Transcriptome Profiling in Susceptible Model and Natural Host Systems in Response to *South African cassava mosaic virus*

Erica Pierce

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of

Doctor of Philosophy

Johannesburg, 2013

DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

_____

Erica Pierce

4th day of October 2013

ACKNOWLEDGEMENTS

I would like to extend my gratitude to my supervisor Professor Chrissie Rey, for her advice, guidance and patience throughout the project.

A special thank-you to Nicky Olivier and Jeanne Korsman for their microarray assistance and technical advice, and to the members of the Cassava Biotechnology Lab, past and present, for the good memories shared.

I would also like to thank the National Research Foundation, The National Bioinformatics Network, and the University of the Witwatersrand Postgraduate Merit Award for the financial support.

To all my friends and family, thank-you for the constant support and encouragement throughout the course of my studies.

RESEARCH OUTPUTS

Publications

Pierce, E.J., and Rey, M.E.C. (2013) Assessing global transcriptomic changes in response to *South African cassava mosaic virus* [ZA-99] infection in susceptible *Arabidopsis thaliana.* PLoS ONE 8(6): e67534

Oral Presentations

Rey, M.E.C., **Pierce, E.J**., and Van Schalk, F. (2011) Transcriptome profiling of *South African cassava mosaic virus* – responsive genes in *Arabidopis* and *Nicotiana benthamiana*. Abstracts of the American Society for Virology 30th Annual Meeting, Minnesota, US, Pg 79.

Chetty, C., Moralo, M., Mundembe, R., **Pierce, E.J**., Rogans, S., Van Schalk, F., Vanderschuren, H., and Rey, M.E.C. (2011) Biotechnology solutions for engineering virus resistant cassava. Agricultural Biotechnology for Economic Development: moving towards a Bioeconomy. 6-9 September, Sandton Convention Centre, Johannesburg, South Africa.

Rey, M.E.C., **Pierce, E.J**., Berger, D., and Olivier, N. (2010) Assessing global transcriptome changes in response to *South African cassava mosaic virus* in susceptible *Arabidopsis thaliana*. EMBO workshop: Genomic approaches to Interactions between plant viruses, their hosts and their vectors. 12-16 June, 2010, Fenestrelle, Italy.

Pierce, E. J. (2009) An Introduction to Gene Profiling and Microarray Analysis. Wits University and University of Arizona joint symposium.

Pierce, E.J., Vanderschuren, H., and Rey M.E.C. (2008). Agroinfection optimization – towards improvement of begomovirus infeciton efficiencies in cassava. BIO 08 conference, Grahamstown, 21 -25th January.

Rey M.E.C., **Pierce, E.J**., and Van Schalk, F. (2007) Host responses to infection by South African cassava mosaic virus. 5th international Geminivirus Symposium and 3rd International ssDNA Comparative Virology Workshop, Ouro Preto, Brazil, 20 -26 May.

Pierce, E.J. (2006) Gene expression studies in *Arabidopsis* in response to *South African cassava mosaic virus* infection utilizing microarrays. Bioinformatics and Functional Genomics Workshop held at Wits University

Poster Presentation

Pierce, E.J., and Rey, M.E.C. (2005) Gene expression studies in susceptible *Arabidopsis* upon SACMV infection. Joint Congress of Southern African Society for Plant Pathology. Abstract submitted, pg 108.

ABSTRACT

Geminiviruses causes diseases to many staple food and cash crops of great economic importance worldwide. Currently eight species of Begomoviruses belonging to the Geminivirus family exist, of which *South African cassava mosaic virus* SACMV-[ZA:99] is a member, and is known to cause cassava mosaic disease (CMD). Cassava (*Manihot esculenta*, Crantz) is considered to be an important food crop consumed in many tropical, sub-tropical and African countries, and is increasingly becoming well-known for its ethanol production on a global a scale. Various strategies to control CMD are currently being implemented, one of which is to elucidate mechanisms involved in host-virus interactions with the aim of identifying defence-related genes involved in the disease process. Many defence genes within the plant kingdom are evolutionary conserved, potentially providing methods of control not only to CMD but to other diseases as well. The research outlined in this thesis aimed to identify networks and pathways involved in disease susceptibility between the model plant host system, *Arabidopsis thaliana* and cassava T200 upon SACMV-[ZA:99] infection. Conclusions were also drawn from within host comparisons between susceptible cassava T200 and resistant cassava TME3 cultivars in order to explore if similarities, differences or common patterns of expression existed between genes governing resistance and susceptibility.

Before transcriptomic profiling studies were carried out, it was important to improve *South African cassava mosaic virus* (SACMV-[ZA:99]) and *African cassava mosaic virus* (ACMV-[NG:Ogo:90]) infection efficiencies in recalcitrant crop systems such as cassava. Susceptible cassava cultivars T200, TMS60444, and SM14334 were tested for these purposes following infection with three different *Agrobacterium* strains (C58C1; AGL1; LBA4404). Results demonstrated that an overall increase in infection efficiency was achieved for each genotype and virus tested, although with varying infectivity levels, suggesting that although an improved method was established, basal levels of susceptibility differed between genotypes and therefore it was not possible to achieve 100% infection efficiencies for agroinfection methods.

A 4 x 44k microarray whole genome study was then conducted to identify susceptible host genes involved in the interaction between the model plant system *Arabidopsis thaliana* and SACMV-[ZA:99]. An infectivity assay was carried out across

three time points (14, 24, and 36 dpi), confirming that disease symptoms and virus infectivity levels correlated with an increase in differentially expressed transcripts across time points, with SACMV-[ZA:99] predominantly causing host-gene suppression. Many complex genes and pathways were disrupted and were shown to be involved in categories pertaining to stress and defence responses, phytohormone signalling pathways, cellular transport, metabolism and cell-cycle regulation strongly suggesting an attempt made by SACMV-[ZA:99] to affect homeostasis and antagonize host defence responses. This was the first geminivirus study identifying differentially expressed transcripts across 3 time points.

Next generation sequencing (NGS) using the ABI Solid platform was then carried out on SACMV-[ZA:99] – infected susceptible cassava T200 cultivar at 3 time points (12, 32, and 67 dpi), comparing infection responses to mock-inoculated healthy controls. Similarly to the *Arabidopsis* microarray study, findings from this analysis also revealed a shift from up-regulated to down-regulated genes across time points, once again reflecting virus-specific suppression on host genes suggesting SACMV-[ZA:99] specific alterations were induced in the host, regardless of the host (*Arabidopsis* and cassava T200) or platform (microarray and NGS) used. Genes identified pertaining particularly to the susceptible cassava T200 - SACMV-[ZA:99] interaction such as the disease resistance protein families (TIR-NBS-LRR), RPP1, RPM1, and NHO1 were showing down-regulation demonstrating that SACMV-[ZA:99] pathogenicity proteins may be causing this suppression leading to inactivation of basal immunity. Comparisons between tolerant cassava TME3 and susceptible T200 showed similarities and differences in responses between the cultivars. Many similarities such as cell wall precursor proteins and glutathione-S-transferases were up-regulated in both cultivars, which may be due to the host attempting to mount appropriate defences. Opposite patterns of expression was observed for genes in categories involved in transcription and phytohormone signalling such as WRKY's, NAC, JAZ, and ERF where suppression was evident in susceptible cassava T200, confirming the suppressive nature of SACMV-[ZA:99] to establish a replication-competent environment. Findings in this study contributed to the little that is known about geminivirus disease progression within a previously uncharacterised susceptible host such as cassava.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Literature Review

1.1 Cassava (*Manihot esculenta*, Crantz)

Cassava, (*Manihot esculenta* Crantz, Euphorbiaceae), also commonly known as Manioc, Tapioca, Brazilian arrowroot, and Yuca, is a major carbohydrate food source, producing enlarged tuberous starchy roots for over 500 million people in the tropics, sub-tropical Africa, Asia and Latin America (Alleman and Coertze, 1996; El-Sharkawy, 2004; Olsen, 2004) and has become a major source of carbohydrate in sub-Saharan Africa and the fourth most important tropical crop worldwide (Olsen, 2004). Cassava globally provides food for 800 million people, many of whom subsist on it (Dahniya, 1994; Burns *et al*., 2010).

Cassava is grown on more than 16 million hectares worldwide and thus produces a minimum annual root yield of 170 million tons. During the last 30 years, cassava production has increased by 75% (Anderson *et al.*, 2004). As a monocrop cassava can yield as much as 90 tons of fresh roots per hectare, under favourable experimental conditions (El-Sharkawy, 2004). It is, however, usually grown in poor soils and harsh climates and in association with other crops, such as maize, beans, or cowpeas. Cassava production has accelerated over the past decade, with an estimated global harvest of more than 280 million tonnes in 2012 (FAO, 2013). In Africa in particular, the growing rate of cassava has been equal to that of maize since 2000 (FAO, 2012). In addition, in South, South East and Eastern Asia, the rate has been shown to be almost three times that of rice (FAO, 2012). In addition to human consumption, cassava is also a desirable energy source for biofuel applications.

Cassava is generally grown by small-holding farmers as a subsistence crop. It is not season-bound and can be planted and harvested at any time of the year (Nestel, 1980; Burns *et al*., 2010) where fresh cassava roots are either eaten on the farm, processed for starch extractions, dried for flour production, roasted for human and/or animal feed, and marketed for consumption. These processed food products are commonly known as *farinha da mandioca* in Brazil, *gablek* in Indonesia, and *gari* or *foufou* in West Africa (El-Sharkawy, 2004). For human food consumption, the cassava root is prepared by boiling, baking, frying, as meal, flour as well as in beer, and in many countries such as West and Central Africa, the leaves are eaten as a vegetable (Figure 1). A wide range of sweet and

savoury foods such as crackers, tapioca pearls, noodles, and cheese breads are made from the starch extracted from cassava roots. When utilized as animal feed, the fresh roots provide a first-rate source of carbohydrate and the leaves may be used as a protein supplement for cattle. Concentrates of dried cassava are also used for poultry, pigs, and cattle as well (Alleman and Coertze, 1996). Cassava's role varies in different parts of the world. It is grown mainly by women and used mostly for food in sub-Saharan Africa. As a result, it is an important source of employment and income, since most of the processing of this crop into food is done on a small scale in rural areas (CIAT, 2001_a). In Asia and Latin America, cassava starch is used in industry in the manufacture of many chemical products such as citric acid, sorbitol, mannitol, monosodium glutamate, high fructose syrup, glucose, and alcohol. It may also be used in paper-making, food processing, as a lubricant in oil wells, adhesives and textiles (Alleman and Coertze, 1996).

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

Figure 1.1: A and D) Cassava tubers B) cooked and made into C) flour for human food consumption. Photos taken in Cali, Colombia (South America).

1.2 Cassava Mosaic Disease (CMD)

Cassava mosaic disease (CMD) was first identified in Tanzania in 1894, but it was not evident that a virus was responsible for the disease. It was not until Storey (1936) suggested that a virus may be the causal agent of CMD as it was shown to be transmissible. Mechanical transmission studies from cassava to *Nicotiana benthamiana* and transmission back to a Brazilian cassava cultivar validated these findings (Bock and Woods, 1983). Characteristic symptoms of CMD are stunted plants, with leaves that are distorted, misshapen, and displaying yellow mosaic patterns. Symptoms may vary from plant to plant, due to differences in virus species and strain, sensitivity of the host, plant age, environmental factors as well as mixed infections (Legg and Thresh, 2000; Hillocks and Thresh, 2001) (Figure 1.2).

Figure 1.2: A) Healthy cassava leaves and B) SACMV-[ZA:99] - infected cassava leaves showing characteristic yellow mosaic patterns, leaf reduction, and distortion as a result of CMD

1.3 Geminiviruses

Genome Organization of Geminiviruses

Geminiviruses cause major diseases of great economic importance. The family, *Geminiviridae* is divided into four genera, *Mastrevirus, Curtovirus, Topocuvirus* and *Begomovirus*, which is based on genome organization, host range and insect vector (Gafni and Epel, 2002; Harrison and Robinson, 2002). Mastreviruses infect monocotyledonous plants mainly (those belonging to the family, *Poaceae*) and are transmitted by leafhoppers, an example of which is *Maize streak virus* (MSV). The mastrevirus contains a monopartite circular ss DNA genome of 2.6 kb in size, and encodes four proteins, two on the viral DNA strand and two on its complementary strand (Harrison and Robinson, 2002). *Beet curly top virus* (BCTV) is a Curtovirus infecting dicotyledonous plants (BCTV), also transmitted by leaf hoppers. Like mastreviruses, curtovirus genomes consist of ssDNA, encoding three proteins in the viral strand and four proteins in the complementary strand which are approximately 3kb in size. The genome organization of topocuviruses of which *Tomato pseudo-curly top virus* is a species, resembles that of BCTV but contains only two genes in the viral strand (Harrison and Robinson, 2002).

Of particular interest to this study is the *Begomovirus* genus. Begomoviruses may contain monopartite (ssDNA-A of about 2.6kb) or bipartite (ssDNA-A and DNA-B each about 2.6-2.8kb in size) genomes (Gafni and Epel, 2002) and are transmitted by the whitefly, *Bemisia tabaci* (*B. tabaci*), in a persistent circulative manner. *B. tabaci* is considered to be a species complex occurring worldwide in tropical, subtropical and warm temperate regions (Harrison and Robinson, 2002). The DNA-A and DNA-B share a short “common region” of 200-400 nucleotides that is very similar, even identical between the two DNAs which includes a stem-loop structure (loop containing the nonanucleotide TAATATTAC) that is conserved in all geminivirus genomes. The origin of rolling circle DNA replication is the last A in the nonanucleotide (Harrison and Robinson, 2002) (Figure 1.3).

Bipartite-genome organization of begomoviruses encodes at least four proteins on the DNA-A: the viral strand contains the coat protein (CP or AV1) and the pre-CP (AV2), which is found only in Old World begomoviruses (like SACMV-[ZA:99]). The complementary strand contains four proteins; AC1, AC2, AC3 and AC4 from overlapping open reading frames (ORFs). AC1 is required for initiation of DNA replication and is

termed the replication-associated protein (Rep), AC2 (TrAP) activates transcription in both the DNA-A and DNA-B of the viral sense genes, and AC3 is the DNA replication enhancer (REn) (Gafni and Epel, 2002; Harrison and Robinson, 2002). The function of AC4 protein is unknown but it has been suggested to act as a silencing suppressor (Vanitharani *et al*., 2004). DNA-B encodes two proteins, namely BC1 and BV1 which are involved in intracellular, intercellular and systemic virus movement. BC1 is found on the complementary strand and mediates cell-to-cell movement of the virus. BV1 is the nuclear shuttle protein (NSP) which controls movement of viral DNA between the nucleus and cytoplasm (Gafni and Epel, 2002; Harrison and Robinson, 2002) (Figure 1.3).

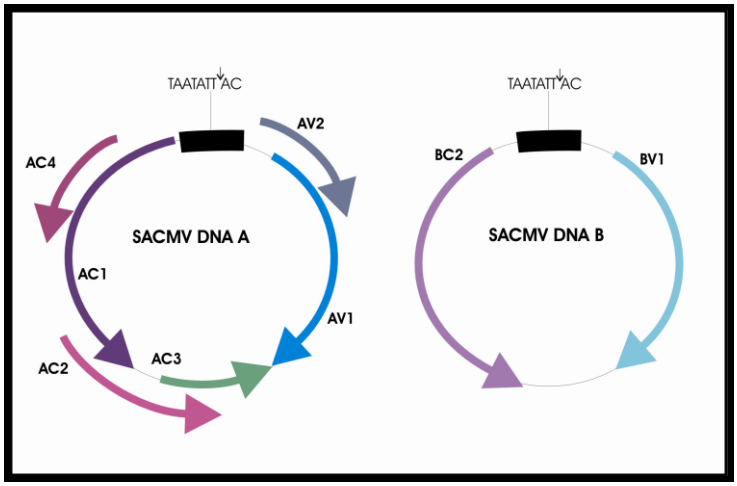


Figure 1.3 Bipartite genome organization of begomoviruses (particular species shown is SACMV-[ZA:99]). DNA-A containing 6 ORFs and DNA-B containing 2 ORFs (indicated in coloured arrows). Direction of transcription depicted by arrows (Diagram modified from Berrie *et al.*, 2001).

Geminivirus DNA Replication

Geminivirus DNA replication occurs in the nucleus of the host, requiring two stages for replication. These stages include: i) ssDNA conversion to dsDNA, and ii) rolling-circle replication (RCR). Viral circular genomic ssDNA (positive strand) is converted into supercoiled covalently dsDNA intermediates through a priming event activating the negative strand origin of DNA replication. These dsDNA intermediates are then amplified through a RCR mechanism (Gutierrez, 1999). The initiation site for RCR has been mapped to the intergenic region which contains an invariant nine nucleotide sequence (TAATATTAC) which is common among all geminiviruses (Figure 1.3). Viral proteins are expressed from the transcriptionally active dsDNA forms. Generally, genes encoded on

the v-sense have movement and structural functions whereas those encoded on the c-sense function in DNA replication, regulation of transcription, and most probably interfere with cellular processes required for replication (Gutierrez *et al.*, 2004). Viruses require both host cellular factors and machinery for replication, systemic spread as well as for suppression of antiviral defence mechanisms (Petty *et al.*, 2000).

In order to complete an infection cycle in a host, geminiviruses disrupt many host processes which involve transcriptional regulation, DNA replication, cell cycle control, and macromolecular trafficking in plants. Certain pathways such as those involved in plasmodesmata structure and function are altered and silencing of defence-related mechanisms occurs (Gutierrez, 2002). In addition, geminivirus pathogenicity proteins interfere with host proteins such as NAC transcription factor domains as well as with retinoblastoma-related (RBR) pathways (Gutierrez, 2002) (Figure 1.4).

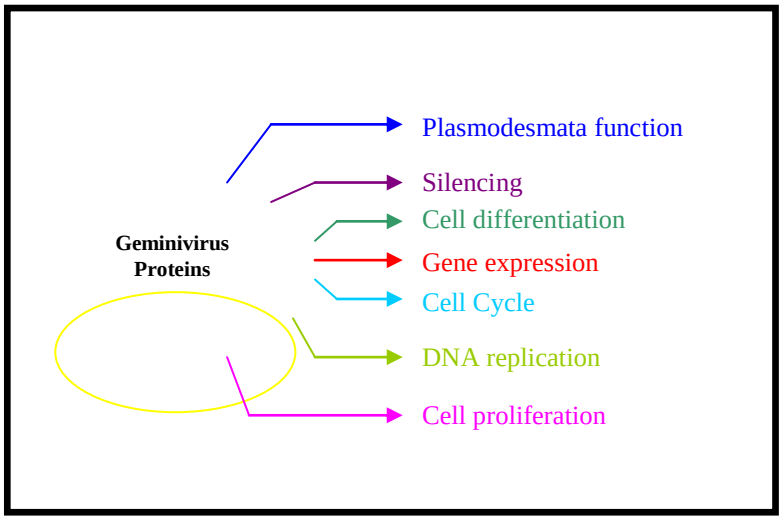


Figure 1.4 Disruption of host processes by geminiviruses (Gutierrez *et al.*,2004).

The Rep/AC1 protein functions by nicking the DNA once it has bound to the stem structure at the replication origin and initiates rolling-circle DNA replication. One problem encountered by geminiviruses is that they infect terminally differentiated cells at the resting state (G_0) which lack factors required for DNA replication. In order to overcome this problem geminiviruses induce host proliferating cell nuclear antigen (PCNA) which is a DNA polymerase accessory factor, normally found in the S-phase. Rep/AC1 is therefore required to induce PCNA. The process involves binding of Rep/AC1 to the viral replication

enhancer (AC3), which then binds to PCNA (Arguello-Astorga *et al.*, 2004). Rep/AC1 in turn, physically interacts with host encoded retinoblastoma-like tumour suppressor proteins (pRbs). In mammals, the pRB protein functions as the G_1 checkpoint regulator which prevents completion of G_1 and entry into the S-phase. Cyclin-dependent kinases, have phosphorylating activity which stops the cell-cycle inhibition activity of pRb, therefore allowing progression into S-phase. It is believed that geminivirus Rep proteins interact with and either inactivate or divert the pRb-like protein in infected cells, allowing S-phase-specific mRNA production, also providing a pool of factors and enzymes required for viral DNA replication (Carrington and Whitham, 1998; Gutierrez, 2000; Egelkrout *et al.*, 2002; Hanley-Bowdoin et al., 2004). Plant geminiviruses are therefore analogous to animal DNA tumour-inducing viruses (such as SV40) and adenoviruses. These viruses also encode proteins which affect cell cycling apparatus (Carrington and Whitham, 1998).

Post-transcriptional gene silencing (PTGS) is a natural defence mechanism plants have against viruses, involving a homology-dependent mRNA degradation process (Voinnet *et al*., 1999). Several plant viruses are known to counter antiviral RNA silencing with silencing suppressor proteins. These proteins interfere with initiation, maintenance, and systemic signalling in the RNA-silencing process. Potyvirus helper component proteinases (HC-Pro) have been shown to interfere with initiation and maintenance of silencing where short interfering RNAs (siRNAs) are produced. siRNAs form an active multi-complex RNA-induced silencing complex which is responsible for homologous RNA cleavage (Voinnet *et al*., 1999; Vanitharani *et al.*, 2004). *Cucumber mosaic virus* (CMV) has been shown to inhibit long-range PTGS-signalling activity, thereby preventing initiation of PTGS in newly formed tissues (Brigneti *et al*., 1998). It has been suggested that AC2 and AC4 of cassava – infecting geminiviruses behave differently. AC2 of *Indian cassava mosaic virus* (ICMV) was shown to be a strong suppressor of PTGS whereas AC2 of *Sri Lankan cassava mosaic virus* (SLCMV) showed weak suppression. ICMV AC4, on the other hand, was not able to suppress locally induced PTGS. In addition, AC4 of the Cameroon strain of *African cassava mosaic virus* (ACMV- [CM]) and SLCMV also revealed PTGS suppression as observed with their rapid symptom induction in hosts (Vanitharani *et al*., 2004).

1.4 *Agrobacterium*-Mediated Gene Transfer (Agroinoculation) of SACMV-[ZA:99] into *Arabidopsis*

One of the problems encountered by begomoviruses such SACMV-[ZA:99], is that they cannot be transmitted mechanically by rubbing healthy plants with virus-infected leaf sap. SACMV-[ZA:99] is transmitted by the whitefly, *Bemisia tabaci*, which delivers the virus directly into the plant's vascular system. One approach for virus introduction into a plant host without the use of the insect vector is by agroinoculation. This is a form of mechanical inoculation in which *Agrobacterium tumefaciens* cells are transformed with a full-length copy of the virus genome which is injected into the plant (Grimsley et al., 1986; Leiser *et al.,* 1992; Mayo *et al.,* 2000).

Agrobacterium tumefaciens is a soil-born pathogen well-known for its ability for *trans*-kingdom DNA transfer (Ward and Zambryski, 2001; Eckardt, 2004). Besides its natural ability to transform host plants with foreign DNA, its primary role is to cause crown gall disease on several important crop species such as grapevines, stone fruit, and nut trees (Eckardt, 2004). *Agrobacterium* has a tumour-inducing (Ti) plasmid which contains a specific DNA fragment called transferred DNA (T-DNA), which is delimited by a left and right T-DNA border (25bp direct repeats) (Tzfira and Citovsky, 2002). It is within this T-DNA region that foreign DNA is inserted which is then transferred to the plant cell and integrated into the plant genome (de la Riva, 1998). The process involves *Agrobacterium* VirD1 and VirD2 proteins which are induced by host signals where both T-DNA borders are nicked, resulting in a single stranded transfer strand (T-strand). A channel is then formed by the VirD4 and VirB4 proteins where the T-strand and several Vir proteins are exported into the cytoplasm of the host cell. Once integrated, a T-DNA transport complex (T-complex) is formed consisting of one VirD2 molecule covalently attached to the 5´end of the T-strand (coated with many VirE2 molecules), which is then translocated into the host nucleus with the aid of VirD2 and VirE2 proteins. This entire process requires both host factors and bacterial Vir proteins for successful transformation of plant cells (Tzfira and Citovsky, 2002). Two different strategies are utilized for foreign gene introduction into the T-DNA region. These involve an indirect method of cloning whereby the gene of interest in *cis* is cloned on the same plasmid as the *vir* genes (co-integrative vectors); or the gene of interest is cloned into the T-DNA region on a separate plasmid from the *vir* genes (*trans*-acting *vir* genes), also known as binary vectors (Gelvin, 2003).

Grimsley *et al.,* 1986, were the first to describe a transformation system for dicotyledonous plants whereby cloned *Cauliflower mosaic virus* (CaMV) DNA was transferred into plants via the Ti plasmid from *Agrobacterium tumefaciens.* This study showed that viral DNA was infectious once excised from the bacterial vector, as it was able to replicate and systemically infect the plant. Geminiviruses have been successfully introduced using agroinoculation methods into different host tissue types. These include: host leaf discs, mainly using model plant systems such as *Nicotiana* spp., germinating seeds, and whole plants, based on direct injection of transformed bacterial cultures into the plant vascular system (Pico *et al.*, 2001).

It has been documented in many studies that cloned geminivirus DNAs are infectious as tandemly repeated copies present on a Ti plasmid and delivered into plants via agroinoculation (Grimsley *et al.*, 1986; Hayes *et al.*, 1988; Stenger *et al.*, 1990). Geminiviral partial tandem repeats are constructed and inserted into the T-DNA region of a binary vector. This construct is then introduced into *A.tumefaciens* and released into the plant genome through mechanism described previously (Jacob *et al.*, 2003). The release of infectious unit-length circular replicative viral genomes into plants from the partial tandem repeats may be explained by two mechanisms. Either intramolecular homologous recombination within the tandem viral genome repeat occurs, which results from a single cross-over event (occurring at random locations), may lead to the release of circular dsDNA (Stenger *et al.*, 1991), or consequently, when the viral origin of replication is duplicated, the viral replication-associated protein replicates the full-length viral genome from the T-DNA region containing the partial tandem repeat portion (Stenger *et al.*, 1991). Agroinoculation has also been used to infect dicotyledonous plants with *African cassava mosaic virus* (ACMV) and *Tomato golden mosaic virus* (TGMV) as well as monocotyledonous plants with *Maize streak virus* (MSV), *Digitaria streak virus* and *Wheat dwarf virus*. Agroinoculation of *Beet curly top virus* (BCTV) has successfully been introduced into *N. benthamiana* and *D. stramonium* (Briddon *et al.*, 1989). Plants where agroinoculation with SACMV-[ZA:99] have been successful are from the species *Phaseolus vulgaris*, *Malva parviflora*, and cassava (*Manihot esculanta* Crantz), however infection efficiencies achieved by agroinoculation were low in cassava (Berrie *et al.*, 2001).

1.5 Defence- Responses and Signalling Pathways

Evolution of Plant-Pathogen Interactions

Close communication between a plant and a pathogen exist where plants are able to detect the presence of a pathogen and mount appropriate defence responses and pathogens in turn, focus on colonization and utilization of host resources. However, due to the high co-evolution of plant and pathogen species, a particular pathogen species may circumvent plant defences, or plants may adapt in such a way that successful pathogens may be blocked by adaptive responses (Staskawicz *et al.*, 1995). When a host is able to specifically recognize and resist pathogen infection through modification of a host receptor, selection favours evolution of the host, however, the pathogen may respond by mutating its avirulence gene (becoming virulent) causing disease, with the host requiring new resistance (*R*) gene specificities for defence. Gene-for-gene evolution between host and pathogen is therefore constantly evolving with a diverse array of avirulence (*avr*) genes found in different pathogen races as well as in *R* genes found in different host species (Staskawicz *et al.*, 1995).

Plants are able to defend themselves against viruses in a mechanism known as post-transcriptional gene silencing (PTGS). In turn, viruses are able to suppress PTGS, as observed in potato virus Y whereby a helper component proteinase (HC-Pro) blocks PTGS in tissues where silencing had already been established. Viruses such as *Cucumber mosaic virus* (Cmv2b) encode a 2b protein capable of preventing PTGS initiation at growing tips of the plant (Voinnet *et al*., 2004). Both HC-Pro and 2b proteins are therefore important for virulence and systemic spread throughout the plant. This mechanism compares to other pathogens that may accumulate avirulence (*avr*) genes to escape recognition from plants in gene – for - gene resistance mechanisms (Li *et al.*, 1999). In gene - for - gene resistance, plant *R* genes are able to detect specific pathogens through *Avr* protein recognition. Some pathogens are therefore able to eliminate these *Avr* proteins to avoid detection by the plant, unless required for pathogen fitness (Feys and Parker, 2000).

Resistance (Incompatibility)

Plants are continuously threatened by a vast number of potential pathogens, and in order to counter the attack by pathogens, they have developed intricate defence

mechanisms to recognize and respond to invading pathogens (Mysore and Ryu, 2004). A broad spectrum of plant defence molecules are activated upon pathogen detection. This early response is controlled by plant disease resistance (*R*) genes which encode proteins that either directly or indirectly recognize pathogen *Avr* proteins in a plant defence mechanism known as gene - for - gene disease resistance. A host plant that carries an *R* gene will give the pathogen that carries the corresponding *avr* gene an avirulent phenotype (Staskawicz *et al.*, 1995; Bent, 1996; Li *et al.*, 1999; Tang *et al.*, 1999; Feys and Parker, 2000; Dangl and Jones, 2001; Feys *et al.*, 2001). It has been hypothesized that *R* proteins "guard" plant proteins targeted by the pathogen *Avr* proteins which trigger the hypersensistive response (HR) and other defence related responses upon *R-avr* detection (Glazebrook, 2001). Induction of the plants defence responses leading to the HR is initiated by elicitor molecules produced by the pathogen and recognized by the plant, leading to the activation of a cascade of host genes (Staskawicz *et al.*, 1995). The HR is associated with plant cell death, inhibiting movement of the pathogen from the infection site, therefore rendering the infection unsuccessful (Li *et al.*, 1999; Feys and Parker, 2000).

There are many examples of *avr* and *R* gene specificities found in fungi, bacteria and viruses. *Cladosporium fulvum* is an extracellular growing fungal pathogen that contains *Avr4* and *Avr9* genes encoding pre-proteins that become small secreted peptides which are able to elicit *R* gene-dependent defence responses. The type III secretory system encoded by bacterial *Hrp* gene cluster (required for HR induction and pathogenesis) is required for movement of *Avr* proteins in plant cells and is responsible for the recognition event of the bacterial *avr* gene products and corresponding plant *R* gene products occurring inside plant cells (Li *et al.*, 1999). Generally, for most cloned bacterial *avr* genes, the *Avr* protein itself is the elicitor which when injected into the plant encodes hydrophilic proteins lacking signal sequences, not inducing a HR. Viral *Avr-R* recognition occurs inside plant cells as viruses enter through existing wounds and replicate intracellularly, encoding genes for replication, movement and encapsidation. The coat protein, RNA replicase, or movement protein are all avirulence determinants. This has been observed in *Tobacco mosaic virus* (TMV) in which the *Avr* protein for which a matching *R* gene, *N*, has been cloned. This *N* gene has been putatively identified as a cytoplasmic protein belonging to the nucleotide binding site, leucine rich repeat (NBS-LRR) family of resistance genes (Li *et al.*, 1999). *RCY1* has been cloned from *Arabidopsis* ecotype C24 which confers resistance to *Cucumber mosaic virus* (CMV) strain Y

(Takahashi *et al.*, 2002). This gene encodes a CC-NBS-LRR (CC, coiled coil domain) protein which is allelic to both *RPP8* and *HRT* proteins which confer resistance to *Peronospora parasitica* and *Turnip crinkle virus* (TCV) (McDowell *et al.*, 1998; Cooley *et al.*, 2000).

Basal Resistance and Tolerance

Another phenomenon in resistant and susceptible responses is that susceptible hosts possess what is called “basal resistance”, an innate defence response that limits but does not stop pathogen growth. Unlike the HR in gene - for - gene resistance occurring at the site of infection, this response results in cell death at the site of infection, which is preceded by a spreading of chlorosis as well as a secondary necrosis in surrounding, uninfected tissue (O´Donnell *et al.*, 2003). As with the resistance responses, susceptible responses also undergo changes in gene expression such as reactive oxygen (ROS) production and cell wall composition; the only difference in the susceptible response is that these changes are delayed. Effective resistance is therefore dependent on the speed in which induced defence reactions occurs in susceptible (compatible) and resistant (incompatible) interactions (O´Donnell *et al.*, 2003; Yang *et al*., 1997). This is all dependent on how quickly the pathogen can replicate and cause disease or how fast the plant may respond with corresponding levels of defences. Immediate recognition by a plant of an invasive pathogen (as in gene -for - gene resistance) results in a rapid defence mechanism (such as a HR). Susceptibility (disease) results when pathogen recognition is not rapid enough to mount the appropriate defence responses required to block pathogen infection. After a HR (resistance) or a successful infection (susceptibility) has been initiated, SAR is induced by the plant to prevent infection from spreading further or to prevent a secondary infection from occurring, resulting from a broad spectrum of pathogens (Dong, 1998). It has been hypothesized that the host may therefore suppress disease in an incompatible response in a regulated manner or alternatively, the host may not immediately recognize the pathogen and induce effective resistance responses, resulting in disease. Disease symptoms are not always associated with pathogen growth, thus in the absence of symptoms, pathogen growth may still be observed in a process referred to as tolerance (O´Donnell *et al.*, 2003). Studies on compatible interactions may explain why certain plants develop mild symptoms (tolerance), while others suffer severe symptoms sometimes leading to plant death. Most plant disease problems are caused by

systemic infection which explains the need to study compatible interactions in plants and pathogens (Scholthof, 2001).

Susceptibility (Compatibility)

Virulent pathogens are able to infect particular plant species or cultivars, reinforcing their highly specialized ability to cause disease. Successful disease formation depends on the pathogen's ability to actively suppress or avoid plant defence responses therefore inducing susceptibility in a host that would normally be either resistant or tolerant to the pathogen (Abramovitch and Martin, 2004). A susceptible reaction is the result of the plant not being able to detect the pathogen, or if detected, the pathogen is able to cope with plant defence mechanisms. Alternatively, the response may either not be activated or activated too late leading to spread of the pathogen (Venisse *et al.*, 2002). Examples of pathogenicity factors include small molecule suppressors from phytopathogenic fungi, phytopathogenic bacteria containing type III effectors and toxins, as well as post-transcriptional gene silencing suppression by plant viruses (Abramovitch and Martin, 2004). O'Donnell *et al.* 2003, suggested that a susceptible host plays a vital role in the interaction with a virulent pathogen due to its participation in basal resistance and disease symptom development. Two signalling intermediates, ethylene (ET) and salicylic acid (SA) have been found to play a role in both susceptible and resistant responses of hosts to pathogens (O´Donnell *et al.*, 2003). In order to establish SA influence on ET- dependent development of disease, increased levels of ET and SA in the *Arabidopsis -Xanthomonas campestris* pv. *campestris (Xcc)* compatible interaction proved that these two signals are essential in disease symptom production, clearly identifying the host's role in regulation of disease symptoms (O´Donnell *et al.*, 2003).

Disease in plants is caused from the expression of specific plant host and viral genes in a compatible interaction. Once viruses have mechanically disrupted the cell wall and plasma membrane, they remain within the symplast until they move into the plasmodesmata where they continue to infect cells. Host susceptibility results when the plant is not able to mount an appropriate defence response, the plant does not detect the pathogen, or activated defence responses are ineffective (Hammond-Kosack and Jones, 1996), the virus is thus able to successfully complete genome replication, local cell - to - cell movement and long distance movement (Carrington and Whitham, 1998). Host plant cells therefore do not die but retain large quantities of virus, supporting a complete

infection “cycle” and allowing progressive spread of the virus to adjacent tissue. The appearance of symptoms is the final result of virus infection in susceptible tissues (Maule *et al.*, 2002). Symptom development causes changes in gene expression resulting in whole plant physiological alterations such as stunting, vein-clearing, mosaics and chlorosis (Geri *et al.*, 2004; Maule *et al.*, 2002). Disease components in a host-virus interaction can therefore be studied individually or in combination by exploring genetic variation between the two genomes of both a plant and pathogen (Cecchini *et al.*, 1998).

Defence Responses and Signalling

Early changes within the plant associated with localized resistance responses (HR) include:- reactive oxygen intermediates (ROI) resulting from an oxidative burst, alterations in cell wall structure, signalling molecules, nitric oxide (NO) production, endogenous SA increase, and the transcriptional activation of defence-related genes, including those encoding pathogenesis-related (PR) proteins (Feys *et al.*, 2001). Similarly, signalling to uninfected portions of the plant, known as systemic acquired resistance (SAR) establishes a heightened resistance throughout the plant acting against a broad spectrum of pathogens. SAR is a long-lasting form of resistance which requires the phenolic signalling molecule, SA. Alternatively, a different form of systemic resistance, induced systemic resistance (ISR), also exists. This form of resistance is independent of SA, requiring plant growth hormones such as jasmonic acid (JA) and ET, which are effective against a broad spectrum of pathogens (Feys and Parker, 2000; Koornneef and Pieterse, 2008; Pieterse *et al*., 2009). A given pathogen can cause a physiological state of elevated defensiveness and potentiation of the defence activation machinery by SAR and ISR (related in function but act independently), resulting in stronger and more rapid responses to subsequent infections (Wan *et al.*, 2002).

1.6 Molecular Studies Identifying and Engineering Resistant/Tolerant Cassava Germplasm

Uncharacterised Crop Systems - Cassava

Six types of resistance to virus diseases have been proposed: (1) immunity; (2) resistance to virus infection; (3) resistance to establishment and spread of the virus in host plants; (4) virus multiplication resistance; (5) tolerance; and (6) vector resistance (Hahn *et*

al., 1980). CMD epidemics have been addressed principally by resistant cassava germplasm which has been achieved through interspecific crosses between cassava cultivars, of diverse origins, and accessions of a wild relative, *Manihot glaziovii*. Resistance has been described as being largely additive in nature, as well as stable, and with a heritability of about 60% (Hahn *et al.*, 1980; Fregene *et al.*, 2000). The identification of genes responsible for different sources of resistance such as CMD1 has been made possible by the genetic map of cassava using marker-assisted genetic analysis of simple pedigreed populations (Fregene *et al.,* 2000). A second mapping population involving the cassava TME3 line, a Nigerian land race, showed a new form of resistance. TME3 revealed near immunity to the west and east African strains of the virus. The original deployed source of resistance was thus recessive (*CMD1*) and the new source was shown to be a single dominant gene (*CMD2*) in the heterozygous state. Simple regression analysis on marker genotypic classes of disease responses revealed that about 50% of phenotypic variance for resistance from the *M.glaziovii* source (CMD1) was found on a region of chromosome D, whereas, the new source of CMD resistance (CMD2) showed a more than 70% phenotypic resistance on a region on chromosome R ($CIAT_b$, 2001).

In the past, mapping population studies have only been used in the genetics of resistance against African cassava mosaic disease (ACMD) and cassava bacterial blight (CBB). Simple sequence repeat (SSR) markers have been utilized to make marker technology widely available in cassava. These markers are found in all eukaryotic genomes and consist of short tandem repeat motifs usually consisting of 1-6 bp of nucleotides. SSRs were made the molecular markers of choice for genetic mapping and diversity studies due to their high levels of heterozygosity as well as the codominant and PCR-based nature of their repeat loci (Mba *et al.,* 2001). A project to genetically map CMD-resistance genes, analysing a special class of markers known as single-dose restriction fragments (SDRF) was carried out in an attempt to simplify the determination of allelism. These DNA markers are known to be absent in one parent and present in the other and segregate in a 1:1 ratio (absence to presence) in the progeny. SDRFs represent the segregation equivalent of an allele at a heterozygous locus in an allopolyploid or diploid genome. F1 population linkage analysis requires the presence of several unique alleles in either or both parents, resulting in two separate linkage maps, based on female and male sources of markers. Through SDRF marker studies, a dominant gene for resistance to CMD was thus found in a F1 cross between resistant and susceptible parents. This single - dominant - gene finding made it particularly useful in breeding for

CMD resistance and also for identifying genetic markers for marker-assisted-selection (MAS), providing an invaluable tool for breeding CMD resistance, particularly in Latin America where disease is not found but the presence of a vector makes it a threat (Akano *et al.,* 2002). The presence of two different sources of CMD resistance provides a means of combining multiple sources of resistance, although the recessive nature of the older source of resistance (*CMD1*) makes it a less attractive approach, given cassava's heterozygous and out-crossing nature (Akano *et al.,* 2002).

Because cassava is vegetatively propagated and is heterozygous in nature, disease resistance breeding is a considerably slow process. It is therefore important to develop varieties carrying as many different genes for resistance as possible. Stable resistance against a broad spectrum of the virus can be obtained through identifying and pyramiding different disease resistance genes, as viruses are continually evolving and/or are accidentally introduced into infected germplasm. Bacterial Artificial Chromosomes (BAC) were previously adopted to construct large insert libraries used for positional cloning, end-sequencing of large clones, development of physical maps, and high-throughput genome sequencing in an attempt to identify disease and pest resistance genes in cassava (Tomkins *et al.*, 2004). Resistance genes identified through BAC clones containing resistance gene analogues (RGAs) were used to identify gene organization, genome location and gene number in cassava TMS30001 (having the deployed source of resistance from *M.glaziovii*), MECU72 (showing high levels of resistance to whiteflies), and TME3 (source of CMD resistance) (Tomkins *et al.*, 2004).

Although breeding for resistance is still a major means for controlling diseases, fairly little is known about cassavas responses at the molecular level to pathogen attack. In addition, it is still unclear what pathways of resistance are used to prevent virus replication and movement. Two methods such as expressed sequence tags (EST's) and serial analysis of gene expression (SAGE) have been employed to elucidate these resistance mechanisms. In a previous study conducted by Lopez *et al*, 2004 and 2005, a large EST collection was developed in cassava in response to the pathogen *Xanthomonas axonopodis* pv. *manihotis.* From this collection, 5700 unique clone sets were used for microarray analysis. Changes in gene expression were found in the resistant cassava variety, MBra685 at various time points investigated. Genes that were differentially expressed were primarily found to be involved in defence responses. This study identified mechanisms underlying the molecular nature of cassava defence when under pathogen

attack (Lopez *et al.*, 2005). EST's therefore provide insight into expression patterns of transcripts as well as coordinated expression of genes. SAGE is another resource which is a sequence-based method providing a quantitative profile of expressed genes under any condition. This method utilizes 10-11 bp tags from a known location of a transcript. Concatemerization occurs followed by sequencing of 25-50 bp concatemers. These tags are then identified by comparisons made with cDNA databases. A study conducted by Fregene *et al.*, 2004 used a SAGE approach to monitor gene expression patterns in 40 CMD resistant and susceptible genotypes. Gene annotation of 30 differentially expressed tags revealed genes primarily involved in resistance, suggesting future implications for genetic transformation in cassava.

Transcriptome Profiling in Model Plant Systems (Arabidopsis thaliana)

To date, model-plant systems such as *Arabidopsis* have remained the plant-host of choice because of its small size, high transformation efficiencies and short generation time (Koornneef and Meinke, 2010; Meinke et al., 1998). In addition, it is the most thoroughly studied organism since whole genome sequences became available through the *Arabidopsis* Genome initiative (AGI) (The *Arabidopsis* Genome initiative, 2000 reviewed in Seki 2009). Since then, readily available community resources such as 1,500,000 expressed sequence tags (EST's) were deposited in the EST database which included sequences from France, United States and Japan, all providing information from different tissues, organs, and developmental stages with the *Arabidopsis* genome. The wealth of information supplied by the *Arabidopsis* interactome allows for more interdisciplinary and multi-investigative studies to take place especially when unreferenced genomes such as cassava are not available. Sequenced genomes can provide information about conserved genes likely to be involved in the same biological process within and across species such as humans (*Homo sapiens)*, yeast (*Saccharomyces cerevisiae*), fruit-fly (*Drosophila melanogaster*), and nematode worm (*Caenorhabditis elegans*) (Geisler-Lee et al., 2007). Based on existing as well as hypothesis driven data, unknown proteins can be added to networks without the need of cost-prohibitive experimental approaches to validate data (Geisler-Lee et al., 2007).

1.7 Main Objectives and Specific Aims

Cassava mosaic disease (CMD) is a devastating disease causing major yield losses to cassava production annually. SACMV-[ZA:99] along with many plant viruses are difficult to control, bringing about the need for effective disease control strategies. One approach for control is to elucidate mechanisms involved in host defence-responses to an invading pathogen. Host defences and signalling pathways are complex and form highly-connected networks with one another, bringing about the need for high-throughput, complete genome profiling studies. Currently, many screening approaches such as transcriptomics, proteomics, and metabolomics are widely used for high-throughput functional genomic analyses. These approaches identify many candidate genes required for functional analysis on a global level as one approach is limiting when it comes to studying highly diverse networks in plants.

The overall objective of this study was to utilize high-throughput technologies such as whole genome microarrays and next generation sequencing to identify candidate genes involved in model and natural host-defence responses which can be exploited for developing resistant germplasm.

The specific aims of the study were:-

1) Optimize agroinfection efficiencies in recalcitrant crop systems such as cassava using different begomovirus species.

2) To identify host genes governing susceptibility in the interaction between the model host plant system, *Arabidopsis* and SACMV-[ZA:99] using whole genome microarrays.

3) Elucidate mechanisms underlying host-pathogen interactions in uncharacterised genomes such as cassava upon SACMV-[ZA:99] infection using next generation sequencing, with host susceptibility as the primary focus.

1.8 References

- Abramovitch, R.B. and Martin, G.B. (2004) Strategies used by bacterial pathogens to suppress plant defenses. Curr Opin Plant Biol. 7: 356-364.
- Akano AO, Dixon AGO, Mba C, Barrera E, and Fregene M (2002) Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. Theor Appl Gen. 105:*521-525*
- Alleman, J. and Coertze, A.F. (1996) The Cultivation of Cassava. *Vegetable and Ornamental Plant Institute, Pretoria, 1-5.*
- Anderson, J.V., Delseny, M., Fregene, M.A., Jorge, V., Mba, C., Lopez, C., Restrepo, S., Soto, M., Piegu, B., Verdier, V., Cooke, R., Tohme, J. and Horvath, D.P. (2004) An EST resource for cassava and other species of Euphorbiaceae. Plant Mol Biol. 56: 527-539.
- Arguella-Astorga, G., Lopez-Ochoa, L., Kong, L., Orozco, B.M., Settlage, S.B. and Hanley-Bowdoin, L. (2004) A novel motif in geminivirus replication proteins interacts with the plant retinoblastoma-related protein. J Virol. 4817-4826.
- Bent, A.F. (1996) Plant Disease Resistance Genes: Function Meets Structure. Plant Cell 8:1757-1771.
- Berrie LC, Rybicki EP, and Rey MEC (2001) Complete nucleotide sequence and host range of *South African cassava mosaic virus:* Further evidence of recombination amongst begomoviruses. J Gen Virol. 82: 53-58.
- Briddon, R.W., Watts, J., Markham, P.G. and Stanley, J. (1989) The coat protein of *Beet curly top virus* is essential for infectivity. Virology 172:628-633.
- Brigneti, G; Voinnet, O; Li, W; Ding, S; and Baulcombe, D (1998) Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. EMBO J, 17(22):6739-6746.
- Burns A, Gleadow R, Cliff J, Zacarias A and Cavagnaro T (2010) Cassava: The Drought, War and Fame Crop in a Changing World. Sustainability 2, 3572-3607.
- Carrington, J.C. and Whitham, S.A. (1998) Viral invasion and host defense: strategies and counter-strategies. Curr Opin Plant Biol. 1: 336-341.
- Cecchini, E., Al-Kaff, N.S., Bannister, A., Giannakou, M.E., McCallum, D.G., Maule, A.J., Milner, J.J. and Covey, S.M. (1998) Pathogenic interactions between variants of cauliflower mosaic virus and *Arabidopsis thaliana.* J Exp Bot. 49(321): 731-737.

- CIAT$_a$ (Centro Internacional de Agricultura Tropical). (2001) Cassava: A Crop for Hard Times and Modern Times. www.ciat.cgiar.org/ciatinfocus/pdf/cassavafocus.pdf.
- CIAT$_b$(Centro Internacional de Agricultura Tropical)(2001) Cassava`s Potential in the 21st Century:Present Situation and Future Research and Development Needs. Proc. 6th Regional Workshop, held in Ho Chi Minh city, Vietnam. Feb 21-25, 2000.
- 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.
- Dahniya, M.T. (1994) An Overview of Cassava in Africa. Afr Crop Sci J. 2(4): 337-343.
- Dangl, J.L. and Jones, D.G. (2001) Plant pathogens and integrated defense responses to infection. Nature 411: 826-833.
- de la Riva, G., González-Cabera, J., Vázquez-Pardrón, R. and Ayra-Pardo, C. (1998) *Agrobacterium tumefaciens*: a natural tool for plant transformation. Electron Journal of Biotechn. 1(3):118-133.
- Dong, X. (1998) SA, JA, ethylene, and disease resistance in plants. Curr Opin Plant Biol. 1: 316-323.
- Eckardt, N.A. (2004) Host proteins guide *Agrobacterium*-mediated plant transformation. Plant Cell 16: 2837-2839.
- Egelkrout, E.M., Mariconti, L., Settlage, S.B., Cella, R., Robertson, D. and Hanley-Bowdoin, L. (2002) Two E2f elements regulate the proliferating cell nuclear antigen promoter differently during leaf development. Plant Cell 14:3225-3236.
- El-Sharkawy, M.A. (2004) Cassava biology and physiology. Plant Mol Biol. 56: 481-501.
- FAO (2012) Food Outlook Global Market Analysis- November 2012. Rome.
- FAO (2013) FAOSTAT Statistical database (http://Faostat.fao.org).
- Feys, B.J. and Parker, J.E. (2000) Interplay of signaling pathways in plant disease resistance. Trends Genet. 16(10): 449-455.
- Feys, B.J., Moisan, L.J., Newman, M. and Parker, J.E. (2001) Direct interaction between the Arabidopsis disease resistance signaling proteins, EDS1 and Pad4. EMBO J. 20(19): 5400-5411.

- Fregene M, Bernal A, Duque M, Dixon A, and Tohme J (2000) AFLP analysis of African cassava (*Manihot esculenta* Crantz) germplasm resistant to the cassava mosaic disease (CMD). Theor Appl Genet. *100:*678-685.
- 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 (CMD). Plant Mol Biol. 56: 563-571
- Gafni, Y. and Epel, B.L. (2002) The role of host and viral proteins in intra- and inter-cellular trafficking of geminiviruses. Physiol Mol Plant Path. 60:231-241.
- Geisler-Lee J, O'Toole N, Ammar R, Provart NJ, Millar AH et al. (2007) A predicted interactome for Arabidopsis. Plant Physiol. 145(2): 317-329.
- Gelvin, S.B. (2003) *Agrobacterium*-mediated plant transformation: the biology behind the “gene-jockeying” tool. Microbiol Mol Biol Rev. 67(1): 16-37.
- Geri, C., Cecchini, E., Giannakou, M.E., Covey, S.N. and Milner, J.J. (1999) Altered patterns of gene expression elicited by cauliflower mosaic virus (CaMV) infection and by a CaMV gene VI transgene. MPMI 12(5): 377-384.
- Glazebrook, J. (2001) Genes controlling expression of defense responses in *Arabidopsis* – 2001 status. Curr Opin Plant Biol. 4: 301-308.
- Grimsley, N., Hohn, B., Hohn, T. and Walden, R. (1986) “Agroinfection,” an alternative route for viral infection of plants by using the Ti plasmid. PNAS 83:3282-3286.
- Gutierrez, C. (1999) Geminivirus DNA replication. Cell Mol Life Sci. 56: 313-329.
- Gutierrez, C. (2000) DNA replication and cell cycle in plants: learning from geminiviruses. EMBO J. 19(5): 792-799.
- Gutierrez, C., Ramirez-Parra, E., Castellano, M.M., Sanz-Burgos, A.P., Luque, A. and Missich, R. (2004) Geminivirus DNA replication and cell cycle interactions. Vet Microbiol. 98: 111-119.
- Hahn SK,Terry ER, and Leuschner K (1980) Breeding cassava for resistance to cassava mosaic disease. Euphytica *29:* 673-683.
- Hammond-Kosack KE, and Jones JDG (1996) Resistance Gene-Dependent Plant Defense Responses. Plant Cell 8:1773-1791.
- Hanley-Bowdoin, L., Settlage, S.B., and Robertson, D. (2004) Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. Mol Plant Pathol. 5(2):149-156.

- Harrison, B.D. and Robinson, D.J. (2002) Green shoots of geminivirology. Physiol Mol Plant Path. 60: 215-218.
- Hayes, R.J., Coutts, H.A. and Buck, K.W. (1988) Agro-infection of *Nicotiana* spp. with cloned DNA of *Tomato golden mosaic virus.* J Gen Virol. 69:1487-1496.
- Hillocks, R.J. and Thresh, J.M. (2001) Cassava Mosaic and Cassava Brown Streak Virus Diseases in Africa: A comparative guide to symptoms and aetiologies. *Roots* 7(1): 36-50.
- Jacob, S.S., Vanitharani, A., Karthikeyan, S., Chincore, Y., Thillaichidambaram, P. and Veluthambi, K. (2003) *Mungbean yellow mosaic virus*-Vi agroinfection by codelivery of DNA A and DNA B from one *Agrobacterium* strain. Plant Dis. 87(3): 247-251.
- Koornneef A, Pieterse CM (2008) Cross talk in defence signalling. Plant Physiol. 146(3): 839-844
- Koornneef M, Meinke D (2010) The development of Arabidopsis as a model plant. Plant J. 61(6): 909-921
- Legg, J.P., Thresh, J.M. (2000) Cassava mosaic virus disease in East Africa: a dynamic disease in a changing environment. Virus Res. 71: 135-149.
- Leiser, R.M., Ziegler-Graff, V., Reutenauer, A., Herbach, E., Lemaire, O., Guilley, H., Richards, K. and Jonard, G. (1992). Agroinfection as an alternative to insects for infecting plants with beet western yellows luteovirus. PNAS 89:9136-9140.
- Li, H, Lucy, A.P., Guo, H., Li, W., Ji, L., Wong, S. and Ding, S. (1999) Strong host resistance targeted against a viral suppressor of the plant gene silencing defence mechanism. EMBO J. 18(10): 2683-2691.
- Lopez, C., Jorge, V., Piegu, B., Mba, C., Cortes, D., Restrepo, S., Soto, M., Laudie, M., Berger, C., Cooke, R., Delseny, M., Tohme, J., and Verdier, V. (2004) A unigene catalogue of 5700 expressed genes in cassava. Plant Mol Biol. *56:*541-554.
- Lopez, C., S., Soto, Restrepo, S., Piegu, B., Cooke, R., Delseny, M., Tohme, J., and Verdier, V. (2005) Gene expression profile in response to *Xanthomonas axonopodis* pv. *manihotis* infection in cassava using a cDNA microarray. Plant Mol Biol. *57:*393-410.
- Mathews, C. (2000) *Cassava production by small holder farmers in Mpumalanga:A case study*. Lowveld Research unit, Department of Agriculture Conservation and Environment, P/Bag x11318, Nelspruit-1200.

- Maule, A., Leh, V. and Lederer, C. (2002) The dialogue between viruses and hosts in compatible interactions. Curr Opin Plant Biol. 5: 1-6.
- Mayo, M., Ryabov, E., Fraser, G. and Taliansky, M. (2000) Mechanical transmission of *Potato leafroll virus.* J GenVirol 81: 2791-2795.
- Mba REC, Stephenson P, Edwaqrds K, Melzer S, Nkumbira J, Gullberg U, Apel K, Gale M, Tohme, and Fregene M (2001) Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theor Appl Genet. *102*:21-31.
- McDowell, J., Dhandaydham, M., Long, T., Aarts, M., Goff, S., Holub, E. and Dangl, J. (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis.* Plant Cell 10: 1861-1874.
- Meinke DW, Cherry JM, Dean CD, Rounsley S, Koornneef M (1998) Arabidopsis thaliana: a model plant for genome analysis. Science 282:662–682
- Mysore, K.S. and Ryu, C. (2004) Nonhost resistance: how much do we know? Trends Plant Sci. 9(2): 97-104.
- Nestel, B. (1980) Optima 1: Cassava, A new outlook for an ancient crop. *United Nations Food and Agricultural Organisation, 53-68*
- 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. Plant J. 33: 245-257.
- Olsen, K.M. (2004) SNPs, SSRs and inferences on casssava's origin. Plant Mol Biol. 56: 517-526.
- Petty, I.T.D., Carter, S.C., Morra, M.R., Jeffrey, J.L. and Olivey, H.E. (2000) Bipartite geminivirus host adaptation determined cooperatively by coding and noncoding sequences of the genome. Virology 277: 429-438.
- Picó, B., Ferriol, m., Dĭez, M.J. and Viňals, F.N. (2001) Agroinoculation methods to screen wild *Lycopersicon* for resistance to *Tomato yellow leaf curl virus*. J Plant Pathol. 83(3): 215-220.
- Pieterse CM, Leon-Reyes A, Van der Ent S, Van Wees SC (2009) Networking by small-molecule hormones in plant immunity. Nat Chem Biol. 5(5): 308-316
- Scholthof, H.B. (2001) Molecular plant-microbe interactions that cut the mustard. Plant Physiol. 127: 1476-1483.

- Staskawicz, B.J., Ausubel, F.M., Baker, B.J., Ellis, J.G. and Jones, J.D.G. (1995) Molecular genetics of plant disease resistance. *Science* 268: 661-667.
- Stenger, D.C., Carbonaro, D. and Duffus, J.E. (1990) Genomic characterization of phenotypic variants of *beet curly top virus.* J Gen Virol. 71: 2211-2215.
- Stenger, D.C., Revington, G.N., Stevenson, M.C. and Bisaro, D.M. (1991) Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. PNAS 88: 8029-8033.
- Storey, H.H. (1936) Virus diseases of East African plants. VI. A progress report on studies of the diseases of cassava. East Afr Agr J. 2: 34-39.
- Takahashi, H., Miller, J., Nozaki, Y., Sukamto, Takeda, M., Shah, J., Hase, S., Ikegami, M., Ehara, Y. and Dinesh Kumar, S.P. (2002) *Arabidopsis thaliana RPP8/HRT* family resistance gene, conferring resistance to *Cucumber mosaic virus* requires salicylic acid, ethylene and a novel signal transduction mechanism. Plant J. 32: 655-667.
- Tang, X., Xie, M., Kim, Y.J., Zhou, J., Klessig, D.F. and Martin GB (1999) Overexpression of *Pto* Activates Defense Responses and Confers Broad resistance. Plant Cell. 11:15-29.
- Tomkins, J., Fregene, M., Main, D., Kim, H., Wing, R., and Tohme, J. (2004) Bacterial artificial chromosome (BAC) library resource for positional cloning of pest and disease resistance genes in cassava (*Manihot esculenta* Crantz). Plant Mol Biol. 56:555-561.
- Tzfira, T. and Citovsky, V. (2002) Partners-in-infection:host proteins involved in the transformation of plant cells by *Agrobacterium*. Trends Cell Biol. 12(3): 121-129.
- 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 suppression of posttranscriptional gene silencing. J Virol. 9487-9498.
- Venisse, J., Malnoy, M., Faize, M., Paulin, J. and Brisset, M. (2002) Modulation of defense responses of *Malus* spp. during compatible and incompatible interactions with *Erwinia amylovora.* MPMI 15(12): 1204-1212.
- Voinnet, O., Pinto, Y., and Baulcombe, D. (1999) Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. PNAS 96:24, 14147-14152.
- Wan, J., Dunning, F.M. and Bent, A.F. (2002) Probing plant-pathogen interactions and downstream defense signaling using DNA microarrays. Funct Integr Genomics 2: 259-273.

- Ward, D.V. and Zambryski, P.C. (2001) The six functions of *Agrobacterium* VirE2. PNAS 98(2): 385-386.

Chapter 2

A Comparative Study of Improved Begomovirus Agroinfection Efficiencies in Cassava (*Manihot esculenta* Crantz)

2.1 Abstract

Agroinfection of plant cells with cloned DNA of geminiviruses involves the transfer of replicative infectious dimers to plant cells. A common difficulty with inoculation methods is the low infection efficiencies obtained in recalcitrant crop systems such as cassava. The purpose of this research was to improve agroinfection efficiencies of *South African cassava mosaic virus* (SACMV-[ZA:99]) and *African cassava mosaic virus* (ACMV-[NG:Ogo:90]) and to compare infectivity between three susceptible cassava cultivars (T200, TMS60444, and SM14334) following infection with three different *Agrobacterium* strains (C58C1; AGL1; LBA4404). Adaptation of existing protocols resulted in a 66% increase in infection efficiency in cassava genotype T200 for SACMV-[ZA:99] (C58C1) and 45% for SACMV-[ZA:99] (AGL1). For ACMV (LBA4404), a 20% infection increase was achieved in genotype TMS60444, while for ACMV (C58C1) a 62% increase was obtained, showing an overall improvement to the method. Infectivity comparisons between and among genotypes revealed consistently high infection rates for ACMV-[NG:Ogo:90] (C58C1) in all three genotypes and varying levels of infection efficiencies for SACMV-[ZA:99] (C58C1), suggesting basal levels of susceptibility differ between genotypes. SACMV-[ZA:99] (AGL1) infections were consistently higher (78%) than SACMV-[ZA:99] (C58C1) (52%), confirming the advantage of using the hypervirulent *Agrobacterium* strain AGL1. Inoculum levels also demonstrated significant differences in infection. This study demonstrated that virus agroinfection efficiencies are significantly influenced by differences among cassava cultivars (genotypes), geminivirus species, and *Agrobacterium* strains.

2.2 Introduction

Geminiviruses are emerging plant pathogens, severely affecting many staple food and cash crops of great economic importance worldwide (Mansoor *et al*., 2006, Vanderschuren *et al*., 2007). Biological and molecular properties such as genome organization, insect vector and host range separate geminiviruses into four genera, namely *Mastrevirus, Curtovirus, Topocuvirus,* and *Begomovirus* (Fauquet *et al*., 2008; Gafni and Epel, 2002; Harrison and Robinson, 2002). Begomoviruses infect dicotyledonous plants and are transmitted by the whitefly, *Bemisia tabaci (Genn)*, which is considered to be a species complex occurring in warm temperate, tropical and subtropical regions worldwide (Harrison and Robinson, 2002). Cassava (*Manihot esculenta* Crantz) is a major carbohydrate source for over 500 million people in the tropics, sub-tropical Africa, Asia and Latin America, with an estimated production of 233 million tonnes in 2008 (FAOSTAT website http://faostat.fao.org). Cassava Mosaic Disease (CMD) is caused by eight distinct species of begomoviruses, six on the African continent and two from the Indian subcontinent (Fauquet *et al*., 2008). These species all contain single-stranded circular bipartite genomes (DNA-A and DNA-B) of approximately 2.7kb in size. DNA-A encodes proteins associated with replication and transcriptional activation and DNA-B encodes proteins involved in movement (Harrison and Robinson, 1999). They use a double-stranded (ds) DNA replication intermediate in the infected plant cell nucleus and replicate by means of a rolling circle mechanism (Hanley-Bowdoin *et al*., 2000) and/or recombination-dependent replication (Jeske *et al*., 2001).

Inoculation methods currently used for geminivirus infection studies include mechanical and insect transmission, biolistics, grafting, and agroinfection. More recently, a DNA abrasion technique has proven to be successful among geminiviruses that are not phloem-limited (Ascencio-Ibañez and Settlage, 2007). Agroinfection, originally introduced by Grimsley *et al*. (1986) has been widely used for geminiviruses that are recalcitrant to mechanical inoculation such as *Maize streak virus* (MSV) in maize (Grimsley *et al*., 1987), *Beet curly top virus* (BCTV) in *Nicotiana benthamiana, Beta-vulgaris,* and *Datura stramonium* (Briddon *et al*., 1989), *Tomato yellow leaf curl virus* (TYLCV) in *Lycopersicon* species (Picó *et al*., 2001), *Wheat dwarf virus* in *Triticum aestivum* (Hayes *et al*., 1988a), *Tomato golden mosaic virus* in *N. benthamiana* (Hayes *et al*., 1988b), and yellow mosaic viruses infecting legumes (Usharani

et al., 2005). The method has also shown success for plant viruses with linear plus-strand RNA genomes (Leiser *et al*., 1992).

Agroinfection is widely used to infect experimental plants with viruses to study host resistance, transgenic plant resistance, plant-virus interactions, and virus replication. It involves the construction of full-length or partial dimers into the T-DNA (transfer DNA) of a binary vector which is then mobilized into *Agrobacterium tumefaciens*. Following inoculation into plants, the viral dimer is released into plant cells via T-DNA mediated transfer (Grimsley *et al*., 1986, 1987; Hayes *et al.*, 1988a, b). In geminiviruses, unit-length circular replicative viral copies are released by one of two mechanisms. For full-length or partial dimers, a replication-associated protein (Rep) initiates replication at the origin of replication (ori) and host DNA polymerase replicates the full-length circular genome when two oris are present. However, when one ori is present, homologous recombination occurs, allowing the release of unit-length double-stranded forms of viral DNA between the tandem repeats (Rigden *et al*., 1996; Shivaprasad *et al*., 2006; Stenger *et al*., 1991).

Agroinfection offers many practical advantages over other inoculation methods such as insect transmission, due to its simplicity and ease (Ascencio-Ibañez and Settlage, 2007; Elmer *et al*., 1988). It is also difficult to apply uniform inoculum pressures with insect vectors, such as whitefly, as transmission efficiencies may vary due to host genotypes, vector biotypes, and growth conditions (Harrison and Robinson, 1999). Another limiting factor to this method is plant resistance, whereby an insect vector can mask the effect of virus resistance, as seen in wild *Lycopersicon* species (Chanarayappa *et al*., 1992). Higher infection rates have also been achieved with agroinfection compared to mechanical inoculation procedures, as some viruses such as *Tomato golden mosaic virus* (TGMV), are phloem limited and are thus transmitted at low frequencies (Elmer *et al*., 1988), or not at all, as in *Potato leaf roll virus* (Mayo *et al*., 2000). In cassava particularly, graft inoculation procedures have proven to be efficient but with resultant delays in symptom development (Ariyo *et al*., 2003). Mechanical transmission has shown to be unsuccessful with *South African cassava mosaic virus* (SACMV-[ZA:99]) in cassava, and agroinfection efficiencies in cassava has been low (20-66%) compared to biolistic inoculation (40-84%) (Berrie *et al*., 2001; Vanderschuren *et al*., 2009). Particle bombardment has shown success in cassava but is an expensive procedure and has shown variation in infection efficiencies among genotypes (Ariyo *et al*., 2003; Briddon

et al., 1998; Rothenstein *et al*., 2005). The more recent DNA abrasion technique has shown to be extremely successful in solanaceous species but this technique is limited to certain plant species (Ascencio-Ibañez and Settlage, 2007).

Agroinfection is particularly effective for geminiviruses that are insect transmitted, since needle injection techniques mimic insect stylets through the apoplastic cell wall components, directly targeting and puncturing phloem cells which consist of sieve elements (SE), companion cells (CC), or vascular parenchyma (VP) cells. This enables geminivirus delivery specifically into tissues that may support their replication (Wege and Pohl, 2007). Agroinfection has proven to be successful for many geminivirus plant systems, making it the inoculation procedure of choice, albeit some crops demonstrating low infection rates. This study was conducted to improve agroinfection efficiencies in recalcitrant crop systems such as cassava by adapting existing protocols in order to obtain higher infection percentages. We showed susceptibility differences among three cassava genotypes (T200, TMS60444, and SM14334) to agroinoculation by two begomovirus species, ACMV-[NG:Ogo:90] and SACMV-[ZA:99]. Significant comparative virulence differences between *Agrobacterium* strains AGL1 and C58C1 were also demonstrated.

2.3 Materials and Methods

Plant Genotypes, Acclimatization Procedure and Growth Conditions

Susceptible cassava genotypes and breeding origins used in the study are shown in Table 1. Plants were grown *in vitro* and acclimatized to growth chamber conditions as follows: Nodal cuttings were placed onto Murashige and Skoog basal salts containing vitamins (Murashige and Skoog, 1962) (Highveld Biological (PTY) LTD), supplemented with 20g sucrose and reduced agar (6.8g) and grown until the emergence of roots and shoots (1-2 weeks). The plantlets were then transferred to Jiffy Peat Pellets (Jiffy Products International) on trays which were covered in plastic wrap for adaptation to chamber conditions. Depending on the vigour of the genotype, plantlets were carefully exposed to chamber conditions by small razor-like slits in the plastic wrap to avoid air flow around individual plants. This procedure takes approximately 2-4 weeks. All cassava cultivars were maintained at 26°C under a 12h day at an intensity of 150 $\mu Em^{-2}sec^{-1}$.

Agrobacterium Strains and Begomovirus Combinations

A list of *Begomovirus* species, *Agrobacterium* strains and cloning vector types used in the infectivity assays are presented in Tables 1, and 2. Full-length head-to-tail dimers of DNA-A and DNA-B components of SACMV-[ZA:99] (GenBank accession **AF155806**) were constructed separately in pBIN19 (Berrie *et al*., 2001) and mobilized into *Agrobacterium tumefaciens* strains C58C1pMP90 (Koncz and Schell, 1986) and AGL1 (Lazo *et al*., 1991) by the freeze-thaw method of Holsters *et al*., 1978. The infectious clones of ACMV-[NG:Ogo:90] DNA-A and DNA-B (GenBank accession **AJ427910**) (Liu *et al*., 1997) were constructed in pCAMBIA1300 as tandem repeats (Vanderschuren *et al*., 2009) and mobilized separately into *Agrobacterium tumefaciens* C58C1 by electroporation.

Table 1: Cassava genotypes and breeding origins, Begomovirus strains, cloning vector and dimer constructs, *Agrobacterium* strains and total number of plants infected for testing with protocol 3

Cassava genotype	Cultivar origin	Dimer construct and cloning vector		*Agrobacterium* strain	Total no. of plants tested for protocol 3
T200	Local landrace, South Africa	SACMV-[ZA:99] DNA A + B	Full-length head-to-tail dimers in pBIN19	C58C1(pMP90)	12
T200	Local landrace,South Africa	SACMV-[ZA:99] DNA A + B	Full-length head-to-tail dimers in pBIN19	AGL1	23
T200	Local landrace,South Africa	ACMV-[NG:Ogo:90] DNA A+B	Tandem repeats in pCAMBIA1300	C58C1(pMP90)	16
SM14334	CIAT,South America*	SACMV-[ZA:99] DNA A + B	Full-length head-to-tail dimers in pBIN19	C58C1(pMP90)	11
SM14334	CIAT,South America*	SACMV-[ZA:99] DNA A + B	Full-length head-to-tail dimers in pBIN19	AGL1	17
SM14334	CIAT,South America*	ACMV-[NG:Ogo:90] DNA A +B	Tandem repeats in pCAMBIA1300	C58C1(pMP90)	17
TMS60444	IITA,NigeriaAfrica*	SACMV-[ZA:99] DNA A + B	Full-length head-to-tail dimers in pBIN19	C58C1(pMP90)	10
TMS60444	IITA,NigeriaAfrica*	SACMV-[ZA:99] DNA A + B	Full-length head-to-tail dimers in pBIN19	AGL1	13
TMS60444	IITA,NigeriaAfrica*	SACMV-[ZA:99] DNA A + B	Full-length head-to-tail dimers in pBIN19	LBA4404	6
TMS60444	IITA,NigeriaAfrica*	ACMV-[NG:Ogo:90] DNA A +B	Tandem repeats in pCAMBIA1300	C58C1	23
TMS60444	IITA,NigeriaAfrica*	ACMV-[NG:Ogo:90] DNA A +B	Tandem repeats in pCAMBIA1300	LBA4404	6
TMS60444	IITA,NigeriaAfrica*	ACMV-[NG:Ogo:90] DNA A +B	Tandem repeats in pCAMBIA1300	GN54D	6

*CIAT: Center for International Tropical Agriculture, Colombia; IITA: International Institute for Tropical Agriculture, Nigeria

Table 2: *Agrobacterium* strains used in the study showing chromosomal backgrounds and Ti-plasmids from which they are derived

***A.tumefaciens* strain**	**Chromosomal background**	**Ti-plasmid**	**Antibiotic resistance**	**Reference**
AGL1	C58	pTiBO542	Rifampicin, Carbenicillin	Lazo *et al*.,1991
C58C1pMP90	C58	pTiC58	Rifampicin, Gentamicin	Koncz and Schell., 1986
LBA4404	Ach5	pTiAch5	Rifampicin	Hoekema *et al*., 1983

Protocol Comparisons

Three protocols for virus infectivity were tested in different susceptible cassava genotypes, and named protocol 1, 2 and 3 accordingly (Table 3). Protocol 1 (H. Vanderschuren, unpublished laboratory protocol) and protocol 2 (Berrie *et al*., 2001; L. Berrie, unpublished laboratory protocol) were pre-existing protocols. Protocol 3 was adapted in this study from protocol 2 and agroinfection efficiencies tested in different cassava cultivars (genotypes) with different *Agrobacterium* strains and begomoviruses, ACMV-[NG:Ogo:90] and SACMV-[ZA:99] (Table 1) .

Table 3: Comparative agroinfection protocols tested in cassava

Protocol 1 (H. Vanderschuren, unpublished laboratory protocol)	**Protocol 2** (L. Berrie, unpublished laboratory protocol and Berrie *et al*.,2001)	**Protocol 3** (Improved protocol in this study)
Step1: Day 1: Streak Agro-dimer A and Agro-dimer B on selected YEB plates containing appropriate antibiotics and incubate at 28 ^{0}C Step 2: Day 3: Transfer over-grown Agro-dimer A and Agro-dimer B to eppendorf tubes. Mix equal amounts of Agro-dimer A and Agro-dimer B by weighting. Vortex to homogenize. Step 3: Use 4-6 weeks old cassava plantlets and make 1 cm longitudinal incisions (4 per plant) on the apical part of the stem. Step 4: Cover each incision with the Agrobacterium mixture using an inoculation loop. Keep the plants for 1 hour at 22°C before bringing them back to greenhouse conditions. Symptoms should appear 2 to 3 weeks post-infection	Step 1:Day1: Inoculate 200µl of Agro-dimer A into 5ml LB (containing appropriate antibiotics) and incubate overnight at 30°C Step 2:Day 2: Inoculate 500ul of overnight culture into 5ml LB and incubate cultures until an OD600 of +/-0.4 is reached Step3:Spin down 1ml cultures at 8000 rpm for 1min to pellet cells, remove supernatant Step 4:Add 1ml sterile water, mix and spin at 8000rpm for 1min Step 5:Resuspend in 200ul of LB Step 6:Mix equal volumes of each culture Step7:Inoculate each plant approx 2-4 times with 10ul at diff places along stem from bottom to top (alot leaks out) Incubate plants and disease symptoms should appear in 2-3 weeks	Step 1: Day1: Inoculate 500µl of Agro-dimer A into 5ml of LB and 500µl of Agro-dimer B into 5ml of LB (each containing appropriate antibiotics). Incubate at 30°C overnight. Step 2:Day2: Obtain OD_{600} 1.8/2.0 (+/-18h). Sub-inoculate 4ml of each culture into 30ml LB (each containing appropriate antibiotics). Step 3: Day3: Obtain OD_{600} 1.8/2.0 (+/-24h). Spin down 1ml aliquots of each culture at full speed (13 400rpm) for 1 minute to pellet cells. Remove supernatant. Step 4: Add 1ml sterile water, mix and spin at full speed for 1 minute. Step 5: Resuspend in 200µl LB Step 6:Mix equal volumes of each culture (A+B) Step 7: Wound stems by needle puncture first, then inject each plant with approximately 100ul (for a 10cm plant height) of culture along the stem, concentrating primarily on plant meristem (apex) Step 8:Cover plants for approx 2 days and re-acclimatize (i.e. small holes in covering daily until plants have readjusted to growth chamber conditions Symptoms should start appearing from 14 to 21 dpi; plants fully-susceptible by 28 days post infection

Optical Density and Direct Cell Enumeration Comparisons between Protocol 2 and Protocol 3

Agrobacterium C58C1pMP90 cultures containing both SACMV-[ZA:99] DNA-A or DNA-B dimers were grown in Luria Broth media supplemented with antibiotics, Rifampicin (50µg/ml) and Kanamycin (100µg/ml). In order to evaluate the effect of bacterial inoculum (cell numbers) on agroinfection, serial dilutions were performed to compare colony forming units between protocol 2, OD_{600} at 0.4 and protocol 3, OD_{600} at 1.8/2.0. At the end of each respective OD, 1ml of culture was transferred to a saline solution containing 0.85% NaCl. Cell numbers were then determined by duplicate spotting (50µl) of 10-fold serial dilutions using the droplet technique developed by Becker and colleagues (1994).

Symptom Evaluation

Cultivars T200, SM14334, and TMS60444 were assessed for symptom severity once all genotypes displayed symptoms 28 days post infection (dpi). Symptom severity was based on a 5 class CMD scoring system of Hahn *et al*., 1980. A score of 1 was assigned for no symptoms, and 5 for extreme symptoms such as severe distortion, mosaic and reduction of entire leaflet.

Total Nucleic Acid (TNA) Extraction and PCR

In order to confirm phenotypic infectivity status, TNA was extracted from 45 leaf samples according to the CTAB (cetyltrimethyl ammonium bromide) method of Doyle and Doyle (1987). Concentrations were determined using the Nanodrop® ND-1000 Spectrophotometer. PCR was carried out using SACMV-[ZA:99] BC1 primer pair:-
BC1385 (5' GGACTAGTACCCAGGTTTAGCCCACA 3')
BC2184 (5' GAAGATCTTGGACGCCCAATTTACCG 3')

Virus-free (healthy) leaf material was included as a negative control. PCR cycling conditions included an initial denaturation at 94°C for 5 minutes, followed by 30 cycles of denaturation at 94°C for 30 seconds, annealing at 50°C for 30 seconds, primer extension at

72°C for 30 seconds, and a final extension of 72°C for 7 minutes. All reactions were carried out in a MyCycler™ Thermal Cycler (Bio-Rad).

Statistical Analyses

All statistical test's such as t-test, F-test, Analysis of variance (ANOVA) (for virus infectivity comparisons, bacterial virulence and genotype susceptibility), and Bland-Altman plots for protocol comparisons were carried out using GraphPad Prism® Version 5.02 (www.graphpad.com). All values were based on mean infection percentages. Significance testing was based on two hypotheses. The null hypothesis (H_0), where virus infectivity or bacterial virulence differences are equal or similar across and/or within each cassava genotype tested, or the alternate hypothesis (H_a), where they are not equal, and are thus significantly different. ANOVA post-test's such as Bonferroni correction (Graphpad Prism®) was applied to test significance in infectivity or virulence between two begomovirus or *Agrobacterium* strains within a genotype. Additionally, The Bland-Altman method (Graphpad Prism®) compared virus infectivity between two protocols using the average and differences between two virus/bacterial combinations. Difference versus average was calculated as follows:

The average between two data points: A + B/2;

The difference between two data points: A-B;

Where:

A denotes SACMV-[ZA:99] (C58C1) and B, SACMV-[ZA:99] (AGL1) in T200 for protocol 2 versus protocol 3.

A denotes ACMV (LBA40444) and B, ACMV (C58C1) in TMS60444 for protocol 1 versus protocol 3.

Values were then constructed to create the Bland-Altman plots.

2.4 Results

Method Comparisons

Agroinfection efficiencies (percentage) obtained from protocol 1 and 2, were compared to protocol 3 (Table 4). Protocol 2 achieved an approximate 25% (SACMV-[ZA:99] in

Agrobacterium strain C58C1pMP90) and 33% (SACMV-[ZA:99] in *Agrobacterium* strain AGL1) virus agroinfection efficiency in local cassava landrace T200 (Table 4). Protocol 3 achieved a significantly higher infection rate of 78% (SACMV-[ZA:99] in *Agrobacterium* strain AGL1) and 91% (SACMV-[ZA:99] in *Agrobacterium* strain C58C1pMP90) compared to protocol 2 (Figure 1A). Protocol 3 was also compared with protocol 1 using ACMV-[NG:Ogo:90] in cassava genotype, TMS60444. For ACMV (in LBA4404), a slight increase from 80% (protocol 1) to 100% (protocol 3) was achieved, while infection with ACMV-[NG:Ogo:90] (in C58C1) resulted in a significant infection increase from protocol 1 (25%) to protocol 3 (87%) (Figure 1C). The Bland-Altman method (Graphpad Prism®) was applied at the 95% significance level to test for significant differences (Y-axis) in average virus infectivity (X-axis) between two protocols. In Figure 1B (T200), mean values for infectivity using protocol 2 was 29% (blue dot) and protocol 3, 85% (pink dot), indicating a significant difference in infection percentages between the two protocols. Similarly, protocol 3 also displayed a significantly higher infection rate of 94% (purple dot) when compared to protocol 1 (53%) (green dot) in TMS60444 (Figure 1D). The significant increase in agroinfection efficiency achieved using protocol 3 is clearly demonstrated in Figure 1. Additionally, Table 4 shows comparisons made between protocols 1, 2 and 3 with SACMV-[ZA:99] (in C58C1pMP90) infection in genotype TMS60444. Protocol 1 and 2 achieved a 20% infection efficiency, which was increased to 40% with protocol 3.

Table 4: Protocol efficiency comparisons between two Begomovirus/*Agrobacterium* combinations and two susceptible cassava genotypes

Cassava genotype	**Begomovirus/*Agrobacterium* strain combination**	**Total No. of plants infected / No. of plants inoculated (%)**		
		Protocol 1	**Protocol 2**	**Protocol 3**
T200	SACMV(C58C1)	NT*	1/4 (25%)	10/11 (91%)
T200	SACMV(AGL1)	NT*	2/6 (33%)	18/23 (78%)
TMS60444	SACMV(C58C1)	2/10 (20%)	2/10 (20%)	4/10 (40%)
TMS60444	ACMV(C58C1)	3/12 (25%)	NT*	20/23 (87%)
TMS60444	ACMV(LBA4404)	8/10 (80%)	NT*	6/6 (100%)

*NT – Not tested

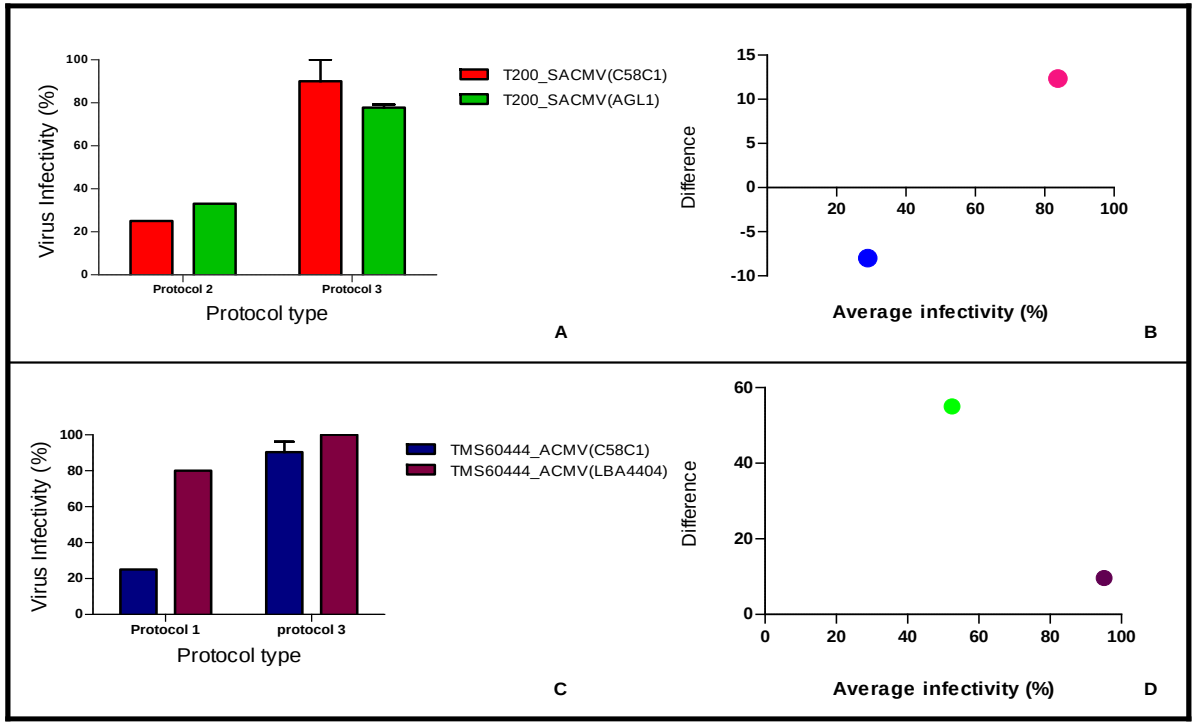


Figure 1: A) Histogram showing improvement in infection from 25% (protocol 2) to 91% (protocol 3) for SACMV (C58C1), and 33% (protocol 2) to 78% (protocol 3) for SACMV (AGL1) in T200. Bars indicate standard error. B) Bland-Altman plot showing mean infectivity differences of 29% for protocol 2 and 85% for protocol 3. C) Histogram showing improvement in infection from 25% (protocol 1) to 87% (protocol 3) for ACMV (C58C1), and 80% (protocol 1) to 100% (protocol 3) for ACMV (LBA4404) in TMS60444. Bars indicate standard error. D) Bland-Altman plot showing mean infection differences of 53% for protocol 1 and 94% for protocol 3.

Viable Agrobacterium Cell Enumeration Comparisons between Protocol 2 and 3

One of the major differences between protocol 2 and protocol 3 was the optical density readings (0.4 and 1.8/2.0). In order to determine if cell number was a major factor contributing to agroinfection efficiency, *Agrobacterium* viable cell counts (cells/ml) were compared at each OD respectively. *Agrobacterium* strain C58C1pMP90 containing SACMV-[ZA:99] constructs was used for protocol comparisons. From Figure 2A, it is evident that *Agrobacterium* cells are approximately 10 fold higher at an OD of 1.8/2.0 than an OD of 0.4 for *Agrobacterium* containing either SACMV-[ZA:99] DNA-A or DNA-B dimer. At an OD of 0.4, 4.5×10^8 cells/ml were present and 3.3×10^9 cells/ml at OD 1.8/2.0 for SACMV-[ZA:99] DNA-A. For DNA-B, 2.5×10^8cells/ml were present at OD 0.4 and 3.0×10^9 cells/ml at OD 1.8/2.0. Results from an unpaired one- tailed t-test showed that OD values and hence viable cell counts were significantly higher at an OD of 1.8/2.0 than at an OD of 0.4 ($p < 0.05$).

Symptom Development and Scoring

Visual assignment of severity scores are depicted in Figure 3, based on the 1-5 scoring system of Hahn *et al*., 1980. All infected cassava genotypes displayed a severity index score of 4 out of 5 (two thirds distortion of a leaflet, mosaic and reduction in leaf size) to 5 out of 5 (severe distortion, mosaic and reduction of entire leaflet). Mock-inoculated controls were assigned a 1 out of 5 score (healthy leaf material showing no symptoms).

PCR Validations

In order to confirm SACMV-[ZA:99] replication in plants showing visual symptoms, PCR was carried out on 45 leaf samples. A list of samples that were tested and confirmed for infection status is observed in Table 6. SACMV-[ZA:99] BC1 primers successfully amplified a 799 base pair region from SACMV-[ZA:99] DNA-B genome for all infected plants phenotypically displaying geminivirus symptoms (Table 6).

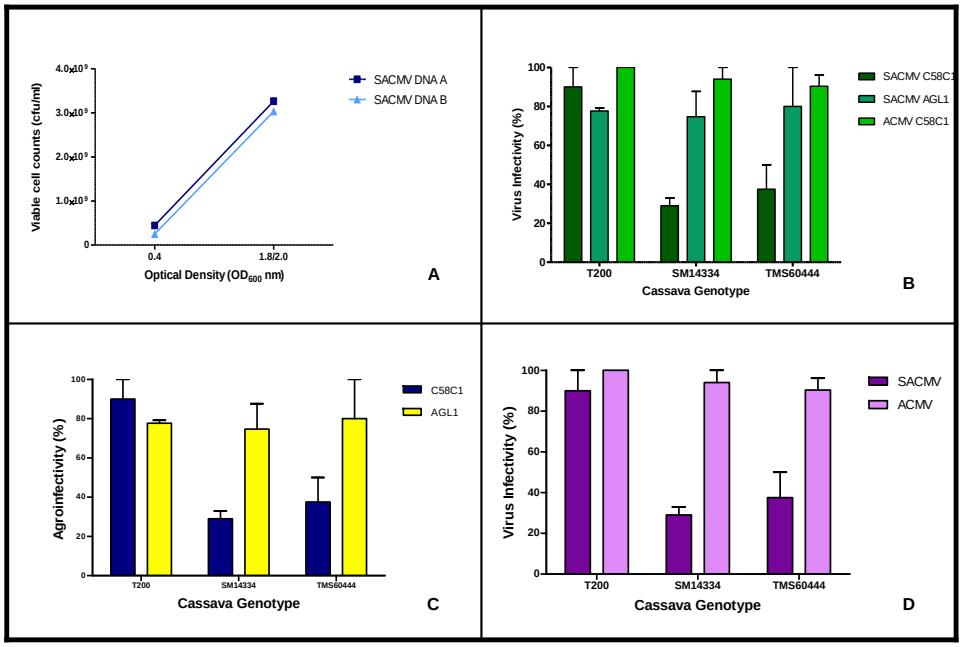


Figure 2: A) Viable cell counts of *Agrobacterium tumefaciens* strain C58C1 containing SACMV DNA-A and DNA-B dimers, taken at OD_{600} 0.4 and OD_{600} 1.8/2.0, respectively. Data represents average cell counts from 2-3 replicate experiments; bars indicate the standard error. B) Infectivity percentage comparisons between and within cassava genotypes T200, SM14334, and TMS60444. C) Bacterial virulence (in percentage) comparisons between C58C1 and AGL1 (containing SACMV dimers) & between and within genotypes T200, SM14334, and TMS60444; bars indicate standard error. D) SACMV and ACMV virulence differences between and within genotypes T200, SM14334, and TMS60444. Bars indicate the standard error.

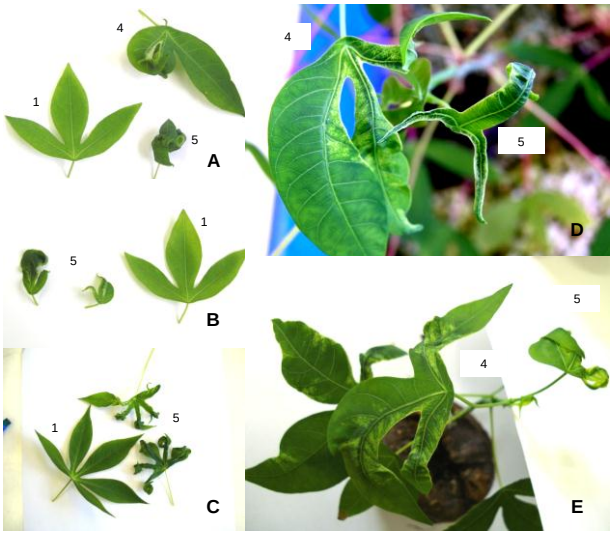


Figure 3: Disease severity phenotypes (score 1, no symptoms to 5, severe symptoms) of infected cassava leaves in A) T200 (SACMV in C58C1pMP90), B) SM14334 (SACMV in AGL1), and C) TMS60444 (ACMV in C58C1). D, and E) Typical SACMV symptoms in cassava T200 plants showing cassava mosaic disease symptoms.

Infectivity Assays Utilizing Protocol 3

Virus Species and Genotype Infection Comparisons

Virus infection efficiency (percentage) comparisons between begomoviruses SACMV-[ZA:99] (in C58C1pMP90), SACMV-[ZA:99] (in AGL1), and ACMV-[NG:Ogo:90] (in C58C1) in each of the cassava genotypes T200, SM14334 and TMS60444, were performed. Mean virus infection percentages and standard error values are displayed in Table 5, illustrated in Figure 2B, respectively. Virus infection comparisons across three genotypes (mean) revealed that ACMV-[NG:Ogo:90] (in C58C1) had the highest infection efficiency of 95%, followed by SACMV-[ZA:99] (in AGL1), 78%, with SACMV-[ZA:99] (in C58C1) displaying the lowest infectivity of 52% ($p= 0.0002$; unmatching two-way ANOVA). Susceptibility differences between genotypes T200, SM14334 and TMS60444 were compared by obtaining mean virus infection percentages for SACMV-[ZA:99] (in C58C1), SACMV-[ZA:99] (in AGL1), and ACMV-[NG:Ogo:90] (in C58C1). Of the three genotypes tested, T200 was most susceptible to virus infection (89%), followed by TMS60444 (69%), and SM14334 (66%) ($p = 0.0125$) (Figure 2B).

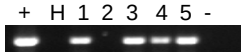
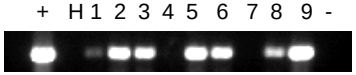
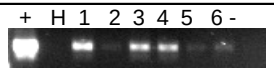
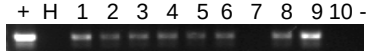
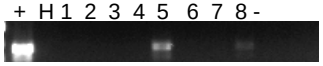
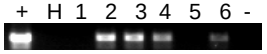
Bonferonni post test's were carried out in order to determine significance in infectivity between two virus/bacterial combinations within a particular genotype (Figure 2B). A 12% infection difference between SACMV-[ZA:99] (C58C1) (90%) and SACMV-[ZA:99] (AGL1) (78%) was not significant ($p>0.05$) in T200. In SM14334, SACMV-[ZA:99] (C58C1) (29%) and SACMV-[ZA:99] (AGL1) (75%) revealed a highly significant difference in infectivity of 46% ($p<0.01$). A significant difference of 42% was also observed in TMS60444 ($p<0.05$) for SACMV-[ZA:99] (C58C1) (38%) and SACMV-[ZA:99] (AGL1) (80%). Virus infectivity differences (10%) between SACMV-[ZA:99] (C58C1) (90%) and ACMV (C58C1) (100%) in T200 was not significant ($p>0.05$). However, a highly significant difference of 65% ($p<0.001$) was observed between SACMV-[ZA:99] (C58C1) (29%) and ACMV (C58C1) (94%) in SM14334. Similarly, a 52% significant difference ($p<0.01$) was obtained for SACMV-[ZA:99] (C58C1) (38%) and ACMV (C58C1) (90%) in TMS60444. No significant infection differences ($p>0.05$) were obtained for SACMV-[ZA:99] (AGL1) (78%) and ACMV (C58C1) (100%) in T200, or SACMV-[ZA:99] (AGL1) (75%) and ACMV (C58C1) (94%) in SM14334, as well as in TMS60444 with SACMV-[ZA:99] (AGL1) (80%) and ACMV (C58C1) (90%) (Figure 2B).

Table 5: Agroinfectivity results from replicate experiments utilizing protocol 3

Cassava genotype	Begomovirus species	Agrobacterium Strain	No. of plants infected /no. of plants inoculated (%)			Mean infection (%)	Standard Error (SEM) (%)	Total No. of Replicates
			Replicate exp1	Replicate exp2	Replicate exp3			
T200	SACMV	C58C1	4/5 (80%)	6/6 (100%)	NT*	90	10	2
T200	SACMV	AGL1	7/9 (78%)	8/10 (80%)	3/4 (75%)	78	1.5	3
T200	ACMV-NOg	C58C1	7/7 (100%)	6/6 (100%)	3/3 (100%)	100	0	3
SM14334	SACMV	C58C1	2/8 (25%)	1/3 (33%)	NT*	29	4	2
SM14334	SACMV	AGL1	4/6 (67%)	4/4 (100%)	4/7 (57%)	75	13	3
SM14334	ACMV-NOg	C58C1	3/3 (100%)	3/3 (100%)	9/11 (82%)	94	6	3
TMS60444	SACMV	C58C1	3/6 (50%)	1/4 (25%)	NT*	38	12.5	2
TMS60444	SACMV	AGL1	3/3 (100%)	6/10 (60%)	NT*	80	20	2
TMS60444	ACMV-NOg	C58C1	8/10 (80%)	2/2 (100%)	10/11 (91%)	90	5.8	3

*NT – Not Tested

Table 6: PCR showing 799 bp SACMV BC1 amplicons from agroinfected cassava genotypes T200 and SM14334 showing symptoms

Cassava cultivar	**Virus/*Agrobacterium* combination**	**Phenotypic symptom observation (plants infected/plants inoculated)**	**PCR infection result (BC1 primers)**
T200 Replicate 1	SACMV C58C1(pMP90)	4/5	4/5 
T200 Replicate 1	SACMV AGL1	7/9	7/ 9 
T200 Replicate 2	SACMV C58C1(pMP90)	6/6	6/6 
T200 Replicate 2	SACMV AGL1	8/10	8/10 
SM14334 Replicate 1	SACMV C58C1(pMP90)	2/8	2/8 
SM14334 Replicate 1	SACMV AGL1	4/6	4/6 

Key: (+) positive control, SACMV DNA-B in pBS (+/- 100ng/ul), H denotes healthy virus-free negative control.
(-) negative control, water.

Numbers to the left in column 4 (PCR results) indicate a positive correlation between number of PCR amplicons obtained and number of agroinfected cassava showing visible symptoms (column 3)

Comparisons in Infection Efficiencies between Agrobacterium Strains

Virulence comparisons between *Agrobacterium* strain C58C1 and AGL1, containing SACMV-[ZA:99] constructs were determined by obtaining the mean across genotypes T200, SM14334 and TMS60444 (Figure 2C). C58C1 was determined to be less virulent, inducing 52% infection, whereas AGL1 displayed higher infectivity (78%). The difference in virulence (as determined by their ability to induce infection) between the two strains was considered significant (F-test, $p<0.05$). Within genotype comparisons showed AGL1 to be more virulent (75%) than C58C1 (29%) in SM14334. Similarly, in TMS60444, AGL1 demonstrated 42% higher infectivity of 80% compared to C58C1 (38%). However, in T200, the difference in agroinfection efficiency between AGL1 (78 %) and C58C1 (90%) was much lower for AGL1 (Figure 2C).

Comparisons between Begomoviruses, ACMV-[NG:Ogo:90]* and *SACMV-[ZA:99]* in *Agrobacterium* C58C1 *in Cassava Genotypes T200, TMS60444, and SM14334

Significant differences in virus infectivity (mean average percentages) between SACMV-[ZA:99] and ACMV-[NG:Ogo:90] were tested using an unmatched two-way ANOVA (Figure 2D). ACMV-[NG:Ogo:90] and SACMV-[ZA:99] displayed a 95% and 52% mean infection efficiency across all three genotypes, respectively. Virus infectivity differences was shown to be significant at $p = 0.0012$. Genotype T200 was significantly more susceptible to infection (95%) by SACMV-[ZA:99] and ACMV-[NG:Ogo:90], followed by TMS60444 (64%), and SM14334 (62%), $p<0.0001$(Figure 2D). Bonferonni post test's were applied for testing significance between virus species within a single genotype. A 10% infection difference was obtained between SACMV-[ZA:99] (90%) and ACMV-[NG:Ogo:90] (100%). This slight difference in infectivity was not considered significant ($p>0.05$). Significance was however observed in genotypes SM14334 and TMS60444. SM14334 displayed an infection difference of 65% between SACMV-[ZA:99] (29%) and ACMV-[NG:Ogo:90] (94%), $p<0.001$. Similarly, for TMS60444, a 52% difference was obtained between SACMV-[ZA:99] (38%) and ACMV-[NG:Ogo:90] (90%), $p<0.01$ (Figure 2D).

2.5 Discussion

Infectivity studies of begomoviruses in plant hosts, such as cassava, and screening of genetically modified crops for virus resistance rely on efficient and reliable methods for

virus challenge. This study was initiated to improve agroinfection efficiencies in the recalcitrant crop cassava, through comparative evaluation of virulence between begomoviruses SACMV-[ZA:99] and ACMV-[NG:Ogo:90] in three *Agrobacterium* strains in three susceptible cassava genotypes. We have demonstrated increased agroinfection efficiencies by using an improved and adapted method (protocol 3) of protocol 2 (L.C Berrie, unpublished laboratory protocol; L.C Berrie *et al*., 2001) in three cassava genotypes. Criteria that were critically evaluated in protocol 3 were optical density (OD) and thus cell number determination (step 3), method of injection (additional step concentrating on apical meristematic tissue) (step 7), and inclusion of an incubation/acclimatization period of 2 days post agroinfection (step 8).

Results from this study clearly illustrated increased agroinfection efficiency of protocol 3 compared to protocol 1 and 2 (Table 4). A significant increase in infectivity of 66 % was observed between protocol 2 and protocol 3 for SACMV-[ZA:99] in *Agrobacterium* strain C58C1 and 45% for SACMV-[ZA:99] in AGL1 in cassava genotype T200 (Figure 1A). Infection percentages also increased from 25 to 87% for ACMV (C58C1) and by 20 % for ACMV in LBA4404 from protocol 1 to protocol 3 in genotype TMS60444 (Figure 1C). Significant differences in infection rates between protocols were also confirmed by Bland-Altman plots (Figures 1 B and D), where mean values for infectivity (average between SACMV-[ZA:99] (C58C1) and SACMV-[ZA:99] (AGL1)) using protocol 2 was 29% and protocol 3, 85% in T200. Similarly, in TMS60444, protocol 3 also displayed a significantly higher agroinfection efficiency of 94% than in protocol 1, obtaining 53% (average between ACMV (C58C1) and ACMV (LBA4404)). Finally, protocol 3 showed a 20% infection increase for SACMV-[ZA:99] (C58C1) in TMS60444 (Table 4). One contributing factor for higher infection efficiencies achieved in protocol 3 may be due to *Agrobacterium* cell number increase. Cell number has been shown to be an important determinant for improved agroinfection efficiencies in other crops (Cruz *et al*., 1999; Hayes *et al*., 1988b). An approximate 10 fold increase was observed from ~10^8cells/ml at an OD of 0.4 to ~10^9 cells/ml (OD 1.8/2.0) for both SACMV-[ZA:99] DNA-A and DNA-B (Figure 2A), suggesting that an increase in bacterial inoculum concentration (cell numbers/ml) contributes to an increase in infection rates. An increase in bacterial cell numbers may allow for more viral DNA to be delivered via T-DNA transfer into plant cells, resulting in rapid virus replication and systemic movement. According to Hayes *et al*. (1988b), an 80% infection rate was achieved for cells ranging from 2 x 10^5 to 2 x 10^9 cells/ml whereas a 0% infection rate was achieved for 2 x 10^3 cells/ml. Cruz *et al*., 1999, also tested bacterial inoculum levels in

three genotypes of rice against *Rice tungro bacilliform virus* (RTBV). At inoculums levels of 0.2 x 10^{12} cells/ml, infection rates never achieved anything greater than 39% whereas at higher (2 fold) inoculum level of 0.5 x 10^{12} cells/ml, up to 100% infection was achieved for susceptible rice varieties. This demonstrated the importance of high inoculums levels required for recalcitrant crops.

We suggest that another possible attribute leading to increased agroinfection efficiencies in cassava , obtained in this study, was the application of needle injection along the stem (as suggested in protocol 2), but concentrating at the apex of the plant (not suggested in protocol 2). This is an important application as geminivirus replication must occur in actively growing, dividing cells (Elmer *et al*., 1988). It has been shown in tomato that stem agroinfection is more effective than leaf agroinfection as it allows direct injection into vascular tissues thereby allowing long-distance transport and systemic spread, targeting the shoot apex, meristematic tissues, and roots from the inoculation point (Picó *et al*., 2001). Thus, by targeting the apex and meristematic tissue directly, a more effective infection process is suggested as T-DNA transfer, virus replication and spread can be initiated, bypassing the problem of injecting into non-dividing cells. Studies conducted in meristematic tissue of maize (*Zea mays*) have shown that bacterial inoculum containing *Maize streak virus* (MSV) injected into meristematic tissues close to or at the apex of the coleoptilar node achieved higher infection efficiencies (83%) than when injected into non-meristematic regions (5-16%) (Grimsley *et al*.,1988). In addition, younger plants contain more meristematic tissue and are thus more susceptible to infection, particularly at the apex as it contains a region which differentiates to produce germ cells, rendering it more effective than leaf inoculation methods (Grimsley *et al*., 1988).

Covering of cassava plantlets after inoculation in order to increase temperature and incubation time for *Agrobacterium* (optimal growth temperature is 30°C) as performed in protocol 3 (step 8), proved to add to improved agroinfection rates. Temperature has been shown to be a signalling factor that aids in promoting the expression of *vir* genes (Jones *et al*., 2005). A temperature increase may encourage bacterial replication at inoculation sites or inside the host (before T-DNA release), enabling multiplication of virus copies. Phenotypic symptoms in SACMV-[ZA:99] - agroinoculated cassava were observed at 28 dpi when plants were fully-susceptible. Most infected plants displayed an index score of 4 out of 5 to 5 out of 5 in the three genotypes tested, indicating severe infection (Figure 3). Symptom development in susceptible cassava varieties has been shown to correlate with

virus replication and is detected by PCR (Briddon *et al*., 1998). Virus detection was confirmed in our study by a positive correlation between PCR amplicons obtained and agroinfected plants showing symptoms (Table 6).

For infectivity assays utilizing protocol 3, across genotype comparisons resulted in ACMV-[NG:Ogo:90] (C58C1) displaying the highest infection efficiency of 95%, followed by SACMV-[ZA:99] (AGL1) with 75%, and SACMV-[ZA:99] (C58C1), 52% (Table 5, Figure 2B). Bonferonni post test's for within genotype comparisons showed significant differences in virus infectivity between SACMV-[ZA:99] (C58C1) and SACMV-[ZA:99] (AGL1) in genotypes SM14334 and TMS60444, but not in T200 (Table 5, Figure 2B). Similarly, SACMV-[ZA:99] (C58C1) and ACMV (C58C1) infection differences were significant in SM14334 and TMS60444, but not significant in T200 (Table 5, Figure 2B). No significance in virus infectivity was observed for SACMV-[ZA:99] (AGL1) and ACMV (C58C1) in T200, TMS60444, and SM14334 (Table 5, Figure 2B). Significance thus illustrated that begomoviruses have different infectivity (virulence) levels in certain genotypes. This study therefore clearly illustrated that certain begomovirus combinations were equally/similarly effective in initiating infection in particular genotypes, while in other genotypes, significant differences in infectivity (virulence) levels were observed. Differences in virulence between virus species and differential genotype levels of resistance or susceptibility to cassava mosaic disease (CMD) have been shown to exist among cassava genotypes (Ariyo *et al*., 2003). The reason why 100% infection was not consistently obtained for all infection assays may be attributable to *Agrobacterium* delivery to cells where virus is not able to replicate and thereby spread (Briddon *et al*., 1989). Another factor that may pose a problem is that for bipartite genomes it is essential that both genomic components A and B are delivered to the same cell. This holds true for Begomoviruses as both genomes are required for viral replication (DNA A) and systemic movement (DNA B) (Elmer *et al*., 1988; Gafni and Epel, 2002; Harrison and Robinson, 2002). In some biological systems such as cowpea (*Vigna unguiculata*), 100% infection was observed with *Cowpea mosaic virus* (CPMV) as was demonstrated with the presence of both RNA components within a cell (Liu and Lomonossoff, 2002). Thus, it is evident that 100% efficiencies may be attained but the effectiveness of each method is dependent on virus, bacterial strain and host genotype combination used as well as efficient delivery of DNA components into cells (Ascencio-Ibañez and Settlage, 2007).

It has been reported that success of T-DNA transfer may be dependent on specific receptors such as glycoproteins on the cell wall of plant cells (Schläppi and Hohn, 1992). According to Marks *et al*., 1989, certain *Agrobacterium* strains are able to bind more efficiently to a plant cell wall than others. Observations made in maize immature embryos suggest that cell wall receptors may be produced at specific stages during shoot apical meristem development (Schläppi and Hohn, 1992). Competence for T-DNA transfer would therefore be dependent on genotype and developmental stage (Schläppi and Hohn, 1992). In this study, competence was based primarily on genotype, as all young plantlets were inoculated at the same developmental stage. Bacterial competence for T-DNA transfer was illustrated with SACMV-[ZA:99] (in AGL1) in genotypes SM14334 and TMS60444, that would normally be less permissive to infection as shown with SACMV-[ZA:99] (in C58C1) (Figure 2B). However a further study to test the age of cassava plantlets and agroinfection competence would be informative.

Increased bacterial virulence, as demonstrated by the hypervirulent AGL1 (Lazo *et al*., 1991) strain also played a role in infection efficiencies, as revealed by higher agroinfection rates of SACMV-[ZA:99] in AGL1 (78%) compared to SACMV-[ZA:99] in C58C1 (52%) across genotypes (Figure 2C). Within genotype comparisons, AGL1 was 46% more virulent than C58C1 in SM14334 and 42% in TMS60444, but was less virulent (by 12%, which is not considered significant) in T200. The hypervirulence of *Agrobacterium* strain AGL1 may be sufficient to overcome inhibition to infection caused by non-permissive host barriers. It has been documented that differences in agroinfection competence between genotypes is dependent on factors specific to a certain genotypes such as (among others) *vir*-inducing substances, endogenous hormone levels, and availability of receptors to aid in bacterial attachment to meristematic cells (Schläppi and Hohn, 1992). Cell wall receptors have also been shown to contain a regulation mechanism as seen in Pinto bean where efficient attachment of *Agrobacterium* is prevented (Schläppi and Hohn, 1992). We could speculate that lower infection efficiencies obtained for SACMV-[ZA:99] (C58C1) in genotypes TMS60444 and SM14334 compared with high infection efficiencies obtained for T200 (Figure 2B) were related to cell wall receptors and efficient *Agrobacterium* attachment, however this would need to be explored. On the other hand, host barrier differences may also be attributed to inefficient virus attachment or replication leading to reduced movement (Surendranath *et al*., 2005).

All T-DNA events have the ability to allow replication and systemic spread of viruses, but limitations arise with factors such as host range determinants. Vir induction has been shown to be important in host range determination, particularly VirA. VirA induces phenolic compounds and works directly with CHVE, a sugar-binding protein, forming a complex that is required for agroinfection in maize (Heath *et al*., 1997). *VirA* and *virG* genes have been shown to increase the induction of *vir* genes of the hypervirulent AGL1 strain, suggesting that improvements in transformation efficiency relates to increased induction of *vir* genes in AGL1 (Chabaud *et al*., 2003). It was also documented that some *Agrobacterium* strains may be highly efficient in transforming one plant species or cultivar of the same species and may not be as efficient in another (Heath *et al*., 1997). This was illustrated with SACMV-[ZA:99] (C58C1) in cultivars T200, TMS60444, and SM14334, where a high infection rate of 90% was obtained in T200, but was low in TMS60444 (38%) and SM14334 (29%) (Table 5, Figure 2B). Efficient T-DNA transfer also relies on the repression of the host silencing machinery that targets small RNAs against the T-DNA at the initial stage of infection (Dunoyer *et al*., 2006). Differences in infection efficiencies may also be connected to complex interplay between the *Agrobacterium* strain and the host silencing machinery.

Alternatively, the low infections rates observed for SACMV-[ZA:99] (C58C1) in TMS60444 and SM14334 may be attributable vector construction as well as the chromosomal background of the *Agrobacterium* strain. Both SACMV-[ZA:99] (C58C1) and SACMV-[ZA:99] (AGL1) were constructed in pBIN19, a low copy, broad host range plasmid that contains an RK2 replicon. ACMV-[NG:Ogo:90] (C58C1) constructs were made in a pCAMBIA1300 vector which contains a pVS1 replicon. It has been reported that deletions of viral sequences have been observed in binary plasmids containing RK2 replicons in partial dimers, thus suggesting a more tolerant nature of pVS1 replicons for replicating repeated sequences. The chromosomal background of *Agrobacterium* C58 strains has been shown to support deletions of unit-length viral sequences. *Agrobacterium* strains with C58 chromosomal backgrounds and binary vectors with RK2 replicons should thus be avoided (Shivaprasad *et al*., 2006). This may be an explanation as to why SACMV-[ZA:99] (C58C1) had extremely low infection rates in TMS60444 and SM14334, as it contains both the RK2 replicon and a C58 chromosomal background. The high infection rate achieved for T200 could possibly be attributed to its highly susceptible nature, being a local South African and natural host to SACMV-[ZA:99] under field

conditions, suggesting a high affinity between virus and host plant, even with the presence of the RK2 replicon and C58 chromosomal background.

Efficiency of agroinfection is thus dependent on the *Agrobacterium* strain, *vir* functions and strength of promoters upstream from the replication-associated protein found on DNA-A (Jacob *et al*., 2003). Another factor worth mentioning is that systemic infection is due to the rate of secondary spread of the virus from cells once inoculated and is thus not due to continuous infection of cells by *Agrobacterium* (Elmer *et al*., 1988). ACMV-[NG:Ogo:90] achieved a higher infection efficiency of 95% than SACMV-[ZA:99], 52% across all genotypes (Table 5, Figure 2D). Genotype susceptibility also impacts on virus infection efficiency, which would be linked to several factors including the rate of systemic spread. The role of DNA B-encoded movement proteins may play a role in this regard. Movement proteins are diverse among the cassava geminiviruses, and differences in virus movement efficiencies between CMGs may be a factor (Hehnle, *et al*., 2004). T200 showed the highest susceptibility to infection (95%), followed by TMS60444 (64%), and SM14334, 62% (comparisons made between SACMV-[ZA:99] C58C1 and ACMV C58C1, ruling out bacterial differences) (Figure 2D). ACMV-[NG:Ogo:90] consistently achieved a higher infection rate across all three cassava genotypes compared to SACMV-[ZA:99] (Figure 2D) despite being cloned into *Agrobacterium* with a similar C58 chromosomal background as SACMV-[ZA:99]. Plant vector genetic backgrounds between ACMV-[NG:Ogo:90] (pCAMBIA1300) and SACMV-[ZA:99] (pBIN19) are different and may have some impact on efficiency of transfer.

This research demonstrated an efficient method to improve geminivirus infection efficiencies in the recalcitrant crop, cassava, by obtaining an optimal OD for virus infection, including a method of injection at the apex of the plant, and an incubation/acclimatization step. The study further highlighted the complex interplay between virus genetic factors involved in infection, host genotype factors supporting virus replication and movement, plant binary vector genetic background, and *Agrobacterium* virulence factors, that govern agroinfection efficiencies in crops. Suggestions for future improvements involve testing a variety of *Agrobacterium* strains and binary vectors in recalcitrant crop systems. Heath *et al*. (1997) has shown that constitutively expressing *vir* genes in mutant derivatives of *Agrobacterium* is effective in some plant systems that are difficult to transform, although expressed at lower levels than wild-type systems. Additionally, other factors such as testing a variety of recombinant strains enabling a broader host range, and development of

more virulent *Agrobacterium* strains, may be useful for crops systems that are recalcitrant to transformation.

2.6 References

- Ariyo, O.A., Koerbler, M., Dixon, A.G.O., Atiri, G.I., and Winter, S. (2003). Development of an efficient virus transmission technique to screen cassava genotypes for resistance to cassava mosaic disease. Conference on International Agricultural Research for Development, Deutscher Tropentag, Göttingen, October 8-10.
- Ascencio-Ibañez, J.T., and Settlage, S.B. (2007). DNA abrasion onto plants is an effective method for geminivirus infection and virus-induced gene silencing. J Virol Methods 142, 198-203.
- Becker, B., Holzapfel, W.H., and von Holy, A. (1994). Effect of pH and the bacteriocin carnocin 54 on growth and cell morphology of two *Leuconostoc* strains. Lett. Appl. Microbiol. 19, 126-128.
- Berrie, L.C., Palmer, K.E., Rybicki, E.P., and Rey, M.E.C. (1998). Molecular characterization of a distinct South African cassava infecting geminivirus. Arch Virol. 143, 2253-2260.
- Berrie, L.C. (2000). Characterisation of a new distinct geminivirus infecting cassava in South Africa, and development of virus resistance in tobacco and cassava. PhD Thesis
- 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 of recombination among begomoviruses. J Gen Virol. 82, 53-58.
- Briddon, R.W., Watts, J., Markham, P.G., and Stanley, J. (1989). The coat protein of *Beet curly top virus* is essential for infectivity. Virology 172, 628-633.
- Briddon, R.W., Liu, S., Pinner, M.S., and Markham, P.G. (1998). Infectivity of *African cassava mosaic virus* clones to cassava by biolistic inoculation. Arch Virol. 143, 2487-2492.
- Chabaud, M., de Carvalho-Niebel, F., and Barker, D.G. (2003). Efficient transformation of *Medicago truncatula* cv. Jemalong using the hypervirulent *Agrobacterium tumefaciens* strain AGL1. Plant Cell Rep. 22, 46-51.

- Chanarayappa, C., Shivashankar, G., Muniyappa, V., and Frist, R.H. (1992). Resistance of *Lycopersicon* species to *Bemisia tabaci*,a tomato leaf curl virus vector. Can J Bot. 70, 2184-2192.
- de Tarso de Oliviera Ferreira, P., Lemos, T.O., Nagata, T., Inoue-Nagata, A.k. (2008). One-step cloning approach for construction of agroinfectious begomovirus clones. J Virol Methods 147, 351-354.
- Doyle, J.J., and Doyle, J.L. (1987). A rapid isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19, 11-15.
- Dunoyer, P., Himber, C., Voinnet, O. (2006). Induction, suppression and requirement of RNA silencing pathways in virulent Agrobacterium tumefaciens infections. Nat Genet. 38(2): 258-63.
- Elmer, J.S., Sunter, G., Gardiner, W.E., Brand, L., Browning, C.K., Bisaro, D.M., and Rogers, S.G. (1988). *Agrobacterium* – mediated inoculation of plants with tomato golden mosaic virus DNAs. Plant Mol Biol. 10, 225-234.
- El-Sharkawy, M.A. (2004). Cassava biology and physiology. Plant Mol Biol. 56,481-501.
- Fauquet, C.M., and Fargette, D. (ed.) (1988). Proceedings of the International Seminar: African cassava mosaic disease and its control. CTA/ORSTOM, Ede, Holland.
- 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*, an updated list of begomovirus species. Arch Virol.148,405-421.
- Fauquet, C.M., Briddon, R.W., Brown, J.K., Moriones, E., Stanley, J., Zerbini, M., and Zhou, X. (2008). Geminivirus strain demarcation and nomenclature. Arch Virol. 153, 783-821.
- Gafni, Y., and Epel, B.l. (2002). The role of host and viral proteins in intra- and inter-cellular trafficking of geminiviruses. Physiol Mol Plant Pathol. 60, 231-241.
- Grimsley, N., Hohn, B., Hohn, T., and Walden, R. (1986). “Agroinfection”, an alternative route for viral infection of plants by using the Ti plasmid. PNAS. 88,3282-3286.
- Grimsley, N., Hohn, T., Davies, J.W., and Hohn, B. (1987). *Agrobacterium*-mediated delivery of infectious maize streak virus into maize plants. Nature 325,177-179.

- Grimsley, N.H., ramos, C., Hein, T., and Hohn, B. (1988). Meristematic tissues of maize plants are most susceptible to agroinfection with *Maize streak virus*. Nature Biotechnol. 6, 185-189.
- Hahn, S.K., Terry, E.R., and Leuschner, K. (1980). Breeding cassava for resistance to Cassava Mosaic Disease. Euphytica 29, 673-683.
- Hanley-Bowdoin, L., Settlage, S.B., Orozco, B.m., Nagar, S., Robertson, D. (2000). Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. Crit. Rev. Biochem. Mol. Biol. 35, 105-140.
- Harrison, B.D., and Robinson, D.J. (1999). Natural genomic and antigenic variation in whitefly-transmitted geminiviruses (Begomoviruses). Ann. Rev. Phytopathol. 37,369-398.
- Harrison, B.D., and Robinson, D.J. (2002). Green shoots of geminivirology. Physiol Mol Plant Path. 60, 215-218.
- Hayes, R.J., Macdonald, H., Coutts, R.H.A., and Buck, W. (1988a). Agroinfection of *Triticum aestivum* with cloned DNA of Wheat dwarf virus. J Gen Virol. 69, 891-896.
- Hayes, R.J., Coutts, R.H.A., and Buck, K.W. (1988b). Agroinfection of *Nicotiana* spp. With cloned DNA of Tomato golden mosaic virus. J Gen Virol. 69, 1487-1496.
- Heath, J.D., Boulton, M.I., Raineri, D.M., Doty, S.L., Mushegian, A.R., Charles, T.C., Davies, J.W., and Nester, E.W. (1997). Discrete regions of the sensor protein Vir A determine the strain-specific ability of *Agrobacterium* to agroinfect maize. MPMI. 10(2), 221-227.
- Hehnle, S.,Wege, C., and Jeske, H. (2004). Interaction of DNA with the movement proteins of geminiviruses revisited. J Virol. 78(14), 7698-7706.
- Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J., and Schilperoort, R.A. (1983). A binary plant vector strategy based on separation on vir and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. Nature 303, 179-180.
- Holsters, M., de Waele, D., Depicker, A., Messens, E., van Montagu, M., and Schell, J. (1978). Transfection and transformation of *Agrobacterium tumefaciens.* Molecular Gen Genet 163, 181-187.
- Jacob, S.S., Vanitharani, R., Karthikeyan, A.s., Chinchore, Y., Thillaichidambaram, P., and Veluthambi, K. (2003). *Mungbean yellow mosaic virus*- Vi agroinfection by co-delivery of DNA A and DNA B from one *Agrobacterium* strain. Plant Dis. 87(3), 247-251.

- Jeske, H., LÜtgemeier, M., and Preiß, W. (2001). DNA forms indicate rolling circle and recombination dependent replication of *Abutilon* mosaic virus. EMBO J. 20(21), 6158-6167.
- Jones, H.D., Doherty, A., and Wu, H. (2005). Review of methodologies and a protocol for the *Agrobacterium*-mediated transformation of wheat. Plant Methods 1: 5 doi:10.1 186/1746-4811-5.
- Koncz, C., and Schell, J. (1986). The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Mol Gen Genet. 204, 383-396.
- Lazo, G.R., Stein, P.A., and Ludwig, R.A. (1991). A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. Biotechnol. 9, 963-967.
- Leiser, R.M., Ziegler-Graff, V., Reutenauer, A., herrbach, E., Lemaire, O., Guilley, H., Richards, K., and Jonard, G. (1992). Agroinfection as an alternative to insects for infecting plants with beet western yellows luteovirus. Proc. Natl. Acad. Sci. 89, 9136-9140.
- Liu, L., and Lomonossoff, G.P. (2002). Agroinfection as a rapid method for propagating *Cowpea mosaic virus*-based constructs. J Virol Methods. 105, 343-348.
- Mansoor, S., Zafar, Y., Briddon, R.W. (2006). Geminivirus disease complexes: the threat is spreading. Trends Plant Sci. 11(5), 209-212.
- Marks, M.S., Kemp, J.M., Woolston, C.J., and Dale, P.J. (1989). Agroinfection of wheat: a comparison of *Agrobacterium* strains. Plant Sci. 63, 247-256.
- Mayo, M., Ryabov, E., Fraser, G., and Taliansky, M. (2000). Mechanical transmission of *Potato leafroll virus*. J Gen Virol. 81, 2791-2795.
- Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol.Plant. 15,473.
- Olsen, K.M. (2004). SNPs, SSRs and inferences on cassava's origin. Plant Mol biol. 56, 517-526.
- Picó, B., Ferriol, M., Diez, M.J., and Viñals, F.N. (2001). Agroinoculation methods to screen wild *Lycopersicon* for resistance to *Tomato yellow leaf curl virus*. J Plant Pathol. 83(3), 215-220.
- Rigden, J.E., Dry, I.B., Krake, L.R., and Rezaian, M.A. (1996). Plant virus DNA replication processes in *Agrobacterium:* Insight into origins of geminiviruses? Proc. Natl. Acad. Sci. 93, 10280-10284.

- Rothenstein, D., Briddon, R.W., haible, D., Stanley, J., Frischmuth, T., and Jeske, H. (2005). Biolistic infection of cassava using cloned components of *Indian* cassava mosaic virus. Arch Virol. 150, 1669-1675.
- Schläppi, M., and Hohn, B. (1992). Competence of immature maize embryos for Agrobacterium-mediated gene transfer. Plant Cell 4, 7-16.
- Shivaprasad, P.V., Thomas, M., Balamani, V., Biswas, D., Vanitharani, R., Karthikeyan, A.S., and Veluthambi, K. (2006). Factors contributing to deletion with *Mungbean yellow mosaic virus* partial dimmers in binary vectors used for agroinoculation. J Virol Methods. 137(1), 72-81.
- Sta.Cruz, F.C., Boulton, M.I., Hull, R., and Azzam, O. (1999). Agroinoculation allows the screening of rice for resistance to *Rice tungro bacilliform virus*. J. Phytopath. 147, 653-659.
- Stenger, D.C., Revington, G.N., Stevenson, M.C., and Bisaro, D.M. (1991). Replicational release of geminivirus genomes from tandemly repeated copies: Evidence for rolling-circle replication of a plant viral DNA. Proc. Natl. Acad. Sci. 88, 8029-8033.
- Surendranath, B., Usharani, K.S., Nagma, A., Victoria, A.K., and Malathi, V.g. (2005). Absence of interaction of genomic components and complementation between *Mungbean yellow mosaic India virus* isolates in cowpea. Arch. Virol. 150, 1833-1844.
- Usharani, K.S., Surendranath, B., Haq, Q.M.R., and Malathi, V.G. (2005). Infectivity analysis of a soybean isolate of *Mungbean yellow mosaic india virus* by agroinoculation. J Gen Plant Pathol. 71, 230-237.
- Vanderschuren, H., Stupak, M., Futterer, J., Gruissem, W., and Zhang, P. (2007). Engineering resistance to geminiviruses—review and perspectives. Plant Biotechnol J. 5(2), 207–220.
- Vanderschurem, H., Alder, A., Zhang, P., and Gruissem, W. (2009). Dose-dependent RNAi-mediated geminivirus resistance in the tropical root crop cassava. Plant Mol. Biol. 70, 265-272.
- Wege, C., and Pohl, D., 2007. *Abutilon mosaic virus* DNA B component supports mechanical virus transmission, but does not counteract begomoviral phloem limitation in transgenic plants. Virol. J. 365, 173-186

Chapter 3

This chapter has been submitted and accepted for publication:
Pierce, EJ and Rey, MEC (2013) Assessing Global Transcriptomic Changes in Response to *South African cassava mosaic virus* [ZA-99] Infection in Susceptible *Arabidopsis thaliana.* PLoS ONE, 8(6):e67534

3.1 Abstract

In susceptible plant hosts, co-evolution has favoured viral strategies to evade host defenses and utilize resources to their own benefit. The degree of manipulation of host gene expression is dependent on host-virus specificity and certain abiotic factors. In order to gain insight into global transcriptome changes for a geminivirus pathosystem, *South African cassava mosaic virus* [ZA:99] and *Arabidopsis thaliana*, 4 x 44K Agilent microarrays were adopted. After normalization, a 2-fold change filtering of data ($p<0.05$) identified 1,743 differentially expressed genes in apical leaf tissue. A significant increase in differential gene expression over time correlated with an increase in SACMV accumulation, as virus copies were 5-fold higher at 24 dpi and 6-fold higher at 36 dpi than at 14 dpi. Many altered transcripts were primarily involved in stress and defense responses, phytohormone signalling pathways, cellular transport, cell-cycle regulation, transcription, oxidation-reduction, and other metabolic processes. Only forty-one genes (2.3%) were shown to be continuously expressed across the infection period, indicating that the majority of genes were transient and unique to a particular time point during infection. A significant number of pathogen-responsive genes were suppressed during the late stages of pathogenesis, while during active systemic infection (14 to 24 dpi), there was an increase in up-regulated genes in several GO functional categories. An adaptive response was initiated to divert energy from growth-related processes to defense, leading to disruption of normal biological host processes. Similarities in cell-cycle regulation correlated between SACMV and *Cabbage leaf curl virus* (CaLCuV), but differences were also evident. Differences in gene expression between the two geminiviruses clearly demonstrated that, while some global transcriptome responses are generally common in plant virus infections, temporal host-specific interactions are required for successful geminivirus infection. To our knowledge this is the first geminivirus microarray study identifying global differentially expressed transcripts at 3 time points.

3.2 Introduction

In a compatible host, plant viruses manipulate and recruit host metabolites for translation and replication of their genomes and silence host responses through suppressors, despite attempts by the host to mount a defense response (Nagar *et al*., 1995; Whitham *et al*., 2003; Havelda *et al*., 2003; Trinks *et al*., 2005; Whitham *et al*., 2006; Dardick 2007; Agudelo-Romero *et al*., 2008; Babu *et al*., 2008a; Babu *et al*., 2008b). Virus infection causes host cells to over- or under - express certain pathways, causing both physiological and phenotypic changes in the host (Havelda *et al*., 2003; Whitham *et al*., 2003; Whitham *et al*., 2006; Babu *et al*, 2008a; Babu *et al*., 2008b; Ascencio-Ibanez *et al*., 2008; Owens *et al*., 2012). The degree of transcriptome change that a particular host undergoes will change spatially and temporally, and will depend on the compatibility and adaptibility of the pathogen. This host-genotype combination thus determines the severity and type of symptoms displayed (Agudelo-Romero *et al*., 2008; Babu *et al*, 2008a; Babu *et al*., 2008b, Elena *et al*., 2012). Disease formation is the outcome once a virus has successfully completed genome replication, spread through the plasmodesmata to neighbouring cells and colonised distal tissues by vascular dependent long-distance movement in the host plant (Carrington and Whitham, 1998; Maule *et al*., 2002; Agudelo-Romero *et al*., 2008).

Viral proteins are able to accumulate to much higher levels than host proteins in order to fufill their required tasks in replication, movement and suppression of host defences (Whitham *et al*., 2006). This in turn has a huge impact on host cells and causes abnormalities in plant growth and development. Not all changes in host gene expression and metabolism are initiated by specific interactions between virus and host proteins, and alterations can also be consequences of general accumulation of viral proteins and subversion of cellular components (Whitham *et al*., 2003). Plant viruses are biotrophic pathogens which cause alterations (either by induction or repression) to a wide array of cellular processes, at transcriptional, translational or posttranslational levels (Lozan-Duran *et al*., 2011). These processes include, among others, hormonal regulation, cell-cycle control and endogenous transport of macromolecules (Hevelda *et al*., 2003; Whitham *et al*., 2003; Whitham *et al*., 2006; Babu *et al*, 2008a; Babu *et al*., 2008b; Ascencio-Ibanez *et al*., 2008; Pallas and Garcia, 2011). From an evolutionary perspective, a constant battle between plant defense and virus infection exists. Plants are capable of counteracting the effects of virus attack with pre-existing physical and chemical barriers (constitutive

defense), which if overcome by the virus, activate signalling pathways (induced responses) as the next line of defense. Constitutive (preformed) defences are usually non-specific and are effective against a wide array of abiotic and biotic stresses. Induced responses are more targeted and are triggered upon herbivorous insect or microbial pathogen attack. These specific responses are co-ordinated by defense-related hormones involved in signalling pathways (Whitham *et al*., 2003; Whitham *et al*., 2006; Ascencio-Ibanez *et al*., 2008; Pallas and Garcia, 2011; Ballare, 2007). Upon pathogen attack, induced defences rely on energy resources which are critical to plant fitness. In order to minimise fitness costs and maximise defense responses, plants possess regulatory mechanisms to coordinate pathogen-specific defense responses, which involve signalling molecules that act systemically throughout the plant (Koornneef and Pieterse, 2008). Salicyclic acid (SA), jasmonic acid (JA) and ethylene (ET) are the main signalling pathways responsible for regulating responses to biotic and abiotic stresses. In addition, abscissic acid (ABA), auxins, cytokinins, gibberellins, and brassinosteroids have also been implicated (Koornneef and Pieterse, 2008; Pieterse *et al*., 2009). Once activated, these signalling molecules are responsible for reallocating resources away from plant growth and development towards defense. The specificity of plant defense responses is determined by the quantity, composition, and timing of these signal molecules and varies across plant species. The replication and defense strategy of the pathogen determines which defense-related genes are triggered by the plant (Koornneef and Pieterse, 2008; Pieterse *et al*., 2009). Following pathogen infection, antagonistic or synergistic cross talk between signalling pathways enables the plant to devise optimal resistance strategies in order to minimise fitness costs and activate specific defenses. Generally, SA-mediated defenses are usually induced by biotrophic pathogens, whereas necrotrophic pathogens and herbivorous insects are more sensitive to JA/ET mediated defenses (Koornneef and Pieterse, 2008). Pathogens on the other hand, are also capable of manipulating these signalling networks as well as suppressing induced defenses for their own benefit, resulting in host susceptibility (Pallas and Garcia, 2011; Koornneef and Pieterse, 2008).

South African cassava mosaic virus [ZA:99] (SACMV) infects an important food security crop, cassava (*Manihot esculenta* Crantz), in Sub-Saharan Africa, and causes extensive damage to the crop, resulting in Cassava mosaic disease (CMD) (Berrie *et al*., 2001). SACMV is a member of the genus *Begomovirus*, and belongs to the *Geminiviridae* family, whose members are transmitted by the whitefly, *Bemisia tabaci* (Gennadius) (Harrison and Robinson, 2002; Gorovits *et al*., 2007). Its genome is bipartite, consisting of

a DNA-A and DNA-B segment of 2800 nt and 2760 nt, respectively (Berrie *et al*., 2001). The bipartite-genome of SACMV encodes at least four proteins on the DNA- A: the viral strand contains the coat protein (CP or AV1) and the pre-CP (AV2). The complementary strand contains four proteins; AC1, AC2, AC3 and AC4 from overlapping open reading frames (ORFs). AC1 is required for initiation of DNA replication and is termed the replication-associated protein (Rep), AC2 (TrAP) activates transcription in both the DNA-A and DNA-B of the viral sense genes, AC3 is the DNA replication enhancer (REn), and the function for AC4 is unknown. DNA-B encodes two proteins, namely BC1 and BV1 which are involved in intracellular, intercellular and systemic virus movement. BC1 is found on the complementary strand and mediates cell-to-cell movement of the virus. BV1 is the nuclear shuttle protein (NSP) which controls movement of viral DNA between the nucleus and cytoplasm (Harrison and Robinson, 2002; Gafni and Epel, 2002; Fontes *et al*., 2004). Geminiviruses have been implicated in many host-responsive processes such as transcriptional regulation, DNA replication, control of the cell cycle, cell proliferation and differentiation, and macromolecular trafficking in whole plants (Fontes *et al*., 2004; Mariano *et al*., 2004; Ascencio-Ibanez *et al*., 2008; Jeske, 2009; Mills-Lujan and Deom, 2010). In order to complete infection in a host, geminiviruses need to modify certain host-cell pathways. Such changes include:- modulation of plasmodesmata structure and function, host silencing-related defense mechanisms, interactions with proteins such as NAC-domain (*NAM,ATAF1/ATAF2*, and *CUC2)* containing proteins which are involved in growth and development regulation, host gene expression changes, and retinoblastoma-related (RBR) pathway interference (Gutierrez, 2000; Gutierrez, 2002; Hanley-Bowdoin *et al*., 2004; Jeske, 2009; Mills-Lujan and Deom, 2010).

Global analyses of exceptionally large datasets are emerging from transcriptome, protein-protein interaction and regulatory, developmental and metabolic pathway studies in order to construct networks that systematically categorize function and interaction between molecules or organisms at differing levels of complexities (Ma *et al*., 2007). This rapidly increasing area of systems biology, where networks are formed from underlying signalling and regulatory control, as well as cellular function, is referred to as “interactomics” (Geisler-Lee *et al*., 2007). While deep sequencing and whole-genome tiling assays have recently become more important technologies in plant biology (Busch and Lohmann, 2007), microarrays and qRT-PCR remain accurate and invaluable tools in expression profiling of host-virus interactions. Plant gene-expression networks have been elucidated through microarray technology by identifying global gene expression changes in a host,

infected, in most instances, with positive-sense RNA viruses (Whitham *et al*., 2003; Whitham *et al*., 2006; Babu *et al*, 2008a; Babu *et al*., 2008b; Agudelo-Romero *et al*., 2008). In a study by Postnikova and Nemchinov (2012), a comparative analysis of all published microarray data sets of compatible interactions in *Arabidopsis*, with 11 plant viruses (9 RNA, 1 ds DNA and one ssDNA geminivirus), showed that there was a greater variety of up-regulated genes as compared with repressed genes in the course of viral pathogenesis. Furthermore, each virus-host interaction is unique in terms of altered expression levels, but at the same time, there are some shared genes affected by all viruses. Only one whole genome microarray gene expression study has been conducted on a DNA geminivirus, *Cabbage leaf curl virus* (CaLCuV), at 12 days post infection (dpi) in *Arabidopsis* (Ascencio-Ibanez *et al*., 2008).

The *Arabidopsis* experimental system remains the host of choice due to its adaptable and favourable genetic nature, and is the most thoroughly studied organism providing readily available community resources. This allows for more interdisciplinary and multi-investigative studies to take place (Koornneef and Meinke, 2010). The *Arabidopsis* interactome, in particular, can provide information about conserved genes likely to be involved in the same biological process across species such as humans (*Homo sapiens)*, yeast (*Saccharomyces cerevisiae*), fruit-fly (*Drosophila melanogaster*), and nematode worm (*Caenorhabditis elegans*). In addition, knowledge of signalling pathways and protein complexes has increased existing *Arabidopsis* experimental data by adding previously unknown proteins into existing networks. Based on the predicted *Arabidopsis* interactome, hypothesis-driven data can be added to the current knowledge of signalling and cellular function without the need of a cost-prohibitive, high-throughput experimental approach to validate data (Geisler-Lee *et al*., 2007).

Since annotation of the cassava genome is currently incomplete (www.phytozome.org), and no transcriptome studies have been carried out in cassava (except for a study conducted by Fregene *et al*., 2004, using serial analysis of gene expression (SAGE) of host-plant resistance to Cassava mosaic disease), the model plant system, *Arabidopsis*, was chosen to conduct a susceptibility study with SACMV. A temporal study across 36 days post infection (3 time points) was performed to identify co-regulated defense and stress mechanisms activated by SACMV for establishing infection, and also to identify transient or persistent genes expressed across the course of infection. Global monitoring of gene expression was essential to distinguish if host alterations were

SACMV-specific and/or a general biotic stress response. Results from this study, and correlations with other plant viruses, has provided further insight into the little that is known about geminivirus gene expression changes in compatible hosts. This is the first reported geminivirus gene expression microarray study identifying progressive differential transcription during a compatible time course of infection.

3.3 Materials and Methods

This SACMV-[ZA:99]-[ZA:99] – *Arabidopsis* microarray study is MIAME compliant and has been deposited in the Gene Expression Omnibus (GEO) of NCBI (www.ncbi.nlm.nih.gov/geo/) (Edgar *et al*., 2002). The accession number GSE43282 has been assigned to the project and the data is publicly available.

Agroinfection of Plants and Virus Detection and Copy Number Determination

Arabidopsis thaliana (ecotype Columbia-0) seeds were planted in seed trays containing peat pellets (Jiffy Products International), covered with plastic wrap and placed at 4°C for 1 day to eliminate dormancy and ensure uniform germination. These plants were then transferred to growth chambers (Binder Growth Cabinets) operating at 22°C under a 10 h photoperiod, in a humid environment, at a light intensity of 100 μm^{-2} sec^{-1}. In order to acclimatize the plants, two-to-three cuttings were made in the plastic covering approximately 2 weeks after planting. This procedure was repeated daily for ten days in order to maintain humidity and avoid air flow around the plants. Once acclimatized, the plastic covering was removed and plants were fertilized and watered as required, until ready for virus inoculations.

Eight-week-old *Arabidopsis* plants were co-inoculated with full-length head-to-tail SACMV DNA-A and DNA-B dimers (Berrie *et al*., 2001), mobilized into *Agrobacterium tumefaciens* strain AGL1 according to the improved agroinfection protocol of Pierce *et al*., unpublished. Briefly, five hundred microlitres of *Agrobacterium* cultures (containing SACMV DNA-A and DNA-B) were separately inoculated into 5 ml of LB (containing a final concentration of 100 µg/ml of Carbenicillin and Kanamycin), and incubated at 30°C overnight. Once an OD_{600} of 1.8/2.0 was reached (approximately 18 h), 4 ml of culture was sub-inoculated into 30 ml LB with antibiotics for approximately 24 h. One millimetre of each culture (OD of 1.8/2.0) was spun down and the supernatant removed. Sterile water was

then added, mixed and spun for 1 min. The pellet was then resuspended in 200 µl LB and equal volumes of DNA-A and DNA-B were mixed together. Approximately 100 µl (for a 10 cm high plant) was used to wound the stems by needle puncture, and the inoculum was then injected along the stem, concentrating on the apex. Plants were covered for 2 days and re-acclimatised to adapt to chamber conditions. Healthy control plants were mock-inoculated with AGL1 cultures only. Virus inoculations and harvesting of leaves was done at the same time of day in order to maintain consistency between time points and to minimize variations in gene expression patterns due to abiotic factors.

Total nucleic acid (TNA) was extracted from SACMV-infected and mock-inoculated *Arabidopsis* plants according to the CTAB (cetyltrimethylammonium bromide) method of Doyle and Doyle (1987). Fifty milligram young leaf samples were ground in liquid nitrogen and TNA was extracted by the addition of 0.5 ml pre-heated CTAB extraction buffer (2% CTAB, 20 mM EDTA, 1.4 M NaCl, 100 mM Tris pH 8.0) and ß-mercaptoethanol (to a final concentration of 0.1% v/v). The aqueous layer containing the TNA was extracted using chloroform:isoamyl (24:1) in a two-step process and the nucleic acids precipitated with an equal volume of isopropanol. The pellet was then washed with 70% ice-cold ethanol, vacuum dried and resuspended in 50 µl 1 X TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) containing 20 µg/ml RNase A.

PCR was carried out using BV1 primers that amplify a 168 bp region on SACMV DNA-B genome component. BV1 primers consisted of the following sequences:
BV1 Forward 5'TACGGCATGCCTAGGTTGAAGGAA3', and
BV1 Reverse 5'ATCCACATCCTTGAACGACGACCA3'.
Approximately 1 µg of TNA was added to each reaction consisting of 0.1 volume 10 X Taq buffer ($NHSO_4$), 10 mM dNTPs, 0.04 volumes of 25 mM $MgCl_2$, and 1.25 U Taq DNA Polymerase, Recombinant (Fermentas) of which 10 µM of each primer was added, making up a final reaction volume of 50 µl. Amplification was carried out utilizing the MyCycler™ Thermal Cycler (Bio-Rad) with cycling conditions programmed for 1 cycle at 95°C for 1 min, followed by 30 cycles at 93°C for 30 sec, 60°C for 30 sec, and 72°C for 30 sec, this was followed by a final extension step for 7 min at 72°C.

In order to determine SACMV copy number, absolute quantification was performed. Rolling circle amplification of SACMV DNA-B was carried out using the Illustra™ Templiphi™ 100 Amplification kit (GE Healthcare) according to the manufacturer's

instructions. A standard curve was constructed (in duplicate) using 5 known concentrations of SACMV DNA-B RCA products spiked with 200 ng of healthy *Arabidopsis* TNA. In order to obtain a curve where SACMV DNA-B was present at 100 000, 10 000, 1000, and 10 copies, the following calculations were followed:

1. Calculating mass of a single DNA-B molecule

   $m = (n)(1mole/6{,}023x10^{23}\ molecules\ (bp))(660g/mole)$

   $= (n)(1.096x10^{-21}\ g/bp)$

   Where:

   n = DNA size (bp)

   m = mass

   Avogadro's no. = $6.023x10^{23}$ molecules/ 1 mole

   Average MW of a double-stranded DNA molecule = 660g/mole

2. Calculating the mass of DNA-B required to achieve the copy no. of interest

   Copy no. of interest x mass of single DNA-B molecule = mass of DNA-B required

   Where copy no. = 100 000, 10000, 1000, 100, and 10 virus copies

   Mass of single DNA-B molecule = that obtained from point 1 above.

3. Calculating the concentration of DNA-B required to achieve copy no. of interest

   Mass (g) (step 2) / volume pipetted in each reaction

The cartridge-purified BV1 primer pair (explained in SACMV detection section) was used for absolute quantification real-time PCR. Quantitative PCR was performed using the Maxima® SYBR Green qPCR Master Mix (2x) kit (Fermentas). Three biological replicates and two technical replicates were carried out at each time point. Target samples were prepared in LightCycler capillaries (Roche Applied Science) containing 10 µl of Maxima® SYBR Green qPCR Master Mix (2x) with a final $MgCl_2$ of 2.5 mM, 0.5 mM of each primer, and 2 µl template DNA (200 ng) in a final volume of 20 µl. RCA DNA-B standards were prepared as above with the addition of 200 ng of healthy *Arabidopsis* TNA spiked into each reaction in order for the standards to be homologous to the target samples. Cycling conditions consisted of an activation mode of 95°C for 10 min, followed by 32 amplification cycles run at 95°C for 15 sec, 55°C for 30 sec, and 72°C for 30 sec for a single acquisition (fluorescence detection at 520nm at the end of the elongation phase

for each cycle). A melting curve was then performed by heating to 95°C, cooling to 65°C for 30 sec, and slowly heating to 95°C at 0.1°C/sec with continuous measurement of fluorescence at 520 nm, followed by a final cooling step at 40°C for 10 sec. All quantitative PCR data was analysed using the Roche LightCycler Software Version 4.

VirD2 PCR was carried out in order to detect *A. tumefaciens* AGL1Ti plasmid (TiBo542) presence in healthy and infected *Arabidopsis* leaf samples at 14, 24, and 36 dpi. Primers were designed for the virD2 gene (AF242881) from *A. tumefaciens* AGL1Ti plasmid (TiBo542), containing a C58C1 chromosomal background (Petti *et al*., 2009). This primer pair amplified a 360 bp region of the virD2 gene:
virD2 Forward, 5'GCAGAGCGACCAATCACATA3'
virD2 Reverse, 5'GGCTTCAGCGACATAGGAAG3'
Approximately 1 µg TNA was added to each reaction consisting of 0.1 volume 10 X Taq buffer ($NHSO_4$), 10 mM dNTPs, 0.04 volumes of 25 mM $MgCl_2$, and 1.25 U Taq DNA Polymerase, Recombinant (Fermentas) of which 10 mM of each primer was added, making up a final reaction volume of 50 µl. Amplification was carried out utilizing the MyCycler™ Thermal Cycler (Bio-Rad) with cycling conditions: 1 cycle at 95°C for 4 min, followed by 30 cycles at 95°C for 30 sec; annealing temperatures at of 57°C for 30 sec; an elongation step set at 72°C for 30 sec; followed by a final extension step for 4 min at 72°C.

A standard curve was constructed (in duplicate) using 6 known concentrations of AGL1 Ti plasmid, TiBo542, which is approximately 250 kb in size in order to obtain 100 000, 10 000, 1000, 100, 10, and 1 copy(ies), respectively. In order for standards to be as homologous to the target samples as possible, 200 ng of *Arabidopsis* healthy TNA was spiked into each standard. Calculations were carried out as previously described in SACMV copy number determination section. For quantitative PCR, 3 biological replicates were pooled for healthy and SACMV-infected TNA samples, respectively, at each time point (14, 24, and 36 dpi), and a technical replicate was performed for each biological replicate. Samples were prepared in LightCycler capillaries (Roche Applied Science) containing 10µl of Maxima® SYBR Green qPCR Master Mix (2x) with a final $MgCl_2$ of 2.5 mM, 0.5 mM of each virD2 primer, and 2 µl template DNA (200 ng) in a final volume of 20 µl. Cycling conditions consisted of an activation mode of 95°C for 10 min, followed by 40 amplification cycles run at 95°C for 15 sec, 57°C for 30 sec, and 72°C for 30 sec for a single acquisition (fluorescence detection at 520nm at the end of the elongation phase for each cycle). A melting curve was then performed by heating to 95°C, cooling to 65°C for

30 sec, and slowly heating to 95°C at 0.1°C/s with continuous measurement of fluorescence at 520 nm, followed by a final cooling step at 40°C for 10 sec.

Arabidopsis thaliana (ecotype Columbia-0) seeds were planted in seed trays containing peat pellets (Jiffy Products International), covered with plastic wrap and placed at 4°C for 1 day to eliminate dormancy and ensure uniform germination. These plants were then transferred to growth chambers (Binder Growth Cabinets) operating at 22°C under a 10 h photoperiod in a humid environment at an intensity of 100 μm^{-2} sec^{-1}. In order to acclimatize the plants, two - to - three cuttings were made in the plastic covering approximately 2 weeks after planting. This procedure was repeated daily for ten days in order to maintain humidity and avoid air flow around the plants. Once acclimatized, the plastic covering was removed and plants were fertilized and watered as required, until ready for virus inoculations.

Gene Expression Studies

Extraction, Purification and Quantification of RNA

In order to limit variation in profiling entire organs or tissues, only the rosette leaves closest to the meristem tip, representing cells containing active geminivirus replication) were sampled. Three independent biological replicates and 1 technical replicate (total RNA from biological replicate 1) were carried out. For each biological replicate, total RNA was extracted from pooled SACMV-[ZA:99]-infected and healthy *Arabidopsis* plants at 14, 24, and 36 dpi using a QIAzol lysis reagent modified method originally described by Chomczynski and Sacchi (1987). Uppermost tissue from 2-3 pooled leaves from individual *Arabidopsis* plants in each biological replicate was ground in liquid nitrogen with a mortar and pestle and 1ml of QIAzol (Qiagen) added. Samples were then incubated at 60°C for 5 min followed by centrifugation at 13400 rpm for 10 min at 4°C. The supernatant was then treated with 200 µl of chloroform, vortexed for 15 sec, left at room temperature (RT) for 2-3 min and centrifuged at 13400 rpm at 4°C for 15 min. The aqueous phase was carefully pipetted into a new tube and precipitated by adding isopropanol and 0.8M Sodium Citrate/1.2M NaCl (Sigma), half volume of aqueous phase of each. The tubes were then mixed by gentle inversion and incubated for 10 min at RT, followed by another centrifugation step at 13,400 rpm at 4°C for 10 min. The RNA pellet was washed with 75% ice-cold ethanol, vortexed gently, and centrifuged at 10600 rpm at 4°C for 10 min. The

supernatant was discarded and centrifuged for a further 10600 rpm at 4°C for 2 min. Samples were dried at 37°C for 5-10 min and resuspended in 50 to 100 µl of sterile water (Sabax water for injections, Adcock Ingram),and placed at 55°C for RNA to dissolve. In order to purify the RNA samples, the RNeasy Mini Protocol for RNA cleanup (Qiagen) was performed according to manufacturer's instructions (RNeasy ® Mini Handbook, Qiagen), and 0.5 ul of Ribolock RNAse inhibitor (Fermentas) was added to each 50 ul sample (14 and 24dpi) and 1ul to 100ul for 36dpi samples. Concentration and purity (A_{260}/A_{280} and A_{260}/A_{280} ratios) of the samples after cleanup was assessed on the Thermo Scientific NanoDrop™ 1000 Spectrophotometer. RNA integrity was pre-assessed on a 1% TBE gel (not shown). Stringent RNA quality control was carried out using the Agilent 2100 Bioanalyzer (Eukaryote Total RNA Pico series II chip, version 2.5) (not shown).

DNA Contamination Checking of RNA Samples (Quality Control)

To detect contaminating DNA in the RNA samples, RT-PCR was carried out using primers designed to amplify an exon/intron region from the *Arabidopsis* Ubiquitin gene (AT4G05320). Primer sequences were as follows: - UB Forward 5'ATTTCTCAAAATCTTAAAAACTT3' and UB Reverse 5'TGATAGTTTTCCCAGTCAAC3'. cDNA synthesis was carried out as follows: - Oligo dT primer (0.5 ug/ul) (Invitrogen), 0.5 ul Ribolock RNAse inhibitor (Fermentas) and RNAse free water were added to 1 ug of total RNA (total volume 11.6 ul) and samples heated to 70°C for 10 min and chilled on ice. A 7.8 ul master mix containing 5 X buffer, MgCl2 (2.5 mM), 10 mM dNTPS, and 1 ul ImProm-II™ enzyme (Promega) was added to each reaction and RT was carried out utilizing the MyCycler™ Thermal Cyler (Bio-Rad) consisting of 1 cycle of 25°C for 10 min, 42°C for 60 min, and 70°C for 15 min. PCR using Ubiquitin primers was carried out using 100 ng (5 µl) of Arabidopsis TNA (positive control) and 5ul of RT product, with RNAse free water as a negative control. Reaction mixtures contained 10 X reaction buffer, 10 uM Ubiquitin F and R primer (0.5 uM each final), and 2.5 U Dream taq. Amplification was carried out utilizing the MyCycler™ Thermal Cycler (Bio-Rad) with cycling conditions of DNA denaturation and *Taq* DNA Polymerase activation for 20 secs at 95°C, and then 35 cycles of denaturation for 30 sec at 95°C, annealing for 30 sec at 55°C and extension for 60 sec at 72°C. The amplification products were examined by electrophoresis on a 1% agarose gel stained with ethidium bromide (EtBr) to a final concentration of 10µg/µl in a 1 X TAE electrophoresis buffer containing 50 µg of EtBr run at 75V.

RNA Amplification, Labelling, Microarray Hybridization, and Scanning

Total RNA (1 µg) was amplified using the Amino Allyl MessageAmp™II aRNA Amplification Kit (Ambion) following manufacturer's instructions. Modifications to the protocol included: Adding 2 ul of spike A or B (dilution 3, 1:4 ratio) (Agilent RNA spike-in kit). Spike A labelled with Cy3 and spike B, Cy5, to the first strand cDNA synthesis step, resulting in a 12 ul total volume before incubation. During aRNA:Dye coupling, 4 µg of aRNA was vacuum dried at 45°C and resuspended in 5 µl of 0.2 M NAHCO3 (pH 9.0) at RT for 20 min. Two microlitres of each dye (Cy5 or Cy3) was added, incubating for 2 h at RT. Dye labelled aRNA purification was carried out using the RNEASY MinElute Kit (Qiagen). Dye incorporation (into aRNA) was measured using a NanoDrop 1000 Spectrophotomer. Microarray hybridization was carried out according to manufacturer's instructions (Agilent). One hundred pmol of each cyanine dye, linearly amplified cRNA was added to a hybridization mix containing 10 x blocking agent and 25 x fragmentation buffer were incubated for 30 min at 60°C to fragment the RNA. Fifty five microliters of 2 x GE buffer was then added to the solution, spun gently and placed on ice, ready for hybridization. One hundred and ten microliters of solution was added onto three Agilent 4 X 44 slides containing containing 37,683 *A.thaliana* probes (Version 3), and placed in a rotating hybridization chamber (Agilent) set at 65°C for 18 h. Slides were then washed using Agilent's Gene Expression Wash Buffers 1 and 2. Briefly, hybridization chambers were disassembled in Wash Buffer 1. The microarray slide was then removed and placed into a 50 ml Greiner tube containing Gene Expression Washer Buffer 1 at room temperature for 1 minute. This step was repeated for each slide (3 times). Each slide was then placed into pre-warmed (37°C) Wash Buffer 2 for 1 minute. Slides were then centrifuged briefly in 50 ml Greiner tubes to remove remaining droplets. Scanning was conducted using a GenePix 4000B scanner (Axon Molecular Devices) at 532 nm for Cy3 and 635nm for Cy5. Spots were scanned using 5 µm resolution. Adjustments to photomultiplier tubes were made to balance intensities between each dye and to increase signal-to-noise ratios. GenePix Pro 6.0 (Axon Molecular Devices) software was used to quantify spot intensities

Relative Quantitative Reverse-Transcription PCR (qRT-PCR)-Microarray Validation

cDNA was synthesized from 1 µg of total RNA in a volume of 20 µl using the iScript™ cDNA Synthesis Kit (Bio-Rad) according to the manufacturer's instructions.

Quantitative RT-PCR was carried out using primer sets selected from the primer library for *Arabidopsis* Pathogen inducible genes (Sigma). Primers for three normalization genes were selected from the library which included:

CBP20 Forward 5'TGTTTCGTCCTGTTCTACTC3'

CBP 20 Reverse 5'ACACGAATAGGCCGGTCATC3',

ACTIN2 Forward 5'GCAAGTCATCACGATTGGTGC3'

ACTIN2 Reverse 5'GCAACGACCTTAATCTTCATGCTG3'

UBC Forward 5'TCAAATGGACCGCTCTTATC3'

UBC Reverse 5'CACAGACTGAAGCGTCCAAG3'

A fourth normalization gene namely, EF1-alpha was cartridge purified and synthesized as follows:-

EF1- alpha Forward 5'GGAGATTGAGAAGGAGCCCAAGTTC3'

EF1-alpha Reverse 5'GTGTGTGTAGATCCGCCACCTC3'

Four reference genes were selected in order to determine the expression stability of each gene through Normfinder (Andersen *et al*., 2004). The top-ranked gene would be the resulting gene with the lowest expression value. For time points 14 and 24dpi respectively, 3 biological replicates were carried out for both healthy and SACMV-[ZA:99]-infected cDNA. In addition, a technical replicate was run for each biological replicate. A master mix was prepared for each gene using the Maxima® SYBR Green qPCR Master Mix (2x) kit (Fermentas), with 2 ul of cDNA in a final reaction volume of 20 ul. Two negative controls were prepared which included: - a no-template control to ensure that no primer dimer formation was detected, and a no-RT control was included to ensure that no detectable genomic DNA was present in the sample. Standard curves were prepared at both 14 dpi and 24 dpi by pooling equal amounts of both healthy and SACMV-[ZA:99]-infected cDNA for each time point, respectively. Six dilutions were prepared for each curve containing the following concentrations: 150 ng, 30 ng, 6 ng, 1.2 ng, and 0.24 ng. In order to account for PCR inhibition, 100 pg of of the 18S gene from *N.tabacum* (AY079155.1) was spiked into every sample in order to detect a 139 bp amplicon. 18S primer pairs appeared as follows: -

Forward 5'GGCAAATAGGAGCCAATGAA3' and

Reverse 5'GGGGTGAACCAAAAGCTGTA3'.

Relative quantification real-time RT-PCR reactions were performed on the LightCycler 2.0 System (Roche Applied Science) with thermal cycling conditions consisting of an initial activation step of 95°Cfor 10 min, followed by a cycling step repeated 40 times

consisting of 95°C for 15 sec, 65°C for 30 sec, and 72°C for 30 sec with a single fluorescence measurement. A slight amendment to cycling parameters for the 18S spike-in gene consisted of an annealing temperature of 57°C and 30 cycles, differing slightly to the above-metioned parameters for all other genes tested. A melting curve analysis was then carried out at 95°C for 0 sec, 65°C for 30 sec, and 95°C for 0 sec at a heating rate of 0.1°C per second and a continuous fluorescence measurement. Melting curve analysis was carried out to confirm that the PCR amplicons corresponded to a single cDNA fragment of expected size. A final cooling step was then carried out at 40°C for 10 sec. Crossing Points (CP) were then determined with the LightCycler software version 4.0 (Roche Applied Science). Real-time values were calculated using the relative standard curve method (Applied Biosystems Technical Bulletin). Target quantity (infected leaf material) was determined by interpolating from the standard curve and then dividing by the untreated control (healthy leaf material). Both target quantity and untreated control was normalized to an endogenous control which was determined from the appropriate standard curve. Three biological replicates and two technical replicates were conducted for infected samples and two biological replicates with two technical replicates were performed for healthy, untreated controls. Calculations as follows: Normalized infected sample = target / endogenous control; normalized healthy sample = target / endogenous control; and fold difference in target = normalized target (infected sample) / normalized target (healthy sample).

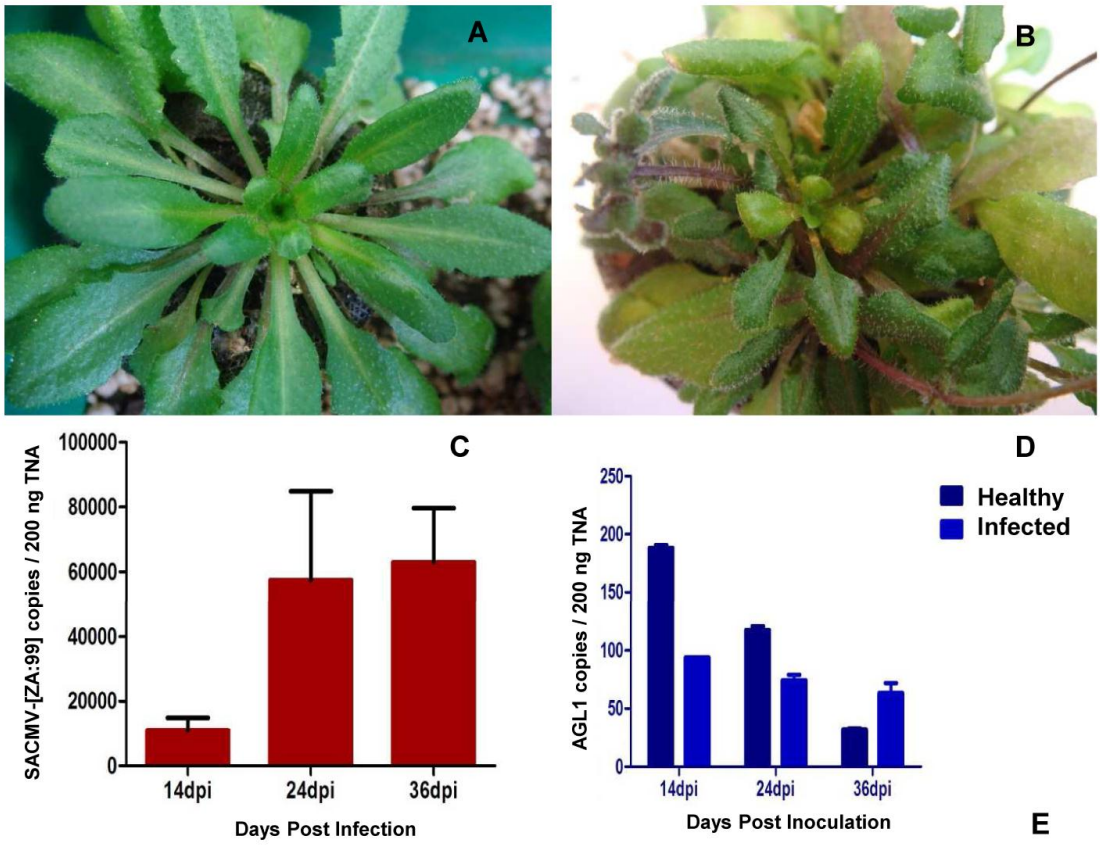
3.4 Results

Arabidopsis Infectivity Assay

Eight-week-old *Arabidopsis* plants were agro-inoculated with SACMV (treatment) and healthy control plants were mock-inoculated with AGL1 cultures to eliminate *Agrobacterium* effects. Symptoms started to appear at 14 dpi and were fully symptomatic at 24 dpi. Overall stunting, slight chlorosis, leaf curl and deformation was observed in infected leaf tissues (Figure 1 B), compared to mock-inoculated controls (Figure 1 A). Viral DNA accumulation was measured in copy number for 3 biological replicates (independent DNA) and mean values obtained at each time point. BV1 primers were designed for quantitative real-time PCR which amplify a 168 bp region on the SACMV DNA-B component. In 200 ng of total nucleic acid, 1.09×10^4 SACMV copies were present

at 14 dpi, 5.75 x 10^4 SACMV copies at 24 dpi, and 6.30 x 10^4 SACMV copies at 36 dpi (Figure 1C). Symptom severity thus correlated with an increase in SACMV copy number.

AGL1, although disarmed, is a pathogen capable of causing gene expression changes in a host (Veena *et al*., 2003). In order to confirm host alterations are a consequence of viral infection and not *Agrobacterium* interference, PCR was performed to detect replicating AGL1 in both healthy (inoculated with AGL1 cultures only) and SACMV - infected leaf tissue. AGL1 levels were measured for each biological replicate at 14, 24, and 36 dpi respectively. Although still detected at each time point (Figure 1 D, E), copy number decreased over time, and was negligible at 36 dpi for both mock-inoculated (32 copies remaining) and SACMV- infected (63 copies remaining) plants. AGL1 mock-inoculated controls in the microarray study were used to eliminate the effects of *Agrobacterium* gene expression.



AGL1 copies / 200 ng TNA in *Arabidopsis* leaf tissue	Days Post Inoculation (Dpi)		
	14	24	36
Healthy (AGL1 only)	189	117	32
SACMV-[ZA:99] - infected	96	75	63

Figure 1. A) Mock-inoculated *Arabidopsis* plants displaying no symptoms (healthy). B) SACMV-[ZA:99]-[ZA:99] – infected leaves displaying leaf curl and deformation. C) SACMV-[ZA:99]-[ZA:99] copy number (copies / 200 ng TNA) over time. Large error bars indicate variability in virus copy number due to biological differences between replicates. D and E) AGL1 detection in 200 ng of TNA from healthy and SACMV-[ZA:99]-[ZA:99] – infected leaf tissue across time points 14, 24, and 36 dpi.

Microarray Gene Expression Analysis in SACMV-[ZA:99] - Infected Arabidopsis

Agilent 4 x 44k *Arabidopsis* gene expression microarray slides were used to establish global profiles of virus-infected plants at 14, 24, and 36 dpi. Labeled cRNA from three biological replicates and 1 technical replicate were analyzed per time point using a direct comparison experimental design. Fluorescence data obtained from the microarray was imported into Limma (linear models for microarray data) (Smyth, 2004) in the R

computing environment, where the data was normalized ('within-array' global loess normalization and 'between-array' quantile normalization), and linear models were fitted in order to contrast SACMV expression values with those of AGL1 mock-inoculated leaf tissue. An output of 13,934 differentially expressed genes was obtained with an adjusted p-value statistic at 0.05 after normalization of data. A total of 1,590 genes were common across the three time points indicated (Figure 2). The number of genes restricted to a particular time point was shown to be 1,456 for 14 dpi, 3 859 for 24 dpi, and 1,570 for 36 dpi indicating unique significant genes at each time point (Figure 2). Gene overlap was highest between 24 and 36 dpi (1,870 corresponding genes), followed by 14 and 24 dpi (1,748 genes showing similarity), with 14 and 36 dpi showing the lowest correlation of 626 genes between the two time points, indicating a large diversion in transcript expression between early and late infection phases. Significantly, maximum levels of gene transcriptional alterations correlated with the peak expression of symptoms, high virus copy number and full systemic virus infection.

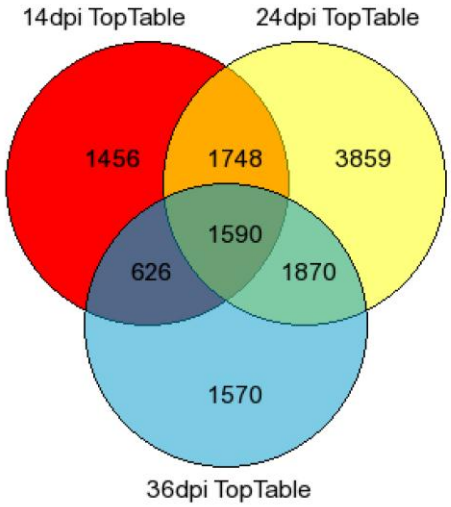


Figure 2. Venn diagram depicting the distribution of 13,934 differentially expressed genes ($p<0.05$) in SACMV-[ZA:99]-[ZA:99] - infected leaf tissue at three time points post infection.

Functional Categorization of 2-Fold Induced and Repressed Genes Across 3 Time Points

A 2-fold cut-off ($p < 0.05$) was then applied to the data resulting in a total of 1,743 highly significant differentially expressed genes (Table S1). The fold change expression data was then assigned to a functional category according to the *Arabidopsis* MIPS (Munich Information Centre for Protein Sequence) functional classification scheme (Figure

3). At each time point, MIPS identified the following number of transcripts: - 203 induced and 194 repressed at 14 dpi, 323 induced and 369 repressed at 24 dpi, and 275 induced and 701 repressed for 36 dpi. Based on Fisher's exact test (Fisher, 1970), putative functions for 24 functional categories were established with the majority of differentially regulated transcripts ($p<0.05$) associated with metabolism, cell cycle and DNA processing, transcription, protein fate (folding, modification, destination), protein binding with binding function or cofactor requirement, cellular transport, transport facilities and transport routes, cellular communication/signal transduction, cell rescue, defense, and virulence, interaction with the environment, systemic interaction with the environment, and sub-cellular localization (Figure 3).

Changes in GO Functional Category Expression Patterns Over The Infection Period

Examination of the patterns of transcript fold changes in GO functional categories (FCs) (Figure 3) over the infection period revealed some interesting results. For the over-represented FCs such as metabolism (1); transcription (11); protein fate (folding, modification, destination); protein binding (16); cellular transport (20); signal transduction and cell communication (30); defense and cell rescue (32); interaction with the environment; abiotic stress (34 and 36); biogenesis of cellular components (42); and subcellular localization (70) (Figure 3), the trend for each FC was a significant increase ($p<0.05$) in the total number of differentially regulated (DE) (repressed and induced) genes from onset of symptoms (14 dpi) to 24 dpi and 24 to 36 dpi (establishment of fully systemic symptoms). Of these DE transcripts, notably the percentage of repressed genes compared to total number of altered genes in each FC also increased as infection progressed. Several RNA plant virus studies [3,4] have indicated that in compatible interactions suppression of host transcription defense responses is a pre-requisite for infection, and this study supports previous findings. Additionally, repression of many host-responsive genes at the later stages of pathogenesis when the geminivirus has successfully established systemic infection, may indicate senescence-related responses, and this trend has also been demonstrated in several plant virus-host interactions in *Arabidopsis* [34]. Interestingly, the pattern of change in up-regulated genes in each FC was not as consistent compared with gene down-regulation. A large number of FCs showed that the percentage of induced genes increased from 14 to 24 dpi, and then remained constant or declined in the later stages (36 dpi) of pathogenesis. The GO FCs for cell cycle and DNA processing, transcription, protein binding and biogenesis of cell

components, all showed a significant ($p<0.05$) increase from 14 to 24 dpi, and this is not surprising since all of these functions would need to be induced in order for SACMV to replicate and move systemically during these early to middle stages of acute infection. Defense and cell rescue related transcripts, representing ~12% of all 2-fold or more differentially expressed genes, while also showing an overall increase in percentage of repressed transcripts across the infection period, interestingly had a steady continuous expression of up-regulated genes (12-16%) over 36 days and did not change significantly. The total number of up-regulated stress/abiotic-related genes (FCs 34 and 36: interaction with the environment; figure 3) declined over the 36 day infection period.

Identification of 2-Fold Induced and Repressed Genes Over 3 Time Points

Once functional categories were established, genes that were continuously expressed across all three time points were identified (Table 1) and a gene tree heat map (Figure 4) was constructed by applying hierarchical clustering using a Euclidean distance metric and average linkage clustering. A total of 41 genes were found to be continuously expressed across time points, 10 showing up-regulation (24.39%), 23 down-regulation (56.10%), 2 down-regulated at 14 dpi then up-regulated at 24 and 36 dpi (4.88%), 4 up-regulated at 14 and 24 dpi, then down-regulated at 36 dpi (9.76%), and 2 up-regulated at 14 dpi then down-regulated at 24 and 36 dpi (4.88%). In addition, we selected the top 20 genes (10 up-regulated and 10 down-regulated) displaying the highest and lowest expression values at each time point to identify which host genes are most reactive to SACMV infection (Table 2). Many transcripts appearing in Table 1 and Figure 4 illustrated that not only were they continuously expressed across time points, but they also appeared in the data listed to have the most highly expressed transcripts (Table 2). Differentially expressed genes were shown to be primarily involved in stress and defense responses as observed with down-regulation of HSP's (Table S2) and up-regulation of defensins, up-regulation and repression of phytohormone signalling pathways, and induction of genes involved in incompatible reactions, transcription, oxidation-reduction responses and other metabolic processes. An interesting trend observed was the redirection of up-regulated genes, at 14 dpi, that represent many phytohormone signalling responses and related defense responses, towards a large number of induced genes involved in metabolic processes such as oxidation-reduction, transport, and cell-wall modification at 24 and 36 dpi. (Figure 1C, 3, and 4, Tables 1 and 2).

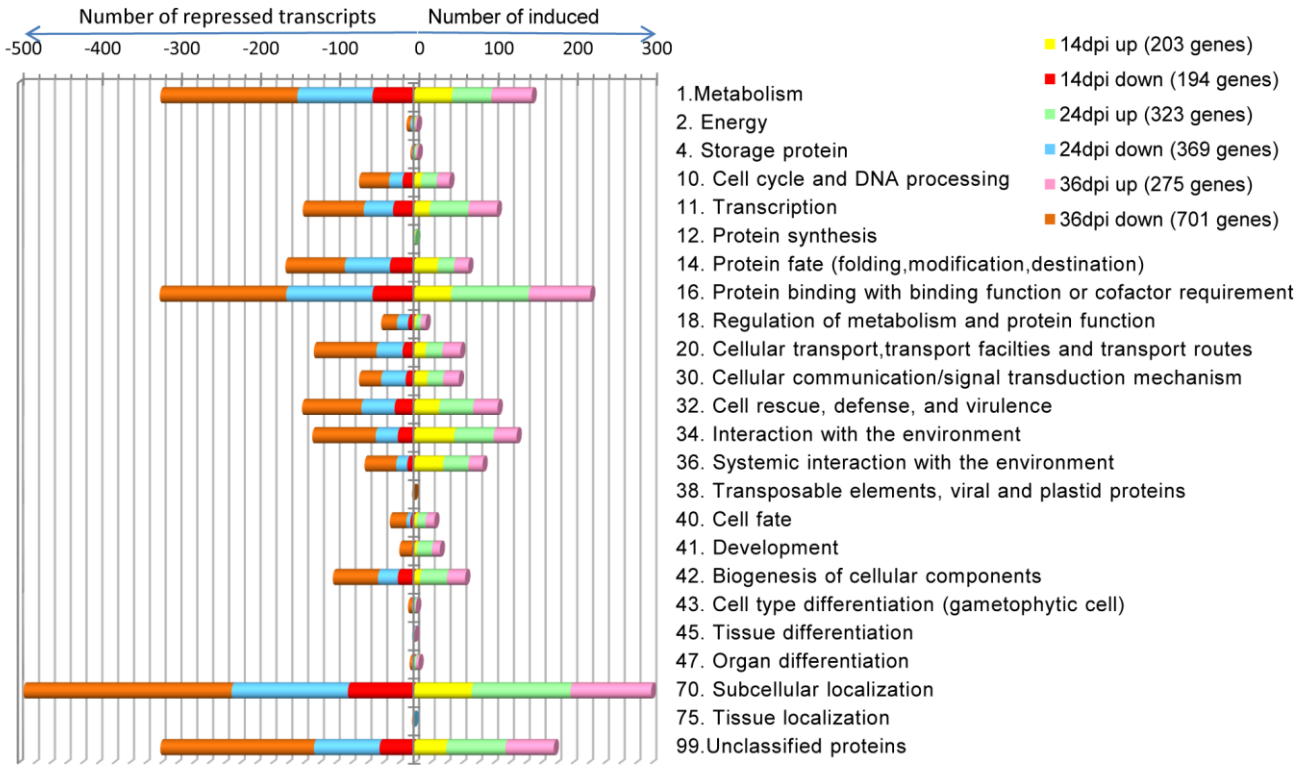


Figure 3. MIPS functional distribution categories of 2-fold differentially expressed transcripts in SACMV-[ZA:99]-[ZA:99] - infected *Arabidopsis* leaf tissues at 14, 24 and 36 dpi.

Table 1. Fold change and adjusted P-values ($p < 0.05$) for 37 transcripts continuously expressed at a 2-fold cut-off across 3 time points post infection (14, 24, and 36 dpi). The four unknown proteins were excluded from the table.

ATG ID	Description	14dpi Fold Change	14dpi Adjusted P-value	24dpi Fold Change	24dpi Adjusted P-value	36dpi Fold Change	36dpi Adjusted P-value
AT5G44430	PDF1.2c (plant defensin 1.2c)	15.82	2.40E-09	10.94	5.80E-202	8.48	4.39E-65
AT2G26020	Arabidopsis thaliana PDF1.2b (plant defensin 1.2b)	14.42	2.40E-09	11.60	2.90E-254	7.91	4.23E-73
AT2G26010	PDF1.3 (plant defensin 1.3)	9.47	2.40E-09	7.25	2.30E-138	6.14	5.48E-47
AT5G07610	F-box family protein	7.48	2.40E-09	7.65	2.40E-175	5.17	1.41E-38
AT2G18193	AAA-type ATPase family protein	4.25	1.14E-08	4.57	9.36E-98	3.51	4.47E-27
AT4G12490	Protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	3.89	2.48E-08	2.32	1.60E-30	2.00	9.44E-09
AT1G52040	MBP1 (MYROSINASE-BINDING PROTEIN 1)	3.07	3.46E-07	2.04	4.82E-22	2.14	2.45E-10
AT4G38860	Auxin-responsive protein, putative	2.98	1.43E-07	3.06	5.39E-53	2.45	4.40E-14
AT4G19700	Protein binding / zinc ion binding	2.66	4.39E-07	2.61	2.89E-39	-2.73	1.62E-17
AT4G32280	IAA29 (indoleacetic acid-induced protein 29)	2.32	1.47E-06	3.27	1.68E-59	2.57	1.60E-15
AT1G61120	Terpene synthase/cyclase family protein	2.22	2.76E-06	-2.28	3.52E-29	-2.86	3.83E-19
AT5G13220	JAS1/JAZ10/TIFY9 (JASMONATE-ZIM-DOMAIN PROTEIN 10)	2.04	2.39E-06	-3.02	7.18E-52	-4.98	6.83E-37
AT5G07580	DNA binding / transcription factor	2.00	1.30E-05	2.24	7.83E-28	-2.26	9.12E-12
AT1G09950	Transcription factor-related	-2.00	3.52E-06	-2.56	1.08E-37	-2.76	6.23E-18
AT5G57015	CKL12 (Casein Kinase I-like 12)	-2.06	2.54E-06	-3.06	5.20E-53	-2.27	5.85E-12
AT1G72210	Basic helix-loop-helix (bHLH) family protein (bHLH096)	-2.09	2.09E-06	-2.08	3.32E-23	-2.42	1.09E-13
AT4G38420	SKS9 (SKU5 Similar 9)	-2.16	1.35E-06	-3.62	5.43E-70	-2.90	1.31E-19
AT5G48070	ATXTH20 (XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE 20)	-2.17	1.63E-06	-2.36	1.48E-31	-2.35	6.23E-13
AT3G02140	TMAC2 (TWO OR MORE ABRES-CONTAINING GENE 2)	-2.19	1.36E-06	-3.90	1.75E-78	-5.99	9.49E-55
AT4G25810	XTR6 (XYLOGLUCAN ENDOTRANSGLYCOSYLASE 6)	-2.20	1.48E-06	-2.03	1.29E-21	-2.73	1.42E-17
AT1G22570	Proton-dependent oligopeptide transport (POT) family protein	-2.26	9.53E-07	-2.62	1.85E-39	-2.22	2.38E-11
AT1G65390	ATPP2-A5 (ATPP2-A5)	-2.36	6.08E-07	-2.35	2.87E-31	-2.29	3.87E-12
AT4G36000	Pathogenesis-related thaumatin family protein	-2.41	7.73E-07	-2.04	5.37E-22	-3.28	2.37E-24
AT5G52020	AP2 domain-containing protein	-2.51	1.32E-06	-5.17	4.00E-114	-3.50	4.96E-27
AT5G45340	CYP707A3 (cytochrome P450, family 707, subfamily A, polypeptide 3)	-2.54	5.57E-07	2.33	1.02E-30	2.36	5.46E-13
AT5G52050	MATE efflux protein-related	-2.60	3.86E-07	-5.16	8.80E-114	-2.08	3.50E-08
AT4G25470	CBF2 (FREEZING TOLERANCE QTL 4)	-2.66	2.03E-07	2.54	3.72E-37	2.92	6.92E-20
AT5G63770	ATDGK2 (DIACYLGLYCEROL KINASE 2)	-3.02	3.06E-07	-4.70	3.80E-101	-2.44	6.02E-14
AT1G74310	ATHSP101 (HEAT SHOCK PROTEIN 101)	-3.09	6.34E-08	-5.18	1.90E-114	-5.47	2.14E-49
AT1G65310	ATXTH17 (XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE 17)	-3.15	6.22E-08	-2.49	1.05E-35	-2.36	4.80E-13
AT2G21510	DNAJ heat shock N-terminal domain-containing protein	-3.30	4.39E-08	-3.58	8.87E-69	-2.23	2.02E-11
AT2G20350	AP2 domain-containing transcription factor, putative	-3.58	3.49E-08	-5.10	4.00E-112	-3.20	2.13E-23
AT1G63030	DDF2 (DWARF AND DELAYED FLOWERING 2) (DDF2)	-3.88	1.63E-08	-3.85	8.35E-77	-2.38	2.78E-13
AT1G54050	17.4 kDa class III heat shock protein (HSP17.4-CIII)	-4.11	1.24E-08	-2.66	8.12E-41	-2.93	5.60E-20
AT1G07400	17.8 kDa class I heat shock protein (HSP17.8-CI)	-5.57	3.11E-09	-4.17	2.22E-86	-5.37	2.09E-40
AT1G22810	AP2 domain-containing transcription factor, putative	-5.96	3.83E-09	-8.59	8.50E-196	-2.74	1.12E-17
AT1G59860	17.6 kDa class I heat shock protein (HSP17.6A-CI)	-6.41	2.40E-09	-6.77	9.00E-155	-3.71	1.13E-24

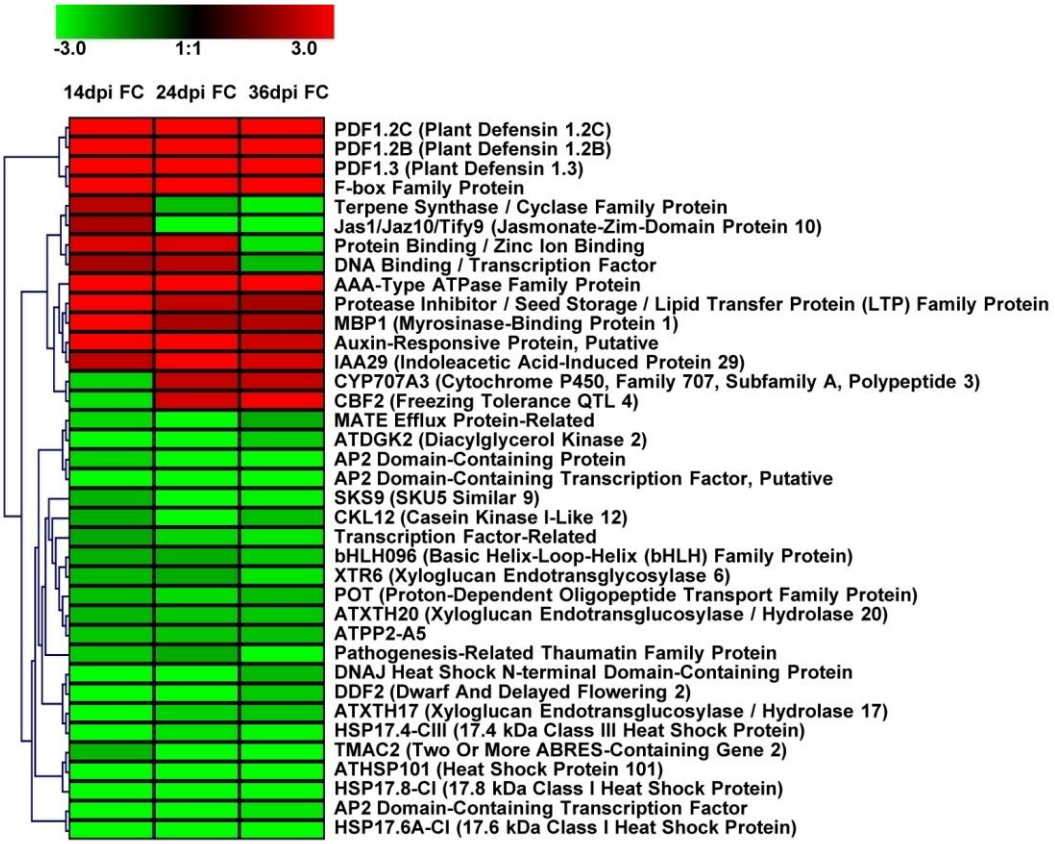


Figure 4. Gene tree heat map showing hierarchical clustering of 37 out of 41 transcripts expressed continuously across time points 14, 24, and 36 dpi (4 unknowns were not displayed). Red bars indicated induction (>2.0) and green bars, repression (<-2.0). Abbreviations: FC (lFold Change).

Table 2. Fold change and adjusted P-values ($p < 0.05$) representing the most significantly induced and repressed (10 up- and 10 down-regulated) A*rabidopsis* genes at 14, 24 and 36 dpi.

***Arabidopsis* acc no.**	**Description**	**Fold Change**	**Adjusted P-Value**
14 dpi			
AT5G44430	PDF1.2c (plant defensin 1.2c) (PDF1.2c)	15.82	2.40E-09
AT2G26020	PDF1.2b (plant defensin 1.2b)	14.42	2.40E-09
AT5G44420	PDF1.2 (Low-molecular-weight cysteine-rich 77)	13.59	2.88E-09
AT2G26010	PDF1.3 (plant defensin 1.3)	9.47	2.40E-09
AT5G07610	F-box family protein	7.48	2.40E-09
AT5G24780	VSP1 (VEGETATIVE STORAGE PROTEIN 1); acid phosphatase (VSP1)	4.78	2.05E-08
AT4G38840	Auxin-responsive protein, putative	4.68	5.26E-09
AT4G25110	ATMC2 (METACASPASE 2)	4.60	6.34E-09
AT1G52400	BGL1 (BETA-GLUCOSIDASE HOMOLOG 1); hydrolase, hydrolyzing O-glycosyl compounds (BGL1)	4.57	4.13E-08
AT2G39030	GCN5-related N-acetyltransferase (GNAT) family protein	4.37	3.49E-08
AT5G13700	APAO/ATPAO1 (POLYAMINE OXIDASE 1); FAD binding / polyamine oxidase (APAO/ATPAO1)	-4.30	1.69E-08
AT4G30280	ATXTH18/XTH18 (XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE 18)	-4.53	1.24E-08
AT3G15210	ATERF-4,Ethylene responsive binding factor 4 DNA binding/ protein binding/transcription factor/transcriptional repressor	-4.64	1.31E-08
AT2G29370	Tropinone reductase, putative / tropine dehydrogenase, putative	-5.36	3.83E-09
AT2G20630	Protein phosphatase 2C, putative / PP2C, putative	-5.37	3.23E-09
AT1G07400	17.8 kDa class I heat shock protein (HSP17.8-CI)	-5.57	3.11E-09
AT3G27540	Glycosyl transferase family 17 protein	-5.65	2.88E-09
AT1G22810	AP2 domain-containing transcription factor, putative	-5.96	3.83E-09
AT1G59860	17.6 kDa class I heat shock protein (HSP17.6A-CI)	-6.41	2.40E-09
AT5G10100	Trehalose-6-phosphate phosphatase, putative	-7.86	2.40E-09
24 dpi			
AT5G45890	SAG12 (SENESCENCE-ASSOCIATED GENE 12); cysteine-type peptidase (SAG12)	13.16	5.01E-281
AT2G26020	PDF1.2b (plant defensin 1.2b)	11.60	2.86E-254
AT5G44430	PDF1.2c (plant defensin 1.2c)	10.94	5.81E-202
AT3G49340	Cysteine proteinase, putative (AT3G49340)	9.42	4.21E-213
AT5G07610	F-box family protein (AT5G07610)	7.65	2.41E-175
AT2G26010	PDF1.3 (plant defensin 1.3) (PDF1.3)	7.25	2.33E-138
AT4G37990	ELI3-2 (ELICITOR-ACTIVATED GENE 3)	5.56	2.31E-124
AT3G44550	Oxidoreductase, acting on the CH-CH group of donors	5.25	3.12E-116
AT2G18193	AAA-type ATPase family protein	4.57	9.36E-98
AT5G44050	ATGEX1/GEX1 (GAMETE EXPRESSED PROTEIN1)	4.50	9.21E-96
AT2G20350	AP2 domain-containing transcription factor, putative	-5.10	4.03E-112
AT5G52050	MATE efflux protein-related (AT5G52050)	-5.16	8.78E-114
AT5G52020	AP2 domain-containing protein	-5.17	3.97E-114
AT1G74310	ATHSP101 (HEAT SHOCK PROTEIN 101); ATP binding / ATPase	-5.18	1.93E-114
AT2G17660	Nitrate-responsive NOI protein, putative (AT2G17660)	-5.25	3.12E-116
AT2G26150	ATHSFA2 (Arabidopsis thaliana heat shock transcription factor A2)	-5.33	2.13E-118
AT1G59860	17.6 kDa class I heat shock protein (HSP17.6A-CI)	-6.77	9.03E-155
AT5G37940	NADP-dependent oxidoreductase, putative	-6.83	2.78E-156
AT1G22810	AP2 domain-containing transcription factor, putative	-8.59	8.46E-196
AT5G37970	S-adenosyl-L-methionine:carboxyl methyltransferase family protein	-10.16	1.36E-227
36 dpi			
AT5G44430	PDF1.2c (plant defensin 1.2c)	8.48	4.39E-65
AT2G26020	PDF1.2b (plant defensin 1.2b)	7.91	4.23E-73
AT2G26010	PDF1.3 (plant defensin 1.3)	6.14	5.48E-47
AT1G31690	Copper ion binding	5.38	1.84E-48
AT1G72920	Disease resistance protein (TIR-NBS class), putative	5.25	4.54E-47
AT5G07610	F-box family protein	5.17	1.41E-38
AT5G21960	AP2 domain-containing transcription factor, putative	4.90	2.84E-43

AT2G40610	ATEXPA8 (ARABIDOPSIS THALIANA EXPANSIN A8)	4.55	2.00E-39
AT2G43590	Chitinase, putative	4.00	5.26E-33
AT2G41180	SigA-binding protein-related	3.74	7.49E-30
AT5G22490	Condensation domain-containing protein	-7.85	1.39E-72
AT1G61820	BGLU46; hydrolase, hydrolyzing O-glycosyl compounds	-8.13	5.34E-75
AT2G38240	Oxidoreductase, 2OG-Fe(II) oxygenase family protein	-8.19	5.36E-63
AT1G43160	RAP2.6 (related to AP2 6); DNA binding / transcription factor	-8.21	1.21E-75
AT3G27170	CLC-B (chloride channel protein B); anion channel/ voltage-gated chloride channel	-8.28	3.15E-76
AT4G12400	Stress-inducible protein, putative	-8.56	1.25E-78
AT5G01380	Transcription factor	-8.68	1.41E-79
AT3G02550	LBD41 (LOB DOMAIN-CONTAINING PROTEIN 41)	-9.73	3.34E-88
AT5G63450	CYP94B1 (cytochrome P450, family 94, subfamily B, polypeptide 1); oxygen binding	-12.27	2.40E-107
AT3G56700	Male sterility protein, putative	-14.30	9.11E-121

Quantitative Reverse-Transcription PCR (qRT-PCR) (Microarray Validation)

Since the greatest differences in fold-change occurred between 14 and 24 dpi, and 24 dpi was our most significant time point in terms of altered gene expression, we chose to validate expression values obtained from microarray data with relative quantification real-time PCR at these time points (Figure 5). At 14 dpi, 3 up-regulated genes, namely BGL2 (AT3G57260), Ankyrin repeat family protein (AT4G03450), and BG3 (AT3G57240), and two down-regulated genes, Transcription factor family (TCP) (AT2G45680) and Ethylene response factor 4 DNA binding /transcriptional repressor (ERF4)(AT3G15210) confirmed expression results obtained from microarray data. Induced genes such as PR4 (AT3G04720) and Glycosyl hydrolase family 17 protein (AT4G16260) and repressed genes such as an Unknown protein (AT2G32200) and AtRABH1c (AT4G39890) showed similarities to microarray data at 24 dpi. In addition, the plant defensin (PDF1.2c) gene was tested at both 14 and 24 dpi, showing similarities in up-regulation to the microarray data. While fold-change patterns correlated, discrepancies in magnitude between the two platforms is not uncommon, and could be attributed to the differences in normalization methods used, where the use of endogenous controls such as CBP20 at 14 dpi and Actin2 at 24 dpi was carried out for normalization of qRT-PCR data, whereas a global normalization was applied to the microarray data. In addition, cDNA was used for qRT-PCR whereas cRNA was used for microarray analysis, suggesting a more efficient fold-change detection method to changes in gene expression for microarray experiments. All qRT-PCR analyses involved 3 biological replicates for SACMV - infected cDNA and 2 biological replicates for AGL1 mock-inoculated controls.

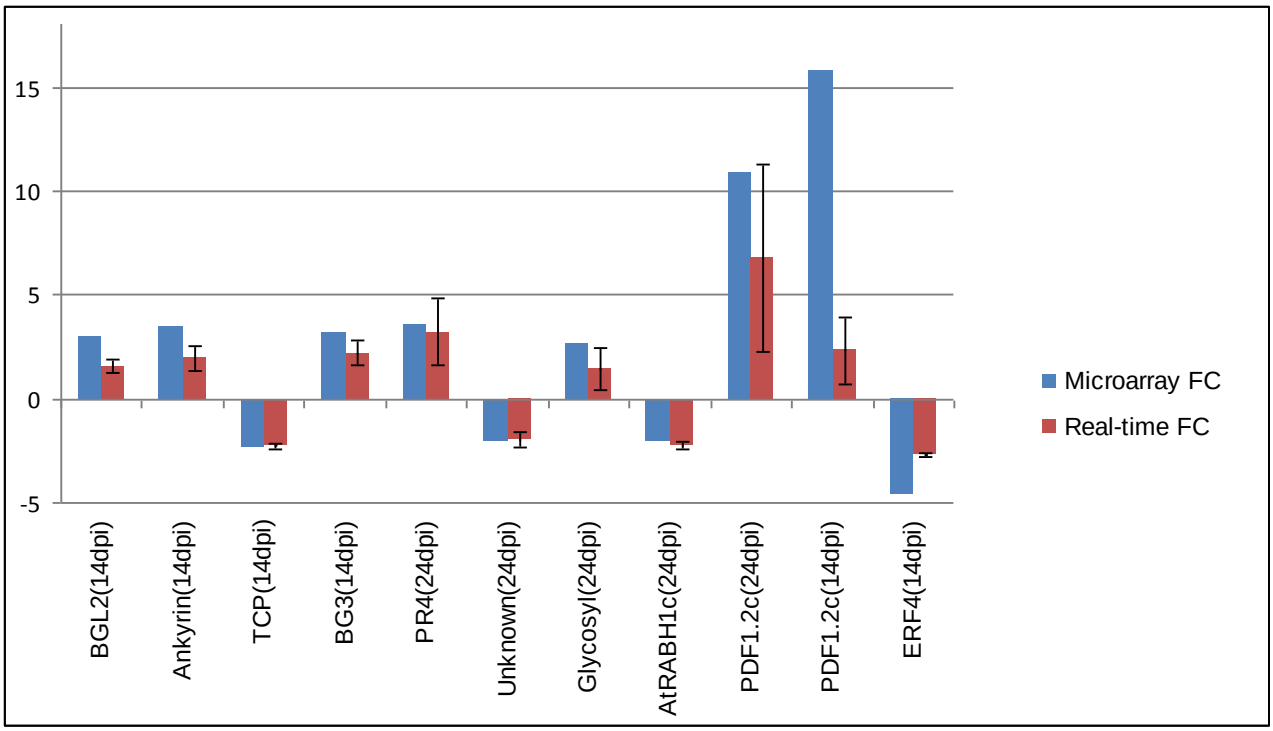


Figure 5. Validation of microarray expression data with relative quantitative real-time RT-PCR (qRT-PCR). Expression changes of 10 selected transcripts (excluding 1 unknown) depicting similarities in expression patterns between the two technologies are shown. Signal intensities for each transcript were normalized with CBP20 for 14 dpi and Actin2 for 24 dpi. The x-axis represents validated genes at time points 14 and 24 dpi. The y-axis represents normalized fold-change expression values for each transcript. The error bars show standard deviation from 3 biological replicates

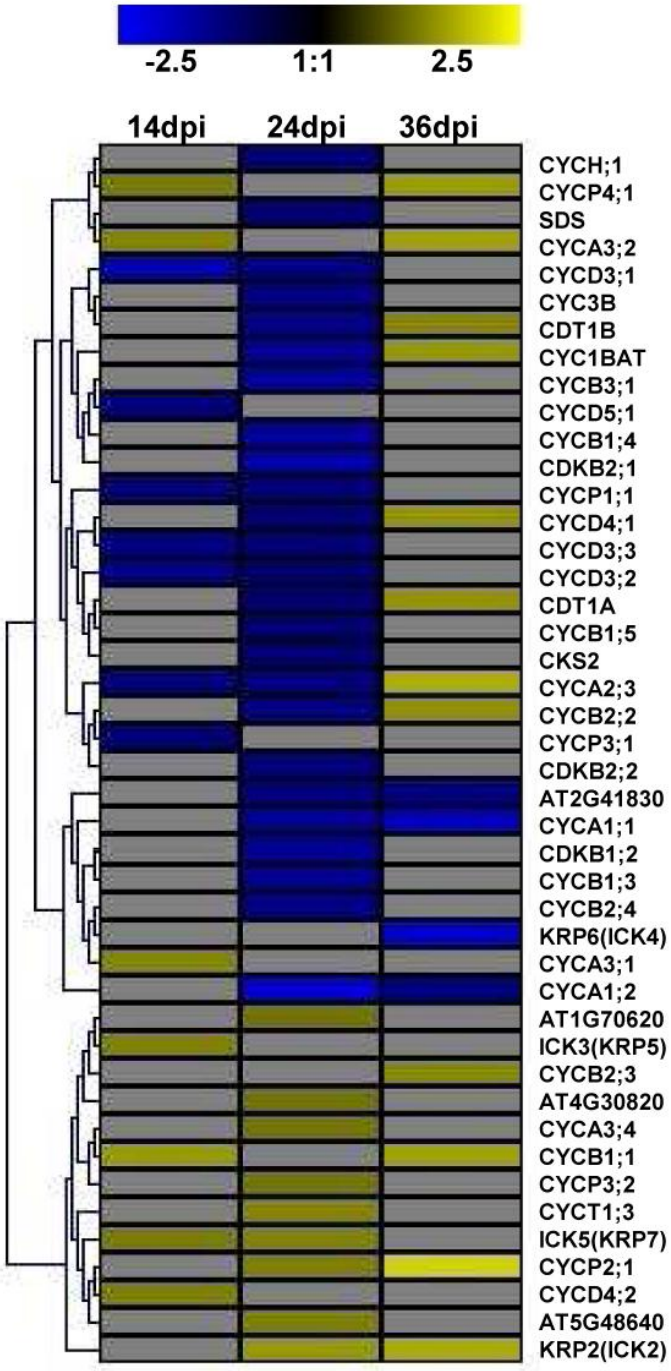


Figure 6. Gene tree heat map of differentially expressed core-cyclin genes in response to SACMV-[ZA:99]-[ZA:99] infection. All listed *Arabidopsis* accession numbers refer to cyclin-related genes.

Table 3. Identification of SACMV-[ZA:99]-[ZA:99]-induced differentially expressed *Arabidopsis* host genes ($p < 0.05$) showing similarities to *Tomato yellow leaf curl Sardinia virus* (TYLCSV) virus in *N. benthamiana* (Lozano-Durán *et al*, 2011).

***Arabidopsis* Acc no.**	**SACMV-[ZA:99]-[ZA:99] Fold Change 14 dpi**	**SACMV-[ZA:99]-[ZA:99] Fold Change 24 dpi**	**SACMV-[ZA:99]-[ZA:99] Fold Change 36 dpi**	**Identity**	**Function**	**Selection criteria for TYLCV**
				Group A: no detected infection effects		
AT2G02560		-1.47		Cullin-associated and neddylation-dissociated (CAND1)	Protein metabolism	TrAP/C2 interaction
AT1G67630		-1.5		DNA polymerase alpha 2 (POLA2)	DNA metabolism	Cellular process
AT5G22220		1.19		E2F transcription factor 1 (E2FB)	Transcription	Cellular process
AT1G21920	1.61	2.38	1.66	Histone 3 K4-specific methyltransferase SET7/9	Unknown	TrAP/C2 interaction
AT3G44110		-1.19		Homologue to co-chaperone DNAJ-like protein (ATJ3)	Protein folding	C3 interaction
AT3G25560	-1.31			NSP interacting kinase 2 (NIK2)	Signal transduction	Phloem over-expression
AT5G03150		-1.47		Putative nucleic acid binding/transcription factor (JDK)	Unknown	TrAP/C2 interaction
AT1G01720		-2.16		Putative transcriptional activators with NAC domain (ATAF1)	Transcription	C3 interaction
AT4G17230	-1.56	-2.41		Scarecrow-like protein (SCL13)	Transcription	Phloem over-expression
AT5G50580		1.24		SUMO activating enzyme (SAE1B)	Protein metabolism	Cellular process
AT4G24440	1.35	1.26		Transcription factor IIA gamma chain (TFIIA-S)	Transcription	Phloem over-expression
AT1G19660	-1.41	1.15		Wound inducive gene (F14P1.1)	Stress	C4 interaction
				Group B: early infection promoted		
AT1G09270			-1.41	Importin alpha isoform 4 (IMPA-4)	Transport	CP interaction
AT1G15380		1.66		Lactoylglutathione lyase (GLO1)	Stress	C3 Interaction
AT1G47128	1.2	1.55		Dehydration responsive 21 (RD21)	Stress	V2 interaction
AT5G22000	-1.21			RING-type E3 ubiquitin ligase (RHF2A)	Protein modification	Transactived by TrAP/C2
AT2G30110	-1.21			Ubiquitin activating enzyme (UBA1)	Protein modification	TrAP/C2 Interaction
				Group C: infection delayed, reduced or prevented		
AT1G51680	1.21	1.4	-2.5	4-coumarate:CoA ligase (AT4CL1)	Metabolism	Phloem over-expression
AT3G25760	1.28	-1.9	-3.05	Allene oxide cyclase (AOC1)	Metabolism	Phloem over-expression
AT5G61430	1.36	1.61		Geminivirus Rep A-binding (GRAB2)	Transcription	Rep interaction
AT2G26560	1.61			Patatin-like protein 2 (PLP2)	Stress	Phloem over-expression
AT1G09840		-1.16		Shaggy-related kinase kappa (SK4-1/SKK)	Protein modification	C4 interaction
AT5G08590		1.19		SKP1-like 2 (ASK2)	Protein modification	Transactived by TrAP/C2

3.5 Discussion

Symptom Development and Virus Accumulation in SACMV-Infected Arabidopsis

Arabidopsis plants were observed to be fully symptomatic at 24 dpi, although symptoms started appearing at 12-14 dpi. Symptoms such as stunting of the entire plant, leaf reduction and deformation were observed in all SACMV - infected *Arabidopsis*, while additionally, chlorosis was observed in approximately 60% of infected plants (Figure 1B). SACMV was detected in all infected plants tested. Chlorotic symptoms may be the direct result of the plants attempt to rescue resources from infected tissues via basal resistance mechanisms. If chlorosis is absent in infected tissues, this usually indicates a loss of basal resistance (O'Donnell *et al*., 2003), and the appearance of mild chlorosis in the majority (60%) of SACMV – infected Arabidopsis leaves suggests a down-regulation of innate basal resistance leading to expected susceptibility to the virus. An increase in SACMV replication was observed between time points 14 and 24 dpi showing a 5-fold increase. Between 14 and 36 dpi, a 6-fold increase was observed (Figure 1C), confirming that an increase in viral titre correlated with symptom development. These findings were also observed in studies conducted by Babu *et al*., 2008b, in soybean [*Glycine max* (L.) Merr] plants infected with *Soybean mosaic virus* (SMV) whereby at 14 dpi virus titer was approximately 2-fold higher than 7 dpi as detected by Northern hybridizations. Similarly, in a gene expression study conducted by Golem and Culver (2003), a greater fold-change increase was also observed in *Tobacco mosaic virus* (TMV) response genes in *Arabidopsis* Shahdara from 4 dpi to 14 dpi, suggesting that higher levels of TMV were present at a later infection time point.

Previous studies have suggested that *Agrobacterium*, although containing a disarmed plasmid, is able to cause changes in host gene expression but at very early stages of infection. These occur between 3-6 h and 30-36 h after initiation of infection (Veena *et al*., 2003). In order to eliminate the effects of *Agrobacterium* in microarray experiments, *Agrobacterium* mock-inoculated controls are commonly used. In this study, qPCR was conducted on AGL1 mock-inoculated control and SACMV-infected plants to rule out the possibility that *Agrobacterium* was persistently replicating in *Arabidopsis* leaf tissues, consequently causing changes in gene expression. qPCR results showed minimal detectable AGL1 copies, showing a decline from 189 copies (14 dpi) to 32 copies (36 dpi)

in mock-inoculated leaf tissue and 96 copies (14 dpi) to 63 (36 dpi) in SACMV- infected leaf tissue (Figure 1D,E). Although *Agrobacterium* AGL1 was still detected by PCR, copy numbers were too low to be considered significant, and most likely represent initial replication following the agroinoculation procedure. Additionally, host gene expression changes in *Arabidopsis* are identified by normalization against mock-inoculated controls, ensuring that alterations are solely due to SACMV.

Differentially Expressed Transcript Data

Gene expression non-filtered data revealed 13,934 significant ($p < 0.05$) differentially expressed genes (including up- and down-regulated transcripts) in response to SACMV infection at three different time points (14, 24, and 36 dpi). Individual gene transcripts were identified at a particular time point and overlap of genes between time points was also observed (Figure 2). Genes expressed transiently at a particular time point may indicate either induction or repression for a specific function or to conserve energy resources in the host (Cheong *et al*., 2002; Koornneef and Pieterse, 2008; Pieterse *et al*., 2009; Baena-Gonzalez, 2010). Those transcripts that appear to show persistent expression (across two or more time points) may be necessary to carry out appropriate function such as stress and defense-like responses for basal resistance to counteract virus attack or alternatively may be induced or repressed by SACMV to aid in its own replication, cell-to-cell spread and systemic movement, as implicated in other studies (Whitham *et al*., 2003; Ascencio-Ibanez *et al*., 2008).

As a first step toward assigning differentially expressed genes to function, the distribution of *Arabidopsis* genes significantly induced or repressed at a 2-fold cut-off in SACMV infected *Arabidopsis* leaves were assigned according to the MIPS (http://mips.gsf.de/proj/thal/db/Arabidopsis) classification scheme. For the purpose of this study we refer to early response genes as 14 dpi (initiation of symptoms), to 24 dpi as fully symptomatic, middle-phase genes, and to 36 dpi as late response genes. A general overview of 1,743 differentially expressed transcripts revealed more up-regulated genes (203) than down-regulated genes (194) at 14 dpi, and a higher number of repressed genes for both 24 dpi (369) and 36 dpi (701) compared with induced genes at 24 dpi (323) and 36 dpi (275), respectively. The margin between induced and repressed genes at 14 dpi was very narrow (difference of 9 genes favouring up-regulation) which increased to a 46 gene difference at 24 dpi, favouring down-regulation. At 36 dpi, a 426 difference in down-

regulated genes was evident (Figure 3). We propose that the higher number of induced genes at 14 dpi may reflect more of a general non-specific innate host response to virus invasion by the activation of stress and defense-like genes, whereas the increase in down-regulated genes at 24 and 36 dpi is indicative of SACMV attempt to hijack many host processes for its own benefit, leading to repression of a large number of genes. The host (*Arabidopsis*) may also be attempting to divert metabolites such as those involved in, among others, glycolysis and gluconeogenesis, pentose-phosphate pathways, and carbohydrate metabolism, away from normal cell function in order to conserve energy, as well as defend itself from SACMV attack (Figure 3).

Comparison of 2-Fold Gene Expression Patterns With Other Datasets

In a comparative plant virus microarray study by Postnikova and Nemchinov (2012), they demonstrated that collectively from eleven *Arabidopsis*-virus interaction studies, 7639 unique genes were significantly changed at least 2-fold, which represents 23 % of the *Arabidopsis* genome. SACMV shared 817 genes (across three time points) in common with the 7639 unigenes (Table S3), and 524 genes (across three time points) in common with the geminivirus, CaLCuV, at 12 dpi (Table S4). Only 19 genes (Table S5) were common to SACMV, CaLCuV and the 7639 unigenes [34]. This was not surprising as only 198 genes were differentially expressed in response to all eleven viruses (9 RNA; 1 dsDNA; 1 ssDNA) in the *Arabidopsis* comparative microarray study [34], pointing to the unique nature of virus-host interactions (Postnikova and Nemchinov, 2012). However, as useful as these comparisons are, one must acknowledge the limitations in comparing individual and combined datasets. Another notable observation was that an estimated 12%, 15% and 22% of responsive genes described in the SACMV, eleven *Arabidopsis*-virus and CaLCuV studies, respectively, were related to abiotic/biotic stress/defense, and over-representation in this functional category is not uncommon in virus-host interactions (Whitham *et al*., 2003, Whitham *et al*., 2006; Postnikova and Nemchinov, 2012).

In the CaLCuV study [10], at 12 dpi (representing prominent symptoms and active viral replication), a significantly (q value <0.005) high number (5365 representing 23% of the Affymetrix total 22,748 gene probes) of genes were found to be differentially expressed, with 3004 being up-regulated and 2631 down-regulated (6% difference). Similarly in this study, at 14 and 21 dpi, differences in numbers of up-regulated and suppressed genes were not significant, but at 36 dpi there was a significant number of

repressed compared with up-regulated genes (difference of 43%). If one compares SACMV at 24 dpi with CaLCuV at 12 dpi (approximate similar stages of infection; fully symptomatic), the number of differentially expressed genes from the total number represented on the arrays, is significantly lower (4% of the Agilent 37,683 array probes) compared with CaLCuV (23%). However, thirty three percent of the 1,743 2-fold altered transcripts were differentially expressed at 24 dpi in this study, compared with 23% at 12 dpi in CaLCuV-infected *Arabidopsis*. This striking difference in gene expression levels, in the identical host, between two different geminiviruses, is hypothesized to be partly attributed to the more virulent nature of CaLCuV in *Arabidopsis*, resulting in a more severe symptom phenotype, and symptoms appearing much earlier, compared with SACMV. This would point to a greater susceptible host response and a higher number of gene alterations associated with cellular processes redirected by CaLCuV, suggesting that CaLCuV may be less adapted to this non-natural host compared to SACMV. Additionally, we consider it reasonable to speculate that different geographical evolutionary patterns of CaLCuV, a New World northern hemisphere geminivirus, and SACMV (southern hemisphere) from the Old World, in relation to the *Arabidopsis*, may also contribute to differences in host response.

Forty-one genes (2.3%) at a 2-fold cut-off were present across all three time points in SACMV infected *Arabidopsis* (Table 1), indicating that most genes were transiently expressed and not sustained throughout virus progression in time. A snapshot of the most significant highly induced and repressed (highest expression values) early-response genes occurring at 14 dpi indicated more signalling-related defense responses, whereas those appearing from middle to late responses (24 and 36 dpi) were primarily involved in metabolic functions (Table 2). As the shift continues from early to middle and late gene expression, host metabolism is altered, which suggests that more host metabolites may be diverted to aid in SACMV replication and cell-to-cell-spread, and at the same time, the host is diverting resources away from normal cell functions to minimize fitness costs in an attempt to defend itself against SACMV. At the 24 and 36 post-infection stage, a more specific defense response appears to be induced, evidenced by the induction of putative stress (AT4G12400) and disease resistance (AT1G72920) proteins (Table 2). Results from Table 1 and 2 provide evidence to support that *Arabidopsis* initiates early signalling and basal innate defense responses, albeit not sufficiently rapid or effective to prevent SACMV establishment.

Phytohormone and Signalling Networks

In order for plants to adapt to both biotic and abiotic stresses in a cost-efficient manner, cross communication between phytohormone signalling pathways must take place. Signalling pathways may be activated at the same time, depending on the type of pathogen or they may function to act synergistically or antagonistically in order to attempt to mount the most effective defense responses possible (Peninckx *et al*., 1998; Cheong *et al*., 2002; Marathe *et al*., 2004; Guo and Ecker, 2004; Llorente *et al*., 2008; Pieterse *et al*., 2009; Baena-Gonzalez, 2010). An example of two such pathways working antagonistically was shown by the suppression of the Jasmonic Acid (JA) pathway by salicylic acid (SA) signalling pathway induction following CaLCuV infection in *Arabidopsis* (Ascencio-Ibanez, 2008). JA and ET are also known to work synergistically with each other as shown by several studies, including Penninckx *et al*. 1998 (Penninckx *et al*, 1998). In contrast to CaLCuV, in our study, SA, JA, and ET appeared to function concomitantly in infected *Arabidopsis* as both up-regulation of PR genes (SA pathway) and defensin (PDF) genes ((JA/ET pathways) (2-fold or more) was evident (Table S1). Several pathogenesis-related (PR) genes were up-regulated at 14, 24 and 36 dpi. These included, PR1, AT2G14610 (24 dpi, 1.86, and 36 dpi, 3.04), PR5, AT1G75040 (14 dpi, 2.18, 24 dpi, 1.42, and 36 dpi, 1.56), PR4, AT3G04720 (14 dpi, 3.19, 24 dpi, 3.56, and 36 dpi, 2.00), PR-1-like, AT2G19990 (24 dpi, 2.45), and PR protein, AT2G19970 (24dpi, 2.15), confirming functioning of the SA pathway. Significant induction of JA/ET responsive genes such as PDF 1.2a,b and c (> 9 fold up-regulation) and VSP1 (4.78 fold change) (Table 1, 2 and Table S2) were also noted. Ethylene response factor 4 DNA binding /transcriptional repressor (ERF4)(AT3G15210) was significantly down-regulated (-4.64) (Table 2), indicating a possible switching on of transcription of ET signalling. Concomitant functioning of jasmonate and ethylene response pathways have been shown in a previous study to be required for induction of a plant defensin gene in Arabidopsis (Penninckx *et al*., 1998). *Cauliflower mosaic virus*, a compatible pathogen of Arabidopsis, has been shown to engage three distinct (ET/JA/SA) defense-signalling pathways (Love *et al*., 1995). PR and PDF transcripts were dominantly prevalent in apical leaves, suggesting that all three pathways, SA, JA and ET, are operational/activated by SACMV in *Arabidopsis* and are acting synergistically with each other, as shown by the induction of marker genes such as PR and PDF (Table S2). However, JA/ET signalling may be favoured over SA pathway since marker genes for JA/ET were more highly induced throughout the study, compared

with SA. A basal type of resistance response is ongoing, but is unable to prevent SACMV replication and systemic movement.

Auxin has been shown to be involved in disease susceptibility to viral pathogens (Padmanabahn *et al*., 2006; Culver and Padmanabahn, 2007; Robert-Seilaniantz *et al*., 2007; Spaepen and Vanderleyden, 2011), for example TMV, where the 126 and 183kDa replicase disrupts interacting Aux/IAA proteins promoting disease development (Padmanabahn *et al*., 2006). In addition, Aux/IAA proteins were also shown to be down-regulated by PPV in *Arabidopsis* (AT5G57420 and AT1G52830) (Babu *et al*., 2008a). SA, on the other hand, is able to affect disease susceptibility by repressing the auxin receptor F-box protein TIR1 (Transport Inhibitor response 1, ubiquitin-protein ligase, AT3G62980) causing enhanced resistance (Dharmasiri *et al*., 2005). This was not evident in this study as TIR1 was not repressed but up-regulated at 24 dpi (1.52). Furthermore, all auxin-responsive genes identified in our >2-fold change category were activated by infection (Tables 1, 2), suggesting that, together with evidence of TIR1 activation, symptom and disease progression was allowed to continue in *Arabidopsis*. Indeed, the auxin-responsive protein, AT4G38860 (SAUR-like auxin responsive), was up-regulated at 14dpi (2.98), 24dpi (3.06), and 36dpi (2.45) and IAA29 (AT4G32280) was also induced at 14, 24, and 36 dpi (2.32; 3.27; and 2.57, respectively). It may be advantageous for a geminivirus to regulate this pathway as a means to create a favourable cellular environment for replication in apical leaves.

Brassinosteroids control many aspects of plant growth and development, and are able to induce broad spectrum resistance, but their connection to SA/JA/ET remains to be established (Nakashita *et al*., 2003; Bari and Jones, 2009). A receptor - like kinase, BAK1 has been shown to interact with receptors that recognize pathogen molecules. BRI1 is one member of a family of leucine-rich receptor-like kinase (LRR-RLK) receptors which interacts with BAK1 upon brassinosteroid perception, initiating the signalling pathway involved in growth - and development related processes (Li *et al*., 2002). Although the roles of BAK1 in immunity and in brassinosteroid signalling seem to function independently and remain to be elucidated, BRI1 (AT4G39400) was down-regulated in our study at 14 dpi (-1.19), and a BKI1 kinase inhibitor (AT5G42750) was shown to be up-regulated at 24 dpi (1.37) indicating SACMV-induced suppression of the BR1 receptor. This in turn would disrupt brassinosteroid signal transduction as transduction requires heterodimerisation of BRI1 and BAK1 to elicit transcriptional activation of responsive genes. In the same way as

C4 of another geminivirus, *Beet curly top virus* (BCTV), may suppress antiviral host defence by disrupting LRR-RLK activity (Piroux *et al*., 2007), prevention of brassinosteroid-associated signal perception and downstream deactivation of the LRR-RLK BRI1 by SACMV may contribute to failure to activate transcription of resistance-related responsive genes.

Signalling and Cell-Cycle Regulation Comparison with The Bipartite Geminivirus, CaLCuV

Several core cell-cycle genes were found to be differentially expressed in this study (Figure 6). Functional links between plant signalling hormones (auxin, ethylene, brassinosteroids and cytokinins), and cell-cycle proteins have been established (Dudits *et al*., 2007; Bari and Jones, 2009), and this is depicted in figure 7. Plant hormones may either directly influence cell-cycle entry and transition or indirectly through developmental regulatory proteins. It has been shown that auxin may stimulate entry into the S-phase, as shown by an increase in histone H4 promoter activity. We believe that SACMV may be responsible for the induction of auxin partly in order to promote S-phase activation. As evidenced by CaLCuV-induced core cell cycle gene transcriptional alterations, geminiviruses manipulate the core cell cycle genes (induce S-phase and G2 genes) in order to provide a replication-enabling environment (Ascencio-Ibanez *et al*., 2008). A similar finding was observed with SACMV, where 44 of the 61 core cell cycle genes (Vandepoele *et al*., 2002) were differentially expressed (Figure 6). We believe this to hold true for SACMV as cyclin genes, such as S-phase CYCA3;2, were induced at both 14 dpi (1.32) and at 36 dpi (1.61). In addition, an auxin-responsive factor protein (AT4G38860) was shown to be up-regulated consistently across time points strongly supporting our hypothesis (Figure 4, Table 1).

CYCB1;1 and CDKB2;1 both promote mitosis and growth in *Arabidopsis*, however opposite effects on expression were noted in both SACMV and CaLCuV studies (Table S6). Down-regulation of CDKB2;1 was noted in both SACMV at 24 dpi (-1.69 fold change) and CalCuV at 12 dpi, while CYCB1;1 was induced by both viruses, and in SACMV-infected Arabidopsis remained induced even at 36 dpi. The SACMV results support the proposal suggested by Ascencio-Ibanez *et al*., 2008, that elevated CYCB1;1 leads to sequestering factors necessary for G2 arrest, while reduced CDKB2;1 expression at the G2/M boundary maintains G2 and blocks entry into the M phase, leading to shut down of

meristem during infection. In an abiotic stress response study, upon gamma-ray (IR) induction (Ricaud *et al*., 2007), G2/M phase inducers such as CYCB2;1 (and CYCB1;4, CYCB2;2, CYCA1;1) and CDKB1;2, were down-regulated, but CYCB1;1 was induced, similar to biotic stresses (CaCuLV) (Ascencio-Ibanez *et al*., 2008) and SACMV, as mentioned above. G2 to M transition takes place with CDK complexes containing CYCA and CYCB cyclins. WEE1 kinases and inhibitory proteins (CKI's) phosphorylate CDK complexes in order to keep them in their inactive states. The CKI protein is released by positive phosphorylation by CAK kinase and an unknown protein at the G2 to M boundary, and the kinase is activated (Andrietta *et al*., 2001). A link in SACMV-infected *Arabidopsis* between CYCB1;1 and auxin is suggested by the observation that the CYCB1;1's promoter contains an auxin response factor (ARF) binding site (Ricaud *et al*., 2007). Negative regulators of CDKA;1, namely WEE1, expressed at S-phase, were shown to be up-regulated upon IR induction, most likely to ensure that cell division is delayed from G2 to M (De Schutter *et al*., 2007). WEE1 (AT1G02970) was also elevated upon SACMV infection, supporting the above-mentioned hypothesis that the G2 phase is maintained by geminiviruses. It is also suggested that, as KRPs (encoding a cyclin-dependent kinase inhibitor) normally function as a negative regulators of cell division (Agudelo-Romero *et al*., 2008), induction of KRP2 and KRP5 by SACMV at 24 dpi and 14 dpi, respectively, may contribute to M phase repression. The interaction between phytohormone signalling and cell cycle gene pathways (Figure 7) illustrates that these pathway genes may be co-ordinately suppressed or induced by geminiviruses when required. Here we suggest that SACMV has a concomitant impact on cell-cycle progression and selected hormones that influence the pathways.

Certain features that control the cell-cycle are conserved among eukaryotes in order to ensure mitosis does not begin until DNA replication is completed (Andrietta *et al*., 2001; Sorrell *et al*., 2001). Cyclin-dependent kinases bind to the various cyclin types according the phase of the cycle they are entering, and are responsible for transit through control points in cell-regulation. It is the cyclin which determines the specificity and sub-cellular localization, as it is the regulatory component of the complex and can be classified into G1, S and G2-phases (Andrietta *et al*., 2001; Sorrell *et al*., 2001). In addition, CDKs are also regulated by interacting proteins and posttranslational modifications (Figure 7) (Andrietta *et al*., 2001). In general, G1 to S transition phases are controlled by CDK containing D-type cyclins which function to release E2F transcription factors in order for transcription of genes necessary for G1 to S transition to occur. They do this by

phosphorylating the retinoblastoma protein (RBR) (Andrietta *et al*., 2001; Ascencio-Ibanez *et al*., 2008). It was demonstrated that CaLCuV-infected *Arabidopsis* cells only pass through the early G1 phase since genes such as CYCD1;1 and CYCD3;2 were down-regulated (Ascencio-Ibanez *et al*., 2008). Differentially expressed core cell cycle genes detected in the SACMV-*Arabidopsis* array were not always picked up in the CaLCuV-*Arabidopsis* hybridization. However a comparison between differentially regulated gene expression between the two geminiviruses (Table S6) showed some similarities. While CYCD1;1 was not detected in the SACMV study, CYCD3;2 was also reduced by SACMV at 14dpi (-1.38) and at 24 dpi (-1.15), indicating it is likely that geminivirus-infected cells only transit through late G1 (Ascencio-Ibanez *et al*., 2008). Additionally, late G1 cyclin CYCD4;2 was induced by both SACMV and CaLCuV (Table S6). CaLCuV AC1 binding to RBR causes changes to E2F (E2FA and E2FC) expression by bypassing the G1 phase leading to induction of the endocycle. CYCD3's normal function is to promote the mitotic cycle and prevent endocycle (Ascencio-Ibanez *et al*., 2008). Thus, down-regulation of CYCD's prevent the mitotic cycle from taking place. In addition, genes such as CYCD3;1, CYCD3;2, and CYCD3;3 mutants showed severe symptoms at 12 dpi in CaLCuV suggesting that CaLCuV replicates in endocycling cells. In this study, SACMV infection led to a similar response compared with CaLCuV, as down regulation of CYCD3;1, CYCD3;2 and CYCD3;3 was persistent at 14 and 24 dpi.

The above listed similarities in cell cycle regulation which occur upon biotic stresses such as CalCuV and SACMV infection provided some insight into what is required for geminiviruses to establish a replication-efficient environment, and in addition, similarities shown between abiotic stresses, such as IR induction, confirms that certain cell-cycle regulators are conserved, as previously suggested in other studies.

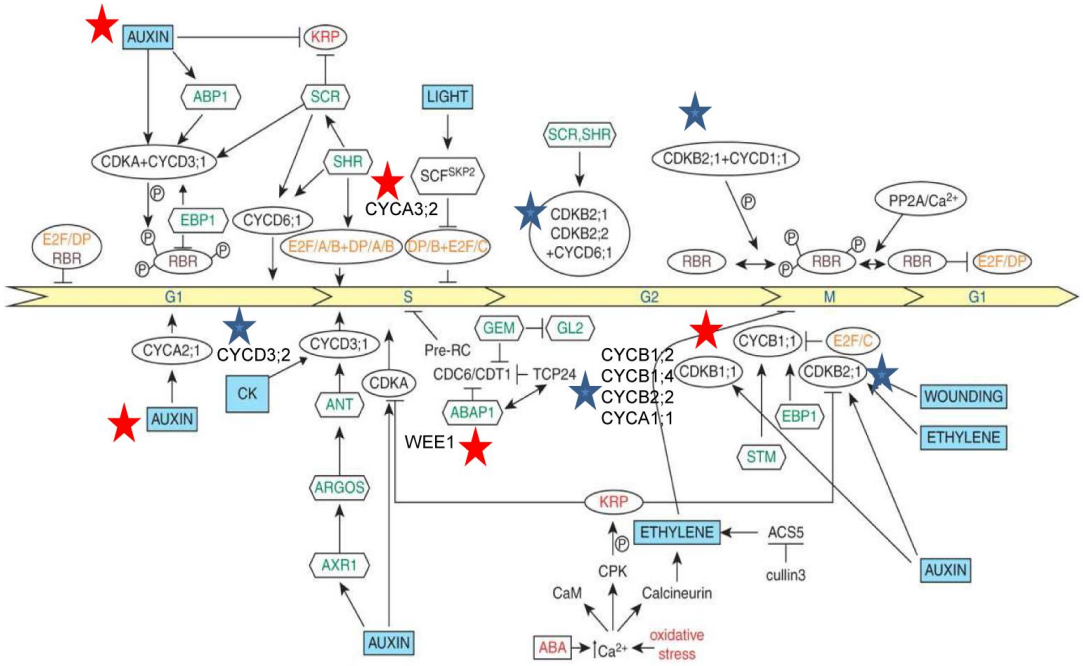


Figure 7. **Potential links between hormonal signals and cell cycle regulators.** Abbreviations: CK, cytokinin; E2F/DP, transcription factors; RBR, retinoblastoma-related protein; P, phospho-protein; CYC, cyclin; CDK, cyclindependent kinase; PP2A, phosphatase; SCR, SCARECROW; SHR, SHORT ROOT; SCF, SKP1 + CULLIN + F-box (SKP2); EBP1, plant homologue of epidermal growth factor-binding protein; SKP2, F-box protein; STM, SHOOT MERISTEMLESS; KRP, CDK inhibitor; CaM, calmodulin; CPK, calmodulin-like domain protein kinase; ABAP1, armadillo BTB Arabidopsis protein 1; TCP24, transcription factor; CDT1, DNA replication-licensing factor; ABP1, auxin binding protein 1; ANT, aintegumenta; ARGOS, auxin-regulated gene in organ size; AXR1, RUB1-activating enzyme; ABA, abscisic acid; GL2, GLABRA (root hair); GEM, GL2 expression regulator; ACS5, 1-aminocyclo-propane-1-carboxil acid synthase [72]. Stars depict SACMV-[ZA:99]-[ZA:99] involvement in hormone signals and cell cycle regulators. Red stars show up-regulation, while blue stars show down-regulation.

Comparison of Data Between SACMV and The Monopartite Geminivirus Tomato Yellow Leaf Curl Virus (TYLCV)

In a comparative investigation of gene expression changes induced by TYLCV in *Nicotiana benthamiana* (Lozano-Duran *et al*., 2011), we identified 27 common genes with SACMV (Table 3). Many of these genes were shown to have either no effect on infection

by TYLCV, or were involved in promotion of earlier infection or in a delay or reduction of infection. The three genes with the highest fold change in SACMV-infected *Arabidopsis* were histone 3 K4-specific methyltransferase (2.38 fold change), which was up-regulated, and two genes which were significantly down-regulated, namely a putative transcriptional activator with NAC domain (-2.16) and a scarecrow-like protein (SCL13) (-2.41). Histone 3 K4-specific methyltransferase and the putative transcriptional activator with NAC domain protein (ATAF1) have been shown to interact with monopartite geminiviral proteins, namely TrAP/C2 and C3, respectively, while the scarecrow-like protein has been found to be a transcription factor, and overexpressed in phloem (Vilaine *et al*., 2003). Histone 3 K4-specific methyltransferase is located in the chloroplast but its function is not known. A NAC domain protein (SINAC1) was shown to be induced by *Tomato leaf curl virus* (TLCV), interact with the replication enhancer protein of TLCV in tomato, and promote replication (Selth *et al*., 2005). Furthermore, interaction of TMV replicase protein with a NAC domain transcription factor (ATAF2) has also been shown to be associated with suppression of systemic host defences, promoting systemic virus accumulation (Wang *et al*., 2009). In SACMV, down-regulation of ATAF1 at 24 dpi would appear to behave in contradiction to the TMV and TLCV study, and it would be interesting in future to ascertain whether it can bind to SACMV AC2/AC3 proteins.

Genes such as NSI, GRAB2, and RPA32 were also shown to modify TYLCSV infection in *N. benthamiana* (Table 3) [15]. In SACMV-infected *Arabidopsis*, GRAB2 was up-regulated at 14dpi (1.36) and at 24dpi (1.61), respectively. GRAB2 is a Rep A binding protein whose exact role in replication initiation is unclear. An increase in expression was shown to cause inhibition of replication of the monopartite geminivirus, *Wheat dwarf virus* (WDV) (Xie *et al*., 1999), whereas in contrast, down-regulation of GRAB2 caused inhibition of TYLCSV infection indicating that GRAB2 is required for complete infectivity but that the appropriate expression levels are critical (Lozano-Duran *et al*., 2011). According to the TYLCSV study by Lozano-Durán *et al*., 2011, 8 of the 18 differentially expressed genes involved in protein modifications, were associated with ubiquitination, acetylation, protein folding, phosphorylation and rubylation, four of which were involved in ubiquitination (UBA1, RHF2A, ASK2, and CSN3). UBA1 was found to be down-regulated by SACMV at 24dpi (-1.21). This gene is involved in many levels of plant defense, one of which is virus resistance. Down-regulation of this gene by both a monopartite and bipartite geminivirus, TYLCSV and SACMV, respectively, favours the proposal that a geminiviral protein interaction, C2 protein in the case of TYLCSV, inhibits UBA1-mediated ubiquitination of

possible viral proteins or host protein(s) linked to a resistance-associated response, which would favour progression of infection. Silencing of UBA1 resulted in early TYLCSV infection, supporting this theory. RFH2A was also silenced by TYLCSV, prolonging virus infection, and this gene was also found to be repressed by SACMV at 14dpi (-1.21) (Table 3) confirming its likely role in sustaining virus infection. It has also been suggested that this gene may be involved in counteracting plant defense, as it was up-regulated by CaLCuV in *Arabidopsis* at 12 dpi (Ascencio-Ibanez *et al*., 2008). Genes identified in biotic stress responses (RD21, GLO1, and PLP2) upon TYLCSV infection were also induced by SACMV at 14dpi and/or 24dpi, demonstrating that geminiviruses, in addition to RNA plant viruses in general (Whitham *et al*., 2003), initiate basal innate plant defense responses, and that this is not unique to a particular group of pathogens. AOC1, involved in JA biosynthesis was differentially expressed at all 3 time points upon SACMV infection [up-regulated at 14 dpi (1.28) and down-regulated at 24 dpi (-1.87) and 36 dpi (-3.05)], but up-regulation early in infection (14 dpi) suggests an early non-specific JA-associated broad defense host response, as discussed previously. In contrast, AOC1 was reduced by CaLCuV infection, correlating with its suppression of the JA pathway and the induction of the SA pathway.

Selected Genes of Interest With More Than 2-Fold Expression Changes

Plant defensins are cationic antimicrobial peptides, belonging to classes four and five, and are involved in plant innate immunity (Broekaert *et al*., 1995). The *Arabidopsis* defensins are divided into three families. PDF1-3 (Thomma *et al*., 2002) and expression of defensins are highly regulated, usually linked to the ET and JA pathways (Penninckx *et al*., 1998). For example, PDF1.2a (AT5G44420) which is a low molecular weight cysteine-rich protein, is highly responsive to ET and JA, and is involved in JA- and ET-.dependent systemic resistance. This PDF is not responsive to salicylic acid and is located in the cell wall and extracellular region. PDF1.2b (AT2G26020) and PDF1.2c (AT5G44430) encode for pathogenesis-related (PR) proteins involved in the ET-mediated signalling pathway, and are also cell wall and extracellularly located. PDF1.3 is a PR-protein which is involved in innate defense responses (Thomma *et al*., 2002). PDF1.2a, b, and c, and PDF1.3 represented some of the most highly up-regulated genes (6.14 -15.82 fold changes) across all time points in this study (Tables 1 and 2). Transcription factors ERF1 and ORA59 form part of the APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF) superfamily. The AP2/ERF domains bind to a GCC promoter box of stress-responsive

genes, and can act as either activators or repressors of stress responsive genes [54,59,78](Guo and Ecker, 2004; Mishra *et al*., 2009; Spaepen and Vanderleyden, 2011). AP2 domain-containing transcription factors were down-regulated across all time points at a log 2 fold cut-off (Figure 4, Table 1). In an abiotic stress response study conducted by Brini *et al*., 2011, down-regulation of AP2 domain-containing transcription factors and up-regulation of plant defensin genes such as PDF1.2 was evident, illustrating a common trend in expression patterns to both abiotic and biotic stress responses. Plant defensin genes were highly up-regulated in our study suggesting that JA/ET signalling pathways were acting synergistically or concomitantly, leading to up-regulation of these genes in response to SACMV-[ZA:99].

Toll-interleuken-1-receptor/nucleotide binding site/leucine rich repeat (TIR-NBS-LRR) is a disease resistance protein which confers specific resistance to viral diseases. This was up-regulated (10.84) in *Arabidopsis* protoplasts by the RNA virus, *Plum pox virus* (PPV) (Babu *et al*., 2008a), but was down-regulated by SACMV in *Arabidopsis* leaves. Repressed TIR-NBS-LRR disease resistance proteins for SACMV infection in *Arabidopsis* were as follows:- AT5G41740 (-2.76 (14 dpi), -2.47 (24 dpi)), AT3G44630 (-2.08, 24 dpi), AT4G19520 (-2.30 (14 dpi), -2.24 (24 dpi)), AT5G41550, -2.48 (24 dpi), AT5G18360 (-2.32, 24 dpi), AT5G22690 (-2.98, 24 dpi), AT5G58120 (-2.03, 24 dpi), AT1G56510 (-2.89, 24 dpi), and AT1G56540 (-2.02, 24 dpi)]. TIR-NBS-LRR protein down-regulation supports a model that SACMV suppresses these disease resistance proteins in order to allow for replication and spread.

Little is known about cell-to-cell movement of geminiviruses, and we were keen to identify putative host proteins known to play a role in RNA virus movement (Boevink and Oparka, 2005). ß-1,3-glucanase (BGL2) (AT3G57260), BGLU46 and BGL1 (Table 2) were found to be up-regulated by SACMV at all three time points, especially at 14 dpi (3.01) [24dpi (1.73), and 36dpi (1.36)], with 14 dpi showing the highest expression. Callose deposition/removal and ß-1,3-glucanase activity have been associated with plasmadesmatal (Pd) gate modifications (Levy *et al*., 2007; Epel, 2009). Degradation of callose by ß-1,3-glucanases increases the Pd size exclusion limit (SEL), and has been implicated in facilitating cell-to-cell movement of RNA viruses (Levy *et al*., 2007; Epel, 2009). RNA viruses (TVCV, ORMV, PVX, CMV, and TuMV) all demonstrated elevated ß-1,3-glucanase activity at 2,4,5 DAI (days after infection), increasing exponentially over the time course of infection (Whitham *et al*., 2003). Another interesting gene, 4CL1, is

responsible for channelling carbon flow in the phenylpropanoid metabolic pathway. It appears to be involved in cell wall modification as silencing of this gene caused increased cellulose and decreased lignin in general (Guerra-Peraza *et al*., 2005; Jin and Villegas, 2006). 4CL1 was shown to be up-regulated at 14 dpi (1.21) and 24 dpi (1.40), and significantly down-regulated at 36 dpi (-2.50) by SACMV, indicating a possible synergistic role, along with ß-1,3-glucanase, in SACMV cell-to-cell movement via cell wall modifications. Up-regulation of ß-1,3-glucanase and callose breakdown, along with decreased lignin production in this SACMV-*Arabidopsis* interaction, strongly supports involvement in cell wall modification at the Pd location in facilitating geminivirus cell-to-cell movement, and may argue for a cell-wall "loosening" associated mechanism and Pd gate expansion model as a general conserved plant response to many RNA and DNA virus infections.

Two important protein families of interest in virus-host interactions are those belonging to the proteosome-related and heat shock protein (HSPs) associated pathways (Aranda *et al*., 1996; Whitham *et al*., 2003; Wang *et al*., 2004; Jin and Villegas, 2006; Whitham *et al*., 2006; Babu *et al*., 2008a; Camborde *et al*., 2010). In *Plum pox virus* (PPV) infection study (Babu *et al*., 2008a), genes associated with the 26S proteasome were found to be highly significantly ($Q < 0.05$), up-regulated, one of which being AAA-ATPAse. The 26S proteosome functions to control degradation of regulatory target proteins such as virus-encoded movement proteins, suggesting an involvement in resistance (Babu *et al*., 2008a). In this study, AAA type ATPase family protein (AT2G18193) was shown to be highly up-regulated across three time-points [4.25(14dpi), 4.57(24dpi), and 3.51(36dpi)] (Table 1). This suggests that a basal resistance may be activated but is not sufficient enough to counteract SACMV attack as an increase in virus titre across the time line was evident, resulting in a susceptible interaction (Figure 1C, and Figure 4,Table 1).

HSP's are involved in a wide range of functions in both abiotic and biotic cellular stress and in plant growth and development, and are controlled at the transcriptional level (Aranda *et al*., 1996; Whitham *et al*., 2003; Wang *et al*., 2004; Whitham *et al*., 2006; Scarpei *et al*., 2008; Li *et al*., 2011). In many plant studies with RNA viruses, HSP's are shown to be up-regulated as a general stress response upon virus attack (Whitham *et al*., 2003; babu *et al*., 2008a). Little is known about HSP's associated with host responses to DNA viruses, but mention was made to induction of HSP70 in response to the geminivirus, *Beet curly top virus* (Escaler *et al*., 2009). In this study, we were surprised to observe that

many HSP's were down-regulated at a 2-fold cut-off (Table S2) and several small class III heat shock proteins (HSP17.4-CIII); HSP17.8-Cl) and HSP17.6A-Cl were also found to be highly repressed across all time points (Table 1). *Arabidopsis* cytosolic HSP17.6A was shown to be a chaperone protein, induced by heat and osmotic stress (Sun *et al*., 2001), and HSP17.8 functions as an AKR2A cofactor in targeting the chloroplast outer membrane proteins in Arabidopsis (Kim *et al*., 2011). Since many HSPs are up-regulated by abiotic and biotic stress, opposite findings in our study suggest multiple roles for HSPs in both general and geminivirus-specific stress responses and possibly virus replication. Li *et al*., 2011 recently identified a heat shock protein 70 (HSP70) which may play multiple roles in virus replication of influenza A, such as interaction with the influenza virus ribonucleprotein (RNP) complex, which is involved in negative regulation of influenza A transcription and replication in infected cells. HSP70 may also assist with subcellular localization and membrane insertion of viral replication proteins and assembly of viral replicase (Wang *et al*., 2004; Li *et al*., 2011).

In *Arabidopsis,* heat shock proteins were induced by five RNA viruses (ORMV, TVCV, CMV, Potato virus X and TuMV) and by SYMV and INSV (negative-strand RNA viruses) in *N. benthamiana* (Whitham *et al*., 2003). Of the HSP's (HSP70 and HSP90) showing chaperone activity in the Agudelo-Romero *et al*. 2008 TEV study, one of the HSP's (HSP70,AT3G12580) in particular was also identified in our SACMV-*Arabidopsis* study, but showed opposite expression. HSP70 (AT3G12580) was up-regulated by TEV and down-regulated by SACMV (-1.98 at 14dpi, and -2.36 at 24dpi). This finding, again supports the earlier suggestion that HSP70 may play different roles at different times in virus-infected plants and that differential regulation of HSP's is not always a general stress response but may be specifically targeted by a geminivirus at a particular stage of infection for its own benefit, for example replication or cell-to-cell movement, where HSP70 family chaperones may well be exploited in general folding of movement protein-nucleic acid complexes (Boevink and Oparka, 2005), or regulation of host defenses directly or indirectly through interactions with J-domain proteins (Kanzaki *et al*., 2003). It has been suggested that one of the replicase, movement or 16-KDa proteins encoded by RNA1 of *Pea early browning virus* (PEBV) was possibly the elicitor for induction of HSP70 expression (Escaler *et al*., 2009). If this is the case, we suggest that if a movement protein is capable of eliciting HSP's (in particular HSP70) then it is also capable of suppressing HSP expression which is evident with significantly ($p<0.05$) down-regulated HSP's identified at a 2-fold cut-off in SACMV infection. Down-regulation of HSPs was also maintained across

the 36 day infection period. We think it not unreasonable to argue that down regulation may be mediated by SACMV in order to suppress innate immune responses, and redirect cellular pathways for its own replication and movement, and also suggest that some geminiviruses may not have an absolute requirement for heat shock for infection progression.

Conclusions

In conclusion, the large number of genes unveiled in this study provided valuable insight into the little that is known about geminivirus-host interactions. The GO results in this study are consistent with the hypothesis that plant virus stress leads to a transition from normal host growth processes to altered metabolic pathways geared for defense responses. Both similarities and differences were identified between SACMV and the geminiviruses, CaLCuV in *Arabidopsis* and TYLCV in *N. benthamiana*, and other RNA viruses, identifying general as well as virus-specific responses in a host. Importantly, we also demonstrate that different altered gene profiles occur at early, middle and late infection stages, and that a limited number of genes are differentially expressed across the entire infection period. Differences between geminiviruses in the same host, *Arabidopsis*, demonstrate that many host responses in a compatible interaction are geminivirus-specific, and differences in expression patterns may in part be a reflection of different adaptation and evolutionary histories of the viruses and their hosts. This is supported by the comparative microarray study of Arabidopsis, where, while some overlap in altered expression between different viruses in this host occurred, virus-host interactions were essentially unique (Postnikova and Nemchinov, 2012). It is evident that many host defense layers exist which viruses need to overcome in order to establish successful infection. The suppressive nature of SACMV on many host genes revealed that in a compatible interaction, basal defences are induced but are not capable of inhibiting viral replication and spread, as demonstrated by the progressive increase in symptom severity, virus titre and high number of repressed genes over the infection period. Identifying gene interactions in signalling pathways is a step closer toward identifying master transcription factors controlling these networks. A more systems biology approach will be adopted in further studies to connect these networks. Host-responsive genes may also be grouped or clustered based on their co-expression pattern or chromosomal location, and this also needs to be investigated. Functional testing of candidate genes and transcription factors

through a reverse genetics approach, RNA silencing, VIGS and miRNA studies, will also be the next step in expanding on our knowledge of geminivirus-host interactions.

3.6 References

- Agudelo-Romero P, Carbonell P, de la Iglesia F, Carrera J, Rodrigo G *et al*. (2008) Changes in the gene expression profile of Arabidopsis thaliana after infection with *Tobacco etch virus*. Virol J 5: 92.b
- Agudelo-Romero P, Carbonell P, Perez-Amador MA, Elena SF (2008) Virus adaptation by manipulation of host's gene expression. PLoS One 3(6): e2397.a
- Andersen CL, Jensen JL, Orntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 64(15): 5245-5250.
- Andrietta MH, Eloy NB, Hermerly AS, Ferreira PCG (2001) Identification of sugarcane cDNAs encoding components of cell cycle machinery. Gen Mol Bio 24(1-4):61-68.
- Aranda MA, Escaler M, Wang D, Maule AJ (1996) Induction of HSP70 and polyubiquitin expression associated with plant virus replication. Proc Natl Acad Sci U S A 93(26): 15289-15293.
- Ascencio-Ibanez JT, Sozzani R, Lee TJ, Chu TM, Wolfinger RD *et al*. (2008) Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant Physiol 148(1): 436-454.
- Babu M, Gagarinova AG, Brandle JE, Wang A (2008b) Association of the transcriptional response of soybean plants with soybean mosaic virus systemic infection. J Gen Virol 89(Pt 4): 1069-1080.b
- Babu M, Griffiths JS, Huang TS, Wang A (2008a) Altered gene expression changes in Arabidopsis leaf tissues and protoplasts in response to Plum pox virus infection. BMC Genomics 9: 325.a
- Baena-Gonzalez E (2010) Energy signaling in the regulation of gene expression during stress. Mol Plant 3(2): 300-313.
- Ballare CL (2011) Jasmonate-induced defenses: a tale of intelligence, collaborators and rascals. Trends Plant Sci 16(5): 249-257.

- Ballare CL (2011) Jasmonate-induced defenses: a tale of intelligence, collaborators and rascals. Trends Plant Sci 16(5): 249-257.

- Bari R, Jones JD (2009) Role of plant hormones in plant defence responses. Plant Mol Biol 69(4): 473-488.
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D *et al*. (2007) NCBI GEO: mining tens of millions of expression profiles—database and tools update. Nucleic Acids Res 35(Database issue): D760-765.
- Berrie LC, Rybicki EP, Rey ME (2001) Complete nucleotide sequence and host range of *South African cassava mosaic virus*: further evidence for recombination amongst begomoviruses. J Gen Virol 82(Pt 1): 53-58.
- Boevink P, Oparka KJ (2005) Virus-host interactions during movement processes. Plant Physiol 138(4): 1815-1821.
- Brini F, Yamamoto A, Jlaiel L, Takeda S, Hobo T *et al*. (2011) Pleiotropic effects of the wheat dehydrin DHN-5 on stress responses in Arabidopsis. Plant Cell Physiol 52(4): 676-688.
- Broekaert WF, Terras FRG, Cammue BPA, Osborn RW (1995) Plant defensins: novel antimicrobial peptides as components of the host defense system. Plant Physiol 108:1353-1358.
- Busch W, Lohmann JU (2007) Profiling a plant: expression analysis in Arabidopsis. Curr Opin Plant Biol 10(2): 136-141.
- Camborde L, Planchais S, Tournier V, Jakubiec A, Drugeon G *et al*. (2010) The ubiquitin-proteasome system regulates the accumulation of Turnip yellow mosaic virus RNA-dependent RNA polymerase during viral infection. Plant Cell 22(9): 3142-3152.
- Carrington JC, Whitham SA (1998) Viral invasion and host defense: strategies and counter-strategies. Curr Opin Plant Biol 1(4): 336-341.
- Cheong YH, Chang HS, Gupta R, Wang X, Zhu T *et al*. (2002) Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis. Plant Physiol 129(2): 661-677.
- Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162(1): 156-159.
- Culver JN, Padmanabhan MS (2007) Virus-induced disease: altering host physiology one interaction at a time. Annu Rev Phytopathol 45: 221-243.

- Dardick, C (2007) Comparative expression profiling of Nicotiana benthamiana leaves systemically infected with three fruit tree viruses. Mol Plant Microbe Interact 20(8): 1004-1017.
- De Schutter K, Joubes J, Cools T, Verkest A, Corellou F *et al*. (2007) Arabidopsis WEE1 Kinase Controls Cell Cycle Arrest in Response to Activation of the DNA integrity Checkpoint. Plant Cell 19: 211–225.
- Dharmasiri N, Dharmasiri S, Estelle M (2005) The F-box protein TIR1 is an auxin receptor. Nature 435(7041): 441-445.
- Doyle JJ, Doyle JL (1987) A rapid isolation procedure for small quantities of fresh leaf tissue.Phytochemical Bulletin 19:11-15.
- Dudits D, Abraham E, Miskolczi P, Ayaydin F, Bilgin M *et al*. (2007) Cell-cycle control as a target for calcium, hormonal and developmental signals: the role of phosphorylation in the retinoblastoma-centred pathway. Ann Bot 107(7): 1193-1202.
- Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30(1): 207-210.
- Elena SF, Carrera J, Rodrigo G (2011) A systems biology approach to the evolution of plant-virus interactions. Curr Opin Plant Biol 14(4): 372-377.
- Epel BL (2009) Plant viruses spread by diffusion on ER-associated movement-protein-rafts through plasmodesmata gated by viral induced host beta-1,3-glucanases. Semin Cell Dev Biol 20(9): 1074-1081.
- Escaler M, Aranda MA, Thomas CL, Maule AJ (2000) Pea embryonic tissues show common responses to the replication of a wide range of viruses. Virology 267(2): 318-325.
- Fisher R A (1970) Statistical Methods for Research Workers. 14th edition, Hafner Publishing, 1970, p. 96.
- Fontes EP, Santos AA, Luz DF, Waclawovsky AJ, Chory J (2004) The geminivirus nuclear shuttle protein is a virulence factor that suppresses transmembrane receptor kinase activity. Genes Dev 18(20): 2545-2556.
- Fregene M, Matsumura H, Akano A, Dixon A, Terauchi R (2004) Serial analysis of gene expression (SAGE) of host-plant resistance to the cassava mosaic disease (CMD). Plant Mol Biol 56(4): 563-571.

- Gafni Y, Epel BL (2002) The role of host and viral proteins in intra- and inter-cellular trafficking of geminiviruses. Physiological and Molecular Plant Pathology 60:231-241.
- Geisler-Lee J, O'Toole N, Ammar R, Provart NJ, Millar AH *et al*. (2007) A predicted interactome for Arabidopsis. Plant Physiol 145(2): 317-329.
- Golem S, Culver JN (2003) Tobacco mosaic virus induced alterations in the gene expression profile of Arabidopsis thaliana. Mol Plant Microbe Interact 16(8): 681-688.
- Gorovits R, Akad F, Beery H, Vidavsky F, Mahadav A *et al*. (2007) Expression of stress-response proteins upon whitefly-mediated inoculation of Tomato yellow leaf curl virus in susceptible and resistant tomato plants. Mol Plant Microbe Interact 20(11): 1376-1383.
- Guerra-Peraza O, Kirk D, Seltzer V, Veluthambi K, Schmit AC *et al*. (2005) Coat proteins of Rice tungro bacilliform virus and Mungbean yellow mosaic virus contain multiple nuclear-localization signals and interact with importin alpha. J Gen Virol 86(Pt 6): 1815-1826.
- Guo H, Ecker JR (2004) The ethylene signaling pathway: new insights. Curr Opin Plant Biol 7(1): 40-49.
- Gutierrez C (2000) DNA replication and cell cycle in plants: learning from geminiviruses. Embo J 19(5): 792-799.
- Gutierrez C (2002) Strategies for geminivirus DNA replication and cell cycle interference. Physiological and Molecular Plant Pathology 60: 219-230.
- Hanley-Bowdoin L, Settlage SB, Robertson D (2004) Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. Mol Plant Pathol 5(2): 149-156.
- Harrison BD, Robinson DJ (2002) Green shoots of geminivirology. Physiological and Molecular Plant Pathology 60: 215-218.
- Havelda Z, Hornyik C, Crescenzi A, Burgyan J (2003) In situ characterization of Cymbidium Ringspot Tombusvirus infection-induced posttranscriptional gene silencing in *Nicotiana benthamiana*. J Virol 77(10): 6082-6086.
- Jeske H (2009) Geminiviruses. Curr Top Microbiol Immunol 331: 185-226.
- Jin H, Li S, Villegas A, Jr. (2006) Down-regulation of the 26S proteasome subunit RPN9 inhibits viral systemic transport and alters plant vascular development. Plant Physiol 142(2): 651-661.

- Kanzaki H, Saitoh H, Ito A, Fujisawa S, Kamoun S *et al*. (2003) Cytosolic HSP90 and HSP70 are essential components of INF1-mediated hypersensitive response and non-host resistance to Pseudomonas cichorii in Nicotiana benthamiana. Mol Plant Pathol 4(5): 383-391.
- Kim DH, Xu ZY, Na YJ, Yoo YJ, Lee J *et al*. (2011) Small heat shock protein Hsp17.8 functions as an AKR2A cofactor in the targeting of chloroplast outer membrane proteins in Arabidopsis. Plant Physiol 157:132-146.
- Koornneef A, Pieterse CM (2008) Cross talk in defense signaling. Plant Physiol 146(3): 839-844.
- Koornneef M, Meinke D (2010) The development of Arabidopsis as a model plant. Plant J 61(6): 909-921.
- Levy A, Guenoune-Gelbart D, Epel BL (2007) Beta-1,3-Glucanases: Plasmodesmal Gate Keepers for Intercellular Communication. Plant Signal Behav 2(5): 404-407.
- Li G, Zhang J, Tong X, Liu W, Ye X (2011) Heat shock protein 70 inhibits the activity of Influenza A virus ribonucleoprotein and blocks the replication of virus in vitro and in vivo. PLoS One 6(2): e16546.
- Li J, Wen J, Lease KA, Doke JT, Tax FE *et al*. (2002) BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. Cell 110(2): 213-222.
- Llorente F, Muskett P, Sanchez-Vallet A, Lopez G, Ramos B *et al*. (2008) Repression of the auxin response pathway increases Arabidopsis susceptibility to necrotrophic fungi. Mol Plant 1(3): 496-509.
- Love AJ, Yun BW, Laval V, Loake GJ, Milner JJ (2005) Cauliflower mosaic virus, a compatible pathogen of Arabidopsis, engages three distinct defense-signaling pathways and activates rapid systemic generation of reactive oxygen species. Plant Physiol 139(2): 935-948.
- Lozano-Duran R, Rosas-Diaz T, Luna AP, Bejarano ER (2011) Identification of host genes involved in geminivirus infection using a reverse genetics approach. PLoS One 6(7): e22383.
- Ma S, Gong Q, Bohnert HJ (2007) An Arabidopsis gene network based on the graphical Gaussian model. Genome Res 17(11): 1614-1625.
- Marathe R, Guan Z, Anandalakshmi R, Zhao H, Dinesh-Kumar SP (2004) Study of Arabidopsis thaliana resistome in response to cucumber mosaic virus infection using whole genome microarray. Plant Mol Biol 55(4): 501-520.

- Mariano AC, Andrade MO, Santos AA, Carolino SMB, Oliveira ML *et al*. (2004) Identification of a novel receptor-like protein kinase that interacts with a geminivirus nuclear shuttle protein. Virology 318:24–3.
- Maule A, Leh V, Lederer C (2002) The dialogue between viruses and hosts in compatible interactions. Curr Opin Plant Biol 5(4): 279-284.
- Mills-Lujan K, Deom CM (2010) Geminivirus C4 protein alters Arabidopsis development. Protoplasma 239(1-4): 95-110.
- Mishra A, Godavarthi SK, Maheshwari M, Goswami A, Jana NR (2009) The ubiquitin ligase E6-AP is induced and recruited to aggresomes in response to proteasome inhibition and may be involved in the ubiquitination of Hsp70-bound misfolded proteins. J Biol Chem 284(16): 10537-10545.
- Nagar S, Pedersen TJ, Carrick KM, Hanley-Bowdoin L, Robertson D (1995) A geminivirus induces expression of a host DNA synthesis protein in terminally differentiated plant cells. Plant Cell 7(6): 705-719.
- Nakashita H, Yasuda M, Nitta T, Asami T, Fujioka S *et al*. (2003) Brassinosteroid functions in a broad range of disease resistance in tobacco and rice. Plant J 33(5): 887-898.
- O'Donnell PJ, Schmelz EA, Moussatche P, Lund ST, Jones JB *et al*. (2003) Susceptible to intolerance - a range of hormonal actions in a susceptible Arabidopsis pathogen response. Plant J 33(2): 245-257.
- Owens RA, Tech KB, Shao JY, Sano T, Baker CJ (2012) Global analysis of tomato gene expression during Potato spindle tuber viroid infection reveals a complex array of changes affecting hormone signaling. Mol Plant Microbe Interact 25(4): 582-598.
- Padmanabhan MS, Shiferaw H, Culver JN (2006) The Tobacco mosaic virus replicase protein disrupts the localization and function of interacting Aux/IAA proteins. Mol Plant Microbe Interact 19(8): 864-873.
- Pallas V, Garcia JA (2011) How do plant viruses induce disease? Interactions and interference with host components. J Gen Virol 92(Pt 12): 2691-2705.
- Penninckx IA, Thomma BP, Buchala A, Metraux JP, Broekaert WF (1998) Concomitant activation of jasmonate and ethylene response pathways is required for induction of a plant defensin gene in Arabidopsis. Plant Cell 10(12): 2103-2113.
- Petti C, Wendt T, Meade C, Mullins E (2009) Evidence of genotype dependency within *Agrobacterium tumefaciens* in relation to the integration of vector backbone

sequence in transgenic *Phytophthora infestans*-tolerant potato. J Biosci Bioeng 107(3): 301-306.

- Pieterse CM, Leon-Reyes A, Van der Ent S, Van Wees SC (2009) Networking by small-molecule hormones in plant immunity. Nat Chem Biol 5(5): 308-316.
- Piroux N, Saunders K, Page A, Stanley J (2007) Geminivirus pathogenicity protein C4 interacts with Arabidopsis thaliana shaggy-related protein kinase AtSKeta, a component of the brassinosteroid signalling pathway. Virology 362(2): 428-440.
- Postnikova OA, Nemchinov LG (2012) Comparative analysis of microarray data in Arabidopsis transcriptome during compatible interactions with plant viruses. Virology J 9:101.
- Ricaud L, Proux C, Renou JP, Pichon O, Fochesato S *et al*. (2007) ATM-mediated transcriptional and developmental responses to gamma-rays in Arabidopsis. PLoS One 2(5): e430.
- Robert-Seilaniantz A, Navarro L, Bari R, Jones JD (2007) Pathological hormone imbalances. Curr Opin Plant Biol 10(4): 372-379.
- Scarpeci TE, Zanor MI, Valle EM (2008) Investigating the role of plant heat shock proteins during oxidative stress. Plant Signal Behav 3(10): 856-857.
- Selth L, Dogra SC, Rasheed MS, Healy H, Randles JW *et al*. (2005) NAC Domain Protein Interacts with *Tomato leaf curl virus* Replication Accessory Protein and Enhances Viral Replication. Plant Cell 17: 311-325.
- Smyth GK (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology Volume 3, Issue 1, Article 3.
- Sorrell DA, Menges M, Healy JM, Deveaux Y, Amano C *et al*. (2001) Cell cycle regulation of cyclin-dependent kinases in tobacco cultivar Bright Yellow-2 cells. Plant Physiol 126(3): 1214-1223.
- Spaepen S, Vanderleyden J (2011) Auxin and plant-microbe interactions. Cold Spring Harb Perspect Biol 1;3:a001438..
- Sun W, Bernard C, van de Cotte B, Van Montagu M, Verbruggen N (2001) At-HSP17.6A, encoding a small heat-shock protein in Arabidopsis, can enhance osmotolerance upon overexpression. Plant J 27:407-415.
- Thomma BPHJ, Cammue BPA, Thevissen K (2002) Plant defensins. Planta 216:193-202.

- Trinks D, Rajeswaran R, Shivaprasad PV, Akbergenov R, Oakeley EJ *et al*. (2005) Suppression of RNA silencing by a geminivirus nuclear protein, AC2, correlates with transactivation of host genes. J Virol 79(4): 2517-2527.
- Vandepoele K, Raes J, De Veylder L, Rouze P, Rombauts S *et al*. (2002) Genome-wide analysis of core cell cycle genes in Arabidopsis. Plant Cell 14(4): 903-916.
- Veena, Jiang H, Doerge RW, Gelvin SB (2003) Transfer of T-DNA and Vir proteins to plant cells by *Agrobacterium tumefaciens* induces expression of host genes involved in mediating transformation and suppresses host defense gene expression. Plant J 35(2): 219-236.
- Vilaine F, Palauqui J-C, Amselem J, Kusiak C, Lemoine R *et al*. (2003) Towards deciphering phloem: a transcriptome analysis of the phloem of *Apium graveolens*. Plant J 36: 67-81.
- Wang W, Vinocur B, Shoseyov O, Altman A (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci 9(5): 244-252.
- Wang X, Sameer P, Goregaoker P, Culver JN (2009) Interaction of the *Tobacco Mosaic Virus* Replicase Protein with a NAC Domain Transcription Factor is Associated with the Suppression of Systemic Host Defenses. J of Virol 83: 9720-9730.
- Whitham SA, Quan S, Chang HS, Cooper B, Estes B *et al*. (2003) Diverse RNA viruses elicit the expression of common sets of genes in susceptible Arabidopsis thaliana plants. Plant J 33(2): 271-283.
- Whitham SA, Yang C, Goodin MM (2006) Global impact: elucidating plant responses to viral infection. Mol Plant Microbe Interact 19(11): 1207-1215.
- Xie Q, Sanz-Burgos AP, Guo H, Garcia JA, Gutierrez C (1999) GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. Plant Mol Biol 39(4): 647-656.
- Zhong R, Ye ZH (2009) Transcriptional regulation of lignin biosynthesis. Plant Signal Behav 4(11): 1028-1034.

Chapter 4

Transcriptome Profiling of Susceptible Cassava T200 Compared with Susceptible Model- Host, *Arabidopsis thaliana*, and Tolerant Cassava LandraceTME3 Infected with *South African cassava mosaic virus*

4.1 Abstract

Cassava suffers major yield losses from diseases such as Cassava Mosaic Disease (CMD) which is caused by geminiviruses of which SACMV-[ZA:99] is a member. Transcriptome profiling has become a useful strategy for the global identification of candidate susceptibility and defence-related genes from previously uncharacterised genomes such as cassava, and model-plant systems like *Arabidopsis.* In order to explore transcriptome changes in networks and pathways involved in host response to SACMV-[ZA:99] infection, next generation deep sequencing using the ABI SOLiD™ platform was carried out on virus-infected cassava cultivar T200 at three timepoints post agroinoculation, compared to mock - inoculated (*Agrobacterium*) plants. Assembled scaffolds and corresponding *Arabidopsis* accession numbers were identified in Phytozome. A 2.5 log2 fold cut-off ($p < 0.05$) was applied to the cassava T200 transcript data and differentially expressed genes were assigned to the Munich Information for Protein Sequences (MIPS) in order to identify host defence categories. The number of transcripts per time point increased from 12 dpi to 32 and 67 dpi, which correlated with an increase in virus titre levels (1.4E+00 molecules of DNA at 12 dpi to 2.19E+03 at 32 dpi, and 4.43E+05 at 67 dpi, respectively). Time point comparisons revealed a general shift from up-regulated genes (61%) at 12 dpi to down-regulation (72%) at 32 dpi with an approximate even number of differentially expressed genes at 67 dpi (49% suppressed). This shift from up- to down- regulation suggested disruption of host homeostasis as SACMV-[ZA:99] anatagonized host defence responses, and manipulated resources to aid in virus replication and spread. Functional categories assigned by MIPS for cassava T200 defence genes displayed similarities to the *Arabidopsis* - SACMV-[ZA:99] pathosystem, as categories such as subcellular localization, protein with binding function, and metabolism contained the highest number of differentially expressed genes between the two susceptible hosts, suggesting that SACMV-[ZA:99] specific responses were activated. In addition, signalling pathways such as SA, JA, and ET were working synergistically with each other in cassava T200, as demonstrated previously in *Arabidopsis* as well.

Comparisons between susceptible cassava T200 and tolerant TME3 identified 3 interesting heat map clusters, 2 of which were showing differences in expression patterns and 1 cluster was showing similarities between the cultivars. Transcription factors and genes involved in signalling were up-regulated in TME3 but down-regulated in T200. Protein kinases were shown to be repressed in TME3 but induced in T200, and transcripts such as PAR1 and PPR superfamily genes were shown to be up-regulated at 32 dpi for both TME3 and T200, but down-regulated at 67 dpi. These findings suggest that similar transcripts were activated and suppressed for virus-specific diversion of host resources, regardless of the genetic background of the host. Differences revealed that basal defences initiated by T200 were not fast or effective enough to limit SACMV-[ZA:99] replication and spread, resulting in disease as the final outcome.

4.2 Introduction

Plants have developed both highly specialized defence responses to prevent and limit disease. Many disease responses are activated locally at the site of infection, and can spread systemically when a plant is under pathogen attack (Mysore and Ryu,2004; Whitham *et al*., 2006; Agudelo-Romero *et al*., 2008; Pallas and Garcia, 2011).This initial response is usually termed basal or broad immunity which may be sufficient to combat the viral pathogen, or may lead to futher specific resistant responses, namely induced resistance, which is often triggered by specific recognition and interaction between virus and host resistance proteins encoded by R genes (Staskawicz *et al*., 1995; Feys *et al*., 2001, Jones and Dangl, 2006). Transcriptional reprogramming occurs both temporally and spatially within the plant leaves and other organs on a global level, and depending on the outcome, a resistance or susceptible response is initiated (Agudelo-Romero *et al*., 2008; Babu *et al*., 2008a; Elena *et al*, 2011). Specific induced resistance is usually associated with direct pathogen recognition, resulting in limited or inhibited pathogen spread, or programmed cell death, or hypersensitive response (HR), often followed by systemic signalling and systemic acquired resistance (SAR) (Feys *et al*, 2001; Durrant and Dong, 2004). In susceptible hosts, basal defences are initiated but are not fast or effective enough to limit pathogen growth, allowing the pathogen to replicate and spread systemically. Activated defence responses result from several possible signalling pathways, including reactive oxygen species (ROS), signalling molecules, and pathogenesis-related protein (PR proteins) induction, which causes biochemical and morphological alterations in the host plant such as cell-wall reinforcement (Fagard, 2007;

Blomster *et al*., 2011). The outcome between susceptibility and resistance depends on the pathogen-host genotype combination (Abramovitch *et al*., 2004), speed of host response, and specific virus pathogenicity determinants which recognize and interact with host-specific proteins (Jones and Dangl, 2006; Pavan *et al*., 2010). With plant viruses in particular, including geminiviruses, the pathogen has to suppress basal immune systems such as RNA silencing. Many virus-encoded proteins act as host defence reponse suppressors such as HC-Pro of potyviruses and AC2, AC3 and AC4-ORF-encoded proteins of geminiviruses (Ascencio-Ibanez *et al*., 2008; Lozano-Duran *et al*., 2011).

In the cascade of events which take place, host stress responses may vary between species, as well as within the actual stress response itself, not only activating or repressing genes common to both abiotic and biotic stresses, but also pathogen targeting-specific genes as well (Nagar *et al*., 1995; Havelda *et al*., 2003; Owens *et al*., 2012). The optimal defence of a plant is to initiate all defence mechanisms in order for at least some responses to be effective against the invading pathogen. This activation may be to the detriment of the plant as fitness costs may often outweigh the benefits, because energy and resources are redirected toward defence, and normal cellular processes such as growth and yield are affected (Bolton, 2009). In many cases, in the absence of a speedy, effective and persistent basal immnune response, plants will be suscetible, unless virus-specific R genes are present in that plant species/cultivar/variety. In order to minimise fitness costs, signalling molecules and pathways coordinating pathogen-specific defences are activated. Signalling molecules are predominantly regulated by salicylic acid (SA), jasmonic acid (JA), and ethylene (ET) pathways which are known to act synergistically or antagonistically with each other in order to minimise fitness costs (Bolton, 2009; Koornneef and Pieterse, 2008; Pieterse *et al*., 2009).

Successful pathogens use specialized tactics to induce disease by actively repressing plant defence responses or by avoiding them completely (Abramovitch and Martin, 2004). Once a virus has successfully entered and completed virus replication in initial cells, it spreads through the leaf tissue or other tissues, via plasmodesmata and colonises distal tissues in a plant, leading to a susceptible interaction resulting in disease as the final outcome (Carrington and Whitham, 1998; Maule *et al*., 2002, Agudelo-Romero *et al*., 2008). Geminiviruses have been implicated in many host-responsive processes such as transcriptional regulation, DNA replication, control of the cell-cycle, cell proliferation and differentiation, and macromolecular trafficking in whole plants (Gutierrez,

2002; Mariano *et al*., 2004; Fontes *et al*., 2004, Ascencio-Ibanez *et al*., 2008). In addition, the geminivirus AC2, AC3 and AC4 - encoded proteins have been implicated as a pathogenicity factors that assist in infection (Ascencio-Ibanez *et al*., 2008; Bolton, 2009; Lozano-Duran *et al*., 2011). AC3, in particular has been shown to affect transcriptional activation of a NAC transcription factor (Lozano-Duran *et al*., 2011). Overexpression of the NAC transcription factor was shown to cause enhanced viral replication as observed by specific viral interacting proteins from the geminivirus, *Tomato yellow leaf curl virus* (TLCV) in a yeast-two hybrid experiment (Selth *et al*., 2005 reviewed in Lozano-Duran *et al*.,2011).

Many strategies toward disease control have involved identifying geminivirus proteins which interact with a diverse set of host factors in *Arabidopsis thaliana*, *Solanum lycopersicum* and *Nicotiana benthamiana* (Jeske, 2009), however, these findings have been limited to single or a small subset of genes. Those studies encompassing large differentially expressed gene sets involved a whole genome microarray study with *Cabbage leaf curl virus* (CaLCuV) in *Arabidopsis* (Asencio-Ibáñez *et al*., 2008), serial analysis of gene expression (SAGE) in geminivirus-resistant cassava (Fregene *et al*., 2004), and a reverse genetics approach to identify genes involved in *Tomato yellow leaf curl virus* resistance (Eybishtz *et al*.,2009). Obtaining information from key transcripts and pathways at the molecular level thus provides insight into host genes responsible for regulating defence responses. “Interactomics”, is a new and rapidly increasing area in systems biology where networks governing signalling pathways, regulatory control as well as cellular function are assessed (Busch and Lohmann, 2007). High-throughput technologies such as next generation sequencing (NGS) has allowed for global analyses of exceptionally large datasets from transcriptomic, proteomic, metabolic, regulatory and developmental pathways to create networks that categorize interactions and function of organs or molecules at varying complexity levels (Ma *et al*, 2007). To date, the *Arabidopsis* interactome remains the model plant system of choice for comparative studies against largely unreferenced genomes such as cassava, as it is the most thoroughly studied organism with readily available community resources, providing interdisciplinary and multi-investigative resources (Koorneef and Meinke, 2010).

Cassava belongs to the family *Euphorbiaceae* in the Fabid superfamily, and its genome comprises an estimated 770 Mb (Awoleye *et al*., 1994) in N=18 chromosomes. With the availability of more sequenced plant genomes, especially in more closely related Fabid superfamily, such as *Ricinus communis* (castor bean) (Chen *et al*, 2010) and

Populus trichocarpa (Tuskan *et al*., 2006); more comparative genomic studies on natural hosts are emerging. A draft genome assembly and partial annotation of cassava from a single accession was released at the end of 2009 (Prochnik *et al*., 2012), with a total of 22.4 billion bp of raw data, enough to cover the genome ±29 times (Prochnik *et al*., 2012). The genome assembly is in 12, 977 scaffolds, with a total scaffold length of 532.5 Mb. Ninety six percent of the putative transcripts from the publically available cassava EST database (http://cassava.igs.unmaryland.edu/blast/db/EST_asmbl_and_single.fasta) can be mapped to the genome, making this a powerful tool for functional genomic studies. To date 30,666 protein-coding loci have been predicted (Prochnik *et al*., 2012).

In this study, whole genome transcript profiling using NGS (Applied Biosystems SOLiD™ Platform) was carried out in the susceptible cassava T200 cultivar to examine biotic stress-responsive genes, and differential expression of genes potentially involved in promoting systemic disease induced by SACMV-[ZA:99], at 3 time points post infection. In order to determine if similar global patterns or trends in differentially expressed genes occurred between hosts, comparisons were made to a previous study conducted by Pierce and Rey, 2013 using the susceptible *Arabidopsis*-SACMV-[ZA:99] pathosystem. In addition, a comparison of transcriptome patterns between T200 and a tolerant cassava cultivar TME3 was also assessed. Since large amounts of data was obtained in this study (supplementary tables), details and discussions are limited to selected genes associated with signalling and altered expression,and stess and cellular pathways of interest linked to virus-induced host responses. This is the first report identifying global differentially expressed transcripts in geminivirus-challenged cassava at three stages during the course of infection, namely, pre-symptom (14dpi), fully-susceptible (32 dpi) and late infection (67 dpi).

4.3 Materials and Methods

Plant Growth and Virus Inoculations

Cassava tissue culture plantlets, T200, an elite South African susceptible cultivar was obtained from the International Institute for Tropical Agriculture (IITA). T200 nodal cuttings were placed onto Murashige and Skoog basal salts containing vitamins (Murashige and Skoog, 1962) (Highveld Biological (PTY) LTD), supplemented with 20g sucrose and reduced agar (6.8g), and grown for approximately 1-2 weeks until roots and

shoots were visible. Plantlets were then transferred to trays containing Jiffy Peat Pellets (Jiffy Products International) and covered in plastic wrap for adaptation to chamber conditions. In order to acclimatize the plantlets, small razor-like slits were made in the plastic wrap on a daily basis for approximately 2 – 4 weeks to avoid air flow and introduce the plantlets to chamber conditions accordingly. Plants were maintained at 28°C under a 16h day at an intensity of 150 $\mu Em^{-2}sec^{-1}$.Cassava T200 plants were grown to the 4 -6 leaf stage and were co-inoculated with full-length head-to-tail SACMV DNA-A and DNA-B dimers (Berrie *et al*., 2001), mobilized into *Agrobacterium tumefaciens* strain AGL1 according to the improved agroinfection protocol outlined in detail in chapters 2 and 3 of this thesis.

SACMV Infection Validation (Total Nucleic Acid (TNA) Extraction and PCR)

Total nucleic acid (TNA) was extracted from SACMV-[ZA:99] - infected and mock-inoculated *Arabidopsis* plants according to the CTAB (cetyltrimethylammonium bromide) method of Doyle and Doyle (1987). Fifty milligrams of young leaf samples were ground in liquid nitrogen and TNA was extracted by the addition of 0.5 ml pre-heated CTAB extraction buffer (2% CTAB, 20 mM EDTA, 1.4 M NaCl, 100 mM Tris pH 8.0) and ß-mercaptoethanol (to a final concentration of 0.1% v/v). The aqueous layer containing the TNA was extracted using chloroform: isoamyl (24:1) in a two-step process and the nucleic acids precipitated with an equal volume of isopropanol. The pellet was then washed with 70% ice-cold ethanol, vacuum dried and resuspended in 50 µl 1 X TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) containing 20 µg/ml RNase A. PCR was carried out using core-coat protein (CCP) primers that amplify a 550 bp region on SACMV-[ZA:99] DNA-A genome component. CCP primers consisted of the following degenerate primer sequences:
Forward primer: (V524) 5’ GCCHATRTAYAGRAAGCCMAGRAT 3’, and
Reverse primer: (C1048) 5’ GGRTTDGARGCATGHGTACANGCC 3’.
Approximately 500 ng of TNA was added to each reaction consisting of 1 x DreamTaq Buffer (Fermentas), 200 µM dNTPs, and 2 U DreamTaq DNA polymerase (Fermentas) of which 0.4 µM of each primer was added, making up a final reaction volume of 50 µl. Amplification was carried out utilizing the MyCycler™ Thermal Cycler (Bio-Rad) with cycling conditions programmed for 1 cycle at 95°C for 5 min, followed by 35 cycles at 95°C for 30 sec, 55°C for 30 sec, and 72°C for 45 sec, this was followed by a final extension step for 5 min at 72°C.

Experimental Design, RNA extraction, Quantification and Quality Assessment

Total RNA was extracted from 6 biological replicates from SACMV-[ZA:99] (n=6) and mock-inoculated (n = 6) cassava T200 plants at 12, 32 and 67 dpi using a modified high molecular weight polyethylene glycol (HMW-PEG) procedure (Gehrig *et al*., 2000). For each biological replicate, 2 newly emerged, but fully expanded T200 leaves were used for pooling experiments at each time point, resulting in one SACMV-[ZA99] – infected and one mock-inoculated control sample at each time point. Briefly, one microgram of cassava leaf tissue was ground in liquid nitrogen with a pestle and mortar, and was added to a pre-heated (65°C) 5ml GHCL buffer containing 0.1g HMW-PEG (Mr: 20 000, Sigma). The samples were vortexed for 5 sec and centrifuged at 10 000 rpm for 10 min at 4°C. The supernatant was then transferred to a new tube where 0.1 ml (1M) sodium citrate (pH 4.0), 0.2 ml (2M) NaCl, and 5 ml phenol, chloroform, isoamyl alcohol (25:24:1) was added. The tubes were then shaken vigorously for 5 min at room temperature (RT) and centrifuged at 10 000 rpm for 10 min at 4°C. Five millilitres of isopropanol was then added to the supernatant in a new tube, mixed and incubated for 1 h at -20°C, and centrifuged at 10 000 rpm for 30 min at RT. The pellet was then washed with 75% ice-cold ethanol, and centrifuged at 10 000 rpm for 10 min at 4°C. Once the supernatant was discarded, the pellet was dried at 37°C, and 100 µl of pre-heated (55°C) RNase-free water was added to resuspend the pellet. The samples were then incubated at 55°C for 5 min to allow the RNA to dissolve. One microlitre of RNase inhibitor (Fermentas) was added and samples were quantified using the Thermo Scientific NanoDrop™ 1000 spectrophotometer. Stringent RNA quality control was carried out using the Agilent 2100 Bioanalyzer.

Next Generation Sequencing (NGS) of The Cassava Transcriptome

The generation of cassava sequence reads was carried out using the SOLiD™ v4 sequencer (Applied Biosystems) and was run in paired-end mode (50 + 35 bp). Two csFasta and two quality files were generated, and reads for each llibrary were mapped to the genome assembly (Phytozome, *Manihot esculenta* 147) using the Lifescope software from LifeTech. Samtools (http://samtools.sourceforge.net/) was used to prepare, sort and index SAM/BAM alignment files. The BAM data was then aligned with the genome annotations available in Phytozome as a GTF/GFF3 file, describing genes, transcripts and their exons with the genomes coordinates. rnaSeqMap library of Bioconductor **(**Leśniewska and Okoniewski, 2011) was used to transform alignments to counts, and

counts from all genes annotated were analyzed using DESeq (Anders and Huber, 2010). In addition, in order to find probable regions of novel transcription, unknown to the curators (of the annotations) in Phytozome, the procedure for finding significant expression regions for intergenic spaces was also performed.

Functional Analysis Tools

Cassava scaffolds were mapped and aligned to *Arabidopsis thaliana* sequences in Phytozome (www.phytozome.net) in order to assign putative identities to cassava T200 transcripts of unknown function. In addition, The *Arabidopsis* Information Resource (TAIR) (www.arabidopsis.org) database was utilized for these purposes, as it is also a web-based retrieval system. To select for highly expressed transcripts, a 2.5 Log2 fold cut-off with a p-value of 0.05 was applied to the data. In order to classify and assign transcripts into functional groups, analysis tools from the Munich Information of Protein Sequences (MIPS) (http://mips.gsf.de/proj/thal/db/Arabidopsis) was used. By assigning genes into groups, transcripts involved in disease, virulence and defence were identified. Heat Maps were constructed using TIGR Multi Experiment Viewer (MEV) (www.tigr.org/software) by applying Pearson correlation to identify unique and common patterns among differentially expressed transcripts in susceptible and resistant cassava cultivars ($p < 0.05$).

4.4 Results

Infectivity Assay of SACMV-[ZA:99] - Infected CassavaT200

SACMV-[ZA:99] induced symptoms were assessed over a time course of 67 dpi. Symptoms started to appear at 12 dpi and cassava T200 plants were shown to be fully symptomatic at 32 dpi. Severe symptoms such as leaf distortion, deformation, leaf curl, yellow mosaic, and overall stunting of leaves was observed in SACMV-[ZA:99] infected cassava T200 leaves (Figure 1 B,C), compared to mock-inoculated control leaves (Figure 1 A). The appearance and severity of symptoms correlated with an increase in virus titre as SACMV-[ZA:99] replication levels were 1.4E+00 molecules of DNA/ng of total plant nucleic acid (TNA) at 12 dpi, 2.19E+03 molecules of DNA/ng of TNA at 32 dpi, and 4.43E+05 molecules of DNA/ng of TNA at 67 dpi (Van Schalk and Rey, unpublished).

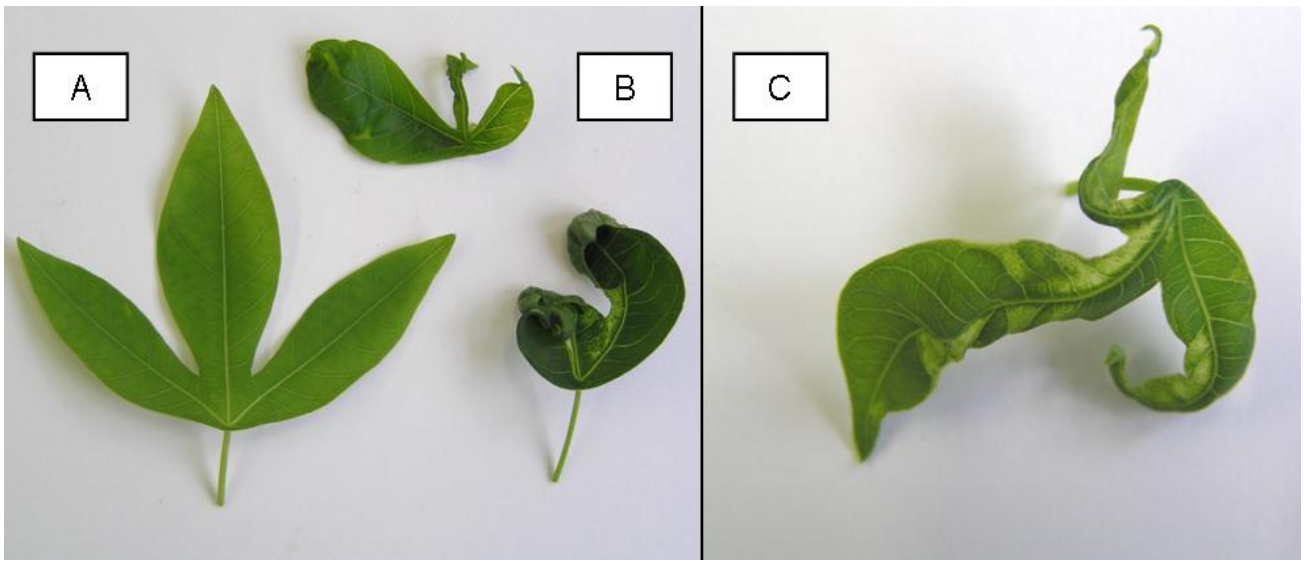


Figure 1: A) Mock – inoculated cassava T200 leaf displaying no symptoms. B and C) SACMV-[ZA:99] – infected cassava T200 leaves displaying typical geminivirus symptoms such as mosaic, leaf curl, deformation, and reduction in size

Functional Categorization of Global Differentially Expressed Transcripts Across Time Points 12, 32, and 67 dpi

Cassava T200 scaffolds were assembled in Phytozome and putative annotations were assigned using the model plant system, *Arabidopsis thaliana*. This model plant system was used in a previous whole genome microarray study for identification of differentially expressed susceptible genes upon SACMV-[ZA:99] infection (Pierce and Rey, 2013).

A 2.5 log2-fold filter was applied to the data ($p < 0.05$) and differentially expressed genes were assigned according to the *Arabidopsis* MIPS (Munich Information Centre for Protein Sequence) functional classification scheme (http://mips.gsf.de/proj/thal/db/Arabidopsis).

Time point comparisons revealed 225 differentially expressed transcripts at 12 dpi (138 induced and 87 repressed), 633 at 32 dpi (175 induced and 458 repressed), and at 67 dpi, 634 genes differentially expressed genes (322 induced and 312 repressed) were identified (Figure 2A). A complete list of differentially expressed genes from susceptible cassava T200 at each time point may be viewed in supplementary data_TableS1. Based on Fisher's exact test (Fisher 1970), putative functions for 24 functional categories were assigned. Results from the MIPS functional category revealed a similar trend in expression patterns between platforms (microarray and NGS) and two susceptible host systems (Cassava T200 and *Arabidopsis*) upon SACMV-[ZA:99] infection (Figure 2A and B). A shift

from gene induction to repression was evident in both pathosystems for the early to middle time points. The majority of transcripts (61%) were induced at 12 dpi, but at 32 dpi, most transcripts (72%) were reduced. At 67 dpi, the number of differentially expressed transcripts was approximately even in both plant hosts (51% up-regulated and 49% down-regulated). In the *Arabidopsis*-SACMV-[ZA:99] pathosystem, a similar trend in differentially expressed transcripts was observed, 51% were induced at 14 dpi, 53% repressed at 24 dpi, however, a significantly larger number of transcripts were repressed at 36 dpi (72%). Correlations between time points 32 dpi in cassava T200 and 36 dpi in *Arabidopsis* were evident, as shown by the large amount (72%) of repressed transcripts, at 32 and 36 dpi, respectively. In addition, both pathosystems displayed similar trends in gene function, as categories containing the highest number of transcripts were the same (Figure 2A and B). For example, most transcripts fell into categories such as metabolism, cell cycle and DNA processing, transcription, protein with binding function or cofactor requirement (structural or catalytic), and subcellular localization indicating stress-specific responses upon pathogen (SACMV-[ZA:99]) infection, regardless of the host and platform in question.

In order to establish if a shift from general growth and development to defence would prevail, the top three differentially expressed categories at each time point were identified, this included, metabolism, subcellular localization, and protein with binding function or cofactor requirement (structural or catalytic) (Table 1). Classification patterns revealed an increasing shift between time points where the majority of transcripts involved in metabolism, subcellular localization, and protein with binding function or cofactor requirement (structural or catalytic) were up-regulated at early stages of infection (at 12 dpi), whereas at 32 and 67 dpi, more transcripts in those categories were down-regulated (Table 1), indicating that a definite shift from induced to repressed responses was evident.

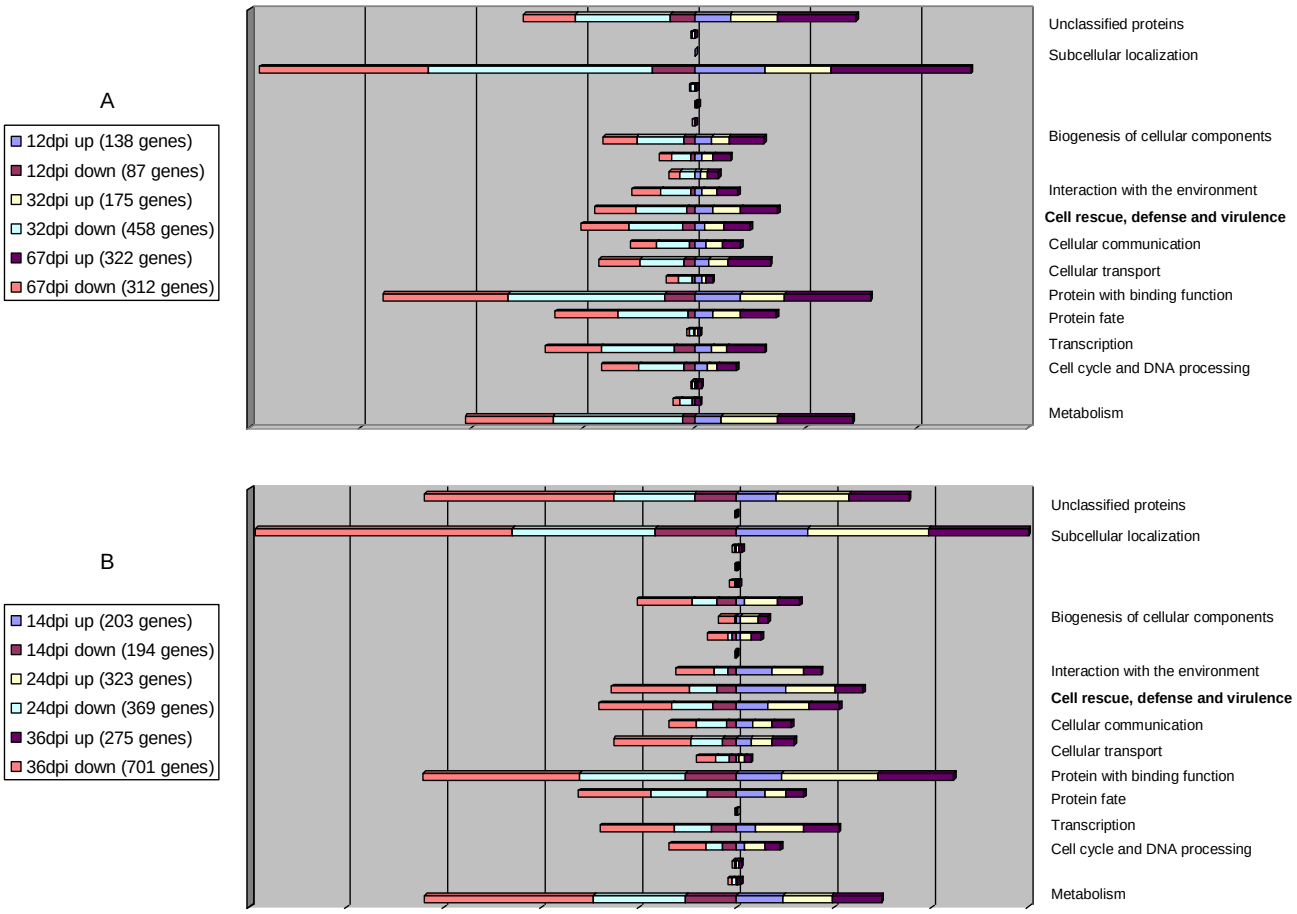


Figure 2: MIPS distribution of 2.5 $\log_2$ fold differentially expressed genes across 3 time points, displaying similar trends in assigned functional categories between two susceptible host systems, A) cassava T200, and B) *Arabidopsis thaliana* upon SACMV-[ZA:99] infection

Table 1: Classification of transcripts identified from MIPS displaying the top 3 functional categories at time points 12, 32 ,and 67 dpi in cassava T200 upon SACMV-[ZA:99] infection

Time Point (Total transcript number)	**Functional Category** (Transcript number)		
	Subcellular localization	**Protein with binding function**	**Metabolism**
12dpi			
Up (138)	63 (45.7%)	41 (29.7%)	24 (17.4%)
Down (87)	38 (43.7%)	27 (31%)	11 (12.6%)
32 dpi			
Up (175)	60 (34.3%)	40 (22.9%)	51 (29.1%)
Down (458)	201 (43.9%)	140 (30.6%)	116 (25.3%)
67 dpi			
Up (322)	125 (38.8%)	77 (23.9%)	67 (20.8%)
Down (312)	152 (48.7%)	113 (36.2%)	79 (25.3%)

Within context to the study, the category, cell rescue, defence, and virulence was assigned by the MIPS program for functional transcript classification. We are aware that plants themselves cannot be virulent but this category title is assigned by default in MIPS and for ease of comparisons between studies and for further functional analyses we chose to keep selected category names standard. Differentially expressed genes were thus assigned to stress response categories in MIPS in the following manner :- virulence and defence; oxidative stress; electromagnetic waves; heat shock; cold shock; osmotic and salt stress, and other sub-categories such as glutathione conjugation and peroxidase reaction. An interesting observation from Table 2 revealed that of the 75 cassava T200 scaffolds involved in defence responses, approximately 68% were down-regulated. Those transcripts showing down-regulation were many disease resistance-associated proteins such as the disease resistance protein (TIR-NBS-LRR) class, RPP1, RPM1, MLO2,

MLO12 and NHO1 resistance proteins. Also among the repressed transcripts were transcription factors such as WRKY, as well as HSP's and heat shock factors which are involved in defence (Table 2). In addition, transcripts such as ERF5 and JAR1 involved in phytohormone signalling were also repressed. Metabolic genes such as those involved in the phenylpropanoid metabolic process (4CL3 and CHS), alcohol dehydrogenase 1, class 1 glutamine amidotransferase-like superfamily protein, and phenylalanine ammonia lyase 1 and 2 were up-regulated.

Table 2: Disease, virulence, and defence genes identified in cassava T200 upon SACMV-[ZA:99] infection across 3 time points post infection (12, 32 and 67dpi)

Arabidopsis Accession no.	**Gene Name**	**Time Point**	**Expression (Up/Down-regulated)**	**Stress Response (MIPS)**
AT1G65060	4-coumarate:coa ligase 3 (4cl3)	32, 67	Up	Electromagentic waves
AT4G33300	adr1-like 1 (adr1-l1)	67	Down	Disease,virulence and defence
AT1G77120	alcohol dehydrogenase 1 (adh1)	67	Up	Osmotic and salt stress
AT5G28540	bip1	32	Down	Heat shock
AT4G34150	calcium-dependent lipid-binding (calb domain) family protein	32,67	Down	Cold shock
AT3G07370	carboxyl terminus of hsc70-interacting protein (chip)	32,67	Down	Osmotic and salt stress
AT4G18780	cellulose synthase 8 (cesa8); irregular xylem 1 (irx1); leaf wilting 2 (lew2)	67	Up	Osmotic and salt stress
AT3G55120	chalcone flavanone isomerase (cfi); chalcone isomerase (chi); transparent testa 5 (tt5)	32	Up	Electromagentic waves
AT5G13930	chalcone synthase (chs); transparent testa 4 (tt4)	32,67	Up	Oxidative stress
AT2G23970	class i glutamine amidotransferase-like superfamily protein	32,67	Down	Disease,virulence and defence
AT5G07990	cytochrome p450 75b1 (cyp75b1); transparent testa 7 (tt7)	12,32,67	Up	Electromagnetic waves
AT4G23690	dirigent protein 6 (dir6)	32,67	Up	Disease,virulence and defence
AT2G34930	disease resistance family protein / lrr family protein	32	Down	Disease,virulence and defence

AT1G69550	disease resistance protein (tir-nbs-lrr class)	32,67	Down	Disease,virulence and defence
AT3G04220	disease resistance protein (tir-nbs-lrr class) family	32,67	Down	Disease,virulence and defence
AT4G12010	disease resistance protein (tir-nbs-lrr class) family	67	Down	Disease,virulence and defence
AT5G18350	disease resistance protein (tir-nbs-lrr class) family	32,67	Down	Disease,virulence and defence
AT5G17680	disease resistance protein (tir-nbs-lrr class), putative	32,67	Down	Disease,virulence and defence
AT1G58170	disease resistance-responsive (dirigent-like protein) family protein	12	Up	Diseases,virulence and defence
AT1G65870	disease resistance-responsive (dirigent-like protein) family protein	32	Up	Disease,virulence and defence
AT3G12610	dna-damage repair/toleration 100 (drt100)	32,67	Up	Disease,virulence and defence; Electromagnetic waves; Resistance proteins
AT1G24620	encodes a ef-hand calcium-binding protein family member	67	Up	Cold shock
AT1G70250	encodes a protease inhibitor/seed storage/ltp family protein	67	Down	Disease,virulence and defence
AT5G47230	ethylene responsive element binding factor 5 (erf5)	12,32,67	Down	Cold shock
AT5G20630	germin-like protein 3 (ger3,glp3)	12,67	Up	Cold shock
AT5G41210	glutathione s-transferase theta 1 (gstt1)	32	Up	Glutathione conjugation reaction
AT1G16030	heat shock protein 70b (hsp70b)	12,32	Down	Heat shock
AT2G26150	heat shock transcription factor a2 (hsfa2)	32	Down	Heat shock; Oxidative stress
AT1G07400	hsp20-like chaperones superfamily protein	67	Down	Heat shock; Oxidative stress
AT1G53540	hsp20-like chaperones superfamily protein	67	Up	Heat shock
AT2G46370	jasmonate resistant 1 (jar1); far-red insensitive 219 (fin219)	67	Down	Oxidative stress
AT4G02380	late embryogenensis abundant like 5 (lea5); senescence-associated gene 21 (sag21)	12,67	Down	Cold shock; Oxidative stress

AT1G17420	lipoxygenase 3 (lox3)	12,32,67	UP,Down, Down	Diseases,virulence and defence
AT2G02100	low-molecular-weight cysteine-rich 69 (lcr69); pdf2.2	67	Up	Disease,virulence and defence
AT2G39200	mildew resistance locus o 12 (mlo12)	32	Down	Disease,virulence and defence
AT1G11310	mildew resistance locus o 2 (mlo2); powdery mildew resistant 2 (pmr2)	32	Down	Disease,virulence and defence
AT3G45640	mitogen-activated protein kinase 3 (mapk3, mpk3)	32,67	Down	Oxidative stress; Osmotic and salt stress; Cold shock
AT3G13080	multidrug resistance protein 3 (mrp3); atp-binding cassette c3 (abcc3)	67	Down	Glutathione conjugation reaction
AT2G31180	myb domain protein 14 (myb14)	32,67	Down	Osmotic and salt stress
AT3G23250	myb domain protein 15 (myb15)	32,67	Down	Osmotic and salt stress
AT1G22640	myb domain protein 3 (myb3)	32,67	Up	Osmotic and salt stress
AT2G16720	myb domain protein 7 (myb7)	12	Up	Osmotic and salt stress
AT2G35980	ndr1/hin1-like 10 (nhl10); yellow-leaf-specific gene 9 (yls9)	12,32,67	Down	Diseases,virulence and defence
AT1G80460	nonhost resistance to p. s. phaseolicola 1 (nho1); gli1	32	Down	Disease,virulence and defence
AT3G12500	pathogenesis-related 3 (pr3); basic chitinase (chi-b)	67	Down	Disease,virulence and defence
AT3G11820	penetration1 (pen1); syntaxin of plants 121 (syp121); syntaxin related protein 1 (syr1)	32,67	Down	Disease,virulence and defence
AT1G14550	peroxidase superfamily protein	32	Down	Oxidative stress
AT2G34060	peroxidase superfamily protein	67	Up	Oxidative stress; Peroxidase reaction
AT2G37130	peroxidase superfamily protein	67	Down	Disease,virulence and defence; Peroxidase reaction
AT2G39040	peroxidase superfamily protein	32	Down	Oxidative stress
AT2G41480	peroxidase superfamily protein	32	Down	Oxidative stress
AT4G16270	peroxidase superfamily protein	32	Down	Oxidative stress
AT5G39580	peroxidase superfamily protein	32	Down	Disease,virulence and defence

AT2G37040	phe ammonia lyase 1 (pal1)	32,67	Up	Oxidative stress; Disease,virulence and defence
AT3G53260	phenylalanine ammonia-lyase 2 (pal2)	32,67	Up	Oxidative stress; Disease,virulence and defence
AT1G29340	plant u-box 17 (pub17)	67	Down	Disease,virulence and defence
AT4G35100	plasma membrane intrinsic protein (pip2;7,pip3;pip3a)	67	Up	Osmotic and salt stress
AT1G61190	putative cc-nb-lrr resistance gene	67	Down	Disease,virulence and defence
AT1G05260	rare cold inducible gene 3 (rci3)	32	Up	Oxidative stress; Osmotic and salt stress; Peroxidase reaction
AT3G44480	recognition of peronospora parasitica 1 (rpp1); cog1	32,67	Down	Disease,virulence and defence
AT3G07040	resistance to p. syringae pv maculicola 1 (rpm1); resistance to pseudomonas syringae 3 (rps3)	32	Up	Disease,virulence and defence
AT3G07040	resistance to p. syringae pv maculicola 1 (rpm1); resistance to pseudomonas syringae 3 (rps3)	67	Down	Disease,virulence and defence
AT3G07040	resistance to pseudomonas syringae 3 (rps3); resistance to p. syringae pv maculicola 1 (rpm1)	32	Down	Disease,virulence and defence
AT5G37260	reveille 2 (rve2); circadian 1 (cir1)	12	Down	Osmotic and salt stress
AT1G66350	rga-like 1 (rgl1)	12,67	Down	Osmotic and salt stress
AT1G27730	salt tolerance zinc finger (stz); zat10	12,32,67	Down	Osmotic and salt stress; Cold shock
AT2G40140	salt-inducible zinc finger 2 (szf2); czf1	32,67	Down	Disease,virulence and defence; Cold shock
AT1G60940	sucrose nonfermenting 1-related protein kinase 2-10 (snf1-related protein kinase 2.10) (snrk2-10)	32	Up	Osmotic and salt stress
AT3G14440	sugar insensitive 7 (sis7); salt tolerant 1; nine-cis-epoxycarotenoid dioxygenase 3	32	Down	Osmotic and salt stress

	(nced3)			
AT5G44510	target of avrb operation1 (tao1)	67	Down	Disease,virulence and defence
AT3G16720	toxicos en levadura 2 (tl2)	12,32,67	Down	Diseases,virulence and defence
AT1G15690	v-ppase 3 (avp-3), avp1, atvhp1;1, fugu5	32	Up	Osmotic and salt stress
AT2G38470	wrky dna-binding protein 33 (wrky33)	32,67	Down	Disease,virulence and defence
AT1G80840	wrky dna-binding protein 40 (wrky40)	32,67	Down	Disease,virulence and defence
AT3G56400	wrky dna-binding protein 70 (wrky70)	67	Down	Disease,virulence and defence
AT5G57560	xyloglucan endotransglucosylase/hydrolase 22 (xth22); touch 4 (tch4)	32	Down	Cold shock

Susceptibility genes in cassava T200 were then further analysed at 2.5 log2 fold cut-off at $p < 0.05$ in order to determine which genes were differentially expressed over 67 dpi during the course of infection. Common genes across time points 12, 32, and 67 dpi were identified (Supplementary data_TableS3), and putative annotation and function for each transcript was identified in Phytozome, TAIR and MIPS. Induced transcripts such as pectin-lyase superfamily proteins and plant invertase/pectin methylesterase inhibitor superfamily proteins, involved in cell wall degradation were induced. Addtionally, transcripts involved in secondary metabolism such as serine carboxypeptidase-like 45 and those involved in protein/peptide degradation such as eukaryotic aspartyl protease family proteins which are involved in protein/peptide degradation were also up-regulated across time points. Transport genes showing differential expression were those genes involved in cation transport such as the up-regulated potassium transporter 2 protein whereas the heavy metal transport/detoxification superfamily protein was down-regulated across the 3 time points. Sugar transport proteins such as the major facilitator superfamily protein were up-regulated whereas Cytochrome P450, family 71, subfamily B,polypeptide 37 and Cytochrome P450, family 76, subfamily G,polypeptide 1, all involved in electron transport were down-regulated across all three time points.

Common up-regulated gene transcripts in cassava T200 across 3 time points, involved primarily in metabolism, were EMB3004, MEE32 (dehydroquinate dehydratase/

shikimate dehydrogenase) and UGT84A1 which are involved in c-compound and carbohydrate metabolism. In addition, genes such as EMB3004, MEE32 and CYP75B1, D501, TT7, involved in secondary metabolism were induced across time points. Genes such as HAD and PERK10 that are involved in phosphate metabolism were down-regulated across time points. HAD was also shown to be involved in metabolism of energy reserves such as glycogen and trehalose. Transcription factors such as WRKY41, signalling molecules such as ERF5, and other defence response genes such as TL2 and LEA were also repressed (Supplementary data_TableS3). A very interesting finding was the up-regulated Cyclin P4:1 gene which is involved in cell cycle and DNA processing, as Geminiviruses have been shown for to interfere with cell cycling in a host (Ascencio-ibanez *et al*, 2008; Andrietta, 2001).

Within host comparisons were then performed to identify similar patterns of genes differentially expressed in susceptible cassava T200 and tolerant TME3 (P-value 0.05). The complete list may be viewed in Supplementary data_Table S2. Pearson correlation using K-means average linkage clustering was carried out in order to group genes and experiments into 10 clusters according to their expression patterns. Significant differences in expression patterns between T200 and TME3 were observed but similarities were also evident. Three interesting clusters were chosen for further analyses, two of which displayed different expression patterns, while the 3rd cluster showed a similar pattern (Figure 3) between T200 and TME3 cultivars. Cluster 1 contained 26 genes (10%) and displayed transcripts predominantly appearing at 32 and 67 dpi. These transcripts were up-regulated in tolerant TME3 but down-regulated in susceptible T200 (Figure 3, cluster 1). This cluster contained many signalling and transcription factor type genes. Transcripts such as those involved in JA (JAZ1, JAZ8) and ET (ERF1, ERF4) signalling were present in this group. In addition, signalling molecules such as Redox response proteins, LEA5, and Cytochrome 450 as well as transcription factors such as NAC, WRKY, bHLH, and MAPK19 were also identified in this cluster.

Twenty-nine genes were identified in cluster 2 (11%) (Figure 3, cluster 2) with transcripts displaying opposite expression patterns to cluster 1. Genes at 32 and 67 dpi were repressed in TME3 but induced in T200. Transcripts which fell into this category were pectin-lyase proteins, protein kinases, major facilitator superfamily protein, and peroxidises, many of which are involved in cell wall defences. Cluster 3 (Figure 3, cluster 3) contained 17 genes (6%) and exhibited similar patterns of expression (up- or down-)

between TME3 and T200. Transcripts exhibited up-regulation at 32 dpi and down-regulation at 67 dpi for both cultivars. This cluster contained many defence-type responsive genes such as subtilisin-like serine endopeptidase, PAR1, pentatricopeptide repeat (PPR) superfamily protein, and the transcription factor family protein, AP2/B3 which has a regulatory function in ET signalling.

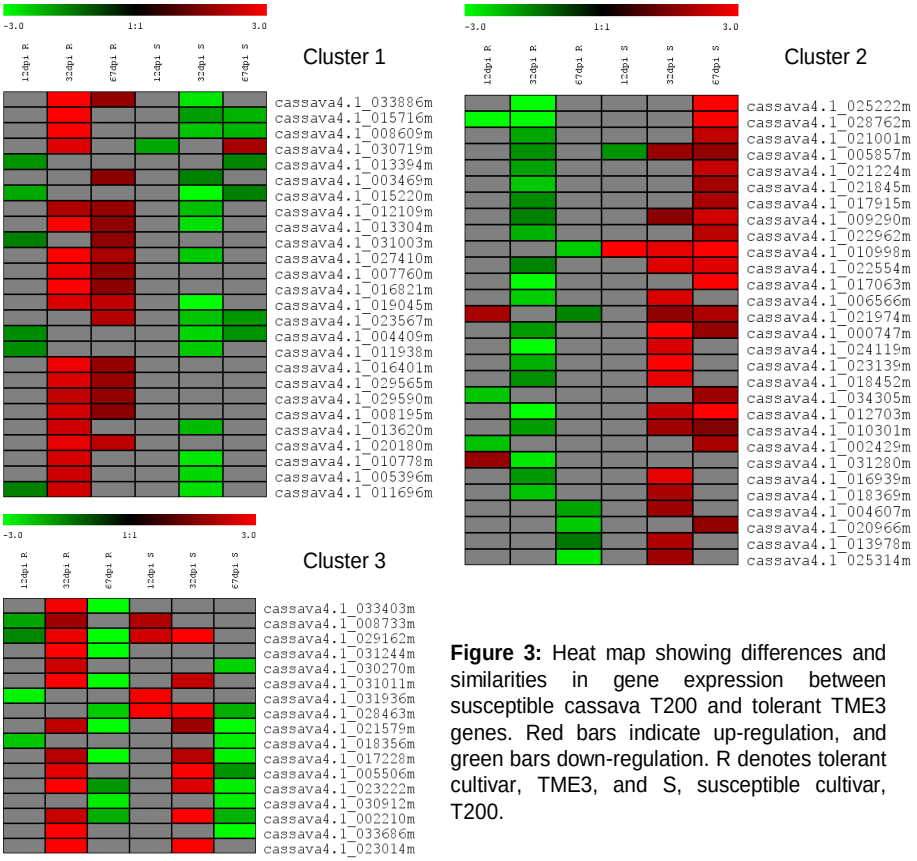


Figure 3: Heat map showing differences and similarities in gene expression between susceptible cassava T200 and tolerant TME3 genes. Red bars indicate up-regulation, and green bars down-regulation. R denotes tolerant cultivar, TME3, and S, susceptible cultivar, T200.

Cassava T200 Cell-Cycle Genes

A pre-requisite to geminivirus replication is reprogramming of plant gene expression, inducing quiescent cells to re-enter the cell cycle to support DNA replication (Hanley-Bowdoin *et al*., 2004). The cassava genome is annotated at an estimated ± 65%, indicating that the total number of cyclin genes identified in the published sequence database in Phytozome is not known, and the functionality of identical cyclins (BLASTED against *Arabidopsis* for annotation) on different scaffolds have thus not been elucidated.

In this study, only 11 differentially expressed core cyclin transcipts were identified in T200 (Table 3), compared to *Arabidopsis* where 44 of the 61 core cell cycle genes were altered in response to SACMV-[ZA:99] (previous chapter 3). Of the 11 core cyclin genes identified, 9 were up-regulated and 2 down-regulated in T200. Additionally, there was little correlation in the up and down gene expression patterns between the two hosts at different time points post- inoculation (Table 3). Only CYCP4;1 was up-regulated in both susceptible hosts, at all time points in T200, and at 14 and 36 dpi in *Arabidopsis.*

Table 3: Cyclin genes identified in cassava T200 and compared to *Arabidopsis* cyclin genes following SACMV:[ZA:99] infection

Cassava ID	**Log2 Fold Change**	**P value (0.05)**	***Arabidopsis* Acc**	**Cyclin Gene**	**Cassava T200 Time Point (dpi)**	**Expression**	***Arabidopsis* Time Point (dpi)**	**Expression**
Cassava 4.1_008064m	3.42	1.17E-02	AT1G76310	CYCB2;4	12, 67	Up	24	Down
Cassava 4.1_009258m	-2.70	2.45E-02	AT5G43080	CYCA3;1	32	Down	14	Up
Cassava 4.1_009524m	2.92	3.51E-02	AT1G16330	CYCB3;1	12	Up	24	Down
Cassava 4.1_009919m	1.55	5.16E-02	AT4G34160	CYCD3;1	12	Up	14, 24	Down
Cassava 4.1_010147m	1.59	5.20E-02	AT5G67260	CYCD3;2	67	Up	14, 24	Down
Cassava 4.1_016519m	2.51	3.25E-02	AT2G44740	CYCP4;1	12, 32, 67	Up	14, 36	Up
Cassava 4.1_020415m	-3.05	5.17E-03	AT1G27630	CYCT1;3	67	Down	24	Up
Cassava 4.1_021243m	2.32	3.33E-02	AT4G37630	CYCD5;1	12	Up	14	Down

Cassava 4.1_026 079m	2.30	5.03E-02	AT1G44110	CYCA1;1	12	Up	24, 36	Down, Up
Cassava 4.1_032 656m	3.72	9.00E-05	AT5G11300	Mitotic-like cyclin 3B from *Arabidopsis*	32, 67	Up	24	Down

4.5 Discussion

In order to map the interaction between SACMV-[ZA:99] and its natural host, cassava, a study involving whole genome transcriptomics was conducted in a susceptible South African cassava elite cultivar, T200, in order to identify putative genes involved in virus replication and movement (pathogenicity), and basal defence and signalling pathways during the disease process, over 67 days post agroinoculation. The purpose of this study was to understand host responses to a geminivirus, in particular SACMV-[ZA:99], towards informing decisions for future engineering of cisgenic cassava for broad spectrum resistance. SACMV-[ZA:99] – infected tolerant cultivar TME3, a Nigerian landrace (Dixon *et al*., 2001; Fregene *et al*., 2001a), was used for comparative purposes, and appropriate selective comparisons with a study in a model host, *Arabidopsis thaliana* using a 4 x 44K Agilent microarray, is also presented. Analysis of differential transcript alterations demonstrated several basal defence reponses elicited in both susceptible (T200) and tolerant (TME3) cultivars, but differences in patterns and timing was clearly evident. In T200, large numbers of transcripts involved in basal immunity were downregulated, especially at 32 dpi (full systemic infection), which resulted in persistent virus infection and susceptibility. Some similar and different patterns in defence-related gene expression between T200 and *Arabidopsis* were noted, both being susceptible to SACMV-[ZA:99], but in TME3 supression of many transcripts appeared at an earlier time point, 12 dpi, which suggests a rapid response to SACMV-[ZA:99].

Defence and Signalling Pathways

In general, a shift from up-regulated genes at an early time point (12 dpi), to down-regulated genes from middle (32 dpi) to late (67 dpi) time points (Figure 2, Table 1) was evident in susceptible cassava T200. This phenomenon is not uncommon when large

amounts of virus nucleic acid and proteins are produced during viral replication in plants tissues, causing normal cellular processes to be redirected toward viral replication (Bolton, 2009; Culver and Padmanabhan, 2007; Whitham *et al*., 2006). From the infectivity assay, an increase in virus titre was observed from 12 dpi (1.4E+00 molecules of DNA/ng of TNA) to 32 dpi (2.19E+03 molecules of DNA/ng of TNA) and 67 dpi (4.43E+05 molecules of DNA/ng of TNA) in cassava T200, strongly suggesting, that an increase in virus titre may correlate with host gene suppression. A study by Pierce and Rey, 2013 (chapter 3), using the *Arabidopsis*-SACMV-[ZA:99] pathosystem also confirmed these findings. It was evident that in cassava T200 at 32 and 67 dpi and in *Arabidopsis* at 24 and 36 dpi, both systems represented high virus replication levels and full systemic infection. SACMV-[ZA:99] titre levels was observed to increase from 1.09 x 10^4 copies at 14 dpi, 5.75 x 10^4 copies at 24 dpi, and 6.30 x 10^4 copies at 36 dpi in *Arabidopsis* (reviewed in chapter 3), and plants were shown to be fully symptomatic at 24 dpi. The higher SACMV-[ZA:99] replication levels observed in cassava T200 could be attributed to the fact that T200 is a natural host to SACMV-[ZA:99], providing a more favourable replication-competent environment. It was also evident that SACMV-[ZA:99] was able to maintain repression at later time points as virus infection persisted. Viruses have been shown to cause host gene shut-off in an attempt to inhibit broad spectrum defence responses activated by the plant (Havelda and Maule, 2000; Maule 2002; Pallas and Garcia, 2011). Although host shut-off was previously described as transient, more recently, Conti *et al*., 2012 demonstrated that gene-specific and persistent shut-off was evident in *Nicotiana tabacum* upon *Tobacco mosaic virus* (TMV) infection. Similarly, this was also evident in both *Arabidopsis*-SACMV-[ZA:99] (Pierce and Rey, 2013) and cassava T200-SACMV-[ZA:99] (this study) pathosystems, as shown by persistent down-regulation of many genes across 3 time points (Table 2, Table 3, and Supplementary Tables S1 and S2).

General findings in the study revealed that extensive overlap between SACMV-[ZA:99] stress and abiotic stress responses (oxidative stress, heat shock, cold shock, and wounding) existed (Oktem *et al*., 2008; AbuQamar *et al*., 2009; Walley *et al*., 2007; An *et al*., 2012). Biotic and abiotic stresses cause a reduction in plant growth and yield, and in order for a plant to survive, signalling pathways and molecules are activated to mount appropriate defence responses against these stresses (Oktem *et al*., 2008). With reference to biotic stresses in particular, upon pathogen recognition, basal innate immune responses are intitiated, whether the plant is a susceptible or resistant host. Specific induced defence mechanisms are triggered following the primary response depending if

there is direct recognition between compatible pathogen avirulence or pathogenicity gene products and host resistance genes (R) genes. SACMV-[ZA:99] was shown to cause transcriptional reprogramming in cassava T200 which caused activation of basal broad defences such as ROS and the oxidative burst, and functional defence proteins such as PR proteins, heat shock proteins (HSP's) and secondary metabolites were also affected. Hormone (JA, SA, and ET) signalling pathways responsible for regulating these functional proteins as well as transcription factors were also showing differential patterns of expression (mainly down-regulation) demonstrating the severity that SACMV-[ZA:99] infection impacts on it's susceptible host cultivar, cassava T200. The signalling molecules and pathways affected by SACMV-[ZA:99] infection in T200 may be viewed in a basic schematic diagram in Figure 4.

However despite up-regulation of several basal immunity transcripts, T200 remains susceptible to SACMV-[ZA:99], as these responses are not efficient to prevent virus movement and replication. As previously mentioned, cassava is a vegetatively propagated perennial and virus persistence and disease occurs throughout the life-cycle of the plant until it is harvested, which is different to *Arabidopsis* and other annual seed model crops. One would thus expect a continuous fluctuation in cassava T200 responsive genes as SACMV-[ZA:99] spreads systemically to new apical leaves, where geminiviruses prefer to replicate (Jeske, 2009). Transcriptome changes in T200, were consistently monitored in upper leaves below the apex, where SACMV-[ZA:99] is actively replicating in order to avoid inconsistencies across older leaves and to minimize spatial variations. Therefore, there would be cyclic activation and suppression of responses and interactions as virus and host co-exist; the "arms" race where host attempts to mount a basal defence and the geminivirus overcomes this by suppression. This study of a natural host, cassava, infected with a geminivirus, SACMV-[ZA:99] further illustrates that in addition to virus-spcific responses, general responses to different viruses in susceptible hosts occur (Whitham *et al*., 2003; Whitham *et al*., 2006). General responses which include genes such as BGL2 (ß-1,3-glucanase) or heat shock proteins involved in defence (although not always exclusively involved in defence), have been shown to require signalling hormones such as salicylic acid (SA) for these purposes, which demonstrates and confirms that compatible host-virus interactions elicit basal defence pathways (Huang *et al*., 2005).

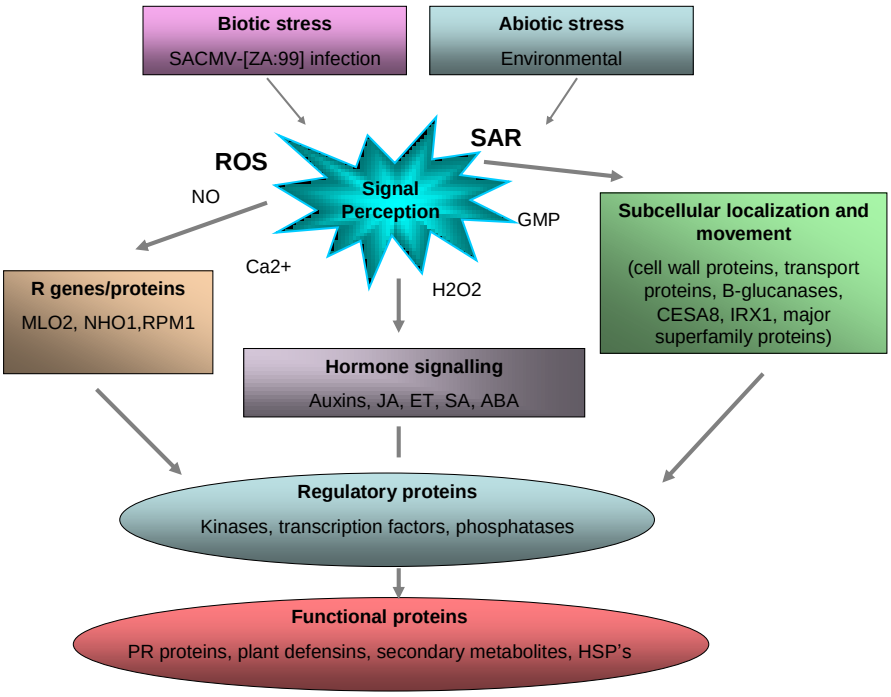


Figure 4: Schematic diagram depicting signalling molecules and pathways activated in response to SACMV-[ZA:99] infection

ROS and Hormone Signalling

Reactive oxygen species (ROS) are considered to be hazardous by-products of metabolism, but also show diverse roles as signalling molecules. Many pathogens activate plasma membrane-localized NADPH and cell-wall peroxidases to produce apoplastic ROS (Blomster *et al*., 2011). Seven peroxidase superfamily proteins were identified by MIPS classification within the disease, virulence, and defence category in cassava T200 at a 2.5 $\log_2$ fold cut-off (Table 2), 6 of which were showing down-regulation, 5 (AT1G14550, AT2G39040, AT2G41480, AT4G16270,AT5G39580) at 32 dpi and 1 (AT2G37130) at 67 dpi. The only up-regulated peroxidase (AT2G34060) was found at 67 dpi. Matching cassava T200 scaffolds to the above-listed *Arabidopsis* accession numbers may be viewed in Supplementary data_Table S1. Repression of peroxidases (87%) suggested an inhibitory effect that SACMV-[ZA:99] may be exerting on proteins involved in the activation of signalling pathways in defence, in order to promote its own systemic infection. Suppression of the auxin signalling hormone was also evident in this study. Auxin has been shown to mediate pathogen and abiotic stress tolerance by SA antagonism (Blomster *et al*., 2011). An auxin - like resistant protein (AT2G21050) was identified in cassava T200 showing down-regulation at 67 dpi as well as in TME3 at 12 dpi

(Supplementary data_Table S2). In addition, a thioredoxin superfamily protein (AT1G28480) was also suppressed in T200 at 12, 36, and 67 dpi, and at 12 dpi in TME3 (Supplementary data_Table S2) indicating that SACMV-[ZA:99] could possibly be disrupting auxin homeostasis and development. Studies revealing interactions between auxins and oxidative stress containing mutant thioredoxins and glutathiones have shown disruption to plant homeostasis and development (Bashandy *et al*., 2010).

Systemic Acquired Resistance (SAR) and JA Signalling

A report conducted by Truman *et al*, 2007 suggested that SAR transcriptional responses have a strong similarity to basal defences, and JA is central to systemic defences as seen by the activation of SAR (Truman *et al*., 2007). *Arabidopsis* systemic immunity uses conserved defence signalling pathways and is induced by jasmonates. In addition, the JA signalling pathway is common to both abiotic and biotic stress signals as seen by an overlap between local herbivory and wounding responses. In the previous *Arabidopsis*-SACMV-[ZA:99] study (Pierce and Rey, 2013, chapter 3), JA was shown to act synergistically with other signalling pathways, illustrating it's activation during defence. In this study, we found JAZ1 (AT1G19180), JAZ8 (AT1G30135), and JAZ12 (AT5G20900) differentially expressed in both resistant cassava TME3 and susceptible cassava T200 (Supplementary data_Table S2). In cassava T200, JAZ1, JAZ8, and JAZ12 exhibited down-regulation at 32 dpi and/or 67 dpi, whereas in tolerant TME3, JAZ1 and JAZ8 were up-regulated at 12 dpi, but down-regulated at 32 and/or 67 dpi. In addition, JAZ12 was also repressed at 32 dpi (Supplementary data_Table S2). Down-regulation of JAZ could possibly be attributed to the SCF (Skp1-Cullin-F-box) complex which mediates degradation of JAZ proteins, and causes JAZ degradation in order to relieve JA repression (Pieterse *et al*., 2009; Ballare, 2011). This is evident in our study, as JAZ proteins were down regulated in susceptible T200 but up-regulated in TME3, suggesting that the JA pathway is required for defence in a resistant or tolerant plant. Similar findings were also evident in the *Arabidopsis* -SACMV-[ZA:99] study (Pierce and Rey, 2013) where JAZ proteins such as JAZ10 (AT5G13220) was down-regulated at 24 and 36 dpi. It is clear from these results that the JA pathway is activated upon SACMV-[ZA:99] perception by cassava host genes, and a defence response is mounted in TME3, whereas down-regulation in T200 results in disease as the final outcome, confirming the differences between resistant and susceptible cultivars.

SA, JA and ET Signalling

Within this study, it was difficult to assess what role SA signalling played in response to SACMV-[ZA:99] infection. Expression of transcription factors such as WRKY70 (AT3G56400) and WRKY33 (AT2G38470), and the PR3 (AT3G12500) marker gene (Table 2), indicated presence of the SA pathway, but no clear indication was evident on how active this pathway was as not many SA-responsive genes were identified in comparison to those that were in JA and ET pathways. Genes which belong to the AP2/EREB transcription factor family such as, ATERF-5 (AT5G47230) showed down-regulation at 12, 32, and 67 dpi in cassava T200, with ATERF-9 (AT5G44210), ATERF-4 (AT3G15210), and ATERF-1 (AT3G23240 and AT4G17500), predominantly showing repression in the same cultivar, but differential expression patterns were revealed in TME3 (Table 2, Table 3, Supplementary data_Table S2). This activation and repression of such genes suggested a clear role of ethylene signalling in cassava responses to SACMV-[ZA:99] infection albeit with exact functions unknown. Differences in expression patterns revealed possible differences in timing and activation of pathways during SACMV-[ZA:99] progression in both resistant and susceptible cassava cultivars. ET is very influential in mediating the outcome of synergism or antagonism between JA and SA signalling. It is able to bypass key regulator genes such as NPR1 in SA signalling during SA/JA crosstalk, preventing suppression of JA signalling (Pieterse *et al*., 2009). Pierce and Rey, 2013 reported that JA/ET signalling pathways were favoured over SA signalling since marker genes for JA and ET were more prevalent and highly expressed throughout the study compared to SA, confirming speculations made in our cassava study. SA may be active at certain stages of infection, but JA and ET are predominantly the pathways used in SACMV-[ZA:99] defence.

Cell Wall and Trans-Membrane Proteins

The plant cell wall is very rigid as it is made up of a complex matrix of pectin, cellulose or hemicellulose, and many plant pathogens, such as bacteria and fungi, are known to modify the cell wall in order to promote penetration and invasion, while viruses modify the cell wall in order to faciltitate cell - to - cell movement (Guerra-Peraza *et al*., 2005; Zhong and Ye, 2009). In turn, plant hosts may modify their cell wall in several ways in response to pathogen attack. In cassava T200, pectin lyases (AT5G04310, AT4G13710) and pectin-methyl-esterases (PME) (AT5G47500) were shown to be up-

regulated across 3 time points during the course of infection (Table 3). Up-regulation of these enzymes may indicate that SACMV-[ZA:99] could be positively regulating these genes to allow for an increase in the size exclusion limit of the plasmodesmata (Pd) for efficient virus replication and spread to take place. Pectin is enriched around the Pd, and PME is an enzyme involved in pectin de-esterification, and has been shown to interact with virus movement proteins (Levy *et al* 2007; Epel, 2009). It has been hypothesised that PME may act as a receptor protein which may be hijacked by plant viruses to aid in cell to cell movement. Concomitantly, many induced cell wall precursor proteins (AT2G20870) found in TME3 (67 dpi) and T200 (12, 32, and 67 dpi) (Supplementary table_S2), plasma membrane intrinsic proteins found in T200 at 67 dpi (Table 2), and Glutathione S-transferases (TAU19) (AT1G78380) were up-regulated in both TME3 and T200 at 12 dpi, suggesting that multiple defences were attempted to overcome SACMV-[ZA:99] infection, as also observed by an early induction of GST at 12 dpi. This was proven ineffective in limiting pathogen spread, especially in susceptible T200, as virus progression was evident. Alternatively, as previously suggested, SACMV-[ZA:99] may require activaton of cell-wall and plasmamembrane based genes for replication and movement. Other plasma membrane proteins identified in T200 also showed up-regulation, an example of which was CESA8 (AT4G18780) (Table 2). Cellulose synthases (CESA) are large membrane-bound complexes synthesized at the plasma membrane. Many CESAs, including CESA8 are required for cell wall formation, and mutations in these proteins have shown increased resistance to pathogens (Hernandez *et al*., 2007). Collectively, overexpression of CESA8, and other cell-wall associated proteins may indicate an attempt by cassava T200 to intitate defence responses at the cell boundary to inhibit viral replication but was unsuccesful as SACMV-[ZA:99] systemic spread was still evident .

Signalling and Regulatory Proteins

Calmodulin-like genes 23 (AT1G66400), calmodulin-like 37 (AT5G42380) and calmodulin-like 42 (AT4G20780) were down-regulated in both susceptible T200 and tolerant TME3 cassava cultivars upon SACMV-[ZA:99] infection, except at 32 dpi calmodulin-like 42 was induced in cassava TME3 (Supplementary data_Table S2). It has been reported in many studies that calmodulin-like proteins are involved in defence and signalling against pathogen and insect attack and function in pathogen resistance (Cheong *et al*., 2002) In T200, down-regulation of calmodulin 23, calmodulin 37, and calmodulin 42 was observed at 32 and 67 dpi, and similarly in TME3 at 12 dpi, but up-regulation was

observed at 32 dpi for calmodulin-like 42 in the tolerant cultivar (supplementary data_S2). Induction of calmodulin-like 42 at 32 dpi in TME3 could indicate that an appropriate defence response had been mounted once SACMV-[ZA:99] replication had reached levels detectable by TME3. Van Schalk and Rey, 2013, unpublished, reported that no SACMV-[ZA:99] replication was observed at 12 dpi in TME3 but an increase in SACMV-[ZA:99] replication levels from 32 dpi (1.79E+02) to 67 dpi (4,23E+03) was observed. Calmodulin-like genes were showing a higher reduction in expression in T200 at 32 dpi and 67 dpi than TME3 at 12 dpi. For example Calmodulin-like 23 was down-regulated -1.7 log2 fold in TME3 whereas in T200, it was down-regulated -3.6 log2 fold at 32 dpi and -2.8 log2 fold at 67 dpi, indicating that SACMV-[ZA:99] may be exerting a greater inhibitory effect toward cellular responses in susceptible T200, more so than it is capable of in resistant TME3. In addition, an increase in replication levels across time points strongly supports that TME3 is more tolerant to cassava infecting begomoviruses causing CMD, such as SACMV-[ZA:99], rather than being resistant as the plant recovers from infection at later time points, and does not display hallmarks of resistance such as a hypersensitive response (HR). TME3 has been reported in the literature to be resistant, harbouring the putative CMD2 gene which confers dominant specific resistance (Hahn *et al*., 1980; Akano *et al*., 2002; Lokko *et al*., 2005; Okogbenin *et al*., 2007), but it is unlikely that this gene alone is involved in CMD resistance. In fact, TME3 appears to be a 'recovery phenotype' suggesting the involvement of multiple genes and RNA silencing.

Transcription Factors (TFS)

A general overview of 2.5 log2-fold altered transcription factors identified in this study indicated that the majority of these transcription factors were down-regulated in both cassava T200 and TME3 possibly due to the suppressive effect SACMV-[ZA:99] exerts on its host (Table 2, Table 3, Supplementary Table_S2).

WRKY

WRKY TF's have been implicated in many stress-responses as fungal elicitors, pathogen responses, and in SA signalling (Cheong, 2002). WRKY transcription factors identified in this study were WRKY33 (AT2G38470), WRKY40 (AT1G80840), WRKY41 (AT4G11070) and WRKY70 (AT3G56400) (Table 2 and Supplementary data_Table S3), and all were down-regulated in the susceptible cassava cultivar, T200. Expression patterns for WRKY40 and WRKY41 in tolerant TME3 and susceptible cassava T200 revealed up-

regulation in TME3 at 32 and 67 dpi, except for WRKY40 which was down-regulated at 12 dpi. Suppression at 12 dpi may indicate an attempt was made by SACMV-[ZA:99] for host-gene shut-off but this attempt was rendered unsuccessful as up-regulation of this transcription factor is observed at later time points and virus replication was possible. Possibly, once viral replication has reached detectable levels, a defence response was successfully initiated, resulting in induction at 32 dpi and 67 dpi, and hence activation of appropriate defence responses. Down-regulation of both WRKY40 and WRKY41 in T200 may be attributed to virus-specific inhibition, and even though virus levels are detected, basal host defences may not be strong enough to combat SACMV-[ZA:99] infection, resulting in susceptibility as the final outcome.

Currently, 8 WRKY TFs have been shown to be involved in defence in *Arabidopsis*. AtWRKY18, AtWRKY38, AtWRKY53, AtWRKY54, AtWRKY 58, AtWRKY59, AtWRKY66, and AtWRKY70 were identified as targets for NPR1 which is an essential component in SA signalling. WRKY70 was identified in susceptible cassava T200 and was shown to be down-regulated at 67 dpi (Table 2). WRKY70 has been shown a positive regulator of SA-mediated defences while repressing JA signalling (Li *et al*., 2003; Koorneef and Pieterse, 2008). It was suggested that repression of this TF may indicate an attempt by SACMV-[ZA:99] to suppress the SA pathway, to subvert an induced resistance response. Additionally, AtWRKY18 and AtWRKY40 which are closely related to WRKY1 and WRKY2 have been shown to negatively affect EDS1 (SA pathway) expression but positively regulate JA signalling (Ishihama and Yoshioka, 2012). WRKY40 was identified in susceptible cassava T200 but was shown to be down-regulated at 32 dpi and 67 dpi (Ishihama and Yoshioka, 2012).

MAPKs

MAP kinases primary role is to transfer sensors to cellular responses (koornneef and Pieterse, 2008). The MAPK signalling pathway is evolutionary conserved, and can therefore show functional similarities among eukaryotes. In tobacco for instance, salicylic acid induced protein kinase (SIPK) and wound induced protein kinases (WIPK) have been shown to be important regulators of immune responses and in *Arabidopsis*, MPK3, MPK4, and MPK6 have been identified as pathogen responsive MAPKs (Andreasson *et al*., 2005; Nakagami *et al*., 2006; Ishihama and Yoshioka, 2012; Meldau *et al*., 2012). *Arabiopsis* MPK3 is orthologous to tobacco WIPK and MPK6 orthologous to SIPK. In addition, MAPK3 and MAPK6 which are found downstream to MKK4/MKK5 have also been shown

to be regulate auxin and ROS signalling (Blomster *et al*, 2011). MAPK3 (AT3G45640) was found to be down-regulated at 32 and 67 dpi in cassava T200 (Table 2), illustrating an attempt made by SACMV-[ZA:99] to avoid host defences.

MPK4 has been identified as important regulator in defence (Ascencio-Ibanez *et al*., 2008). MPK4 is a negative regulator of SA signalling but a positive regulator of JA signalling (Koorneef and Pieterse, 2008). It has been shown that AtWRKY33 and MPK4 form an indirect interaction with each other through the Map Kinase 4 Substrate 1 (MKS1) complex. We have identified this association in cassava T200 as WRKY33 was identified and was shown to be repressed at 32 and 67 dpi (Table 2). MKS1 functions not only as an adaptor protein but has been shown to enhance the DNA-binding activity of AtWRKY33. Upon bacterial pathogen perception, a complex formed with MPK4 and its upstream kinases, MKK1/MKK2 and MEKK1, cause dissociation and release of WRKY33 and MKS1 from the complex, allowing for MKS1-AtWRKY33 to bind to the promoter region of PHYTOALEXIN DEFICIENT 3 (PAD3), which is responsible for activating expression of antimicrobial camalexin. Down-regulation of WRKY33 shown by SACMV-[ZA:99] may suggest that dissociation of the complex and hence expression of PAD3 was not possible, allowing for disease progression. The identification and down-regulation of WRKY33 in this study may provide some insight (a snapshot) of what could possibly be happening in a cascade of events where pathogens are actively trying to avoid or fight defence (Ishihama and Yoshioka, 2012).

MYB

MYB transcription factors have been shown to be involved in the regulation of flavonoid genes (Meissner *et al*., 1999 and Borevitz *et al*., 2000). Differential expression patterns were observed with MYB transcription factors in cassava T200 and TME3 cultivars. As shown with wounding responses in *Arabidopsis*, MYB15 (AT3G23250) was up-regulated in TME3 at 32 dpi (Supplementary data_TableS2), but was down-regulated in T200 at 32 dpi (Table 2, Supplementary data_TableS2). This is a significant finding as MYB15 has been shown to be a positive regulator of secondary metabolism which is involved in activation of defence responses (Cheong *et al*., 2002). MYB3 transcription factors have been shown to be repressors of the phenylproanoid pathway and secondary metabolism (Cheong *et al*., 2002), therefore up-regulation of MYB3 in cassava T200 (Table 2) suggested that repression of secondary metabolism led to inhibition of defence responses as secondary metabolism is involved in basal defences (Edreva *et al*., 2008)

Geminivirus Pathogenicity Proteins and Host Responses

Specific host-pathogen interactions in resistant hosts are mediated by direct R-avr protein recognition and interaction leading to induced resistance (Pavan *et al*., 2010). However, in susceptible hosts, lacking specific R genes, or lacking recognition between R gene proteins and pathogen proteins (avirulence or effector proteins), disease will occur (Agudelo-Romero *et al*., 2008; Babu *et al*., 2008a; Elena *et al*, 2011). Two layers of pathogen perception exist in plant immunity. One layer detects pathogen-associated molecular patterns (PAMPs) called PAMP triggered immunity (PTI) which attempts to prevent potential pathogen colonization and the second layer recognizes pathogen effectors or avirulence proteins which triggers immunity, known as effector triggered immunity (ETI). This layer is perceived to be a faster and a more effective defence response, often resulting in localized cell death in the presence of resistance proteins (Ishihama and Yoshioka, 2012). Viruses have genome-encoded proteins with multiple functions that can act as pathogenicity proteins which function to suppress basal innate immunity, including RNA silencing.

Extensive overlap exists in pathways induced by viral, fungal, and bacterial pathogens (Pavan *et al*., 2010). For example, the Geminivirus AC4 protein has been shown to interfere with host processes (Bolton, 2009). In *Arabidopsis,* expression of the C4 gene of *Beet curly top virus* (BCTV-B) showed that there was altered tissue layer organization and disruption of the vascular system (Mills-Lujan *et al*., 2010). Exogenous application of brassinosteroid and abscisic acid weakly rescued the C4-phenotype, while seedlings were hypersensitive to kinetin and gibberellic acid, suggesting that C4 can interact with many hormonal pathways and alter plant development. In another study BCTV C4 was shown to interact with two members of the shaggy-related protein kinase family (AtSKŋ and AtSKξ) and a putative leucine-rich repeat receptor-like kinase (LRR-RLK) in a yeast two-hybrid assay (Piroux *et al*., 2007). *Tomato golden mosaic virus* (TGMV) AC4 was also shown to bind to these two shaggy-related protein kinases. These kinases were shown to phosphorylate BCTV C4 and TGMV AC4 with differing efficiencies, and BCTV C4 interacts with the the brassinosteroid signalling pathway via its interaction with AtSKŋ (Piroux *et al*., 2007).

Geminivirus AC3 has also been shown to interact with host proteins such as DNA-J like proteins which are involved in protein folding and NAC transcription factors (NAC). NAC transcription factors such as NAC002 (AT1G01720), NAC086 (AT5G17260), and NAC090 (AT5G22380) showed suppression in cassava T200 at 12, 32, or 67 dpi, whereas in cassava TME3, down-regulation was observed at 12 dpi, but up-regulation at 32 and 67 dpi. NAC has been shown to regulate JA-induced expression (Bu *et al*., 2008). Down-regulation of NAC in the susceptible cultivar could suggest suppression of host defences caused by SACMV-[ZA:99], whereas up-regulation in the TME3 cultivar (at later time points 32, and 67 dpi) indicated appropriate defence response activation once SACMV-[ZA:99] virus replication levels were detected.

Baliji *et al*, 2010, reported that interactions between geminivirus pathogenicity proteins such as TGMV AL2, and *Beet curly top virus* (BCTV) C2 and adenosine kinase (ADK) leads to increased cytokinin-responsive genes that act to promote cell cycle progression. Down-regulation of ADK suppresses antiviral RNA silencing, an innate defence mechanism of plants (Baliji *et al*., 2010), which in turn increases the expression of cytokinins involved in promoting cell cycle progression. ADK has been shown to be linked to SnRK1 whereby inhibition of ADK by geminivirus TGMV AL2 and BCTV C2 prevented inactivation of subgroup I SNF-related protein kinase (SnRK1), leading to enhanced susceptibility. In cassava T200, we identified at SnRK2-10 type gene (AT1G60940) which is similar to SnRK1, however, it was shown to be up-regulated at 32 dpi (Table 2) which does not correlate to the TGMV and BCTV studies mentioned above (Baliji *et al*., 2010). SnRK1 and ADK may be targeted by geminiviruses to facilitate infection, and in addition inhibition of ADK by AL2/C2 pathogenicity proteins prevents phosphorylation of cytokinin, causing an increase in cytokinins necessary for cell-cycle progression which is required for geminivirus genome replication (Baliji *et al*., 2010).

Cell-Cycle Regulation

Functional links between plant signalling hormones (auxin, ethylene, brassinosteroids and cytokinins), and cell-cycle proteins have been established (Andrietta, 2001; Dudits *et al*., 2007; Bari and Jones, 2009) Geminiviruses such as *Cabbage leaf curl virus* (CaLCuV) induce S-phase and G2 genes in order to establish a replication-competent environment (Ascencio-Ibanez *et al*., 2008). We believe this to hold true for SACMV-[ZA:99] in *Arabidopsis* as cyclin genes such as S phase CYCA3;2 were induced

at both 14 dpi (1.32) and at 36 dpi (1.61). In addition, an auxin-responsive factor protein (AT4G38860) was shown to be up-regulated consistently across time points strongly supporting our hypothesis (Chapter 3). Activation of the S phase is accomplished by either blocking transit into the M phase or bypassing the M phase as part of an endocycle (Ascencio-Ibanez er al., 2008). In cassava, the most highly expressed cyclin transcripts were CYCB2;4 (>3.0 log2 fold change) at 12 and 67 dpi and CYCP4;1 at 32 and 67 dpi (5.22 and 5.34 log2 fold change, respectively). The two most down-regulated cyclins were CYCA3;1 at 32 dpi and CYCT1;3 at 67 dpi (-2.70 and -3.05 log2 fold change, respectively). CYCD4,2, which is a late G1 cyclin was highly upregulated in *Arabidopsis* by CaLCuV, supporting the hypothesis that other CDK type D-cyclins phosphorylate RBR releasing E2F transcription factors promoting the S phase. While CYCD4;2 was not detected in T200, CYCD4,1 was highly expressed across all timepoints, indicating a critical role in cell cycle modulation.

CYCD3;1 and CYCD3;2 were both up-regulated in cassava T200 at 12 dpi and 67 dpi, respectively. In *Arabidopsis*, CYCD3;2, an early activator of G1 (Andrietta *et al*., 2001; Dudits *et al*., 2007), was shown to be down-regulated, suggesting that only infected cells transit through late G1, but this was not the case in cassava. Several plant genes, including CYCD3;1 and CYCD3;2 influence the balance of mitotic and endoreduplication or the number of endocycles (Andrietta *et al*., 2001; Dudits *et al*., 2007, Ascencio-Ibanez *et al*., 2008). Differential expression of these two genes in T200 may indicate a balance between mitosis (M) and plant growth of cassava, entry into G1, and shifts to the S phase, depending on where the virus is located. In general, G1 to S transition phases are controlled by CDK containing D-type cyclins which function to release E2F transcription factors in order for transcription of genes necessary for G1 to S transition to occur. They do this by phosphorylating the retinoblastoma protein (RBR). CYCD3,2 is also part of the RBR/EF2 regulatory network which regulates transcription at the G1/S boundary (Gutierrez *et al*., 2002; Hanley-Bowdoin *et al*., 2004). In cassava T200, this cyclin was upregulated at 67 dpi suggesting that this may influence, regulation of this network throughout SACMV-[ZA:99] infection. It is hypothesised that upregulated CYCD3;1 and CYCD3;2 play an important role in cell cycle modulation at 12 and 67 dpi stages of T200 infection, promoting the mitotic cycle therefore resulting in larger numbers of cells to infect as the apical growth tip. CYCD3's normal function is to promote the mitotic cycle and prevent endocycle (Gutierrez *et al*., 2002; Hanley-Bowdoin *et al*., 2004). Notably, 6 cyclin genes were upregulated at 12 dpi, namely, CYCB2;4 CYCB3;1; CYCD3;1; CYCD4;1; and

CYCD5;1 and CYCA1;1, which strongly indicate a critical role in early infection and virus multiplication and spread. At 32 dpi CYCA3;1 and CYCT1;3 were the only two cyclins downregulated. At this stage (32 dpi), SACMV-[ZA:99] is systemically spreading through T200, and symptoms are now well established.

In summary, as shown with other geminivruses (Gutierrez *et al*., 2002; Hanley-Bowdoin *et al*., 2004; Ascencio-Ibanez *et al*., 2008), SACMV-[ZA:99] regulates the cell cycle in order to replicate in the nuclei of infected cells. However, while transcription was monitored in young leaves below the growing apex, it is also important to note that studies done on cell cycle genes in geminiviruses have been performed at fully symptomatic stages (±12-24 dpi) in infection. From this cassava T200 study, it is clear that most differentially expressed (DE) cyclin genes detected (likely not representing all DE cyclins in cassava as annotations are not complete) were upregulated at 12 and/or 67 dpi. Since cassava continues to grow vegetatively, and SACMV-[ZA:99] continues to replicate and spread, in younger and older leaves (older leaves also show symptoms), the virus may not exclusively and always promote the endocycle, as cyclins, such as CYCD3;1 and CYCD3;2, which are associated with promotion of the mitotic cycle were also up-regulated. However since we have no information on CDKs or other cyclins, at this stage in cassava we can only speculate based on the results obtained in this study. However some differences between the two geminiviruses in the SACMV-[ZA:99]-*Arabidopsis* study (chapter 3) and the CaLCuV-Arabidopsis study (Ascencio-Ibanez *et al*., 2008), and some differences between two hosts, cassava and *Arabidopsis* with the same geminivirus, SACMV-[ZA:99] suggest that regulation of cell cycle by geminiviruses is complex and not uniform across different host-virus pathosystems. More genes involved in cell cycle regulation in cassava need to be explored in future.

Conclusions

This study was undertaken to improve the little that is known about mechanisms underlying host-pathogen interactions especially for crops systems with previously uncharacterised genomes such as cassava. Until recently, only marker assisted breeding was available for identifying resistant germplasm, but this remains a slow process as not many resistance genes have been identified. With the availability of technologies such as next generation sequencing, we were able to map the interaction between SACMV-[ZA:99] and its susceptible host, cassava T200, identifying putative disease susceptibility genes

involved in defence. Many candidate genes involved in defence will be selected for functional studies with a specific goal of engineering resistant or tolerant cassava varieties.

4.6 References

- Akano A , Dixon AGO , Mba C , Barrera E, Fregene M (2002) Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. Theor App Genet. 105:521-525.
- An D, Yang J, 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: doi:10.1186/1471-2164-13-64.
- Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol. 11:R106.
- Andrietta MH, Eloy NB, Hermerly AS, Ferreira PCG (2001) Identification of sugarcane cDNAs encoding components of cell cycle machinery. Gen Mol Bio. 24(1-4):61-68.
- Ascencio-Ibanez JT, Sozzani R, Lee TJ, Chu TM, Wolfinger RD *et al*. (2008) Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant Physiol. 148(1): 436-454.
- Awoleye F, Duren M, Dolezel J *et al* (1994) Nuclear DNA content and in vitro induced somatic polyploidization cassava (Manihot esculenta Crantz) breeding. Euphytica 76:195–202.
- Babu M, Griffiths JS, Huang TS, Wang A (2008) Altered gene expression changes in Arabidopsis leaf tissues and protoplasts in response to Plum pox virus infection. BMC Genomics 9: 325.a
- Balijib S, Lacatusc G, Suntera G (2010) The Interaction between Geminivirus Pathogenicity Proteins and Adenosine Kinase Leads to Increased Expression of Primary Cytokinin Responsive Genes. Virology 5:402(2):238-247.
- Ballare CL (2011) Jasmonate-induced defences: a tale of intelligence, collaborators and rascals. Trends Plant Sci 16(5): 249-257.
- Bari R, Jones JD (2009) Role of plant hormones in plant defence responses. Plant Mol Biol 69(4): 473-488.

- Bashandy T, Guilleminot J, Vernoux T, Caparros-Ruiz D, Ljung K, Meyer Y, Reichheld JP (2010) Interplay between the NADP-linked thioredoxin and glutathione systems in Arabidopsis auxin signaling. Plant Cell 22:376–391.
- Berrie LC, Rybicki EP, Rey MEC (2001) Complete nucleotide sequence and host range of South African cassava mosaic virus: further evidence for recombination amongst begomoviruses. J Gen Virol 82(Pt 1): 53-58.
- Blomster T, Saloja¨rvi2 J, Sipari N, Brosche´ M, Ahlfors R, Keina¨nen M, Overmyer Kangasja¨rvi J (2011) Apoplastic reactive oxygen species transiently decrease auxin signaling and cause stress-induced morphogenic response in *Arabidopsis*. Plant Physiology 157:1866-1883.
- Bolton, B (2009) Primary Metabolism and Plant Defense – Fuel for the Fire. MPMI 22:5, 487-497.
- Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12:2383-2393.
- Bu Q, Jiang H, Li CB, Zhai Q, Zhang J, Wu X, Sun J, Xie Q, Li C (2008) Role of the Arabidopsis thaliana NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses. Cell Res. 18(7):756-67.
- Busch W, Lohmann JU (2007) Profiling a plant: expression analysis in Arabidopsis. Curr Opin Plant Biol 10(2): 136-141.
- Carrington JC, Whitham SA (1998) Viral invasion and host defence: strategies and counter-strategies. Curr Opin Plant Biol 1(4): 336-341.
- Chen H, Zhang Z, Teng K, Lai J, Zhang Y, Huang Y, Li Y, Liang L, Wang Y, Chu C, Guo H, Xie Qi (2010) Up-regulation of LSB1/GDU3 affects geminivirus infection by activating the salicylic acid pathway. Plant J. 62(1):12-23.
- Conti G, Rodriguez MC, Manacorda CA,Asurmendi S (2012) Transgenic expression of *Tobacco mosaic virus* capsid and movement proteins modulate plant basal defense and biotic stress responses in *Nicotiana tabacum.* MPMI 25(10):1370-1384.
- Culver JN, Padmanabhan MS (2007) Virus-induced disease: altering host physiology one interaction at a time. Annu Rev Phytopathol 45: 221-243.
- Dixon AGO, Whyte JBA, Mahungu NM, Ng SYC (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, NJ Taylor, F Ogbe and CM Fauquet (eds), pp. S8-05, Donald Plant Sciences Centre, St Louis, USA.

- Doyle JJ, Doyle JL (1987) A rapid isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11-15.
- Dudits D, Abraham E, Miskolczi P, Ayaydin F, Bilgin M *et al*. (2007) Cell-cycle control as a target for calcium, hormonal and developmental signals: the role of phosphorylation in the retinoblastoma-centred pathway. Ann Bot 107(7): 1193-1202.
- Durrant WE, and Dong X (2004) Systemic acquired resistance. Ann Rev Phytopathol. 42, 185-209.
- Edreva A, Velikova V, Tsonev T, Dagnon S, Gürel A, Aktas L, Gesheva E (2008) Stress-protective role of secondary metabolites: Diversity of functions and mechanisms. Gen. Appl. Plant Physiology, Special Issue, 34 (1-2), 67-78
- Elena SF, Carrera J, Rodrigo G (2011) A systems biology approach to the evolution of plant-virus interactions. Curr Opin Plant Biol. 14(4): 372-377.
- Epel BL (2009) Plant viruses spread by diffusion on ER-associated movement-protein-rafts through plasmodesmata gated by viral induced host beta-1,3-glucanases. Semin Cell Dev Biol 20(9): 1074-1081
- Erik Andreasson E,Jenkins T, Brodersen P, Thorgrimsen S,Petersen NHT, Zhu S, Qiu J Micheelsen P,Rocher A, Petersen M, Newman M, Nielsen HB, Hirt H, Somssich I, MattssonO, Mundy J (2005) The MAP kinase substrate MKS1 is a regulator of plant defense responses. The EMBO Journal 24:2579-2589.
- Eybishtz A, Peretz Y, Sade D, Gorovits R, Czosnek H (2010) Tomato yellow leaf curl virus infection of a resistant tomato line with a silenced sucrose transporter gene LeHT1 results in inhibition of growth, enhanced virus spread, and necrosis. Planta 231:537-548.
- Feys BJ, Moisan, LJ, Newman, M, Parker, JE (2001) Direct interaction between the Arabidopsis disease resistance signaling proteins, EDS1 and Pad4. *The EMBO Journal* 20(19): 5400-5411.
- Fisher R A (1970) Statistical Methods for Research Workers. , 14th edition, Hafner Publishing, 1970, p. 96.
- Fontes EP, Santos AA, Luz DF, Waclawovsky AJ, Chory J (2004) The geminivirus nuclear shuttle protein is a virulence factor that suppresses transmembrane receptor kinase activity. Genes Dev. 18(20): 2545-2556.

- Fregene M, Matsumura H, Akano A, Dixon A, Terauchi R (2001a) Progress towards cloning the single dominant gene conferring immunity to cassava mosaic disease (CMD). In: Cassava, an ancient crop for modern times: Food, health, culture. NJ Taylor, F Ogbe, CM Fauquet (eds), pp. S5-09, Donald Danforth Plant Sciences Centre, St Louis, USA.
- Gehrig H, Winter K, Cushman J, Borland A, Taybi T (2000) An improved isolation method for succulent plant species rich in polyphenols and polysaccharides. Plant Mol Biol Rep. 18(4):369-376.
- Guerra-Peraza O, Kirk D, Seltzer V, Veluthambi K, Schmit AC *et al*. (2005) Coat proteins of Rice tungro bacilliform virus and Mungbean yellow mosaic virus contain multiple nuclear-localization signals and interact with importin alpha. J Gen Virol 86(Pt 6): 1815-1826.
- Gutierrez C (2002) Strategies for geminivirus DNA replication and cell cycle interference. Physiol Mol Plant Path. 60: 219-230.
- Hahn SK, Terry ER, Leuschner K (1980) Breeding cassava for resistance to cassava mosaic disease. Euphytica 29:673-683.
- Hanley-Bowdoin L, Settlage SB, Robertson D (2004) Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. Mol Plant Pathol 5(2): 149-156.
- Havelda Z, Hornyik C, Crescenzi A, Burgyan J (2003) In situ characterization of Cymbidium Ringspot Tombusvirus infection-induced posttranscriptional gene silencing in Nicotiana benthamiana. J Virol 77(10): 6082-6086.
- Havelda Z, Maule AJ (2000) Complex spatial responses to *Cucumber mosaic virus* infection in susceptible *Cucurbita pepo* cotyledons. Plant Cell 12:1975-1986.
- Herna´ndez-Blanco C, Feng, DX, Hu J, Sa´ nchez-Vallet,a Laurent Deslandes A, Llorente,a Marta Berrocal-Lobo F, Keller H, Barlet X, Sa´ nchez-Rodrı´guez C, Anderson LK,Somerville S,Marco Y,Antonio Molinaa A (2007) Impairment of cellulose synthases required for Arabidopsis. Plant Cell 19:890-903.
- Huang Z, Yeakley JM, Garcia EW, Holdridge JD, Fan JB, Whitham SA (2005) Salicylic acid-dependent expression of host genes in compatible *Arabidopsis*–virus interactions. Plant Physiol. 137:1147-1159.
- Ishihama N, Yoshioka H (2012) Post-translational regulation of WRKY transcription factors in plant immunity. Current Opinion in Plant Biology 2012, 15:431–437.
- Jeske H (2009) Geminiviruses. Curr Top Microbiol Immunol 331: 185-226.

- Jones, JDG, and Dangl, JL (2006) The plant immune system. Nature 444,323-329.
- Koornneef A, Pieterse CM (2008) Cross talk in defence signalling. Plant Physiol 146(3): 839-844.
- Koornneef M, Meinke D (2010) The development of Arabidopsis as a model plant. Plant J 61(6): 909-921.
- Leniewska A, Oboniewski MJ (2011) rnaSeqMap: a Bioconductor package for RNA sequencing data exploration. BMC Bioinformatics 12:200.
- Levy A, Guenoune-Gelbart D, Epel BL (2007) beta-1,3-Glucanases: Plasmodesmal Gate Keepers for Intercellular Communication. Plant Signal Behav 2(5): 404-407.
- Li J, Wen J, Lease KA, Doke JT, Tax FE *et al*. (2002) BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signalling. Cell 110(2): 213-222.
- Lokko Y, Danquah EY, Offei SK, Dixon AGO , Gedil MA Molecular markers associated with a new source of resistance to the cassava mosaic disease. African Journal of biotechnology (2005) 4(9):873-881.
- Lozano-Duran R, Rosas-Diaz T, Luna AP, Bejarano ER (2011) Identification of host genes involved in geminivirus infection using a reverse genetics approach. PLoS One 6(7): e22383.
- Ma S, Gong Q, Bohnert HJ (2007) An Arabidopsis gene network based on the graphical Gaussian model. Genome Res 17(11): 1614-1625.
- Mariano AC, Andrade MO, Santos AA, Carolino SMB, Oliveira ML, Baracat-Pereira MC,Brommonshenkel SH, Fontes EPB (2004) Identification of a novel receptor-like protein kinase that interacts with a geminivirus nuclear shuttle protein. Virology 318:24–3.
- Maule A, Leh V, Lederer C (2002) The dialogue between viruses and hosts in compatible interactions. Curr Opin Plant Biol 5(4): 279-284.
- Meissner RC, Jin H, Cominelli E, Denekamp M, Fuertes A, Greco R, Kranz HD, Penfield S, Petroni K, Urzainqui A (1999) Function search in a large transcription factor gene family in Arabidopsis: assessing the potential of reverse genetics to identify insertional mutations in R2R3 MYB genes. Plant Cell 11:1827-1840.
- Meldau S, Ullman-Zeunert L, Govind G, Bartram S, Baldwin IT (2012) MAPK-dependent JA and SA signalling in Nicotiana attenuata affects plant growth and fitness during competition with conspecifics. BMC Plant Biol. 12:213.

- Mills-Lujan K, Deom CM (2010). Geminivirus C4 protein alters Arabidopsis development. Protoplasma 239(1-4): 95-110.
- Mysore, K.S. and Ryu, C. (2004) Nonhost resistance: how much do we know? Trends Plant Sci. 9(2): 97-104.
- Nagar S, Pedersen TJ, Carrick KM, Hanley-Bowdoin L, Robertson D (1995) A geminivirus induces expression of a host DNA synthesis protein in terminally differentiated plant cells. Plant Cell 7(6): 705-719.
- Nakagami H, Soukupová H, Schikora1 A, Žárský V, Hirt HA (2006) Mitogen-activated protein kinase kinase kinase mediates reactive oxygen species homeostasis in Arabidopsis http://www.jbc.org/cgi/doi/10.1074/jbc.M605293200
- Okogbenin, E, Porto MCM, Egesi C, Mba C, Espinosa E, Santos LG, Ospina C, Marín J (2007) Marker-assisted introgression of resistance to Cassava Mosaic Disease into Latin American germplasm for the genetic improvement of cassava in Africa. Crop Science 47:1895-1904.
- Oktem HA, Eyidogan F, Selcuk F, Oz MT, da Silva J, Yucel M (2008) Revealing response of plants to biotic and abiotic stresses with microarray technology. Genes, Genomes, and Genomics 2(1):14-48.
- Owens RA, Tech KB, Shao JY, Sano T, Baker CJ (2012) Global analysis of tomato gene expression during Potato spindle tuber viroid infection reveals a complex array of changes affecting hormone signalling. MPMI 25(4): 582-598.
- Pallas V, Garcia JA (2011) How do plant viruses induce disease? Interactions and interference with host components. J Gen Virol 92(Pt 12): 2691-2705.
- Pavan S, Jacobsen E, Visser R, Bai Y (2010) Loss of susceptibility as a novel breeding strategy for durable and broad-spectrum resistance. Mol Breeding 25:1-12.
- Pierce, EJ, and Rey, MEC (2013) Assessing global transcriptomic changes in response to South African cassava mosaic virus [ZA:99] infection in susceptible Arabidopsis thaliana. PLoS ONE, 8(6): e67534.
- Pieterse CM, Leon-Reyes A, Van der Ent S, Van Wees SC (2009) Networking by small-molecule hormones in plant immunity. Nat Chem Biol 5(5): 308-316.
- Piroux N, Saunders K, Page A, Stanley J (2007) Geminivirus pathogenicity protein C4 interacts with Arabidopsis thaliana shaggy-related protein kinase AtSKeta, a component of the brassinosteroid signalling pathway. Virology 362(2): 428-440.

- Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, & Mohiuddin M,Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar DS, & Rounsley S (2012) The Cassava Genome: Current Progress, Future Directions. Tropical Plant Biol 5:88-94.
- Seki M, Shinozaki K (2009) Functional genomics using RIKEN Arabidopsis thaliana full -length cDNA's. J Plant Res. 122(4):355-366.
- Staskawicz, B.J., Ausubel, F.M., Baker, B.J., Ellis, J.G. and Jones, J.D.G. (1995) Molecular genetics of plant disease resistance. Science 268: 661-667.
- Truman B, Bennett MH, Kubigsteltig I, Turnbull C, Grant M (2007) Arabidopsis systemic immunity uses conserved defense signalling pathways and is mediated by jasmonates. PNAS 104, 1075-1080.
- Tuskan GA, Difazio S, Jansson S *et al* (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313:1596–1604.
- Walley JW, Coughlan S, Hudson ME, Covington M, Kaspi R, Banu G,Harmer SL, Dehesh K (2007) Mechanical Stress induces biotic and abiotic stress responses via a novel cis-element. PLOS Gen. 3(10):1800-1812.
- Whitham SA, Quan S, Chang HS, Cooper B, Estes B *et al*. (2003) Diverse RNA viruses elicit the expression of common sets of genes in susceptible Arabidopsis thaliana plants. Plant J 33(2): 271-283.
- Whitham SA, Yang C, Goodin MM (2006) Global impact: elucidating plant responses to viral infection. MPMI 19(11): 1207-1215.
- Zhong R, Ye ZH (2009) Transcriptional regulation of lignin biosynthesis. Plant Signal Behav 4(11): 1028-1034.

Chapter 5

Overall Summary and Conclusions

The overall goal of this study was to identify novel host transcripts induced or repressed by SACMV-[ZA:99] infection. The study was undertaken in order to identify host defence genes governing susceptibility involved in host – pathogen interactions. The virulence of SACMV-[ZA:99] was assessed in both susceptible *Arabidopsis* and cassava T200 and many genes involved in defence pathways were identified in this study. Before transcriptome profiling studies could take place, it was essential to address the low agroinfection efficiencies that were previously obtained in recalcitrant crop systems such as cassava in order to establish an efficient infection method that can be broadly used for both model and host pathosystems. In addition, comparisons in virulence between begomoviruses, SACMV-[ZA:99] and ACMV-[NG:Ogo:90] in three *Agrobacterium* strains in three susceptible cassava genotypes were also drawn from the study. Agroinfection efficiency is thus dependent on the *Agrobacterium* strain, *vir* functions and strength of promoters upstream from the replication-associated protein found on DNA-A genome of the virus (Jacob *et al*., 2003). This study highlighted a complex interplay between virus genetic factors and host defences involved in infection such as host genotype factors supporting virus replication and movement, plant binary vector genetic background, and *Agrobacterium* virulence factors that govern agroinfection efficiencies in crops.

Once agroinfection procedures were optimal for SACMV-[ZA:99] penetration into it's host plant, transcriptome profiling was carried out in the susceptible model plant system, *Arabidopsis thaliana*, in order to identify host response factors involved in defence against SACMV-[ZA:99] infection. A whole genome 4 x 44k microarray was used at 3 time points (14, 24, and 36 dpi) where 3 biological replicates and 1 technical replicate were run at each time point. Results form the infectivity assay revealed that an increase in virus titre across time points correlated with an increase in differentially expressed transcripts. Studies such as those conducted by Babu *et al*., 2008b, and Golem and Culver, 2003, confirmed these findings. It was also necessary to confirm that agroinoculation procedures using *Agrobacterium* strain AGL1 would have no effect on host gene expression and appropriate control measures were thus carried out for these purposes. Quantitative PCR (qPCR) revealed a decrease in the low levels of AGL1 copies already present across time points, strongly suggested that *Agrobacterium* AGL1 was not persistently replicating in

Arabidopsis leaf tissues. In addition mock-inoculated AGL1 controls were run in the experiment to confirm that expression changes were SACMV-[ZA:99] specific. Microarray analysis of 13,934 significant ($p < 0.05$) differentially expressed genes (including up- and down-regulated transcripts) revealed that genes expressed transiently at a particular time point could function to conserve energy resources in *Arabidopsis* (Cheong *et al*., 2002; Koornneef and Pieterse, 2008; Pieterse *et al*., 2009; Baena-Gonzalez, 2010) as the host was under immense stress due to pathogen (SACMV-[ZA:99]) attack. Alternatively, these results may also be suggesting that transiently expressed genes (up- or down-regulated) may be targeting downstream genes for activation of host defence responses. Those genes showing continuous expression across time points were shown to be involved primarily in basal defences as the plant was continuously attempting to mount appropriate defence responses in order to inhibit SACMV-[ZA:99] infection. Alternatively, SACMV-[ZA:99] may be manipulating host genes for its own replication, movement and systemic spread (Whitham *et al*., 2006; Ascencio-Ibanez *et al*., 2008). MIPS functional categories results revealed a higher number of induced genes present at 14 dpi suggesting activation of non-specific innate host responses to virus invasion by induction of stress and defence-like genes. Host-gene shut-off in *Arabidopsis* was observed with an increase in down-regulated genes at 24 and 36 dpi. SACMV-[ZA:99] was actively replicating at this stage as observed with increased virus titre levels and attempt to hijack many host processes for its own benefit was evident. Another explanation for down-regulation may be host-directed diversion of metabolites such as glycolysis and gluconeogenesis, pentose-phosphate pathways, and carbohydrate metabolism away from normal cell function in order to conserve energy, as well as defend itself from SACMV-[ZA:99] antagonism.

Pathways of significance identified in the study were those involved in phytohormone signalling. Results conveyed that primary pathways such as salicylic (SA), ethylene (ET) and Jasmonic acid (JA) were working synergistically with each other. Marker genes such as EDS1, PAD4, MPK4, and NPR1 were identified and expressed in the SA pathway, and JA responsive genes such as PDF1.2 and VSP1, and those involved in ET signalling such as EIN2 and EBF were also expressed. The geminivirus, *Cabbage leaf curl virus* (CaLCuV) showed contradictory results to these findings as authors established that the SA pathway was working antagonistically with the JA pathway as shown by the repression of JA marker gene, PDF1.2 (Ascencio-Ibanez *et al*., 2008) However, similar findings were observed with *Cauliflower mosaic virus* (CaMV) and SACMV-[ZA:99] by PDF1.2 up-regulation. This gene was highly up-regulated in our study confirming JA

activation and its importance in *Arabidopsis* defence against SACMV-[ZA:99] infection. Evidence revealed that some changes in host-gene function may be virus-specific or common to certain viruses but differences also exist in host manipulation among viruses in general (Havelda and Maule, 2000; Maule 2002; Pallas and Garcia, 2011).

The next approach to the study was to apply whole genome transcriptomics in a previously uncharacterised susceptible South African cassava landrace, T200, in order to identify putative genes involved in SACMV-[ZA:99] replication and movement (pathogenicity), as well as in defence and signalling pathways during the disease process, over 67 days post agroinoculation. Next generation deep sequencing using the ABI Solid system was utilized for this aspect of the project. The overall goal of this study was to identify genes involved in defence in a natural host system such as cassava T200 in order to establish if similarities in SACMV-[ZA:99] infection exist in susceptible hosts with different genetic backgrounds such as in the model plant, *Arabidopsis*. Within host gene comparisons between susceptible cassava T200, and a tolerant TME3 cultivar were also carried out. As with the susceptible *Arabidopsis* - SACMV-[ZA:99] microarray study previously described, SACMV-[ZA:99] replication levels were also shown to increase across time points (12, 32, and 67 dpi) in cassava T200, also correlating with a down-regulation of expressed genes. These findings substantiate previous results suggesting that host-gene shut-off by SACMV-[ZA:99] was evident, and in addition similar MIPS functional categories were established between the two systems, suggesting a common mechanism of host defence against SACMV-[ZA:99] infection and in turn a virus-specific manipulation of host genes for viral replication and spread. Host-gene shut-off is a common mechanism used by viruses to establish infection (Havelda and Maule, 2000; Maule 2002; Pallas and Garcia, 2011), but the type of genes and pathways altered are virus and host specific, as observed with the differences in expression patterns identified. As was shown in the *Arabidopsis* study, SA, JA, and ET signalling pathways were activated in cassava T200 working synergistically with each other, identifying a common pattern of defence against SACMV-[ZA:99]. The current study (cassava T200), however, identified more regulatory proteins involved in defence than what the *Arabidopsis* study showed, as many transcription factors such as WRKY, MAPK's, and MYB were expressed. In addition, many viral pathogenicity proteins were suggested to cause down-regulation of disease resistance proteins such as the TIR-NBS-LRR class at 32 and/or 67 dpi, respectively. A disease resistance protein (TIR-NBS-LRR) was shown to be highly expressed in *Plum pox virus* (PPV) – infected *Arabidopsis* (Babu *et al*., 2008a), and is

known to confer resistance to viral diseases. Down-regulation of TIR-NBS-LRR in the *Arabidpopsis* and cassava T200 - SACMV-[ZA:99] pathosystem confirmed that a susceptible environment for SACMV-[ZA:99] was established.

Conclusions

Whole genome transcriptome profiling studies in susceptible model and natural host systems provided valuable information into the little that is known about geminivirus pathogenicity as well as a better understanding into the molecular mechanisms underlying uncharacterised plant genomes such as cassava was achieved. Findings in this study identified and illustrated how defence signalling pathways and networks are highly interconnected and complex. In addition, the shut-off of host genes by SACMV-[ZA:99] demonstrated how viruses manipulate host defences and genes to their own benefit in order to establish a replication-competent environment. Future studies will entail testing candidate host genes and transcription factors through VIGS, si- and miRNA silencing and reverse genetic approaches, as well as identifying virus pathogenicity factors involved in suppression of host defence genes. Additional approaches toward identifying candidate genes will not only involve transcriptomics, but also proteomic and metabolomic approaches as well, in order to further elucidate SACMV-[ZA:99] – induced transcriptional responses of cassava and compare these to model crops. Different approaches toward understanding molecular mechanisms underlying diverse pathways and networks in crop systems on all levels would assist strategic approaches to engineer resistance to geminiviruses.

5.1 References

- Ascencio-Ibanez JT, Sozzani R, Lee TJ, Chu TM, Wolfinger RD *et al*. (2008) Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant Physiol 148(1): 436-454.
- Babu M, Gagarinova AG, Brandle JE, Wang A (2008) Association of the transcriptional response of soybean plants with soybean mosaic virus systemic infection. J Gen Virol 89(Pt 4): 1069-1080b.
- Babu M, Griffiths JS, Huang TS, Wang A (2008) Altered gene expression changes in Arabidopsis leaf tissues and protoplasts in response to Plum pox virus infection. BMC Genomics 9: 325a.
- Baena-Gonzalez E (2010) Energy signaling in the regulation of gene expression during stress. Mol Plant 3(2): 300-313.
- Cheong YH, Chang HS, Gupta R, Wang X, Zhu T *et al*. (2002) Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis. Plant Physiol 129(2): 661-677.
- Golem S, Culver JN (2003) Tobacco mosaic virus induced alterations in the gene expression profile of Arabidopsis thaliana. Mol Plant Microbe Interact 16(8): 681-688.
- Havelda Z, Maule AJ (2000) Complex spatial responses to *Cucumber mosaic virus* infection in susceptible *Cucurbita pepo* cotyledons. Plant Cell 12:1975-1986.
- Jacob, S.S., Vanitharani, R., Karthikeyan, A.s., Chinchore, Y., Thillaichidambaram, P., and Veluthambi, K. (2003). *Mungbean yellow mosaic virus*- Vi agroinfection by co-delivery of DNA A and DNA B from one *Agrobacterium* strain. Plant Disease 87(3), 247-251.
- Koornneef A, Pieterse CM (2008) Cross talk in defense signaling. Plant Physiol 146(3): 839-844.
- Maule A, Leh V, Lederer C (2002) The dialogue between viruses and hosts in compatible interactions. Curr Opin Plant Biol 5(4): 279-284.
- Pallas V, Garcia JA (2011) How do plant viruses induce disease? Interactions and interference with host components. J Gen Virol 92(Pt 12): 2691-2705.
- Pieterse CM, Leon-Reyes A, Van der Ent S, Van Wees SC (2009) Networking by small-molecule hormones in plant immunity. Nat Chem Biol 5(5): 308-316.

- Whitham SA, Yang C, Goodin MM (2006) Global impact: elucidating plant responses to viral infection. Mol Plant Microbe Interact 19(11): 1207-1215.