in a well-balanced and regulative manner during plant growth and development (Culver and Padmanabhan, 2007). Any change from normal levels of phytohormones, such as those caused by infection with virus pathogens, could significantly alter physiological processes and morphology, resulting in symptoms such as stunting and leaf deformation, as was observed in our study. One striking observation for both T200 and TME3 across timepoints is the absence of altered salicylic-acid (SA) expression, as well as genes that are reported to activate and regulate the SA signalling pathway such as ENHANCED DISEASE SUSCEPTIBILITY 1 (EDS1) and PHYTOALEXIN DEFICIENT 4 (PAD4). This is particularly interesting, especially for TME3, reported to be resistant, since numerous studies have shown that (SA) plays an essential role in signal transduction pathways leading to the dramatic accumulation of pathogenesis-related (PR) transcripts culminating in a disease resistance response (Durner et al., 1997; Dempsey et al., 1999). Furthermore, transcripts encoding proteins involved in the synthesis and signal transduction of hormones such as ethylene (ET) and jasmonate (JA) were found to be consistently suppressed by 2-fold or more during systemic infection by SACMV-[ZA:99] in T200 AND TME3, respectively, but gibberellin-regulated family protein were consistently upregulated in T200 plants, particularly at 32 and 67dpi. Furthermore, a gene for the enzyme lipoxygenase (cassava4.1_001238m.g), involved in the early steps of JA synthesis, was also found to be downregulated. Other suppressed genes included regulators of JA signalling, Jasmonate ZIM domain proteins (JAZ, cassava4.1_009349m.g, cassava4.1_015456m.g, cassava4.1_018315m.g) and ethylene-responsive element binding factors (cassava4.1_012714m.g, cassava4.1_015499m.g, cassava4.1_012714m.g). The similar expression patterns of JA and ET is not all that surprising as ET and JA pathways, in many instances, have been shown to regulate similar type of defence genes (Schenk et al. 2000; Lorenzo and Solano, 2004).

Ethylene-responsive element binding factors (ERF) proteins are plant-specific transcription factors that play important roles in plant responses to various hormones or environmental changes. The function and mediation of signals of ERFs has been demonstrated in *Arabidopsis* in response to various stimuli, including ethylene (Fujimoto et al., 2000), salt stress (Park et al., 2001) and pathogen infection (Schenk et al., 2000; Park et al., 2001). In addition, the role of ERFs has also been

demonstrated in other higher plants such as tobacco and tomato, illustrating the induction of ERFs following infection by *Pseudomonas syringae* (Zhou et al. 1997; Thara et al. 1999) and *Tobacco mosaic virus* (Horvath et al., 1998). A report in 2005, by Love et al., showed that ethylene-signalling mutants reduced virus titers of *Cauliflower mosaic virus* and hindered long-distance movement of the virus. This finding is contradictory to what we observed in our study, particularly for T200, as downregulation of ethylene-responsive element binding factors did not appear to hinder virus replication or movement. For TME3, data was contrasting as no ethylene-responsive element binding factors were found to be significantly changed across any of the three timepoints. Results of this study suggest that SACMV-[ZA:99] infection in cassava T200 is supported by evasion of basal host defence via negative regulation of JA and ET signaling pathways and lack of host elicitation of SA pathway dependent resistance.

4.5.3 Changes in cell wall- and plasmodesmata-associated genes

In a study by Shimizu et al (2007), it was reported that *Rice dwarf virus* (RDV) infection in rice plants resulted in the repression of several cell-wall related genes, leading to dwarfing symptoms. Differential expression of at least three of the cell-wall enzymes (cellulose synthase (cassava4.1_001280m.g), pectin methylesterase (cassava4.1_004357m.g), and arabinogalactan protein (cassava4.1_018722m.g , cassava4.1_029618m.g) identified in the RDV study was also observed in this study for T200 but the expression of these genes were found to be upregulated as compared to downregulation observed in RDV-infected rice plants at 9 dpi. Down regulation of these enzymes was linked to the dwarfing phenotype observed in RDV-plants, however, the same phenotype is not observed in T200. In addition, several plasmodesmata-associated transcripts were upregulated, particularly at heightened infection at 32 dpi and 67 dpi in T200 and the same genes were not detected in tolerant cultivar TME3 at either timepoint. These genes include, plant invertase (cassava4.1_016774m.g, cassava4.1_021617m.g), pectin lyase (cassava4.1_005619m.g, cassava4.1_007568m.g, cassava4.1_009002m.g), β -tubulin (cassava4.1_007617m.g, cassava4.1_007632m.g), expansin (cassava4.1_014066m.g, cassava4.1_014407m.g, cassava4.1_014440m.g, cassava 4.1_014489m.g), plasmodesmata callose-binding protein 3 and (cassava4.1_016458m.g, cassava4.1_016746m.g), calreticulin (cassava4.1_008376m.g). The induction of these genes indicate that there are numerous cell wall and plasmodesmata modifications occurring at T200 cell boundaries, but whether these modifications are favourable to the virus is yet to be determined. However, what is true for virus infections, whether in a compatible or incompatible interaction, is the increase in nutrient and energy demands of the host as well as the cellular demands of mounting a defence response. The enhanced expression and activity of cell wall

invertases for example and its role as in plant-pathogen interactions has been reported in several studies (Chou et al., 2000; Fotopoulos et al., 2003; Berger et al., 2004; Swarbrick et al., 2006). Several lines of evidence indicate that an increase in cell-wall invertase for example will result in the cleavage of sucrose into glucose and fructose which serve as the energy molecules that fulfil the carbon and energy demand of the mounting a defence response against the invading pathogen (Roitsch et al. 2003; Truernit et al, 1996). In addition, sugars such as glucose and sucrose serve as signalling molecules (Bolouri-Moghaddam et al. 2010) which will prime the activation of PR genes following infection (Gomez-Ariza et al. 2007). For example, infection of tobacco plants with PVY led to sugar accumulation, which was accompanied by an accumulation of transcripts encoding PR proteins (Herbers et al. 2000). Based on these results it was proposed that sugars act as amplifiers for plant defence responses during plant-pathogen interactions (Herbers et al. 2000). Our study shows an upregulation of invertase at the late stages of infection suggesting that the breakdown of sucrose could play a role as both an energy source and signalling molecule for attempted defence responses against SACMV-[ZA:99].

Also observed in our transcriptome data were the upregulation of β -tubulin, pectin methylesterase (PME), calreticulin and plasmodesmata-callose binding protein. A number of previous studies have implicated a number of cellular components and proteins that are localised to the plasmodesmata (PD) and that play a role in either cell-to-cell communication or movement of molecules across the PD. SACMV-[ZA:99] is a bipartite virus that has a DNA-B component harbouring two movement (BV1 and BC1) genes that encode movement proteins that act in a cooperative fashion to facilitate local and systemic movement of the virus (Sanderfoot and Lazarowitz 1995). These movement proteins are thought to require interaction with a number of host factors in order to aid virus movement throughout the host plant. In a number of studies conducted, it has already been suggested that the viral movement proteins modify the plasmodesmata and alter the plasmodesmal size exclusion limits (SEL) to allow the movement of viral protein–nucleic acid complexes to neighbouring cells (Su et al, 2010). Furthermore, the interaction between viral movement proteins, the PD and the host cytoskeleton has already been scrutinised for many virus-plant systems (Boevink and Oparka, 2005; Heinlein et al, 1995, Heinlein et al, 1998; Kahn et al., 1998; Lucas 2006; McLean et al., 1995; Su et al, 2010). For example, in some studies, PME has been shown to interact with TMV movement protein which assists the virus with cell-to-cell movement during infection (Chen et al. 2000; Dorokhov et al., 1999; Heinlein, 2002). This interaction was confirmed by Chen and Citovsky in 2003, when their work using a reverse genetics antisense approach to reduce PME in tobacco plants successfully demonstrated that reduced levels of PME lead to a considerably delayed systemic movement of TMV. Additionally, the interaction between PME and TMV MP is not an exclusive one. Chen et al, (2000) further demonstrated through a yeast two-hybrid system, that the MPs from two other plant

viruses, *Cauliflower mosaic virus* and *Turnip vein clearing virus* also bind to PME. We therefore speculate, that the induction of PD-associated genes in T200 is favouring cell-to-cell movement of the virus which has can be linked to the increase of SACMV-[ZA:99] replication (virus titers) and spread observed at 32 and 67dpi.

4.5.4 Upregulation of Methyl-CpG-binding domain in TME3

One unique observation made with TME3, but not in T200 at any timepoint, was the upregulation of methyl-CpG-binding domain (MBD, $\text{Log}_2 = 2.478$) specifically at 67 dpi. Twelve methyl-CpG-binding domains (MBD) have been identified and characterized in *Arabidopsis* and these function with chromatin remodelling proteins to inactivate gene expression (Berg et al. 2003; Springer and Kaeppler, 2005). Given that geminiviruses such as SACMV-[ZA:99] have DNA genomes and go through a double stranded intermediate during rolling circle amplification, it has been suggested that plant hosts use viral genome methylation as a defence against geminiviruses (Raja et al. 2008), which culminates in the recovery phenotype. The recovery phenotype is a phenomenon observed in virus-infected host plants exhibiting severe disease symptoms following infection, but symptoms gradually attenuate until the host appears almost symptomless (Hagen et al. 2008). The recovery phenotype has been observed in many plant systems infected with virus, including cassava infected with either *African cassava mosaic virus* (ACMV) and *Sri Lankan cassava mosaic virus* (SLCMV) (Chellappan et al., 2004; Hagen et al. 2008; Patil and Fauquet 2009). However, the recovery phenotypes in these cassava infections were attributed to the production of virus-derived siRNAs through post-transcriptional gene silencing (PTGS). Since alterations in siRNA-associated genes such as Dicer-like ribonuclease (DCL), Argonaute (AGO) and RNA-depednent RNA polymerase (RdRP) were not detected in any of our sequencing data, we cannot draw any firm conclusions linking the recovery phenotype observed in TME3 plants at 67dpi to mechanisms of PTGS. However, with the induction of MBD (cassava4.1_028187m.g) it could be speculated that the upregulation of MBD in this cultivar in particular is targeted at binding to methylated CpG within the SACMV-[ZA:99] genome, therefore inhibiting replication and transcription of SACMV-[ZA:99]. This could be one of the reasons accounting for lower viral titres and the recovery phenotype observed in TME3 at 67dpi as compared with T200.

4.6 Conclusions

In general, the transcriptome alterations induced by SACMV-[ZA:99] in TME3 were significantly lower compared to T200, and also in comparison with *Cabbage leaf curl virus* (CaLCuV) in the susceptible host, *Arabidopsis* (Ascencio-Ibanez et al., 2008). This could be attributed to failure of

SACMV-[ZA:99] to elicit wide transcriptome changes, a feature typical of many studied plant virus-interactions, where such responses are critical in order for the virus to establish disease . The most important finding from this study was the downregulation observed for TIR-NBS-LRR R-genes in SACMV-infected T200 at 12, 32 and 67dpi but not in TME3. This finding strongly suggests that TME3 is not a “resistant” landrace, associated with CMD2. Furthermore, at least 33% (2 out of every six) TME3 plants challenged with SACMV-[ZA:99] displayed a “recovery” phenotype from about 50dpi , which may involve persistent RNA silencing maintenance. The genes identified in this study can be used as markers for subsequent cassava breeding programs and studies involving genetic modification for crop improvement.

Chapter 5

Summary and Future Recommendations

5.1 Overall summary and Discussion

Cassava mosaic disease (CMD) is a devastating disease caused by at least eight species of *Begomovirus*, one of which is *South African cassava mosaic virus*-[South Africa:99] (SACMV-[ZA:99]). This virus has been identified in regions of South Africa and some neighbouring countries including Zimbabwe, Mozambique and Swaziland. Understanding the mechanisms underlying CMD could facilitate control strategies to combat this virus, either through genetic modification approaches or through breeding programs, which could result in conferring resistance or a degree of tolerance to this virus. The work demonstrated in this thesis therefore presents the analysis of gene expression changes associated with SACMV-[ZA:99] infection in *Nicotiana benthamiana*, a model host used for plant virus studies, in particular geminiviruses, and natural host cassava. This work therefore aimed at addressing a number of questions with regards to SACMV-[ZA:99] infection in these different plant hosts. These included:

- 1 – What are the genes that are changed globally in hosts as a result of SACMV [ZA:99] infection,
- 2 – What are the similarities or differences in host responses in different plant systems and finally,
- 3 – How effective are the defence responses that are triggered in these host plants against SACMV-[ZA:99]?

The first goal of this work was to characterize the transcript profile of *Nicotiana benthamiana* leaf tissue infected with SACMV-[ZA:99] at four timepoints post inoculation. *N. benthamiana* was chosen as our model system as it is the most widely-used experimental host in plant virology and has been established in many studies as particularly susceptible to a diverse number of plant viruses (Goodin et al. 2008). Four timepoints were chosen which represented early infection (7dpi), onset of symptoms (14dpi), onset of systemic infection (21dpi) and finally fully symptomatic (35dpi.) At the conception of this study, there were only a limited number of commercially available cDNA and oligo plant microarray platforms, most of which were designed for model host *Arabidopsis thaliana*. Based on comparative heterologous/cross-species work carried out by Horvath et al., and other research groups, we were optimistic that a cross-species microarray approach could be conducted between *N. benthamiana* and *Arabidopsis*. This study was therefore initiated using a custom made cDNA microarray with 5000 ESTs representing an *Arabidopsis* SSH library of pathogen *Ralstonia* -inducible genes. Strong cross-hybridization signals illustrated approximately 70% cross-hybridization between *N. benthamiana* leaf tissue cDNA and *Arabidopsis* probes in an experiment we called the “yellow

experiment”, and indicated that these slides could be suitably used for downstream experiments elucidating gene expression changes in SACMV-infected leaf tissue. Gene expression changes in SACMV-[ZA:99] infected *N.benthamiana* leaf tissue were measured from early infection (7dpi) to full systemic infection (35dpi). We found 73 genes in total that were regulated (1.42% of the SSH chip) throughout the development of SACMV-infection in *N.benthamiana* leaves, where the number of differentially expressed genes increased between 7dpi (15 genes) and 35dpi (44genes).

The CSH study provided useful insights into some of the genes involved in SACMV-infection, but with the progression of this PhD work, an oligo microarray slide (Nb-array) representing 13 014 unigenes of *N. benthamiana*, became available. We wanted to take full advantage of this and therefore decided to conduct a second microarray experiment using systemically infected *N.benthamiana* leaf tissue. Following data analysis, it was found that 611 genes were found to be differentially expressed, 483 of which were upregulated, and 128 were downregulated (4.7% of Nb-array). The genes found to be differentially expressed for systemically infected leaf tissue were similar to the patterns of expression of many genes identified to RNA virus studies dissecting molecular interactions of a pathogen in a plant host. Genes identified in this study were chiefly associated with general defence response, phytohormone signalling and virus movement. While, many of the genes or pathways were similar to genes observed in other RNA virus infections, unique differentially expressed genes were also identified. This suggests that some of the general defence and hormone pathways activated during different virus infections are conserved, but certain activated or repressed pathways, or specific host genes that interact specifically with viral proteins, may be host-virus specific. One of the genes uniquely identified in this study was Rab6A which we hypothesise may be playing a role in the movement of SACMV-[ZA:99] through the plasmodesmata. This conclusion is supported by several plant virus-host interaction studies, which have demonstrated a similar core of altered genes between eleven viruses in *Arabidopsis*, and an increase of altered gene expression as infection progresses, but also unique genes that are virus-specific (Postnikova and Nemchinov, 2012).

While model host plants are useful in studying plant virus-host interactions, they may not always represent precisely what is happening in the virus pathogen’s natural host. In 2009, the genome sequence of cassava (www.phytozome.org) was made publically available. This sequence data was derived from cassava line AM560-2. With the availability of this data and next generation sequencing (NGS), we embarked upon a study of monitoring SACMV-[ZA:99] infection in its natural host.

To characterize the virus-responsive transcriptome of cassava more thoroughly, we employed the high-throughput ABI SOLiD platform in order to identify genes that were differentially regulated in

both susceptible T200 and TME3 cassava landraces. Taken together, the comparative transcriptomic analysis of T200 and TME3 in this study revealed that the two landraces differ greatly in the molecular response to SACMV-[ZA:99] infection, which may somewhat explain their differences in symptom phenotype, particularly at timepoint 67 dpi, where the recovery phenotype was observed in infected TME3 cassava leaves but not in susceptible T200. T200 challenged with SACMV-[ZA:99], led to a significant number (4181 genes, 13.6% of the total cassava transcripts annotated on phytozome) of differentially expressed genes, many involved in transcription, defence response genes, phytohormone signalling and those putatively involved in movement. In contrast, gene expression changes in TME3 were significantly lower (1008 genes, 3.3% of the total cassava transcripts annotated on phytozome) than those observed in T200 at all three time points. Additionally, only a few basal defence responsive proteins such as thaumatin-like protein and cytochrome p450, as well as genes involved in the DNA methylation pathway were activated in the SACMV-challenged TME3 leaf tissue at 32 and 67dpi. Induction of successful systemic infection and symptoms in a compatible plant host requires recognition of the virus, and significant induction of metabolic pathways to support virus replication and movement. We conclude, that poor gene expression response in TME3, compared with the highly susceptible T200, reflects its tolerant genotype, and that the induction of a limited amount of transcripts reflects a limited host response and altered gene expression to successfully support SACMV-[ZA:99]. Also, a particularly important finding for T200, was the repression of ten TIR-NBS-LRR R-genes at 12, 32 and 67 dpi. The downregulation of TIR-NBS-LRR transcripts in susceptible T200 is strong evidence for defence repression, and susceptibility being a consequence of failure of host recognition of viral proteins. The repression of TIR-NBS-LRR genes was not observed TME3 plants. This conclusion is further supported by viral titre measurements for T200 and TME3 across three timepoints, where SACMV-[ZA:99] increased by 202-fold from 32 to 67 dpi, but TME3 only increased by 20-fold between the same two timepoints, further demonstrating and confirming the tolerant nature of TME3. This finding in T200 is also consistent with the increasing viral titre trends observed in Chapters 2 and 3 for infected *N.benthamiana* leave tissue over time, where both T200 and *N. benthamiana* demonstrate a susceptible phenotype. Additional analysis such as the phylogenetic relationship between TIR-NBS-LRR domains between plants species was not part of this study, but it would be interesting to carry out multiple sequence alignments between R genes identified in (eg Rx and N) identified in *Arabidopsis*, rice and other plant systems in order to investigate the homology between these genes. The homology between these genes could provide some insight into the evolutionary conservation of these R-genes.

Moreover, work carried out in our research group (but not presented in this thesis) involving the sequencing and quantification of siRNA and miRNA populations (manuscript currently in

preparation) in SACMV-[ZA:99]-infected T200 vs. TME3 (at 12, 32 and 67 dpi) revealed that not only do the phenotypes and number of miRNA-targeted genes differ between T200 and TME3, but so do the siRNA populations that map to target sequences present on DNA-A and DNA-B of SACMV-[ZA:99]. There are three main siRNA pathways that have been identified in higher plants. These include , small-interfering (si) RNA, micro (mi) RNA and RNA-mediated DNA methylation (RdDM) as described by Baulcombe, (2004) and others (Carrington and Ambros 2003; Vaucheret, 2006; Vazquez, 2006). The RNA silencing pathways are triggered by double stranded RNA structures and each are involved in a number of diverse functions.

Although siRNA populations can range in length between 21- and 26-nt, the 24-nt siRNA range, produced by DCL3 (Herr et al. 2005; Xie et al. 2004), has primarily been associated with siRNA-mediated DNA methylation (RdDM). Notably, the 24-nt siRNA populations targeting SACMV-[ZA:99] are much higher in TME3 ($\pm 22\ 000$ 24-nt siRNA targeting DNA-A, $\pm 8\ 300$ 24-nt siRNAs against DNA-B) as compared to T200 ($\pm 12\ 000$ 24-nt siRNA against DNA-A, ± 5500 24-nt siRNAs against DNA-B) at 67pi. Since the “recovery” phenotype is observed in TME3 at 67dpi, we conclude that TME3 is not a classic “resistant” landrace, associated with the previously identified dominant R gene, CMD2 (Akano et al. 2002; Fregene et al. 1997), but rather that it is tolerant which could, in part, be accounted for by persistent RNA silencing maintenance of genes required by SACMV-[ZA:99], and by silencing of this geminivirus through a transcriptional gene silencing (TGS) i.e. methylation-associated mechanism. Methylation of viral DNA by plant hosts has already been demonstrated as an epigenetic defence against geminivirus (Raja et al. 2008). Raja et al. elegantly demonstrated that *Arabidopsis* mutants defective in a number of genes that are key players in the RdDM pathway (eg *drm1*, *drm2*, *kyp2*, *ago4* and others) results in hypersusceptibility to infection with the geminiviruses *Cabbage leaf curl virus* (CaLCuV) and *Beet curly top virus* (BCTV).

For our study, we were also interested in determining whether there were some similarities between a natural and model host. Since the Nimblegen array and cassava sequence data represented an estimated 35% and 95% of the *N. benthamiana* and cassava genomes, respectively, a direct comparison cannot be done, but rather trends or patterns could be discerned, and specific altered genes at 21 dpi examined. One of the major findings consistent between the Nb-array study (Chapter 3) and the RNA-seq transcriptome profile study (Chapter 4), was the differential expression of WRKY transcription factors in susceptible host *N. benthamiana* and cassava T200. The expression patterns between these two hosts differed though, where WRKY genes were found to be downregulated in *N. benthamiana* at 21dpi but upregulated in T200 at 32 and 67dpi. WRKY transcription factors contain one or two WRKY domains and have a binding preference for the W box, which is found in the promoter of many defence genes, including PR genes (Eulgem et al., 2000). Since the WRKY genes differ for *N. benthamiana* (WRKY 3, WRKY6 and WRKY 30) and T200

(WRKY7 and WRKY70), it can be speculated that the defence –related genes activated downstream by these genes is different, therefore resulting in the activation of different pathways in an attempt to combat SAMCV-[ZA:99] infection. This could further explain the differences in induction vs. suppression of these WRKY genes observed in the two susceptible cultivars. Based on these findings and other published work, it is becoming increasingly evident that a major theme underlying transcriptional regulation in defence response is the combinatorial control of defence-related genes by the positive or negative regulation of various transcription factors. In addition, the potential role of Rab-GTPases in the intercellular trafficking of SACMV-[ZA:99] was also elucidated in this study as three different Rabs (Chapter 2 – Rab GTPase homolog 7 and Rab GTPase Homolog E1B and Chapter 3 - Rab6A) were found to be upregulated in both the CSH and Nb-array hybridization respectively. This finding warrants further investigation possibly using fluorescent tagging proteins in bimolecular fluorescence complementation studies to monitor their possible role in virus movement through the PD.

In conclusion, results from this thesis, while presenting unique genes altered by a geminivirus, SACMV-[ZA:99], also supports other plant virus studies, where common basal defence mechanisms are induced early in infection, irrespective of whether the host is susceptible or resistant/tolerant. The host plant may attempt to reduce virus infection throughout the course of infection, but in compatible hosts there is a more significant increase in gene induction or repression as disease symptoms progress, compared to resistant or tolerant plants, where host-gene response is lower. Alteration in gene expression is specifically required to support virus replication (it has been shown in a comparative microarray study of eleven plant viruses that 23% of *Arabidopsis* genes are differentially expressed during infection; Postnikova and Nemchinov, 2012), and some of the host proteins are specifically required to interact with viral proteins, while other changes are associated with a need for the plant to attempt defence, albeit ineffectively in compatible susceptible plants.

This study has contributed further to our understanding of plant host-virus interactions, and has identified unique responsive genes which may play a role in virus infection. In addition, this is the most comprehensive study performed on a geminivirus, as most studies have focused on RNA viruses (Babu et al. 2008 , Senthil et al. 2005, Whitham et al. 2003). In this thesis, we also demonstrate, to our knowledge, the first study to date using the most comprehensive gene dataset of *N. benthamiana*, and also the only transcriptome study of a geminivirus using the SOLiD NGS platform and the full cassava genomic sequence dataset. Furthermore, to our knowledge we have also for the first time compared a global transcriptome profile between a susceptible and tolerant

cassava landrace. This study has opened up many avenues to further elucidate the complex pathways in a geminivirus-host system.

5.2 Future recommendations

Bearing in mind that the work presented in this thesis utilized transcriptome profiling tools to identify SACMV-[ZA:99] responsive genes in host plants, this study can be considered to be one of gene discovery. In order to fully exploit the data from this work, there are a number of follow-up experiments that could be used in order to start building “interactome” models representing the interaction between SACMV[ZA:99] and host genes/proteins. Since transcriptomics alone is not sufficient to piece together a detailed interaction model, it is necessary to also consider plant-pathogen proteomic and metabolomic studies for downstream analysis.

Firstly, a reverse genetics approach using virus-induced gene silencing (VIGS) or *Arabidopsis* mutants can be used to silence candidate genes of interest followed by infection with SACMV-[ZA:99]. The knock-down of these genes will further demonstrate the role of these genes in plant susceptibility and resistance. Also, the yeast two-hybrid (Y2H) is a screening method that can be used to determine if protein-protein interactions occur in a given system. As a follow on study from this work, it would be interesting to test whether any of the suites of genes identified from the *N.benthamiana* or cassava studies actually directly interact with each other or with any of the SACMV-[ZA:99] proteins. This can be achieved by using *N. benthamiana*, cassava cDNA libraries or SACMV-[ZA:99] as prey. Successful virus infection in plants is not only dependent on the activation or suppression of cohorts of genes or protein-protein interactions between host and pathogen, but may also be fuelled by the presence of primary and secondary metabolites such as lipids and lipases, and the future of metabolomics in this regard will prove interesting.

In conclusion, the work presented here is novel and cutting edge, as only one other global gene expression study has been carried out monitoring geminivirus infection, namely *Cabbage leaf curl virus* (CaLCuV) in the susceptible model host, *Arabidopsis* (Ascencio-Ibanez et al. 2008). Furthermore, using the natural host cassava, this research has also revealed differences in gene expression, which in a broader and more comprehensive bioinformatics study, as well as *in planta* studies, will prove extremely interesting. Gene lists generated from eleven different transcriptome studies of *Arabidopsis*-virus interactions (reviewed in Postnikova and Nemchnov, et al. 2012) were found to overlap significantly, with only minor differences among the significantly overrepresented functional categories, re-emphasizing a general conformity between most virus-host interactions. Downregulated genes tended to be common regardless of the virus, while upregulated genes tended to be more diverse. However, this was shown only in the model host *Arabidopsis*, and more

bioinformatics and genomic studies are needed to compare single virus infections in multiple hosts. With the available sequencing of many more genomes, such as *N. benthamiana* and tomato (Bombarely et al. 2012; The Tomato Genome Consortium (2012). Comparative studies will prove both interesting and valuable. In cassava, many unique host-responsive genes were identified, and differences in TME3 and T200 will form the basis for further work to identify possible resistance mechanisms that can be exploited in crop improvement.

We were able to successfully address the questions that set out the framework of this PhD. This work therefore contributes significantly to the existing limited knowledge specifically underlying gene expression changes in GV-host interactions. Through a better understanding of the effect of pathogens on host transcription, an area that is deserving of more attention in plant systems, my hope is that the findings presented in this PhD will contribute to long term goals of devising new methods of disease control against SACMV-[ZA:99] and understanding the complex interconnected mechanisms involved in virus-host interactomes.

Chapter 6

References

- Agrios. G.N.** (2005) Plant pathology- fifth edition. Elsevier Academic Press, San Diego, California
- Agnew. P, Koella, J.C., Michalakakis, Y.** (2000) Host life history responses to parasitism. *Microbes Infection*. 2: 891-896.
- Ahlquist, P., Noueiry, A.O., Lee, W.M., Kushner, D.B., and Dye, B.T.** (2003) Host factors in positive-strand RNA virus genome replication. *Journal of Virology*. 77: 8181–8186.
- Akano, A.O., Dixon, A.G.O., Mba, C., Barrera, E., and Fregene, M.** (2002) Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theoretical and Applied Genetics*. 105: 521-525.
- Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, PietrellaM, et al.** (2009) Comparative 454 pyrosequencing of transcripts from olive genotypes during fruit development. *BMC Genomic*. 10: 399.
- Alba, R., Fei, Z., Payton, P., Liu, Y., Moore, S.L., Debbie, P., Cohn, J., D'Ascenzo, M, Gordon, J.S., Rose, J.K, Marton, G., Tanksley, S.D., Bouzayen, M., Jahn, M.M. and Giovannoni, J.** (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. *Plant Journal*. 39: 697–714.
- An, D., Yang, J., and Zhang, P.** (2012) Transcriptome profiling of low temperature-treated cassava apical shoots showed dynamic responses of tropical plant to cold stress. *BMC Genomics*. 13: 64
- Anaya-Lopez, J.L., Pérez-Mora, E., Torres-Pacheco, I., Muñoz-Sanchez, C.I., Guevara-Olvera, L., Gonzalez-Chavira, M.M., Ochoa-Alejo, N., Rivera-Bustamante, R.F., Guevara-Gonzalez, R.G.** (2005) Inducible gene expression by pepper huasteco virus in *Capsicum chinense* plants with resistance to geminivirus infections. *Canadian Journal of Plant Pathology*. 27: 276-282.
- Anders, S. and Huber, W.** (2010). Differential expression analysis for sequence count data. *Genome Biology*. 11:R106.
- Anderson, M. and Wilson, F.** (2000) Growth, maintenance and use of Arabidopsis genetic resources, in: *Arabidopsis: A Practical Approach*, pg 1-26.
- Anderson, J .V., Delseny, M, Fregene, M. A, Jorge, V., Mba, C, Lopez, C., Restrepo, S, Soto, S, Piegu, B., Verdier V, Cooke, R, Tohme J, Horvath, D. P.** (2004) An EST resource for cassava and other species of Euphorbiaceae. *Plant Molecular Biology*. 56:527–539.
- Andleeb, S., Amin, I., Bashir, A., Briddon, R.W., and Mansoor , S.** (2010) Transient expression of betaC1 protein differentially regulates host genes related to stress response, chloroplast and mitochondrial functions. *Virology Journal*. 7: 373.
- Aoki, K., Kragler, F., Xoconostle-Cazares, B., and Lucas,W.J.** (2002) A subclass of plant heat shock cognate 70 chaperones carries a motif that facilitates trafficking through plasmodesmata. *Proceedings of the National Academy of Sciences of the United States of America*. 99: 16342–16347.

- Aparicio, F., Thomasm, C.L., Lederer, C., Niu, Y., Wang, D., and Maule, A.J.** (2005) Virus induction of heat shock protein 70 reflects a general response to protein accumulation in the plant cytosol. *Plant Physiology*. 138: 529-536.
- Aranda, M. A., Escaler, M., Wang, D. and Maule, A. J.** (1996) Induction of HSP70 and polyubiquitin expression associated with plant virus replication. *Proceedings of the National Academy of Sciences of the United States of America*. 93: 15289–15293.
- Aranda, M. and Maule, A.** (1998). Virus-induced host gene shutoff in animals and plants. *Virology*. 243: 261–267.
- Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.L., Gomez-Gomez, L., Boller, T., Ausubel, F.M. and Sheen, J.** (2002). MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature*. 415: 977-983.
- Ascencio-Ibáñez, J.T., Sozzani, R., Lee, T., Chu, T., Wolfinger, R.D., Cella, R., and Linda Hanley-Bowdoin** (2008) Global Analysis of *Arabidopsis* Gene Expression Uncovers a Complex Array of Changes Impacting Pathogen Response and Cell Cycle during Geminivirus Infection. *Plant Physiology*. 148: 436–454.
- Avisar, D., Prokhnevsky, A.I., and Dolja, V.V.** (2008) Class VIII myosins are required for plasmodesmatal localization of a closterovirus Hsp70 homolog. *Journal of Virology*. 82: 2836–2843.
- Ayling, S., M. Ferguson, S. Rounsley, and P. Kulakow** (2012) Information resources for cassava research and breeding. *Tropical Plant Biology*. 5(1): 140-15.
- Babu, M., Griffiths, J.S., Huang, T.S., and Wang, A.** (2008) Altered gene expression changes in *Arabidopsis* leaf tissues and protoplasts in response to Plum pox virus infection. *BMC Genomics*. 9: 325.
- Bar-Or, C., Czosnek, H. and Koltai, H.** (2007) Cross-species hybridizations: a developing tool for studying species diversity. *Trends in Genetics*. 23:200–207.
- Bartel D.P.** (2004) MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*. 116: 281–297.
- Beckers G.J.M., Jaskiewicz M., Liu Y., Underwood W.R., He S.Y., Zhang S., Conrath U.** (2009) Mitogen-activated protein kinases 3 and 6 are required for full priming of stress responses in *Arabidopsis thaliana*. *Plant Cell*. 21: 944–953.
- Baulcombe, D.** (2004) RNA silencing in plants. *Nature*. 431: 356-363.
- Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., and Edgar, R.** (2007) NCBI GEO: mining tens of millions of expression profiles–database and tools update. *Nucleic Acids Research*. 35: D760–D765.
- Bayne, E.H., Rakitina, D.V., Morozov, S.Y. and Baulcombe, D.C.** (2005) Cell-to-cell movement of Potato Potexvirus X is dependent on suppression of RNA silencing. *Plant Journal*. 44: 471–482.
- Bent, A.F.** (1996). Plant disease resistance genes: Function meets structure. *Plant Cell*. 8: 1757–1771.

- Bendahmane, A., Ko hm, B. A., Dedi, C. & Baulcombe, D. C.** (1995) The coat protein of potato virus X is a strain-specific elicitor of Rx1-mediated virus resistance in potato. *Plant Journal*. 8: 933–941.
- Bendahmane, A., Kanyuka, K. and Baulcombe, D.C.** (1999). The Rx gene from potato controls separate virus resistance and cell death responses. *Plant Cell*. 11: 781–792
- Berg, A., Meza, T.J., Mahic, M., Thorstensen, T., Kristiansen, K., and Aalen, R.B.** (2003) Ten members of the *Arabidopsis* gene family encoding methyl-CpG-binding domain proteins are transcriptionally active and at least one, AtMBD11, is crucial for normal development. *Nucleic Acids Research*. 31: 5291–5304.
- Berger S., Papadopoulos M., Schreiber U., Kaiser W., Roitsch T.** (2004) Complex regulation of gene expression, photosynthesis and sugar levels by pathogen infection in tomato. *Physiologia Plantaru*. 122: 419–428.
- Berrie, L. C., Palmer, K. E., Rybicki, E. P. and Rey, M. E. C.** (1998). Molecular characterisation of a distinct South African cassava infecting geminivirus. *Archives of Virology*. 143: 2253–2260.
- Berrie, L. C., Rybicki, E. P. and Rey, M. E. C.** (2001) Complete nucleotide sequence and host range of South African cassava mosaic virus: further evidence for recombination amongst begomoviruses. *Journal of General Virology*. 82: 53–58.
- Bieri, S., Potrykus I, Futterer, J** (2002). Geminivirus sequences as bi-directional transcription termination/polyadenylation signal for economic construction of stably expressed transgenes. *Molecular Breeding*. 10:107–17.
- Bisaro, D.M.** (2006.) Silencing suppression by geminivirus proteins. *Virology*. 344: 158–168.
- Blagbrough, I.S, Bayoumi, S.A.L, Rowan, M.G, Beeching, Jr** (2010). Cassava: An appraisal of its phytochemistry and its biotechnological prospects. *Phytochemistry*. 71: 1940 – 1951.
- Boevink, P and Oparka, K.J.** (2005) Virus-host interactions during movement processes. *Plant Physiology*. 138(4):1815–21.
- Bolouri-Moghaddam, M.R., and Van den Ende, W.** (2012). Sugars and plant innate immunity. *Journal of Experimental Botany*. 63(11): 3989–98.
- Bombarely. A., Rosli ,H.G., Vrebalov, J., Moffett, P., Mueller. L., and Martin, G.** (2012). A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Molecular Plant-Microbe Interactions*. 25(12): 1523–1530.
- Borsani, O, Zhu, J.H., Verslues, P.E, Sunkar,R., Zhu,J.K.** (2005). Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell*. 123: 1279–1291.
- Bouche N., Lauressergues D., Gasciolli V., and Vaucheret H.** (2006). An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO Journal*. 25: 3347.

- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C.P., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M.** (2001) Minimum information about a microarray experiment (MIAME)—toward standards for microarray data, *Nature Genetics*. 29: 365-371.
- Brazma, A. Krestyaninov, M and Sarkans, U.** (2006) Standards for systems biology. *Nature Reviews in Genetics*. 7:593–605.
- Briddon, R.W., Pinner, M.S., Stanley, J. and Markham, P.G.** (1990) Geminivirus coat protein gene replacement alters insect specificity. *Virology*. 177: 85-9.
- Brigneti, G., Voinnet, O., Li, W.X., Ji, L., Ding, S., Baulcombe, D.** (1998). Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. *EMBO Journal*. 17: 6739–6746.
- 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 Molecular Biology*. 18: 703–712.
- 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. *Journal of Virology*. 83: 5005–5013.
- Burdon, J. and Thrall, P.** (2009). Coevolution of plants and their pathogens in natural habitats. *Science*. 324(5928): 755 – 756.
- Bustin, S.A., Benes, V., Garson, J.A., Hellems, J., Huggett, J., Kubista, M., Mueller, R, Nolan, T., Pfaffl M.W., Shipley, G.L., Vandesompele, J., and Wiltwer, C.T.** (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*. 55(4): 611-622.
- Cao, H., Bowling, S.A., Gordon, A.S., and Dong, X.** (1994). Characterization of an *Arabidopsis* mutant that is nonresponsive to inducers of systemic acquired resistance. *Plant Cell*. 6: 1583-1592.
- Carrington J.C., Ambros V.** (2003) Role of microRNAs in plant and animal development. *Science*. 301:336–338.
- Carvalho C. M.; Fontenelle M. R.; Florentino L. H.; Santos A. A.; Zerbini F. M.; and Fontes E. P.** (2008a) A novel nucleocytoplasmic traffic GTPase identified as a functional target of the bipartite geminivirus nuclear shuttle protein. *Plant Journal*. 55: 869–880.
- Carvalho C. M.; Machado J. P.; Zerbini F. M., and Fontes E. P.** (2008b) NSP-interacting GTPase: a cytosolic protein as cofactor for nuclear shuttle proteins. *Plant Signaling and Behavior*. 3: 752–754.

- Castillo, A.G, Collinel, D., Deret, S., Kashoggi, A., and Bejarono, E.R.** (2003) Dual interaction of plant PCNA with geminivirus replication accessory protein (REn) and viral replication protein (Rep). *Virology*. 312: 381-394.
- Castillo, A. G.; Kong, L. J.; Hanley-Bowdoin, L. And Bejarano, E. R.** (2004). Interaction between a geminivirus replication protein and the plant sumoylation system. *Journal of Virology*. 78 :2758–2769.
- Ceballos, H., Iglesias, C.A., Perez, .JC., and Dixon, A.G.O.** (2004) Cassava breeding: opportunities and challenges. *Plant Molecular Biology*. 56: 503-516.
- Chan, A.P., Crabtree, J., Zhao Q, et al.** (2010) Draft genome sequence of the oilseed species *Ricinus communis*. *Nature Biotechnology*. 28(9): 951–956.
- Chellappan, P., Vanitharani, R., and Fauquet, C. M.** (2004) Short interfering RNA accumulation correlates with host recovery in DNA virus-infected hosts, and gene silencing targets specific viral sequences. *Journal of Virology*. 78:7465–7477.
- Chen M.H., and Citovsky, V.** (2003) Systemic movement of a tobamovirus requires host cell pectin methylesterase. *Plant Journal*. 35: 386-392.
- Chen, Y-F, Li, L-Q, Xu, Q., Kong, Y.H., Wang, H., Wu, W-H.** (2009) The WRKY6 transcription factor modulates PHOSPHATE1 expression in response to low Pi stress in *Arabidopsis*. *Plant Cell*. 21: 3554–3566.
- Chen, M.H., Sheng, J., Hind, G., Handa, A., and Citovsky, V.** (2000) Interaction between the tobacco mosaic virus movement protein and host cell pectin methylesterases is required for viral cell-to-cell movement. *EMBO Journal*. 19: 913–920.
- Chisholm, S.T., Coaker ,G, Day, B., Staskawicz, B.J.** (2006) Host-microbe interactions: shaping the evolution of the plant immune response. *Cell*. 124: 803–814.
- Chiu, M.H, Chen, I.H., Baulcombe, D.C, Tsai, CH.** (2010) The silencing suppressor P25 of Potato virus X interacts with Argonaute1 and mediates its degradation through the proteasome pathway. *Molecular Plant Pathology*. 11(5):641-9.
- Chong J, Baltz, R, Schmitt, C., Beffa, R., Fritig, B., Saindrenan, P.** (2002) Downregulation of a pathogen-responsive tobacco UDP-Glc:phenylpropanoid glucosyltransferase reduces scopoletin accumulation, enhances oxidative stress, and weakens virusresistance. *The Plant Cell*. 14: 1093–1107.
- Cock, J.H and Reyes, J.A.** (1985) Cassava Research Production and utilization. Cassava Program. The United Nations Development Programme (UNDP)/The International Center for Tropical Agriculture (CIAT). pp 745.
- Collier, S. M. & Moffett, P.** (2009) NB-LRRs work a ‘bait and switch’ on pathogens. *Trends in Plant Science*. 14: 521–529.

- Cooley, M.B., Pathirana, S. Wu, H.J., Kachroo, P., and Klessig, D.F.** (2000) Members of the *Arabidopsis* HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens. *Plant Cell*. 12: 663–676.
- Csorba, T, Pantaleo, V, and Burgyán, J.** (2009) RNA silencing: An antiviral mechanism. *Advanced Virus Research*. 75: 35–71.
- Cui, X., Li, G., Wang, D., Hu, D. and Zhou, X.** (2005) A begomovirus DNA β -encoded protein binds DNA functions as a suppressor of RNA silencing, and targets the cell nucleus. *Journal of Virology*. 79: 10764–10775.
- Culver, J.N., and Padmanabhan, M.S.** (2007). Virus-induced disease: Altering host physiology one interaction at a time. *Annual Reviews in Phytopathology*. 45: 221-243.
- Dahniya, M. T.** (1994) An overview of cassava in Africa. *Crop Science Journal*, Makerere. University, Uganda. pp. 337-343.
- Dardick, C.** (2007) Comparative expression profiling of *Nicotiana benthamiana* leaves systemically infected with three fruit tree viruses. *Molecular Plant-Microbe Interactions*. 20(8): 1004-1017.
- Delaney, T., Friedrich, L., and Ryals, J.** (1995). *Arabidopsis* signal transduction mutant defective in chemically and biologically induced disease resistance. *Proceedings in the National Academy of Science USA*. 92: 6602-6606.
- 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 defense. *Science*. 313: 68–71.
- Dempsey, D.A., Shah, J., and Klessig, D.F.** (1999) Salicylic acid and disease resistance in plants. *Critical Reviews in Plant Science*. 18: 547–575.
- Department of Agriculture, Forestry and Fisheries, Republic of South Africa**, 2010. <http://www.nda.agric.za/docs/Brochures/ProdGuideCassava.pdf>
- Deslandes, L. Olivier, J, Peeters, N, Feng, D.X., Khounlotham. M., Boucher. C., Somssich, I, Genin, S, Marco, Y.** (2003) Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proceedings in the National Academy of Science USA*. 100: 13 8024–8029.
- De Souza, C.R., Carvalho, L.J., de Mattos Cascardo J.C.** (2004). Comparative gene expression study to identify genes possibly related to storage root formation in cassava. *Protein and Peptide Letters*. 11: 577-582.
- De Vos, M., Van Oosten, V.R., Van Poecke, R.M.P., Van Pelt, J.A., Pozo, M.J., Mueller, M.J., Buchala, A.J., Métraux, J-P., Van Loon, L.C., Dicke, M., and Pieterse, C.M.J.** (2005) Signal signature and transcriptome changes of *Arabidopsis* during pathogen and insect attack. *Molecular Plant-Microbe Interactions*. 18(9), 923- 937.
- DeVries, J., and Toenniessen, G.** (2001) Securing the harvest: biotechnology, breeding and seed systems for African crops. CABI publishing, Oxon, UK.

- Dixon, A.G.O., Whyte, J.B.A., Mahungu, N.M. and 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: Cassava, an ancient crop for modern times: Food, health, culture, N.J. Taylor, F. Ogbé and C.M. Fauquet (eds), pp. S8-05, Donald Danforth Plant sciences Center, St. Louis, USA.
- Dong, J., Chen, C., and Chen, Z.** (2003) Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. *Plant Molecular Biology*. 51: 21–37.
- Dong Z., Han M.H., Fedoroff N.** (2008). The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proceeding of the National Academy of Sciences USA*. 105: 9970–9975.
- Dorokhov, Y.L., Makinen, K., Frolova, O.Y., Merits, A., Saarinen, J., Kalkkinen, N., Atabekov, J.G., and Saarma, M.** (1999) A novel function for a ubiquitous plant enzyme pectin methylesterase: the host-cell receptor for the tobacco mosaic virus movement protein. *FEBS Letters*. 461: 223–228.
- Dowd, C., Wilson, I.W., and McFadden, H.** (2004) Gene expression profile changes in cotton root and hypocotyl tissues in response to infection with *Fusarium oxysporum* f. sp. *vasinfectum*. *Molecular Plant-Microbe Interaction*. 17: 654–667.
- Doyle, J.J. and Doyle, J.L.**, (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*. 19: 11-15.
- Dressman, D., Yan, H., Traverso, G., Kinzler, K.W., and Vogelstein, B.** (2003) . Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceeding of the National Academy of Science. U.S.A.* 100: 8817-8822
- Durner, J., Shah, J., and Klessig, D.F.** (1997) Salicylic acid and disease resistance in plants. *Trends in Plant Science*. 2: 266–274.
- Durrant, W.E, and Dong, X.** (2004) Systemic acquired resistance. *Annual Reviews in Phytopathology* 42: 185–209.
- Edgar, R., Domrachev, M., and Lash, A.E.** (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*. 30(1): 207-210.
- Elena, S.F., Carrera, J. and Rodrigo, G.** (2011) A systems biology approach to the evolution of plant virus interactions. *Current Opinion in Plant Biology*. 14: 372-377.
- Ellis J.G., Lawrence G.J., Luck J.E., Dodds P.N.** (1999) The identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *Plant Cell*. 11: 495–506.
- El-Sharkawy, M.A.** (2003) *Cassava* biology and physiology. *Plant Molecular Biology*. 53: 621–641.
- Epel, B.L.** (2009) Plant viruses spread by diffusion on ER-associated movement-protein-rafts through plasmodesmata gated by viral induced host β -1,3-glucanases. *Seminars in Cell and Developmental Biology*. 20: 1074–1081.

- Erickson, F.L., Holzberg, S., Calderon-Urrea, A., Handly, V., Axtell, M, Corr, C, Baker, B.** (1999). The helicase domain of the TMV replicase proteins induces the N-mediated defense response in tobacco. *Plant Journal*. 18: 67–75
- 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 Research*. 25: 3445–3450
- Escaler, M., Aranda, M.A., Thomas, C.L., and Maule, A.J.** (2000) Pea embryonic tissues show common responses to the replication of a wide range of viruses. *Virology*. 267: 318–325.
- Eulgem, T., and Somssich, I.E.** (2007) Networks of WRKY transcription factors in defense signaling. *Current Opinion in Plant Biolog.*, 10: 366-371.
- Eybishtz. A., Peretz, Y., Sade, D., Akad, F., and Czosnek, H.** (2009) Silencing of a single gene in tomato plants resistant to *Tomato yellow leaf curl virus* renders them susceptible to the virus. *Plant Molecular Biology*. 71: 157–171.
- FAO.** (2008) Cassava for food and energy security. FAO Media Center, Rome. <http://www.fao.org/newsroom/en/nes/2008/1000899/index.html>. Accessed September 2012
- FAO.** Cassava varietal improvement for processing and utilization in livestock feeds. <http://www.fao.org/wairdocs/ilri/x5458e/x5458e0b.htm#TopOfPage> Accessed 25 September 2013
- Falk, A., Feys, B.J., Frost, L.N., Jones, J.D.G. and Daniels, M.J. and Parker, J.E.** (1999) eds.1, an essential component of R gene-mediated resistance in Arabidopsis has homology to eukariotic lipases. *Proceeding from the National Academy of Science USA*. 96: 3292- 3297.
- Fauquet, C. M. & Fargette, D.** (1988) *Proceedings of the International Seminar: African Cassava Mosaic Disease and its Control*. Ede, Netherlands: CTA/ORSTOM.
- Fauquet, C.M., and Stanley, J.** (2003) Geminivirus classification and nomenclature: progress and problems. *Annual Applied Biology*. 142: 165–189.
- Fauquet, C.M., Bisaro, D.M., Briddon, R.W., Brown, J.K., Harrison, B.D., Rybicki, E.P., Stenger, D.C. and Stanley, J.** (2003). Revision of taxonomic criteria for species demarcation in the family Geminiviridae, and an updated list of begomovirus species. *Archives of Virology*. 148: 405-421.
- Fedurco, M., Romieu, A., Williams, S., Lawrence, I., and Turcatti, G.** (2006) B TA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Research*. 34: e22.
- Fernandez Calvino, L.; Faulkner, C., and Maule, A.** (2011) Plasmodesmata as active conduits for virus cell-to-cell movement. In Recent Advances in Plant Virology, eds Caranta C, Aranda M-A, Tepfer M, Lopez Moya J-J (Caister Academic Press, Norkfolk), pp 47-74.
- Feulner, Gretchen.** “The Modulation of *Nicotiana benthamiana* Gene Expression by Red Clover Necrotic Mosaic Virus”. Master's thesis, North Carolina State University, (2007).

- Flor, H.H.** (1971) Current status of the gene-for-gene concept. *Annual Reviews in Phytopathology*, 9: 275–296.
- Fotopoulos, V., Gilbert, M.J., Pittman, J.K., Marvier, A.C., Buchanan, A.J., Sauer, N., Hall, J.L., Williams L.E.** (2003) The monosaccharide transporter gene, AtSTP4, and the cell-wall invertase, Atbetafruct1, are induced in *Arabidopsis* during infection with the fungal biotroph *Erysiphe cichoracearum*. *Plant Physiology*, 132: 821–829.
- Frank, S.A.** (1992). Models of plant-pathogen coevolution. *Trends in Genetics*. 8(6):213-9.
- Franssen, S.U., Shrestha, R.P., Bräutigam, A., Bornberg-Bauer, E, Weber, A.P.M.** (2011) Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *BMC Genomics*. 12:227.
- Fregene, M., Angel, F., Gomez, R., Rodriguez, F., Chavarriaga, P., Roca, W., Tohme, J., Bonierbale, M.** (1997) A molecular genetic map of cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics*. 95: 431-441.
- Fregene, M., Matsumura, H., Akano, A., Dixon, A., and Terauchi, R.** (2004) Serial analysis of gene expression (SAGE) of host-plant resistance to the cassava mosaic disease. *Plant Molecular Biology*. 56: 563-571.
- Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H., Ohme-Takagi, M.** (2000) *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell*. 12: 393–404.
- Fullwood, M.J., Wei, C., Liu, E.T. and Ruan, Y.** (2009) Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses, *Genome Research*. 19 (4): 521-532.
- 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 Reports*. 7: 1168–1175.
- Gassmann, W., Hinsch, M.E., and Staskawicz, B.J.** (1999) The *Arabidopsis* RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *Plant Journal*. 20, 265–277.
- García-Marcos, A., Pacheco, R., Martiáñez, J., González-Jara, P., Díaz-Ruiz, J.R., and Tenllado, F.** (2009) Transcriptional changes and oxidative stress associated with the synergistic interaction between Potato virus X and Potato virus Y and their relationship with symptom expression. *Molecular Plant-Microbe Interactions*. 22: 1431-1444.
- Garcia-Neria, M. A., and Rivera-Bustamante, R. F.** (2011) Characterization of Geminivirus Resistance in an Accession of *Capsicum chinense* Jacq. *Molecular Plant-Microbe Interactions*. 24(2): 172-182.
- Garga R, Patel RK, Tyagi AK, Jain M.** (2011) *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research*. 18: 53–63.

- Gehrig, H.H., Winter, K., Cushman, J., Borland, A. and Taybi, T.** (2000) An improved RNA isolation method for succulent plant species rich in polyphenols and polysaccharides. *Plant Molecular Biology Reporter*. 18 (4): 369-376.
- Gilbertson, R.L., and Lucas, W.J.** (1996) How do viruses traffic on the 'vascular highway'? *Trends in Plant Science*. 1: 260–268.
- Glotzer, J.B., Saltik, M., Chiocca, S., Michou, A.I., Moseley, P., Cotten, M.** (2000) Activation of heat-shock response by an adenovirus is essential for virus replication. *Nature*. 407: 207–211.
- Golem, S., and Culver, J. N.** (2003) Tobacco mosaic virus induced alterations in the gene expression profile of *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions*. 16: 681-688.
- Goodin, M.M., Zaitlin., Naidu, R. A., And Lommel, S.A.** (2008) *Nicotiana benthamiana*: It's history and future as a model for plant-pathogen interactions. *Molecular Plant-Microbe Interactions*. 21(8): 1015-1026.
- Gomez-Gomez, L., and Boller, T.** (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Molecular Cell*. 5: 1003–1011.
- Gómez-Ariza, J., Campo, S., Rufat, M., Estopà, M., Messeguer, J., San Segundo, B.; Coca, M.** (2007) Sucrose-mediated priming of plant defense responses and broad-spectrum disease resistance by overexpression of the maize pathogenesis-related PRms protein in rice plants. *Molecular Plant-Microbe Interactions*. 20: 832–842.
- González, I., Martínez, L., Rakitina, D.V., Lewsey, M.G., Atencio, F.A., Llave, C., Kalinina, N.O., Carr J.P., Palukaitis, P., Canto, T.** (2010). Cucumber mosaic virus 2b protein subcellular targets and interactions: their significance to RNA silencing suppressor activity. *Molecular Plant-Microbe Interactions*. 23: 294–303.
- Goto, K., Kobori, T., Kosaka, Y., Natsuaki, T., Masuta, C.** (2007). Characterization of silencing suppressor 2b of cucumber mosaic virus based on examination of its small RNA-binding abilities. *Plant Cell Physiology*. 48: 1050–1060.
- Grattapaglia, D.; Silva-Junior, O.B.; Kirst, M.; de Lima, B.M.; Faria, D.A.; and Pappas, G.J.** (2011). High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. *BMC Plant Biology*. 11: 65.
- Greenberg, J.** (1997) Programmed cell death in plant-pathogen interactions. *Annual Reviews in Plant Physiology and Plant Molecular Biology*. 48: 525–545.
- Greenberg, J. and Yao, N.** (2004). The role and regulation of programmed cell death in plant-pathogen interactions. *Cell Microbiology*. 6: 201–211.
- Gutierrez, C.** (2002). Strategies of geminivirus DNA replication and cell cycle interference. *Physiology and Molecular Biology of Plants*. 60: 219-230.
- Gutierrez, C., Ramirez, P.E., Castellano, M.M., Sanz-Bugos, A.P., Luque, A., and Missich, R.** (2004). Geminivirus DNA replication and cell cycle interactions. *Veterinary Microbiology*. 98: 111-119.

- Hagen, C., M. R. Rojas, T. Kon, and R. L. Gilbertson.** (2008) Recovery from *Cucurbit leaf crumple virus* (family Geminiviridae, genus *Begomovirus*) infection is an adaptive antiviral response associated with changes in viral smallRNAs. *Phytopathology*. 98:1029–1037.
- Hamilton, A. and Baulcombe, D.** (1999) A species of small antisense RNA in post-transcriptional gene silencing in plants. *Science*. 286: 950-952.
- Hanley-Bowdoin, L., Settlege, S.B., Orozco, B.M., Nagar, S. and Robertson, D.** (2000) Gemiviruses: models for plant DNA replication, transcription and cell cycle regulation. *Critical Reviews in Biochemical and Molecular biology*. 35: 105-140
- Hanley-Bowdoin, L, Settlege, S.B, Robertson, D.** (2004) Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. *Molecular Plant Pathology*. 5: 149–156.
- Hanssen, I.M., Peter van Esse, H., Ballester, A.R., Hogewoning, S.W., Parra, N.O., Paeleman, A., Lievens, B., Bovy, A.G. and Thomma, B.P.H.J.** (2011) Differential tomato transcriptomic responses induced by Pepino mosaic virus isolates with differential aggressiveness. *Plant Physiology*. 156: 301-318.
- Harries P.A., Park J.W., Sasaki N., Ballard K.D., Maule A.J., Nelson R.S.** (2009). Differing requirements for actin and myosin by plant viruses for sustained intercellular movement. *Proceedings of the National Academy of Sciences of the United States of America*. 106: 17594–17599.
- Harries, P.A., Schoelz, J.E, and Nelson, R.S.** (2010) Intracellular transport of viruses and their components: utilizing the cytoskeleton and membrane highways. *Molecular Plant-Microbe Interactions*. 23: 1381-1393.
- Harrington, C. A., Rosenow, C., and Retief, J.** (2000) Monitoring gene expression using DNA microarrays. *Current Opinion in Microbiology*. 3: 285–291
- Havelda, Z., Várallyay, E., Válóczy, A., and Burgyán, J.** (2008) Plant virus infection-induced persistent host gene downregulation in systemically infected leaves. *Plant Journal*. 55: 278–288.
- Hays, J.B.** (2002) *Arabidopsis thaliana*, a versatile model system for study of eukaryotic genome-maintenance functions. *DNA Repair*. 1: 579-600.
- Heinlein, M.** (2002). The spread of *Tobacco mosaic virus infection*: insights into the cellular mechanism of RNA transport. *Cellular and Molecular Life Sciences*. 59: 58–82.
- HerbersK, Takahata, Y, Melzer, M, Mock, H.P, Hajirezaei, M, Sonnewald, U.** (2000) Regulation of carbohydrate partitioning during the interaction of potato virus Y with tobacco. *Molecular Plant Pathology*. 1:51–59.
- Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C.** (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science*. 308: 118-120.
- Hillier, L.W., Marth, G.T., Quinlan, A.R., Dooling, D., Fewell, G., Barnett, D., Fox, P., Glasscock, J.I., Hickenbotham, M., Huang, W., Magrini, V.J., Richt, R.J., Sander, S.N., Stewart,D.A.,**

- Stromberg, M., Tsung, E.F., Wylie, T., Schedl, T., Wilson, R.K. and Mardis, E.R.** (2008) Whole-genome sequencing and variant discovery in *C. elegans*. *Nature Methods*. 5:183–188
- Hillocks.R.J,** (1997). Cassava virus diseases and their control with special reference to southern Tanzania, *Integrated Pest Management Reviews*. 2:125-138.
- Hodgson, R.A.J., Beachy, R.N., and Pakrasi, H.B.** (1989). Selective inhibition of photosystem II in spinach by *tobacco mosaic virus*: an effect of the viral coat protein. *FEBS Letters*. 245: 267–270.
- Hofmann, C., Niehl, A., Sambade, A., Steinmetz, A., and Heinlein, M.** (2009). Inhibition of Tobacco mosaic virus movement by expression of an actin-binding protein. *Plant Physiology*. 149: 1810–1823.
- Horvath, D.M. and Huang, D.J.and.Chua, N.-H.** (1998) Four classes of salicylate-induced tobacco genes. *Molecular Plant-Microbe Interactions*. 11: 895- 905.
- Horvath, D. P., R. Schaffer, M. West, and E. Wisman.** (2003) *Arabidopsis* microarrays identify conserved and differentially-expressed genes involved in shoot growth and development from distantly related plant species. *Plant Journal*. 34: 125–134.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A.** (2009a) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*. 4 (1): 44–57.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A.** (2009b) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*. 37(1): 1-13.
- Hückelhoven ,R.** (2007). Cell Wall-Associated Mechanisms of Disease Resistance and Susceptibility. *Annual Reviews in Phytopathology*. 45S: 101-127.
- Huffaker A.,Pearce G., Ryan C.A.** (2006). An endogenous peptide signal in *Arabidopsis* activates components of the innate immune response. *Proceedings of the National Academy of Science USA*. 103: 10098–10103.
- Iglesias, V.A. and Meins., F.** (2000). Movement of plant viruses is delayed in a β -1,3-glucanase-deficient mutant showing a reduced plasmodesmatal size exclusion limit and enhanced callose deposition. *Plant Journal*. 21: 157–166.
- Innes, R.W.** (2001). Mapping out the roles of MAP kinases in plant defense. *Trends in Plant Sciences*. 6: 392–394.
- Itaya, A., Matsuda, Y., Gonzales, R. A., Nelson, R. S., and Ding, B.** (2002). Potato spindle tuber viroid strains of different pathogenicity induces and suppresses expression of common and unique genes in infected tomato. *Molecular Plant-Microbe Interactions*. 15: 990-999.
- Jeske, H.** (2009) Geminiviruses. In *Torque Teno Virus: the Still Elusive Human Pathogens*, pp. 185–226. Edited by H. zur Hausen & E.M. de Villiers. Berlin: Springer

- Jaillon O, Aury JM, Noel B, et al.** (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 449 (7161): 463–7.
- Jaramillo, G., Morante, N., Perez, J.C., Calle, F., Ceballos, H., Aria, B. and Bellotti, A.A.** (2005). Diallel analysis in cassava adapted to the mid-altitude valleys environment. *Crop Science*. 45: 1058-1063.
- Jia, M.A., Li, Y., Lei, L., Di, D., Miao, H., and Fan, Z.** (2012). Alteration of gene expression profile in maize infected with a double-stranded RNA fijivirus associated with symptom development. *Molecular Plant Pathology*. 13: 251–262.
- Johnson, C.S., Kolevski, B., and Smyth, D.R.** (2002). TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor. *Plant Cell*. 14: 1359–1375.
- Jones, J.D.G., and Dangl J.L.** (2006). The plant immune system. *Nature* 444: 323–329
- Kahn, T.W., Lapidot, M., Heinlein, M., Reichel, C., Cooper, B., Gafny, R., and Beachy, R.N.** (1998). Domains of the TMV movement protein involved in subcellular localisation. *Plant Journal*. 15: 15–25
- Kasajima, I.** (2010). WRKY6 is involved in the response to boron deficiency in *Arabidopsis thaliana*. *Plant Physiology*. 139: 80-92.
- Kawano, K.** (1978). Genetic improvement of cassava (*Manihot esculenta* Crantz) for productivity. *Tropical Agricultural Research Series*. 11: 9-21.
- Keen, N.T.** (1992). The molecular biology of disease resistance. *Plant Molecular Biology*. 19: 109–122.
- Khraiwesh, B., Zhu, J.K., Zhu, J.H.** (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica et Biophysica Acta*. 819: 137–148.
- Kolomiets, M.V., Hannapel, D.J., Chen, H., Tymeson, M., Gladon, R.J.** (2001) Lipxygenase is involved in the control of potato tuber development. *The Plant Cell*. 13: 613–626.
- Kong, L., Orozco, B.M., Roe, J.L., Nagar, S., Ou, S., Feiler, H.S., Durfee, T., Miller, A.B., Gruissem, W., Robertson, D., and Hanley-Bowdoin, L.** (2000) A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *EMBO Journal*. 19: 3485-3495.
- Kuella, M., Hsiao, L., Yoshida, T., Randall, J.D., Chow, G., Sarang, S.S., Roderick, J.J. and Gullans, S.R.** (2001) DNA microarray analysis of complex biological systems. *Journal of American society of Nephrology*. 12: 1072-1078.
- Kunkel, B.N. and Brooks, D.M.** (2002) Cross talk between signaling pathways in plant defense. *Current Opinion in Plant Biology*. 5: 325–331.
- Lakatos, L., Szittyá, G., Silhavy, D., Burgyan, J.** (2004) Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. *EMBO Journal*. 23:876–884.

- Lakatos, L., T. Csorba, V. Pantaleo, E. J. Chapman, J. C. Carrington, Y. P. Liu, V. V. Dolja, L. F. Calvino, J. J. Lopez-Moya, and J. Burgyn.** (2006). Small RNA binding is a common strategy to suppress RNA silencing by several viral suppressors. *EMBO Journal*. 25: 2768-80.
- Lamb, C and Dixon R.A.** (1997). The oxidative burst in plant disease resistance. *Annual Reviews of Plant Physiology and Plant Molecular Biology*. 48: 251-275.
- Laminski, S. Robinson, E.R. and Gray, V.M.** (1997). Application of molecular markers to describe South Africa elite cassava cultivars. *African Journal of Root and Tuber Crops*. 2: 132-134.
- Larionov, A.; Krause, A.; and Miller, W.** (2005). A standard curve based method for relative real time PCR data processing. *BMC Bioinformatics*. 6: 62.
- Latham, J. R., Saunders, K., Pinner, M. S. and Stanley, J.** (1997) Induction of plant cell division by beet curly top virus gene C4. *Plant Journal*. 11: 1273–1283.
- Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S., Schell, J. & Gronenborn, B.** (1995) In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proceeding of the National Academy of Sciences, U S A*. 92: 3879–3883.
- Lazarowitz S.G., and Beachy, R.N.** (1999) Viral movement proteins as probes for intracellular and intercellular trafficking in plants. *Plant Cell*. 11. 535-548.
- Lee, Y.R. and Liu, B.** (2004) Cytoskeletal motors in Arabidopsis: sixty-one kinesins and seventeen myosins. *Plant Physiology*. 136: 3877–3883
- Legg, J.P, and Fauquet, C.M.** (2004) Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology*. 56: 585-599.
- Legg, J.P. and Thresh, J.M.** (2003) Cassava virus diseases in Africa. Proceedings of the First International Conference on Plant Virology in Sub-Saharan Africa. 4-8 June, Ibadan, Nigeria.
- Legg, J.P., bB. Owor, bP. Sseruwagib and J. Ndunguru,** (2006) Cassava mosaic virus disease in East and Central Africa: Epidemiology and management of a regional pandemic. *Advanced Virus Research*. 67: 355-418.
- Legg, J. P.** (2008). African Cassava Mosaic Disease. *Encyclopedia of Virology*. 3: 30-36.
- Lellis, A. D., K. D. Kasschau, S. A. Whitham, and J. C. Carrington.** (2002) Loss-of susceptibility mutants of *Arabidopsis thaliana* reveal an essential role for eIF(iso)4E during potyvirus infection. *Current Biology*, 12:1046-1051.
- Leniewska, A., and Okoniewski, M.J.** (2011) rnaSeqMap: a Bioconductor package for RNA sequencing data exploration. *BMC Bioinformatic*. 12:200.
- Levy, A., Guenoune-Gelbart, D., and Epel, B.L.** (2007) β -1,3-Glucanases: Plasmodesmal gate keepers for intercellular communication. *Plant Signaling and Behavior*, 2, 404–407.

- Lewsey M., Surette M., Robertson F. C., Ziebell H., Choi S. H., Ryu K. H., Canto T., Palukaitis P., Payne T.** (2009). The role of the Cucumber mosaic virus 2b protein in viral movement and symptom induction. *Molecular Plant- Microbe Interactions*. 22: 642–654.
- Lewis, J. D., and Lazarowitz, S. G.** (2010). *Arabidopsis* Synaptotagmin AtSYTA Regulates Early Endosome Formation and Virus Movement Protein Cell-to-Cell Transport. *Proceedings of the National Academy of Sciences of the United States of America*. 107: 2491-2496.
- Life Technologies (2008)** : SOLiD Software
<http://www.lifetechnologies.com/global/en/home/technical-resources/software-downloads/solid-software.html>
- Li, Y.F., Gourierrec, J.L., Moez, T, Kim, Y.J, Guerineau, F., Zhou D.X.** (1999) Characterization and functional analysis of ArabidopsisTFIIA reveal that the evolutionarily unconserved region of the large subunit has a transcription activation domain. *Plant Molecular Biology*. 39:515–525.
- Li, Y.Q., Mareck, A., Faleri, C., Moscatelli, A., Liu, Q. and Cresti, M.** (2002) Detection and localization of pectin methylesterase isoforms in pollen tubes of *Nicotiana tabacum* L. *Planta*. 214: 734–740.
- Li H and Durbin R.** (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25: 1754-1760.
- Li, J, Brader ,G., and Palva, E.T.** (2004) The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *Plant Cell*. 16: 319-331.
- Liu, J.Z., Blancaflor, E.B., and Nelson, R.S.** (2005) The Tobacco mosaic virus 126-kilodalton protein, a constituent of the virus replication complex, alone or within the complex aligns with and traffics along microfilaments. *Plant Physiology*. 138: 1853–1865.
- Liu, X., Jin, D.Y., McManus, M.T., and Mourelatos, Z.** (2012) Precursor MicroRNA-Programmed Silencing Complex Assembly Pathways in Mammals. *Molecular Cell Press*. 46: 507–517.
- Liu,L., Li,Y., Li, S., He, Y,** (2012) Comparison of Next-Generation Sequencing Systems, *Journal of Biomedicine and Biotechnology*, vol. 2012, Article ID 251364
- Lokko Y, Anderson, J. V, Rudd S, Raji A, Horvath, D., Mikel, M. A, Kim R, Liu L, Hernandez A, Dixon A G, Ingelbrecht I L.** (2007) Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. *Plant Cell Reporter*. 26:1605–1618.
- Lopez, C., Soto-Suarez, M., Restrepo, S., Piegu,B., Cooke, R., Delseny, M., Tohme, J., and Verdier, V.** (2005) Global transcriptome analysis of cassava responses to *Xanthomonas axonopodis* pv.*manihotis* infection using a cassava cDNA microarray. *Plant Molecular Biology*. 57:393–410.
- Lorenzo, O., and Solano, R.** (2005). Molecular players regulating the jasmonate signalling network. *Current Opinion in Plant Biology*. 8: 532–540.

- Love, A. J., Yun, B. W., Laval, V., Loake, G. J., and Milner, J. J.** (2005) Cauliflower mosaic virus, a compatible pathogen of *Arabidopsis*, engages three distinct defence signalling pathways and activates rapid systemic generation of reactive oxygen species. *Plant Physiology*. 139: 935-948.
- Lucas, W.J.** (2006). Plant viral movement proteins, Agents for cell-to-cell trafficking of viral genomes. *Virology*. 344: 169-184.
- Marathe, R., Guan, Z., Anandalakshmi, R., Zhao, H., and Dinesh-Kumar, S.P.** (2004). Study of *Arabidopsis thaliana* resistome in response to cucumber mosaic virus infection using whole genome microarray. *Plant Molecular Biology*. 55: 501–520.
- Mardis, E.R.** (2008) The e impact of next-generation sequencing technology on genetics, *Trends in Genetics*. 24(3): 133-141.
- Mathur, J. and Hülkamp, M.** (2002) Microtubules and microfilaments in cell morphogenesis in higher plants. *Current Biology*. 12: R669–R676.
- Martin, G. B., Bogdanove A.J., and Sessa,G.** (2003) Understanding the function of plant disease resistance proteins. *Annual Reviews in Plant Biology*. 54: 23–61.
- Maule, A.J.** (1991). Virus movement in infected plants. *Critical Reviews in Plant Sciences*. 9: 457–473.
- McLean, B.G., Zupan, J., and Zambryski, P.** (1995) Tobacco mosaic virus movement protein associates with the cytoskeleton in tobacco plants. *Plant Cell*. 7: 2101–2114.
- Melcher, U.** (2000). The '30K' superfamily of viral movement proteins. *Journal of General Virology*, 81, 257-266.
- Menke, F.L., van Pelt, J.A., Pieterse, C.M., and Klessig, D.F.** (2004) Silencing of the mitogen-activated protein kinase MPK6 compromises disease resistance in *Arabidopsis*. *Plant Cell*. 16: 897-907.
- Méndez-Lozano J., Torres-Pacheco I., Fauquet C.M., Rivera-Bustamante, R.F.** (2003). Interactions between geminiviruses in a naturally occurring mixture: Pepper huasteco virus and Pepper golden mosaic virus. *Phytopathology*. 93: 270–277.
- Mészáros, T., Helfer, A., Hatzimasoura, E., Magyar, Z., Serazetdinova, L., Rios, G., Bardóczy, V., Teige, M., Koncz, C., Peck, S., and Bögre, L.** (2006) The *Arabidopsis* MAP kinase kinase MKK1 participates in defence responses to the bacterial elicitor flagellin. *Plant Journal* . 48: 485-498.
- Miguel, A.A., Escaler, M., Wang, D and Maule, A.J.** (1996) Induction of HSP70 and polyubiquitin expression associated with plant virus replication. *Proceedings of the National Academy of Science USA*. 93: 15289-15293.
- 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 signaling in *Arabidopsis*. . *Proceedings of the National Academy of Science. U.S.A.* 104: 19613–19618.

- Moreno, I., Gruissem, W. and Vanderschuren, H.,** (2011) Reference genes for reliable potyvirus quantitation in cassava and analysis of Cassava brown streak virus load in host varieties. *Journal of Virological Methods*. 177: 49-54.
- Morilla, G., Castillo, A. G., Preiß, W., Jeske, H. & Bejarano, E. R.** (2006) A versatile trans replication-based system to identify cellular proteins involved in geminivirus replication. *Journal of Virology*, 80: 3624–3633.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B.** (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*. 5: 621–8.
- Moy, P., D. Qutob, B.P. Chapman, I. Atkinson and M. Gijzen.** (2004) Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. *Molecular Plant-microbe Interactions*. 17: 1051–1062.
- Murashige, T. and F. Skoog.** (1962) A revised medium for rapid growth and bioassays with tobacco cultures. *Plant Physiology*. 15: 473-497.
- Mysore, K.S., Tuori, R.P. and Martin, G.B.** (2001) Arabidopsis genome sequence as a tool for functional genomics in tomato. *Genome Biology* 2(1):reviews1003.1-1003.4.
- Naidoo, S., Denby, K.J., and Berger, D.K.** (2005). Microarray experiments: considerations for experimental design. *South Africa Journal of Science*. 101:347-354.
- Nairn, C.J., Lewandowski, D.J. and Burns, J.K.** (1998). Genetics and expression of two pectinesterase genes in Valencia orange. *Plant Physiology*. 102: 226–235.
- Nash T. E., Dallas, M.B., Reyes, M.I., Buhrman, G.K., Ascencio-Ibañez, J.T., and Hanley-Bowdoin L** (2011). Functional analysis of a novel motif conserved across geminivirus Rep proteins. *Journal of Virology*. 85:1182–1192.
- Nassar, N.M.A.** (2002). Cassava, *Manihot esculenta* Crantz, genetic resources: origin of the crop, its evolution and relationships with wild relatives. *Genetic and Molecular Research*. 1: 298-305.
- Nelson, R.S. & Citovsky, V.** (2005) Plant viruses: invaders of cells and pirates of cellular pathways. *Plant Physiology*. 138: 1809-1814.
- Nemchinov, L.G, Shabala, L., Shabala, S.** (2008) Calcium efflux as a component of the hypersensitive response of *Nicotiana benthamiana* to *Pseudomonas syringae*. *Plant Cell Physiology*, 49: 40-46.
- Nimchuk, Z., Eulgem, T., Holt, B. F., III, and Dangl, J. L.** (2003) Recognition and response in the plant immune system. *Annual Reviews in Genetics*. 37: 579-609.
- Nweke, F. I., Spencer, D. D. C. and Lynam, J. K.** (2002) The Cassava Transformation, Michigan State University Press, East Lansing, pp. 3, 116, 124, 178 – 182.
- Nelson, R.S. & Citovsky, V.** (2005) Plant viruses: invaders of cells and pirates of cellular pathways. *Plant Physiology*. 138: 1809-1814.

- Nicot, N., Hausman, J.F., Hoffmann, L., and Evers, D.** (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany*. 56: 2907–2914.
- Novaes, E, Drost ,D.R., Farmerie, W.G, Pappas. G.J Jr, et al.** (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics*. 9: 312.
- O'Donnell, P.J., Schmelz, E.A., Moussatche, P., Lund, S.T., Jones, J.B. and Klee, H.J.** (2003) Susceptible to intolerance - a range of hormonal actions in a susceptible *Arabidopsis* pathogen response. *The Plant Journal*. 33: 245-257.
- Okinaka, Y., Yang, C.H., Perna, N.T., and Keen, N.T.** (2002) Microarray profiling of *Erwinia chrysanthemi* 3937 genes that are regulated during plant infection. *Molecular Plant-Microbe Interactions*. 15: 619-629.
- Omarov R, Sparks K, Smith L, Zindovic J, Scholthof HB.** (2006) Biological relevance of a stable biochemical interaction between the tombusvirus-encoded P19 and short interfering RNAs. *Journal of Virology*. 80: 3000–3008.
- Oparka, K.J.** (2004) Getting the message across: How do plant cells exchange macromolecular complexes? *Trends in Plant Sciences*. 9: 33–41.
- Overall, R.L., White, R.G., Blackman, L.M., and Radford, J.E.** (2000) Actin and myosin in plasmodesmata. In *Actin: A Dynamic Framework for Multiple Plant Cell Functions*, Staiger, C.J., Baluska, F., Volkmann, D., and Barlow, P.W., eds (Dordrecht: Kluwer Academic), pp. 1–19.
- Pallas, V. and Garcia, J.A.** (2011) How do plant viruses induce disease? Interactions and interference with host components. *Journal of General Virology*. 92: 2691-2705.
- Pallas, V., Genoves, A., Sánchez-Pina, M. A. and Navarro J. A.** (2011) Systemic movement of viruses via the plant phloem. In *Advances in Plant Virology*, p. 470. Edited by Caranta C., Aranda M. A., Tepfer M., Lopez-Moya J. J. . Norwich, UK:Caister Academic Press.
- Park, W., Li, J., Song, R., Messing, J., and Chen, X.** (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Current Biology*. 12: 1484–1495.
- Patil, B.L. and Dasgupta, I** (2006) Defective interfering DNAs of plant viruses. *Critical reviews in Plant Sciences*. 25: 47–64.
- Patil, B.L., and Fauquet, C.M.** (2009) Cassava mosaic geminiviruses: actual knowledge and perspectives. *Molecular Plant Pathology*. 10: 685–701.
- Pilartz, M. & Jeske, H.** (2003) Mapping of Abutilon mosaic geminivirus minichromosomes. *Journal of Virology*. 77 :10808–10818.
- Pitzschke, A., Schikora, A., and Hirt H.** (2009) MAPK cascade signalling networks in plant defence. *Current Opinion in Plant Biology*. 12: 421–426.

- Podolyan, A., White, J., Jordan, B., Winefield, C.** (2010). Identification of the lipoxygenase gene family from *Vitis vinifera* and biochemical characterisation of two 13-lipoxygenases expressed in grape berries of *Sauvignon Blanc*. *Functional Plant Biology*. 37: 767–784.
- Pompe-Novak, M., Gruden, K., Baebler, S., Krecic-Stres, H., Kovac, M., Jongsma, M. & Ravnikar, M.** (2006). Potato virus Y induced changes in the gene expression of potato (*Solanum tuberosum* L.). *Physiological and Molecular Plant Pathology*. 67: 237–247.
- Postnikova, O. and Nemchinov, L.** (2012). Comparative analysis of microarray data in *Arabidopsis* transcriptome during compatible interactions with plant viruses. *Virology Journal*. 9: 101
- Prochnik, S., Marri, P.R., Desany, B., Rabinowicz, P.D., Kodira, C., Mohiuddin, M., Rodriguez, F., Fauquet, C., Tohme, J., Harkins, T., Rokhsar, D.S., and Rounsley, S.** (2012) The Cassava Genome: Current Progress, Future Directions. *Tropical Plant Biology*. 5(1): 88-94.
- Prokhnevsky, A.I., Peremyslov, V.V., Napuli, A.J., Dolja, V.V.** (2002). Interaction between long-distance transport factor and Hsp70-related movement protein of Beet yellows virus. *Journal of Virology*. 76: 11003–11011.
- Qiu Q., Ma T., Hu Q., Liu B., Wu Y., Zhou H., Wang Q., Wang J., Liu J.** (2011). Genome-scale transcriptome analysis of the desert poplar, *Populus euphratica*. *Tree Physiology*. 31:452–461.
- Qu, F., Ren, T. & Morris, T. J.** (2003) The coat protein of Turnip crinkle virus suppresses posttranscriptional gene silencing at an early initiation step. *Journal of Virology*. 77: 511–522.
- Rabbi, I.Y., Kulembeka, H.P., Masumba, E., Marri, P.R., and Ferguson, M.** (2012). An EST-derived SNP and SSR genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics*. 125: 329-342.
- Radford, J.E., and White, R.G.** (1998). Localization of a myosin-like protein to plasmodesmata. *Plant Journal*. 14: 743–750.
- Raja, P., Sanville, B.C., Buchmann, R.C., Bisaro, D.M.** (2008) Viral genome methylation as an epigenetic defense against geminiviruses. *Journal of Virology*. 82: 8997–9007.
- Rainer, J., Sanchez-Cabo, F., Stocker, G., Sturn, A. and Trajanoski, Z.** (2006) CARMAweb: Comprehensive R- and bioconductor based web service for microarray data analysis. *Nucleic Acids Research*. 34: 498.
- Rensink, W.A., and Buell, C.** (2005) Microarray expression profiling for plant genomes, *Trends in Plant Science*. 10(12):603-609.
- Riedle-Bauer, M.** (1998). Activities of antioxidant enzymes in cucumber plants infected with cucumber mosaic virus. *Phyton*. 38: 149 – 157.
- Robatzek, S., and Somssich, I.E.** (2001). A new member of the *Arabidopsis* WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defense-related processes. *Plant Journal*. 28: 123–133.
- Roitsch, T., Balibrea, M. E., Hofmann, M., Proels, R., and Sinha, A. K.** (2003). Extracellular invertase: Key metabolic enzyme and PR protein. *Journal of Experimental Botany*. 54: 513-524.

- Rom,M.Y. Antignus, Y.; Gidoni D.; Pilowskyand M.; Cohen,S.** (1993). Accumulation of tomato yellow leaf curl virus DNA in tolerant and susceptible tomato lines. *Plant Disease*. 77: 253—257.
- Rushton, P.J., Somssich, I.E., Ringler, P., and Shen, Q.J.** (2010). WRKY transcription factors. *Trends in Plant Science*. 15: 247-258.
- Sacristán, S. and García-Arenal, F.** (2008). The evolution of virulence and pathogenicity in plant pathogen populations. *Molecular Plant Pathology*. 9(3): 369-384.
- Sakurai, T, Plata, G., Rodriguez-Zapata, F., Seki,M., Salcedo, A., Toyoda, A, Ishiwata, A, Tohme ,J, Sakaki, Y., Shinozaki, K., Ishitani, M.** (2007). Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biology*. 7:66.
- Sánchez-Durán, M.A., Dallas, M.B., Ascencio-Ibañez, J.T., Reyes, M.I., Arroyo-Mateos, M., Ruiz-Albert, J., Hanley-Bowdoin, L., and Bejarano, E.R.** (2011). Interaction between geminivirus replication protein and the SUMO-conjugating enzyme is required for viral infection. *Journal of Virology*. 85: 9789-9800.
- Sanderfoot , A. A., and Lazarowitz, S.G.** (1995). Cooperation in viral movement: The geminivirus BL1 movement protein interacts with BR1 and redirects it from the nucleus to the cell periphery. *Plant Cell*. 7. 1185–1194.
- Sanz, L.C., Perez, A.G., and Olias, J.M.** (1992). Purification and catalytic properties of chickpea lipoxigenase. *Phytochemistry*. 31 : 2967-2972
- Schena, M., Shalon, D., Davis, R. W., and Brown, P. O.** (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 270:467-470.
- Schena, M, Heller, R.A, Theriault, T.P, Konard, K., Lachenmeier, E., Davis, R.W.** (1998). Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnology*. 16: 301–306
- Schenk, P., Kazan, K., Wilson,I., Anderson,J., Richmond,T., Somerville,S. and Manners, J.** (2000) Co-ordinated plant defense responses in *Arabidopsis* revealed in microarray analysis. *Proceedings of the National Academy of Science of the USA*. 97: 11655-11660
- Schlaich, N.L.** (2011) *Arabidopsis thaliana*- the model plant to study host-pathogen interactions. *Current Drug Targets*. 12(7): 955-66.
- Senthil, G., Liu, H., Puram, V. G., Clark, A., Stromberg, A. & Goodin,M. M.** (2005). Specific and common changes in *Nicotiana benthamiana* gene expression in response to infection by enveloped viruses. *Journal of General Virology*. 86: 2615–2625.
- Settlage, S.B., See, R.G. and Hanley-Bowdoin, L.** (2005). Geminivirus C3 Protein: Replication enhancement and protein nteractions. *Journal of Virology*. 79(15): 9885-9895.

- Severin A. J., J. L. Woody, Y. Bolon, B. Joseph, B. W. Diers, A. D. Farmer, G. J. Muehlbauer. (2010). RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. *BMC Plant Biology*. 10: 160
- Shahid.M, Briddon.R.W, Zafar.Y and Stanley.J (2003) Geminivirus disease complexes: an emerging threat, *TRENDS in Plant Science*. 8(3):128-133.
- Shalitin, D. and S. Wolf (2000) Cucumber mosaic virus infection affects sugar transport in melon plants. *Plant Physiology*. 123: 597-604.
- Shalitin, D., Y Wang, Y., Omid, Gal-On, A., and S. Wolf (2002).Cucumber mosaic virus movement protein affects sugar metabolism and transport in tobacco and melon plants. *Plant, Cell & Environment*. 25: 989-998.
- Sharma C. M., Hoffmann S., Darfeuille F., Reignier J., Findeiss S., Sittka A., Chabas S., Reiche K., Hackermuller J., Reinhardt R., Stadler P. F., Vogel J. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*. 464: 250–255.
- Shendure, J., and Ji ,H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*. 26: 1135–1145.
- Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. & Church, G.M. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* (New York, N.Y.), vol. 309, no. 5741, pp. 1728-1732.
- Shimizu, T., Satoh, K., Kikuchi, S. and Omura, T. (2007) The repression of cell wall- and plastid-related genes and the induction of defense-related genes in rice plants infected with *Rice dwarf virus*. *Molecular Plant-Microbe Interactions*. 20: 247–254.
- Smith, L.G., and Oppenheimer, D.G. (2005) Spatial control of cell expansion by the plant cytoskeleton. *Annual Review of Cell and Developmental Biology*. 21: 271–295.
- Spellman PT, Miller M, Stewart J, Troup C, Sarkans U, et al. (2002) Design and implementation of microarray gene expression markup language (MAGE-ML).*Genome Biology*. 3:46
- Springer, N.M and Kaeppler, S.M. (2005). Evolutionary divergence of monot and dicot methyl-CpG-binding domain proteins. *Plant Physiology*, 138(1), 92-104.
- Stanley, J., Bisaro, D.M., Briddon, R.W., Brown, J.K., Fauquet, C.M., Harrison, B.D., Rybicki, E.P. and Stenger, D.C. (2005) Geminiviridae.In: *Virus Taxonomy*, VIIIth Report of the IC(C.M. Fauquet, M.A. Mayo, J. Maniloff, U. Desselberger, and L.A. Ball, eds), 301-327Elsevier/Academic Press, London.
- Su, S., Liu, Z., Chen, C., Zhang, Y., Wang, X., Zhu L, Miao, L., Wang, X.C., and Yuan, M. (2010) *Cucumber Mosaic Virus* movement protein severs actin filaments to increase the plasmodesmal size exclusion limit in tobacco. *Plant Cell*. 22(4): 1373-87.
- Suarez, M.C, Bernal, A., Gutierrez, J, Tohme, J. and Fregene, M. (2000) Developing expressed sequence tags (ESTs) from polymorphic transcript-derived fragments (TDFs) in cassava (*Manihot esculenta* Crantz). *Genome*. 43: 62-67.

- Subbaiah, C.C., Palaniappan, A., Duncan, K., Rhoads, D.M., Huber, S.C., Sachs, M.M.**(2006) Mitochondrial localization and putative signaling function of sucrose synthase in maize. *Journal of Biological Chemistry*. 281: 15625-15635
- Swarbrick P.J., Schulze-Lefert P., Scholes J.D.** (2006). Metabolic consequences of susceptibility and resistance in barley leaves challenged with powdery mildew. *Plant, Cell and Environment*. 29: 1061–1076.
- Swidzinsk, J.A., Sweetlove, L.J., and Leaver, C.J.** (2002) A custom microarray analysis of gene expression during programmed cell death in *Arabidopsis thaliana*. *Plant Journal*. 30(4) : 431-446
- Takai, Y., Sasaki, T., and Matozaki, T.** (2001) Small GTP-binding proteins. *Physiological Reviews*. 81: 153–208.
- Taliansky, M., Torrance, L., and Kalinina, N.O.** (2008) Role of plant virus movement proteins. *Methods in Molecular Biology* . 45: 33-54.
- Tao Y., Xie Z., Chen W., Glazebrook J., Chang H.S., Han B., Zhu T., Zou G., Katagiri F.**(2003). Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell*. 15: 317–330.
- Tao, L-Z, Alice Y. Cheung, A.Y and Hen-ming Wu, H-M** (2002).Plant Rac-Like GTPases Are Activated by Auxin and Mediate Auxin-Responsive Gene Expression. *Plant Cell*. 14(11): 2745–2760.
- Tecsi, L.I., Maule, A.J., Smith, A.M., and Leegood, R.C.** (1996) A spatial analysis of physiological changes associated with infection of cotyledons of marrow plants with cucumber mosaic virus. *Plant Physiology*. 111: 975–985.
- Thara, V.K, Tang X, Gu Y.Q, Martin, G.B, Zhou, J.M.** (1999) *Pseudomonas syringae* pv tomato induces the expression otomato EREBP-like genes Pti4 and Pti5 independent oethylene, salicylate and jasmonate. *Plant Journal*. 20: 475–483.
- Thomma, B., K. F. M. Tierens, I. Penninckx, B. Mauch-Mani,W. F. Broekaert, and B. P. A. Cammue.** (2001) Different microorganisms differentially induce *Arabidopsis* disease response pathways. *Plant Physiology and Biochemistry*. 39: 673–680.
- The Tomato Genome Consortium** (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*. 485 (7400): 635–641.
- Truman, W, Bennett, M.H, Kubigsteltig, I, Turnbull, C., Grant, M.** (2007) *Arabidopsis* systemic immunity uses conserved defense signaling pathways and is mediated by jasmonates. *Proceedings of the National Academy of Science USA*. 104: 1075–1080.
- Truman, W.M, Bennett, M.H, Turnbull, C.G.N, Grant, M,R.** (2010) *Arabidopsis* auxin mutants are compromised in systemic acquired resistance and exhibit aberrant accumulation of various indolic compounds. *Plant Physiology*. 152: 1562–1573.
- Trinks, D., Rajeswaran, R., Shivaprasad, P. V., Akbergenov, R., Oakeley, E. J., Veluthambi, K., Hohn, T., and Pooggin, M. M.** (2005). Suppression of RNA silencing by a geminivirus nuclear protein, AC2, correlates with transactivation of host genes. *Journal of Virology*. 79 : 2517-2527.

- Truernit, E., Schmid, J., Eppe, P., Illig, J., and Sauer, N.** (1996) The sink-specific and stress-regulated *Arabidopsis* STP4 gene: Enhanced expression of a gene encoding a monosaccharide transporter by wounding, elicitors, and pathogen challenge. *Plant Cell*. 8: 2169-2182.
- Tsuda K., Katagiri F.** (2010). Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current Opinion in Plant Biology*. 13: 459–465.
- Tzfira, T., Rhee, Y., Chen, M.H., Kunik, T., and Citovsky, V.** (2000) Nucleic acid transport in plant-microbe interactions: The molecules that walk through the walls. *Annual Reviews of Microbiology*. 54:187-219.
- Uemura, T, Sato, M.H, Takeyasu K.** (2005) The longin domain regulates subcellular targeting of VAMP7 in *Arabidopsis thaliana*. *FEBS Letters*. 579:2842–2846.
- Uemura, T, Ueda T, Ohniwa, R.L, Nakano, A, Takeyasu, K, Sato M.H.** (2004) Systematic analysis of SNARE molecules in *Arabidopsis*: dissection of the post-Golgi network in plant cells. *Cell Structure and Function*. 29:49– 65.
- Utsumi, Y., Tanaka, M., Morosawa, T., Kurotani, A., Yoshida, T., Mochida, K., Matsui, A., Umemura, Y., Ishitani, M., Shinozaki, K., Sakurai, T. and Seki, M.** (2012) Transcriptome analysis using a high-density oligo microarray under drought stress in various genotypes of cassava, an important tropical crop. *DNA Research*. 19(4): 335-345.
- Van Der Biezen, E.; and Jones, J. D. G.** (1998) Plant disease-resistance proteins and the gene-for-gene concept. *Trends in Biochemical Sciences*. 23 (12): 454–451.
- van der Hoorn, R.A.L. and Kamoun, S.** (2008) From guard to decoy: a new model for perception of plant pathogen effectors. *Plant Cell*. 20: 2009-2017.
- Vanitharani, R, Chellappan, P., and Fauquet, C.M.** (2005) Geminiviruses and RNA silencing. *Trends in Plant Science*. 10: 144–151.
- Vanitharani, R., Chellappan, P., Pita, J.S., and Fauquet, C.M.** (2004) Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and posttranscriptional gene silencing suppression. *Journal of Virology*. 78: 9487–9498.
- van Wezel, R., Dong, X., Liu, H., Tien, P., Stanley, J. & Hong, Y.** (2002) Mutation of three cysteine residues in tomato yellow leaf curl virus-China C2 protein causes dysfunction in pathogenesis and posttranscriptional gene-silencing suppression. *Molecular Plant-microbe interactions*. 15: 203–208.
- Vaucheret H.** (2006) Post-transcriptional small RNA pathways in plants: Mechanisms and regulations. *Genes and Development*. 20: 759.
- Vazquez F** (2006) *Arabidopsis* endogenous small RNAs: highways and byways. *Trends in Plant Science*. 11: 460–468

- Ventelon-Debout M., Delalande F., Brizard J. P., Diemer H., Van Dorsselaer A., Brugidou C. (2004)** Proteome analysis of cultivar-specific deregulations of *Oryza sativa indica* and *O. sativa japonica* cellular suspensions undergoing rice yellow mottle virus infection. *Proteomics*, 4: 216–225.
- Vidavski, F., Czosnek, H., Gazit, S., Levy, D., Lapidot, M. (2008)** Pyramiding of genes conferring resistance to *Tomato yellow leaf curl virus* from different wild tomato species. *Plant Breeding*. 127: 625-631.
- Voinnet, O., Pinto, Y. M. & Baulcombe, D. C. (1999)** Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. *Proceeding for the National Academy of Science USA*. 96: 14147–14152.
- Volkman, D., Mori, T., Tirlapur, U.K., Konig, K., Fujiwara, T., Kendrick-Jones, J., and Baluska, F. (2003)** Unconventional myosins of the plant-specific class VIII: endocytosis, cytokinesis, plasmodesmata/pit-fields, and cell-to-cell coupling. *Cell Biology International*. 27:289–291.
- Wan, J.G., Dunning, F.M., and A.F. Bent. (2002)** Probing plant-pathogen interactions and downstream defense signaling using DNA microarrays. *Functional and Integrative Genomics*. 2: 259–273.
- Wang, W., Vinocur, B., Shoseyov, O., Altman, A. (2004)** Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science*. 9(5): 244-252.
- Wassenegger, M., Heimes, S., Riedel, L. and Sanger, H. L. (1994)** RNA-directed de novo methylation of genomic sequences in plants. *Cell*. 76: 567-576.
- Waterhouse, P.M., Wang, M.-B., and Lough, T. (2001).** Gene silencing as an adaptive defence against viruses. *Nature*. 411: 834–842.
- Whitham, S., Dinesh-Kumar, S.P., Choi, D., Hehl, R., Corr, C., and Baker, B. (1994).** The product of the tobacco mosaic virus resistance gene N: Similarity to Toll and the interleukin-1 receptor. *Cell*. 78: 1101–1115
- Whitham, S. A., Quan, S., Chang, H-R., Cooper B., Estes, B., Zhu, T., Wang, X. and Hou, Y-M. (2003)** Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. *Plant Journal*. 33: 271-283.
- Whitman, S.A., Yang, C and Goodin, M.M. (2006)**Global impact : Elucidating plant responses to viral infection. *Molecular Plant-Microbe Interactions*. 19(11):1207-1215
- Woollard, A. A. D. and Moore, I. (2008).** The functions of Rab GTPases in plant membrane traffic. *Current Opinion in Plant Biology*. 11: 610-619.
- Xie, A.D., Zilberman, D., Jacobsen, S.E., and Carrington J.C. (2004).** Genetic and functional diversification of small RNA path-ways in plants. *PLoS Biology*. 2: E104.
- Xu, X.; Pan, S.; Cheng, S.; Zhang, B.; Mu, D.; Ni, P.; Zhang, G.; Yang, S. et al. (2011)** Genome sequence and analysis of the tuber crop potato. *Nature*. 475 (7355): 189–195.

- Yang, P., Wang, Z., Fan, B., Chen, C., and Chen, Z.** (1999). A pathogen- and salicylic acid induced WRKY DNA-binding activity recognizes the elicitor response element of the tobacco class I chitinase gene promoter. *Plant Journal*. 18: 141–149.
- Yang K.-Y., Liu Y., Zhang S.** (2001) Activation of a mitogen-activated protein kinase pathway is involved in disease resistance in tobacco. *Proceedings of the National Academy of Science USA*. 98: 741–746.
- Yang, S-J., Carter, S.A., Cole, A.B., Cheng, N-H., Nelson, R.S.** (2004) A natural variant of a host RNA-dependent RNA polymerase is associated with increased susceptibility to viruses by *Nicotiana benthamiana*. *Proceeding of the National Academy of Science USA*. 101(16): 6297–6302.
- Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R., and Nishikura, K.** (2006) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nature Structural & Molecular Biology*. 13: 13–21.
- Yang, C., Guo, R., Jie, F., Nettleton, D., Peng, J., Carr, T., Yeakley, J.M., Fan, J.B. and Whitham, S.A.** (2007). Spatial analysis of *Arabidopsis thaliana* gene expression in response to Turnip mosaic virus infection. *Molecular Plant-Microbe Interactions*. 20(4): 359-370
- Yang, J., An, D., Zhang, P.** (2011) Expression profiling of cassava storage roots reveals an active process of glycolysis/gluconeogenesis. *Journal of Integrated Plant Biology*, 53, 193-211
- Yang, Y. H., and Speed, T. P.** (2003) Design and analysis of comparative microarray experiments. In T. P. Speed (ed.), *Statistical Analysis of Gene Expression Microarray Data*, Chapman & Hall/CRC Press, pages 35-91.
- Ye, J., Qu, J., Zhang, J. F., Geng, Y. F. & Fang, R. X.** (2009). A critical domain of the Cucumber mosaic virus 2b protein for RNA silencing suppressor activity. *FEBS Letters*. 583: 101–106
- Zeidler D., Zähringer U., Gerber I., Dubery I., Hertung T., Bors W., et al.** (2004). Innate immunity in *Arabidopsis thaliana* lipopolysaccharides activate nitric oxide synthase NOS and induce defense genes. *Proceeding of the National Academy of Science USA*. 101: 15811–15816.
- Zhang, Y., Goritschnig, S., Dong, X., and Li, X.** (2003). A gain-of-function mutation in a plant disease resistance gene leads to constitutive activation of downstream signal transduction pathways in suppressor of npr1-1, constitutive 1. *Plant Cell*. 15: 2636–2646.
- Zhang, X., Yuan, Y. R., Pei, Y., Lin, S. S., Tuschl, T., Patel, D. J. & Chua, N. H.** (2006). Cucumber mosaic virus-encoded 2b suppressor inhibits *Arabidopsis* Argonaute1 cleavage activity to counter plant defense. *Genes and Development*. 20: 3255–3268.
- Zhou, J., Tang, X., Martin, G.B.** (1997) The Pto kinase confer-ring resistance to tomato bacterial speck disease interacts with proteins that bind a cis-element of pathogenesis-related genes. *EMBO Journal*. 16: 3207–3218

Zhu, J.K. (2008). Reconstituting plant miRNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 105 (29): 9851-9852.