

https://dspace.org/

The transcription factor interacting network of tolerant TME3 and susceptible T200 cassava landraces infected with SACMV

Freeborough, Warren

2019

https://hdl.handle.net/10539/29438

Freeborough, Warren (2019) The transcription factor interacting network of tolerant TME3 and susceptible T200 cassava landraces infected with SACMV, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/29438>

Downloaded from a DSpace repository. https://dspace.org/

School of Molecular and Cell Biology
University of the Witwatersrand
Johannesburg

THE TRANSCRIPTION FACTOR INTERACTING NETWORK OF TOLERANT TME3 AND SUSCEPTIBLE T200 CASSAVA LANDRACES INFECTED WITH SACMV

A Dissertation submitted to the Faculty of Science of the University of Witwatersrand, Johannesburg, in full fulfilment of the requirements for the degree of Master of Science

March 2019

by

Warren Freeborough

723388

ORCID ID: 0000-0003-3825-3401

Supervisor:	Prof Chrissie Rey
Co-Supervisor:	Dr Nikki Gentle
Advisor:	Dr Vanessa Meyer

Declaration

I, Warren Freeborough (723388), am a student registered for the degree of Master of Science (MSc): in the academic year 2019.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise.
- I have not submitted this work before for any other degree or examination at any other University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature of candidate  25th day of March 2019

Acknowledgments

I would like to demonstrate my sincere gratitude to my supervisor, Prof Chrissie Rey, and to my co-supervisor: Dr Nikki Gentle. I would like to thank you both for the guidance you have provided me throughout this project. In particular, I would like to thank both of you for pushing my limits and growing skills and knowledge to heights that would have been previously unattainable. Thank you Prof. Rey for the imagination that created this project and seeing the potential where others failed previously. Thank you Dr. Gentle for taking this project and laying the foundations and making an idea into an achievable goal.

Special thanks are extended to Dr Y. Wei who assisted this project by providing us additional supplementary information from their own research. In doing so, helped the completion of this research and furthered science as a whole.

I would like to thank my advisor Dr Vanessa Meyer, who performed her responsibilities with both excitement and a kind heart. Thank you for providing technical advice regarding completing my studies and listening to my concerns.

To my parents, Lesley and Robert Freeborough, who have shown their support throughout my academic career. Thank you both, who have provided me with the freedom to choose my goals and the work ethic needed to achieve them, To Dominque, my loved sister, thank you for supporting me through this process and being my friend in my time of need.

To all of my friends I have accumulated over my years of study, I say thank you all for helping me through these times whilst creating good memories. Among these friends, special thanks to members of both the plant biotechnology and the bioinformatics laboratories who provided tremendous technical support on this project. Special thanks to my close friends who have provided unconditional support and friendship both in and outside the lab. Lastly, I would like to thank the National Research Foundation and the University of Witwatersrand for providing financial support for the duration of this project.

Abstract

Cassava, *Manihot esculenta* Crantz, is categorized as a food security crop, producing large starchy tubers that are gaining interest from both international and local agro-processing industries for products such as bioethanol, textiles, and food additives. However, cassava is currently under threat from a group of begomoviruses that cause cassava mosaic disease (CMD) in all countries in sub-Saharan Africa where cassava is cultivated. CMD can result in up to 100% crop loss. South African cassava mosaic virus (SACMV) is particularly a threat to the growing cassava industry in southern Africa. Despite extensive breeding programs over the past 70 years to develop CMD-resistant farmer-preferred cassava landraces, total resistance has not been achieved. Furthermore, the high mutational rates of begomoviruses, and mixed infections in the field, have exacerbated the problem. TME3 is a West African landrace that displays tolerance to begomoviruses, including SACMV. Infection of TME3 by SACMV leads to recovery, hallmarked by low virus loads and milder symptoms compared to a susceptible southern African landrace T200. The molecular processes that govern tolerance in crops, including cassava, are not well understood. However, systemic immune responses, which are controlled by hormone-responsive transcription factors (TFs), are required by the plant to successfully combat an invading pathogen. Two different branches of systemic immunity have been described, namely systemic acquired resistance (SAR), facilitated by salicylic acid (SA) signalling, and induced systemic resistance (ISR), which is induced through jasmonic acid (JA) and ethylene (ET) signalling in the presence of beneficial rhizobacteria. In 2014, Allie *et al*. compared global transcriptomic responses occurring in TME3 and the T200 during early 12 days' post inoculation (dpi), middle (32 dpi) and late (67 dpi) stages of SACMV infection.

In order to give greater context to transcriptomic data, which is inheritably large and complex, network analysis may be implemented. By placing the differentially expressed (DE) gene homologs/orthologs identified from the cassava transcriptome datasets into protein-protein networks, functions of SACMV-responsive genes, interacting partners, and potential hubs, can be derived. Cassava gene functions are based on the model crop Arabidopsis thaliana, as despite the sequencing of the cassava genome, the annotations are incomplete. The aim of this study was to identify potential candidate TFs, and their associated hormones and other network partners, that confer either tolerance (TME3) or susceptibility (T200) to SACMV.

Abstract

Differentially expressed (DE) TFs belonging to the bZIP, ERF, NAC and WRKY families (well documented to play a role in biotic stress responses) were identified from global transcriptomic data (Allie *et al.*,2014). Predicted protein-protein interactions between these identified DE TFs and their interacting partners were identified using STRING. By combining this information with the biological function gene ontologies (GO), construction of functional networks in both landraces at each time point was enabled. Chapter 2, focused on the well-studied WRKY family due to its predominant role in pathogen-induced plant host responses. Analysis clearly revealed that in T200 the down-regulation of both *At*WRKY53 and 70 identified orthologs played a large part in the down-regulation of hormone signalling and defences, while in TME3, *At*WRKY40 orthologs were continuously up-regulated throughout SACMV infection and were linked to the enriched JA responsiveness. *At*WRKY33 and 40 orthologs in T200 during middle and late infection are proposed to be the hubs targeted by SACMV to establish infection. In TME3, in addition to *At*WRKY40 orthologs, the uniquely up-regulated *At*WRKY6 and 28 identified orthologs are also suggested to play a role in CMD-tolerance in cassava. While this study has for the first time enabled insights into WRKY TFs and the hormone response in during begomovirus infection of susceptible and tolerant cassava landraces, these candidate genes remain to be empirically verified.

In chapter 3, four additional TFs associated with plant pathogen responses were investigated. At all three stages of infection, a total of 110 DE TFs belonging to the five families were identified in T200, whereas only 49 DE TFs were found in TME3. Analysis of enriched GO terms corresponding to hormone signalling revealed suppression of JA, SA and ET in susceptible T200, whereas in contrast in tolerant TME3, JA and ET were the dominant hormone responses during early and middle infection. In both T200 and TME3, the key TFs that appear to be playing a large role in determining susceptibility or tolerance are the nine identified orthologs that show contrasting expression (*Me*WRKY3, 10 and 11, *Me*MYB151 and 165, *Me*NAC29 and *Me*ERF18,74 and 106). In addition to these TFs, the down-regulated homologs of *At*WRKY53 and 70, as shown in chapter 2, *At*MPK3 and *At*ERF5 and 6 are thought to aid in establishing and maintaining infection in T200. In contrast, the up-regulated Arabidopsis homologs belonging to *At*ETR1 and 2, *At*JAZ1 and 8, and *At*MYC2 are thought to induce defense pathways that lead to tolerance in TME3. When examining the networks common between T200 and TME3, it is noted that most of the similarities occur at 12 dpi before symptoms appear, and become increasingly

more divergent as SACMV infection progresses. When scrutiny is place on the full networks, the prevailing feature found in T200 is that the networks are highly interconnected and are comprised of mostly down-regulated proteins which have functions pertaining to either JA or SA regulation or defense. There was very little change to the overall network in T200 between middle and late infection. In contrast, TME3 shows reduced networks with more up-regulated proteins, most of which are contributing to defense processes or JA and ET orientated hormone responses. Up-regulation of SA genes were notably absent in TME3 at 12 dpi. At 67 dpi, where tolerance manifests as symptom recovery, the prevailing hallmarks are the continued up-regulation of JA responsiveness which is found to be linked to *At*MYB15 (*Me*MYB27) *At*WRKY40 (*Me*WRKY10) orthologs.

Is this a new summary paragraph? In summary, establishment of infection is achieved by SACMV through suppression of key hub TFs (*At*WRKY53,70 and *At*ERF5 and 6 orthologs) that regulate hormones, and initial disruption of the circadian clock (*At*RVE1 orthologs). During middle and late infection, it appears that SACMV reprograms the hosts' transcriptome to maintain infection of cells. This likely achieved through the suppression of programmed cell death (PCD) pathways which result in maintenance of infected cells and allowing viral movement to other uninfected cells. In contrast, in TME3 during early infection, the induction of the cassava orthologs of *At*WRKY40 (*Me*WRKY40 and 79), and other notable responses such as the suppression of ET responses and the induction of JAZ proteins are implicated in the induction of ISR-associated protein networks leading to recovery and tolerance to SACMV. This is the first detailed study, to the best of our knowledge, on the role of hormones and transcription factors in begomovirus infected cassava. Evidence of a possible rhizobacterium-independent induction of ISR-like mechanisms in a virus-plant host system is presented and requires further exploration. Confirmation for a role for ISR in SACMV tolerance needs to be empirically confirmed through studies such as *in vivo* cassava gene editing studies.

Outputs of the study

1. Poster Presentation at Plant Cold Spring Harbor, Plant Genomes & Biotechnology: From Genes to Networks conference: Freeborough, W., Gentle, N., and Rey, MEC (2017). Hub WRKY transcription factors and Interacting Partners in Response to Cassava Mosaic Geminiviruses.
2. Poster presentation at the South African Microbiology Society: Freeborough, W., Gentle, N., and Rey, MEC (2018). Hub WRKY transcription factors and Interacting Partners in Response to Cassava Mosaic Geminiviruses.
3. Oral presentation at the joint South African Genetics Society and the South African Society of Bioinformatics: Freeborough, W., Kanyika, BN., Gentle, N., and Rey, MEC (2018). The importance of hub WRKY transcription factors in the establishment of tolerance and susceptibility against plant viruses.
4. Two journal articles are predicted to be submitted for publication to ISI rated journals in the following months.

Table of Contents

Declaration i

Acknowledgments ii

Abstract iii

Outputs of the study vi

Table of Contents vii

List of Figures xi

List of Tables xiii

List of Abbreviations xiv

Chapter outline xvi

Chapter 1 1

1.1 Cassava as a food security crop 1

1.2 Cassava Diseases 1

1.3 Resistance, tolerance and susceptibility 4

1.4 Tolerance and susceptibility 5

1.4.1 TME3 and T200 cassava landraces 5

1.4.2 Key mechanisms associated with the development of tolerance 6

1.5 Plant immunity 7

1.5.1 PTI and ETI 7

1.5.2 ROS 8

1.5.3 Mitogen-Activated Protein Kinases (MAPKs) 9

1.6 Hormone signalling in plant immunity 11

1.7 Transcription factors 15

1.7.1 Basic leucine zipper contain domain proteins 15

Table of Contents

1.7.2 Ethylene-responsive element binding factors 16
1.7.3 Myeloblastosis related proteins 17
1.7.4 No apical meristem, Arabidopsis transcription activation factor and cup-shaped cotyledon 18
1.7.5 WRKYs 21
1.8 Induction of the plant immune system 22
1.8.1 Systemic Acquired Resistance and Induced Systemic Resistance 22
1.8.2 RNA Interference (RNAi) 24
1.9 Transcriptomics and systems biology 26
1.9.1 Cassava transcriptomic data 27
1.9.2 Systems biology 27
1.10 Working with non-model organisms 29
1.11 Rationale for Study 29
1.11.1 Aim 31
1.11.2 Objectives 31
Chapter 2 32
2.1 Background 32
2.2 Methods 37
2.2.1 Identification of differentially expressed genes 37
2.2.2 GO enrichment 39
2.2.3 Identification of interacting partners 40
2.3 Results 41
2.3.1 Comparison between differentially expressed (DE) *Me*WRKYs in cassava TME3 and T200 at 12, 32 and 67 dpi 41
2.3.2 Phylogenetic analysis of differentially expressed WRKYs in Arabidopsis, cassava T200 and TME3 45

2.3.3 Gene ontology functional enrichment analysis of biological processes occurring in T200 and TME3 47

2.3.4 Comparative identification and functions of DE *Me*WRKY-interacting partners in T200 and TME3 based on an Arabidopsis *At*WRKY33 and 40 homolog network 49

2.3.5 Analysis of other *Me*WRKYs and DE *Me*WRKY-interacting partners in T200 and TME3 based on homologs in Arabidopsis networks 50

2.3.6 Predicted pathways and hormone responses of the differentially expressed (DE) T200 and TME3 *Me*WRKY homologs in Arabidopsis 52

2.4 Discussion 59

2.4.1 General 59

2.4.2 Phylogenetics 60

2.4.3 Gene Ontology 62

2.4.4 Hormone signalling 63

2.4.5 Central protein-protein network 64

2.4.6 Early 66

2.4.7 Middle 69

2.4.8 Late 76

2.4.9 WRKYs in literature 77

2.5 Conclusions 81

Chapter 3 83

3.1 Introduction 83

3.2 Material and Methods 86

3.2.1 Identification of differentially expressed TFs in cassava and their Arabidopsis homologs 86

3.2.2 Gene Ontology enrichment 87

3.2.3 Pathway analysis of protein-protein interactions surrounding the identified TFs 87

Table of Contents

3.3 Results 88
3.3.1 Global transcription factor expression in T200 and TME3 infected with SACMV 88
3.4 Discussion 108
3.4.1 General Introduction 108
3.4.2 The contrast between susceptibility (T200) and tolerance (TME3) 108
3.4.3 The downfall of susceptible T200 111
3.4.4 The rise of tolerance in TME3 120
3.4.5 The systemic response in cassava 126
3.4.6 Proposed ISR model in TME3 134
3.4.7 Transcription factors in literature 136
Chapter 4 138
4.1 T200, hallmarks of susceptibility 139
4.2 TME3, the model of Induce Systemic Resistance 141
4.3 Contribution to the field 143
4.4 Study Limitations 145
4.5 Future Work 147
References 149
Appendix 171

List of Figures

Figure 1-1:The bi-partite genome of SACMV showing the open reading frames and their positions 3

Figure 1-2:A MAPK cascade used in the recognition of PAMPs involving MPK3,4 and 6 to induce a rapid defence response in the host (Adapted from Cristina *et al*., 2010). 10

Figure 1-3: Overview of the signal transduction involved in biotic and abiotic stress responses in plants and the various components involved with mounting the appropriate response. ..... **Error! Bookmark not defined.**

Figure 1-4: A structural overview of the subgroups within the AP2/ERF TF family with their associated conserved motifs and biological functions. 16

Figure 1-5: : A generalised overview of the NAC TF family showing the (i) typical domain figuration of the NAC domain that is further divided into five different repetitive domains (A-E). 20

Figure 1-6:A generalised overview of the conserved domains present in the three subgroups encompassing the WRKY TF family. Adapted from Wang *et al*., 2014**Error! Bookmark not defined.**

Figure 1-7: Overview of RNAi and the proteins involved. RNAi can either be induced from dsRNA derived from exogenous viral RNA (blue) or dsRNA synthesised in the nucleus. 25

Figure 2-1:Overhead view of cassava landraces infected with SACMV at 67 dpi.. 34

Figure 2-2: Differential expression of *Me*WRKY transcription factors common to T200 and TME3. A.

Figure 2-3: A phylogenetic tree of WRKY transcription factors from Arabidopsis thaliana (At) and cassava [Manihot esculenta (Me)] sorted into taxonomic groups based on amino acid sequences (bootstrap value of a 1000). 46

Figure 2-4: Top 10 most enriched biological processes associated in up- and down-regulated genes in T200 and TME3 throughout SACMV infection. 48

Figure 2-5: STRING protein-protein interaction network centring around AtWRKY33 and 40 and their interacting partners.. 51

Figure 2-6: Predicted pathways and hormone responses of the differentially expressed (DE) WRKYs identified in (A) T200 and (B) TME3 at 12 dpi. 54

Figure 2-7: Predicted pathways and hormone responses of the differentially expressed (DE) WRKYs identified in (A) T200 and (B) TME3 at 32 dpi. 55
Figure 2-8: Predicted pathways and hormone responses of the differentially expressed (DE) WRKYs identified in (A) T200 and (B) TME3 at 67 dpi. 56
Figure 2-9: The number of up- and down-regulated DEGs involved in phytohormone processes in TME3 and T200. 58
Figure 3-1:The global representation of the number of DE TFs belonging to the bZIP, ERF, MYB, NAC and WRKY families in T200 and TME3 during the time course of SACMV infection.. 89
Figure 3-2:The comparison between (A) the number of different DE members belonging to the five TF families of interest and (B) the expression profiles of the common TFs DE at the same time point in both T200 and TME3 landraces. 91
Figure 3-3: Protein-protein interaction network centring around the Arabidopsis homologs of identified DE TF in T200 at 12 dpi. 94
Figure 3-4: Protein-protein interaction network centring around the Arabidopsis homologs of identified DE TF in T200 at 32 dpi. 98
Figure 3-5: Protein-protein interaction network centring around the unique Arabidopsis homologs only identified to be DE in T200 at 67 dpi.. 101
Figure 3-6: Protein-protein interaction network centring around the Arabidopsis homologs of identified DE TFS in TME3 at 12 dpi. 103
Figure 3-7: Protein-protein interaction network centring around the Arabidopsis homologs of DE TFS in TME3 at 32 dpi. T 105
Figure 3-8: Protein-protein interaction network centring around the Arabidopsis homologs of DE TF in TME3 at 67 dpi. 107
Figure 3-9: The proposed model of the network in TME3 at 32 that induces an ISR-like response which leads to SACMV tolerance 135

List of Tables

Table 1-1:Differences between CBSD and CMD. Information was acquired from Legg *et al*., 2011) 2
Table 2-1:The Log2Fold change in expression identified *Me*WRKY transcription factors in SACMV infected T200 and TME3 landraces at 12, 32 and 67 dpi and the function of their Arabidopsis homologs. 42
Table 2-2: Summary of other available global transcriptomic studies focusing on Geminivirus and Ipomovirus interactions in a susceptible (Green) or tolerant (Blue) plant species: 78
Table 3-1:Characteristics of TFs associated with biotic stress 85
Table 3-2:Transcription factors in T200 that form the centre of the underlying protein- protein interacting network at 12 dpi 93
Table 3-3:Identified TFs in T200 that form the centre of the underlying protein-protein interacting network at 32 dpi. 95
Table 3-4:Identified TFs in T200 that form the centre of the underlying protein-protein interacting network at 32 dpi. In red are genes that are uniquely found in T200 at 67 dpi and not 99
Table 3-5: Identified TFs in TME3 that form the centre to the underlying protein-protein interacting network at 12 dpi. 102
Table 3-6: Identified TFs in TME3 that form the centre of the underlying protein-protein interacting network at 32 dpi. 105
Table 3-7: Identified TFs in TME3 that form the centre of the underlying protein-protein interacting network at 67 dpi. 106

List of Abbreviations

Abscisic acid	ABA
Basic leucine zipper contain domain proteins	bZIP
Cabbage leaf curl virus	CalCuV
Cassava brown streak virus	CBSV
Cassava mosaic disease	CMD
CCR4-associated transcription factor	CAF
Days post inoculation	dpi
Dicer-like	DCL
Differenitally expressed genes	DEGs
Differentially expressed	DE
Effector Triggered immunity	ETI
Ethylene responsive factor	ERF
Ethylene	ET
Gene ontology	GO
Hypersensitive Response	HR
Hypersensitive response	HR
Jasmonate-zim	JAZ
Jasmonic acid	JA
Mitogen activated kinase 3	MAPK3
Myeloblastosis related proteins	MYB
NAM, ATAF and CUC,	NAC
Non-expressor of pathogenesis related protein 1	NPR1
nucleotide binding-leucine rich repeat	NBs-LRR
pathogen recognition receptors	PRR
Pathogen recognition receptors	PRRs
Pathogen triggered Immunity	PTI
Pathogenesis-related	PR
Rare cold-inducible gene 3	RCI3

List of Abbreviations

Reactive oxygen species	ROS
Reverse transcriptase quantitative PCR	RT-qPCR
Salicylic acid	SA
Small RNA	sRNA
South African Cassava mosaic virus	SACMV
Suppressor of BIR1	SOBIR1
Systemic aquired resistance	SAR
Tomato leaf curl New Delhi virus	TolCNDV
Tomato yellow leaf curl Sardina virus	TYLCSV
Tomato yellow leaf curl virus	TYLCV
Tomato yellow leaf curl virus	TYLCV
Transcription factors	TFs
Transcription factors	TFs
Ugandan cassava brown streak virus	UCBSV

Chapter outline

The structure for the following MSc dissertation is as follows:

Chapter 1 functions as a broad introduction to plant immunity and pathogen response in plant and familiarises readers with the previous studies to which this research is an extension of. Furthermore, the rationale and aim behind the need for this research is stated along with the and objective that were required to achieve

Chapter 2 places focus on the DE WRKY TF family and seeks to compare and contrasts the interactions occurring in T200 and TME3. Focus was placed on this family of TFs as there is a wealth of articles linking the WRKY family to both biotic and abiotic stress tolerances. However, there is limited observations surrounding WRKYs in cassava and no formal link has been presented between the WRKY family and tolerance to SACMV. In this chapter, the DE hub WRKYs are linked to the induction and suppression of hormone signalling in TME3 and T200 respectively. The consequences regarding induction and suppression of hormone signalling is subsequently linked to the induction of an immune response, connecting hub WRKYs to immunity in T200 and TME3.

Chapter 3 aims to expand the study on WRKYs from chapter 2 to four other transcription factors of interest in plant-pathogen interactions, namely the bZIP, ERFs, MYBs and NAC families. Here a more comprehensive analysis of SACMV-induced responses in T200 and TME3 is presented. The results suggest induced systemic resistance as a possible mechanism for tolerance in TME3, and is discussed herein.

Chapter 4 gives context to the study by linking the findings and conclusion presented in chapters 2 and 3 with regards to disease establishment in T200and the progression to tolerance over the course of SACMV infection in TME3. Furthermore, contributions of this study made to the field are stated, study limitation are discussed, and the solution to these limitations are proposed as future work.

Chapter 1

Literature Review

1.1 Cassava as a food security crop

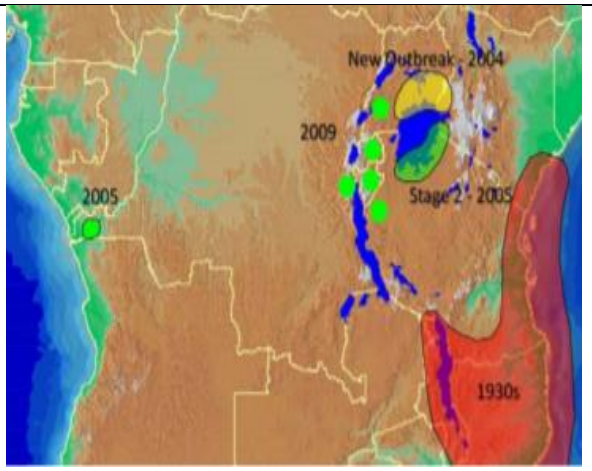
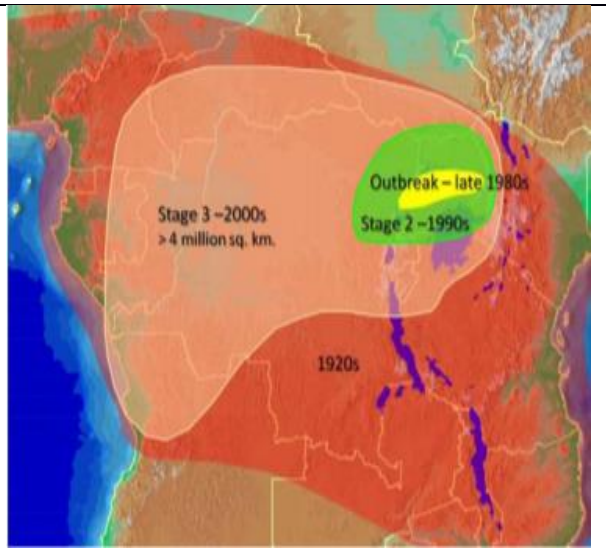
The *Euphorbiaceae* plant family shows high diversity in its physical characteristics. Notably within this family, there are several economically important crops, such as castor oil plant (*Ricinus communis*), rubber tree (*Hevea brasiliensis*) and cassava (*Manihot esculenta Crantz*) (Mwine *et al.*, 2010). Of the three aforementioned plants, considerable interest has been placed on cassava. This semi-perennial plant, originating from Latin-America, produces edible leaves and large starchy (20-40%) roots (Olsen and Schaal, 1999). Much of cassava's importance lies in the fact that it provides a source of calories to over 750 million people every year (FAO, 2011). The reliance on cassava within the tropics has led to cassava being named the third most important source of calories, following rice and maize (FAO, 2011). In addition to cassava being a valuable food source, there has been growing interest in the industrial sector in the use of cassava as a feedstock to produce biofuels. Much of cassava's success can be contributed to its ability to resist abiotic (environmental) stresses, granting it the title of food security crop (Ceballos *et al.*, 2010). Currently plant pathogens and pests account for approximately 40% of crop loss worldwide, and particularly begomoviruses pose the largest threat to cassava production (Varma and Malathi, 2003).

1.2 Cassava Diseases

There are two major diseases that afflict cassava, namely Cassava mosaic disease (CMD) and Cassava brown streak disease (CBSD) (Legg *et al.*, 2011). Despite both sharing the same viral vector, the whitefly (*Bemisia tabaci* Genn.), and both first being reported in Uganda, there are large differences between the two diseases (Table 1-1). Both CBSD and CMD have caused pandemics in the past 30 years. At its peak in 2005, CMD resulted in an estimated annual loss of one third of all cassava produced in Africa, across nine countries (Legg *et al.*, 2006). While the losses associated with CBSD have not been documented, it is known that CBSD can result in up to 100% losses in yield (Hillocks and Jennings, 2003). Currently, CMD poses a greater threat to cassava production in Africa than CBSD, due to the greater spread of the disease. This can be seen in both Angola and Cameroon, where there are published reports of CMD s from the first

pandemic to the present time (Kumar *et al.*, 2009, Akinbade *et al.*, 2010). Furthermore, CMD-causing viruses have also been identified in southern Africa, particularly in Zimbabwe and in South Africa (Berry and Rey, 2001). Due to this, the threat presented to cassava by CMD is ever present in South Africa and the viruses that lead to CMD need to be controlled.

Table 1-1Differences between CBSD and CMD (from Legg *et al*., 2011)

Disease	Virus family responsible	Pandemic	Symptoms	Parts affected in Africa
CBSD	*Potyviradae* (+)ssRNA	Multiple isolated hotspots form independently	Root necrosis Chlorosis of veins Brown streaks on stem	
CMD	*Geminiviradae* ssDNA	Spreading front originating from a single hotspot	Chlorosis Leaf curl Stunted growth	

CMD is caused by viruses belonging to the *Geminiviradae* family, named for their distinct twin icosahedral structure, that contains either a mono-partite or bi-partite ssDNA genome (Rybicki *et al.*, 2000). South African cassava mosaic virus is one of the nine species

belonging to the *Begomovirus* genus and was first identified within South Africa. SACMV is a viral recombinant between Tomato yellow leaf curl virus and East African cassava mosaic virus (Berrie *et al.*, 2001). SACMV has a bipartite genome consisting of DNA-A (2800 nt) and DNA-B (2760 nt), which encode at least six and two proteins, respectively (Figure 1-1). Both genomic components share a common region containing the origin of replication (Bisaro, 2006).

The proteins found in DNA-A can be categorised into three groups based on their function, namely DNA replication initiation (AC1 and AC3), regulation of host suppressions (AC4, AC2 and AV2) and viral assembly (AV1). DNA B has two proteins dedicated to viral transport, namely BV1 that encodes a nuclear shuttle protein and BC1 that codes for a long distance movement protein (Berrie *et al.*, 2001).

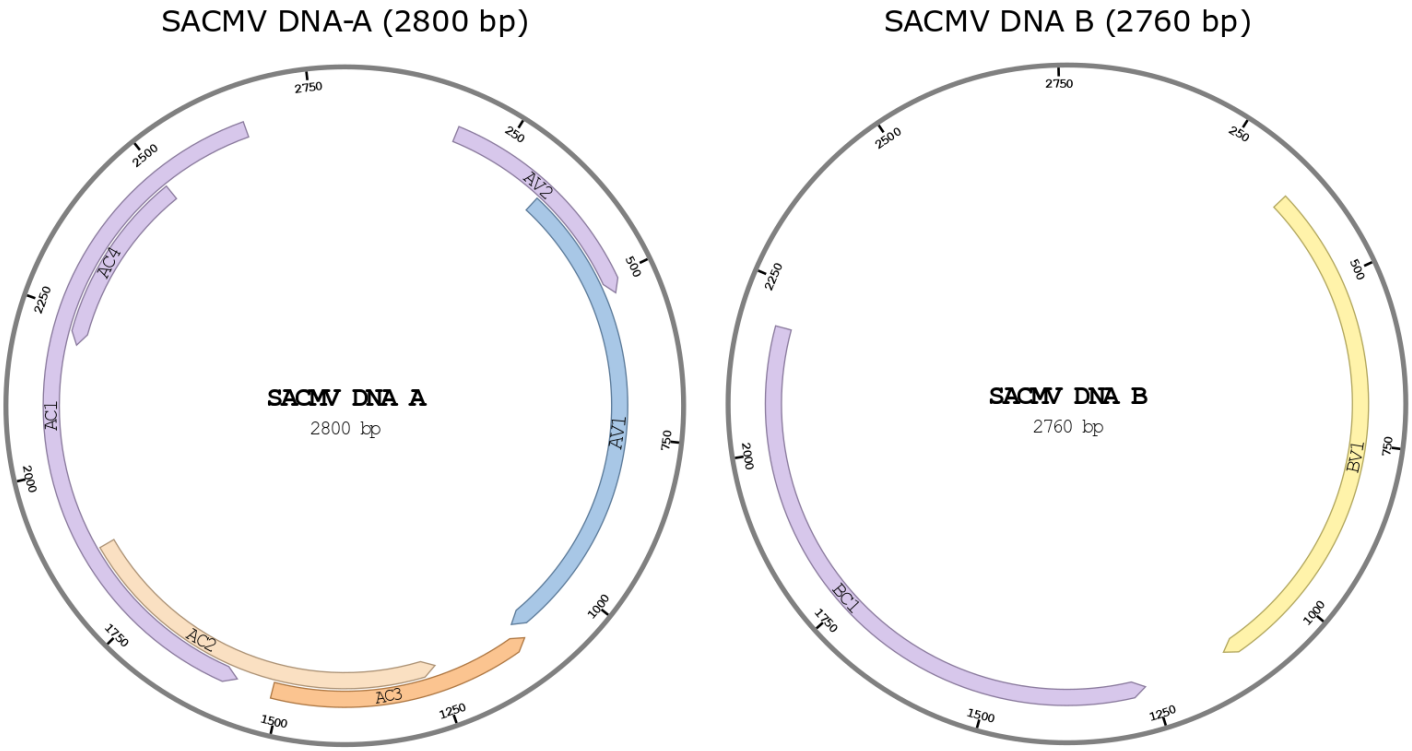


Figure 1-1The bi-partite genome of SACMV showing the open reading frames and their positions

South African cassava mosaic virus, much like all geminiviruses, is transmitted through the whitefly *Bemisia tabaci* and exclusively infects dicotyledonous plants leading to CMD (Briddon and Stanley, 2006, Rey *et al.*, 2012). The polyphagous nature of the whitefly is one of the contributing factors to the severity seen in CMD infections. The is due to the severity of CMD increasing when a mixed infection is present due to synergism occurring between viral species. Indeed, this was observed when East African cassava mosaic virus and African cassava mosaic virus co-infected cassava in the pandemic during the 1990s (Pita *et al.*, 2001). Furthermore, the whitefly vector facilitates the increase in viral finiteness observed among cassava begomoviruses through mixed infection. Although this is not directly due to the whitefly, the fact that it feeds on different dicotyledonous plants increases the likelihood of mixed infections. These mixed infections can increase viral fitness in two ways. Firstly, whitefly may introduce a species of begomovirus into an area already colonised by a different species of begomovirus. Alternatively, whiteflies may act as a vector for two separate begomovirus species simultaneously, by feeding on two different plant sources colonised by separate viral species (Legg and Fauquet, 2004). Insecticides were initially the method of choice for controlling CMD through targeting the vector, however chemical pesticides lead to environmental pollution and are shown to be neither cost- nor time efficient (Boulton, 2003). This has led to the need for an alternative strategy to deal with CMD. At present, there is considerable interest in the use of biotechnology and breeding programs to produce CMD resistance crops.

1.3 Resistance, tolerance and susceptibility

Using the genetic reservoir available in the wild species of cassava, Manihot glaziovii, resistance to CMD has been bred into commercial cassava lines (Jennings and Iglesias, 2002, Ntawuruhunga *et al*., 2006, Dixon *et al*., 2008, Nweke, 2009, Dixon *et al*., 2010). Due to the imprecise nature of breeding, the level of resistance in these cassava plants varies. These plants can be described as either resistant, tolerant or susceptible. Resistance is defined as the ability to limit and disrupt parasite replication (Fraile and García-Arenal, 2010). Susceptibility is defined as the plant lacking any resistance capabilities, thereby failing to mount an immune response against the virus leading. This leads to a successful, often chronic, infection (Bengyella *et al.*, 2015). In contrast to resistance, tolerance is defined as the ability of a host to reduce the replication of the parasite and the damages associated with pathogen infection (Fraile and

García-Arenal, 2010). For this reason, viral tolerance is understood to form part of the plant's basal immunity (Bengyella *et al*., 2015). This manifests in tolerant cassava varieties or landraces that show successful disease establishment and progression but, show milder symptoms and often exhibit recovery later in infection. Recovery is experienced when new expanding leaves at the top of the plant do not show symptoms despite the lower plant continuing to show symptoms. Interestingly in cantaloupe (Cucumis melo), when recovered leaves are re-infected with Cucurbit leaf crumple virus (a geminivirus), the virus fails to re-infect the new leaves (Hagen *et al.*, 2008). However, recovered leaf material showed successful disease establishment when infected with a different virus and showed greater symptom severity than leaf material infected with either virus separately (Hagen *et al*., 2008). From this study, it was concluded that tolerance is specific, and is a reversible process.

1.4 Tolerance and susceptibility

1.4.1 TME3 and T200 cassava landraces

TME3 is a cassava landrace originating from West Africa. This landrace is the product of resistance breeding programs initiated in response to the CMD pandemic in the surrounding areas. Traditionally, TME3 has been described as a resistant cassava landrace due to the presence of the CMD2 gene. CMD2 is an autosomal dominant gene in cassava that conveys resistance against CMD (Akano *et al*., 2002). Due to the repetitive nature of the cassava genome, sequencing of the CMD2 locus has proven challenging, and to date, the genes associated with this locus that may play a role in CMD resistance have not been elucidated (Prochnik *et al.*, 2012b, Bredeson *et al.*, 2016). Recently however, through genome-wide association analysis of 6128 genotyped, resistant cassava plants, some insights were revealed. It was found that two peroxidases and a thioredoxin were found to be associated with resistance. These genes were localized to chromosome 8 (now located on chr12 cassava phytozome v6.1), and are possibly thought to contribute to tolerance in TME3 (Wolfe *et al.*, 2016). Despite the presence of CMD2, it is proposed that TME3 should be designated as tolerant rather than resistant, as it demonstrates recovery, viral replication and persistent mild symptoms (Allie *et al.*, 2014).

TME3 was bred for resistance, however, resistance is not the only characteristic desirable to farmers. Features such as plant size, crop production, taste and availability would influence the use of the crop by farmers. T200, a South African landrace, is considered more desirable to

farmers with regards to crop production as it is known to produce a higher starch content, which would provide a better feedstock for bioprocessing and more calories for subsistence farmers and their livestock alike. However, T200 is susceptible to SACMV which leads to greater crop losses. These differences in phenotypes between T200 and TME3 are important to note, as the phenotype is the physical manifestations of the underlying molecular mechanisms occurring. By understanding how the molecular interactions relate to the contrasting phenotype, allows for the inference of the possible mechanisms that govern tolerance and susceptibility in T200 and TME3.

1.4.2 Key mechanisms associated with the development of tolerance

In addition to tolerant plants showing these phenotypic differences, tolerance in cassava also shows distinct molecular differences, compared with susceptibility in the same host. Firstly, tolerance is associated with a lower initial viral load upon detection(Gasura and Mukasa, 2010). However, it should be noted that decreasing viral load is not necessarily linked to recovery (Allie *et al*., 2014, Bengyella *et al*., 2015). Secondly, RNA silencing is associated with recovery, and if the virus successfully inhibits RNA silencing, the host will fail to recover (Bengyella *et al*., 2015). RNA silencing forms part of the plant's immune system, and is found in numerous wild plants where recovery is largely due to virus-specific degradation of mRNA (Al-Kaff *et al.*, 1998, Ratcliff *et al.*, 1999). Virus-derived RNA, that targets viral encoded genes, has been shown to increase in number during recovery *Nicotiana benthamiana* infected with African cassava mosaic virus (ACMV) (Chellappan *et al.*, 2004) and cassava infected with SACMV (Rogans *et al.*, 2016).

Upon pathogen invasion of a host, transcriptional reprogramming is observed both in susceptible and tolerant hosts (Allie *et al*., 2014). In a susceptible host, this is due to the virus reprogramming host function in order to optimize and prioritize its own replication (Hanley-Bowdoin *et al.*, 2004). Alternatively, the transcription reprogramming observed in a tolerant host is the consequence of the plant undergoing processes that allow it to adequately respond and fight the invading pathogen. In addition, recovery and reduced symptoms are associated with lower transcription reprogramming compared to susceptible hosts (Agudelo-Romero *et al.*, 2008, Allie *et al.*, 2014, Bengyella *et al.*, 2015). This is due to the host limiting transcriptional reprogramming by the pathogen, and in doing so reducing symptoms and viral replication. It is

hypothesized that tolerance is a culminations of various immune processes, therefore understanding plant immunity precedes the understanding of what mechanisms govern tolerance.

1.5 Plant immunity

1.5.1 PTI and ETI

Plant immunity can be largely divided into two main branches, pathogen-triggered immunity (PTI) and effector-triggered immunity (ETI). PTI acts primarily in basal defence, using pathogen recognition receptors (PRR) on the surface of the cell to detect pathogen-associated molecular patterns (PAMPs) (Zipfel and Felix, 2005). However, in order for pathogens to increase the success of infection, specific effectors have evolved to interfere with the mechanisms that govern PTI (Jones and Dangl, 2006). This, in turn, required plants to evolve a means to counter these pathogen effectors, leading to ETI. In contrast, ETI functions predominately within the cell through the use of resistance proteins (nucleotide binding-leucine rich repeat,, NLR), which recognize viral effectors (Matzinger, 2002). Together, PTI and ETI form a large part of the plants immune response, one functioning broadly to most pathogens, the other responds in a specific manner towards specific pathogens. However, this is not to be confused with basal immunity which is described as the combination of strongly induced PTI and weakly induced ETI, without effector-triggered susceptibility (Jones and Dangl, 2006).

PTI is found to be effective against pathogens that have not yet undergone host-specific adaption. This has been termed non-host resistance (Dodds and Rathjen, 2010). The recognition of these PAMPs is by PRR receptors, which fall into one of two classes, namely receptor-like kinases or transmembrane receptor-like proteins (Zipfel, 2008). PAMPs are traditionally the essential components that are present in a wide range of pathogens, such as flagellin in bacteria and chitin in fungi. Viral PAMPS and their cell membrane receptors have not been conclusively identified but are assumed to be viral proteins such as the coat protein. However, it should be noted that not all PAMPs are detectable by all species of plants. This is apparent in the detection of the bacterial elongation factor (EF-Tu), which only is detectable in the *Brassicaceae* family (Zipfel *et al.*, 2006). Upon detection of PAMPs, the host undergoes an increase in Ca2+ ions and reactive oxygen species (ROS) and the induction of mitogen-activated protein kinases (MAPKs). This then leads to transcriptional reprogramming (Boller and Felix, 2009). This transcriptional reprogramming can induce defence responses that are typically, but not entirely, geared towards

a wide range of pathogens. This is seen in the expression of chitinases against fungal expression and callose deposition in cell walls to prevent pathogen entry within the cell (Bednarek *et al.*, 2009) .

Although many of the PTI and ETI signalling mechanisms and responses overlap, the response produced by ETI is generally stronger and more rapid (Boller and Felix, 2009, Tsuda *et al.*, 2009). If PTI is considered to be a generalised response, then ETI can be regarded as a specific immune response, directed against pathogens adapted to the host (Dodds and Rathjen, 2010). This is seen in the highly diverse number of intracellular receptors, known as resistance proteins (R proteins) that either directly or indirectly recognise pathogen effectors (Van Der Biezen and Jones, 1998, Dangl and Jones, 2001). Unlike most protein families, these R proteins are characterised by their function as detectors of effectors rather than their structure. However, both the leucine-rich repeat (LRR) and nuclear binding site (NBS) domains are common in most R proteins (Hammond-Kosack and Parker, 2003, Liu and Ekramoddoullah, 2004). The LRR serves in protein-protein interactions and functionally serve as the receptor for viral proteins (Jones and Jones, 1997). The NBS serves as the region in which ATP binding and hydrolysis occurs, undergoing activation (Tameling *et al.*, 2002). The activation of R proteins often leads to signal transduction pathway leading to many of the defence pathway seen in ETI, in addition to the hypersensitive response- a form of PCD (Dangl and Jones, 2001). Both these branches of immunity exist as a network of signalling molecules. These signalling molecules serve to transduce the stress signal either within a cell or between cells.

1.5.2 ROS

The ROS regulatory system is conserved in plants, and is used as both a means to control plant defence and to reduce damage in cells (Pandey *et al*., 2015). ROS is both beneficial and harmful to plant cells. On the one hand, evidence has suggested that ROS integrates signals from multiple pathways, such as MAPKs, Ca2+ and phytohormones (Mittler, 2017, Suzuki *et al*., 2013). On the other hand, if left unchecked, ROS can lead to damage within the cell. This is seen with H_2O_2, which is able to attack fatty acids and degrade nucleic acids, ultimately leading to cell death (Sharma *et al.*, 2012). Due to the wide range of functions in which ROS participates, it is important to focus in on the defence-related processes. An example of ROS involvement in defence response can be seen in the kinase, LecRK-IX.2. Upon stimulation of an upstream PRR

protein, this kinase leads to defence mechanisms. However, LecRK-IX may initiate phosphorylation of a ROS NADPH oxidase, thereby initiating ROS production via Ca2+ dependent kinases (Luo *et al.*, 2017). This would mean that ROS is induced by pathogen recognition and likely is involved in its modulation. In addition to this, ROS may act as a modulator of immunity by interacting with phytohormones. ROS dependent signalling may be involved with distal signals that go about enhancing abscisic acid (ABA) synthesis in response to heat (Suzuki *et al.*, 2013). Although this enhancement is due to an abiotic stimulus, it has well been established that there exists a convergence between biotic and abiotic signalling pathways (Suzuki and Katano, 2018).

1.5.3 Mitogen-Activated Protein Kinases (MAPKs)

MAPKs are ubiquitous molecules found throughout eukaryotes and play a key role in signal transduction. Upon detection, receptors phosphorylate MAPK4KKs (MAP kinase kinase kinase kinases), which then phosphorylate MAPK3KKs sequentially until MAPKs are phosphorylated and activated (Dan *et al.*, 2001). The term MAPKs refers to all these kinases that form this signalling network, making up the MAPK cascade. In the evolution of plant pathways, signalling molecules like the MAPKs evolved more rapidly at the beginning (near the periphery of the cell), and at the end of the pathway (Ichimura *et al.*, 2002). This evolution at the start of the pathway led to the diversification of signal perception in a plant, whereas the diversification downstream lead to the ability to transduce the signal to multiple responsive pathways. However more importantly, the lack of diversity in molecules at the centre of the network led to overlap in signalling pathways, like those experienced between biotic and abiotic response pathways.

This is evident from the functions of *At*MAPK3, -4 and -6. MPK3 and 6 have been extensively studied in Arabidopsis, in relation to their partially redundant function in innate immunity (Tena *et al.*, 2011). MPK3 and 6 function both in the induction of short-term rapid response as well as long term-responses. With regard to rapid responses, both MPK3 and 6 are found in the pathway responsible for the induction of defence against PAMPs, like flagellin 22 (Figure 1-2, Boller and Thomas, 2009). Here the activation of MP3 and 6 leads to the phosphorylation of ACC synthases, increasing its stability, and resulting in the accumulation of ethylene (ET) (Liu and Zhang, 2004).

In contrast, long-term responses in *At*MPK3 and 6 occur via different pathways. *At*MPK3 and 6 demonstrate prolonged and enhanced expression leading to camalexin biosynthesis (Ren *et al.*, 2008). Much like *At*MPK3 and 6, *At*MAPK4 is also induced by PAMP recognition, however, *At*MAPK4 follows a different signalling pathway (Felix *et al.*, 1999). Similarly, *At*MPK4 is able to induce camalexin synthesis by forming the MSK1 complex with the transcription factor (TF), *At*WRKY33 and other proteins in the nucleus (Figure 1-2, Anderson *et al.*, 2005). However, where MPK3 and 6 lead to the accumulation of ET, *At*MPK4 negatively regulates salicylic acid (SA) accumulation and biosynthesis, and consequently, is a suppressor of systemic acquired resistance (SAR) (Petersen *et al.*, 2000). From this it is evident how *At*MAPKs form the cornerstone of intracellular signalling, transducing signals induced by pathogen infection to the nucleus.

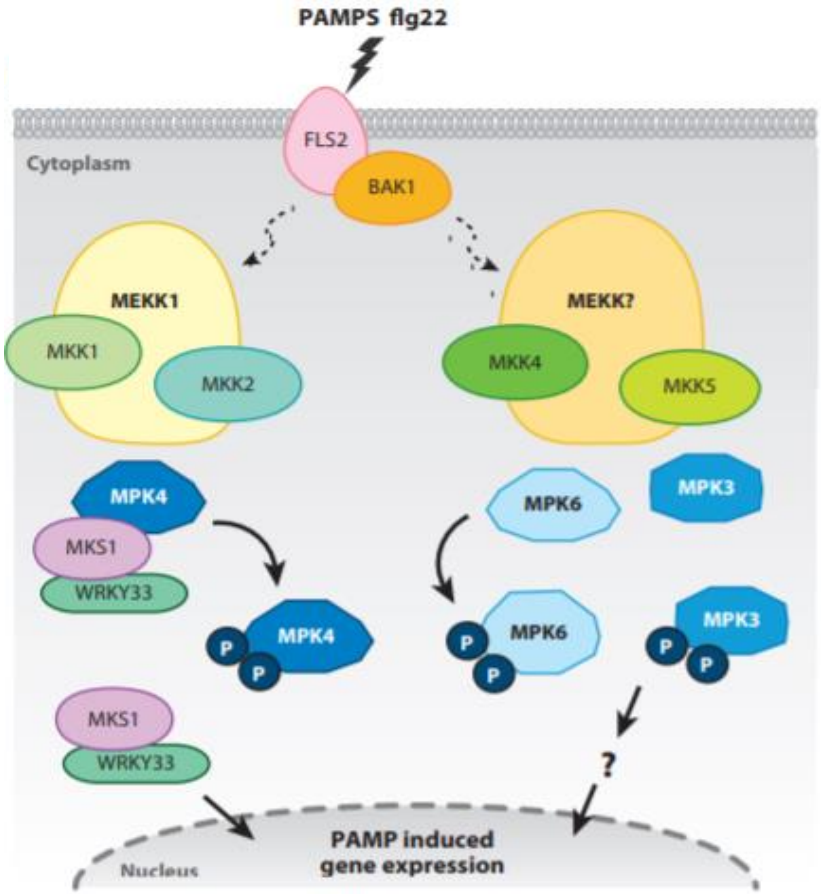


Figure 1-2 A MAPK cascade used in the recognition of PAMPs involving MPK3,4 and 6 to induce a rapid defence response in the host (adapted from Cristina *et al.*, 2010).

However, signal transduction is not a process that is confined within a cell. In order for a plant to mount an appropriate defence response, intercellular signalling is essential to induce the systemic resistance required for tolerance. Phytohormones serve as these systemic intercellular signalling molecules within plants.

1.6 Hormone signalling in plant immunity

Phytohormones function as systemic signalling molecules, sending messages from one part of the cell to another (Coupe *et al.*, 2005). In particular, phytohormones play a considerable role in the regulation and signalling of biotic (and abiotic) stress responses (Fujita *et al.*, 2006). This is achieved in plants through hormone cross-talk, which entails the summation of the synergistic and antagonistic interactions occurring between phytohormones (Bostock, 2005). Signal antagonism evolved to allow plants to undergo signal prioritisation (i.e. when pathway A is prioritised through inhibition of pathway B) (Berens *et al.*, 2017). In contrast, pathway synergism is a strategy employed by networks to increase the accuracy of signalling by requiring the two different signals concurrently in order to initiate the pathway (Berens *et al.*, 2017). These sort of interactions can be seen particularly in the four most prominent hormone signalling networks in plant immunity, namely abscisic acid ABA, ET, JA and SA.

Of these four hormones, SA and JA are considered the most central to immune responses, whereas ABA and ET are seen to be modulators of immunity, through interaction with either JA and/or SA. In order to regulate the dynamics of immunity within plants, cross-talk between JA and SA takes place in an antagonistic manner (Glazebrook, 2005). This antagonism allows the plant to prioritize responses to either a necrotrophic or biotrophic response through either JA or SA signalling, respectively (Glazebrook, 2005). This prioritization is essential because of the hypersensitive response (HR), a type of PCD, which is used to fight biotrophic pathogens, but would be beneficial to necrotrophic pathogens (Glazebrook, 2005). It should be noted that for any hormones to contribute to a defence response, they are required to either interact with JA or SA (Pieterse *et al.*, 2009). This means that JA and SA are the two most important hormones needed for the induction of a defence response.

JA biosynthesis comprises a multistep process that begins with 12- oxophytodienoic, which is a lipid-derived component originating from the chloroplast membrane (Wasternack and Hause, 2013). However, unlike the other hormones, JA requires further processing before it may become functionally active within the plant. When isoleucine is conjugated to JA (JA-Ile), through the JAR1 receptor, JA is perceived and becomes functionally active as a signalling molecule (Wasternack and Hause, 2013). In Arabidopsis, JA-Ile perception is performed by *At*COI1 which leads to 26S proteasome ubiquitination and degradation of various proteins in the *At*JAZ family. The *At*JAZ family functions as inhibitors of JA signalling (Wasternack and Hause, 2013). Once the JAZs are degraded, MYC TFs, particularly *At*MYC2, are able to bind to genes and induce expression of JA response genes (Wasternack and Hause, 2013). Through MYC2 and other TFs, JA is able to induce plant defence, and inhibit growth, seed germination and plant development (Davies, 2010).

In contrast with JA, SA can be produced from one of two pathways, either the phenylalanine ammonium lyase pathway or the isochorismate pathway (D'Maris Amick Dempsey *et al.*, 2011). Genes involved in both pathways are found in a moss (*Physcomitrella patens*), suggesting that SA signalling is conserved within all land plants (Ponce De Leon *et al.*, 2012, Berens *et al.*, 2017). Pathogenic infection of plants leads to the accumulation of SA in eudicots but not monocots (Yang *et al.*, 2011, Kim *et al.*, 2014). This accumulation of SA leads to the conserved induction of immune response against biotrophic and hemibiotrophic pathogens in angiosperms (Berens *et al.*, 2017). Interestingly, this induction of SA signalling can be induced through the external deposition of either SA or a SA analogue (such as benzothiadiazole), consequently inducing an immune response (Grellet-Bournonville *et al.*, 2012). The main function of SA is to induce SAR through PRRs (Cao *et al.*, 1994, Fu and Dong, 2013) . However, SA is also known to inhibit ET, reverse the effects of ABA signalling, and increase the longevity of flowers (Davies, 2010). Alternatively, plants may undergo ISR which is phenotypically similar to SAR but is independent of SA signalling, and is rather established through JA and ET signalling (Pieterse *et al.*, 1996, Pieterse *et al.*, 1998).

ABA is a modulator of immunity, and is synthesized from carotenoid precursors. It is primarily involved in the abiotic stress response, through highly conserved signalling pathways (McAdam *et al.*, 2016). Due to the overlap observed between biotic and abiotic signalling

pathways, ABA is also known to contribute to pathogen responses and defence. For example, ABA signalling is associated with stomatal closure in the presence of the flagellin in Arabidopsis. However, this response requires SA signalling to undergo full stomatal closure (Melotto *et al.*, 2006). Despite providing stomatal immunity with regards to bacterial invasions, ABA shows antagonism and suppression of SA, thereby reducing resistance (Yasuda *et al.*, 2008). In fact, SA silencing by ABA is so widespread within angiosperms, that some pathogens establish infection by exploiting host ABA production or producing their own ABA (Gong *et al.*, 2014). It appears that ABA may be traditionally associated with JA and not SA signalling. In Arabidopsis, *At*MYC2 functions both as a positive regulator for ABA signalling and JA-induced in responses to herbivores (Abe *et al*., 2003, Anderson *et al*., 2004). This is notable as *At*MYC2 is described as the master regulator of most aspects of JA signalling. The induction of ABA by *At*MYC2 may serve as another possible mechanism by which SA is suppressed, allowing JA signal transduction to occur. It is therefore clear that ABA plays an interesting role in biotic stress response, acting as a modulator of immunity through interacting with JA and SA

It is known that ET is synthesized from S-adenosylmethionine, and functions in both plant development and the stress response (Ju *et al.*, 2015). Although ET is involved with stress responses, ET can either act as a positive or negative regulator of immunity, depending on which pathogen is infecting the plant system. In soybean, ET is associated with the increased disease severity towards the necrotrophic bacterium, *Rhizoctonia solani*, but in rice, promotes resistance towards *R. solani* through ACS2 (Hoffman *et al*., 1999, Helliwell *et al*., 2013). This contrast in the effect of ET signalling is likely due to the effect ET has on modulating JA and SA rather than the direct effect derived from ET signalling (Berens *et al*., 2017). Interestingly, ET is able to regulate both JA and ET through suppression of ET signalling pathways (Tsuda *et al*., 2009). However, since ET signalling has been shown to develop after both JA and SA signalling (Wang *et al.*, 2015), it has been suggested that JA and SA evolved to integrate ET as a critical immune modulator (Beren *et al*., 2017).

All of these hormones are able to circulate through the plant in order to establish immunity throughout the plant, however, systemic immunity has only been described through ISR and SAR. Exploring the nuances that govern these two systemic immune signalling mechanisms is necessary to potentially understand tolerance. Hormones may alter gene

expression by transducing the signal downstream through various kinases and phosphatases, then induce TFs to initiate transcriptomic reprogramming and activate stress responses (Figure 1-3, Taj *et al.*, 2010, Hernandez-Garcia and Finer, 2014).

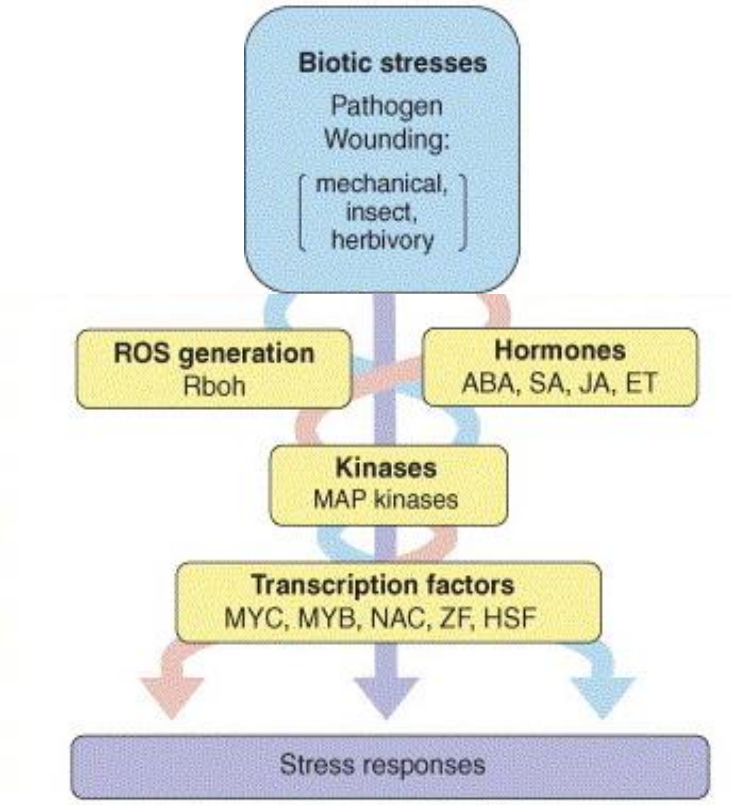


Figure 1-3: Overview of the signal transduction involved in biotic and abiotic stress responses in plants and the various components involved with mounting the appropriate response. Both biotic and abiotic stress pathways incorporate the use of ROS, hormones and kinases to transduce the signal to TFs. It is then up to TFs to lead to changes in expression of targeted to genes in order to respond to the stress in an appropriate manner (adapted from Fujita *et al*., 2006).

1.7 Transcription factors

TFs are defined as master regulatory proteins that alter gene expression in response to stress or signalling stimuli. They achieve this through binding to upstream cis-regulatory elements in their specific targets (Chen *et al*., 2002, Riechmann, 2006). Conversely, TFs may alter downstream hormone levels through the expression of target genes that either inhibit or induce hormone biosynthesis (Schweizer *et al*., 2013, Zhou *et al*., 2011). Upon infection, plants undergo transcriptomic reprogramming, leading to either susceptibility or tolerance towards the disease. TFs can be further divided into families based on the presence of certain conserved amino acid sequences. The common TF families that are involved in regulating pathogen response (PR) genes are WRKYs, basic leucine zipper contain domain proteins (bZIPs), myeloblastosis related proteins (MYBs), ethylene-responsive element binding factors (ERFs), and no apical meristem, Arabidopsis transcription activation factor andcup-shaped cotyledon (NACs) (Chen *et al.*, 2002, Singh *et al.*, 2002, Van Verk *et al.*, 2009).

1.7.1 Basic leucine zipper contain domain proteins

The bZIPs are one of the largest TFs families in the plant kingdom. They uniquely contain the BZIP domain, which they used to bind to the G-box element within the promoters of genes that respond to both biotic and abiotic stress (Duplessis *et al.*, 2009).Currently, 75 bZIP TFs have been identified in Arabidopsis, falling into 13 subgroups (Corrêa *et al.*, 2008). In contrast to Arabidopsis, 77 bZIPs were identified in cassava that fall into 10 subgroups. Furthermore, it is suggested that in cassava, the bZIPs in subgroup G and D function as the ancestral genes that gave rise to the other bZIPs in cassava (Hu *et al*., 2016). bZIP TFs are known to have many functions ranging from seed maturation and floral development, to the regulation of biotic and abiotic stress (Jakoby *et al.*, 2002). However, in A. thaliana, most of the bZIPs that are involved in both biotic and abiotic stress responses likely reside within the A subgroup, as most of the members are known to contribute to ABA signalling (Guan *et al.*, 2009). However, much of the regulation seen in bZIPs are from self-regulatory networks, with other interacting proteins regulating their subcellular location, activity and defence response functions (Jakoby *et al.*, 2002, Fietto *et al.*, 2013). The bZIP TFs appear to be more prevalent in abiotic response, however, this may still be relevant as abiotic signalling and biotic signalling are known to overlap. For instance, *At*bZIP17 is proteolytically processed upon salt stress, leading to it entering the nucleus

and up-regulating salt tolerant genes (Liu *et al.*, 2007). while *At*bZIP10 is found to be involved with regulating basal defences and is a positive regulator of the hypersensitive response, through LSD1 interactions (Kaminaka *et al.*, 2006).

1.7.2 Ethylene-responsive element binding factors

The AP2/Ethylene response factor TF family is diverse, housing several subfamilies such as the dehydration responsive element binding (DREB) and ERF. The ERF subfamily contain the 58-59 aa ERF DNA binding domain (Figure 1-4), which is able to bind a diverse array of cis-regulatory elements, however, show a greater affinity for GCC-boxes (A/GCCGAC) and DRE/CRT elements (Büttner and Singh, 1997). DREBs are more commonly associated with abiotic stress responses, following an ABA-independent signalling pathway, and also bind to DRE/CRT elements (Sakuma *et al.*, 2002). Extensive research into this TF family has revealed 122 ERFs in Arabidopsis, 139 in rice, and 147 in cassava (Fan *et al.*, 2016, Nakano *et al.*, 2006). Interestingly, GCC boxes have been found in multiple PR genes, in addition to JA and SA inducible genes (López *et al.*, 2008). This means that ERFs are particularly involved in the regulation of immunity in plants. ERFs function both modulating and responding to ET signalling. This is seen where ERFs enhance the functionality of both ACS and AC0, two critical enzymes in ET biosynthesis (Gu *et al.*, 2017). Similarly, *At*ERFs have been found to be the

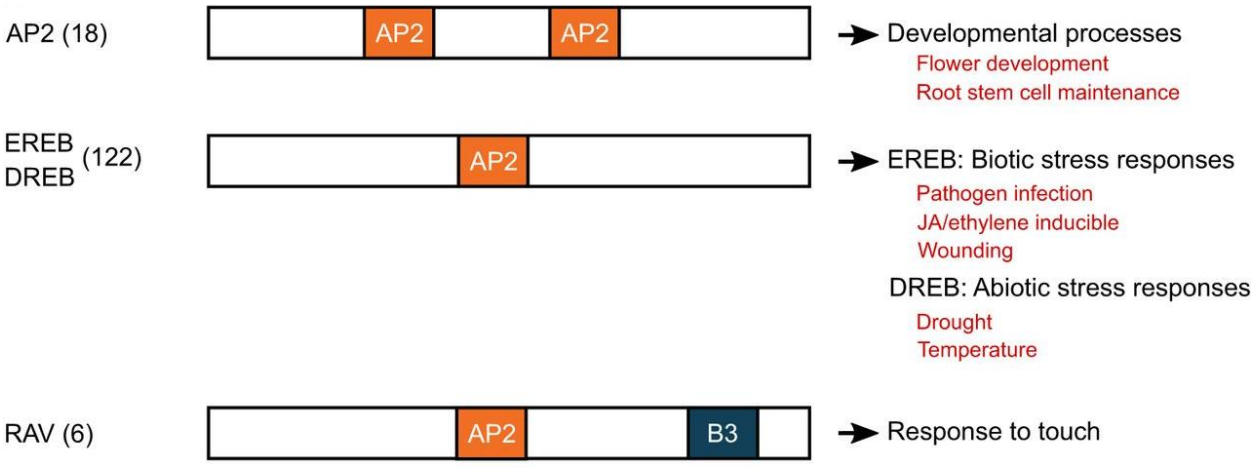


Figure 1-4: A structural overview of the subgroups within the AP2/ERF TF family with their associated conserved motifs and biological functions. The number in brackets shows the number of proteins identified in A. thaliana in each group (adapted from Heyman *et al.*, 2018).

convergent point between the JA, SA and ET hormone signalling pathways (Zhang *et al.*, 2004a). In particular, this has been demonstrated by the interaction between *At*ERF4 and downstream defence-related genes within the ABA and ET signalling pathways (Yang *et al.*, 2005). This idea of ERFs being involved in defence responses is compounded when one considers how closely ERFs and WRKYs interact with each other.

In Arabidopsis, it has been found that *At*ERFs play a key role in the functions of several WRKYs, as seen in *At*ERF6 and *At*WRKY33, which both regulate the accumulation of the antimicrobial compound, camalexin (Meng *et al.*, 2013). Furthermore, in Arabidopsis, WRKY22, WRKY29 and ERF104 are all regulated by MPK6, which suggests that there is an overlap between WRKYs and ERFs in terms of function (Bethke *et al.*, 2009).

*At*ERF1B is a well-studied TF that requires both ET and JA signalling in order to undergo transcriptional activation, and in doing so, confers resistance towards the necrotrophic pathogen, *B. cinerea* (Lorenzo *et al*., 2003). Similarly, *At*ERF9 contributes to resistance against *B.cinerea*, by functioning as a repressor of PDF1.2, a marker for JA/ET signalling (Maruyama *et al.*, 2013). *At*ERF5 and 6 are found to overlap, and therefore exhibit redundant functions in modulating defence against *B. cinerea* in Arabidopsis (Moffat *et al*., 2012). The defence roles of *At*ERF5 and 6 seem to be a consequence of the induction of JA/ET-responsive genes, coupled with the suppression of SA signalling, thereby prioritising JA-mediated defence responses (Moffat *et al*., 2012). *At*ERF109, much like *At*ERF5 and 6 is also involved with hormonal cross-talk. However, *At*ERF109 regulates JA-auxin signalling. *At*ERF109 acts as a highly responsive JA-induced TF that binds to the promoters of the auxin biosynthesis genes ASA1 and YUC2 (Cai *et al.*, 2014). Although *At*ERF109 is generally not associated directly with immunity, auxin signalling is linked with negative regulation of resistance (Robert-Seilaniantz *et al.*, 2011).

1.7.3 Myeloblastosis related proteins

MYBs are not unique to plants, as they were first identified as an oncogene (with retroviral origins) in animals (Zhang *et al.*, 2010). Despite being found in these two divergent kingdoms, all MYBs share a conserved MYB domain comprising up to three different types of repetitive sequences (Ogata *et al.*, 1996, Ito, 2005) . MYBS are classified into different subgroups based on the number and types of repeat elements present within their MYB domains, namely the R1R2R3 group, the R2R3 group and the MYB-like group (which contain only a single repeat

element) (Yanhui *et al.*, 2006). Of these three groups, however, the R2R3 MYB subgroup is involved in a variety of plant-specific functions. These functions range from metabolism, developmental processes, cell differentiation and response to biotic and abiotic stresses (Dubos *et al*., 2010).

In Arabidopsis, 126 MYB TFs belonging to the R2R3 subgroup have been identified, whereas a recent study has found that there are 166 in cassava (Zhang *et al.*, 2010, Liao *et al.*, 2016). To date, there is a greater wealth of knowledge surrounding the role of *Me*MYBs in abiotic, rather than biotic stress responses, in cassava. Despite this, the literature regarding the *At*MYBs is more comprehensive compared to the *Me*MYBs. In these studies have shown that *At*MYB30, *At*MYB60 and *At*MYB96 are involved in stress responses, but *At*MYB60 and *At*MYB90, in particular, exert their function through ABA signalling (Cominelli *et al.*, 2005). Interestingly, *At*MYB genes are regulated by both miRNAs, a component in the plant immune system, and transacting silencing RNAs (ta-siRNAs), which are small interfering RNA that regulate gene expression through post-translational gene silencing (Vazquez *et al.*, 2004).

1.7.4 No apical meristem, Arabidopsis transcription activation factor and cup-shaped cotyledon

NAC TFs traditionally have a conserved 150 aa DNA binding domain found at the N-terminus and a variable protein-binding domain found towards the C-terminus that contributes to its role in regulating transcription of genes (Figure 1-5, Hu *et al*., 2015, Kikuchi *et al*., 2000, Olsen *et al*., 2005). Interestingly, unlike the other TFs of interest, there are fewer identified NACs in cassava (96) than in Arabidopsis (117) (Nuruzzaman *et al.*, 2010, Hu *et al.*, 2015). The *Me*NAC TFs fall into at least seven different subfamilies, based on their functions. However, based on phylogenetic analysis, NACs can be placed into 16 subgroups (Pei *et al.*, 2013, Hu *et al.*, 2015). *At*NAC TFs have many well-characterised functions, some of which are plant development, cell division, and response to biotic and abiotic stresses (Olsen *et al.*, 2005, Kim *et al.*, 2006b, Wu *et al.*, 2009b).

In particular, in biotic stress, *At*NAC TFs contribute such a significant role to plant immunity that pathogens such as *Phytophthora infestans* have developed the effector, LxLR, to suppress functioning of NTP1 and NTP2 (McLellan *et al.*, 2013). Viruses, too, have developed means to suppress NAC TFs by inactivation by binding to their viral proteins (Ren *et al.*, 2000,

Wang *et al.*, 2009). However, it has been found that NACs can interact directly with viral proteins, and in doing so, suppress their function (Ren *et al*., 2000). *At*ATAF1 is a key TF in modulating defence responses in Arabidopsis, through repression of ABA signalling during pathogen invasion (Mauch-Mani and Flors, 2009). Since ABA signalling plays a role in abiotic stress responses, it has been further suggested that *At*ATAF1 functions as the molecular switch between biotic and abiotic stress responses in the host (Mauch-Mani and Flors, 2009). *At*NAC19 and 55 both promote resistance against *B.cinerea*, in Arabidopsis (Bu *et al.*, 2008). It was further determined that these two NAC TFs are induced by JA signalling via the main *At*COI1 and

*At*MYC2 dependent pathways, which is how they regulate downstream defence response genes (Bu *et al*., 2008).

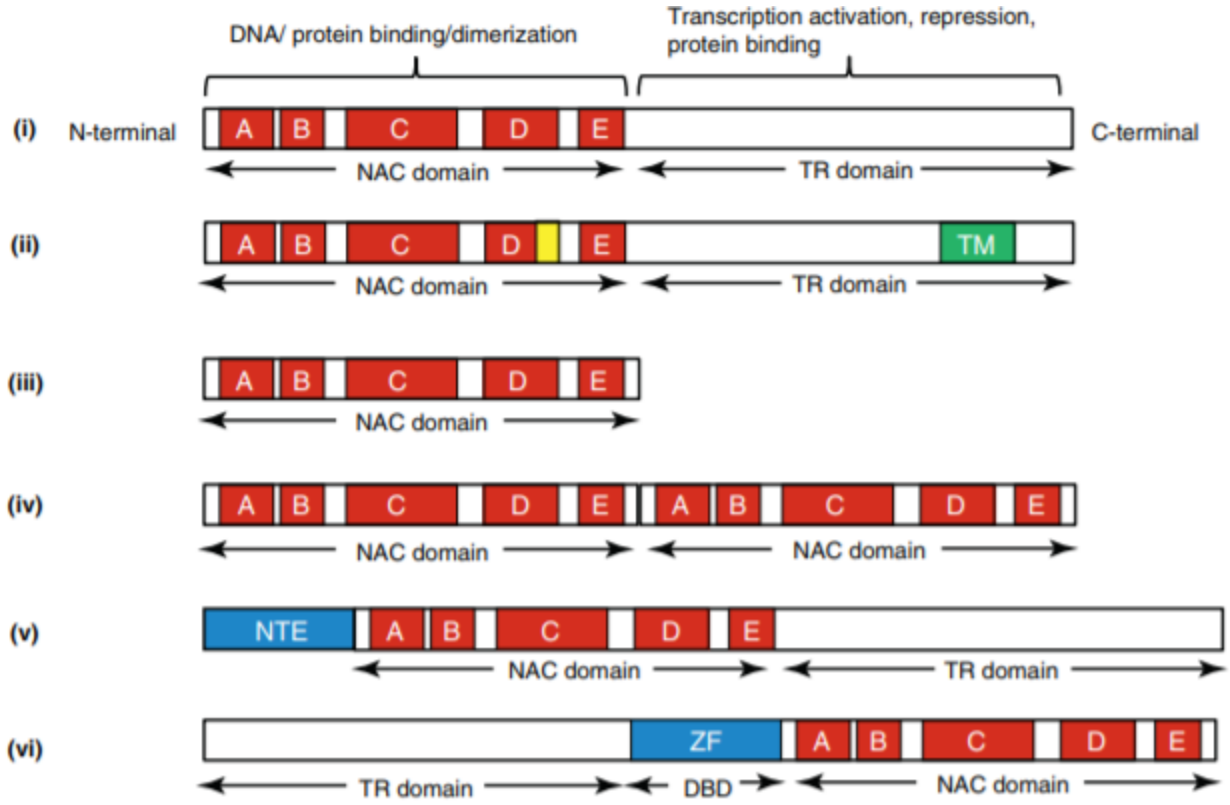


Figure 1-5: An overview of the NAC TF family showing the (i) typical domain figuration of the NAC domain that is further divided into five different repetitive domains (A-E). Sometimes variation can be seen in the family, where some TFs display (ii) a repressor domain (yellow) and a transmembrane domain (TM) at the C terminal. Further variation can be seen in NAC proteins containing (iii) only the NAC domain or (iv) a duplicated NAC domain. Lastly (v) N-terminally extended (NTE) and (vi) zinc finger (ZF) motifs have also been known to make an appearance with some NAC proteins (Prochnik *et al.*, 2012b).

1.7.5 WRKYs

WRKYs have been identified in multiple plant species, with the fewest (45) being currently characterized in barley (*Hordeum vulgare*), whereas soybean (*Glycine Max*) currently has the most identified with 182 (Mangelsen *et al.*, 2008, Bencke-Malato *et al.*, 2014). In 2014, 85 WRKYs were identified in cassava, 11 more than those found in Arabidopsis (Ulker and Somssich, 2004, Wei *et al*., 2014). While WRKYs are a TF family found in slime moulds and protists, they are mostly exclusive in plants, and contain the conserved WRKYGQK motif and zinc-finger-like motifs binding to the W-box (TTGAC(C/T)) cis-regulatory element (Rushton *et al*., 2010, Pandey and Somssich, 2009). Currently, there are three major groups of WRKYs based on the domain composition, with several sub-families (Eulgem *et al*., 2000). Group I is characterised as having two WRKY domains and two C2H¬2 zinc-finger domains. Group II and III both have a single WRKY domain, however, the zinc finger domains differ between the two groups. Group II shares the same C2H¬2 zinc-finger domain as Group I, but it is only present as a single copy, whereas Group III has a single C_2HC domain (Figure 1-6, Eulgem *et al*., 2000).

WRKY TFs are critical to plants as they function as negative regulators of pathogen responses within plants via PTI and ETI (Rushton *et al*., 1996). Since 1996, a number of studies have shown that WRKYs are involved in innate immunity and defence gene activation against biotic stress (Pandey *et al.*, 2010, Bakshi and Oelmüller, 2014). Furthermore, it has been seen that numerous members of the WRKY family interact with various hormone signalling molecules required in signalling responses (Rushton *et al*., 2010, Duan *et al*., 2007, Pandey *et al*., 2010). Within the WRKY family, there are several key TFs involved in regulating defence responses. *At*WRKY18 and 40 are found to be rapidly induced by ABA signalling, where WRKY40 acts as a transcriptional repressor of *At*WRKY60 (Chen *et al*., 2010). *At*WRKY40 is particularly fascinating, as it is able to interact with ABA, JA and SA (Wang *et al.*, 2008, Chen *et al.*, 2010, Pandey *et al.*, 2010). In addition, *At*WRKY33 and 53 are both phosphorylated by MAPKs, and play an essential role in signal transduction during defence responses (Miao *et al.*, 2008, Xie *et al.*, 2010). Their phosphorylation by these signalling molecules implies their downstream application in activating defence pathways. *At*WRKY70 is quite a prominent activator of SA-inducible genes and equally a suppressor of JA-responsive genes (Li *et al*., 2004). This would make *At*WRKY70 involved in the signal prioritization seen in the JA-SA

antagonism in the plant kingdom. It has also been implicated in the coordination of basal resistance against *P. synringae* (Hu *et al*., 2012). Interestingly, *At*WRRKY53 and 70 seems to form part of a mutually regulated negative feedback loop with *At*WRKY22 and *At*WRKY6 (Miao *et al*., 2013). Both *At*WRKY22 and *At*WRKY6 are involved in cold responses. *At*WRKY6 is most closely associated with abiotic stress and senescence, whereas *At*WRKY22 is regulated by H_2O_2 and MAPKs (Thomas *et al*., 1999, Rosenvasser *et al*., 2006, Miao *et al*., 2007).

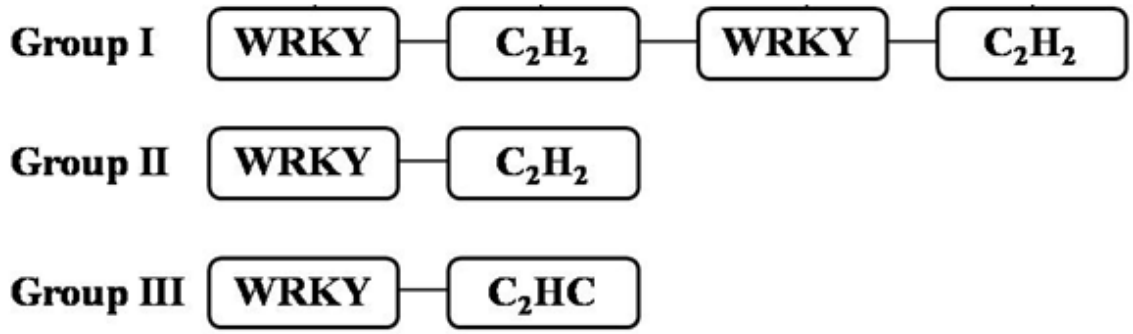


Figure 1-6: An overview of the conserved domains present in the three subgroups encompassing the WRKY TF family (adapted from Wang *et al*., 2014).

As mentioned above, TFs play a considerable role in mounting the appropriate response towards pathogens. It is this fine interplay, seen in regulatory systems and feedback loops between TFs and hormones, that tightly regulates downstream processes.

1.8 Induction of the plant immune system

1.8.1 Systemic Acquired Resistance and Induced Systemic Resistance

SAR is defined as the enhanced resistance experienced in a plant following a localised microbial infection, centring around the enhanced expression of PR proteins and SA signalling (Pieterse *et al*., 1996, Fu and Dong, 2013). This enhanced resistance is directed towards a broad range of pathogens, particularly hemi- and biotrophic fungi and viruses (Sticher *et al*., 1997). SAR has been shown to be initiated through both PTI and ETI, and the pathways that lead to SAR tend to overlap across different pathogens (Jing *et al*., 2011). SA and pipecolic acid function as activators of SAR, as these molecules form the long distance signalling feedback system between primary infected leaves and secondary uninfected leaves (Návarová *et al*., 2012). Notably, SA

leads to the downstream expression of SAR related genes through the SA receptor: NON-EXPRESSOR OF PR GENES1 (NPR1) (Wu *et al*., 2012, Fu and Dong, 2013). In addition to increasing the systemic resistance of a plant post infection, SAR enables priming of the plant, facilitating rapid responses against future pathogenic invasions (Návarová *et al.*, 2012) .

It has been found that when a plant is grown with growth promoting rhizobacteria (most of these belonging to the Pseudomonas spp.) that the plants are better equipped to handle diseases (Hammerschmidt, 1999). When the possible mechanisms were investigated, it was found that beneficial bacteria were not responsible for producing direct antibiotic effects (Van Peer *et al*., 1991), and this type of resistance was termed ISR. This observation was supported when ISR was induced when lipopolysaccharides and heat-killed bacteria was used in place of the beneficial rhizobacteria (Van Peer and Schippers, 1992, Leeman *et al.*, 1995). This led the authors to theorise that ISR is the product of an enhanced defence response, triggered by the presence of the beneficial bacteria. Currently, there are thought to be two different forms of ISR: one mediated and induced by SA signalling, and another through JA and ET signalling. In SA-dependant ISR, accumulation of SA is observed locally, systematically and in the phloem (Métraux *et al*., 1990, Malamy *et al*., 1990). Whereas in JA and ET dependant ISR, Pieterse *et al*. (1998) demonstrated that ISR was still induced in *A. thaliana* expressing the nahG, which would cause the inability of the plant to accumulate SA. The role JA plays in ISR has been demonstrated in *At*MYC2 mutants in tomato. These plants show defects in rhizobacteria-induced ISR against *Pseudomonas syringae* and *Hyaloperonospora parasitica*. In addition, when Arabidopsis is colonised by rhizobacteria, many of the genes that are induced in the roots contain G-box like sequences in their promoters, suggesting regulation by *At*MYC2 (Ponzo *et al*., 2008).

There has been difficulty in making a list of general patterns involving ISR, largely based on studies using different hosts and pathogens. However, there have been a number of associated traits observed in ISR expressing plants. Firstly, it has been found that ISR expressing plants are primed against pathogens. This manifests itself as PR protein expression in uninfected tissues (Heil and Bostock, 2002). Upon infection, however, these tissues then undergo a greater and more rapid response towards the infecting pathogen, through processes like lignification and phytoalexin synthesis (Heil and Bostock, 2002). Furthermore, it has been shown that enhanced peroxidase activity is associated with ISR against *C. lagenarium* in cucumber (Hammerschmdt *et*

al., 1982). Unfortunately, the mechanisms by which this priming is governed are currently not understood. Additionally, it appears that ISR expressing plants have a greater ability to convert ACC to ET (Pieterse *et al*., 2000). This means that there would naturally be an elevated degree of ET signalling experienced in ISR.

1.8.2 RNA Interference (RNAi)

RNA silencing is an important component of gene regulation, as well as a defence against invading nucleic acids. RNA silencing can occur through either post-transcriptional gene silencing (PTGS) or transcriptional gene silencing (TGS) (Voinnet, 2005, Zhang *et al.*, 2016) (Figure 1-7). Although, RNA interference (RNAi) was not understood at the time, the effects were first described in 1928, in tobacco infected with Tobacco Ring spot virus (Wingard, 1928). In this study, the upper leaves of the tobacco plant showed recovery and immunity to secondary infections (Wingard, 1928). Interestingly, the observations made by Wingard coincide with the phenotype observed in tolerant cassava landrace TME3. Despite being found in other forms of life, RNA silencing in plants, in particular, plays an important role in regulating both developmental and environmental responses (Khraiwesh *et al.*, 2012). Furthermore, small RNA (sRNA) also play a critical role in plant immunity (Vanderschuren *et al.*, 2012, Kuria *et al.*, 2017)

Plants utilize a number of methods in order to precisely regulate gene expression in response to environmental stresses, which includes both transcriptional and post-transcriptional gene regulation (Shukla *et al.*, 2008, Zhou and Luo, 2014, Zhang *et al.*, 2016). A number of genes involved in stress tolerance are regulated by PTGS, where small RNA (sRNA) bind to the mRNA and prevent translation (Sunkar *et al.*, 2012). Genes are also regulated through TGS, where genes are suppressed are through histone modification (Irvine *et al.*, 2006).

RNA silencing is mediated by small RNA (sRNA), which are divided into two major classes based on their biosynthesis and mode of action, namely small interfering RNA (siRNA) and microRNA (miRNA). MicroRNA is transcribed from MIR genes, producing partially double-stranded RNA (dsRNA), by RNA polymerase II to form a precursor miRNA (known as a pre-miRNA) and is then further processed by DROSHA-like proteins to form a pri-miRNA. The resulting pri-miRNA is exported from the nucleus where it is further processed by Dicer-like (DCL) protein to form the miRNA/miRNA* duplex. The miRNA is then released from the

duplex and bound to the RISC complex, where, with perfect homology to its mRNA target, leads to cleavage and degradation (Budak and Akpinar, 2015).

Whereas siRNAs are produced from completely dsRNA, produced by RNA-dependant RNA polymerase, and cleaved through DCL (1-4) proteins produce 21-24nt siRNA. Similarly, to miRNA, this 21-24nt siRNA are incorporated into the RISC complex containing AGO1 or AGO4/6, leading to PTGS or TGS respectively (Khraiwesh *et al*., 2012).

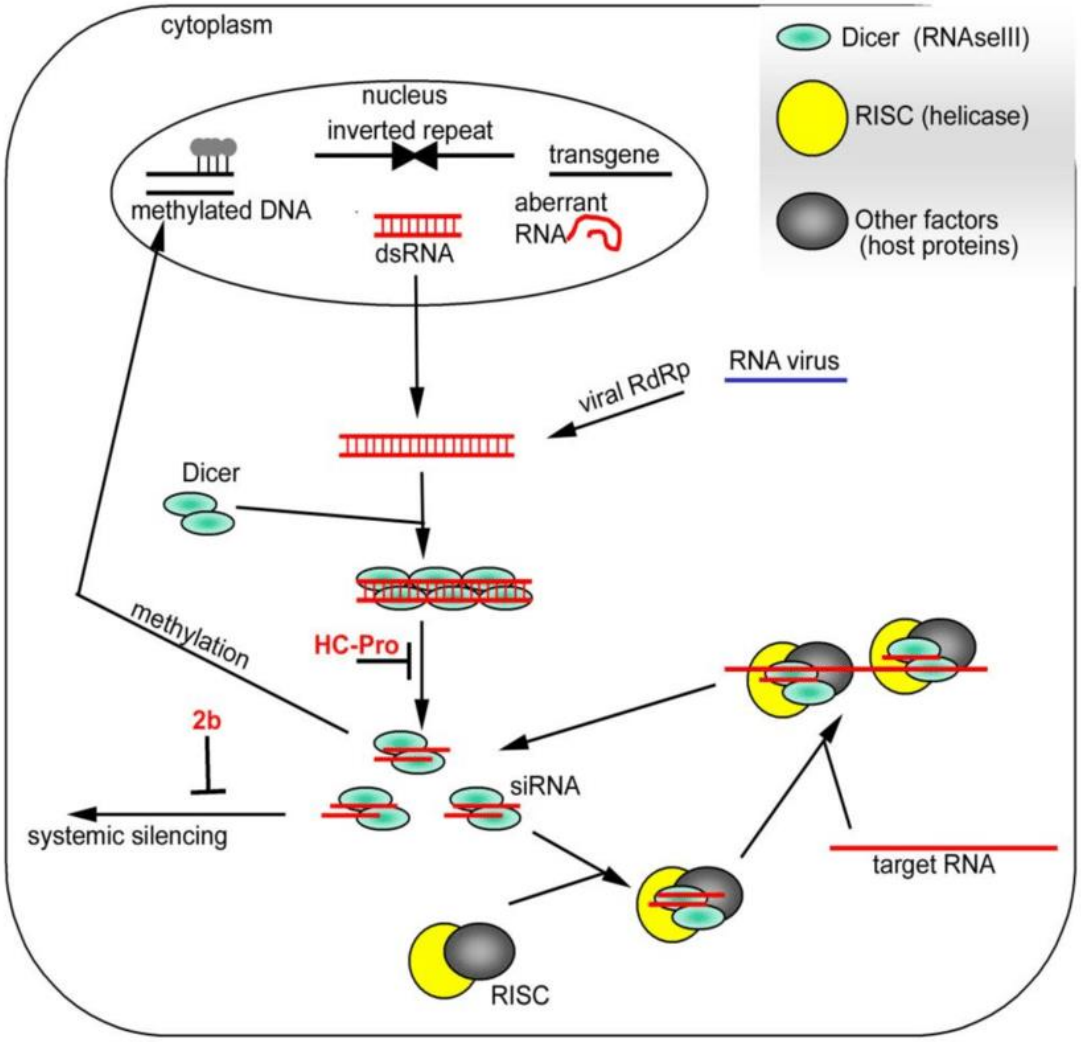


Figure 1-7: Overview of RNAi and the proteins involved. RNAi can either be induced from dsRNA derived from exogenous viral RNA (blue) or dsRNA synthesised in the nucleus. This either leads to PTGS, once the processed RNA is incorporated into the RISC complex, or TGS can occur leading to methylation of targeted DNA. (SURGONDA, 2015)

It is known that one of the means in which a plant responds to and combats pathogenic infection is through RNAi using siRNA and miRNA (Ding and Voinnet, 2007). One of the means by which the plant can do so, is through producing sRNA targeting the auxin receptor (TIR1), thereby reducing the response to auxin, which is traditionally negatively associated with plant defence (Navarro *et al*., 2006). However, in order to improve the likelihood of disease establishment, viruses have evolved means to suppress the RNAi system(Voinnet *et al.*, 1999). One example of this in geminiviruses is the use of the Rep protein to suppress a methyltransferase1, one of the proteins involved with TGS (Rodríguez-Negrete *et al.*, 2013).

To summarize, the plant immune system is a multifaceted and complicated system to study. All the proteins and pathways involved in mounting immune responses do not act in isolation, and perturbation of a single element has a knock-on effect on other parts in the network system. Consequently, in order to study the many aspects regarding immune regulation, a large amount of biological data pertaining to the thousands of genes are required.

1.9 Transcriptomics and systems biology

The scale and precision required to answer questions relating to the tolerant nature of TME3 can be found in RNA sequencing (RNA-Seq) technologies. RNA-seq entails the conversion of an RNA library to cDNA through the use of non-random primers and PCR. The cDNA library is then prepared for sequencing through the ligation of adapters. The adapters and technology used to sequence the cDNA are dependent on the sequencing method selected. The most prevalent technology currently in use is Illumina™ sequencing, which uses fluorescent dyes to detect nucleotides and determine the sequence. It should be noted that this sequencing technology constitutes “high-throughput sequencing”, which entails the highly paralleled sequencing of millions of DNA molecules simultaneously. Such an experiment requires a huge undertaking, requiring large amounts of funding, time and expertise. Thankfully, transcriptomic data that would allow us to study the immune response of TME3 and T200, in order to understand the similarities and differences seen between these two landraces is currently available.

1.9.1 Cassava transcriptomic data

In 2014, a study was performed in order to quantify the global transcriptomic responses in the tolerant TME3 and susceptible T200 cassava landraces, in response to SACMV, over the course of disease progression (Allie *et al*., 2014). Leading up to this, both TME3 and T200 were agro-inoculated, where expanding leaf material below the apex was picked at 12, 32 and 67 days' post inoculation (dpi), representing the early, middle and late stages of infection respectively. RNA extraction, cDNA synthesis and high-throughput SoLID sequencing were performed. Genes deemed to be significantly differentially expressed between the two infected landraces were then identified. These findings were validated using real-time reverse transcriptase quantitative PCR (RT-qPCR) of selected genes. From this research, the authors went about commenting broadly on many aspects regarding the transcriptomic trends in T200 and TME3 during SACMV infection. In doing so, the authors were able to provide evidence supporting the paradigm that viruses are able to manipulate molecular interactions so that they can adapt to the host background (Allie *et al*., 2014). Despite the authors noting the expression of several TFs and hormones, due to the large multi-dimensional nature of transcriptomic data, a comprehensive analysis and discussion surrounding all the TFs of interests was lacking in this study. With this realisation, comes the opportunity to expand upon existing knowledge through reanalysing the data with greater scrutiny and in greater detail, using a systems biology approach.

1.9.2 Systems biology

Systems biology entails the creation of biological networks using nodes, which represent genes, proteins or intermediates, and edges, which describe the type of interaction that occurs between two nodes (Figure 1-8). Systems biology can function as a means to organise high dimensional and complicated data, such as transcriptomic data so that they can better visualised and understood. There are multiple means to transition isolated genes into fully fledged networks. The most common are to derive known protein-protein interactions occurring between your DEGs of interest. Alternatively, gene co-expression networks can be deduced, through the expression profiles of genes. Doing so enables the inference of gene-gene interactions. Additionally, determining the enrichment among biological function gene ontology (GO) terms may be undertaken. This method will determine the processes in which the DEGs are contributing most towards. Indeed, only when DEGs are viewed within biological networks can

a better understanding be derived surrounding their function. For example, it was only through the study of these network interactions that it was found that viruses optimise their strategy of reprogramming host systems by targeting hub proteins in signal transductions pathways (Mukhtar *et al.*, 2011).

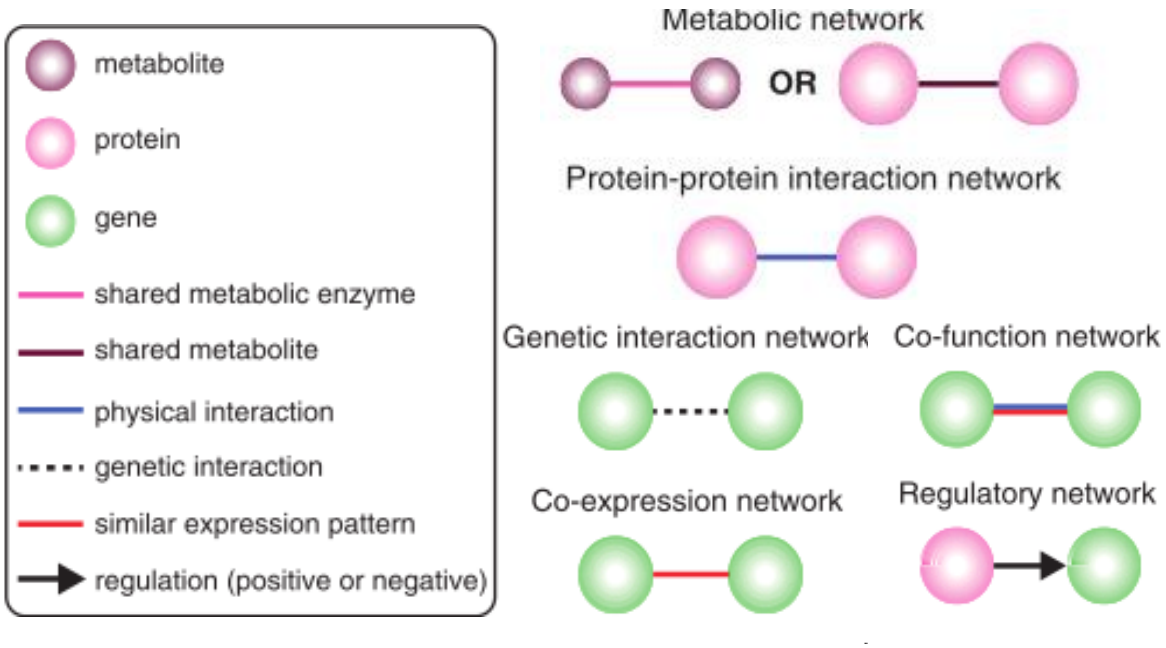


Figure 1-8: Schematic summary of the different types of nodes (circles) and edges (lines) which make up various biological networks. Adapted from Chae *et al.*, 2012.

Just as one would find conserved motifs within proteins that perform a specific function, network motifs also exist. A hub is a term used to define a system motif that is characterised as being a highly interconnected node. When targeted, these hubs produce a larger degree of transcriptional changes, as would occur when viruses target these hubs. It is to be rationalised that the resulting hubs found within the underlying networks in T200 and TME3 are to contribute to either tolerance or susceptibility. Therefore, hub discovery and functional elucidation are potential means to discover candidate genes involved in plant immunity. Effective and meaningful network construction and placing genes into these networks requires a large wealth of information, ranging from expression data, functional annotation, annotated gene names, and characterisation studies. Without some of these, network construction is still possible, however, the information that one may gleam from it would be limited and unable to answer questions

regarding tolerance in cassava. Currently cassava, much like many non-model organisms, suffers from lacking much of the information required to form these networks. In order to combat, there are alternative strategies one may use to study non-model organisms.

1.10 Working with non-model organisms

It is to be noted that cassava is not considered a model organism. Consequently, the wealth of knowledge around cassava is not as extensive as other organisms, like *Arabidopsis thaliana*. This creates several problems when it comes to studying the genome of these non-model organisms. The first problem is the limited knowledge base to work from, as the many of the genes within the cassava genome are either unannotated or have been assigned a vague function. This limitation affects not only curated databases, but also available literature. Secondly, there are regular major revisions that occur within these genomic databases. Assembling a genome de novo is a computationally intense and difficult to perform, resulting in a number of scaffolds not being able to be placed within the context of the chromosomes. When the genome position of these scaffolds is revised and updated, there can be considerable changes in the positions of genes, the number of genes (some genes are merged together while others are known to be split) and the gene accession nomenclature.

A common strategy around this limitation is to make assumptions regarding homology and the guilt-by-association principle. Homology follows the paradigm that sequences govern structure and function, therefore two sequences that share high similarity are assumed to be homologs that share similar functions. Guilt-by-association refers to the notion that genes that share related functions tend to form similar protein interaction partners or expression profiles (Oliver, 2000). These two strategies enable the identification of genes that are likely to share functions in another organism, *Arabidopsis thaliana* offers the best candidate to infer information regarding cassava from, due to its model crop status, meaning that the information surrounding its genome annotation is the most complete compared to other plants.

1.11 Rationale for Study

Cassava is becoming an increasingly important crop in a world faced with a rapidly growing population, rampant climate change and poverty. CMD is the single biggest threat facing cassava production within the African continent and requires addressing in order to ensure cassava

production is able to satisfy demand. This interest has reached South Africa, and there are aims to build a significant bio-economy to contribute significantly to job and wealth creation by 2030. To ensure South Africa reaches this goal in the expected time frame, appropriate safeguards are required to be set in place. An important safe guard would be to ensure high production of cassava.

Currently, the two best proactive strategies available are resistance or tolerance to CMD via breeding or RNA silencing-mediated genetic engineering (Rey and Vanderschuren, 2017). Some would argue that engineering and breeding resistance, as opposed to tolerance, would be more beneficial. However, there are advantages to apply tolerance rather than resistance. Resistance would prevent infection of the engineered crop but does not alter the viral reservoir in the surrounding crops. This would lead to a high degree of selective pressure being exerted on the virus to overcome the resistance barriers in the resistant crop. This would mean that the virus would eventually be able to infect the plant again requiring additional research into devising alternative strategies. On the other hand, since tolerance still enables lower levels of infection by the virus, the resulting selective pressure on the virus would not be as great as seen in naturally or engineered resistant plants, offering a more cost-effective long-term anti-viral strategy. Currently, the precise molecular mechanisms that govern tolerance are not well understood. It is through understanding this natural resistance in cassava that could allow the development of cisgenic plants though gene editing technologies such as CRISPR-Cas9 (Fondong and Rey, 2018).

Since tolerance manifests as non-symptomatic or mild symptoms in recovered emerging leaves, it can be concluded that tolerance is a systemic rather than localized response. This suggests that tolerance uses systemic signals. However, before a systemic response can occur, a localized response by infected cells in primary leaves is first required in order to initiate a defence response. The molecular mechanisms of tolerance are hypothesized to include basal immunity in addition to other mechanisms which may involve resistance nucleotide binding – leucine-rich repeat (NB-LRR) proteins (Bengyella *et al*., 2015). Other signalling mechanisms for plant virus tolerance/resistance, such as transcription factors (TFs), may also be involved in tolerance to CMD in cassava.

1.11.1 Aim

The aim of this research was to identify TFs, and place them into predicted networks, so as to infer their function in the development of tolerance and susceptibility in cassava. Amongst TF families, the WRKY family has been widely reported to be involved in pathogen stress responses, so consequently, focus was first placed on the WRKY family of TFs to (a) identify differentially expressed(DE) TFs, (b) compare and contrast their expression between landraces, and stages of SACMV infection, and (c) construct protein-protein networks. Following this, other TF families thought to play a role in plant immunity, namely the bZIP, ERF, MYB and NAC families, were investigated. It is suggested that through??, elucidating some of the TF-associated host responses, some of the mechanisms that govern tolerance and susceptibility in SACMV-infected cassava TME3 and T200, respectively, may be further elucidated. These host responses can be exploited as candidates for future cassava improvement biotechnology strategies. Furthermore, while many pathogen response studies have been conducted on bacteria and fungi, there are few reports on viruses, particularly geminiviruses. These results will also contribute to knowledge enhancement and provide recommendations to spearhead genetic engineering strategies in the future.

1.11.2 Objectives

In order to address the aim, the following objectives are proposed:

- To identify all the DE TFs belonging to the five families of interest (bZIP, ERF, MYB, NAC and WRKY) in transcriptomic data from susceptible T200 and tolerant TME3 landraces infected with SACMV at 12, 32 and 67 dpi.
- To identify and compare predicted protein-protein interactions occurring between the identified DE TFs and other DEGs in both T200 and TME3 infected with SACMV at 12, 32 and 67 dpi.
- To identify the biological processes associated with the up- and down-regulated DEGs in both T200 and TME3, at each time point.
- To identify hubs within protein-protein interaction networks that are associated with either susceptibility (T200) or tolerance (TME3).

Chapter 2

*Me*WRKY transcription factors and phytohormone networks play pivotal roles in regulation of tolerance and susceptibility to cassava mosaic disease

2.1 Background

Cassava, *Manihot esculenta* Crantz, is a perennial food security crop largely grown in developing countries, with worldwide cassava productions measuring over 277 million tons in 2016 (http://www.fao.org). Cassava's food security crop status is attributed to its natural drought tolerance, which makes it a desirable crop for subsistence farmers (El-Sharkawy, 2004). However, cassava in sub-Saharan Africa and Asia is currently under threat from several pests and diseases, one of the most important being CMD (Legg *et al*., 2015, Rey and Vanderschuren, 2017). CMD presents as a spectrum of symptoms such as leaf curl, stunted growth, mosaic and chlorosis (Bock and Woods, 1983). This subsequently leads to a reduction in cassava crop production, which can be as high as a 100% loss in yield (Hillocks, 2003). In recent years, CMD was classified as a severe pandemic after 11 countries were affected in East and central Africa (Legg *et al*., 2014). Several species and genetic variants of begomoviruses are responsible for CMD, and these are classified under the family *Geminiviridae*, genus *Begomovirus* (Brown *et al*., 2015). Cassava mosaic geminiviruses are bipartite, composed of two circular ssDNA components (DNA A and B) that are packaged separately into twinned geminate virion particles (Stanley 1990). Each DNA component ranges between 2.5-2.9 kb in length, and collectively codes for all the proteins needed for replication, cellular movement, vector transmission and the establishment of infection in the host (Gutierrez, 1999). *South African cassava mosaic virus* is one of the species causing CMD (Berrie *et al.*, 2001), and occurs in South Africa, Madagascar and Mozambique (Rey *et al.*, 2012) SACMV is transmitted by the whitefly vector, (*Bemisia tabaci* (Genn.) (Legg *et al*., 2014). If CMD remain unchecked, high losses are expected in cassava production in southern Africa. This problem will be exacerbated in future by large-scale climate change and increasing population numbers on the sub-Saharan continent and may coerce farmers to become more dependent on security crops like cassava.

Numerous breeding programs for CMD resistance, including gene introgression from wild *Manihot* relatives such as *M. glaziovii*, have been undertaken over the years (Jennings and Iglesias, 2002, Ntawuruhunga *et al*., 2006, Dixon *et al*., 2008, Nweke, 2009, Dixon *et al*., 2010). These efforts led to variable polygenic recessive resistance (designated CMD1) to the disease, yielding cassava with moderate resistance to CMD (TMS series) (Hahn *et al.*, 1980). A major locus (CMD2) obtained from local Nigerian landraces putatively confers monogenic dominant resistance (Akano *et al.*, 2002). A genome-wide association mapping study on 6128 cassava breeding lines revealed a single CMD2 locus on chromosome 8 (now relocated to chromosome 12) (Phytozome V12 *Manihot esculenta* V6.1), and two additional epistatic loci/ or multiple resistance alleles, associated with CMD resistance (Wolfe *et al*., 2016). Two peroxidases (pathogenesis-related proteins) (van Loon *et al.*, 2006) and thioredoxin (protein disulfide isomerase-like 2-2 homolog) were suggested to be strong candidate genes associated with resistance. While TME3 is regarded to be resistant due to the presence of the CMD2 locus, this West African landrace shows an infection phenotype more akin to tolerance (Allie *et al.*, 2014, Bengyella *et al.*, 2015, Paudel and Sanfaçon, 2018). Tolerance in TME3 is characterized by having a lower detectable viral titer compared with susceptible T200 (Allie *et al.*, 2014). An additional hallmark is symptom recovery at later stages of infection, approximately 67 dpi, and milder or no symptoms in recovered leaves which persists in the long term (Figure 2-1). In the susceptible T200 landrace, however, symptoms continue to be chronic for the lifetime of the host, and a high virus load is present in leaves (Bengyella *et al.*, 2015). While the molecular mechanisms of tolerance are not well-understood, a study (Allie *et al.*, 2014)has revealed significant differences in the transcriptomic response to SACMV in TME3 compared with T200. Recovery of virus symptoms has also been associated with antiviral RNA silencing mechanisms (post-transcriptional and transcriptional gene silencing, Rogans and Rey, 2016).

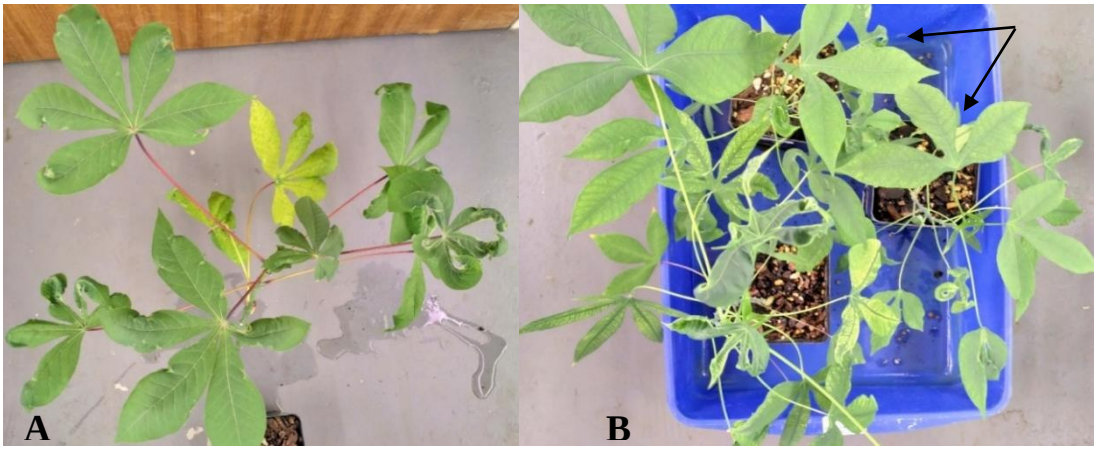


Figure 2-1:Overhead view of cassava landraces infected with SACMV at 67 dpi. A) T200 maintains severe chlorotic and leaf curl symptoms at 32 and 67 dpi. B) TME3 demonstrates tolerance which is apparent from the lower symptom severity in the apical leaves (arrows).

The plant immune system comprises of two interconnected mechanisms, namely pathogen triggered immunity (PTI) and effector-triggered immunity (ETI) (Jones and Dangl, 2006). PTI acts primarily in basal defence of the plants against infection, using pathogen recognition receptors (PRR) on the cells membrane surface to detect and respond to conserved pathogen molecules (Zipfel and Felix, 2005). ETI functions predominately within the cell through nucleotide binding-leucine rich repeat (NB-LRR) resistance proteins which recognize viral effectors, initiating downstream signalling and defense responses, such as PR proteins (Ting *et al*., 2008, Cheng *et al*., 2009) amongst others (Matzinger, 2002). SAR is where the whole plant becomes immunized following salicylic acid (SA)-associated signalling which is induced by a prior localized pathogen or microorganism (Choudhary *et al.*, 2007). Most pathogens trigger a common complex interconnected plant signalling network, suggesting that differences between SAR, PTI and ETI in different host-virus interactions are both quantitative and temporal (Eulgem and Somssich, 2007), and due to fine-tuning and cross-talk between regulatory networks.

Phytohormones are used by plants respond to a diverse range of stimuli, including biotic stress from pathogens. In particular, the most central hormones associated with biotic stress are JA and SA which exist in an antagonistic relationship (Glazebrook, 2005). Traditionally, JA and ethylene (ET) modulate necrotrophic pathogens, while biotrophic and hemibiotrophic pathogens respond to SA (Glazebrook, 2005). Furthermore, abscisic acid (ABA) is shown to be involved in

the mediation of JA-SA interactions. For example, AtMYC2 functions both as a positive regulator for ABA signalling and the central regulatory hub in JA signalling (Abe *et al*., 2003, Anderson *et al*., 2004, Kazan and Manners, 2013, Kazan and Manners, 2013). JA and SA are considered the most important hormones as all other phytohormones modulate plant immunity through interacting either with JA or SA (Pieterse *et al.*, 2009). In particular, SAR is closely associated with SA signalling (Cao *et al.*, 1994, Fu and Dong, 2013). Hormones function both upstream and downstream of transcription factors TFs. Hormones may alter genes expression by transducing the signal downstream through various kinases and phosphatases that induce transcription factors TFs to initiate transcriptomic re-programing and activate stress responses (Hernandez-Garcia and Finer 2014, Taj *et al*., 2010). Conversely, the TFs may alter downstream hormone level through the expression of target genes that either inhibit or induce hormone biosynthesis (Zhou *et al.*, 2011, Schweizer *et al.*, 2013). Consequently, these networks of cross-talk between hormones and TFs enable the plant to respond adequately and flexibly to stress. These networks form an integral part of the plant and form the basis of regulation of plant functions. Regulatory networks comprise two components: namely nodes (representing genes and their products) and edges. Three types of interactions between genes/proteins occur, namely simple direct, indirect or triplets. These regulatory networks have considerable importance when considering virus-host interactions. This is because viruses optimise their strategy of reprogramming the host system by targeting hub proteins in signal transductions pathways (Albert *et al*., 2000). These hub proteins are a network motif describing a node that is disproportionally interconnected compared to neighbouring nodes, and through targeting hubs, viruses increase the number of induced perturbations in the plant hosts (Kitano, 2003). Conversely, it can also be rationalised that many defence genes are hubs as they are required to induce a large degree of transcriptional reprogramming and in doing so the response to the pathogen. All TFs are examples of hubs that are able to undergo large scale transcriptional changes.

Responses to plant pathogens lead to large-scale transcriptional reprogramming, mediated by TFs principally belonging to the ERF, NAC MYB, bZIP and WRKY families. Due to the central role of TFs in regulatory networks, there is potential around exploiting these families for improving biotic and abiotic stress response in plants (Erpen *et al.*, 2018). Of these TFs, in particular, the WRKY family play a pivotal role in regulating plant defences (Eulgem *et al.*,

2000, Eulgem and Somssich, 2007, Pandey and Somssich, 2009). The WRKY family is a large family of TFs found in higher plants (Eulgem *et al.*, 2000), and are implicated in several physiological processes such as seed development, dormancy and senescence (Rushton *et al*., 2010). WRKYs were previously regarded as plant-specific, but their identification in lower organisms such as a slime mould and protist, imply an earlier origin (Ulker and Somssich, 2004). Despite this, most recently, the link between animals and plant WRKYs has been found. Interestingly, NF-κB in mice, a TF with functions in basal immunity, was found to bind the analogous WT-boxes similar to WRKYs studied in Arabidopsis (Kanofsky *et al.*, 2018). WRKY proteins contain the conserved DNA-binding domain, comprising the WRKYGQK peptide and zinc-finger-like motifs, that binds to the W-box (TTGAC(C/T)) cis-regulatory element of target genes (Pandey and Somssich, 2009) Rushton *et al*., 2010). Three broad subgroups exist in the WRKY TF family. Group I is characterised by having two WRKY domains and two C_2H_2 zinc finger domains. Group II and III both contain a single WRKY domain but contain different zinc finger domains (C_2H_2 and C_2HC respectively) (Eulgem *et al*., 2000). Group II is further divided into II a, b,c, d and e based on their amino acid sequence or phylogenetic data (Zhang and Wang, 2005). WRKYs were first shown to respond to pathogen infection in parsley, functioning as both positive and negative regulators of pathogen response in plants via PTI and ETI (Rushton *et al*., 1996). Since then, the number of identified WRKYs has increased in multiple plant species, with the fewest (45) being in barley (*Hordeum vulgare*), while soybean (*Glycine max*) has the highest number (182) characterized (Mangelsen *et al*., 2008, Bencke-Malato *et al*., 2014). More recently, 85 WRKYs were characterized in cassava, 11 more than those found in *Arabidopsis thaliana* (Ulker and Somssich, 2004, Wei *et al*., 2014). It has been shown that WRKYs are involved in innate immunity and defence gene activation in response to biotic stress (Bakshi and Oelmüller, 2014). A recent study reported that *Me*WRKY75 positively regulates disease resistance against cassava bacterial blight (*Xanthomonas axonopodis* pv *manihotis*) (Liu *et al*., 2018). While several studies have been performed on bacterial and fungal pathogens (Zheng *et al.*, 2006, Pandey and Somssich, 2009, Hu *et al.*, 2012, Liu *et al.*, 2012, Bencke-Malato *et al.*, 2014), very few virus-responsive WRKYs have been reported. WRKY8 has been shown to play a role in defence to crucifer-infecting tobacco mosaic virus (TMV-cg) in Arabidopsis by mediating both abscisic acid and ethylene signalling (Chen *et al*., 2013). Modulating the function of transcription factors through interactions with regulatory proteins is a critical process in the repression or

activation of signal transduction pathways (Van Verk *et al*., 2009). WRKY regulatory proteins form a complex interconnected regulatory network with themselves, and other TFs and interacting regulatory partners, such as Mitogen-Activated Protein (MAP) kinases and the key SA regulator, *NPR1,* amongst others (reviewed in Chi *et al*., 2013, Alves *et al*., 2014). Furthermore, it has been seen that numerous members of the WRKY family interact with various hormones required in signalling defence responses (Rushton *et al*., 2010, Duan *et al*., 2007, Pandey *et al*., 2010). More recently, sRNA (small interfering and microRNAs) have also emerged as key regulators of WRKY networks via PTGS(Zhang *et al.*, 2008, Berri *et al.*, 2009).

A previous study was performed quantifying the global transcriptomic response in CMD-tolerant TME3 and susceptible T200 cassava landrace. SACMV-induced disease progression was monitored over the early (12 dpi), middle (32 dpi) and late (67 dpi) infection stages (Allie *et al*., 2014). No symptoms were observed at the early stage of infection in T200 or TME3, but SACMV could be detected by RT- in both landraces. At 32 dpi, CMD symptoms were observed in both infected landraces, but symptoms and viral load were significantly elevated in T200. Recovery leading to no or milder symptoms in emerging apical leaves (Figure 2-1A) hallmarked TME3, while in T200 severe symptoms persisted (Figure 2-1B). Persistent lower virus titers in TME3 compared with T200 were recorded. Significantly, the authors noted the differential expression of the homologs belonging to *At*WRKY33 and 70, which are involved in the suppression of SA signalling and immunity in susceptible T200 (Allie *et al*., 2014), however, the potential roles of the other *Me*WRKY networks in susceptibility or tolerance were not investigated further. In order to elucidate putative *Me*WRKY candidates and network interacting partners involved in susceptibility and tolerance in SACMV infected cassava, a thorough functional bioinformatics interrogation of the RT-qPCR validated transcriptome data was conducted. Overall, results demonstrate, for the first time, significant differences in hormone and *Me*WRKY regulatory networks between a CMD-susceptible and tolerant cassava landrace. These findings could contribute to rethinking novel approaches to cassava mosaic disease control strategies.

2.2 Methods

2.2.1 Identification of differentially expressed genes

The lists of differentially expressed genes (DEGs) were taken from the RT-qPCR validated supplementary data in the Allie *et al*. study (Allie *et al*., 2014). In this study, leaf material was

collected in triplicate from both infected and mock-inoculated T200 and TME3 cassava landrace, agro-inoculated with a SACMV construct. Total RNA was extracted from the samples using a modified high molecular weight polyethylene glycol protocol before pooling the biological triplicates (Allie *et al*., 2014). cDNA libraries were prepared according to the SOLiD Total RNA-Seq kit. The cDNA libraries were sent off for paired-end sequenced using the SOLiD v4 sequencing platform on a single flow-cell. The libraries from the infected material were normalized against the mock-inoculated libraries. Alignment to the cassava reference genome (http://www.phytozome. net/cassava.php, *Manihot esculenta* 147, v4.1) using Lifescope (Lifetech). Transformation on alignments was performed using the rnaSeqMap R package (v2.7.12) (Leśniewska and Okoniewski, 2011). Differential expression was performed using the DEseq (v1.4.1) where intergenic spaces were included for any novel transcripts in determining the transcript counts (Anders and Huber, 2010). Differential expression criteria for the genes required a p-value < 0.05 and a $\log_2$-fold > |1.5|. Although $\log_2$-fold > |2| is traditionally chosen, a lower threshold was chosen as it is known that relevant genes do not necessarily undergo such large $\log_2$-fold changes to be biologically significant in producing the cognate phenotype. For this reason, a lower threshold of $\log_2$-fold > |1.5| was chosen as a cut-off for DE genes for this study.

It is to be noted that cassava is not considered a model organism, consequently, the wealth of knowledge around cassava is not as extensive as other organisms, like *Arabidopsis thaliana*. This creates several problems when it comes to studying the genome of the non-model organisms based on the amount the type of data available to work with. The first problem is the limited knowledge base to work from as the many of the genes within the cassava genes are either unannotated or are given a vague function. This limitation extends not only to curated databases but into literature too. Secondly, there are major revisions that occur within the genomic databases. Assembling a genome *de novo* is a computationally intense and difficult methodology to perform, resulting in many scaffolds not being able to be placed within the context of the chromosomes. When the genome scaffold is revised and updated, there can be considerable changes in positions of genes, the number of genes (some genes are merged together while others are known to be split) and the gene accession nomenclature. It should be noted that the cassava genome on phytozome (https://phytozome.jgi.doe.gov/pz/portal.html) was

updated from version 4.1 to 6.1 in 2016, meaning the gene annotations of the data produced by Allie *et al*. is out of date.

A solution for both the update in the cassava reference genome and the lack of comprehensive annotation is the process of cross-mapping. This is the process where an alternative accession ID is assigned either representing the same gene (cassava v4.1 to cassava v6.1) or, alternatively, the gene homolog (cassava v6.1). Cross-mapping was performed by downloading the annotations from phytozome, using the biomart tool, of all genes from version 6.1 with their version 4.1 synonym (Smedley *et al.,* 2009). Biomart in Phytozome (Genome database v.10 and v.11), was subsequently used to cross-map to Arabidopsis, thereby deriving the associated homologs. Identification of differentially expressed WRKYs.

Using the Arabidopsis homolog IDs, the TFs and the WRKYs could now be determined. The plant Ensembl database (release 42) was used to annotate the genes with information regarding their GO terms, and their PFAM IDs (v 31.0) through biomart (Plant genes v40). GO terms refer to the various features of a gene pertaining to its location in a cell, the molecular function it performs and its biological function, whereas PFAM IDs are assigned to specific conserved domains found within a gene. TFs were identified from the list by filtering all genes that had the GO term "transcription factor activity, sequence-specific binding". From this list of TFs, all WRKYs were identified through further filtering this list for genes which contain the PF03106 PFAM ID, which denotes the conserved WRKY domain specific to the WRKY TF family.

2.2.2 GO enrichment

GO enrichment was performed using the clusterprofiler package (v2.7) in R (v3.3.1) (Yu *et al*., 2012). First, genes were separated based on their expression into two separate lists: up-regulated genes and down-regulated genes for each landrace per time point. This was done in order to determine with greater specificity what processes are being up- or down-regulated in response to SACMV. Each of the gene lists was separately compared to the background list of genes which comprised of all the genes in *A. thaliana* acquired from the plant Ensembl database (TAIR 10). GO terms were deemed enriched if they had a p-value < 0.05 and a q-value < 0.01 when the Benjamini-Hochberg correction was applied. In order to determine the number of genes contributing to hormone signalling in each host and per time point, the list of enriched GO terms was first scanned for all instances of the hormones interests, namely ABA, ET, JA and SA. The

hormones responsive genes that were found contributing to the GO term from the GO enrichment results were tallied, excluding any duplicates. This was done separately for each of the gene lists that underwent GO enrichment analysis.

2.2.3 Identification of interacting partners

The cross-mapped IDs for the identified WRKY TFs, in each landrace per time point, were separately queried against the STRING (v10.1) database for *A. thaliana* interactions (https://string-db.org/). The confidence threshold was set to high (0.6, Supplementary Methods), to only include interactions that are extensively valid. This is necessary as these are inferred interaction in cassava from an Arabidopsis dataset, which entails only interactions with high confidence are likely to be homologous between the two plant species. The second shell of interactions was set to a maximum of 50, which means that, at most, each of the proteins that make up the first shell (WRKYs used in the query) will be connected to 50 proteins. Using the information from STRING and the GO enrichment, networks were formed around the identified WRKY TFs. The resulting list of interactions and genes from STRING were filtered to only include DEGs that were present in the host and time point with the identified WRKYs that formed the search query. Furthermore, the enriched GO terms were filtered to only include those that contained either the DEGs or the WRKYs, Lastly, the genes and GO processes were annotated to show if it was either up, down or up- and down-regulated at the given time point in each host. GO processes were deemed either up- or down-regulated based on if the term was found to be enriched in the lists containing only the up- or down-regulated genes. The network was visualized and annotated within Cytoscape (v3.5.1).

2.3 Results

2.3.1 Comparison between differentially expressed (DE) *Me*WRKYs in cassava TME3 and T200 at 12, 32 and 67 dpi

In total, differential expression (up- or downregulated) of forty-three *Me*WRKY TFs was observed in both SACMV infected susceptible T200 and tolerant TME3 landraces (Table 1). A general pattern emerges where both landraces express the least number of DE MeWRKYs at 12 dpi, with a significant increase in numbers at 32 dpi, and then declining at 67 dpi. However, in contrast, T200 showed a considerably higher number of DE *Me*WRKYs compared with TME3, as the majority (76.74%) of the identified DE *Me*WRKYs were in T200. Furthermore, the overall expressions (72.27%) of T200 *Me*WRKYs are down-regulated in response to SACMV, whereas in TME3, the majority (80%) of *Me*WRKYs are upregulated over the SACMV infection period (67 days). Of the 43 DE TFs, only 23 different *Me*WRKY transcripts were detected in T200 and TME3 at all three time-points (Figure 2-2). Notably, TME3 expressed four unique TFs (MeWRKY17, 40, 58 and 81) not found in T200, whereas four WRKYs (*Me*WRKY3,10, 11 and 79) were the only TFs to be DE in both landraces (Figure 2-2B). Notably, $\log_2$-fold change in expression of the four common *Me*WRKYs in T200 and TME3 at 12, 32 and 67 dpi show opposite (down- versus up-regulation) expression (Figure 2-2B). Interestingly, of the four *Me*WRKYs DE in both landraces (*Me*WRKY3,10,11 and 79), all except for *Me*WRKY3 are annotated as a homolog of *At*WRKY40. In addition to *At*WRKY40, both *At*WRKY53 and 70 appear to have multiple DE *Me*WRKYs homologs, particularly in T200 (Table 2-1)

Table 2-1:The Log2Fold change in expression identified *Me*WRKY transcription factors in SACMV infected T200 and TME3 landraces at 12, 32 and 67 dpi and the function of their Arabidopsis homologs.

Landrace	Gene name	Homolog	12 dpi	32 dpi	67 dpi	Homolog Function	Ref
TME3	*Me*WRKY17	*At*WRKY06		2.2849		Induced by SA, JA and ET. Phosphate tolerance	Scarpeci *et al*., 2008, Pieterse and Loom, 1999 Nilsson *et al*. 2010
	*Me*WRKY58	*At*WRKY07			-1.8375	Negative regulator of defence signalling	Journot-Catalino *et al*., 2006
	*Me*WRKY81	*At*WRKY28			-3.5317	Positive regulator of JA signalling	Hu *et al*., 2013
	*Me*WRKY10	*At*WRKY40		2.0828	2.0828	Positive regulator of ETI Transcriptional repressor of ABA genes Modulator of JA signalling	Schon *et al*., 2012 Pandey *et al*., 2010 Liu *et al*., 2012
	*Me*WRKY11			2.2978			
	*Me*WRKY40		4.2497				
	*Me*WRKY79		2.4377	2.4377			
	*Me*WRKY3	*At*WRKY41		1.8995		Confers positive resistance towards *P.syringae*	Hisagi *et al*., 2008
T200	*Me*WRKY44	*At*WRKY07	2.302	2.302	2.302	Negative regulator of defence signalling	Journot-Catalino *et al*., 2006
	*Me*WRKY70	*At*WRKY12		2.2103		Positive regulator of defence against *P.carotovorum*	Kim *et al*., 2014
	*Me*WRKY57	*At*WRKY14		-4.019		Regulate flower development	Li *et al*., 2016
	*Me*WRKY32	*At*WRKY20	2.3434			Response to hydrogen peroxide	Chen *et al*., 2010
	*Me*WRKY30	*At*WRKY22			2.2109	Regulates leaf senescence	Zhou *et al*., 2011
	*Me*WRKY68	*At*WRKY33		-2.6021	-2.6021	Positive regulator of resistance against *B.cinerea*	Lai *et al*., 2011
	*Me*WRKY54	*At*WRKY35			2.7873	Unknown function	-
	*Me*WRKY10	*At*WRKY40		-2.2446		Positive regulator of ETI Transcriptional repressor of ABA genes Modulator of JA signalling	Pandey *et al*., 2010 Liu *et al*., 2012
	*Me*WRKY11			-2.4773	-2.4773		
	*Me*WRKY79			-2.6405			
	*Me*WRKY3	*At*WRKY41	-2.5359	-2.5359	-2.5359	Confers positive resistance towards *P.syringae*	Hisagi *et al*., 2008

*Me*WRKY50	*At*WRKY44		1.9265	1.9265	Root hair development and seed coat maturation	Rushton *et al* 2010
*Me*WRKY42	*At*WRKY51		-3.8314		Mediates SA and represses JA signalling	Gao
*Me*WRKY15	*At*WRKY53		-2.5636	-2.5636	Basal resistance against *P. syringae*	Hu *et al*., 2012
*Me*WRKY55		-1.8897	-1.8897	-1.8897	Positive regulator of senescence	Miao *et al*., 2007
*Me*WRKY60	*At*WRKY57		-2.6006	-2.6006	ABA induced draught tolerance regulator	Jiang *et al*., 2012
*Me*WRKY59		-1.844	-1.844	-1.844	Regulates JA and SA antagonism	
*Me*WRKY83	*At*WRKY70		-1.5359	-1.5359	Modulates secretion of signalling hormones	Li *et al*., 2004 Li *et al*., 2006
*Me*WRKY61	*At*WRKY72		-3.2255		Basal defence against root-knot nematodes	Bhattarai *et al*., 2010

.

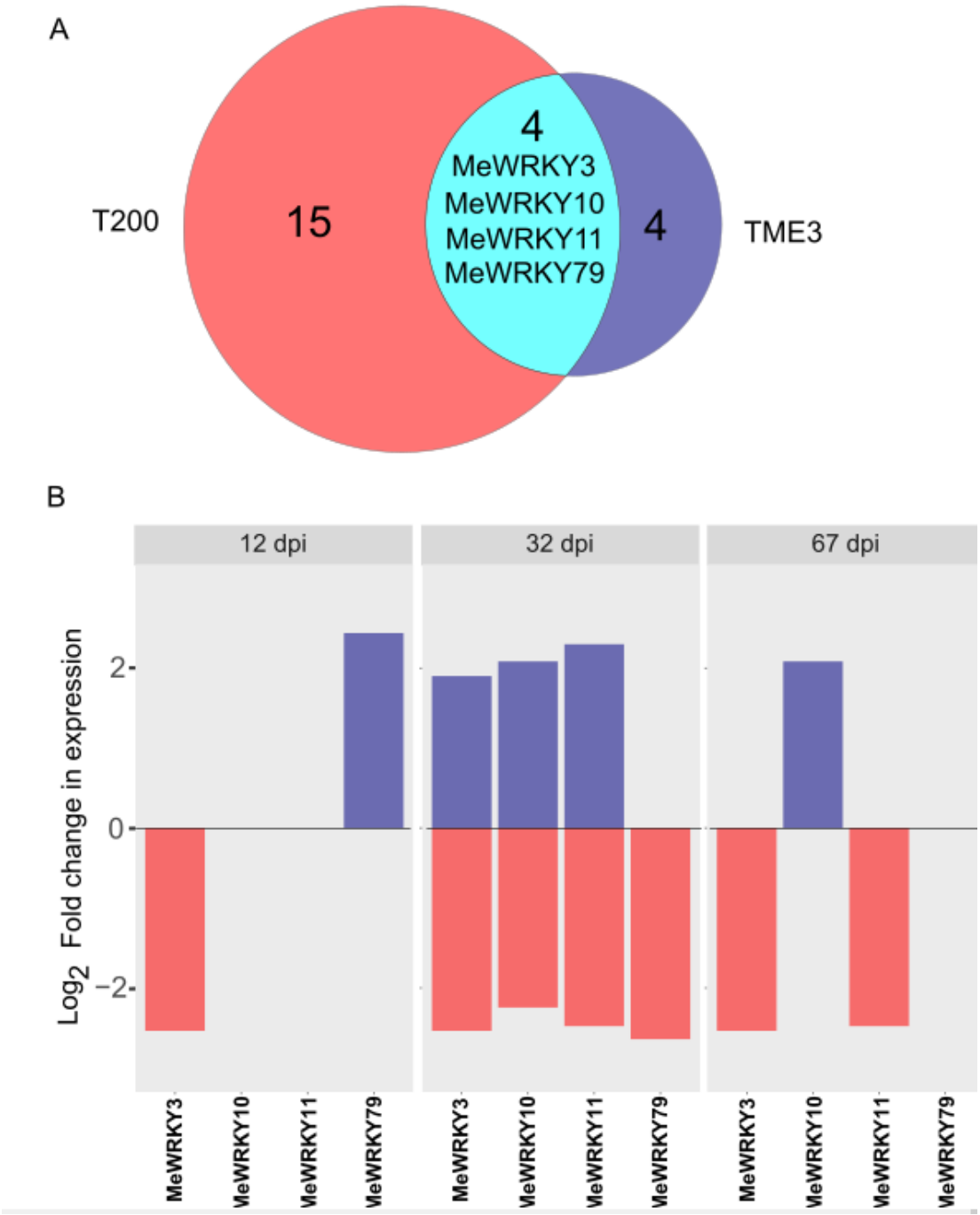


Figure 2-2: Differential expression of *Me*WRKY transcription factors common to T200 and TME3. A. The total number of identified DE MeWRKYs in T200 (15), TME3 (4), and those common (4) to both landraces, landrace throughout disease progression. **B**. Log_2-fold change in expression of the four common *Me*WRKYs identified in T200 and TME3 at 12, 32 and 67 dpi.

2.3.2 Phylogenetic analysis of differentially expressed WRKYs in Arabidopsis, cassava T200 and TME3

With regards to phylogeny, most of the homologs described by Phytozome are shown to lie topologically close to their cassava counter-part. Furthermore, many of the branches are unresolved on the basis of low bootstrap values. Notably a large number (15/23) of the *Me*WRKYs, identified in T200 and TME3, fall into Group 2 which is the largest subgroup within the WRKY family (Figure 2-3). Other than groups 1 and 2e, which contain only DE *Me*WRKYs from T200, there doesn't appear to be much segregation within the different WRKY groups of *Me*WRKYs from T200 or TME3 within a particular subgrouping. Interestingly, most of the DE *Me*WRKYs found in both T200 and TME3 fall specifically within group 2e.

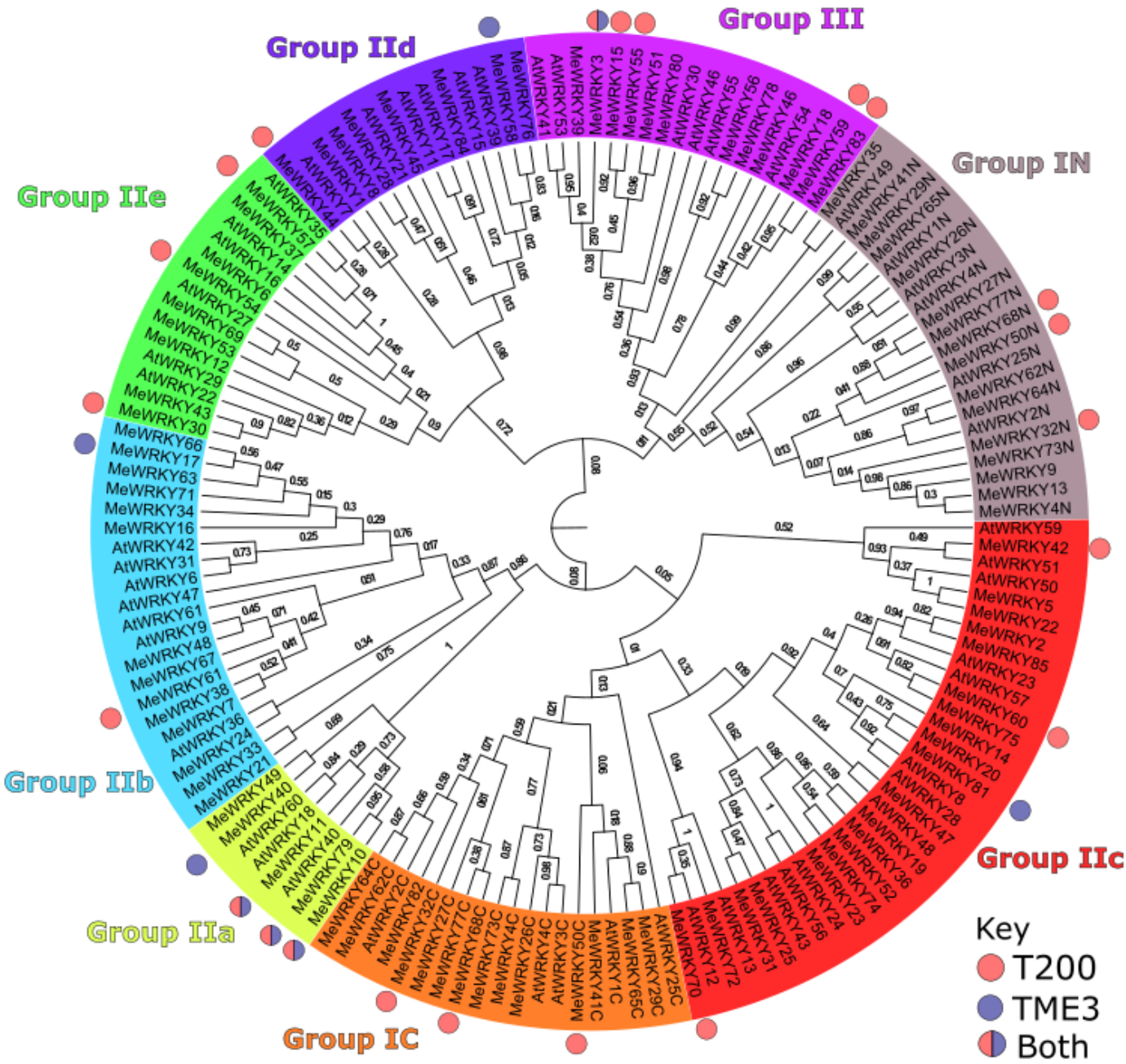


Figure 2-3: A phylogenetic tree of WRKY transcription factors from *Arabidopsis thaliana* (At) and cassava [*Manihot esculenta* (Me)] sorted into taxonomic groups based on amino acid sequences (bootstrap value of a 1000). The tree demonstrates the three main groups that WRKY transcription factors belong to. Group I is characterized by two WRKY domains, Group II by one WRKY domain and a C2H2 zinc finger motif, and Group 3 has one WRKY domain and a C2HC Zn finger motif. Group 1 has been subdivided into two sub-groups (Group IC and Group IN) representing the alignment of the C and N terminus of the protein respectively. The tree is annotated (dots) with the differentially expressed WRKYs in SACMV-infected Arabidopsis. The phylogenetic tree was provided by Hu *et al*., 2016 and altered in the Inkscape program to show DE WRKYs in T200 and TME3 at all 3 time points.

2.3.3 Gene ontology functional enrichment analysis of biological processes occurring in T200 and TME3

Due to the incomplete annotation of genes and their function in the cassava genome, the inferred function was performed through homologous sequences identified from the Arabidopsis genome. The Arabidopsis homologs of each of the genes were used to identify the top 10 most enriched biological processes amongst up- and down-regulated genes in T200 (Figure 2-4A) and TME3 (Figure 2-4B). In T200, the enriched biological processes of down-regulated genes showed higher gene counts than those of up-regulated genes (Figure 2-4A). This coincides with the global transcriptomic data (Allie *et al*., 2014), where the expression patterns of the *Me*WRKYs which showed a greater proportion of down-regulated genes in T200 compared with TME3. Particularly in T200 many of these down-regulated processes at 32 and 67 dpi are involved in immune function, such as responses to wounding and jasmonic acid. Several other notable enriched down-regulated GO processes in T200 at all three infection time points were responses to chitin and organo-nitrogen compounds. Furthermore, at 67 dpi, regulation of innate immune responses was highly represented by down-regulated genes, associated with susceptibility persistence in T200. In comparison, many of the highly represented up-regulated processes were involved with plant development and growth, such as the regulation of meristem growth, regulation of growth and cell wall biogenesis. The overall GO enrichment pattern observed in tolerant TME3 differed considerably compared with T200 (Figure 2-4B). This is most notable at 32 and 67 dpi where the number of up-regulated compared to down-regulated processes was disproportionately higher, as there was no enrichment found among down-regulated at 32 dpi and only 3 processes at 67 dpi. In contrast with T200, a higher representation of hormonal regulation-associated GOs, belonging to ABA, ET, and JA, was observed at the disease symptom stage (32 dpi) in TME3, whereas in T200, only JA is seen to be among the top down-regulated enriched biological processes at 12 and 32 dpi. TME3 exhibited very low GO enrichment in biological processes at 12 and 67 dpi compared with 32 dpi, with only three down-regulated enriched processes, namely GOs pertaining to the transport of amides, peptides and oligopeptides, observed at 67 dpi.

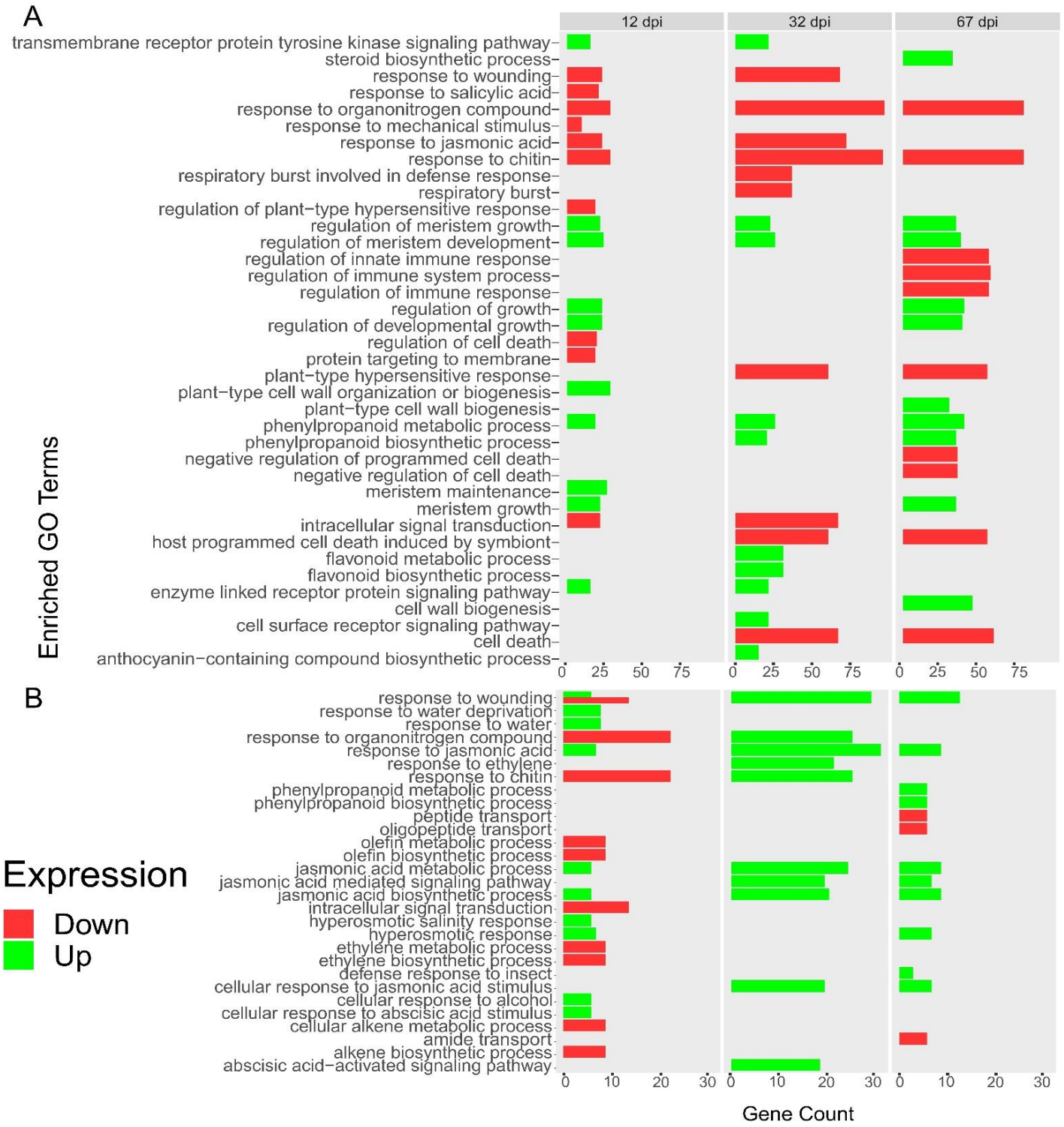


Figure 2-4: Top 10 most enriched biological processes associated in up- and down-regulated genes in T200 and TME3 throughout SACMV infection. The top 10 biological GO terms were taken based on the significance (p-value) of the GO terms in (A) T200 and (B) TME3, plotted against the counts of genes that fell into the GO category. Green represents the process found to be enriched among the up-regulated genes and red are processes enriched among down-regulated genes.

2.3.4 Comparative identification and functions of DE *Me*WRKY-interacting partners in T200 and TME3 based on an Arabidopsis *At*WRKY33 and 40 homolog network

In order to adequately quantify the similarities and differences between the WRKY and interacting partner networks occurring in T200 and TME3 a standard single network was constructed around the two most interconnected DE *At*WRKY homologs (*At*WRKY33 and 40). Arabidopsis homologs of DE *Me*WRKYs in T200 and TME3 at 12, 32 and 67 dpi were mapped to the single network . Functions of WRKYs and DE partners have identified accordingly.

At the early asymptomatic stage of infection (12 dpi), *At*WRKY53 and 70 homologs were up-regulated in T200, while only *At*WRKY40 was up-regulated in TME3. However, despite no common DE WRKYs between T200 and TME3 at early infection, there were considerable commonalities in the homologs of DE interacting partners observed in the central network (Figure 2-5). Notably, a C2H2 zinc finger TF (STZ), a dicarboxylate carrier (DIC2) and a heat shock protein (HSPRO2) are commonly down-regulated at 12 dpi in both landraces, while TET8, MPK3, ERF5 and 6 are uniquely found in T200. In contrast, only a CCR4 associated factor (CAF1b) and the Jasmonate-zim domain 1 (JAZ1) (the only up-regulated interacting partner common between both landraces at 12 dpi) are unique to TME3. It should be noted despite sharing the same homologs (DIC2) in Arabidopsis, the DE cassava genes in T200(cassava4.1_011745m.g) differs from that found DE in TME3 (cassava4.1_011793m.g). This is also the case for HSPRO2 (cassava4.1_006703m.g and cassava4.1_006693m.g in T200 and TME3 respectively). At the symptomatic stage of infection (32 dpi), virus loads increase (Allie *et al*., 2014), and the divergence between TME3 and T200 *At*WRKY53-70 centred networks becomes most apparent. Only three DE interacting partners are common, namely JAZ1, ERF1 and HSFA4A. Interestingly, these three genes show opposite expression between the landraces, where they are are all down-regulated in T200 while being up-regulated in TME3. As is the case with DIC2, the cassava homologs of both *At*JAZ1 and *At*HSFA4A differ between the two landraces. However it should be noted that although the a single Arabidopsis homolog shows contasting expression in the networks, it is actually two separate genes that show differential expression in T200 and TME3. In the instance of JAZ1, cassava4.1_014096m.g and cassava4.1_013620m.g is DE in T200 and TME3 respectively, whereas with HSFA4A, cassava4.1_034415m.g and cassava4.1_008616m.g in T200 and TME3, respectively.

Significantly at 32 dpi, down-regulation of both the homologs of *At*WRKY33 and 40 in T200 is accompanied by a considerable increase in the numbers of DE interacting partners, which in contrast is lacking in TME3 where up-regulation the homolog of *At*WRKY40 persists. The only unique DE interacting partner in TME3 appears to be the homolog of Mitogen-activated kinase 4 MPK4 (cassava4.1_010005m.g), which is up-regulated. At symptom recovery (67 dpi) in TME3 virus load and symptoms decline whereas continuing virus replication and symptoms in SACMV-infected T200 persist. The contrast between the two landrace networks surrounding the two WRKYs is significant (Figure 2-5). The network in T200 persists approximately as it was at 32 dpi, with only minor changes (SIB1 and JAZ1),and all genes are down-regulated whereas in TME3, no differential expression is seen in the interacting partners surrounding the homolog of up-regulated *At*WRKY40.

2.3.5 Analysis of other *Me*WRKYs and DE *Me*WRKY-interacting partners in T200 and TME3 based on homologs in Arabidopsis networks

The STRING network was constructed around both WRKY33 and 40, represented by oval nodes from the *Arabidopsis thaliana* database. These two transcription factors were chosen due to the high number of connections to differentially expressed genes (DEGs) in both TME3 and T200. Red and green nodes represent up- and down-regulation of the DEGs, respectively, at each time point, where the intensity of the colour refer to the change in $\log_2$-fold expression. When WRKY33 and 40 are DE their interactions are highlighted according to their expression.

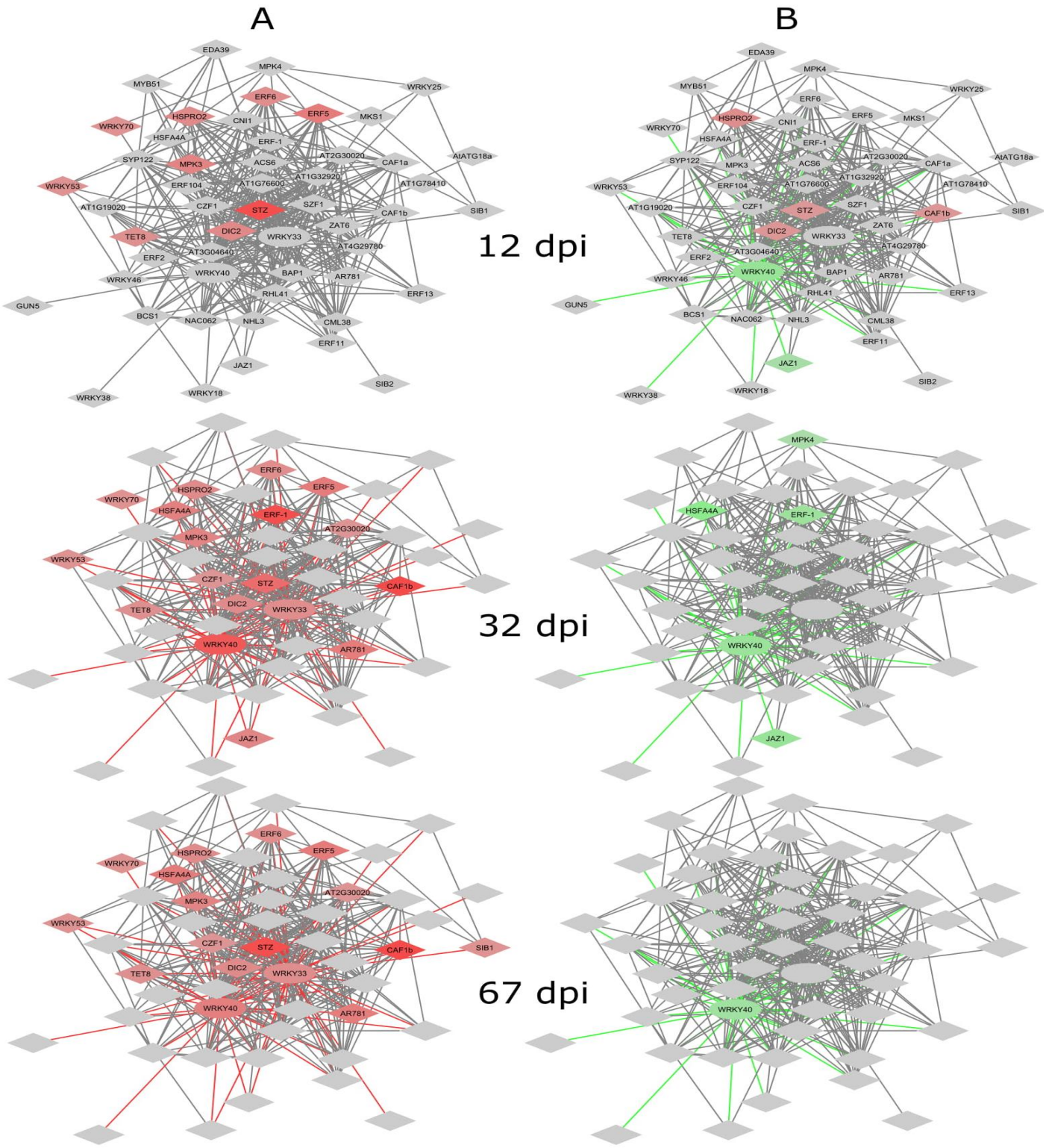


Figure 2-5: STRING protein-protein interaction network centring around AtWRKY33 and 40 and their interacting partners. Differentially expressed (DE) genes from T200 (A) and TME3 (B) at early, middle and late infection are mapped to the network. Red and green genes refer to instances of down- and up-regulation respectively, while shading represents the magnitude of differential expression. Genes that were not DE in 32 and 67 dpi were removed from the network to simplify it.

2.3.6 Predicted pathways and hormone responses of the differentially expressed (DE) T200 and TME3 *Me*WRKY homologs in Arabidopsis

At the early symptomless stage (12dpi) of SACMV infection in T200, a total of five WRKYs homologs are differentially expressed, namely *At*WRKY7, 20, 41, 53 and 70 (Figure 2-6A). Despite having more differentially expressed WRKYs than TME3 (Figure 2-6B), T200 is also observed to have fewer interacting homolog protein partners with only MAPK3 and Suppressor of BIR1 (SOBIR1) connected to *At*WRKY53. *At*WRKY70 and *At*WRKY53. However, through MAPK3 and SOBIR1 T200 is connected to multiple down-regulated phytohormone processes pertaining to JA and SA signalling. The only two up-regulated *At*WRKYs (7 and 20) do not appear to contribute to any downstream defence phytohormone signalling. In contrast, only *A*tWRKY40 is upregulated in TME3, and its predicted interacting network comprises greater diversity and number of differentially expressed interacting partners and downstream hormone processes (Figure 2-6B). From the 11 interacting partners of *At*WRKY40, 10 are down-regulated. These include three uncharacterised genes (AT4G24380, AT4G27280 and AT3G57450), STZ, ERF109 (RRTF1), CAF1b, calmodulin-like protein 37 (CML37), DIC2, and HSPRO2. All these down-regulated genes are connected to ET, ABA, JA, and SA. Notably, *At*WRKY40 was connected to two up-regulated interacting partners, belonging to the Jasmonate-ZIM domain (JAZ) gene family, namely JAZ1 and JAZ8. Both of these are shown to connect to several signalling hormone processes.

At the systemic symptomatic stage of infection (32 dpi), the differences between the WRKY-hormone network inT200 and TME3 became more apparent (Figure 2-7). Notably, in T200, all WRKY DEGs are down-regulated and are highly connected to their interacting partners and hormone responses (enriched GO biological processes) compared with early infection (Figure 2-7A). *At*WRKY53 (*Me*WRKY15 and 55) and 70 (*Me*WRKY59 and 83) remain down-regulated at 32 dpi in T200, but it is the additional down-regulated *At*WRKY33 (*Me*WRKY68) and 40 (*Me*WRKY10,11,79) TFs which seem to be responsible for the majority of down-regulated interacting gene partners in the functional network and GO processes (Figure 2-7A). Particularly, down-regulated *At*WRKY33 and 40 showed the highest degree of network connections, with the *At*WRKY33-interacting partners connected to all the DEGs except *At*WRKY70 and the uncharacterized phosphatase AT2G30020. *At*WRKY33, 40, 53 and 70, and interacting network, partners such as CAF1b, DIC2, STZ, and *At*ERF1 (homolog of *Me*ERF100), are down-regulated in T200. There are a number of interacting partners unique to T200 at this

time point such as the hypothetical protein TET8, the calcium-binding protein (AT4G27280), the putative yeast pheromone receptor-like protein AR781, two zinc-fingers: a C3HC4-type zinc finger (RDUF2) and a putative CCCH-type (CZF), HSFA4A, and two ethylene response factors *At*ERF5 and *At*ERF6. Interestingly, *At*WRKY44 and AT1G23710 are the only genes that do not show down-regulation, however, *At*WRKY44 is unlikely to play a role in the predicted network due to the lack of connections.

In contrast to T200, at 32 dpi where TME3 shows less severe symptoms compared to T200, all the interactors in the network are up-regulated (Figure 2-7B). The most interesting observation is that *At*WRKY40 remains up-regulated during middle infection, however, the orthologs have changed. Whereas in 12 dpi, *Me*WRKY40 and 79 was DE, in 32 dpi *Me*WRKY10, 11 and 79 all show up-regulation. Additionally, *At*WRKY6 is uniquely represented at this stage of infection and is up-regulated. *At*JAZ1 and 8 remain up-regulated at 12 dpi is now up-regulated, however, the number of homologs belonging to these two genes has increased during middle infection. Indeed, at 12 dpi, only a single ortholog belonging to each of the JAZs was up-regulated, whereas, at 32 dpi, there were three homologs of *At*JAZ1 and *At*JAZ8 up-regulated.

At 67 dpi, in TME3 there appears to be no differentially expressed interacting proteins, except for JAZ8 which is not seen to be linked to any hormone process among enriched hormone processes (Figure 2-8B). In contrast, T200 exhibits a similar, but a reduced, network topography at 67 dpi as at 32 dpi, with the majority of represented partners downregulated. The only exception is seen with the homolog of *At*WRKY22. Notably, almost all the genes found down-regulated in T200 at 67 can be found also in T200 at 32 dpi, the exception being AT4G34150, WRKY22 and SIB1. Excluding the interacting partners, T200 shows down-regulation in many of the hormone processes found at 32 dpi, demonstrating very little change overall.

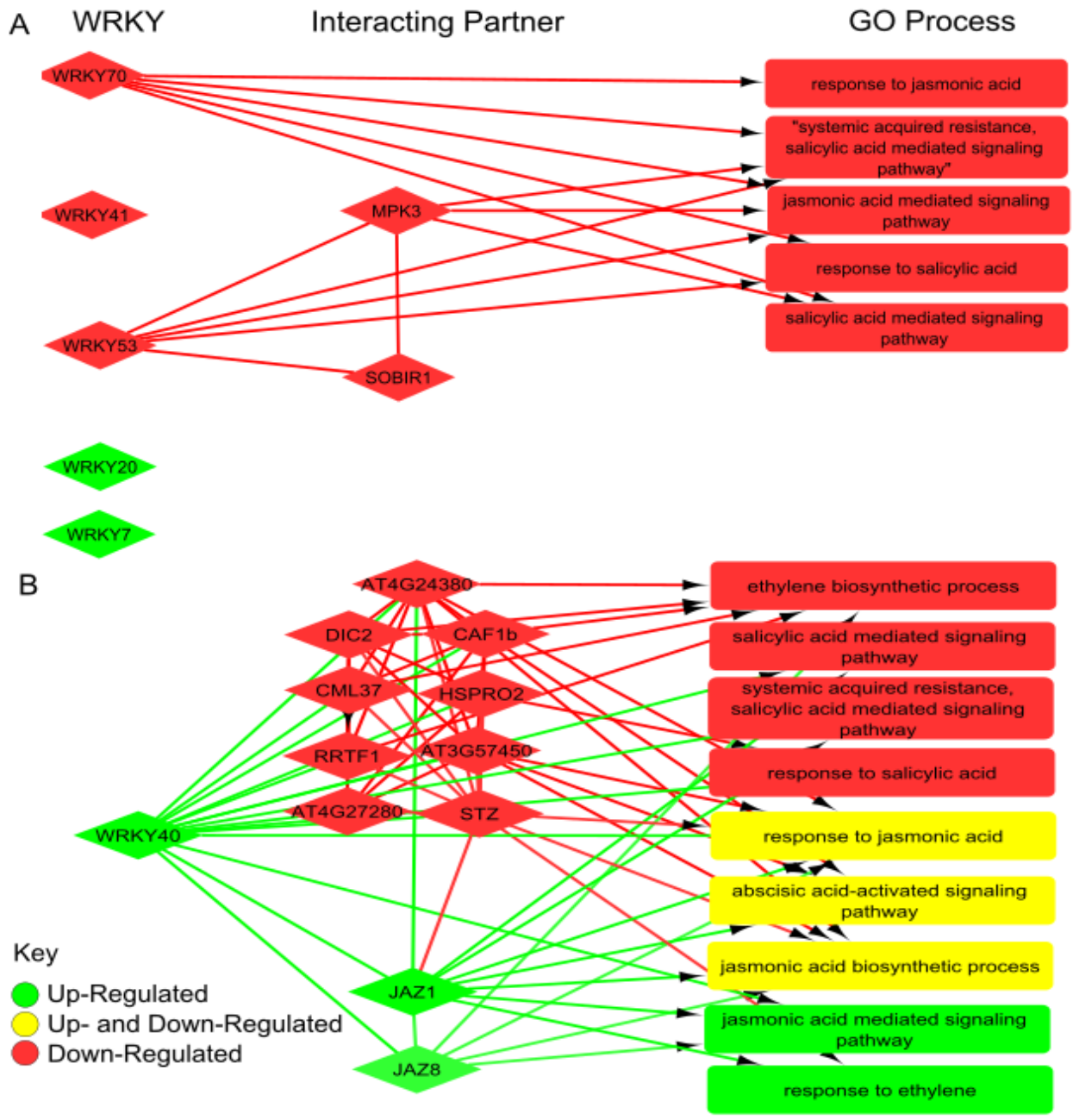


Figure 2-6: Predicted pathways and hormone responses of the differentially expressed (DE) WRKYs identified in (A) T200 and (B) TME3 at 12 dpi. From left to right: the DE WRKY TFs identified at 12 dpi, their differentially expressed interacting partners from STRING, and on the right the enriched GO biological process (hormone responses) derived from the DE WRKYs and DEGs. Red and green represent the relative Log_2-fold change expression, down-regulation and up-regulation, respectively, at 12 dpi at each of the nodes, and the GO terms enriched with only the down-regulated genes. Yellow represents GO processes that were found to be enriched in both up- and down-regulated genes at 12 dpi.

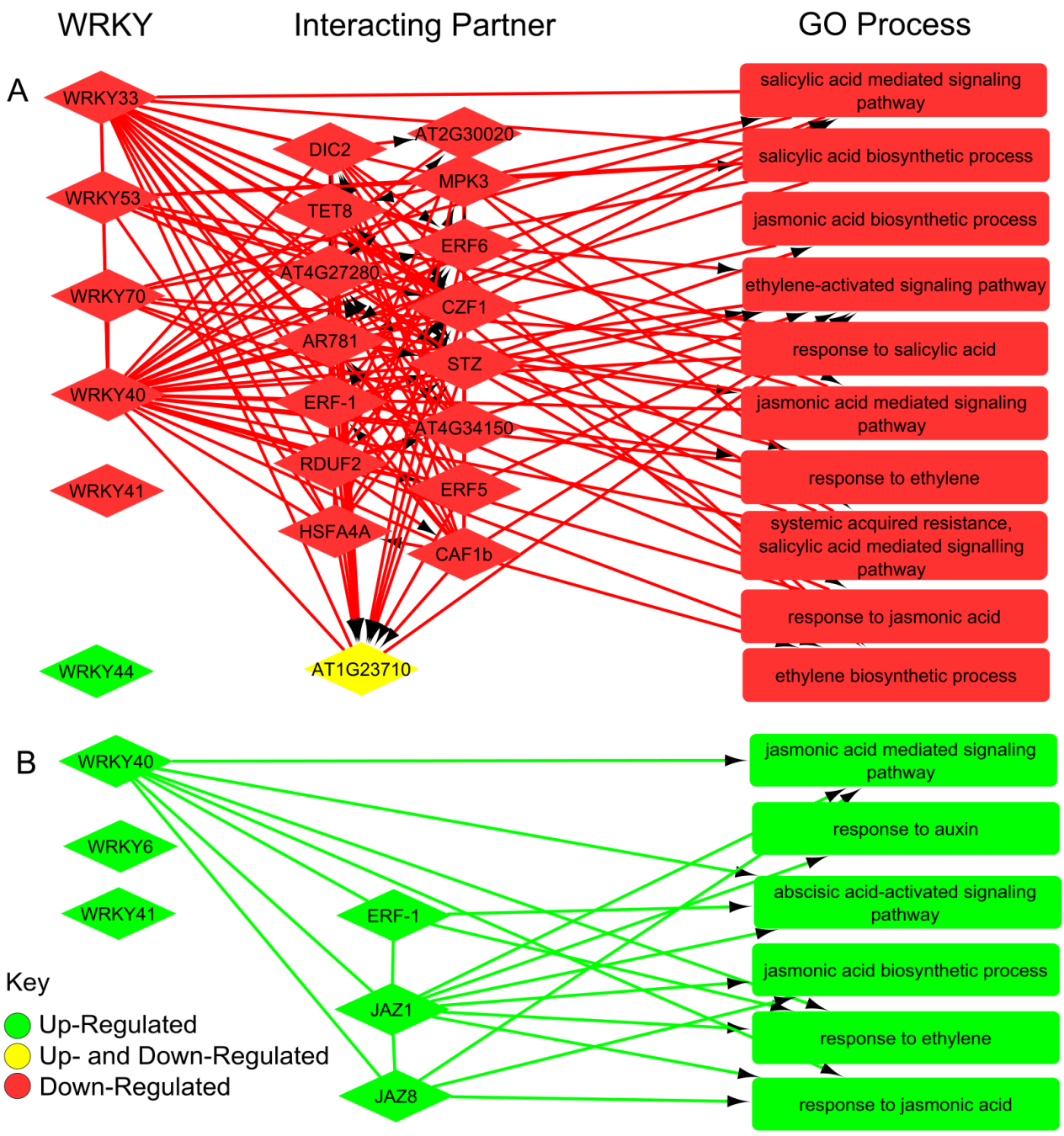


Figure 2-7: Predicted pathways and hormone responses of the differentially expressed (DE) WRKYs identified in (A) T200 and (B) TME3 at 32 dpi. From left to right: the DE WRKY TFs identified at 12 dpi, their differentially expressed interacting partners from STRING, and on the right the enriched GO biological process (hormone responses) derived from the DE WRKYs and DEGs. Yellow represents a single Arabidopsis gene that has two cassava homologs that show contrasting differential expression.

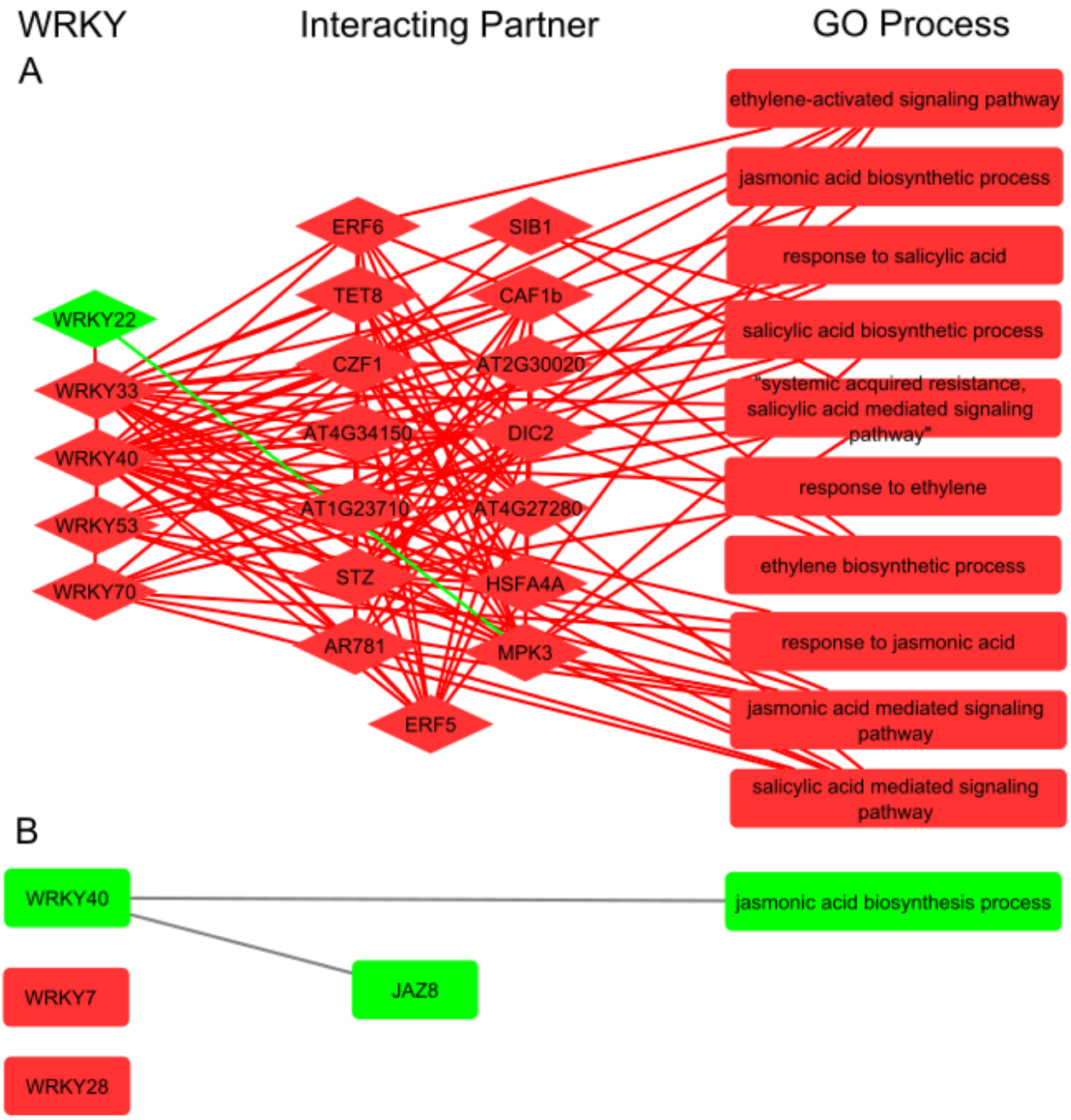


Figure 2-8: Predicted pathways and hormone responses of the differentially expressed (DE) WRKYs identified in (A) T200 and (B) TME3 at 67 dpi. From left to right: the DE WRKY TFs identified at 12 dpi, their differentially expressed interacting partners from STRING, and on the right the enriched GO biological process (hormone responses) derived from the DE WRKYs and DEGs.

Both TME3 and T200 exhibit hormone responses, namely ABA, ET, JA and SA. Significantly, the WRKY-responsive phytohormone- responses are mostly down-regulated in T200 over the entire course of SACMV infection (Figure 2-9). At the early non-symptomatic stage of infection (12 dpi), the number of differentially expressed hormones is lower compared to 32 and 67 dpi, and TME3 exhibited a small number of both up- and down-regulated hormones. At 32 dpi, T200 exhibits more severe symptoms (Figure 2-1) compared with TME3, and viral loads are higher (Allie *et al*., 2014), and there is a significant increase in the number of down-regulated hormones in T200. In contrast, the number of hormone-related genes in TME3 is lower than T200, but notably, they are all up-regulated. At 67 dpi, TME3 shows recovery and starts to exhibit tolerance (milder symptoms and lower viral loads). This is accompanied by an absence of any hormone responses, except for jasmonic acid. T200, on the other hand, displays severe symptoms (Figure 2-1A) and high virus titers at late infection (67 dpi) (Allie *et al*., 2014), and continues to show a high number of down-regulated hormone-related genes.

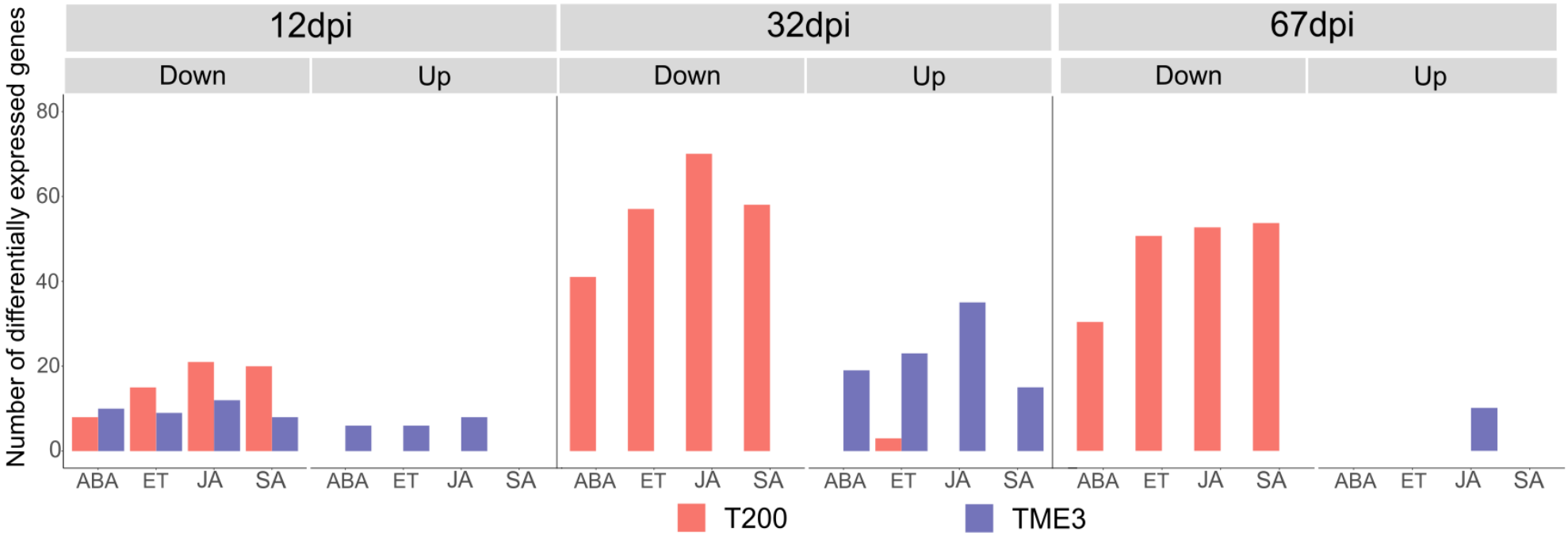


Figure 2-9: The number of up- and down-regulated DEGs involved in phytohormone processes in TME3 and T200. The bar graph shows the number of up- and down-regulated DEGs found to be involved in the ABA, ET, JA, and SA signaling based on the infection stage and relative log_2-fold change in expression. These numbers were determined by counting all unique genes involved in any enriched GO process associated with phytohormones

2.4 Discussion

2.4.1 General

Before comment may be made regarding the STRING networks in both landraces it is first important to discuss the link between gene expression data and a protein interaction network. It is known that genes exert function through a combination of level of expression and the processes their protein products undergo. Although it is possible to infer interaction between genes through a gene co-expression network, the proteins networks in which their gene products contribute to are not as easily inferred. Placing the gene products into a protein interactions network enables for the commentary of function networks that are being influenced by the change in expression observed in the DEGs. To summarize, this method of discussion focuses on the gene products (proteins) and how the functional network in which it belongs to changes with respect to change in gene expression.

WRKY transcription factors and phytohormones have been shown in many studies to be involved in abiotic and biotic stress responses in plants (reviewed by Bakshi and Oelmüller, 2014). While some reports on bacterial and fungal infections are available in experimental and model hosts (Xu *et al*., 2006, Birkenbihl *et al*., 2012, Hu and Dong, 2012), few studies on viruses have been performed, and even less so on non-model crops. In this study, we demonstrate for the first time, that cassava *Me*WRKY transcription factors and phytohormone networks play pivotal roles in the regulation of susceptibility and tolerance to cassava mosaic disease. Significant differences in gene expression profiles of *Me*WRKY transcription factors and hormone network signalling responses between SACMV-infected T200 and TME3 provide important hallmarks for regulation of susceptibility and tolerance, respectively, to cassava mosaic disease. In general, down-regulated *Me*WRKYs are more highly represented in SACMV-susceptible T200 compared to tolerant TME3, supporting previous findings of higher numbers of total suppressed transcripts in T200 compared to TME3 (Allie *et al*., 2014). *At*WRKY40 homologs play a contrasting central role in both T200 and TME3, while *At*WRKY53 and 70 homologs form central hubs in the network only associated with T200. The unique expression of *At*WRKY6 and 28 homologs in TME3 could uncover important regulatory and signalling mechanisms associated exclusively with CMD tolerance, and require further attention. Expression of SA, JA, ET and ABA at 32 dpi in TME3 demonstrate that several signalling pathways triggered during the systemic infection

stage (32 dpi) lead to recovery and tolerance, but that only JA is involved in the maintenance of tolerance at 67 dpi and thereafter. While little is known about molecular mechanisms of tolerance compared to resistance, this study further supports the view that tolerance represents a low energy cost for both virus and plant host during their long term persistent co-existence (Bengyella *et al*., 2015, Paudel and Sanfoçan, 2018). In contrast, the hallmark of susceptibility in T200 is downregulation of almost most WRKYs and hormones ET, SA, JA and ABA at all stages of infection, which would severely impair downstream defence responses.

2.4.2 Phylogenetics

Phylogenetic analyses demonstrated there were multiple instances of homologs in cassava and Arabidopsis grouping together such as *Me*WRKY42 *At*WRKY51, *Me*WRKY47 and *At*WRKY48 (Figure 2-3). From this observation, it is suggested that a number of common WRKYs were present prior to the divergence of cassava and Arabidopsis (Hu *et al*., 2016). Furthermore, it was observed that many of cassava and Arabidopsis WRKYs share a high degree of similarity (*Me*WRKY1 and 44 with *At*WRKY7 and *Me*WRKY69 with *At*WRKY27) (Wei *et al*., 2016). For these reasons, it was feasible to use the functions of Arabidopsis homologs to infer the function of uncharacterized WRKYs in cassava. However, the functions of *Me*WRKYs need to be confirmed in *vivo* experimental studies in the future.

It is interesting that the DE *Me*WRKYs from T200 fall into all three Groups I, II and III whereas those from TME3 were found almost entirely in group II (only *Me*WRKY3 was found in group III). However, the reason that the DE WRKYs in T200 are found in more groups could be the consequence of the T200 experiencing a greater level of transcriptional changes (Allie *et al*., 2014). The distribution of the DE WRKYs in TME3 could be based on chance as group II is the largest of the three groups, alternatively, this could reflect the evolutionary lineage and function of Group II WRKYs. Indeed, it may be that members of Group II arose from a common ancestor that had biotic response functions, and subsequently evolved and diversified into the members seen today. Some evidence supporting this notion is observed in Group IId in Arabidopsis, and likely cassava too, where members have been shown to interact with calmodulin that detects Ca^{2+} signalling in defence responses (Park *et al.*, 2005). Additionally, in Group IIa all the members in Arabidopsis (*At*WRKY18,40 and 60) show cooperation in regulating ABA signalling and modulate resistance and susceptibility to pathogens, implying defence functioning (Liu *et al*., 2012, Xu *et al*., 2006). Furthermore, all the WRKY members of

Group I and IIc interact with a group of proteins containing a conserved VQ motif, which is used to modulate growth, response to stimulus, development, and most importantly defence response (Cheng *et al.*, 2012). This functional link between Groups I and II is likely due to both groups sharing a C_2H_2 zinc finger motif (Wang *et al.*, 2014). However, to date, there is no experimental evidence linking differences in the genes that WRKYs regulate based on the zinc finger motif present.

Differences in Group I, II or III-associated virus-responsive WRKYs between plant hosts are likely to occur. In tomato, for example, six *Soly*WRKY Group III TFs (41, 42, 53,54,80 and 81) were responsive to Tomato yellow leaf curl virus (TYLCV) (Huang *et al*.,2016), and contributed to defence responses. It was found that these six WRKYs in tomato were down-regulated in susceptible tomato but not resistant tomato infected with TYLCV (Chen *et al*., 2013). However, it is interesting to note that there is overlap between the Arabidopsis orthologs between these six tomato WRKYs and those identified in this study. *Soly*WRKY41 is the ortholog of *At*WRKY41, while *Soly*WRKY53,40 and *Soly*WRKY80,81 are the orthologs of *At*WRKY53 and *At*WRKY70 respectively. All three of these orthologs were found to be down-regulated at all three time points in cassava which further supports the notion that the suppression of these genes is contributing to susceptibility in T200.

While highly conserved, WRKYs have evolved to respond to specific viruses in different hosts. During WRKY TF evolution, exon/intron gains, losses or duplications could have occurred leading to functional diversity. However, the suppression of orthologous WRKY genes in the same group (III) by two closely phylogenetically related viruses, namely TYLCV and SACMV, in two different susceptible plant species brings with it an interesting connection. This realization would suggest that these two geminiviruses have evolved similar strategies of establishing disease in different plants through suppressing orthologous proteins.

Significantly, in the West African CMD-tolerant landrace TME3, at least one of the *Me*WRKY orthologs of *At*WRKY40, namely *Me*WRKYs10, 11 and 79, are up-regulated at all three stages of SACMV infection. This finding strongly supports the notion that *At*WRKY40 homologs form an important hub in TME3 that is central to hormone signalling interactions, and downstream associated SACMV-tolerance responses. This is supported by the fact that in the susceptible T200 landrace, the opposite pattern occurs, where *At*WRKY40 homologs are down-regulated at 32 and 67 dpi. In T200, significantly *At*WRKY 53 (*Me*WRKY15 and 55) and 70 (*Me*WRKY59

and 83) homologs are down-regulated throughout infection and form a suggested node contributing to a susceptibility phenotype. Interestingly, *Me*WRKY17 (*At*WRKY6), uniquely expressed in TME3 at 32 dpi, interacts with an ABA-responsive MAPK and via a receptor-like kinase (FRK1), triggering a highly diverse and interconnected hormone signalling and defence response network with several nodes (Figure 2-8). Transcription factors form part of a highly connected network that regulates gene expression, and functional analysis of interactions of TFs with hormones and other proteins is important in elucidating different signalling cascades (Alves *et al*., 2014). WRKY interacting or binding partners have already been identified, and these interactions play roles in signalling, chromatin remodelling and other cellular processes (Chi *et al*., 2013). Many of the interacting partners in the protein interaction WRKY response networks were involved in defence and hormone signalling (Figures 2-6 and 2-7).

2.4.3 Gene Ontology

The top 10 GO process in both T200 and TME3 reveal insights into the different processes occurring during SACMV infection. Coinciding with the hormone response results (Figure 2-9), the divergence in between the two landraces is already apparent starting from 12 dpi. Early on, there is suppression in the two most important hormone response in T200, namely JA and SA. Furthermore, in T200 the up-regulation of various growth and meristem processes point towards SACMV reprogramming host system in order to facilitate a better replicative environment. This has been previously seen with another geminivirus in Arabidopsis (Ascencio-Ibánez *et al*., 2008). This suppression of hormone response and induction of meristem growth suggests that SACMV is employed as a strategy that focuses on disease establishment in T200. However, there appears to be a shift of strategy in SACMV during middle infection in T200 to focus more on disease maintenance and promote subsequent transmission. This is seen in the shift whereby there is a greater suppression of genes involved with defence response (response to JA, wounding and chitin) in addition to the suppression of genes involved in cell death and the hypersensitive response (HR), despite being absent in cassava. Overall this strategy seems to persist into late infection with a more target suppression of genes directly involved with the immune system of the plant. In contrast, TME3 exhibits a comparatively more positive immune response to SACMV at 12 dpi. Despite showing suppression of chitin, a characteristic of PTI, TME3 demonstrates a response to JA signalling and JA biosynthesis which was notably lacking in T200 at the same time point. At 32 dpi, the lack of any down-regulated processes shows a

visible switch in the response by TME3. Here there is a strong induction of the various aspects of defence hormone signalling, particularly in JA, ET and ABA. This is likely to indicate the start of systemic resistance in TME3 as jasmonates, a JA intermediate, is one of the mobile signalling molecules found in Arabidopsis, whereas ET is a modulator of immunity (Chaturvedi *et al*., 2008, Anderson *et al*., 2004). Transitioning to 67 dpi, TME3 starts to display lower levels of enrichment, however, the persistent induction JA related processes implies that JA plays a key role in coordinating the defence response in TME3. Further focus on hormone signalling processes is required to better understand the nuances of systemic hormone signalling experienced in both landraces.

2.4.4 Hormone signalling

The global hormone responses of both landraces reveal some insights surrounding SACMV infection in cassava (Figure 2-9). Due to viruses sharing more similarities with biotrophic rather than necrotrophic pathogens with respect to both depending on living tissue to invade, it was expected that tolerant cassava landrace TME3 would exhibit a predominantly strong SA response since SA is associated with resistance to biotrophic pathogens (Grant and Lamb, 2006). TME3 had a low number of SA-associated genes expressed at 12 dpi, which increased at 32 dpi, but at recovery (67 dpi) there was no SA-related gene expression detected. SA is generally, but not necessarily, required for resistance induction and the HR (Zhang *et al.*, 2004b). SA and ABA are also associated with antiviral RNA silencing (Li *et al.*, 2012, Hunter *et al.*, 2013). Resistant or tolerant cassava does not display a typical HR phenotype, and we propose that SA is induced as a basal biotic stress response which does, along with several other factors, reduce but not inhibit viral replication and spread (Allie *et al*., 2014). At 32 dpi, SA as well as ET, JA and ABA all increase in expression in TME3 (Figure 2-9), a general response we believe to be associated with one of the triggers for recovery and tolerance. Interestingly, expression of JA-associated genes was not expressed at 12 dpi (JAZ1 a repressor of JA biosynthesis was down-regulated) and was only expressed at 67 dpi, and not ET, SA or ABA, suggesting a persistent role for JA in tolerance. The role of JAZ proteins is well studied in the context of JA signalling and defence responses alike. JA-Ile, the active signalling form of JA, binds COI1-JAZ complexes, consequently leading to JAZ degradation through the 26S proteasome complex thereby allowing TFs like MYC2 to transcribe JA responsive genes (Sheard *et al*., 2010, Kazan and Manners, 2013). It has been suggested that JA is associated, albeit not exclusively, with ISR (Pieterse *et*

al., 1998). Virus replication must be permitted if viruses are to induce any host response, and low levels of SACMV (Allie *et al*., 2014) may be required to trigger defence-related responses in TME3, in particular at 32 dpi after which virus levels decrease and recovery and tolerance are established. Low virus level persistence in tolerance is necessary to maintain a host response (Paudel and Sanfacon, 2018). The proposal that tolerance is a mutual co-existence of the virus and host defence and counter-defence mechanisms (Bengyella and Rey, 2015) would be supported by the hormonal response found in TME3. A response of all four hormones at 32 dpi would signal several immune-related mechanisms in TME3 such as RNA silencing (Rogans *et al*., 2016), R gene-mediated responses (Bengyella and Rey, 2015) and other molecular mechanisms (Allie *et al*., 2014) leading to recovery and persistent tolerance. In contrast, in T200 being susceptible, a SA response would not be expected at all, or attenuated, and indeed this was the case. Evidence shows that susceptibility in T200 may be attributed, amongst other factors, to the suppression of global hormone signalling at all stages of SACMV infection, particularly SA and JA, as most of the genes down-regulated are associated with these two hormones. Symptom severity and virus loads during middle symptomatic and late infection stages reflect the hormonal profiles of both landraces accordingly. Therefore, it is hypothesized that the WRKY transcription and hormonal signalling mechanisms contributing to tolerance in TME3, notwithstanding many other interconnected molecular mechanisms involved, encompass three phases: the first begins at the early asymptomatic stage, and the effects begin to manifest at the middle symptomatic infection stage, which is followed by recovery and attenuated symptoms at late infection and beyond.

2.4.5 Central protein-protein network

The central network was derived by creating a STRING network using the most interconnected DE WRKYs as central nodes and observing what connections form when setting the secondary shell to 50. It should be noted that this central network was not constructed to show a conserved topology but demonstrate the responses are different, even in the same network, between T200 and TME3.

The general network was found to be concordant with hormone GO results demonstrating the increasing divergences in the manner in which T200 and TME3 respond to SACMV (Figure 2-5). This is evident in TME3, where commonly down-regulated genes with T200, begins to show contrasting expression at 32 and 67 dpi. Regarding the interconnecting network in TME3

at 12 dpi, the downregulated *At*WRKY40 was linked to repressed genes HSPRO2, STZ and DIC2. Since these three DEGs also showed down-regulation in T200 at 12 dpi, where *At*WRKY40 was not found to be DE, it could be argued that the *At*WRKY40 orthologs in TME3 are not responsive for the down-regulation of the HSPRO2, STZ and DIC2 homologs. Further support for this observation can be found in T200 at 32 dpi, when orthologs of *At*WRKY40 are down-regulated (opposite to TME3), these genes do not show a change in their expression patterns. Notably, *At*STZ is responsive C_2H_2 zinc-finger TF (Kielbowicz-Matuk, 2012) which inhibits JA biosynthesis and oxidative stress (Zhou *et al.*, 2009). It is a transcriptional repressor and can repress stress responsive genes such as *DREB1* (Flower and Thomashow, 2002). Furthermore, STZ is involved in JA early signalling response and may regulate the expression of JAZ1, which is a key repressor in the JA signalling cascade (Pauwels and Goossens, 2008). Although *At*STZ is more commonly associated with abiotic stress response, it has also been implicated in HR against *Cucumber mosaic virus* (Marathe *et al*., 2004). The likely reason for STZ involvement in HR, stems from the paradigm that biotic and abiotic network show overlap as they both use phytohormones to coordinate the appropriate response (Fujita *et al*., 2006). In T200, the comprehensive down-regulation of STZ through SACMV infection would likely contribute to susceptibility indirectly by interfering with signal responses at the point of intersection between biotic and abiotic networks. Alternatively in TME3, it is possible that the down-regulation of STZ and other down-regulation of ABA related genes (*CAF1-b, AMT2, NAC002, DSP3, PP2C5 SAG21, EDL3, GRXC9*) may promote low levels of JA required for tolerance. However, the role of these factors in early infection in TME3 is not clear.

Looking at the common genes between T200 and TME3 at 32 dpi, the contrasting co-expression patterns of *At*ERF-1, *At*HSFA4A and *At*JAZ1 suggest that these genes are positively regulated by *At*WRKY40. However, the absence in the network, in particular, of *At*JAZ1 and *At*ERF-1 at 67 dpi in both landraces despite the *At*WRKY40 orthologs still showing differential expression, suggests one of two possibilities. Firstly, the specific cassava orthologs (*Me*WRKY10,11,40 and 79) of *At*WRKY40 may have different targets, or alternatively, *At*WRKY40 requires other stimuli in order to induce these three genes. These generalized networks around *At*WRKY33 and 40 provided some insight into the global network divergence between the two landraces, in addition to providing evidence for interaction of the orthologs of *At*WRKY40 with *At*JAZ1, *At*HSFA4A and *At*ERF-1. However, if information is to be gained

regarding the specific role that these WRKYs contribute to tolerance or susceptibility, then more detailed networks occurring in each landrace requires scrutiny.

2.4.6 Early

2.4.6.1 T200

Early infection is a time point which marks an absence of symptoms despite the detectable presence of the virus (Allie *et al*., 2014, Bengyella *et al*., 2015). Despite T200 and TME3 sharing the absence of symptoms, it can be observed that each landrace is responding to SACMV infection in contrasting ways. In T200, all ABA, SA, ET and JA associated genes are down-regulated whereas the number of up- and down-regulated DEGs involved in phytohormone processes in TME3 are lower, but show both up- and down-regulation (Figure 2-9). SA and JA responses are down-regulated which is connected to the down-regulation *At*WRKY53 and70 (ortholog of *Me*WRKY55 and 59, respectively), MPK3, and SOBIR1 (Figure 2-6A). *At*WRKY70, *At*WRKY53 is associated with the positive resistance of plants towards necrotrophic fungi and responds to salicylic acid signalling and reactive oxygen species (Zheng *et al*., 2006). In addition, both *At*WRKY53 and 70 functions as positive and negative regulators of senescence respectively (Besseau *et al*., 2012). Although *Me*WRKY55 and 59 show mutual down-regulation in T200, their ortholog *At*WRKY53 and 70 are induced by opposing phytohormones (Miao *et al*., 2013). Whereas *At*WRKY70 is induced by SA but silenced by JA, *At*WRKY53 is silenced by SA and induced by JA (Li *et al*., 2004, Miao and Zentgraf,2007).

Despite this opposing hormone induction response, *At*WRKY53 and 70 have been shown to have overlapping functions as they both confer resistance against the *P. syringae* (Hu *et al*., 2012, Shim *et al*., 2013). Therefore, the early down-regulation in both *At*WRKY53 and 70 in T200 implies that SACMV may be interacting with a common upstream protein of these two TFs. Furthermore, both *At*WRKY53 and 70 are up-regulated SAR tissue in *A.thaliana*, indicating that the down-regulation observed in T200 may contribute to the suppression of SAR leading to enhanced susceptibility to SACMV (Gruner *et al*., 2013). This is further supported when considering that the orthologs of *At*WRKY53 and 70 are down-regulated through SACMV infection.

Connected to *At*WRKY53 are two protein kinases, MPK3 and SOBIR1. Both of these interacting partners show particular importance when considering both are seen to be down-regulated throughout SACMV infection. This suggests that both *At*SOBIR1 and *At*MPK3 are

likely the interacts with either *At*WRKY53 or 70, however, the link has yet to be experimentally determined. *At*MPK3 is well documented for its function in PTI, acting as a primer for PTI which forms the foundations of the basal response in plants (Nühse *et al*., 2000). Furthermore, MPK3 has been further implicated in defence by contributing to both JA and ET biosynthesis (Meng and Zhang,2013, Schweighofer and Meskiene, 2008). The fact that homologs of *At*MPK3 remain down-regulated throughout SACMV infection may suggest why, in part, T200 fails to mount a JA and ET oriented response like that seen in TME3. In contrast, *At*SOBIR1 is a leucine-rich receptor-like kinase. It has been found that *At*SOBIR1 is a suppressor of the negative resistance regulator *At*BIR1 and acts as a signal transducer in cell death and disease resistance pathways (Gao *et al*., 2009). The down-regulation of *At*SOBIR would entail the accumulation of *At*BIR1 leading to suppression in immune response and prevent cell death. Therefore, the continued down-regulation of these two signalling molecules through *Me*WRKY55 or 59 is likely responsible for the suppression of downstream defence pathways. It is clear, starting from early infection, that T200 has a significantly reduced ability to respond to SACMV due to the critical down-regulation of *Me*WRKY55 and 59.

2.4.6.2 **TME3**

In contrast to T200, TME3 experiences a high degree of hormonal antagonism at 12 dpi, as seen up- and down-regulation observed in genes relating to JA, ABA and ET processes (Figure 2-9). This is not surprising as *At*WRKY40 orthologs(*Me*WRKY40 and 79) reside at the centre of three separate hormone signalling pathways, namely ABA, SA, and JA. It is expected that the interacting partners of *At*WRKY40 would be predominately down-regulated as *At*WRKY40 has been described as a transcriptional suppressor (Pandey *et al*., 2010). This is seen in the case of TME3 during early infection, however, when considering the which homologs are found in T200 and TME3 when *At*WRKY40 is DE (DIC2, CCR4-associated transcription factor (CAF)1b, STZ, JAZ1,8) only *At*JAZ1 and 8 show contrasting expression. Furthermore, when this information is taken together regarding the role the JAZ proteins are performing in light of the hormone response data, a proposition can be made. These results propose JAZs positively regulate JA signalling and response in cassava rather than negatively, which is seen in Arabidopsis, and that the orthologs of *At*WRKY40. If this were the to be true, then the up-regulation seen in the *At*WRKY40 orthologs would be promoting JA signalling and contributing to resistance. To support this, it has been demonstrated that *At*WRKY40 plays a role in plant early response to JA

signalling through early induction during hormone treatment (Wang *et al.*, 2008). Furthermore, as previously mentioned, *At*WRKY40 is induced by either SA or ABA, however, if SA induces *At*WRKY40, ABA response genes are suppressed. Based on the hormone responsive gene count, it is evident that *At*WRKY40 is rapidly induced by ABA signalling at early infection and not SA signalling, demonstrating pathway antagonism which leads TME3 to prioritize this pathway.

This is likely the reason that *At*WRKY40 is found up-regulated in TME3 and not T200, which shows considerable and wide-spread suppression of hormone responsive genes. This ABA-*AT*WRKY40-JAZ pathway is unique to TME3 and could suggest it is, in part, responsible for conferring tolerance. Examining the interacting partners of *At*WRKY40 it is clear that the many are also found in T200, which may explain why TME3 still present symptoms. There are a few interacting partners which appear to be contributing to disease establishment in TME3. For example, *At*CML37 is a calcium-binding protein used to perceive Ca^{2+} secondary signalling, and is linked to induction of a defence response through jasmonate accumulation (Scholz *et al*., 2014). The down-regulation of this gene would lead to improved feeding by whitefly and would also reduce JA response by limiting JA accumulation in cells.

Additionally, CAF1b is involved in the deadenylation of stress-responsive mRNA, which in turn regulates the breakdown of these molecules. It was found that *At*CAF1b is induced by ROS, JA, SA, mechanical wounding and strongly by ABA signalling. It was also shown that *At*CAF1b contributes to resistance against *P.syringae* (Liang *et al.*, 2009). This would mean that the down-regulation of *At*CAF1b is contributing to the successful disease establishment in TME3 by reducing its ability to regulate stress response mRNAs. Furthermore, this gene is also targeted in T200 as a strategy to suppress defence responses towards SACMV.

This appears to also be the case with *At*HSPRO2, a nematode resistance protein was observed in TME3. *At*HSPRO2 has been shown to contribute to basal resistance against *P.syringae*, which is induced by SA and suppressed by JA/ET signalling (Murray *et al.*, 2007). Due to this, it is unknown if the down-regulation of *At*HSPRO2 is due to its interaction with ET or JA signals at 12 dpi, or whether its' down-regulation is linked to SACMV trying to suppress immunity. However, much like *At*STZ, *At*HSPRO2 is found down-regulated in both landraces at 12 dpi but continues to be down-regulated throughout infection in T200. This would suggest that in T200 it is used as a target to further suppress defence responses, while in TME3 it is more likely to be down-regulated by JA/ET signalling to prioritize pathways conferring tolerance.

Down-regulation of a gene does not necessarily guarantee that it is contributing to susceptibility in TME3. The expression of *At*RRTF1(*At*ERF-109) coincides with literature where it has been suggested that *At*WRKY40 directly negatively regulates it (Pandey *et al*., 2010). So far, *At*RRTF1 has been implicated in regulating redox homeostasis in ROS signalling thereby influencing stress response (Matsuo *et al.*, 2015). The down-regulation of *At*RRTF1 would, therefore, be indicative of suppressing ROS signalling in TME3 at 12 dpi. However, this reduction in ROS may be used to direct and prioritize the network signals to the adequate tolerance pathways rather than hamper tolerance, which is why its' down-regulation is limited to 12 dpi. In comparison, the homologs of *At*RRTF1 are down-regulated at middle and late infection, showing greater negative $\log_2$ fold values. This would suggest this more directed suppression of the *At*RRTF1 homologs in T200 is serving to reduces the ability of T200 to combat SACMV.

To summarize at 12 dpi, T200 and TME3 differentiate from each other at molecular levels despite sharing the same non-systematic phenotype. This shows that tolerance begins to manifest as soon as the virus is detectable within the plant. Tolerance most notably differentiates from susceptibility by showing signs of hormone response processes (particularly in JA, ET and ABA), whereas in T200, there is a widespread signs of hormone suppression and down-regulation of genes that are positively associated with resistance. Furthermore, the combination of up- and down-regulation among hormone responsive genes in TME3 provides evidence for signal prioritization, which suggest that particular defense response networks are required in tolerance rather than an induction of many undirected defense pathways.

2.4.7 Middle

2.4.7.1 T200

Both TME3 and T200 exhibit disease symptoms at 32 dpi, however, TME3 symptoms are milder and virus loads lower (Allie *et al*., 2014) compared to T200. Symptom presentation in T200 clearly manifests, through a large number of down-regulated partners and GO processes, a far greater degree of immune inhibition. Notably, several cassava orthologs of *At*WRKY33, 40, 53, 70 and 41 were all downregulated in T200 but not in TME3 imply their contribution towards the susceptible phenotype. However, when considering the function of the orthologs (Table 2-1), it is clear to see that the down the down-regulation of these key WRKYs substantially reduces T200 ability to respond adequately to SACMV. It is possible to speculate that downstream associated

defence responses associated with these WRKYs are not activated. WRKYs play a crucial role in ABA-mediated signal transduction in plants (Rushton *et al*., 2010, 2012). Wei *et al*. showed in 2016 that *Me*WRKY83 responded to ABA in cassava and showed higher expression under salt abiotic stress. *Me*WRKY83 is an ortholog of *At*WRKY70 and is downregulated in T200 at 12 and 32 dpi. This provides support a possible role for suppression of downstream ABA responses in T200 susceptibility. Possibly due to ABA function as an inducer for some WRKYs. Little literature is available surrounding the role of *At*WRKY41 in pathogen responses. While the *Me*WRKY3 homolog of *At*WRKY41 was only upregulated in TME3 at 32 dpi, in T200 not only was it down-regulated at 32 dpi but also at 12 and 67 dpi (Wei *et al.*, 2016).

Significantly, *Me*WRKY44 (ortholog of *At*WRKY7) was up-regulated throughout T200, while a putative paralog *Me*WRKY58 in TME3 was down-regulated at 67 dpi. *Me*WRKY58 is an ortholog of *At*WRKY7 as it shows high similarity (Figure 2-3), and has been reported to negatively regulate plant defence against bacterial pathogens (Kim *et al.*, 2006a), possibly indicating a role in tolerance. It is interesting to note at early infection (12 dpi) in T200, that a highly complex functional interaction network of *Me*WRKYs and interacting partners based on Arabidopsis orthologs was linked to *At*WRKY7 via the two calmodulin proteins CAM5 and EDA39. Those familiar with proteins in the calmodulin-like family would know that these family of proteins are associated with resistance through interaction with the secondary signalling molecule Ca^{2+} (Park *et al*., 2004). Interestingly, *At*WRKY7 has been found to bind to Ca^{2+}/calmodulin, whereby loss of function wrky7 mutants demonstrated greater resistance towards *P.syrningae* (Kim *et al*., 2006a). This would mean that the up-regulation of *Me*WRKY44 would contribute to susceptibility seen in T200 and the down-regulation of its paralog, *Me*WRKY58, is contributing towards tolerance. Some interacting members are suggested to contribute to priming the plant for susceptibility. Downstream responses, by persistent up-regulation of *Me*WRKY44 throughout infection, may support virus replication and symptoms associated with susceptibility. Conversely, this may provide strong evidence for a central role of *Me*WRKY44 in the suppression of defence in cassava against SACMV. Comparatively, up-regulation of *Me*WRKY58 in TME3 at 67 dpi may play a role in recovery and tolerance. Therefore, it is highly recommended that more focus is placed on studying the *in planta* functions and interactions of *Me*WRKY44,81,3,42, 15, 55, 59 and 83 (orthologs of AtWRKY7, 41, 51, 53 and 70) in cassava in future studies.

One of the known targets in of *At*MPK3 is *At*WRKY33 which shows consistent down-regulation from middle to late infection (Mao *et al*., 2011). *At*WRKY33 undergoes activation through phosphorylation of MPK4, which leads to the dissociation of *At*WRKY33 and the association factor MKS1, thereby allowing *At*WRKY33 to modulate defence gene expression (Qui *et al*., 2008). MPK4 interacts strongly in the Arabidopsis functional network with another mitogen-activated protein kinase, *At*MPK3, which itself appears to be a node or hub for a larger interconnected network comprising many differentially expressed interacting partners. Furthermore, it has been found that *At*WRKY33, through *At*MPK3 signalling induces camalexin biosynthesis, which is a phytoalexin that is found to confer resistance against necrotrophic pathogens (Mao *et al*., 2011). It is observed that camalexin biosynthesis and metabolism are down-regulated in T200 (Supplementary Table 1) and while camalexin does not directly contribute to defence, its pathway is used by the plant in other defence processes (Alves *et al*., 2014) Interestingly, camalexin within *Arabidopsis* has been linked to insect resistance against the green peach aphid, which means that the down-regulation of camalexin in cassava may be a strategy to increase infection by allowing whiteflies to feed unhindered from infected cassava, enabling subsequent infections (Kettles *et al*., 2013). This further demonstrates the role that *At*MPK3 homologs play in suppressing defences an aiding in establishing new infections at later time points.

Appearing alongside *At*WRKY33 is the highly interconnected *At*WRKY40. *At*WRKY40 is rapidly induced by ABA, where it localizes in the cytosol allowing the expression of ABA-responsive genes *At*ABI4 and *At*ABI5, however, if induced by SA, *At*WRKY40 suppresses these ABA genes (Liu *et al*., 2012). Furthermore, it has been demonstrated that *At*WRKY40 is able to modulate the levels of JA and SA during the early response to *G.orontii*. Therefore, the down-regulation experienced at 32 dpi, would lead to a reduced ability in T200 to regulate JA and SA signalling and with it the downstream defence responses. Furthermore, *wrky18 wrky40* double mutants exhibited elevated levels of multiple JAZ family proteins, with *At*JAZ1, 5, 7, 8 and 10 showing particularly high levels of induction. It has been found that these JAZ proteins, which negatively regulate JA signalling, contain the W-box cis-regulatory element, indicating that *At*WRKY40 is able to modulate JA signalling through down-regulating these JAZ proteins (Pandey *et al*., 2010). This is seen in T200 where the homologs of both *At*JAZ1 and 8 do show down-regulation at 32 dpi (not shown). According to literature the down-regulation of these

*At*JAZ proteins would intern lead to JA responsiveness by liberating proteins like *At*MYC2 once undergoing JA dependent degradation (Katsir *et al.*, 2008). However, the expression of hormone-responsive genes in T200 suggests that, despite the down-regulation of these JAZs, JA signalling is being suppressed in T200.

In addition to JA and SA phytohormone down-regulation in T200, *At*ERF (1,5and 6) down-regulation is also present which would lead to the suppression of ET response too. *At*ERF-1, 5 and 6 all belong to the family of AP2/ERF TF family comprising of four sub-divided families, DREB, ERF AP2 and RAV (Sakuma *et al*.,2002). In particular, the ERF group has been found to function synergistically in defence response (Büttner and Singh, 1997). *At*ERF-1A and *At*ERF94 (ORA59) are one of the points to where JA and ET signalling are integrated. *At*ERF5 and *At*ERF6 are interesting as *At*ERF5 has been showed to interact with *At*ERF6, however, redundancy in function relating to JA/ET-mediated defence response has also been shown (Moffat *et al.*, 2012, Son *et al.*, 2012). Interestingly, *At*ERF5 has been implicated in response to chitin, thereby forming part of the innate network in Arabidopsis and together with *At*ERF6 are shown to contribute resistance towards *B.cinerea* (Son *et al*., 2012, Moffat *et al*., 2012). Much like *At*WRKY33, it is suggested that the down-regulation *At*MPK3 is responsible, at least in part, for the down-regulation of the *At*ERF5 and 6 orthologs as MPK3 is known to phosphorylate and activate both these TFs (Meng *et al.*, 2013). Consequently, the down-regulation *At*MPK3 is shown to be further responsible for the suppression of ET responsiveness in T200, and with it the downstream defence pathways.

The suppression of defence responses are not the only transcriptomic changes associated with an invading pathogen. There are also developmental changes required by the plant to facilitate viral replication. An example of this can be seen with AT4G27280 (KRP1), an ABA-induced calcium binding protein whose function has not been fully characterized to date (Li *et al.*, 2017a). What is known, however, is that *At*KRP1 is a cyclin-dependent kinase inhibitor, thereby exerting control over the cell cycle at the S and M phases (Verkest *et al.*, 2005). Furthermore, it was found that its overexpression leads to inhibition of a cell to transit between the M and S phases (Verkest *et al*., 2005). This suggests that its down-regulation in TME3 and T200 would not necessarily contribute to susceptibility but is likely to be hijacked by SAMCV to induce cellular division to enable replication in the nucleus. Indeed, it is well documented that geminiviruses control the cell cycle in order to switch the cell into the active S-phase due to

nuclear requirements for replication (Hanley Bowdoin *et al*., 2013). Functional links between plant signalling hormones, including ethylene and auxins, and cell cycle proteins have been established (Dudits *et al*., Ann Bot 2007), and in SACMV-infected Arabidopsis, it was suggested that auxin activation was partly responsible for the promotion of the S-phase (Pierce and Rey, 2013). In conclusion, the additional down-regulation observed of defence response hub genes such as the orthologs of *At*WRKY33 and 40, most likely by *At*MPK3 homologs are likely contributing to susceptibility. It is suggested that these two hub genes contribute to susceptibility through the suppression of other hormone-responsive genes like *At*ERF-1, 5 and 6, in addition to the continued level of suppression seen by both *At*WRKY53 and 70.

Specifically, with regards to AT1G23710 which shows contrasting expression as a single gene homolog. This is an example of a one-to-many or many-to-many relationship which occurs due to gene duplication events that occurred during or after the divergence of species from their last common ancestor (Gabaldón, 2008). These gene duplications may then lead to one of two functional possibilities, either the occurrence of neofunctionalism or subfunctionalism. Neofunctionalisation leads to duplicated genes having a new function to the original gene, whereas subfunctionalisation occurs when an ancestral gene with two or more functions has these functions partitioned between the resulting paralogs (Rastogi and Liberles, 2005). However, in order for neofunctionalisation to occur, sufficient change is first required in the duplicated gene over the course of evolutionary time. In this specific case, the cassava ancestral gene underwent duplication and diversification which subsequently lead to two separate genes showing homology to AT1G23710. However, the until function studies are performed on these two genes, the specifics of their contrasting expressions remain unknown.

2.4.7.2 TME3

Middle infection marks the point in which T200 and TME3 further differentiate from one other molecularly, despite both showing symptoms. This divergence is particularly evident when observing commonly DEG homologs displaying contrasting expression. The most notable are the *At*WRKY40, HsfA4A, both *At*JAZ1 and 8 and *At*ERF-1. As mentioned previously, it is believed that the orthologs of *At*WRKY40 and the JAZs contribute towards JAZ JA signalling and response. What is now seen at 32 dpi is further support for this as now both *At*MYC2 and JA amido synthetase (JAR1) homologs are now observed to be up-regulated. Both of these genes are associated with JA where MYC2 is known as the master regulator of JA responsive genes, JAR1

is used in converting JA to JA-Ile which elicits the strongest JA response from receptors perceiving it (Kazan and Manners, 2013, Wasternack *et al*., 2013). *At*MYC2, a bHLH transcription factor is suppressed by *At*JAZs and has been shown to modulate JA response and negatively regulate the expression of putative defence genes such as *At*PDF1.2 and *At*ERF-1 (Chini *et al*., 2007, Dombrecht *et al*., 2007). This evidence would suggest that JA signalling is at its greatest during middle infection, and is in part due to the WRKY40-JAZ-MYC2 pathway. This further implicates the role of the orthologs of *At*WRKY40 in establishing tolerance.

Previously during 12 dpi, the down-regulation of ROS responsive genes (CAF1b and RRTF) provided evidence suggesting that ROS was suppressed. However, with the increase seen in JA signalling also comes with it the increase in ROS. The *At*HsfA4A TF, which is uniquely up-regulated at 32-dpi, is currently suspected of being an H_2O_2 sensor in Arabidopsis (Miller and Mittler, 2006). The perception of H_2O_2 leads to transcription induction of stress-responsive genes. The down-regulation of this gene in T200 suggest that the ROS perception and induction pathway is suppressed, whereas in TME3 it is contributing initiate a defence response.

Lastly, it should be noted that in addition to both JA and ROS increases seen at 32 dpi, there is also an increase to ET responses demonstrated at 32 dpi in TME3. Most notably, *At*ERF-1 is demonstrating contrasting expression between both landraces and interacts with *At*WRKY40. Over-expression of ERF-1 leads to up-regulation of pathogen-responsive genes thereby conferring increased resistance (Berrocal-Lobo *et al*., 2002, Pré *et al*., 2008). Additionally, *At*ERF-1 is found to be induced synergistically by ET and JA and is linked to the regulation of many defense-related genes (Lorenzo *et al*., 2002). Hormone synergism is a strategy employed by networks to increase the accuracy of a pathway, consequently, the up-regulation of *At*ERF-1 in TME3 likely suggests that TME3 has begun to prime TME3 against mounting a specific defence response to SACMV and not only an innate or basal response (Berens *et al*., 2017). This further supports the notion that both JA and ET are the dominant hormone in tolerance and are both induced at 32 dpi in TME3. The up-regulation of *At*ERF-1 in confer resistance in two manners, firstly as a TF *At*ERF-1 the induction of ET responsive defence genes that combat SACMV, and conferring resistance against insect pests (such as the whitefly) thereby lowering the subsequent infections and transmission. In T200, it is likely that both these defence mechanisms conferred by *At*ERF-1 would be lost upon its' down-regulation which would contribute to susceptibility.

In addition to these common genes showing contrasting expression between the two landraces, it is also important to note genes that are uniquely DE in TME3 at 12 dpi. *Me*WRKY3 (the ortholog of *At*WRKY41) in both TME3 and T200 does not show any interactions in any of the constructed networks, however, this is more likely due to the confidence scores surrounding the interactions falling below the cut-off, necessitating focus being placed on *At*WRKY41 in future studies. Interestingly, *At*WRKY41 has been demonstrated to respond to *P.syrnigae* in Arabidopsis (Higashi *et al*., 2008). Furthermore, the up-regulation of *At*WRKY41 lead to the up-regulation of the *PR5* gene but suppressed *PDF1.2*, a defensin that is induced by jasmonate. This specific interaction between *At*WRKY41 and its partners lead the authors to suggest that it is involved with modulating the cross-talk between JA and SA pathways. Much like *At*WRKY40, *At*WRKY41 is induced by ABA (Liu *et al*., 2015), which would explain why the orthologs of these two WRKYs experience up-regulation uniquely in TME3 only. This coupled with the contrasting nature of expression between the landraces leads to the belief that *At*WRKY41 is playing a part in pathogen response and It is recommended that focus is placed on this protein to further characterize its function and pathways.

*At*WRKY6 and 28 (homolog of *Me*WRKY17 and 81 respectively) are the only two unique differentially expressed WRKY TFs in TME3 at 32 dpi. Although there is no link between *At*WRKY6 and hormone signalling, it has been showed to have both functions in both biotic and abiotic signalling (Robatzek and Somssich, 2001, Li *et al.*, 2017b). In addition, *At*WRKY6 is found to be induced by ROS, SA, JA, and ET (Pieterse and Van Loon, 1999, Scarpeci *et al.*, 2008). Therefore, it is likely *At*WRKY6 was induced in TME3 specifically linked to events linked to priming for recovery. It is likely that *AtWRKY6* is playing a role in the expression of specific defence genes rather than aiding in the induction of a hormone signalling response. Not much work has been done on *At*WRKY6 and elucidation of downstream targets may prove useful to the identification of novel defence genes associated with recovery and tolerance.

From these observations surrounding T200 and TME3, it can be reasoned that while the differences were starting to form during early infection, it during middle infection that these differences become more apparent. T200 shows large scale down-regulation among defense response genes, which further enables its downward spiral into persisting infection. This down-regulation appears to pivot on a few key hub WRKY orthologs (*At*WRKY33,40 and 53) which

suggests that these hubs form targets of SACMV to create a greater amount of perturbations in T200. In contrast in TME3, the opposite is seen where the up-regulation of many defense responsive genes seem to pivot on the up-regulation of only hub WRKY, namely the *At*WRKY40 orthologs. The down-stream effect of the induction surrounding the *At*WRKY40 orthologs is apparent in induced hormone response which shows signs of pathway synergism, adding additional nuance to the tolerance defense response in TME3. Whereas in T200, the consequences of this comprehensive down-regulation of defense genes is further apparent in the increased suppression seen amongst hormone signalling pathways.

2.4.8 Late

2.4.8.1 T200

While early and infection reveals insights into the mechanisms leading up to the successful establishment of disease or tolerance, late infection provides insight into the fate of both landraces infected with SACMV. In T200, the overall topology of the functional WRKY-associated network and the hormone responsive genes demonstrate small changes between middle and late infection, and the majority of TFs remain differentially expressed. The only two unique genes found in the late infection network and not the middle infection is the uncharacterized AT1G23710 and sigma factor binding protein 1 (SIB1). Notably, the single Arabidopsis homolog AT1G23710 corresponds to two different cassava DEGs that show up- and down-regulation. *At*SIB1 is induced by SA and is observed to play a role in disease resistance against *P.syringae (*Narusaka *et al*., 2008, Xie *et al*., 2010). The down-regulation of *At*SIG1 would further hamper the efforts of T200 to attempt to mount a defence response during late infection. Although, there is a reduction in the number of DEGs at 67 dpi, the very little change in the interacting partners around the DE WRKYs implies that there is a stability in the suppression of host defences by SACMV. This is predicted to remain past 67 dpi, whereby T200 will never be able to initiate a defence response against SACMV which manifests as a chronic infection.

2.4.8.2 TME3

During recovery and tolerance, strong evidence for a role of *At*WRKY40 and JAZs in tolerance is provided by the persistent up-regulation of *At*WRKY40 orthologs in CMD-tolerant cassava TME3. Phylogenetic analysis (Figure 2-4) shows that *Me*WRKY10 and 79 are more closely related to *At*WRKY40, and therefore suggests these could have similar functions. *Me*WRKY10

is in particular noteworthy as out of the four *Me*WRKY40 orthologs (10,11,40,79) as it is the only one up-regulated at both 32 and 67 dpi. *Me*WRKY10 is worthy of further investigation as *At*WRKY40 has been reported to play roles in positively regulation ETI, and is a modulator of JA signalling, the most dominant of hormones in TME3 (Pandey *et al*., 2010, Liu *et al*., 2012). The other WRKY of particular interest is *At*WRKY28 (*Me*WRKY81) as it is uniquely down-regulated in TME3 and only at 67 dpi. *At*WRKY28 is rapidly and highly induced after oxidative stress (Scarpeci et. al. 2008) and was found to positively regulate oxidative stress (Babitha *et al.*, 2013). This down-regulation suggests that the respiratory burst experienced with ROS signalling is localized to middle infection in TME3, as persistent ROS signalling would result in oxidative damage to the cell.

Very little comment can be made regarding 67 dpi in T200 and TME3, but for different reasons. Entering late infection, there is very little change observed in T200 which suggests that SACMV has entered a stable relationship with T200 whereby the immune system is suppressed. This stable immune suppression likely results in persistent infection seen in T200 and not in TME3. Not much can be said for TME3 during 67 dpi as the resulting network and hormone processes are considerably reduced compared to other time points. What this suggests is that while T200 is maintained in a diseased state by the stable SACMV infection, TME3 is approaching its pre-infected state. However, the maintained induction of JA and JAZ8 would further support the hypothesis that JA in principally involved in tolerance and is required by TME3 to return to a non-symptomatic state. Lastly, the return of TME3 to a non-symptomatic state seen in 67 dpi further supports the hypothesis that tolerance occurs at middle infection and recovery is the product of these systems induced around 32dpi.

2.4.9 WRKYs in literature

Currently, there are other global transcriptomic studies available profiling the response in resistant/tolerant and susceptible plants towards ipomoviruses and geminiviruses (Table 2-1). The provirus studies involving both *Cassava brown streak virus* (CBSV) and Ugandan cassava brown streak virus (UCBSV) enables the comparison of a single host (cassava) multiple pathogens. Whereas, the geminivirus studies demonstrated the use in a large range of viruses, such as SACMV, TYLCV, Cabbage leaf curl virus (CalCuV), Tomato yellow leaf curl Sardina virus (TYLCSV) and Tomato leaf curl New Delhi virus (TolCNDV). These studies demonstrate the different responses elicited in different plants towards phylogenetically related viruses.

Table 2-2: Summary of other available global transcriptomic studies focusing on Geminivirus and Ipomovirus interactions in a susceptible (Green) or tolerant (Blue) plant species:

Family	Virus	Species	Landrace	Time point (dpi)	Stage of infection	Reference
Geminivirus	SACMV	*Manihot esculenta*	T200	12,36,67	Early/Mid/Late	(Allie *et al.*, 2014)
Geminivirus	SACMV	*Arabidopsis thaliana*	Ecotype Columbia-0	14,24,36	Early/Mid/Late	(Pierce and Rey, 2013)
Geminivirus	TYLCV	*Solanum lycopersicum*	TMXA48-4-0	3,5,7	Early	(Chen *et al.*, 2013)
Geminivirus	CalCuV	*Arabidopsis Thaliana*	Ecotype Columbia-0	12	Late	(Ascencio-Ibáñez *et al.*, 2008)
Geminivirus	TYLCSV	*Solanum lycopersicum*	MoneyMaker	42	Late	(Miozzi *et al.*, 2014)
Ipomovirus	CBSV	*Manihot esculenta*	cv60444	28	Early	(Anjanappa *et al.*, 2017)
Ipomovirus	CBSV	*Manihot esculenta*	Albert	365	Late	(Maruthi *et al.*, 2014)
Ipomovirus	UCBSV	*Manihot esculenta*	Albert	6[1],1,2,5,8,45,5	Early/Late	(Amuge *et al.*, 2017)
Geminivirus	TolCNDV	*Solanum tuberosum*	Kufri Pukhraj	15	Late	(Jeevalatha *et al.*, 2017)
Geminivirus	SACMV	*Manihot esculenta*	TME3	12,36,67	Early/Mid/Late	(Allie *et al.*, 2014)
Geminivirus	TYLCV	*Solanum lycopersicum*	CLN2777A	3,5,7	Early	(Chen *et al.*, 2013)
Ipomovirus	CBSV	*Manihot*	Kaleso	365	Late	(Anjanappa

		esculenta				*et al.*, 2017)
Ipomovirus	CBSV	*Manihot esculenta*	KBH2006/18	28	Early	(Maruthi *et al.*, 2014)
Ipomovirus	UCBSV	*Manihot esculenta*	Namikonga	6[1],1,2,5,8,45,54	Early/Late	(Amuge *et al.*, 2017)
Geminivirus	TolCNDV	*Solanum tuberosum*	Kufri Bahar	15	Late	(Jeevalatha *et al.*, 2017)

If these proposed candidate genes were to be associated with tolerance or susceptibility, then it is expected that these genes would show similar expression profiles in other susceptible or tolerant/resistant plants. However, it has been shown that a single pathogen in two different hosts shows a large difference in response by the host(Pierce and Rey, 2013, Allie *et al.*, 2014). Furthermore ,it has been shown that a single host will elicit a varied response towards different viruses, even those that show a close phylogenetic relationship(Rodrigo *et al.*, 2012). This means that there is no single defence gene responsible for mounting a successful defence response to all viruses in all plant species. This variability would none of the proposed candidate genes will be found DE in all the respective (susceptible or resistant/ tolerant) hosts.

However, despite this inherent variability, there is still considerable overlap with the WRKY results and these other transcriptomic studies. Notably, *At*WRKY70, which was down-regulated in T200 and linked to susceptibility by suppressing hormone responses was found to be down-regulated seven out of the nine instances of identification among the other studies (up-regulated against CBSV and CalCuV). This would further support the hypothesis that the orthologs of *At*WRKY70 positively regulates resistance and upon down-regulation leads to enhanced susceptibility in the hosts. *At*WRKY40 was identified the most following *At*WRKY70, however, it was found to be equally up-regulated as it was down-regulated (four up and down) in datasets. However, the observation that its down-regulated in response to SACMV in cassava, TYLCV and TYLCSV implies that targeting of *At*WRKY40 is a strategy recently evolved in the Geminiviridae family and is confined to these three phylogenetically related viruses. *At*WRKY53 was noted in five instances and was only up-regulated to CalCuV, and much like *At*WRKY70,

[1] The first measurement taken by Amuge *et al*. (2017) was taken 6 hours post infection.

further supports the results suggesting that *At*WRKY53 confers susceptibility when down-regulated. Furthermore, the expressions (mostly down-regulation) seen for *At*WRKY033 and 41 in the other transcriptomic studies coincided what was found within the Allie *et al*., (2014) data set. Lastly, AtWRKY07 was found to be exclusively upregulated towards SACMV and CBSV in cassava, potentially implicating AtWRKY07 is a cassava specific susceptibility factor.

The resistant/tolerant dataset show fewer numbers of DE in WRKYs, possibly due to the lower transcriptional perturbations observed in resistant landraces, much like what is observed in TME3. *At*WRKY40 was the only notable WRKY among those identified in the datasets. The *At*WRKY40 orthologs were found all resistant cassava datasets, and only down-regulated in response to UCBSV. This would support the proposition that the up-regulation of *At*WRKY40 orthologs confers tolerance/resistance in cassava.

The frequency of the identified WRKY homologs in susceptible compared to resistant/tolerant dataset reveals something interesting. It appears that in susceptible plants, despite show greater transcriptional perturbation in response to invading pathogens demonstrates a several commonly DE *At*WRKY orthologs. Whereas, in tolerant/resistant datasets show the fewer number and no discernable pattern forming. This would suggest that there is a high degree of diversity present in the mechanisms that govern tolerance in these plants. In contrast, the susceptible data suggest that these viruses may be demonstrating similar strategies by exploiting similar proteins to establish successful viral infection.

The particular study performed by Pierce *et al*. offers an interesting comparison between defense responses in two different plants to a single pathogen: namely SACMV (Pierce and Rey, 2013). In this study gene expression was measured at early (14 dpi), middle (24 dpi) and late (36 dpi) in *A. thaliana* infected with SACMV using a microarray. Since *A. thaliana* demonstrates susceptibility towards SACMV enabling comparison with T200. When comparing the expression of the DE Arabidopsis WRKYs orthologs in cassava, it is interesting to see very little overlap occurring. Only three of the DE WRKYs orthologs (*At*WRKY12,22 and 51) in Arabidopsis show DE in T200, however, in each instance the DEGs show contrasting expression (up-regulated in one study and down-regulated in the other) and never at a stage of infection. This provides evidence that the transcriptional changes that occur in a host upon pathogen invasion is dependent both on the host and the pathogen in question. For this reasons it can be reasoned that

the resulting defense pathways initiated by *A. thaliana* would differ to T200 and TME3 which is seen in the differences in DE WRKY TFs.

However, if one would the phylogenetic relationship between the DE WRKYs in *A. thaliana* and T200, an interesting trend begins to emerge. The phylogenetic tree (Figure 2-3) shows that *At*WRKY50 and WRKY51 lie close topologically to *Me*WRKY22 De in T200. Furthermore, all these three WRKYs all display up-regulation, albeit only *At*WRKY50 and 51 were up-regulated at a different infection stage than *Me*WRKY22. This trend is also present between *At*WRKY12, 28 and *Me*WRKY14,20 and 72 which all show down-regulation (at varied stages of infection) and lie topologically closer to each other. This means that closer evolutionary linked WRKYs show similar patterns of DE in two different host when challenged by a single pathogen. This suggests that SACMV has evolved to target specific proteins in order to establish infection and that evolutionary similar proteins found in other hosts are also susceptible to form interactions with SACMV. This may be partially why plant viruses have such a wide host range.

2.5 Conclusions

Distinct differences in expression profiles of WRKY transcription factors and functional network interacting partners were demonstrated in this study which was observed from 12 dpi. These differences contribute, amongst many other molecular mechanisms, to the susceptibility and tolerance to one of the cassava mosaic disease-associated begomoviruses, namely *South African cassava mosaic virus*. The first noticeable difference was observed in the hormone signalling between the landrace. In T200 suppression of ABA, ET, JA and SA were evident throughout SACMV. The three *At*WRKYs orthologs (41, 53 and 70) that were consistently suppressed throughout infection, therefore it can be suggested that these are critical transcription factors in inducing and maintaining disease symptoms in T200 across the course of infection. Specifically, much of the suppression seen in hormonal responses associated with susceptibility seems to centre around several down-regulated WRKYs, namely *Me*WRKY10,11,15,55,59,68,79 which are the orthologs of *At*WRKY33,40,53 and 70. Notably, it is *Me*WRKY55 and 59 seem to be responsible for the initial suppression of both JA and SA signalling early in infection, whereas, *Me*WRKY10,11,68 and 79 appear to be responsible for the comprehensive hormone suppression in T200 at middle and late infection stages. One of the key proteins involved in this suppression, through interactions with the orthologs of *At*WRKY33, *At*ERF-1, 5 and 6 is the homolog of

*At*MPK3. This would lead to a severe reduction in the ability of T200 to respond and mount a defence response at each of the three time points.

While most virus-responsive immune responses trigger SA and the hypersensitive response. SA is induced at low levels in TME3 at 12 and 32 dpi, but is notably absent at 67 dpi, and supports, along with other transcriptome evidence (Allie *et al*., 2014), the proposal that SA is not key in defence in the non-domesticated perennial crop cassava. SA is likely to play a regulatory cross-talk role in hormone signalling in T200 and TME3, but the lack of a SA-associated hypersensitive response in TME3 and other cassava landraces, suggests other hormones are more important in susceptibility and tolerance/resistance to CMD. This then leads to the suggestion that JA was the dominant hormone which was then modulated through ET. It is hypothesized that the up-regulation orthologs of *At*WRKY40, were likely induced by ABA signalling which in turn regulates down-stream JA and ET processes through JAZ and ERF proteins. This implicates *Me*WRKY10,11,40 and 70 as putative proteins involved in recovery and tolerance in TME3.

Through *Me*WRKY10,11 and 79, it is proposed that the response to JA and ET is further developed at 32 dpi leading to the induction of the defence TFs *Me*WRKY17 (*At*WRKY6) and the *Me*ERF-100 (*At*ERF). Additionally, at 67 dpi when tolerance is established it is the uniquely down-regulated *Me*WRKY81 (*At*WRKY28) which may play an additional supporting role, but whose function at this point is not known. Taken together, the results from TME3 suggest that recovery is initiated during the systemic infection stage (32 dpi) in TME3 and perpetuated as a tolerance phenotype at 67 pi and beyond.

WRKYs and hormones are a small component of the complex and multi-layered nature of molecular mechanisms which govern cassava tolerance or susceptibility to SACMV. Further research is required to identify what potential role *Me*WRKY3 (ortholog of *At*WRKY41) plays in tolerance, and additionally, further attention is needed to determine the functions of the cassava WRKYs whose functions are not known in this crop, or even other plants studied. Moreover, elucidation of the interacting partners in the functional network and identification of master hubs or nodes could prove invaluable in strategic approaches to engineering natural resistance.

Chapter 3

Protein-Protein Interaction Networks cantered around five families of important pathogen-responsive transcription factors in cassava T200 and TME3 landraces

3.1 Introduction

Previously, *Manihot glaziovii* resistance towards CMD was bred into cassava landraces through breeding (Ntawuruhunga *et al.*, 2006, Dixon *et al.*, 2008, Nweke, 2009, Dixon and Hopkins, 2010). Cassava landraces or landraces that demonstrate resistance to CMD have the CMD2 locus (Akano *et al*., 2002) on chromosome 12. Cassava TME3 is currently proving challenging to sequence due to its repetitive nature(Prochnik *et al.*, 2012a) .A genome-wide association study of 6128 resistant cassava landraces revealed 3 putative genes: two peroxidases and a thioredoxin associated with the CMD2 locus (Wolfe *et al.*, 2016). TME3 is a cassava landrace that has the CMD2 locus, however, displays tolerance towards SACMV and CMD-associated geminiviruses (CMGs), such as SACMV rather than resistance (Bengyella *et al.*, 2015), due to TME3 displaying reduced symptoms when infected with SACMV, and also demonstrating recovery of symptoms during late infection (Allie *et al.*, 2014). Recovery presents as newly formed apical leaves showing continuously reduced symptoms. Despite the identification of the CMD2 locus, and other contributing alleles associated with defence responses to CMGs, the precise mechanisms governing tolerance to SACMV in TME3 is not well understood.

Although the molecular mechanisms surrounding tolerance have yet to be defined, there is literature surrounding the characteristics of tolerance. Firstly, it should be stated that tolerance is associated with an initial detectable viral load which is lower compared to susceptible hosts (Gasura and Mukasa, 2010, Allie *et al.*, 2014). At recovery, TME3 displays milder to no leaf symptoms compared to susceptible T200 landrace, and low virus persistence associated with tolerance (Bengyella *et al.*, 2015). Secondly, tolerant cassava hosts show lower transcriptional changes in response to SACMV , which may explain why tolerant hosts display less severe symptoms compared to a susceptible host (Allie *et al.*, 2014). Tolerance may be regarded as a balance between virus and host defence and counter-defense mechanisms hallmarked by lower host responses and is energy costly to the plant (Gaudel and Sanfaçon, 2018). Additionally, RNAi has been linked to recovery, which is seen in *Nicotiana benthamiana* infected with African cassava mosaic virus (Chellappan *et al.*, 2004) and cassava TME3 infected with

SACMV (Rogans and Rey, 2016). The characteristics of recovery seen in all new apical leaves would suggest the occurrence of systemic signalling. To date, there are two forms of systemic signalling found in plants, namely SAR and ISR which overlap in terms of responses, but also each with their own distinct characteristics and deterministic mechanisms.

Systemic acquired resistance is the most common systemic immune defence response to pathogens and is one of the consequences of the HR, a specialized form of PCD (Ross, 1961). SAR is characterized by the enhanced resistance experienced in a plant following a localized microbial infection, centring around the enhanced expression of PR proteins and salicylic acid (SA) signalling (Fu and Dong, 2013). SAR is directed towards a broad range of pathogens but is shown to be particularly effective against hemi/biotrophic pathogens (Sticher *et al.*, 1997). SA and pipecolic acid are the two molecules that induce SAR and serve as the systemic signals, notably SA which leads to the downstream expression of NON-EXPRESSOR PROTEIN 1 (NPR1) which is considered the master regulator in SAR (Návarová *et al.*, 2012, Wu *et al.*, 2012, Fu and Dong, 2013). SAR leads to both the increase of systemic resistance against the invading pathogen, in addition to priming the plant against the invasion of the same pathogen (Návarová *et al.*, 2012). The other systemic response found in plants is known as induced systemic immunity. ISR to pathogens is induced in plants that are colonized by beneficial rhizobacteria (Van Loon *et al.*, 1998). ISR exhibits contrasting characteristics which requires its recognition as being different to SAR. Firstly, unlike SAR, ISR does require beneficial rhizobacteria, however plants treated with crude extracts of rhizobacteria also exhibit ISR induction suggesting that ISR is induced by bacterial elicitor molecules or compounds (Van Peer and Schippers, 1992, Kagale *et al.*, 2011). Furthermore, ISR is not associated with the accumulation of PR proteins in infected tissues as much as SAR, and most notably involves JA E) signalling (Pieterse, 1998, Van Loon *et al.*, 1998, Hammerschmidt, 1999). Similarly to SAR, ISR has been shown to induce ‘priming’ in plants whereby PR protein expression is seen in uninfected cells post-localized infection (Heil and Bostock, 2002). The determination of whichever strategy of systemic resistance is initiated in SACMV-infected TME3 would be the first step to understanding what potential mechanisms occur that lead to recovery and tolerance. However, understanding systemic signalling further requires knowledge of interconnected transcription factors (TFs) as they control both systemic responses as well as expressing downstream genes responding to the systemic signals.

In plants, TFs are responsible for initiating the transcriptional changes required to respond to an external stimulus. Currently, there is substantial literature available surrounding TFs involved with defence responses. The TF families that appear most frequently regarding biotic stress/pathogen response are the bZIP, ERF, MYB, WRKY and the NAC TF families (Table 3-1). Recently, all these TFs have been characterized and annotated in cassava with links to abiotic stress responses (Hu *et al.*, 2015, Fan *et al.*, 2016, Hu *et al.*, 2016, Wei *et al.*, 2016, Ruan *et al.*, 2017a). This provides the foundation to study these TF families in SACMV challenged cassava in order to determine if they play a role in tolerance or susceptibility.

Table 3-1:Characteristics of TFs associated with biotic stress

Family	Identified in cassava	Conserved domain	Binds to	Reference
bZIP	77	Basic region and leucine zipper region	G-box	(Hu *et al.*, 2016)
ERF	147	58-59 aa AP2 domain	GCC-box/ DRE region	(Fan *et al.*, 2016)
MYB	299	50 -53 aa MYB domain	AC1 element	(Ruan *et al.*, 2017b)
NAC	96	150 aa NAC domain	CATGTG motif	(Tran *et al.*, 2004, Hu *et al.*, 2015)
WRKY	85	WRKYGQK domain	W-box	(Wei *et al.*, 2016)

In 2014, Allie *et al*. characterized the global transcriptional changes occurring in both the tolerant TME3 and the susceptible T200 landraces during early (12 dpi), middle (32 dpi) and late (67 dpi) SACMV infection. This study produced detailed and validated data on how a susceptible and tolerant cassava system responds to SACMV, one of the CMGs causing CMD. From this data, the authors made broad observations of each landrace in regard to SACMV-responses, and were able to provide evidence supporting the paradigm that viruses are able to manipulate their molecular interaction with host plant so that they can adapt to the host genetic background (Allie *et al*., 2014). However, due to the large nature of the datasets, the individual nuances surrounding the involvement of the five TF families of interest were not adequately

addressed or discussed. This offers the opportunity to revaluate, interpret and compare the datasets, in T200 and TME3, with focus placed specifically on the five TF families of interest.

One of the most important concepts to understand is that genes and their products do not operate and function in isolation. They form complex networks and feedback loops that signal and cross-talk between each other. This notion led to the creation of systems biology to better understand and study the effect of genes in a pathway. Systems biology entails the creation of biological networks, and in doing so, gives context to the higher dimensional data created from transcriptomic studies like that generated by Allie *et al*. However, the pre-requisite to creating these systems is adequate knowledge surrounding the networks components (genes or proteins). While many cassava genomes have been sequenced, only one genome, namely AM560-2, is publically available on phytozome v.11. Additionally, due to cassava's non-model plant status, information on its genome is considerably lacking, which requires the use of alternative model plants like *Arabidopsis thaliana* to glean gene function. Using gene homology, phylogeny and the guilt by association principle, it is possible to infer functional annotation onto unannotated cassava genes thereby enabling their placement into network systems (Oliver, 2000).

Host immunity is driven by a complicated network of multiple TF families, that regulate host responses and initiate down-stream effects. It is plausible that these networks are, therefore, involved in the development of tolerance in TME3. In order to further investigate this, differentially expressed genes (DEGs) belonging to the bZIP, ERF, MYB, NAC and WRKY families were placed into interaction networks within the context of SACMV-challenged cassava. In this way, further insight into the role of these TFs in the underlying signalling mechanisms priming tolerance of TME3 to SACMV were gained.

3.2 Material and Methods

3.2.1 Identification of differentially expressed TFs in cassava and their Arabidopsis homologs

In order to determine the global transcription factor expression in T200 and TME3 requires first the identification of DEGs. Data derived from Allie *et al*. (2014) was used to perform differential gene expression analysis. Genes were considered to be significantly DE below a p-value threshold of $\alpha = 0.05$ and if they demonstrated a $\log_2$ fold change in expression greater than |1.5|. Several previous studies have previously identified putative bZIP(Hu *et al.*, 2015), ERF(Fan *et*

al., 2016), MYB(Ruan *et al.*, 2017b), NAC(Wang *et al.*, 2017) and WRKY(Wei *et al.*, 2016). TFs in cassava

This allowed these genes to be extracted from the list of DEGs using their Phytozome (v. 11) accession number (Cassava v4.1). These were then cross-referenced with their Arabidopsis homologs using the BioMart tool in Phytozome (Smedley *et al.*, 2009), in order to annotate them with their corresponding TAIR accession number (Release 12). As a second form of quality control, the Arabidopsis homologs corresponding to these TAIR accession numbers were confirmed to be similarly classified as belonging to the appropriate TF family by confirming the presence of the relevant Pfam domains. This was done by assigning PFAM accessions (Pfam 31.0) to the list of TAIR IDs using BioMart in the Ensembl Plant database (release 41). Only genes containing the conserved domain associated with each TF family (bZIP:PF00170, ERF:00847, MYB:PF00249, NAC:PF01849, WRKY: PF03106), were retained.

3.2.2 Gene Ontology enrichment

In order to evaluate which biological processes are enriched between up- and down-regulated gene lists GO enrichment was performed using the clusterprofiler package (v2.7) in R (v3.3.1) (Yu *et al*., 2012). First, genes were separated based on their expression into two separate lists: up-regulated genes and down-regulated genes for each landrace per time point. This was done in order to determine with greater specificity what processes are being up- or down-regulated in response to SACMV. Separately, each of the gene lists was compared to the background list of genes which comprised of all the genes in *A. thaliana* acquired the plant Ensembl database (TAIR 10). GO terms were deemed enriched if they had a p-value < 0.05 and a q-value < 0.01 when the Benjamini-Hochberg correction was applied.

3.2.3 Pathway analysis of protein-protein interactions surrounding the identified TFs

In order to construct a predicted network surrounding each TF, the first step requires the identification of interaction occurring between the TF of interest and other proteins. All TFs identified in each landrace, at each time point, were collectively queried against the STRING (v10.1) database using their TAIR IDs. The confidence threshold was set between median and high (0.6, Supplementary Methods), to only include relatively high confidence interactions. The second shell of interactions, which comprises of the proteins that are known to interact with the query list of proteins, was set to a maximum of 500. The disconnected nodes that remained in each of the networks were removed.

This information was used, together with the results obtained from GO enrichment, to construct predicted protein-protein interaction networks surrounding the identified TFs. In each landrace, at each time point, the list of interactions was filtered to include only interactions between genes identified as being DE.

The genes contained within GO-enriched biological processes were then annotated onto the network, as well as those genes significantly up- or down-regulated at each time point, for each host. This was performed using Cytoscape (v3.5.1).

3.3 Results

3.3.1 Global transcription factor expression in T200 and TME3 infected with SACMV

In total, 110 TFs were found to be differentially expressed in T200 at all three time points post SACMV infection, whereas there are only 49 DE TFs were identified in TME3 (Supplementary Table 2). Furthermore, amongst the identified TFs, time-specific changes in expression are seen throughout infection in both T200 and TME3, and the overall pattern is significantly different between the two landraces (Figure 3-1).

Among the TF families in T200, ERFs are the family with the highest number of cumulative down-regulated TFs over all three time points. Whereas MYB demonstrates the greatest diversity in DE TF with 56 different MYB being identified through SACMV infection.

At 12 dpi, the highest number of DEGs falls under the MYB TF family and with ERF and WRKY make up the majority of TFs at 12 dpi. Notably, both the bZIP and NAC families are the only ones to display only up- and down-regulation respectively. A noticeable trend can be seen at 32 dpi, is an increase in the number of down-regulated TFs is observed across all families. Furthermore, in all cases except for the NAC TF family, the number of down-regulated TFs in T200 is always at its highest during middle infection. Furthermore, the bZIP TFs are the only family to demonstrate more up-regulated than down-regulated DEGs. Moving into 67 dpi in T200, the general pattern is that there is a general decrease in the number of DEGs in each TF familiy, but remain relatively high when compared to TME3. There is a step wise increase in up-regulated DEGs in MYB throughout SACMV infection, while the opposite is true for NAC where there is a step wise increase in the number of down-regulated genes.

TME3 does not display a similar general pattern as T200 whereby the number of DE TFs peak at 32 dpi. Notably, comparatively fewer TFs from the five families are differentially expressed in TME3 as in T200 at all stages of infection (Figures 3-1 and 3-2A). Notably, differential expression of bZIP TFs is largely absent throughout SACMV infection, with only two down-regulated genes at 32 dpi. Overall, the most pronounced response in TME3 is up-regulation of ERFs, MYBs, NACs and WRKYs at 32 dpi when symptoms are showing and virus replication is detected, which is completely the opposite in T200 (mostly down-regaulated). There is a noticeable switch occurring in the ERF family between early and middle infection where the majority of DE ERFs are down-regulated at 12 dpi, whereas at 32 dpi there are only up-regulated ERF DEGs. Interestingly, the MYBs demonstrate a homeostatic relationship at 12 and 32 dpi where their number of up- and down-regulated genes are equal. NAC, similarly to

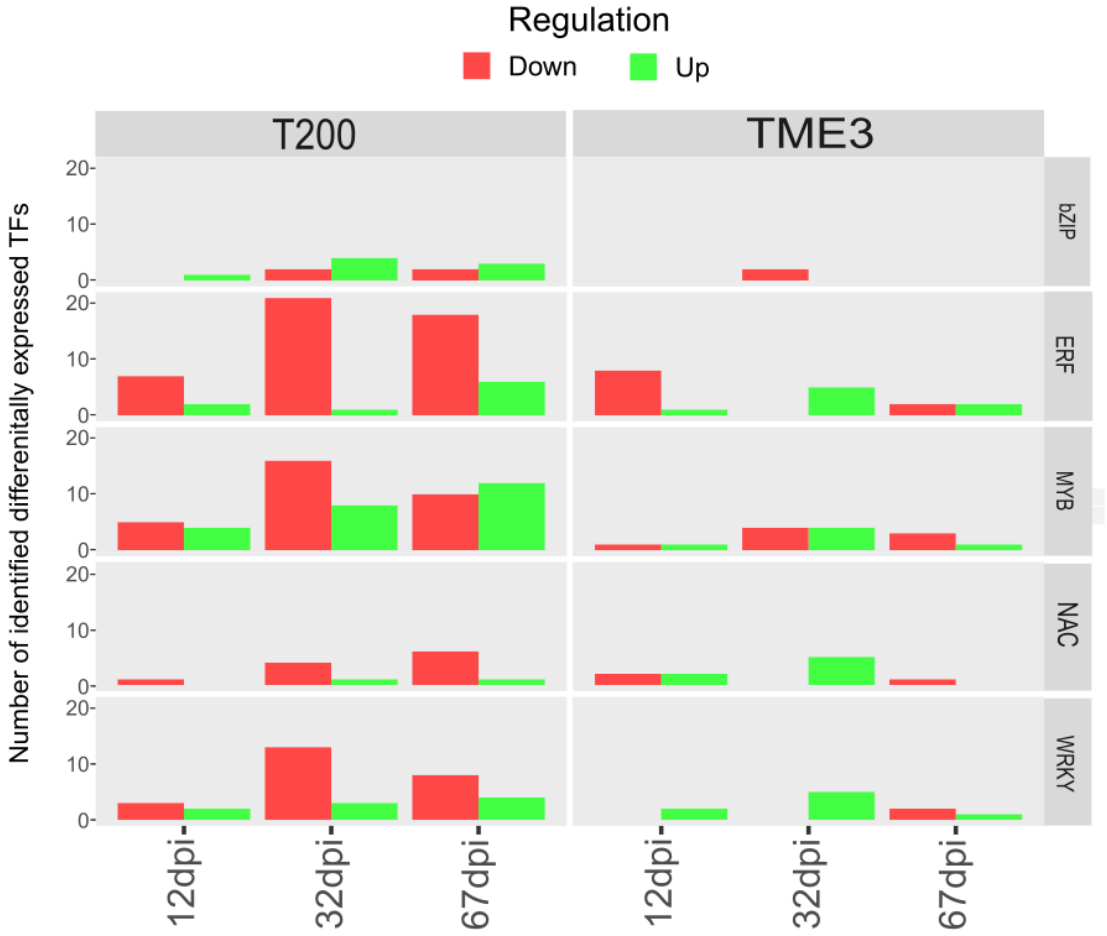


Figure 3-1:The global representation of the number of DE TFs belonging to the bZIP, ERF, MYB, NAC and WRKY families in T200 and TME3 during the time course of SACMV infection. Green and red represent up- and down-regulation, respectively.

ERF, also demonstrates a switch where an equal number of up- and down-regulated genes at 12 dpi later transitions into 32 dpi with only up-regulated NAC TFs. Only four families of WRKYs in TME3 show up-regulation at early, middle and late infection (see chapter 2). Diversity and commonalities amongst TF expression

The number of identified TFs expressed in each family at each time point (Figure 3-2A) does not necessarily reflect the overall levels of diversity. Interestingly, the TF families that show the greatest diversity differ between the susceptible and tolerant landraces. The MYB TF family has 41 (37.27%) different DE MYBs throughout infection in T200 but far fewer in TME3 with only 13 (26.53%). In contrast in TME3, ERFs showed the greats number of DE TFs with a total of 17 (34.69%), whereas in T200, ERFs were the second most diverse of the TFs with 37 (33.63). It should be noted that the ERF family shows the greatest degree of similarity compared to the other four TF families, with 12 commonly DE between T200 and TME3. Expectedly, due to their low numbers, the bZIP TFs showed the lowest overall diversity in both landraces, with only eight and two found in T200 and TME3 respectively, throughout the late infection period (67 dpi). In order to quantify the differences between T200 and TME3, it is important to acknowledge the similarities shared by both landraces. In order to do so, the expression profiles of the shared DEGs from the five families were placed under scrutiny (Figure 3-2B). At 12 dpi, three TFs, namely *Me*NAC80, *Me*WRKY45 and *Me*WRKY132, were DE in both T200 and TME3. Interesting, all three of these DEGs showed down-regulation in early response to SACMV in T200 and TME3, even though no symptoms were visible. In contrast, at 32 dpi, a considerably higher number of commonly DE genes was observed in both landraces, however, all but one of these TFs (*Me*MYB028) were exclusively up-regulated in TME3 and down-regulated in T200. At recovery (67 dpi) in TME3, but continued infection in T200, only a single common down-regulated TF (*Me*ERF30) was observed.

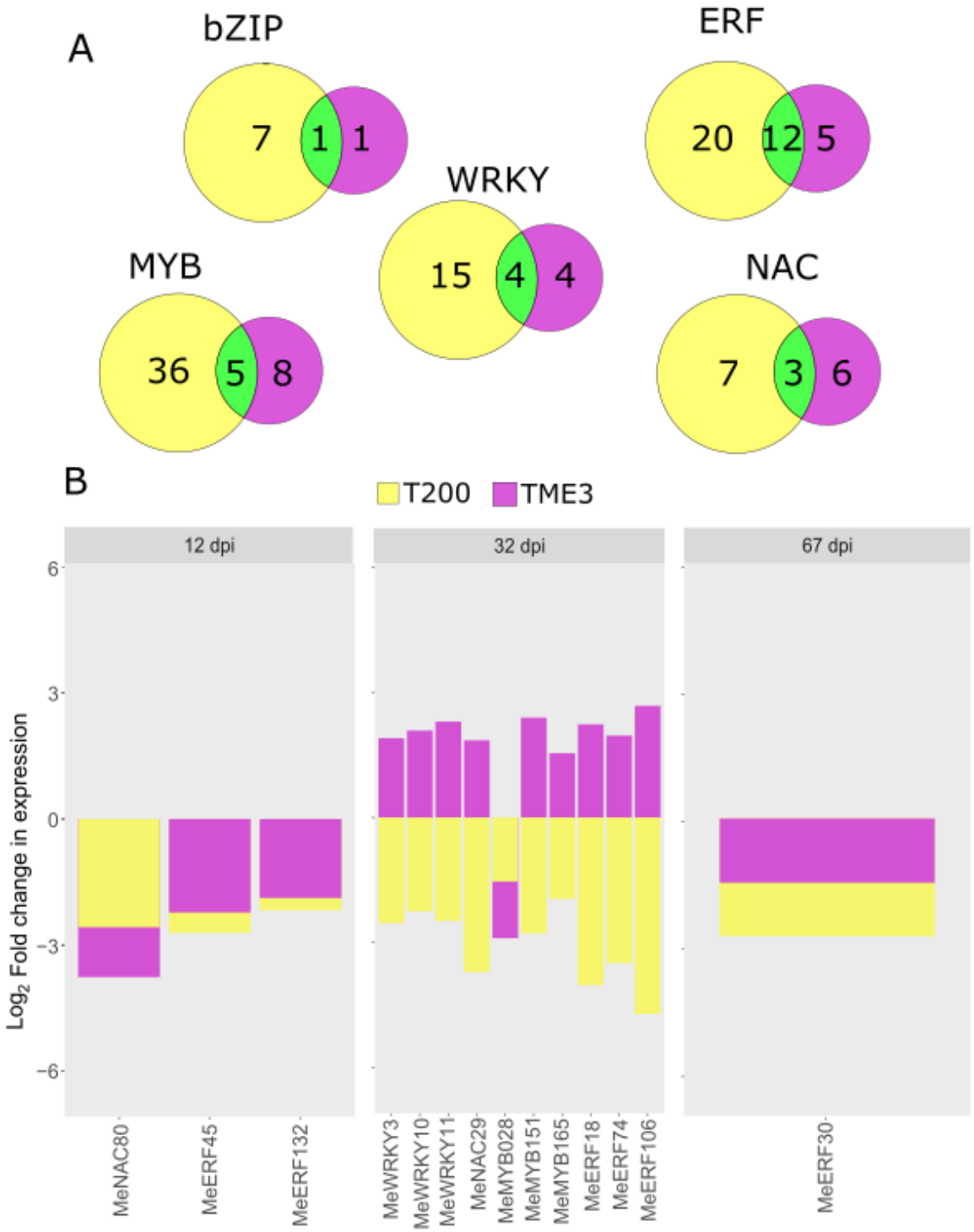


Figure 3-2:The comparison between (A) the number of different DE members belonging to the five TF families of interest and (B) the expression profiles of the common TFs DE at the same time point in both T200 and TME3 landraces.

Transcription factor networks in T200 and TME3 in response to SACMV infection

Due to the incomplete nature of genome annotation and functional characterization surrounding cassava, the means by which the genes that can be placed into a functional pathway are limited. In order to overcome this limitation, the Arabidopsis homologs/orthologs of the cassava genes were further analyzed to infer cassava gene function, and protein-protein networks were constructed.

During early infection, an interconnected network is established around nine different TFs and a TF-like protein that is homologous to *At*AP2 in T200 (Table 3-1, Figure 3-3). Notably, most of these TFs exhibit functions relating to either stress response or directly to regulation of immunity. The majority (76.66%) of the proteins seen in the network surrounding these TFs show down-regulation, whereas the proteins that show up-regulation are mostly connected to each other through two edges. In particular, proteins such as AT5G51190, *At*STZ, *At*RDUF1, *At*DIC2, AT1G23710, *At*ERF5 and 6 show the highest level of interconnectivity within the network. While most of the genes appear to be situated in the main central network, there is a small network surrounding the *At*RVE1 TF where all the associated genes are down-regulated.

Table 3-2:Transcription factors in T200 that form the centre of the underlying protein-protein interacting network at 12 dpi

TF [2]	Expression	Homolog	STRING ID	Arabidopsis Homolog Functions
*Me*ERF20 *Me*ERF19	-2.357246 -2.582828	AT5G51190	AT5G51190	Response to chitin
*Me*ERF17	-2.600037	AT5G47230	ERF5	Resistance to *B.cinerea*
*Me*ERF134	-2.027565	AT4G17490	ERF6	Resistance to *B.cinerea*
*Me*WRKY55	-1.889706	AT4G23810	WRKY53	Basal resistance against *P.syringae*
*Me*MYB218	-2.398403	AT5G17300	RVE1	Response to cold
*Me*WRKY59	-1.844029	AT3G56400	WRKY70	Regulates JA and SA antagonism
*Me*MYB313	3.73088	AT3G08500	MYB83	Negative regulator of plant defence against nematode
*Me*MYB295	1.996281	AT5G15310	MYB16	Negative regulator in cuticle formation
-	1.792281	AT4G36920	AP2*	Flower development and stem cell maintenance

[2] The TF refers to the name of the identified cassava gene and its relative expression compared to the mock-inoculated. The Arabidopsis homolog accession is shown with its corresponding name from the STRING database.

*AP2 was a Arabidopsis homolog that was identified as TF belonging to the ERF family however the cassava ortholog was not identified as an ERF TF.

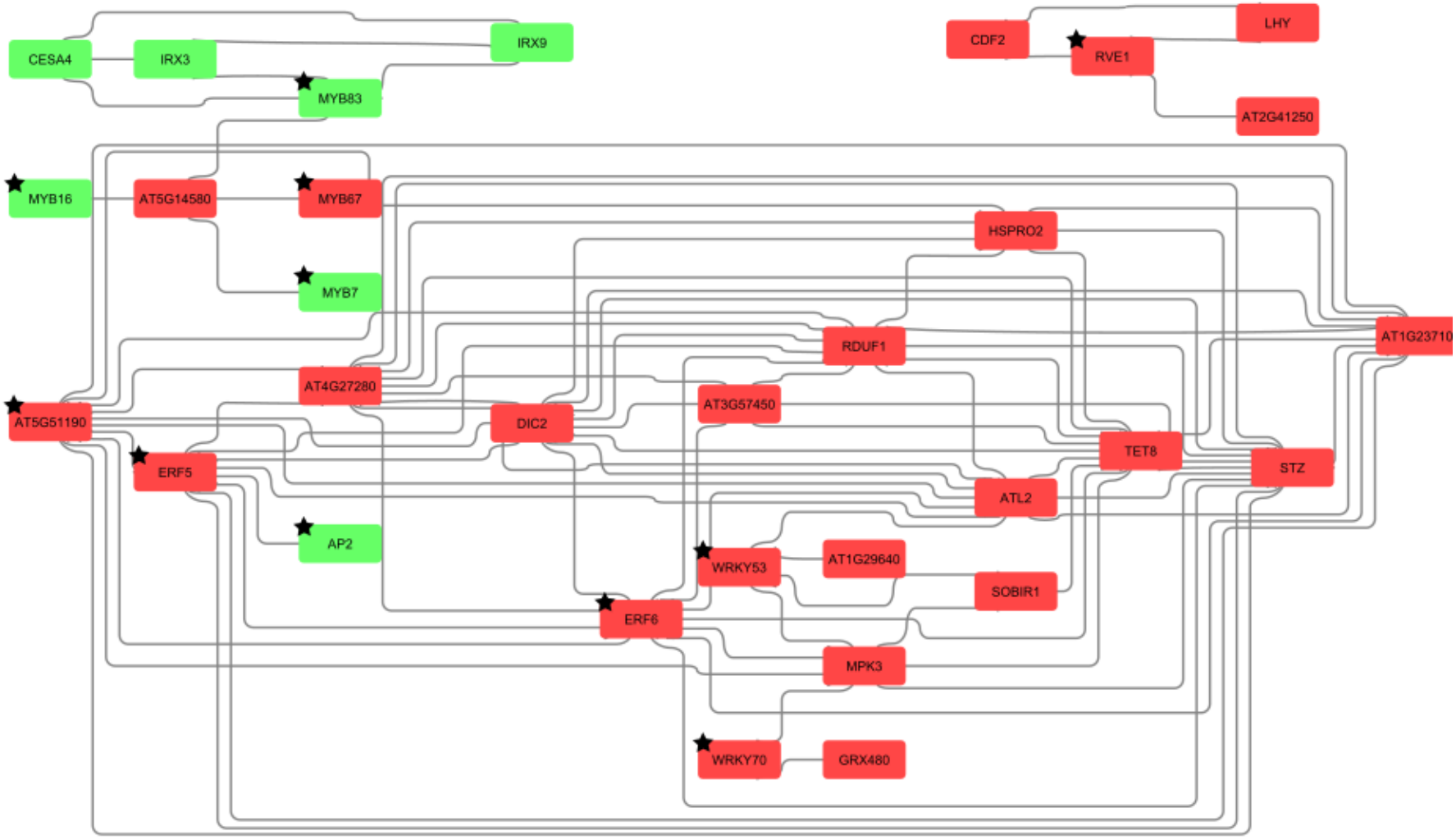


Figure 3-3: **Protein-protein interaction network centring around the Arabidopsis homologs of identified DE TF in T200 at 12 dpi.** TFs in the network are designated with a black star. Red genes represent down-regulation while green represents up-regulation of genes in response to SACMV. Only interactions occurring between at least two DEGs are shown and all unconnected DEGs were removed from the network

Following the trend seen in T200 at 32 dpi where disease symptoms become visible and virus replication is increasing, the number of DE TFs and the size and complexity of the resulting network increases considerably (Figure 3-1, Table 3-2)). Notably, with 49 different DE TFs, the resulting network is too large to adequately visualize (Supplementary Figure 1). For this reason, the focus herein will examine only a subset of the network. The limited network shows only the TFs factors that are commonly expressed in both TME3 and T200 at 32 dpi (Figure 3-2B and Figure 3-4). From the expression dataset, it is apparent that almost all the proteins in the constructed network are down-regulated (AP2 is the only exception) (Figure 3-2B). The next notable feature is how densely interconnected all the nodes are which is a consequence of the increased number of DE TFs and interacting protein partners that make up the network. However, some nodes do show greater levels of interconnectivity such as STZ, WRKY33, WRKY40, CZF1, TET8, AR781 and CAF1b. Interestingly, two of the homologs of the common DEGs, CBF4 and ATAF1, can be seen to be directly connected to the up-regulated AP2 protein.

Table 3-3:Identified TFs in T200 that form the centre of the underlying protein-protein interacting network at 32 dpi.

TF [4]	Expression	Homolog	STRING name	Arabidopsis Homolog Functions
*Me*ERF25	-6.288226	AT4G34410	RRTF1	Regulates redox responsiveness
*Me*ERF117	-4.3186			
*Me*ERF36	-5.138779	AT5G51990	CBF4	Regulation of drought tolerance
*Me*ERF42	-1.88383			
*Me*ERF106	-4.710917			
*Me*ERF74	-3.488525			
*Me*MYB173	-4.362994	AT3G23250	MYB15	Response to ABA, draught tolerance
*Me*MYB151	-2.778031			
*Me*ERF18	-4.019039	AT4G17500	ERF-1	Primary response to JA
*Me*WRKY42	-3.831362	AT5G64810	WRKY51	Mediates SA and represses JA signalling

[4] The TF name refers to the name of the identified cassava gene and its relative expression compared to the mock-inoculated.

***AP2** and *TT2* were Arabidopsis homologs that was identified as TF belonging to the ERF and MYB family respectively however the cassava orthologs were not identified as an TFs.

*Me*WRKY42	-3.831362			
*Me*WRKY79	-2.640528	AT1G80840	WRKY40	Positive regulator of ETI
*Me*WRKY11	-2.477297			
*Me*WRKY10	-2.244599			
*Me*ERF107	-3.406063	AT5G19790	RAP2.11	Modulates response to low-potassium stress
*Me*ERF38	-3.10319	AT3G15210	ERF4	Response to JA,ABA, ET
*Me*ERF30	-2.807985			
*Me*ERF17	-2.600037	AT5G47230	ERF5	Resistance to *B.cinerea*
-	-2.59105	AT5G35550	TT2*	Seed development
*Me*ERF19	-2.582828	AT5G51190	AT5G51190	Response to chitin
*Me*ERF20	-2.357246			
*Me*MYB271	-2.520234	AT3G28470	TDF1	Pollen development
*Me*MYB300	-2.303493	AT5G08520	AT5G08520	Positive regulator of sugar and ABA signalling,
*Me*MYB216	-2.148675	AT4G16110	RR2	Enhancer of SA/NPR1 defense against *P.syringae*
*Me*NAC3	-2.116867	AT5G24590	TIP	Basal defense
*Me*WRKY68	-2.602054	AT2G38470	WRKY33	Positive regulator of resistance against *B.cinerea*
*Me*ERF134	-2.027565	AT4G17490	ERF6	Resistance to *B.cinerea*
*Me*MYB165	-1.949856	AT4G37260	MYB73	Resistance to *B.oryzae*, response to JA,SA and ET
*Me*MYB086	-1.663987			
*Me*WRKY55	-1.889706	AT4G23810	WRKY53	Basal resistance against *P.syringae*
*Me*WRKY15	-2.563566			
*Me*bZIP16	-1.876595	AT5G28770	BZO2H3	Integrates ABA and sugar signalling
*Me*WRKY59	-1.844029	AT3G56400	WRKY70	Regulates JA and SA antagonism
*Me*MYB201	-1.561902	AT5G65790	MYB68	Negative regulator of lignin deposition in roots
-	1.762320	AT4G36920	AP2*	Regulation of flower development and stem cell maintenance
*Me*WRKY50	1.926513	AT2G37260	TTG2	Seed development
*Me*MYB295	1.996281	AT5G15310	MYB16	Negative regulator in cuticle formation
*Me*MYB294	-2.447883	AT1G69560	MYB105	Stamen development
*Me*MYB082	3.159448	AT1G22640	MYB3	Response to salt stress
*Me*MYB198	-4.153789	AT2G31180	MYB14	SA and JA response
*Me*MYB028	-1.537023	AT3G12720	MYB67	Unknown
*Me*bZIP1	-1.940088	AT2G40950	BZIP17	Salt tolerance
*Me*MYB220	2.488755	AT1G79430	APL	Vascular tissue development

*Me*bZIP55	3.064729	AT1G75390	bZIP44	Seed germination
*Me*bZIP68	2.235059			
*Me*MYB227	2.627819	AT2G38090	AT2G38090	Unknown
*Me*MYB179	2.409854			
*Me*NAC29	-3.708052	AT1G01720	ATAF1	Biotic and abiotic stress response

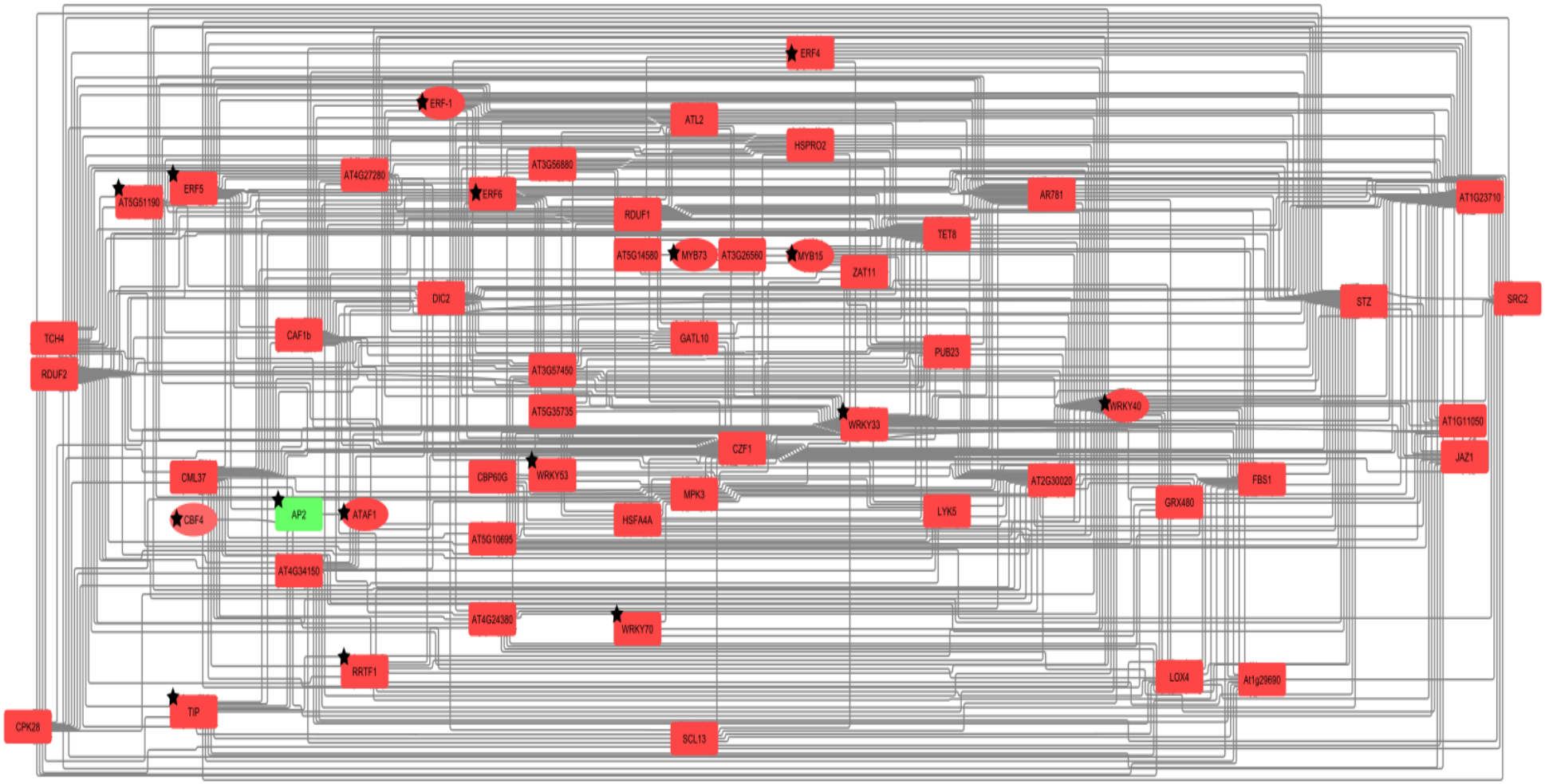


Figure 3-4: Protein-protein interaction network centring around the Arabidopsis homologs of identified DE TF in T200 at 32 dpi. TFs in the network are designated with a black star. Red genes represent down-regulation while green represents up-regulation of genes in response to SACMV. Only interaction occurring between at least two DEGs are shown and all unconnected DEGs were removed from the network. The oval nodes represent the TFs that are commonly DE in both landraces at 32 dpi.

The most notable feature about the DEGs in T200 at 67 dpi, is the large overlap experienced with those found in T200 at 32 dpi (Table 4). In fact, only four TFs, namely: AtbZIP61, AtERF1, *At*MYB83 and AtWRKY35 are found to be uniquely DE at 67 dpi. All the other TFs orthologs show the same expression previously at 32 dpi (Supplementary Figure 1). Interestingly in previous time points, there were multiple cassava genes that corresponded to a single Arabidopsis ortholog, whereas during 67 dpi, each cassava genes only has a single Arabidopsis ortholog. Due to this large degree of similarity between the two time points in T200, it necessitated the need to only look at the unique proteins centring around the four unique TFs that were DE at 67 dpi (Figure 3-5).

In comparison to the previous two networks in T200, the most notable feature is that the majority of the proteins found in the network experienced up-regulation in response SACMV. Furthermore, whereas the network at 32 dpi, showed a single large interconnected network, the unique network at 67 dpi is comprised of multiple smaller networks of roughly three proteins. Among this network, the largest is found to lack any of the unique TFs from the five families of interest. Interestingly, the up-regulation of an ortholog of *At*ERF1 is seen in both T200 at 67dpi and TME3 at 32 dpi. However, that's where the similarities end with T200 at 67 dpi and TME3.

Table 3-4:Identified TFs in T200 that form the centre of the underlying protein-protein interacting network at 32 dpi. In red are genes that are uniquely found in T200 at 67 dpi and not

TF [5]	Expression	Homolog	STRING name	Arabidopsis Homolog Functions
-	1.792281	AT4G36920	AP2*	Regulation of flower development and stem cell maintenance
*Me*MYB220	1.73088	AT1G79430	APL	Vascular tissue development
*Me*MYB227	2.627819	AT2G38090	AT2G38090	Unknown
*Me*bZIP3	-1.77588	AT4G38900	AT4G38900	Unknown

[5] The TF name refers to the name of the identified cassava gene and its relative expression compared to the mock-inoculated

*Me*ERF19	-2.58283	AT5G51190	AT5G51190	Response to chitin
*Me*NAC29	-3.70805	AT1G01720	ATAF1	Biotic and abiotic stress response
*Me*bZIP51	1.930054	AT3G58120	BZIP61	Boron deficiency response
*Me*ERF42	-4.71092	AT5G51990	CBF4	Regulation of draught tolerance
*Me*ERF40	1.976347	AT3G23240	ERF1	Disease resistance
*Me*ERF38	-3.10319	AT3G15210	ERF4	Response to JA,ABA, ET
*Me*ERF17	-2.60004	AT5G47230	ERF5	Resistance to *B.cinerea*
*Me*ERF134	-2.02757	AT4G17490	ERF6	Resistance to *B.cinerea*
*Me*MYB198	-4.15379	AT2G31180	MYB14	SA and JA response
*Me*MYB151	-4.36299	AT3G23250	MYB15	Response to ABA, draught tolerance
*Me*MYB082	3.159448	AT1G22640	MYB3	Response to salt stress
*Me*MYB262	-2.59215	AT5G06100	MYB33	Anther Development
*Me*MYB236	3.143457	AT4G12350	MYB42	Unknown
*Me*MYB313	-1.62503	AT3G08500	MYB83	Regulation of secondary cell biosynthesis
*Me*MYB139	-1.6002	AT5G67300	MYBR1	Response to ABA
*Me*NAC37	-3.64053	AT2G17040	NAC036	Positive regulators of SA-mediated responses
*Me*MYB186	2.391385	AT3G46640	PCL1	Circadian clock
*Me*MYB216	-2.14867	AT4G16110	RR2	Enhancer of SA/NPR1 defense against *P.syringae*
*Me*ERF25	-6.28823	AT4G34410	RRTF1	Regulates redox responsiveness
*Me*NAC3	-2.11687	AT5G24590	TIP	Basal defense
*Me*WRKY50	1.926513	AT2G37260	TTG2	Seed development
*Me*WRKY30	2.21092	AT4G01250	WRKY22	Abiotic stress and senescence
*Me*WRKY68	-2.0461	AT2G38470	WRKY33	Positive regulator of resistance against *B.cinerea*
*Me*WRKY54	2.787313	AT2G34830	WRKY35	Unknown
*Me*WRKY79	-2.4773	AT1G80840	WRKY40	Positive regulator of ETI
*Me*WRKY55	-1.88971	AT4G23810	WRKY53	Basal resistance against *P.syringae*
*Me*WRKY59	-1.84403	AT3G56400	WRKY70	Regulates JA and SA antagonism

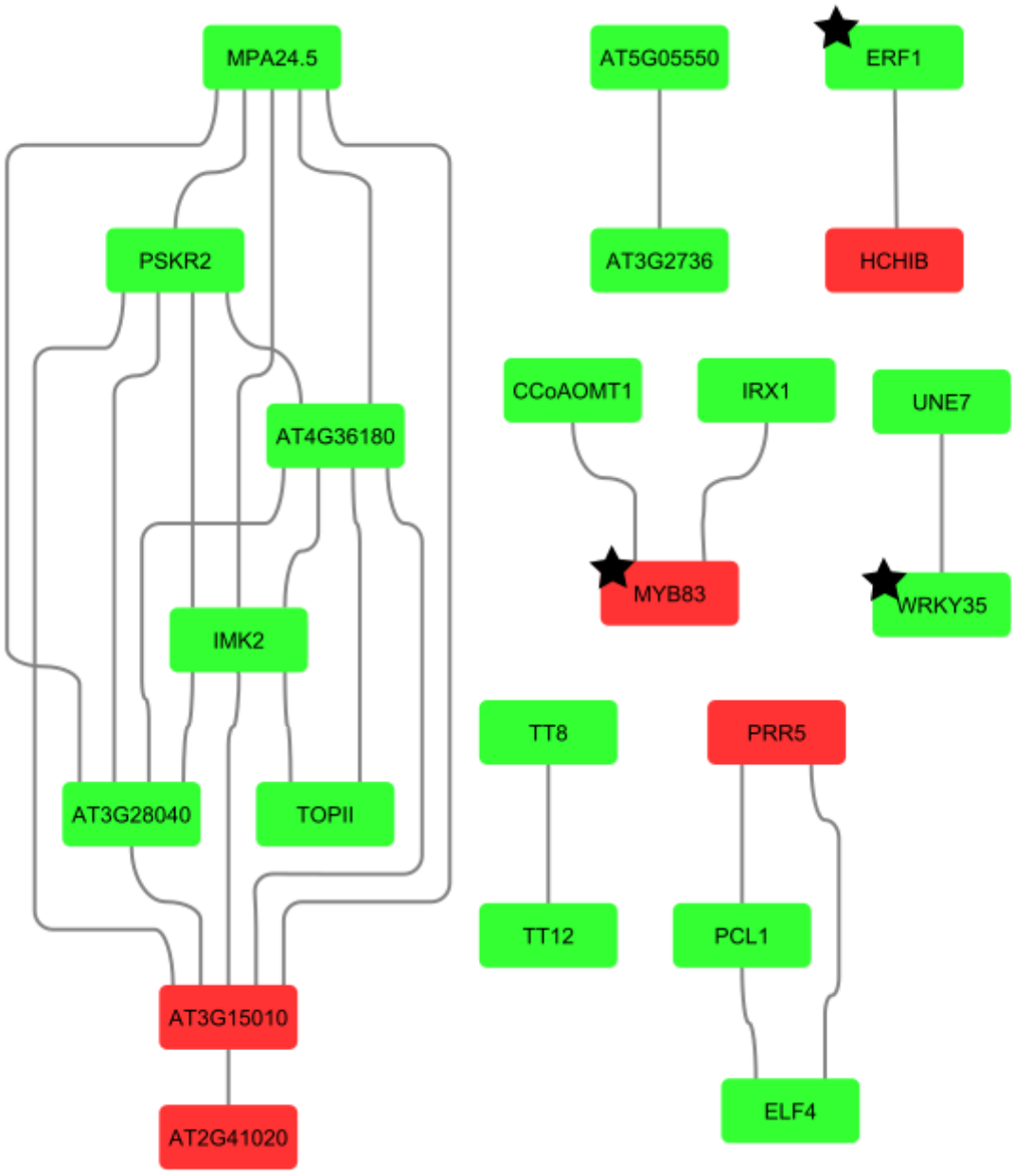


Figure 3-5: Protein-protein interaction network centring around the unique Arabidopsis homologs only identified to be DE in T200 at 67 dpi. TFs in the network are designated with a black star. Red genes represent down-regulation while green represents up-regulation of genes in response to SACMV. Disconnected nodes were removed. Only interaction occurring between at least two DEGs are shown and all unconnected DEGs were removed from the network.

In contrast to the dense complex network in T200, TME3 shows a less connected and smaller network at 12 dpi (Figure 3-5), compared with T200, due to the network is built around only two different TF homologs whose expression was altered in response to SACMV, namely *At*WRKY40 (*Me*WRKY40 and 79) and *At*RRTF1 (*Me* ERF25) (Table 3-3). Despite the size of the network, there are still proteins that show high levels of interconnectivity among nodes, namely: *At*WRKY40, *At*STZ and *At*DIC2 and *At*CML37. From these highly interconnected nodes, STZ and DIC2 (in addition to HSPRO2 and GRX480) would be familiar as they also featured in T200 at 12 dpi. However, it is important to note that the cassava homologs that correspond to these genes are not the same in T200 and TME3 at 32 dpi, but rather their Arabidopsis homologs are shared. Similarly, the TME3 network also shows mostly (88.3%) down-regulation in response to SACMV during early infection. Interestingly the only two genes up-regulated (*At*WRKY40 and *At*JAZ1) are unique to TME3 at 12 dpi. Furthermore, two different cassava TFs (*Me*WRKY79 and 40) are both represented by *At*WRKY40. Notably, of the homologs that were determined to be targets of *At*WRKY40 (Pandey *et al.*, 2010), only the homolog of JAZ1 did not match the expected expression patterns following up-regulation of the homologs of *At*WRKY40.

Table 3-5: Identified TFs in TME3 that form the centre to the underlying protein-protein interacting network at 12 dpi.

TF[6]	Expression	Homolog	STRING name	Arabidopsis Homologs Function
*Me*WRKY79	2.437730721	AT1G80840	WRKY40	Positive regulator of ETI
*Me*WRKY40	4.249705184			
*Me*ERF25	-2.44389418	AT4G34410	RRTF1	Regulates redox responsiveness

[6] The TF name refers to the name of the identified cassava gene and its relative expression compared to the mock-inoculated

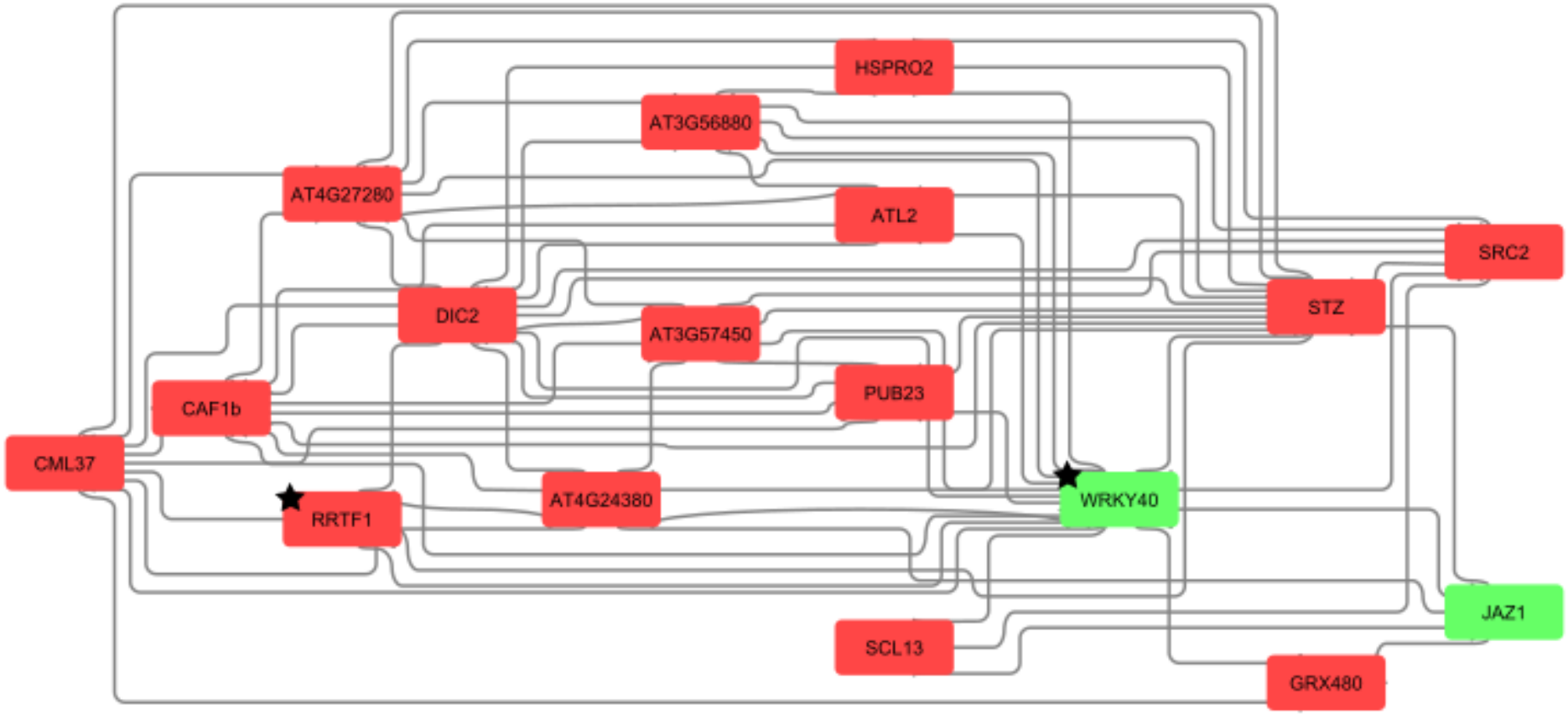


Figure 3-6: Protein-protein interaction network centring around the Arabidopsis homologs of identified DE TFS in TME3 at 12 dpi. TFs in the network are designated with a black star. Red genes represent down-regulation while green represents up-regulation of genes

Despite the additional DE TFs used to build the network, the total number of nodes remain unchanged between the 12 and 32 dpi networks in TME3 (Table 3-5: Figure 3-6). However, it should be noted that whereas at 12 dpi the proportion of down-regulated genes was high (88.3%), whereas at 32 dpi the proportion of up- regulated TFs and interacting partners in TME3 is now high (88.3) and down-regulated genes low (12.7%). Notably, in TME3, there seems to be very little overlap between the DEGs in the 12 and 32 dpi networks. Those that are common between these two networks are JAZ1 and WRKY40, where their expression remained up-regulated at both time points. Comparing the networks generated at 32 dpi in T200 and TME3, there are a few interesting contrasting features that can be observed (Figure 3-4). Firstly, there are very little commonalities between the two networks. For instance, despite DE of MYB15, ERF-1 and WRKY40 in both landraces at 32 dpi, the cassava homologs (*Me*MYB151, *Me*ERF18 and *MeWRKY10*, 11 respectively) of these genes show contrasting expression (Figure 3-2B). Furthermore, the nodes connected to these three Arabidopsis homologs differ between networks. The only commonality is up-regulation of ETR 2 in T200 and TME3 (Figure 3-6).

Table 3-6: Identified TFs in TME3 that form the centre of the underlying protein-protein interacting network at 32 dpi.

TF[7]	Expression	Homolog	STRING name	Arabidopsis Homologs Function
MeERF40	2.019579999	AT3G23240	ERF1	Disease resistance
*Me*WRKY11	2.297786717			
*Me*WRKY79	2.437730721	AT1G80840	WRKY40	Positive regulator of ETI
*Me*WRKY10	2.082773826			
*Me*ERF18	2.230270538	AT4G17500	ERF-1	Primary response to JA
*Me*MYB151	2.392395025	AT3G23250	MYB15	Response to ABA, draught tolerance
*Me*NAC33	3.560821122	AT5G22380	NAC090	Positive regulator of SA-mediated response

[7] The TF name refers to the name of the identified cassava gene and its relative expression compared to the mock-inoculated

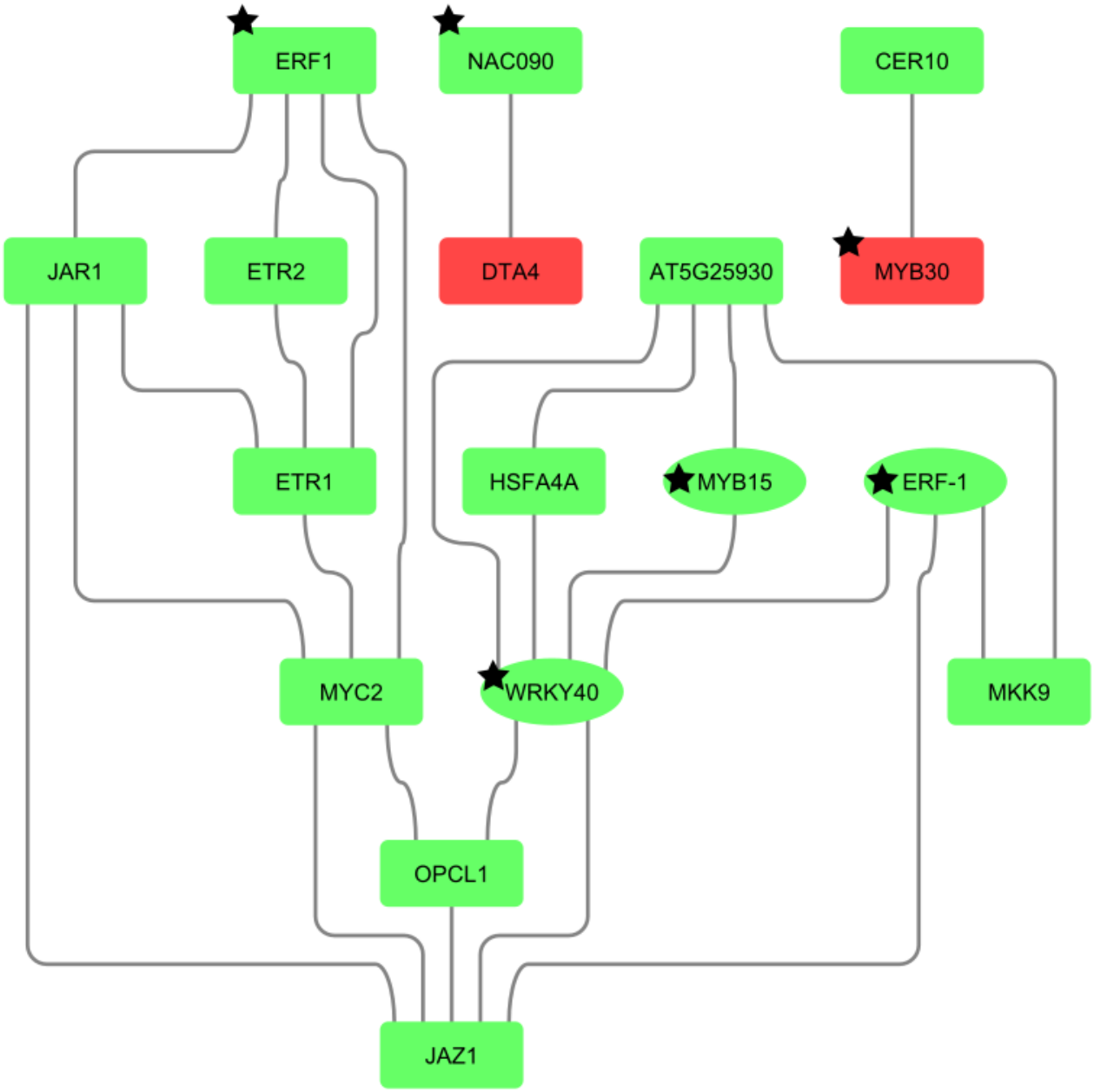


Figure 3-7: **Protein-protein interaction network centring around the Arabidopsis homologs of DE TFS in TME3 at 32 dpi**. TFs in the network are designated with a black star. Red genes represent down-regulation while green represents up-regulation of genes in response

.

At 67 dpi, during this phase of recovery which leads to tolerance, there is a remarkable reduction in the network interconnectivity surrounding DE TFs (**Figure 3-7**). Only AtWRKY40 and MYB15 remain up-regulated, while only ERF4 and a tRNA-synthase (AT5G10695) were down-regulated (**Table 3-5**). Interestingly, ERF4 is down-regulated in both T200 and TME3. In particular, AtWRKY40 is the most interconnected as it connects to each of the other three genes. Furthermore, AtWRKY40 in conjunction with MYB15 are seen to contribute to both the response to wounding and JA biosynthesis which is derived from the previous GO term enrichment analysis.

Table 3-7: Identified TFs in TME3 that form the centre of the underlying protein-protein interacting network at 67 dpi.

TF[8]	Expression	Homolog	STRING name	Arabidopsis Homolog Functions
*Me*MYB027	3.123624	AT3G23250	MYB15	Response to ABA, draught tolerance
*Me*WRKY10	2.082774	AT1G80840	WRKY40	Positive regulator of ETI
*Me*ERF30	-1.53273	AT3G15210	ERF4	Response to JA,ABA,ET

[8] The TF name refers to the name of the identified cassava gene and its relative expression compared to the mock-inoculated

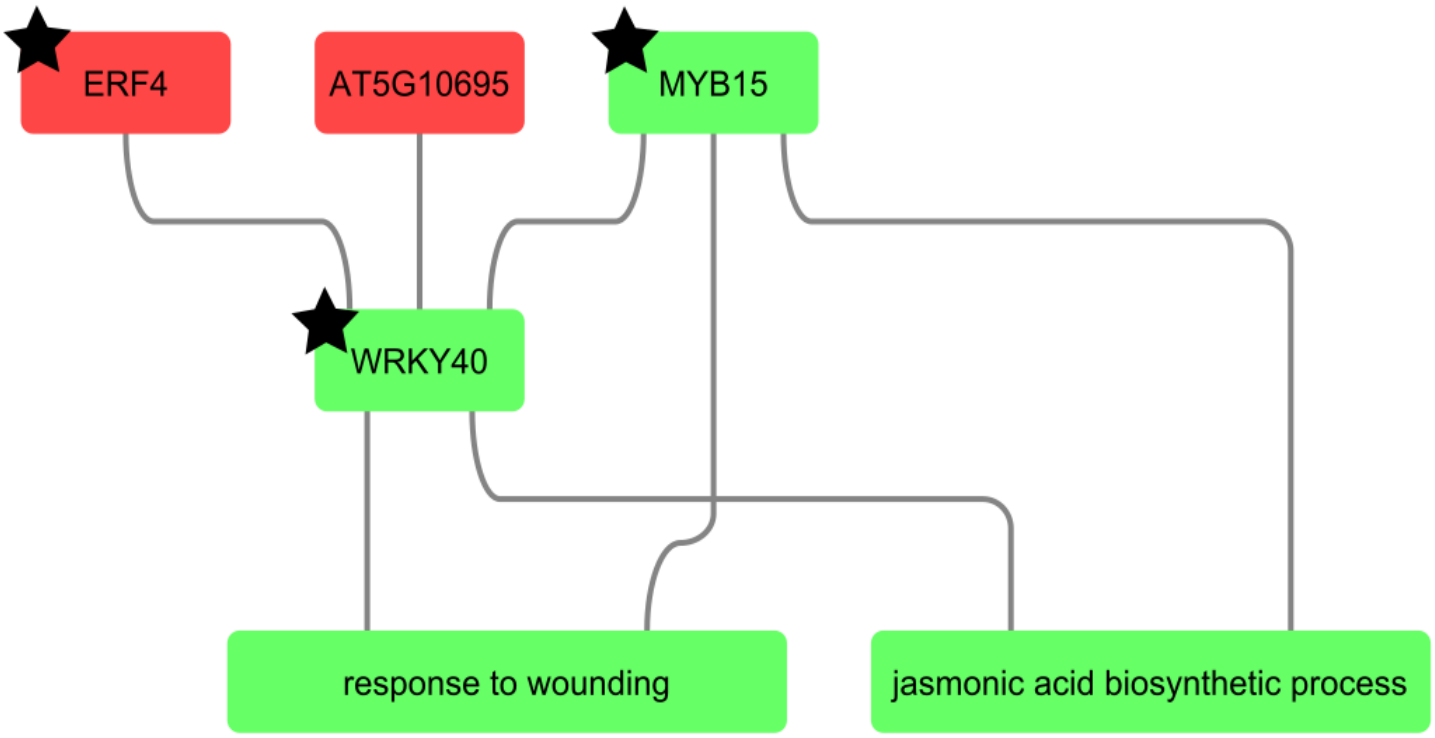


Figure 3-8: Protein-protein interaction network centring around the Arabidopsis homologs of DE TF in TME3 at 67 dpi. TFs in the network are designated with a black star. Red genes represent down-regulation while green represents up-regulation of genes in response to SACMV. Only interaction occurring between at least two DEGs are shown and all unconnected DEGs were removed from the network. Enriched GO processes among up-regulated genes at 67 dpi are shown at the bottom, edges connecting the nodes and the GO processes show which genes contributing to the enrichment.

3.4 Discussion

3.4.1 General Introduction

Transcription factors are widely known to be associated with plant responses to pathogens (Golldack *et al.*, 2011). Most studies have focused on fungal and bacterial pathogens, so any information in regard to viruses would prove invaluable. This research is the first to date to dissect five transcription factor families in a susceptible and tolerant plant host in response to a viral pathogen. Results demonstrate that TF responses (up- or down-regulation) between a susceptible and tolerant host are uniquely different, highlighting the importance of TFs is regulating immune or susceptibility responses. Furthermore, this study is the first to examine five TF responses to a geminivirus, South African cassava mosaic virus, in the host cassava, and to compare TFs between a susceptible T200 and tolerant TME3 landrace. Importantly, the numbers of MYB and ERF TFs were the highest in the two cassava landraces compared with TF members of the WRKY, NAC and bZIP families, and interestingly TF-interacting partner networks differed significantly between T200 and TME3 at all three stages of infection.

3.4.2 The contrast between susceptibility (T200) and tolerance (TME3)

While many studies have been undertaken on resistance to plant viruses (reviewed in Rey and Fondong, 2017), little is known about the molecular mechanisms of tolerance (Bengyella *et al.*, 2015, Gaudel and Sanfaçon, 2018). When considering tolerance to cassava mosaic disease, it is important to focus on the differences observed between T200 and TME3 post-SACMV-infection in relation to gene expression. Furthermore, each time point must be scrutinized independently to be able to dissect the infection stages leading to recovery and tolerance in the TME3 landrace (Allie *et al.*, 2014). Regulation of virus-host interactions by transcription factors and hormones is well documented (Alazem and Lin, 2015, Collum and Culver, 2016). This study aimed to interrogate and compare transcriptome datasets pertaining to signalling (transcription factors and hormones) in T200 and TME3 and to use the information to build a cohesive model for signalling associated with priming tolerance.

The most striking observation is the comparatively low numbers and diversity of DE TFs in TME3 in relation to T200 post-SACMV infection. From this result, it can be concluded that SACMV in TME3 leads to lower perturbations of the host system.

Replication of viruses requires global transcriptional disturbances leading to physiological perturbations, a pre-requisite for hi-jacking of the host by the virus, leading to symptom development (Mandadi and Scholthof, 2013). It has been demonstrated that TME3 shows a lower viral titre at all stages of infection compared with T200, consequently leading to milder symptoms during middle and late infection (Allie *et al.*, 2014). It is unknown whether this observation is a direct consequence of the underlying active mechanisms of tolerance, such as an active immune response, or due to passive features in TME3, such as incompatible receptors preventing viral entry thereby limiting replication. It has been shown that RNA silencing (PAMP-triggered immunity) (Rogans *et al.*, 2016) and nucleotide binding-leucine rich repeat (NB-LRR) R genes (Bengyella and Rey, 2015) play a role in tolerance in TME3, but no studies on signalling networks have been performed to date. The most notable changes seen among the five TF families (particularly ERF, NAC and WRKY) were observed at 32 dpi suggesting a mechanistic switch occurring between 12 and32 dpi as virus infection progresses. This coincides with the observation in the previous chapter surrounding hormone signalling in TME3 showing that ABA, ET, JA and SA are at their pinnacle during middle infection. From these and other observations, even at the early SACMV infection stage TME3 responds differently to T200 leading, at middle infection, to milder symptoms and lower virus replication, which are hallmarks of tolerance. It is hypothesized that from 12 to 32 dpi several molecular mechanisms associated with TF/hormonal signalling are switched on that activate tolerance and recovery. This hypothesis is based on the expression dataset exhibiting the highest level of up-regulation in TME3 which coincides with the spike in hormone responsive genes all found at 32 dpi.

Interesting points may be drawn regarding the response seen in both landraces to SACMV though examining the genes commonly DE in both T200 and TME3 at the same time point (Figure 3-2). At 12 dpi, the down-regulation of all common TFs (*Me*NAC80, *Me*ERF45 and 132) support the previous chapters' observation that T200 and TME3 are most similar during early infection. This degree of similarity seen between the two landraces would suggest that it is the unique genes found in each host at 12 dpi that is contributing to susceptibility or tolerance. Despite both host display symptoms at 32 dpi, it is clear from the large contrasts in expression between common DE TFs that TME3 and T200 have started to

respond to SACMV. Given the assumption that T200 represents a standard susceptible system, it can be proposed that this difference in gene expression is the consequences of TME3 starting to mount a defence response. Furthermore, support for this proposed idea may be derived from the up-regulation of the common DE TF in TME3, which suggest the induction of the TFs function whereas, in T200, they are suppressed. It is likely the suppression of these TF function is preventing T200 from initiating downstream defence processes thereby contributing to susceptibility. Transitioning to 67 dpi, it is apparent that T200 show the lowest degree of similarity during late infection. This is to be expected considering that during previous time points, both T200 and TME3 display similar phenotypes: non-symptomatic and symptomatic respectively. Therefore, the little similarity seen between the two landrace is likely due to the large differences in the underlying molecular mechanisms which culminate in the recovery or chronic infection. In order to determine the potential mechanisms underlying tolerance and susceptibility, scrutiny needs to be placed on these molecular mechanisms (hormone signalling, defence response by TFs, PR protein induction) in the context of each host.

The significant up-regulation of the transcription repressor: *At*WRKY40 (*Me*WRKY40 and 79) may contribute to a wide degree of down-regulation seen in the TME3 networks at 12 dpi. The up-regulation in *At*WRKY40, according to literature, should lead to the down-regulation of the homolog of *At*JAZ1 (Pandey *et al.*, 2010), however, JAZ1 and 8 were up-regulated in TME3 at 12 dpi. It is possible that the homolog of *At*JAZ1 (cassava4.1_014096m.g) is induced through *Me*WRKY79 in TME3 in cassava, as when *Me*WRKY79 is down-regulated in T200 at 32 dpi, the same *At*JAZ1 homolog (cassava4.1_014096m.g) is concomitantly down-regulated, implying co-expression. It is known that the *At*JAZ family of proteins are involved with the inhibition of JA signalling and through the MYC2 pathway (Thines *et al.*, 2007). However, all the putative cassava JAZ proteins found in TME3 throughout SACMV infection are up-regulated despite JA signalling being the most predominant hormone response in TME3. Despite the up-regulation of these JAZ proteins, the induction of JA signalling is evident from the up-regulation of the two cassava homologs (cassava4.1_009912m.g and cassava4.1_006932m.g) of *At*MYC2 at 32 dpi in TME3, suggesting that these homologs in cassava function as inducers rather than

inhibitors of JA signalling. This notion is further supported by the observation that all the DE *At*JAZ protein homologs in T200 are down-regulated.

3.4.3 The downfall of susceptible T200

The transcriptomic data from T200 allows for the observation of individual molecular processes surrounding the biological processes and how they link to the five TF families, enabling comment on CMD establishment and progression. T200 appears to be destined for failure in resisting SACMV, as evident in the down-regulation hormone signalling responses noted in the previous chapter. This initial down-regulation of hormone responses appears to have many perturbations that direct the manner in which T200 respond to SACMV from as early as 12 dpi.

3.4.3.1 Disease establishment at 12 dpi

The suppression of hormone response genes is crucial for the establishment of infection (Teper *et al.*, 2018), in addition, physiological changes are necessary to facilitate optimal viral replication. From the results provided it is clear that T200 in undergoing both suppressions of immune response and shows the beginnings of developmental changes, prior to symptom presentation. It is therefore proposed that disease establishment begins in T200 from 12 dpi.

In immune suppression, there are several groups of TFs at 12 dpi that are linked to the suppression of many hormone responses, particularly ET, JA and SA. Particularly, *At*WRKY53 (*Me*WRKY55) and *At*WRKY70 (*Me*WRKY59) are important as they are found to mediate the cross-talk between SA and JA signalling pathways (Li *et al.*, 2004, Miao and Zentgraf, 2007). The down-regulation of *Me*WRKY55 and 59 is proposed to not only disrupt SA signalling but to lead to a prioritization of JA signalling (Li *et al.*, 2004). Further evidence to support JA prioritization by T200 is the down-regulation of the *At*GRX480 homolog, which has been shown to be a negative regulator of JA-ET signalling (Ndamukong *et al.*, 2007). Since both *At*WRKY53 and 70 are involved in SA signalling and the establishment of SAR, the suppression of both their orthologs (*Me*WRKY55 and 59) is the suggested reason that T200 is unable to initiate a SAR directed immune response against SACMV (Hu *et al.*, 2012). Despite T200 showing signs of JA prioritization, the hormone response profile shows that JA is down-regulated like SA, suggested there are other genes

contributing to the suppression of both hormone signals. Furthermore, from the network, there are no particular genes that induce a JA orientated response. This means that although the expression of genes in the TF networks show that T200 is prioritizing a JA response, by suppressing various SA mediated pathways, T200 merely fails to initiate any JA response. Lastly, much like SA signalling, ET signalling shows clear signs of suppression in the main network, with the down-regulation of *At*ERF5 ,6 and 105 (AT5G51190) (Moffat *et al.*, 2012, Son *et al.*, 2012). These are notable as *At*ERF5 has been implicated in response to chitin, thereby forming part of the innate network in Arabidopsis and together with *At*ERF6 are shown to contribute resistance towards *B.cinerea* (Son *et al*., 2012, Moffat *et al*., 2012) Whereas, *At*ERF105 is more closely associated with abiotic stress such as cold resistance (Bolt *et al.*, 2017). However, *At*ERF105 is a member of the IXb group in the ERF families, whose members contain a MAP kinase phosphorylation site which that members of the IXb group control gene expression in response to an extracellular stimulus (Nakano *et al.*, 2006, Moffat *et al.*, 2012). With this information, coupled with the understanding that biotic and abiotic signalling pathways display overlapping signalling molecules suggest that the down-regulation of *At*ERF105 is contributing to the suppression of immunity in T200.

The myeloblastosis related (MYB) TF family is reported to be involved in many processes, including leaf morphology and biotic stress (Dubos *et al*., 2010 Trends in Plant Science). For example, *At*MYB30 is a positive regulator of the hypersensitive response (HR) (Alves *et al*., 2014). The MYB-like TF *At*RVE1 is found to be at the centre of the smaller network that is disconnected from the larger network that contains all the WRKY orthologs (Figure 3-3). In Arabidopsis *At*RVE1 controls auxin production, which would imply the down-regulation in its orthologs (*Me*MYB218) at 12 dpi, would likely lead to the failure to accumulate auxins in T200 (Rawat *et al.*, 2009). In addition to *At*RVE1, the down-regulation of the *At*LHY homolog in T200 would suggest that SACMV disrupts the circadian clock through transcriptional reprogramming (Rawat *et al.*, 2009). This is notable as it has been found that disruption of the circadian clock through *lhy* mutants leads to increased susceptibility towards oocyte pathogens and *Pseudomonas syringae* and decreased responsiveness to auxin (Rawat *et al.*, 2009, Wang *et al.*, 2011). This increased susceptibility may also be due to the fact that *At*LHY responds to both ABA and ET (Kumar *et al.*, 2016).

This implies that *At*LHY may, in turn, regulate immunity through the response of these two hormones, making it a target by SACMV. It seems paradoxical that SACMV would be suppressing auxin signalling considering that the accumulation of auxin is negatively associated with resistance (Robert-Seilaniantz *et al.*, 2011). It is possible disruption of the circadian clock confers a greater advantage to disease establishment than auxin accumulation would. This has been confirmed in Arabidopsis, whereby certain PTI responses (callose deposition) showed greater induction based on the perceived time period in plants infected with *P.syringae* (Bhardwaj *et al.*, 2011). Furthermore, it was found through publically available microarray data that many defence genes demonstrate expression indicative of circadian rhythm regulation (Bhardwaj *et al.*, 2011). Some of which are found within the networks (*At*SOBIR1, *At*MPK3, *At*WRKY70). This suggests that in order for SACMV to establish infection in T200, it requires the disruption of circadian rhythm.

From these three defence genes, the most well-documented gene contributing to initiating defence responses is *At*MPK3 (Meng and Zhang, 2013). Particularly, *At*MPK3 has been found to respond to flg-22 and lead to stomatal closure preventing pathogen entry (Montillet *et al.*, 2013). The down-regulation of the MPK3 prevents this stomatal closure, and the induction of defence response downstream of the MAPK cascade, marking a crucial step in the invasion of T200. Notably, in T200, the *At*MPK3 homolog (cassava4.1_010219m.g) was not only downregulated at 12 dpi but throughout the infection stages, providing evidence for a pivotal role for suppression of defence to SACMV. Additionally, *At*SOBIR1 is a suppressor of the negative resistance regulator *At*BIR1 and acts as a signal transducer in cell death and disease resistance pathways (Gao *et al.*, 2009) . The down-regulation of *At*SOBIR1 would entail the accumulation of *At*BIR1 leading to suppression in immune response and prevent cell death. This would further contribute to the success of SACMV in establishing disease and maintaining replication within cells.

Despite viruses not containing any chitin, response to chitin is representative of many PTI general signals occurring within the host. Notably, the four DE ERFs, *Me*ERF19 and 20 (AT5G51190), *Me*ERF17(*At*ERF5) and *Me*ERF134(*At*ERF6), are shown to be responsive to chitin. It was found that both *At*ERF5 and 6 interact with *At*MPK3, however specifically *At*ERF5 is believed to regulate the innate defence of the host through modulating SA

signalling (Son *et al.*, 2012). The down-regulation of these three genes suggests that SACMV begins to suppress basal immunity in T200 even before symptoms start to appear. This notion is further supported by two other gene homologs *At*RDUF2 and *At*ATL2. The down-regulation of the chitin responsive ubiquitin E3 ligase *At*RDUF2 would lead to a reduction in the indirect immune responses it initiates (Matthes *et al.*, 2010). This coupled with the down-regulation of *At*ATL2, a zinc-finger protein that functions as an early elicitor of a defence response, would likely lead to further limitations in the initial early immune response in T200. From the above, it is evident that early infection in T200 is hallmarked by numerous examples of suppression of early biotic stress and hormonal response.

Furthermore, successful disease establishment also requires developmental changes in a host to facilitate viral replication (Whitham and Wang, 2004, Culver and Padmanabhan, 2007, Mandadi and Scholthof, 2013). Despite no symptoms occurring at 12 dpi, viral replication has begun and is detectable in cassava (Allie *et al.*, 2014). This means that in order for the virus to create an optimal environment for replication, it must initiate developmental changes. The beginning stages of these changes are apparent in two particular gene homologs, namely the homologs of *At*CDF2 and *At*AP2 (cassava4.1_030452m.g and cassava4.1_006539m.g respectively) detected in the network (Figure 3-3). Notably, *At*CDF2 is known to interact with the DCL1 protein and in doing so regulate the accumulation of miR172 and miR156 which have functions in flower development (Sun *et al.*, 2015).

Interestingly miR156 and miR176 have been shown to target the SPL TF family, one of whose functions is in leaf initiation (Wu *et al.*, 2009a). It is proposed that down-regulation of the *At*CDF2 genes reduces the levels of miRNAs leading to up-regulation of associated SPL transcription and leaf initiation reduction. This could explain the stunted growth seen with infected cassava compared to the mock-inoculated controls. Additionally, miR172 has been shown to target APETALA2 (*AP2*-like) genes (Aukerman and Sakai, 2003), which notably is the only one of the few up-regulated genes in the T200 network (Figure 3-3). *At*AP2 has various functions in development, including stem cell maintenance, and may be up-regulated as a consequence of the reduced levels of miR172 (Würschum *et al.*, 2006). Furthermore, *AP2* genes also play a role in signalling, stress response and regulation of disease resistance pathways (Gutterson and Reuber, 2004). In contrast, in susceptible and

moderately resistant tomato *AP2* was shown to be down-regulated by BCTIV (Ghanbari *et al.*, 2016) and in Tomato leaf curl New Delhi virus (CoLCNDV miR172 was shown to increase with the number of days after geminivirus inoculation (Naqvi *et al*., 2010). An RNAseq study of small RNAs in SACMV-infected cassava showed no change in expression of miR172 in T200 at 12 dpi (unpublished results), suggesting other factors responsible for AP2 up-regulation at the early infection stage.

It is clear to see that the large central network seen in T200 at 12 dpi, governs defence process while the smaller network around *At*RVE1 is used to disrupt circadian rhythm as a secondary means to reduce the immune response. However, there is one final network constructed around numerous MYB TFs (*At*MYB7,16 and 83) that show mostly up-regulation of its proteins. In this network are three genes, irregular xylem (IRX)3, AtIRX9 and the cellulose synthase *At*CESA4, that coordinate a unifying function of regulating the structural integrity of the plant through xylem and cellulose biosynthesis (Turner and Somerville, 1997, Wu *et al.*, 2010, Kim *et al.*, 2013). The link between why these portions are up-regulated in response to SACMV is obscure. However, it is possible the up-regulation of the three proteins is the consequence of up-regulation seen in processes regarding growth and meristem development discussed in the previous chapter. Therefore, it appears that this smaller network is indirectly linked to the developmental processes not shown in the protein-protein network.

3.4.3.2 32 dpi disease maintenance

Despite both TME3 and T200 displaying leaf mosaic and yellowing symptoms, virus replication and symptoms are milder in TME3. However, T200 demonstrates a considerably higher number of DEGs (Table 3-2) compared to TME3 (Table 3-4) at 32 dpi which is when systemic infection becomes established Remarkably almost all of the TFs (ERFs, WRKYs, NAC and MYBs) and interacting gene partners in the network are down-regulated (Figures 3-1 and 3-4). This provides overwhelming evidence that susceptibility to SACMV is established by suppression of genes related to TF signalling of plant defence. This suppression is reflected in the overall transcriptome repression in T200 reported earlier (Allie *et al*., 2014). Some of the genes induced in susceptible tomato by Tomato yellow leaf curl Sardinia virus include several MYB and NAC domain TFs, and a NAC domain

(AT3G12977) TF was up-regulated at 6 weeks (full symptom expression) (Miozzi *et al*., 2014).

Interestingly, many of the interacting partners in this network are also found in TME3 Furthermore, much of the overlap with interacting partners between T200(32 dpi) and TME3 (12 dpi) are connected to *At*WRKY40. These recurring interacting partners give robust evidence for co-expression of genes and suggest robust interactions providing more confidence in the data as a whole. The differences were seen among common DE TFs that show contrasting expression (Figure 3-2B) appear to make up a dense highly down-regulated network are particularly interesting (Figure 3-4).

Furthermore, there are a number of genes that are DE in both T200 and TME3 at 32 dpi, however, what is most interesting is that most of these genes are down-regulated in T200 and up-regulated in TME3. Regarding the common DE TFs, it is interesting to note that several of them expression indicative of circadian clock regulation. These genes, namely *At*ERF-1, *At*WRKY40 and *At*ATAF1), have therefore been implicated in being regulated by circadian clock process(Bhardwaj *et al.,* 2011). It is therefore suggested that the initial deregulation of the circadian clock in T200 at 12 dpi, may be one of the contributors to the down-regulation of these genes. Furthermore, the fact that the orthologs of these genes are up-regulated in TME3 at the same time point and the absence of other DE genes pertaining to circadian clock regulations, may suggest that circadian clock in TME3 is not disrupted.
*At*ATAF1/NAC002 (*Me*NAC29) is expected to be down-regulated as it is induced by multiple pathways, including, JA, ABA, and wounding (Wu *et al.,* 2009b). These are all noticeably suppressed in T200, whereas JA and ABA signalling are induced, explaining the contrasting expression between the two landraces. Furthermore, Wu *et al*. (2009) further went on to note that *At*ATAF contributes to the perception of ROS and HR. This means the down-regulation of *At*ATAF1 is contributing to susceptibility predominately by suppressing ROS responsive in T200.

The disruptions to the circadian rhythm seen in T200 may be only partially why *At*WRKY40 is down-regulation. *At*WRKY40 has been characterized to be induced by ABA (Liu *et al.*, 2012) and in turn goes on to regulate almost every form of stress signalling, including ABA, ROS, JA and SA (Brotman *et al.*, 2012, Liu *et al.*, 2012). Furthermore,

*At*WRKY40 functions as a positive regulator of ETI (Schön *et al.*, 2013). All these culminate in *At*WRKY40 conferring resistance against both *P.syringae* and *G. orontii* (Pandey *et al.*, 2010, Schön *et al.*, 2013). It is an unlikely coincidence that the down-regulation of *At*WRKY40 coincides with the large increase in the number of down-regulated genes associated with hormone and defence response. It is proposed that the down-regulation of this critical hub TFs is largely responsible for the establishment of disease in T200.

The last of the common TFs regulated by circadian rhythm is *At*ERF-1/*At*ERF1A. *At*ERF-1 is a chitin responsive TF that promotes resistance through acting as a positive regulator of JA signalling (McGrath *et al.*, 2005, Libault *et al.*, 2007). In particular, *At*ERF-1 marks the critical point where JA and ET cross-talk and integration occurs, leading to resistance against invading pathogens (Berrocal-Lobo *et al.*, 2002, Berrocal-Lobo and Molina, 2004, Shinshi, 2008). Therefore, it is suggested that the down-regulation of *At*ERF-1 in T200 is responsible for the perturbations particularly in the JA and ET pathways that contribute to susceptibility.

Transitioning to other common TFs, sees the contrasting expression in *At*WRKY41 (*Me*WRKY3), *At*MYB15(*Me*MYB151), *At*MYB73(*Me*MYB165) and *At*CBF4(*Me*ERF74 and 106). Currently, *At*WRKY41 is not well characterized among literature and because of this, its exact mode of action cannot be inferred. Furthermore, its absence from the networks entails that the guilt by association principle cannot even be used to infer its function to contribute to defence signalling. Therefore, it is suggested that focus be placed on characterizing the function of *Me*WRKY3 in its involvement with conferring tolerance and susceptibility in cassava.

*At*MYB15 is regulated lignification of cells, reinforcing them to subsequent pathogen attack(Chezem *et al.*, 2017). Furthermore, Chezem *et al*., (2017) found it was found that *At*MYB15 leads to increased sensitivity towards ABA signalling during up-regulation. In contrast, *At*MYB73 shows induction by JA, SA and ET treatments and was observed to go on to mediate both SA and JA mediated defence response in Arabidopsis infected with *B. oryzae* (Jiao *et al.*, 2011). Taken together, down-regulation of both the *At*MYB15 and 73 orthologs contribute globally to the suppression of all the main hormone associated with

defence response. Furthermore, due to *At*MYB15 reducing the ABA sensitivity in T200, it is likely contributing to the down-regulation of *At*WRKY40.

Strangely, unlike the previously mentioned common TFs, *At*CBF4 is found to only have functions only relating to abiotic stress response. Specifically, *At*CBF4 has been shown to confer drought and freezing tolerance (Khan, 2011). However, it is not uncommon for such TFs to show both biotic and abiotic stress regulation, like that seen in *At*ATAF1(Wu *et al.*, 2009b). Therefore, just because it *At*CBF4 has yet to be characterized to be associated with resistance, does not discredit it from being a potential candidate for conferring to susceptibility in T200. Therefore, much like *Me*WRKY3, it is recommended that characterisation studies be performed on *Me*ERF74 and 106.

Expanding on *At*CBF4, the most striking feature seen is the number of down-regulated TF belonging to the ERF family. Indeed, all 14 TF belonging to ERF were shown to be down-regulation which suggests suppression of multiple ET response pathways. This contrasts with what is previously seen in TME3, where ET signalling is enriched among up-regulated and all ERFs are up-regulated at 32 dpi. This divergence between the two landraces suggests that ET is important in mounting a defence response against SACMV. However, it is important to note that differences between the manner in which both landraces respond to SACMV were not as apparent in 12 dpi. Therefore, it is proposed that differences seen at 32 dpi are the consequences of the small differences observed between the initial response of T200 and TME3 towards SACMV.

To conclude, TME3 shows its greatest divergences in the contrasting common DE TFs during middle infection. Furthermore, most of these genes identified as critical hubs that confer resistance through various means, including regulating and responding to hormone signals, lignification and transduction of defence pathways in response to invading pathogens. Due to these reasons, it is recommended function studies be performed on these genes to adequately quantify their role in tolerance and to further understand the role of currently uncharacterized genes like *At*WRKY41 and *At*CBF4 orthologs.

3.4.3.3 Chronic infection (67 dpi)

At 67 dpi, T200 is unable to recover from SACMV and essentially suppression of all the five TF families remains greater than up-regulation (**Figure 3-1**), The network pattern of general

high suppression of TFs and interacting partners is similar to 32 dpi which means much of the same defence pathways that are suppressed at middle infection remain so during late infection. This contributes to the continuous high virus replication and severe symptoms which persists through the lifespan of cassava T200, which is a perennial crop.

Interestingly it appears that the ortholog of *At*ERF1 in T200 at 67 dpi is initiating a different response to that found in TME3 at 32 dpi. In T200, the down-regulated PR proteins HCHIB, an endochitinase, is connected to *At*ERF1, whereas in TME3, it appears that *At*ERF1 is involved in ethylene signal transduction. Chitinases are closely associated with the PTI response of fungi as they breakdown the polymers into detectable signals to which the host can then perceive and respond to (Chisholm *et al.*, 2006). The down-regulation of *At*HCHIB would suggest that there is persistent suppression surrounding PTI pathways as two additional chitinases were observed to be down-regulated at 32 dpi.

It appears that most of the genes found in this network relate to developmental processes rather than the defense process, which would explain why many more instances of up-regulation are seen in this network.

Notably, connected to MYB83 is two proteins involved with xylem biosynthesis, Caffeoyl CoA 3-O-methyltransferase1 (CCOAOMT1) and IRX1. Due to the fact that these genes are compartmentalised in different parts of the genome (nucleus and membrane bound), direct interaction is unlikely. Although the interaction is not direct, the MYB protein appears to interact with the IRX proteins indirectly by controlling the expression of the IRX genes. It was noted that in knock-down mutants presented with slightly reduced size and compromised structure in the plant (Do *et al.*, 2007). The up-regulation of the CCOAOMT1 homolog would increase the growth of the plant, however as noted with Allie *et al*. (2014), cassava infected with SACMV show reduced size. This is indicative of SACMV further promoting growth pathways in T200 in order to exploit the resources for its own replication. Similarly, to how the leucine-rich kinases, *At*IRX1 contributes to developmental processes like xylem synthesis. Interestingly has been linked to drought tolerance in Arabidopsis, when mutated accumulated more ABA in the plant (Chen *et al.*, 2005). It is likely the up-regulation of the *At*IRX1 homologs is a secondary consequence of the other up-regulated

developmental genes, however, it may be that targeting of *At*IRX1 in cassava provides SACMV with an advantage through indirectly also suppressing ABA pathways.

This observation of proteins involved in developmental processes extends into the larger network separate to the TFs. Barely any meristem (BAM)1/MPA24.5 is appropriately named as plants demonstrate significantly reduced levels of meristem (DeYoung *et al.*, 2006). The up-regulation of this gene demonstrates that SACMV is continuing to promote its own infection at 67 dpi by increasing meristem in T200 and in doing so increase viral replication. Notably, *At*IMK2 is a leucine-rich repeat that has not been well-characterized in Arabidopsis, however, three of its homologs have been shown to be associated with secondary vascular differentiation in poplar, a close relative to cassava (Song *et al.*, 2011). Secondary vascular differentiation entails the increase of plant diameter whereby meristem tissue divides and differentiates to gives rise to new xylem cells. Although identified in poplar, the precise effect this gene has in regards is still not known. However, following the *At*BAM1 homolog expression, and the known strategies employed by geminiviruses, it is proposed that SAMCV would prevent meristem differentiation, for replicative purposes, by up-regulating the *At*IMK2 homolog.

From the unique proteins found at 67 dpi, and not 32 dpi, it is proposed that SACMV is still actively reprogramming the meristem development to suit its replicative means. Despite these proteins are not found in 32 dpi, it is important to note that the network displayed for 32 dpi, is only a partial one surrounding the common proteins of contrasting expression linked to defence pathways. As the top 10 GO-enriched biological processes demonstrated in the previous chapter, that these developmental process regarding growth are still up-regulated. Therefore, what this network is showing is that SACMV does not only manipulate growth of the plant at 12 and 67 dpi but rather only the real differences were seen in T200 from 32 to 67 dpi are in these describe processes.

3.4.4 The rise of tolerance in TME3

The development of tolerance in cassava TME3 over a 67 dpi period can be viewed as several critical molecular processes at each infection stage, that differ from T200: initiating some defenses at 12 dpi which later curtails virus load, priming of recovery at 32 dpi, followed by establishment and maintenance of tolerance at ~67 dpi. In this study, we

focused on five TF family members and their interacting gene partners associated with these SACMV infection stages.

3.4.4.1 Initiating Tolerance (12 dpi)

It is clear that TME3, despite showing a similar non-symptomatic phenotype, differentiates from T200 starting from early infection. This is evident in the differential expression of TFs (Figure 3-2), the GO-enriched hormone response profile (Figure 2-9) and the protein-protein interaction networks (Figure 3-5, 3-6 and 3-7). It is therefore hypothesized that TME3 begins initiating tolerance during early infection and key genes found in the network are likely involved in establishing tolerance.

At 12 dpi, there are two up-regulated orthologs (*Me*WRKY40 and 79) of *At*WRKY40, furthermore, orthologs of *At*WRKY40 are shown to be uniquely up-regulated through SACMV infection, implicating it establishing tolerance. *At*WRKY40 is an important molecule in immunity as it has demonstrated the ability to modulate the levels of JA and SA during the early response to *G.orontii*, deciding to either respond with JA or SA mediated immunity (Liu *et al.*, 2012). Notably, *At*WRKY40 is rapidly induced by ABA, where it localizes in the cytosol allowing the expression of ABA-responsive genes *At*ABI4 and *At*ABI5, however, if induced by SA, *At*WRKY40 suppresses these ABA genes (Liu *et al.*, 2012). This is important as it may explain why a) the *At*WRKY40 orthologs are induced uniquely in TME3 during early infection and b) why SA is shown to be repressed during 12 dpi in TME3.

While previous experiments have shown that *At*WRKY40 functions generally as a negative regulator, some genes are induced by *At*WRKY40 binding (Pandey *et al.*, 2010), this may not be the case in all hosts and all genes. In Arabidopsis, *At*WRKY40 displays direct suppression of JAZ proteins (Pandey *et al.*, 2010), which in turn, leads to the host responding to JA (Thines *et al.*, 2007, Kazan and Manners, 2012). Despite this, both homologs of *At*JAZ1 and 8 show up-regulation in TME3 at 12 dpi. It is, therefore, proposed that *At*WRKY40 orthologs may serve as positive regulators of gene expression among some members of the JAZ family. Support for this hypothesis is found in the expression data through SACMV infection. In TME3, where *At*WRKY40 orthologs are up-regulated throughout SACMV, up-regulation of JAZ proteins is observed at all three-time points,

whereas in T200, where *At*WRKY40 is down-regulated at 32 and 67 dpi, down-regulation is observed for JAZ proteins during middle and late infection. Even if JAZ proteins are induced rather than suppressed by *At*WRKY40 orthologs, it should not explain, why there is a positive response to JA in TME3 at 12 dpi. This suggests that JAZs may not function as JA response suppressors like that in Arabidopsis and may promote JA signalling. Furthermore, TME3 displays down-regulation of *At*GRX480 which is observed to help regulate JA and SA antagonisms by inhibiting JA signalling pathways (Ndamukong *et al.*, 2007). Unlike T200, TME3 only demonstrate the down-regulation at 12 dpi and not throughout SACMV infection like T200, supporting the notion that TME3 is prioritizing a JA over a SA response at 12 dpi, whereas *At*GRX480 contributes to susceptibility in T200 by suppressing SA signalling

Although *At*WRKY40 appears to play a large role in initiating tolerance, it is important to not disregard other TFs and their possible role in initiating tolerance. One of the targets for suppressed in Arabidopsis is *At*MYB73, however, in TME3, the *At*MYB73 orthologs show up-regulation, despite up-regulation in *At*WRKY40 orthologs. This would further suggest that the specific targets suppressed and induced by WRKY40 TFs are host specific and that in cassava, thereby accounting for the induction of JAZs and MYB73 in TME3. Notably, the up-regulation of *At*MYB73 was found to confer resistance against *B.oryzae* in Arabidopsis and shows signs of regulating both JA and SA defence pathways (Jiao *et al.*, 2011). This would mean that the up-regulation observed in TME3, during early infection, would like to contribute to conferring resistance towards.

Despite the up-regulation of ET-responsive genes, it is most interesting that all ERFs in TME3 are down-regulated at 12 dpi. Notably present in the network, RRTF1 has been implicated in being directly negatively regulated by *At*WRKY40 (Pandey *et al.*, 2010). This behaviour whereby ET signalling is reduced, seen in the down-regulation of DE ERFs, is reminiscent of the pattern seen at the onset of ISR in Arabidopsis during rhizobacteria colonization (Verhagen *et al.*, 2004). Although more evidence is required for ISR, this pattern, in addition to the hormone response profile shows a vague likeness to ISR, suggesting one of the modes of action in tolerance. In summary, from this TF study, it is proposed that upregulation of *Me*WRKY40 homologs and linked repression of many networks interacting partners such as RRTF1, MYB73 (**Figure 3-3**) and significant down-

regulation of ERFs (**Figure 3-1**) stand out as possible hallmarks associated with signalling or initiation of tolerance at 12 dpi. So in conclusion, it is the up-regulation of key hormone and defence responsive genes, such as *At*MYB73 and *At*WRKY40 and the prioritization of a JA response to SACMV that is associated with tolerance initiation at 12 dpi.

3.4.4.2 Priming Recovery (32 dpi)

Due to TME3 developing milder symptoms compared to T200, is it proposed that the reduction of SACMV replication is due to the beginnings of an immune response occurring at 32 dpi. And the evidence for an immune response is evident in the hormone response profile of TME3, where hormone responsive GO terms were up-regulated for ABA, ET, JA and even SA. Furthermore, the expression of the common and contrasting TFs are rather telling.

As discussed previously almost all the genes that show contrasting expression (*At*WRKY40, *At*MYB15,73, *At*ERF-1, and ATAF1) are positively associated with resistance or regulate hormone signalling. Furthermore, their role in susceptibility was further discussed by relating their expression to their various function in literature. Therefore, if most of these contrasting common TFs are leading to the suppression of various hormone and defence pathways in T200 through down-regulation, then equally their up-regulation would entail the induction of these same processes. This reinforces the notion that these genes are strong candidates for conferring tolerance in TME3, because genes that confer increased resistance in literature, also lead to increased susceptibility when their expression is perturbed in the plant. However, notably most of their proteins do no feature in the network at 32 dpi, instead, there are several other unique proteins connected to the TFs. Since the tolerance phenotype in TME3 is unique compared to T200, it's likely that some of the unique proteins are contributing to defence pathways specific to TME3, thereby contributing to resistance.

Similar but not to be confused with *At*ERF-1, *At*ERF1/ERF1B also forms a appoint where JA and ET signals are integrated and go on to induce and regulate plant defences (Lorenzo *et al.*, 2003). Furthermore, up-regulation of *At*ERF1 found that genes associated with defence and resistance were, in turn, up-regulated while genes associated with mounting a wounding response were down-regulated(Lorenzo and Solano, 2005). This would suggest that *At*ERF1, through the integration of the two hormones signals, focuses the plant's

responses towards combating pathogen invasion (biotic stress) by suppression pathways affiliated with responding to wounding (abiotic stress). Therefore, it is suggested that the up-regulation of *At*ERF1 orthologs in TME3 is contributing to tolerance by initiating defence pathways from JA and ET signals in a similar manner to *At*ERF-1.

Although ET responsiveness is controlled mainly by the ERF TF family, it should be stated that this family is not involved in the perception of ET. Present within the network is two ET receptors, *At*ETR1 and 2 (Sakai *et al.*, 1998). Since ET is a modulator of immunity in plants, the reduced ability to perceive ET, experienced through down-regulation, expectedly leads to increase susceptibility in Arabidopsis (O'donnell *et al.*, 2003). Furthermore, it was shown that *At*etr1 mutants lacked jasmonate-derived defence against *P.syringae* and prevented the induction of ISR (Pieterse *et al.*, 1998a). This shows that the increased expression of both these ET receptors is leading to the increased perception of ET in TME3, which would, in turn, increase the induction of multiple ET response pathways, thereby conferring resistance through potentially ISR pathways.

Transitioning from ET to JA signalling in TME3 sees the up-regulation of Jasmonate resistant 1 (*At*JAR1). This protein is critical for pathogen defence as it is the main enzyme that conjugates isoleucine to JA forming JA-Ile , a form of JA that once perceived by coronatine insensitive 1 (COI1) produces a greater and more rapid response than traditional JA (Suza and Staswick, 2008). Unsurprisingly, due to this, *At*JAR1 has been shown to be one of the required for resistance in Arabidopsis against *B.cinerea* (Korolev *et al.*, 2008). Furthermore, *At*JAZ1 has been shown to be required in ISR induction in Arabidopsis against *P.syringae* (Pieterse *et al.*, 1998a). From these functions it is clear that the *At*JAZ1 is conferring increased resistance in TME3 by facilitating JA defence process, furthermore, the together with *At*ETR1, *At*ERF-1 further supporting the hypothesis of ISR in TME3.

Continuing with JA signalling, OPC-8:0 CoA Ligase 1 (OPCL1) is one of the enzymes used in JA biosynthesis, however, is not the only ligase involved in JA synthesis, meaning that JA production is reduced but not abolished in *At*opcl1 mutant lines (Koo *et al.*, 2006). Furthermore, *At*OPCL1 is found to respond to wounding that with JAR1 leads to an increase in the biosynthesis of JA-Ile , (Chung *et al.*, 2008, Suza and Staswick, 2008). However, *At*OPCL1 is not exclusively induced by wounding, it is also found to be up-

regulated in response to other PAMPS, such as chitin and flg22 (Li and Hu, 2015). This implicates *At*OPCL1 in conferring resistance to TME3 through up-regulation will lead to an increased level of JA, and consequently JA-Ile, which in turn facilitates and strengthens JA signalling in TME3 required to mount a systemic response.

There is evidence, given the up-regulation of *At*JAZ1 and *At*JAR1, that there is increased levels of JA and JA-Ile in TME3, which appears to be the reason for the induction of *At*MYC2: the master regulator of JA responses. *At*MYC2 is the target of suppression by JAZ proteins, that after undergoing degradation upon JA signal transduction, allows for *At*MYC2 to bind and transcribe JA responsive genes(Chini *et al.*, 2009). In addition to the JAZs, *At*MYC2 has been shown to be regulated by the circadian clock (Shin *et al.*, 2012). Induction of *At*MYC2 is associated with global transcriptional changes that go on to regulate hormone cross-talk and defence responses (Dombrecht *et al.*, 2007, Kazan and Manners, 2013). Interestingly, *At*MYC2 is required for ISR, not only to allow root colonization by the beneficial bacteria but also is required in ISR signalling and defence induction (Pozo *et al.*, 2008). Therefore, the up-regulation of *At*MYC2 experienced in TME3 is indicative of a healthy circadian clock in TME3, the induction defence responses. Furthermore, the induction of *At*MYC2 further supports the hypothesis that ISR pathways are stimulated in TME3 at 32 dpi, suggesting that ISR is the main mode of action of tolerance in TME3.

3.4.4.3 Establishment of Tolerance (67 dpi)

Tolerance at this time point is associated with lower virus replication and absence or mild symptoms. This is associated by both up and down-regulation of MYB, WRKY and ERF TFs which suggests that the effects of these TFs and their interacting partners on downstream responses provide a balance between processes needed for virus defence and replication, and host counter-defence to reduce fitness costs and co-exist with the virus.

Most notable are, *At*MYB15 and *At*WRKY40 orthologs that remain up-regulated at late infection. It is known that the functions of MYB15 are to increase ABA signalling sensitivity and lignification of plant cells (Chezem *et al.*, 2017). Considering that lignification is a product of priming in hosts undergoing ISR, the homologs of MYB15 appears to be good candidates involved in priming (Heil and Bostock, 2002). Furthermore, while ABA is traditionally associated with abiotic stress, it was previously stated how

*At*WRKRY40 orthologs are induced by ABA and are the likely reasons why they were uniquely induced in TME3 and not T200. It is proposed that through these two mechanisms, MYB15 both actively and passively contributes to tolerance by actively protecting against pathogens through the lignification, but also passively through an increase in the plant's sensitivity to ABA allowing for a greater and more rapid response. This increased ABA response in TME3 is likely the reason for the continued up-regulation of *At*WRKY40, whereupon subsequent viral infection, it is likely to lead to the induction of the already up-regulated TFs. Since *At*WRKY40 has been shown to regulated JA response (JAZs) and biosynthesis (*At*OPCL1), it is proposed that *At*WRKY40 is responsible for the continued up-regulation of JA biosynthesis forming the foundation tolerance during late infection.

3.4.5 The systemic response in cassava

3.4.5.1 Systemic Acquired Resistance

\Most pathways surrounding plant defence usually, effector-triggered immunity (ETI), elicit a SAR response, making this pathway highly studied, particularly in Arabidopsis. SAR is traditionally induced to combat hemi/biotrophic fungal or bacterial pathogens (Sticher *et al.*, 1997). Despite viruses not technically forming part of this group, viruses do share more similarities with biotrophs than necrotrophs, such as requiring a living host cell. For this reason, it is predicted that cassava would undergo a SAR response to combat SACMV. However previously, it has been demonstrated that specifically in T200, SA shows the greatest degree of suppression, while in TME3, most hormone-responsive genes are not linked to SA when infected with SACMV. This begs the question to whether downstream SAR is even induced in either T200 or TME3 during SACMV infection, even though salicylic acid is detected in these cassava landraces. In particular, SAR is most closely associated with *NPR1*, whose induction is absolutely critical as *AtNPR1* is the protein that initiates transcriptomic reprogramming in response to SAR signals in Arabidopsis (Cao *et al.*, 1997). Reviews around the functions and interactions surrounding *At*NPR1 have led to the well-received proposed signalling pathway of SAR in Arabidopsis (Kunkel and Brooks, 2002, Durrant and Dong, 2004).

SAR induction can be separated into two parts, genes that regulate *At*NPR1 through SA and genes that are induced by *At*NPR1 leading to disease resistance. Regarding the

perception of SAR signals, it has been suggested through mutation studies that *DIR1* is necessary for inducing systemic resistance (Maldonado *et al.*, 2002). This protein shows high similarity the lipid transfer proteins (LTP) which are members of the PR-14 group (García-Olmedo *et al.*, 1995). Interestingly, four different cassava genes relating to LTP1,3 and 5 (AT2G38540, AT5G59320, AT3G51600) were all down-regulated over the course of SACMV infection, suggesting that SACMV is actively suppressing the mode of action of initiating systemic signalling in the SAR pathway in susceptible T200. This critical feature is notably absent in tolerant TME3. Examination of the genes that regulate SA biosynthesis in Arabidopsis, the most notorious genes are *SID2* (AT1G74710) and *EDS5* (AT4G39030) (Wildermuth *et al.*, 2001, Nawrath *et al.*, 2002). These two gene-encoded proteins, despite not directly contributing to SAR, are needed in SA biosynthesis which is required for SAR induction (Ferrari *et al.*, 2003). Controlling *SID2*, *EDS5*, and SA signalling, are *EDS1* (AT3G48090), *EDS4* (AT3G52430) and *PAD4* (AT3G52430) (Glazebrook, 2001). Interestingly, none of the cassava homologs for these genes that initiate and control SA synthesis is differentially expressed in either T200 or TME3, and hormone-responsive genes in T200 all show down-regulation.

However, in TME3 there are up-regulated genes that are responsive to SA signalling (*At*JAZ1,*At*WRKY40,*At*MYB73,*At*RLK,*At*MPK4) during middle infection, despite other evidence presented herein which shows SAR is not a major player in TME3 defence. This may suggest that the above-mentioned DE genes involved in SAR signalling are either induced via a SA-independent pathway or more likely demonstrate the overlapping nature of hormone signalling pathways and are induced by another hormone such as JA. However, it is notable that no homologs of *At*NPR1 are found to be DE in T200 or TME3 through SACMV infection. The expectation would be that if *At*NPR1 was induced, that many of the downstream genes would show DE in either T200 or TME3. These genes would typically reside in TGA and WRKY TF families as well as genes falling under the pathogenesis-related PR-1 group (Kinkema *et al.*, 2000, Weigel *et al.*, 2001, Yu *et al.*, 2001). Notably, none of the cassava homologs of *AtPR-1* showed any sign of differential expression in T200 or TME3 post SACMV infection. While *At*NPR1 modulates SAR through TGA and WRKY TFs, it was also determined that NPR1 directly influences secretory pathways of PR proteins

(Wang *et al.*, 2005). However, even if *At*NPR1 does not show differential expression in either landrace, the downstream genes associated with *At*NPR1 should not be overlooked.

The pathways surrounding SA synthesis which then goes on to induce *At*NPR1 is well characterized in literature (Dempsey and Klessig, 2017). In this pathway, genes such as enhanced disease susceptibility 1 (EDS1), phytoalexin deficient 4 (PAD4) and non-race-specific disease resistance 1 (NDR1) is required for SA accumulation which would inevitably trigger(Dempsey *et al.*, 2011). However, none of the homologs connects to these three genes show differential expression in either T200 or TME3. This means that although neither of the two landraces demonstrates suppression of the genes involved in SA accumulation, it also means that there is no evidence suggesting the induction of these genes. However, it important to note that *At*MPK3 is a suppressor of SAR, which means its elevated expression in either landrace would entail a direct effort to suppress this immune pathway. This is found in TME3 at only 32 dpi, confirming that TME3 specifically suppresses SAR.

Despite the lack of evidence surrounding the differential expression of SAR hallmarks. The previously discussed results certainly support the notion that SAR does not occur in T200 as susceptibility to SACMV persists, and nor does it play a dominant role. However, in TME3, the up-regulation of the *At*MPK3 homolog at 32 dpi, in addition to the absence of evidence implicating SAR through SACMV infection, suggests that not only is SAR absent in TME3 but it is suppressed too. The recovery phenotype of TME3 supports the notion that some form of systemic resistance involving other molecular mechanisms is induced and persists in long term tolerance to SACMV in the perennial cassava TME3.

3.4.5.2 Induced Systemic Resistance

There is currently no literature available linking viral tolerance or resistance to the induction of an ISR-like response. Current evidence has suggested a role for RNA silencing (Ghoshal and Sanfaçon, 2015) or R genes (Bengyella *et al.*, 2015), and ISR has only been associated with rhizobacteria (Ku, 1982). However, notably, ISR induced by rhizobacteria, or their extracts can induce resistance to plant viruses, such as cucumber mosaic virus (Ryu *et al.*, 2004). Despite this, it is difficult to ignore the connection between ISR and transcriptomic alterations occurring in TME3 in response to SACMV infection.

The first instance of an association between ISR and cassava mosaic disease CMD can be found in the peroxidase gene family. In 2016, a genome-wide association study performed on 6128 cassava breeding lines found that between 30-66% of resistance towards CMD was conferred from a single region on chromosome 8 (now chr12, cassava 6.1 genomes) (Wolfe *et al.*, 2016), which is known as the CMD2 locus (Akano *et al*., 2002). Furthermore, this study revealed three candidate genes within the region, two of which were peroxidases (cassava4.1_029175m.g and cassava4.1_011768m.g), and thioredoxin. One of these peroxidases, namely cassava4.1_011768m.g is seen to be up-regulated in T200 at both 32 and 67 dpi (Wolfe *et al.*, 2016). However, upon closer examination, it is suspected that SACMV is up-regulating these selected peroxidases to further its own means by increasing infection in T200.

Rare cold-inducible gene 3 (RCI3) ,the Arabidopsis homolog of cassava4.1_011768m.g, has been shown implicated in the regulation of mobile RNA silencing (Liang *et al.*, 2014). This observation appears at first paradoxical, considering that enhanced peroxidase activity is linked to general enhanced resistance and ISR induction in cucumber in infected with *C. lagenarium* (Ku, 1982). In contrast, it was proposed that the peroxidase when up-regulated leads to increases H_2O_2 radicles leading to depolymerisation of the cell wall leading to an increase in plasmodesmata size whereby RNAi signals may travel (Liang *et al.*, 2014). Given the fact that plant viruses are known to spread using the plasmodesmata, it is proposed that SACMV may exploit the up-regulation of peroxidase (cassava4.1_011768m.g) to increase its mobility and spread of disease within the plant, as also proposed by (Epel, 2009).

Alternatively, since this particular peroxidase associated with CMD2 has not been shown experimentally to contribute to resistance to CMD, a number of other peroxidases linked to ISR may be involved in tolerance in TME3, requiring examination of all other DE peroxidases.

When scrutiny is placed on the number of DE peroxidases, it was found that there are 15 and 6 identified peroxidase homologs in T200 and TME3, respectively. Almost half of those identified in T200 (7/15) are down-regulated, whereas all the identified peroxidase homologs in TME3 were up-regulated at all infection stages (Allie *et al*., 2014).

Interestingly, up-regulation in peroxidases seen in TME3 occurred from early infection at 12 dpi, while DE of peroxidases in T200 was only detected at 32 dpi. Peroxidases contribute to plant defence by generating H_2O_2, forming the basis of ROS in plants (Kawano, 2003). This oxidative burst is known to contribute to resistance to a wide range of pathogens by activating ROS dependant defence pathways (Bindschedler *et al.*, 2006). In addition to producing these signalling molecules, peroxidases are associated with cell wall strengthening (Sudisha *et al.*, 2012). Interestingly, up-regulation of peroxidases of the *At*MYB15 ortholog, at 32 and 67 dpi is observed in TME3 suggesting that peroxidases, with *At*MYB15, contribute to facilitating cell wall defence against subsequent infections. However, given that most of the peroxidases are up-regulated at 32 dpi, this contributes to mounting evidence that there were elevated levels of ROS signalling occurring at middle infection. Previously in chapter 2, it was noted that *At*HsfA4A is a sensor for H_2O_2 and at 32 dpi demonstrates up-regulation in TME3, but down in T200. This suggests that ROS is important at middle infection form and that the suppression of it would contribute to susceptibility. In contrast, the elevated levels of H_2O_2 produced from the peroxidases in TME3, coupled with the increased perception by *At*HsfA4A would likely lead to the increased induction of ROS dependent pathways specifically during middle infection. In conclusion, although enhanced peroxidase activity does contribute to resistance, it appears that the proportional levels (up- versus down-regulation in peroxidases) are important, in addition to which genes are being induced that are important in tolerance in TME3. This provides strong evidence that peroxidases may play initiate an ISR-like response early in TME3, but not in T200.

Allie *et al*. 2014 previously commented on the overlying hormone response found in T200 and TME3, stating that both cassava landraces prioritize JA and ET over SA. However, when scrutinized further in T200 it was shown that hormone signalling was suppressed across ABA, ET, JA and SA signalling pathways. Furthermore, TME3 displayed a predominately JA response, inducing the highest number of interacting partner genes and ET was the second largest inducer of genes, exhibiting a dramatic increase at 32 dpi (Figure 2-9). ISR shows a similar pattern regarding hormone signalling. The predominant feature of ISR in contrast to SAR is that ISR is predominantly associated with JA and ET signalling (Pieterse, 1998). However, SA signalling has also been found to contribute to ISR (Malamy

et al., 1990, Métraux *et al.*, 1990) which is demonstrated in TME3 at 32 dpi (Figure 2-9), but fewer genes compared to either JA or ET are induced. This hormone response profile in TME3 favours the hypothesis that TME3 predominantly, but not exclusively, exhibits an ISR-like response rather than a SA-cantered SAR response. When considering the biology of cassava, it is unsurprising that cassava would prioritize JA and ER signalling over a SAR orientated response. Effective SAR is more than often the consequence of a localized hypersensitive response (HR), whereby systemic signals are released from the point of primary infection (Ryals *et al.*, 1994), and has largely been described in annual or domesticated/cultivated crops. However, it is important to note that cassava landraces are unable to initiate the HR, which would necessitate an alternative signalling system such as ISR.

The second link between ISR and SACMV tolerance can be seen in the level of gene expression of PR genes It has been reported that ISR does not necessitate the expression of PR proteins whereas SAR does (Pieterse *et al.*, 1998b, Van Loon *et al.*, 1998, Hammerschmidt, 1999). PR proteins represent a group of proteins that are induced during pathogen infection that may or may not contribute towards perception and defence against the invading pathogen. These proteins can fall into one of 14 categories encompassing broad protein families such as chitinases thionins, lipid transfer proteins, peroxidases, to name a few (Sels *et al.*, 2008). Interestingly, of these specific PR protein categories, including the previously mentioned peroxidases, only eight transcripts were found differentially expressed in TME3, however, only one (cassava4.1_017915m.g) was found to be a homolog of a "pathogenesis-related protein" in Arabidopsis. This cassava PR-homology was observed to be down-regulated at 32 dpi. Whereas, in T200, forty DE expressed genes belonging to the PR protein categories were identified throughout SACMV infection. Of these, 16 were DE homologs to genes annotated as "pathogenesis-related protein" in Arabidopsis, of which 87.5% (14/16) of these genes showed up-regulation in response to SACMV in T200. With regard to these results, two different interpretations may be offered. Firstly, based on the known association of SAR and PR-proteins, the low number of expressed annotated PR gene homologs in TME3supports the notion that TME3 lacks dominant SAR-related processes in comparison to ISR-like characteristics. In contrast, in T200 up-regulation of a large number

of PR-related genes, while potentially associated with a downstream basal defence response, is not effective, as severe SACMV susceptibility is a hallmark of T200 (Allie *et al*., 2014) and SAR is not ultimately established. Indeed, in SACMV-infected Arabidopsis, SA and PR-related genes were shown to be induced, but this was not sufficient to induce SAR (Pierce *et al*., 2013). Cross-talk (both concomitant antagonism or synergism) between different hormone responses were shown to play a role in the final susceptibility outcome in SACMV-infected Arabidopsis, and could also be involved in fine-tuning cassava responses to SACMV. In conclusion, while TME3 does show accumulation of PR proteins, the numbers significantly lower in comparison to T200, where induced PR proteins numbers are five-fold greater than that of TME3, and evidence for SAR in TME3 is not supported. Additionally, many of the genes (*At*WRKY6 and 40, ETR1, ERF-1, JAR1, and MYC2) associated with ISR are found in the SACMV-responsive interacting network at 32 dpi in TME3 (Figure 3-6).

Further support for a role for ISR in SACMV tolerance is provided by previous experiments investigating ISR, which have implicated specifically *At*WRKY6 and 40, *At*JAR1, *At*ETR1, *At*MYC2 and *At*ERF1 (Pieterse *et al.*, 1998a, Verhagen *et al.*, 2004, Pozo *et al.*, 2008, Brotman *et al.*, 2012). In 2008, Pozo *et al*. demonstrated that *At*MYC2 was up-regulated consistently during ISR induced Arabidopsis challenged with *P. syringae*. This is understandable considering the central role JA plays in ISR, and because of this, the involvement of MYC2 can almost be considered critical. Furthermore, a transcriptome analysis of ISR in Arabidopsis (Verhagen *et al*., 2004) revealed two transcription factors that were also found in the TME3 network at 32 dpi, namely *At*WRKY6 and *At*ERF1 (AT3G23240). *At*ERF1 was shown to be down-regulated after 24 hours in Arabidopsis after inoculation of the rhizobacteria (Verhagen *et al.*, 2004), that led the authors to suggest that ET reduction could be indicative of ISR onset. In cassava eight out of the nine identified ERFs showed down-regulation at 12 dpi in TME3. An *At*WRKY6 homolog was also found to be up-regulated in TME3 at 32 dpi where tolerance is primed and may be associated with ISR persistence. Another WRKY TF that showed up-regulation in ISR was *At*WRKY40 (Brotman *et al.*, 2012). In this study, the authors looked at the transcriptional changes of selected genes in both plants inoculated with only the rhizobacteria and plants inoculated

with both the rhizobacteria and *P. syringae*. In both of these instances, *At*WRKY40 was shown to be up-regulated, but less so in plants that lacked the rhizobacteria. This affirmed the notion that WRKY40 was closely linked to ISR to *P. syringae*, and supported the observations that tolerant plants underwent reduced levels of transcriptional changes. (Brotman *et al.,* 2012). Significantly, in TME3, *At*WRKY40 homologs were up-regulated, and persisted, at all stages of infection, but not in susceptibleT200, which provides strong evidence for ISR-like mechanisms in SACMV recovery and tolerance in cassava TME3 landrace. Pieterse *et al*. demonstrated in 1998 that through knockout experiments that *AtJAR1* and *AtETR1* lie upstream of *NPR1* and are required for a SAR-independent systemic response seen with ISR, confirming the role of both these genes in the ISR pathway and further showing the contrast between ISR and SAR (Pieterse *et al.*, 1998b) Both *AtJAR1* and *AtETR1* homologs were up-regulated in TME3 at 32 dpi, providing further support for our hypothesis that a SAR-independent systemic response such as ISR is responsible for CMD tolerance in cassava. While ISR to date is described as enhanced resistance due to the presence of growth promoting rhizobacteria (Hammerschmidt, 1999), this is not the case in TME3 as tolerance to SACMV which appears to be heritable ((Allie *et al.*, 2014, Bengyella *et al.*, 2015) and not specifically dependent on rhizobacteria for induction. However, there is a noticeable absence of studies in the literature to dismiss the possibilities of other "priming" events and ISR induction in other plants. This study provides strong support that ISR-like responses may not only be important in TME3. It is therefore proposed that TME3 is able to independently initiate and induce ISR-like pathways common to reported ISR. Critically, continued low levels of SACMV replication at recovery and tolerance at later stages in cassava TME3 may continue to prime tolerance-associated ISR molecular mechanisms which persist in this perennial crop. The ISR-like model for tolerance proposed in this study (Figure 3-8**)** is supported by the recent recognition that tolerance is considered rather as long term co-existence between pathogen and host (Bengyella *et al*., 2015) and a balance between virus and host defence and counter-defences (Gaudel and Sanfaçon, 2018). The ISR model seems to fit this "co-existence" notion. Naturally, other complementary mechanisms are not dismissed as playing an additional role in cassava, such as PTI-associated RNA silencing (Rogans *et al*., 2016) and ETI-NB-LRR resistance (Bengyella *et al.*, 2015). This is the first

study to date, to our knowledge, which reports ISR-like responses in a plant undergoing a viral infection.

3.4.6 Proposed ISR model in TME3

Based on the network seen in cassava TME3 at 32 dpi, a proposed model for the specific signalling pathways associated with an ISR-like response may be drawn (Figure 3-8). Following the ISR paradigm, the two main branches that are encapsulated in the network are ET and JA. It has been shown that ET signalling is at its highest during middle infection, whereby the ethylene receptors ETR1 and ETR2 perceive the signal and initiate a defence response through both ERF1 and ERF-1. During this time, JA biosynthesis is increased through up-regulation of LOX3, OPR2 and OPCL that have direct functions in JA biosynthesis which is in part mediated through the TME3 homologs of both *At*WRKY6 and 40. The JA signal is perpetuated alongside the two kinases (RLK7 and AT5G25930) linked to both these *At*WRKY6 and 40 TFs, which leads to downstream defence responses. This JA signal from *At*WRKY6 appears to be focused on the EP3 chitinase. Since *At*WRKY6 is associated with priming (Verhagen *et al.*, 2004), it is possible that *At*WRKY6 is achieving this through up-regulating EP3 to become more perceptive to other invading fungal pathogens. Despite JAZ1 being up-regulated in 12 dpi, it is proposed that the up-regulation of its target MYC2 during middle infection implies that MYC2 requires signal synergism between both ET and JA to be induced. MYC2 activates JA-responsive genes which can contribute to ISR. Further evidence for JA playing a central role in tolerance to SACMV can be provided by a study where geminiviruses inhibit jasmonate signalling in Arabidopsis in order to establish infection (Lozano-Duran *et al*., 2011). This is achieved by the geminivirus-encoded C2 protein which subverts ubiquitination by altering COP9 signalosome complex (CSN) mediated derubylation of SCF E3 ligase complexes (Lozano-Duran *et al*., 2011). As SCFs are key regulators of many cellular processes, interference provides a strategy for geminiviruses to establish infection. Although *At*MYB15 is predicted to interact with *At*WRKY40, the exact connection has not presented itself in the literature yet. However, the up-regulation of both the MYB15 and WRKY40 homologs at late infection implies that they are both necessary with the priming associated with ISR, which can be viewed as tolerance.

Additionally, MYB15 may also fine-tune defence through enhancing ABA sensitivity and increasing lignification, thereby contributing to priming recovery (Chezem *et al.*, 2017).

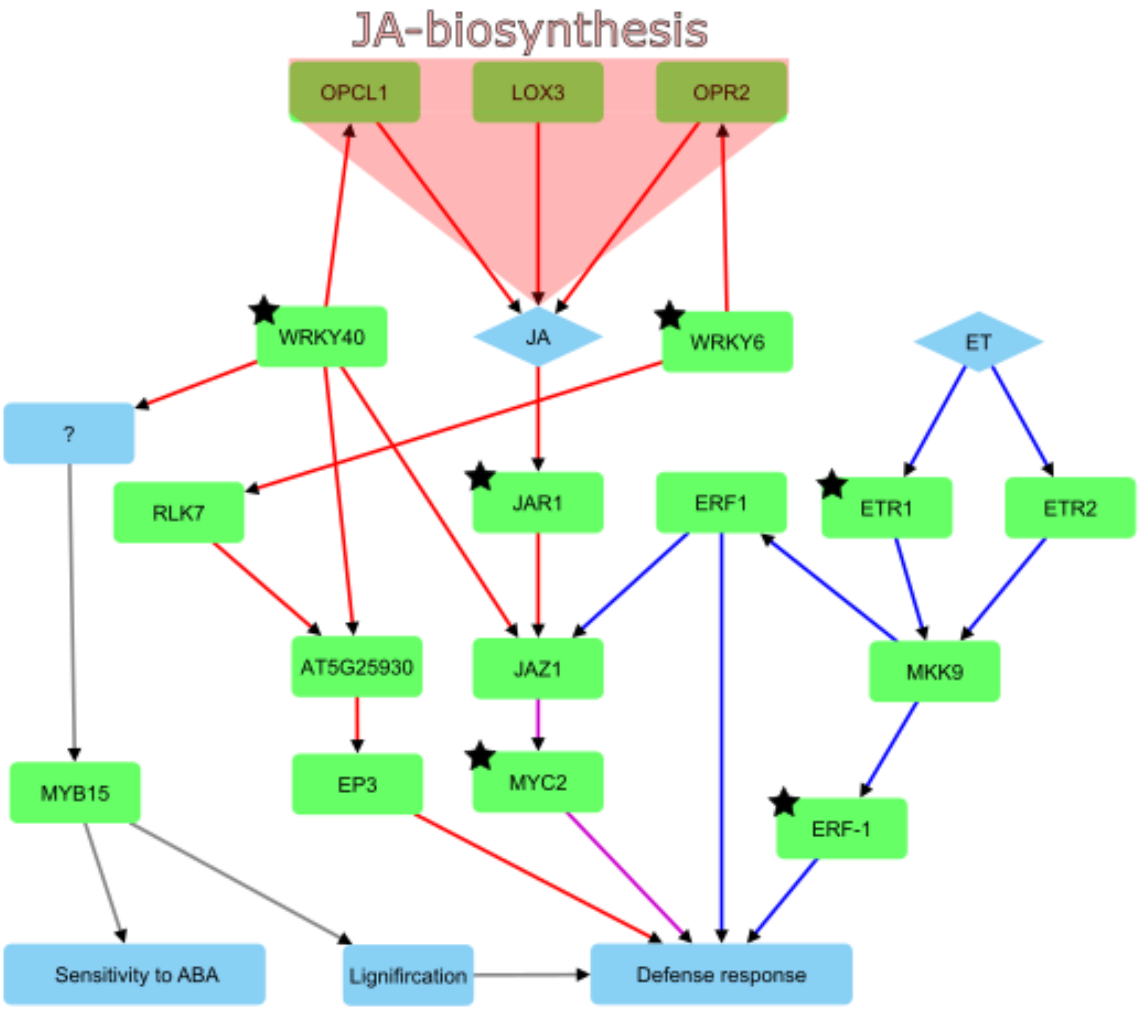


Figure 3-9: **The proposed model of the network in TME3 at 32 that induces an ISR-like response which leads to SACMV tolerance**. Genes shown in green boxes represent up-regulation of the cassava homologs in TME3 at middle infection. Diamond blue nodes represent hormone molecules, while the different colour edges represent the transduction of signal associated with each hormone signaling pathway (red for JA signaling, blue for ET signaling, and purple representing the intersection of both signals). Stars represent not TFs but rather proteins that have been previously described to be associated with ISR.

3.4.7 Transcription factors in literature

As previously discussed in the last chapter, there is value to be gained from determining if whether the identified TFs are implicated in other tolerant/resistant or susceptible plants. Notably, bZIP TF has been absent in the most part regarding contributing to defence networks and in the transcriptome data in general. This appears to also be the case in other transcriptomic studies where the very little pattern is seen among the low number of identified TFs belonging to this family. It can be concluded that the role played by the bZIP family in immunity appears to a minor one if at all.

In this data, MYB15 orthologs have been implicated in positively regulated tolerance/immunity, and it was the most commonly identified MYB TF among TF dataset. Out of the 6 instances of identification, it was showed up-regulation in two (CBSV and TolCNDV), which would further support its candidacy as a gene involved in resistance. In contrast, MYB15 orthologs were found the most after MYB73 homologs (which were not seen in the networks). However, from the four instances of identification, very little can be concluded as it shows equal instances of up- and down-regulation. This may suggest that orthologs of *At*MYB15 are more associated with susceptibility that resistance in other plants. However, in cassava, the contrast in expression seen between T200 and TME3 would lead one to believe that the orthologs of *At*MYB15 are positively linked to resistance against SACMV. Despite *At*MYB73 orthologs showing more down-regulation in tolerant/resistant data sets, it was still notably the most common identified MYB, it is believed to be a candidate hub for tolerance due to its contrasting expression in T200 and TME3.

Most interestingly, the most common NAC ortholog identified among susceptible datasets was a TF completely absent in T200, *At*anac036. Furthermore, in comparison of other datasets, NACs appear to generally under-represented among DEGs in cassava. Whereas in tolerant datasets, the two most common NAC TFs were *At*NAC002 followed by *At*NAC90 where they both show an up-regulation most of the instances of identification. However, there were no sufficiently high confidence interactions meaning they were unable to be placed into networks. In order to comment on the possible association between both TFs more functional characterization studies are required.

Much like both *At*NAC002 and 90, the most common ERF orthologs found in susceptible datasets was *At*ERF025 which was not shown in any networks. Most interestingly, out of the seven instances of identification among susceptible dataset, all of them showed down-regulation. This is compelling evidence for the association between the orthologs of *At*ERF25 and susceptibility. Due to this, it is strongly urged that focus is placed on the orthologs of *At*ERF25 in cassava for future studies regarding immunity. Showing equal number to *At*ERF25, the orthologs of *At*ERF105 were the second most common ERF identified. Following the expression seen in T200, most of the instances depicted down-regulation of the orthologs with only two out of the seven times datasets showing up-regulation (UCBSV and SACMV in Arabidopsis). As mentioned earlier that *At*ERF105 is currently and only recently associated with cold tolerance(Bolt *et al.*, 2017). However, these results would suggest that it is also involved with biotic tolerance too, and like the orthologs of *At*ERF25, it is also recommended that emphasis is also placed on *At*ERF105 orthologs in cassava in future research. Notably, compared to susceptible datasets, there were a proportionally fewer number of *At*ERF orthologs identified.

This would suggest that it is more necessary for viruses to target ERFs in order to establish and maintain infection than it is for ERFs to be involved with a defence response. This is not surprising as ET is used more as an immune modulator and not the principle hormone for immune response. Interestingly, although *At*ERF105 was only the second most common of ERFs TFs, it is notable that it shows up-regulation in all three instances. This shows a dichotomy in the expression for *At*ERF105 homologs in susceptible and tolerant/resistant datasets, thereby reinforcing the recommendation that focus should be placed on the orthologs for future studies. Despite the importance suggested in the up-regulation of both *At*ERF1 and AtERF1 orthologs, *At*ERF-1 was only found in three instances and displayed down-regulation in all except for SACMV in TME3. The reason for this may be that *At*ERF-1 is specific to TME3 which is inducing an ISR-like response, which might not be the means of resistance in other plants to other viruses.

Chapter 4

Summary

In this dissertation, in chapter 1, a background literature review was presented, while chapter 2 focused on a particular unique TF family in plants, namely WRKYs, associated with stress response. Chapter 3 focused on four additional TF families, in addition to WRKYs, associated with pathogen responses in plants. In this study, the mechanisms that contribute to tolerance and susceptibility, in relation to stress-responsive TFs and hormones, in cassava, in response to SACMV, was investigated. Transcriptomic data from an earlier study (Allie *et al.*, 2014) was obtained from a susceptible landrace (T200) and a SACMV-tolerant landrace (TME3), at three stages of infection. These landraces thereby served as model systems to compare host responses. This study successfully identified the differentially expressed TFs and then went about placing them into biological networks, in order to give context to both how they interact with one another, and how these genes may contribute to the global picture of tolerance or susceptibility. Furthermore, protein-protein networks were constructed around the TFs of interest and their DE interacting partners. In doing so, this investigation has unveiled novel information surrounding the mechanisms involved in the development of tolerance, and the nature of systemic resistance in TME3 in response to SACMV. This study showed, for the first time, that CMD susceptible and tolerant cassava landraces respond differently to SACMV, and that several TFs and hormones are uniquely expressed in each landrace, over the course of infection. Identification of DE TF interacting genes and hormones revealed the importance of these networks, in relation to the development of susceptibility and tolerance. T200 displays many of the traditional hallmarks of viral infection, including a high degree of transcript suppression, and follows much of the understood paradigms regarding successful viral invasion. In contrast, TME3 responds differently to virus infection, and in doing so, displays tolerance. In summary, Induced Systemic Resistance (ISR) is proposed to drive tolerance to SACMV in TME3, and a proposed model for tolerance in relation to TFs and hormones is presented.

4.1 T200, hallmarks of susceptibility

Overall, both the WRKY and expanded networks, containing the other four TFs, showed a disproportionately high number of down-regulated genes, implying SACMV is suppressing targeted TFs involved in hormone signalling, and consequently defence responses. From this data, it is proposed that SACMV establishes infection by suppressing basal immunity and hormone signalling. Evidence supporting this can be seen in the downregulation of the two hub proteins, *At*WRKY53 and 70, which are responsive to the two main immune regulators: SA and JA (Li *et al*., 2004, Miao and Zentgraf, 2007). In the case of cassava TME3, JA rather than SA appears to play a major role in tolerance. Additionally, the suppression of multiple homologs belonging to the ethylene responsive factor (ERF) family (*At*ERF5, *At*ERF6 and AT5G1190) would further hamper T200's ability to mount a defence response, as ET is a modulator of this immune response (Pieterse *et al*., 2009). SACMV interferes with the ability of T200 to respond to hormones. Furthermore, in T200 resistance conferring genes are suppressed during SACMV infection. Particularly, *At*MPK3, one of the critical signalling molecules in the PTI pathway (Nühse *et al*., 2000, Meng and Zhang, 2013). In addition, homologs of *At*ATL2 have been characterized as initiators of early defence (Serrano and Guzmán, 2004). The down-regulation of these genes supports the notion that T200 is unable to initiate an adequate early defence response, due to SACMV suppressing these defense conferring genes.

T200 in addition to suppressing basal defence, also SACMV equally prioritizes reprogramming a vast number of developmental processes for its replicative requirements. Support for this argument can be found in the fact that the top GO biological processes enriched among the up-regulated genes in T200 at 12 dpi were classed as regulating developmental function, specifically pertaining to growth and meristem development. A well-studied phenomenon in geminiviruses is the requirement for actively dividing meristem in order to replicate, which was seen with *Cabbage leaf curl virus* in Arabidopsis (Ascencio-Ibáñez *et al*., 2008). Therefore, SACMV may be up-regulating genes involved in meristem growth and development, in order to optimize its replication. Clearly at 12 dpi, in order to optimize infection success, SACMV is required to suppress host basal defence and optimize its replicative environment.

During the middle stages of infection (32 dpi), the presentation of symptoms is due to systemic movement of SACMV, a higher viral load, and concomitant elevated transcriptional reprogramming. From the TF-interacting protein-protein networks, considerable increases in the number of DEGs was apparent. A high level of suppression was observed among hormone-responsive genes. Notably, while previously at early infection the majority of DEGs were up-regulated, the presentation of symptoms corresponds with a shift towards a high percentage of down-regulated DEGs. These results suggest that at 32 dpi, SACMV is displaying a shift from disease establishment to disease maintenance, through increased suppression aimed at defense-related pathways. Evidence for this proposal can principally be found in the GO enrichment data, as well as the interacting TF networks. There, biological GO enrichment results display many of the same defense-related processes and hormone signalling, however, there is a noticeable increase in the gene count. This suggests that there is a greater degree of suppression occurring in these pathways. The culprits that appear to be responsible for this increased number of down-regulated DEGs are the orthologs of the two hub TFs: *At*WRKY33 and 40. In particular, *At*WRKY33 demonstrates the highest degree of connectivity and has been shown to confer resistance through modulation of defence genes, implicating homologs of *At*WRKY33 in susceptibility in cassava T200 (Alves *et al*., 2014).

Interestingly, down-regulation of *At*MYB15 at 32 dpi was noted in T200. *At*MYB is involved in ABA sensitivity, camalexin biosynthesis, and response to wounding in plants, and up-regulation leads to lignification of cells (Chezem *et al*., 2017). Camalexin, while contributing indirectly to defence pathways, also contributes to insect resistance (Alves *et al*., 2014, Kettles *et al*., 2013). Down-regulation of *At*MYB15 at 32 dpi may enable easier feeding by whitefly in the field, and subsequent transmission of geminiviruses such as SACMV. SACMV was also shown to suppress wounding as evidenced in the GO-enriched biological process where homologs of *At*STZ, *At*TET8 and *At*GRXC9 showed down-regulation in T200. Additionally, suppression of 62 different gene homologs that contribute to PCD processes in T200 at 32 dpi was observed. PCD would contribute to virus restriction, and suppression therefore which would favour movement of SACMV and spread of infection.

4.2 TME3, the model of Induce Systemic Resistance

At early infection (12 dpi), differences in SACMV responses between T200 and TME3 are already observed. What this suggests is that even though recovery and tolerance is only observed at 67 dpi, TME3-specific molecular re-programming at early infection, and in fact also at 32 dpi, is significant for downstream induction of tolerance to SACMV. One of the key differences observed was the up-regulation of a number of genes responding to ABA, ET and JA in TME3, but not in T200. The hormone response profile at 12 dpi in TME3 supports triggering of ISR, which is predominately characterized by JA and ET signalling, rather than SAR, which requires SA (Fu and Dong, 2013). Evidence comes from the observation that a large number of the up-regulated DEGs in TME3 fall under JA and ET, whereas SA was represented amongst the down-regulated DEGs at 12 dpi. Despite the up-regulation of ET-responsive DEGs, all members of the ERF family show down-regulation at 12 dpi. This suggests that there are other genes associated to ET are being up-regulated despite the down-regulation of the ERFs. This down-regulation in ERFs may link to another described hallmark of ISR. Transcriptomics of Arabidopsis undergoing ISR revealed an initial reduction of ET signalling before increasing later on during infection (Verhagen *et al*., 2004). The down-regulation of ET signalling would likely lead to down-regulation of the associated responsive ERFs that respond to the signal. Further support for ISR comes from the up-regulation of the *At*WRKY40 orthologs which is first up-regulated at 12 dpi and remains up-regulated throughout infection. This is notable as WRKY40 in Arabidopsis shows up-regulation in plants treated with the ISR inducing *Trichoderma asperelloides* (Brotman *et al*., 2012). This would suggest that the continued induction of *At*WRKY40 orthologs throughout SACMV infection in TME3 is important to ISR induction and maintenance.

The potential role of *At*WRKY40 in inducing the ISR-like defence pathways becomes more apparent at 32 dpi. As *At*WRKY40 forms part of a bigger network that also contains homologs previously reported to be linked to ISR (*At*ETR1, *At*JAZ1, *At*ERF-1, *At*MYC2 and *At*WRKY6) (Pieterse *et al*., 1998, Brotman *et al*., 2012, Verhagen *et al*., 2004, Pozo *et al*., 2008). It can be hypothesised from the network that up-regulated both *At*WRKY6 and 40 transcription factors trigger increased JA synthesis, leading to JAR1 transducing the resulting

hormone signal to the JAZ1 homolog. Here JAZ1 appears to require synergism from ET signals, originating from the ET receptor (*At*ETR1,2) and ERF-1, to induce the homolog of *At*MYC2 that induces the transcription of various defence proteins.

In addition to *At*WRKY 6 and 40, the cassava ortholog *At*MYB15 is also suspected to play a role in ISR in TME3. It is suspected that the up-regulation of the *At*MYB15 homologs actively decrease viral reinfection through lignification, and passively enables cassava to respond faster to subsequent viral infections. In Arabidopsis, MYB15 is known to induce lignification of cells and increase the host's sensitivity to ABA (Heil and Bostock, 2002). ABA accumulations has been reported to be linked positively to abiotic stress tolerance and negatively to biotic stress resistance (Koga *et al.*, 2004, Yamaguchi-Shinozaki and Shinozaki, 2006). In particular, up-regulation of the homologs of *At*MYB108, *At*ETR1, AT5G25930 and *At*UGT74B1 act with MYB15 to deposit addition callose. Strengthening of cell walls through callose deposition and cross-linking is well documented as a defence against pathogens, including viruses (Rampitsch and Bykova, 2012). As previously mentioned, MYB15 increases sensitivity to ABA, explaining why GO-enrichment shows increased levels of ABA-responsive processes. However, given that ABA accumulation is linked to reduced biotic resistance, the precise action to which ABA is involved in tolerance is not obvious and may require the action of other genes.

Another hallmark is that ISR is generally not associated with the accumulation of PR proteins (Pieterse *et al*., 1998, Hammerschmidt, 1999, Van Loon *et al*., 1998). PR proteins are quite diverse, comprising of chitinases, perxoidases, lipid transferases, defensins thionins and protease inhibitors (Van Loon and Van Strien, 1999). While PR protein induction has been well documented in association with SAR-downstream defence to a wide range of pathogens (Sels *et al*., 2008), they are found to be notably absent in ISR. While TME3 did exhibit up-regulation of PR proteins belonging to the peroxidase and chitinase families, it should be noted that their numbers compared to those DE in T200 were considerably lower (18 in T200 vs 5 in TME3 at 32 dpi). Most notably, only one down-regulated gene, that was annotated as "pathogenesis-related", was observed in TME3, whereas 16 were up-regulated in T200 throughout infection. This suggests that ISR-like induction may require the absence of DE PR genes, but rather requires a proportionally lower level of PR induction. In addition,

when comparing genes that are characterised as being “pathogenesis related” in databases, then the down-regulation of the single PR gene further supports the notion of the ISR pathway being induced in TME3.

It is suggested that post-SACMV accumulation and symptoms at 32 dpi (Allie *et al*., 2014), recovery and tolerance in TME3 at 67 dpi which is associated with an attenuation of viral load and symptom severity, is amongst other factors triggered by JA and ET. Notably at 67 dpi, a small network remains containing several up-regulated JA-responsive genes. More importantly both *At*WRKY40 and *At*MYB15 remain up-regulated in this network and are absent in T200 at the same time point. *At*MYB15 increases ABA responsiveness within plant hosts, and *At*WRKY40 also has functions in regulating ABA-responsive genes, and is induced by ABA. Although ABA processes were not observed to be up-regulated at 67 dpi, the up-regulation of *At*MYB15 could result in increased sensitivity of TME3 to ABA responses. This would in turn, lead to the rapid and greater induction of the *At*WRKY40 ortholog upon subsequent pathogen stimulus, thereby eliciting a greater immune response. Based on the network, it is proposed that the homologs of both *At*MYB15 and *At*WRKY40, are contributing to the maintenance of ISR by keeping TME3 in a primed state. ISR is manifested as virus tolerance in cassava TME3 as symptoms and virus replication while reduced are never abolished.

In summary, to date ISR has not been demonstrated to be induced by viral pathogens, and traditionally requires the presence of mutualistic rhizobacteria or a crude extract of their chemicals (Van Loon *et al*., 1998, Kagale *et al*., 2011, Van Peer and Schippers, 1992). However, several of the hormone and transcription factor hallmarks in TME3 suggest that recovery and tolerance in TME3 may share similarities in molecular mechanisms or pathways associated with rhizobacteria-induced ISR.

4.3 Contribution to the field

Many reports in the literature on genes involved in resistance or susceptibility to pathogens focus on individual gene studies, and on bacteria and fungi in the large. The results observed here in T200 and TME3 provide a more in depth study with regards to viruses, geminiviruses in particular, but can logically be extended to other virus-plant host interactions. Furthermore, this study expands on the role of individual genes in host pathogen responses

and attempts, through reliable widely accepted bioinformatics approaches such as STRING, to place these genes in networks. While beyond the scope of this study, building of co-expression networks in cassava is recognized as highly invaluable in identifying molecular players and hubs, and will be performed in future studies. The findings herein support multiple models regarding the characteristics of viral infection in plants and the expected transcriptional changes associated with disease establishment. This research, therefore, contributes to the current body of knowledge, increasing the confidence in these models. This research provides further support to pre-existing knowledge and contributes to novel findings to the field. As mentioned previously, most of the functions derived around defence response genes are based primarily on fungal and bacterial infections. Through the completion of this research, greater context regarding virus-plant interactions has been established. Currently, there are several similar publications available investigating the global transcriptomic changes induced by geminiviruses in tolerant and susceptible plants. However, there is only one other study that thoroughly investigates these pathogen-host interactions through system biology (Rodrigo *et al*., 2011). In this study, Rodrigo *et al*., (2014), examined which genes were DE in Arabidopsis when infected with seven (+) ssRNA viruses and a single ssDNA virus. In doing so, they were able to make comparisons and comment on which genes were commonly targeted by viruses during disease establishment. In contrast this study, this study took their approach and applied it towards T200 and TME3 and in doing so demonstrated the first instance where ISR-like pathways have been shown to be independently induced, without the aid of rhizobacteria or their chemical compounds.

While this research does contribute to both well-established and less-studied scientific literature (pertaining to tolerance in cassava), these findings have potential applications in the agriculture arena in terms of pathogen control. Naturally, before biotechnological strategies may be developed, identification of targets and gaining an understanding of their mechanisms are first required. The genes involved in resistance and tolerance can be targeted by gene editing for engineering cis-genic plants, and this is highly regarded as less risky than genetic modification. This research has successfully identified a reduced number of significant genes from a long list of potential targets for genetic engineering. Promising identified genes that may confer tolerance to other cassava landraces,

may also be investigated and applied to other crops infected with geminiviruses, if common hubs and networks can be recognized. Functional validation of these gene sets would need to be performed in cassava and other crops. Investment in biotechnological strategies may see the development of more desirable CMD tolerant cassava varieties emerge in South Africa, and the other cassava-growing countries in sub-Saharan Africa. T200, which is already preferred by farmers for its high starch yield, would be a strong candidate to benefit from these genetic engineering strategies.

4.4 Study Limitations

Potentially one of the limitations of this study can be found in the lack of knowledge or identification of gene identities and functions from the sequenced cassava genome (AM560-2 which is publically available). In this study, v4.1 was used to identify genes from the cassava transcriptome study (Allie *et al*., 2014). Additionally, while other cassava genomes are being sequenced, these are not publically available. However, since the first released sequence in 2009, there have been vast updates and further improvements, which have allowed more confidence in gene function validation. The sequencing of Arabidopsis (Initiative, 2000), the gold standard in gene identification has provided a suitable platform, but more recent sequencing of the photoneutral rice variety: *Orzya sativa* Kitaake (Goltsman *et al.*, 2017), has contributed to advancing gene function knowledge. Cassava, as mentioned previously, is not a model organism, and currently not all the genes have been functionally annotated. This required that the link between homology and function be established in order to perform sections of the analyses presented. However, these homolog predictions may not necessarily be valid in some instances. However, using phylogeny or/and homology searches with other studied plants, and due to the fact that many of the immune functions in plants are highly conserved, we are confident that most of the observations, and conclusions reached in terms of function in this study, are robust. Arabidopsis is widely used to predict function in most studies in the literature. However, this limitation has been acknowledged by never making non-validated claims surrounding interpretation of function, and homology/orthology of identified cassava genes was established before inferring results. Mechanisms occurring in cassava were based on validated transcriptomic data and from experimentally derived information from the literature.

The second limitation lies in the nature of the transcriptomic data and how it was derived. Transcriptomic data merely functions as a snapshot of the nuanced regulation occurring within a living system. A consequence of this is that the data shows a simplified version of events occurring and does not adequately represent the changes occurring. Due to this, an abductive reasoning is required to interpret the data. Furthermore, the data was produced from SoLID sequencing. This form of sequencing was known to produce lower quality data and shorter reads compared to modern Illumina sequencing. The consequences of this made analysis more challenging as the most used software within the bioinformatics fields are designed around different file formats and better quality of data.

The third limitation lies with the main focus of this study, namely the TFs. It is well known that TFs exert their function primarily on the expression patterns of other genes through binding promoter elements. However, the information surrounding the targets of these cassava TFs are lacking from the transcriptomic data, which means it would be only possible to infer their predicted interactions through differential expression rather than identify their targets. This implies that the only way to derive information regarding the targets, which contributes to the mechanistic action of tolerance, is from literature. Consequently, this limits the degree of specificity in commenting on the potential actions exerted by the TFs. This limitation, too, has been acknowledged by choosing the candidates not on the basis of their transcriptional targets, but rather their expression profiles, function and placement in networks in relation to SACMV infection. By using this method of choosing candidates, it is expected that any important function exerted by TFs downstream through its targets will be reflected in the information presented in the pathway and expression patterns.

Pertaining to STRING, it is important to understand the innate limitation provided by the STRING tool with regards to creating the protein-protein interaction network. The STRING tool produces a network containing predicted interactions, based on information surrounding co-expression data, proximity of genes, experimental validation, protein homology and text-mining of publications for co-occurrence of gene names (Szklarczyk *et al.*, 2015). In particular, the text-mining component in deriving the confidence score for the predicted

interactions is known to inflate the scores resulting in a few false positive interactions. However, this is mitigated through the use of a high confidence score threshold (0.6-0.7) where experimental and functional studies were performed on the genes/proteins of interest. Furthermore, the STRING network is based on Arabidopsis database, which would mean that the downstream specific proteins found in cassava would not be reflected in the network. This results in limited information that can be drawn regarding the downstream partners within a cassava specific context. Despite this, the focus in this study is focused to the general picture and the central regulatory network seen in WRKYs and the other TFs of interest should be conserved. This will still enable comment regarding the immune TFs in cassava in response to SACMV.

Lastly, some methods were not pursued for this study based on the account that they were deemed to be outside the scope of this research assignment. This study was purely a computational based study, containing no lab-based portion whereby the conclusions drawn from this study could be validated. This was deemed to be outside the scope of research as this study formed part of a larger on-going research with the aim to investigate the interactions of viruses and their host and the development of resistance in plant host. This study was, therefore, preliminary to other ongoing research behind transgenic technology which would further investigate the conclusions drawn from this project. Furthermore, it was recommended that MEME be performed to determine enrichment of TFs controlling DEG based on the presence of binding sites within their promoter regions. This route was originally pursued but eventually dropped. The reason for this is that when performed on the WRKYs promoter regions, it was found that 80% of the genome yielded a positive result due to the small size of the W-box motif. Even through consulting literature and reducing the list to only include genes were multiple (three or more) instances were found, 45000 hits still remained. It was for this reason that MEME was dropped from the study as the small binding motifs from the selected TF families posed to be a weakness in the analysis and was predicted not to yield any useful information regarding tolerance candidates.

4.5 Future Work

The aim of this project, which was to identify TFs, and place them into predicted networks, so as to infer their function in the development of tolerance and susceptibility in cassava, was

achieved. However, experimental validation of the functions of selected genes is required in future. It is noted that this study hinged on the use of cross-mapped accession numbers between genome versions (v4.1 and v6.1) and plant species (cassava and Arabidopsis). The reason for cross-mapping was chosen over remapping of the cassava genome was based on the timing when version 6.1 was released. During the course of the study, much of the analysis was performed on the 4.1 version when the updated genome was published and to perform remapping which would entail starting all the analyses over. To start over all the performed analyses was deemed impractical so late into the study, however in order to keep the research relevant, the genes were cross-mapped to the 6.1 version. This is a known limitation to the study and in order to address this, remapping of the most current cassava genome (v7.1) will be performed in future. After remapping, subsequent construction of a gene co-expression network for cassava based on transcriptome data from T200 and TME3 will be performed. In doing so, any updates incurred in the cassava genome between 4.1 to 7.1 will be righted and reflected in the up to date information,

The use of either knock-down or constitutive expression studies of these candidate genes would allow for the identification of other interacting partners in cassava, thereby validating the proposed pathways, and confirming whether these candidates do contribute to tolerance in TME3. The networks in this study focused on protein-protein interactions and do not take into account the downstream targets of the candidate TFs. High-throughput information regarding the binding targets of these candidate TFs would contribute to further knowledge, enabling the molecular mechanisms of tolerance in cassava, and other important crops, to be uncovered. Once validation has been performed and the targets of the candidate TFs have been elucidated, developing natural resistance approaches to control CMD would be possible through gene editing using CRISPR technologies, or genetic modification via gene silencing. Genetic modification or editing of the landrace T200 would be invaluable, as T200 is high in starch and could be developed as an important crop for commercial use in South Africa. However, this would require further investigation in future on biosafety issues in this regard, and also evaluation of potential risks of genetic modification.

References

Agudelo-Romero, P., Carbonell, P., De La Iglesia, F., Carrera, J., Rodrigo, G., Jaramillo, A., Pérez-Amador, M. A. & Elena, S. F. (2008). Changes in the gene expression profile of Arabidopsis thaliana after infection with Tobacco etch virus. *Virology Journal,* 5**,** 92.

Akano, A., Dixon, A., Mba, C., Barrera, E. & Fregene, M. (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theoretical and Applied Genetics,* 105**,** 521-525.

Akinbade, S., Hanna, R., Nguenkam, A., Njukwe, E., Fotso, A., Doumtsop, A., Ngeve, J., Tenku, S. & Lava Kumar, P. (2010). First report of the East African cassava mosaic virus-Uganda (EACMV-UG) infecting cassava (Manihot esculenta) in Cameroon. *New Disease Reports,* 21**,** 2044-0588.2010.

Al-Kaff, N. S., Covey, S. N., Kreike, M. M., Page, A. M., Pinder, R. & Dale, P. J. (1998). Transcriptional and posttranscriptional plant gene silencing in response to a pathogen. *Science,* 279**,** 2113-2115.

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

Amuge, T., Berger, D. K., Katari, M. S., Myburg, A. A., Goldman, S. L. & Ferguson, M. E. (2017). A time series transcriptome analysis of cassava (Manihot esculenta Crantz) varieties challenged with Ugandan cassava brown streak virus. *Scientific reports,* 7**,** 9747.

Anjanappa, R. B., Mehta, D., Okoniewski, M. J., Szabelska, A., Gruissem, W. & Vanderschuren, H. (2017). Early transcriptome response to brown streak virus infection in susceptible and resistant cassava varieties. *BioRxiv***,** 1-22.

Ascencio-Ibáñez, J. T., Sozzani, R., Lee, T.-J., Chu, T.-M., Wolfinger, R. D., Cella, R. & Hanley-Bowdoin, L. (2008). Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. *Plant physiology,* 148**,** 436-454.

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

Babitha, K., Ramu, S., Pruthvi, V., Mahesh, P., Nataraja, K. N. & Udayakumar, M. (2013). Co-expression of AtbHLH17 and AtWRKY28 confers resistance to abiotic stress in Arabidopsis. *Transgenic research,* 22**,** 327-341.

Bakshi, M. & Oelmüller, R. (2014). WRKY transcription factors: Jack of many trades in plants. *Plant signalling & behavior,* 9**,** 1-18.

Bednarek, P., Piślewska-Bednarek, M., Svatoš, A., Schneider, B., Doubský, J., Mansurova, M., Humphry, M., Consonni, C., Panstruga, R. & Sanchez-Vallet, A. (2009). A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science,* 323**,** 101-106.

Bencke-Malato, M., Cabreira, C., Wiebke-Strohm, B., Bücker-Neto, L., Mancini, E., Osorio, M. B., Homrich, M. S., Turchetto-Zolet, A. C., De Carvalho, M. C. & Stolf, R. (2014). Genome-wide annotation of the soybean WRKY family and functional

characterization of genes involved in response to Phakopsora pachyrhizi infection. *BMC plant biology,* 14**,** 236.

Bengyella, L., Waikhom, S. D., Allie, F. & Rey, C. (2015). Virus tolerance and recovery from viral induced-symptoms in plants are associated with transcriptome reprograming. *Plant Molecular Biology,* 89**,** 243-252.

Berens, M. L., Berry, H. M., Mine, A., Argueso, C. T. & Tsuda, K. (2017). Evolution of Hormone Signalling Networks in Plant Defense. *Annual Review of Phytopathology,* 55**,** annurev-phyto-080516-035544.

Berri, S., Abbruscato, P., Faivre-Rampant, O., Brasileiro, A. C., Fumasoni, I., Satoh, K., Kikuchi, S., Mizzi, L., Morandini, P. & Pè, M. E. (2009). Characterization of WRKY co-regulatory networks in rice and Arabidopsis. *BMC plant biology,* 9**,** 120.

Berrie, L. C., Rybicki, E. P. & Rey, M. E. C. (2001). Complete nucleotide sequence and host range of South African cassava mosaic virus: further evidence for recombination amongst begomoviruses. *The Journal of general virology,* 82**,** 53-58.

Berrocal-Lobo, M. & Molina, A. (2004). Ethylene response factor 1 mediates Arabidopsis resistance to the soilborne fungus Fusarium oxysporum. *Molecular Plant-Microbe Interactions,* 17**,** 763-770.

Berrocal-Lobo, M., Molina, A. & Solano, R. (2002). Constitutive expression of ETHYLENE-RESPONSE-FACTOR1 in Arabidopsis confers resistance to several necrotrophic fungi. *The Plant Journal,* 29**,** 23-32.

Berry, S. & Rey, M. E. C. (2001). Molecular evidence for diverse populations of cassava-infecting begomoviruses in southern Africa. *Archives of Virology,* 146**,** 1795-1802.

Bethke, G., Unthan, T., Uhrig, J. F., Pöschl, Y., Gust, A. A., Scheel, D. & Lee, J. (2009). Flg22 regulates the release of an ethylene response factor substrate from MAP kinase 6 in Arabidopsis thaliana via ethylene signalling. *Proceedings of the National Academy of Sciences,* 106**,** 8067-8072.

Bhardwaj, V., Meier, S., Petersen, L. N., Ingle, R. A. & Roden, L. C. (2011). Defence responses of Arabidopsis thaliana to infection by Pseudomonas syringae are regulated by the circadian clock. *PLoS One,* 6**,** e26968.

Bindschedler, L. V., Dewdney, J., Blee, K. A., Stone, J. M., Asai, T., Plotnikov, J., Denoux, C., Hayes, T., Gerrish, C. & Davies, D. R. (2006). Peroxidase-dependent apoplastic oxidative burst in Arabidopsis required for pathogen resistance. *The Plant Journal,* 47**,** 851-863.

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

Bock, K. & Woods, R. (1983). Etiology of African cassava mosaic disease. *Plant Disease,* 67**,** 994-995.

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

Bolt, S., Zuther, E., Zintl, S., Hincha, D. K. & Schmülling, T. (2017). ERF105 is a transcription factor gene of Arabidopsis thaliana required for freezing tolerance and cold acclimation. *Plant, cell & environment,* 40**,** 108-120.

Bostock, R. M. (2005). Signal crosstalk and induced resistance: straddling the line between cost and benefit. *Annu. Rev. Phytopathol.,* 43**,** 545-580.

Boulton, M. I. (2003). Geminiviruses: major threats to world agriculture. *Annals of Applied Biology,* 142**,** 143.

Bredeson, J. V., Lyons, J. B., Prochnik, S. E., Wu, G. A., Ha, C. M., Edsinger-Gonzales, E., Grimwood, J., Schmutz, J., Rabbi, I. Y. & Egesi, C. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nature Biotechnology,* 34**,** 562.

Briddon, R. & Stanley, J. (2006). Subviral agents associated with plant single-stranded DNA viruses. *Virology,* 344**,** 198-210.

Brotman, Y., Lisec, J., Méret, M., Chet, I., Willmitzer, L. & Viterbo, A. (2012). Transcript and metabolite analysis of the Trichoderma-induced systemic resistance response to Pseudomonas syringae in Arabidopsis thaliana. *Microbiology,* 158**,** 139-146.

Bu, Q., Jiang, H., Li, C.-B., 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 research,* 18**,** 756.

Budak, H. & Akpinar, B. A. (2015). Plant miRNAs: biogenesis, organization and origins. *Functional & integrative genomics,* 15**,** 523-531.

Büttner, M. & Singh, K. B. (1997). Arabidopsis thaliana ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. *Proceedings of the National Academy of Sciences,* 94**,** 5961-5966.

Cai, X.-T., Xu, P., Zhao, P.-X., Liu, R., Yu, L.-H. & Xiang, C.-B. (2014). Arabidopsis ERF109 mediates cross-talk between jasmonic acid and auxin biosynthesis during lateral root formation. *Nature Communications,* 5**,** 5833.

Cao, H., Bowling, S. A., Gordon, A. S. & Dong, X. (1994). Characterization of an Arabidopsis mutant that is nonresponsive to inducers of systemic acquired resistance. *The Plant Cell,* 6**,** 1583-1592.

Cao, H., Glazebrook, J., Clarke, J. D., Volko, S. & Dong, X. (1997). The Arabidopsis NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell,* 88**,** 57-63.

Ceballos, G., García, A. & Ehrlich, P. R. (2010). The sixth extinction crisis: loss of animal populations and species. *Journal of Cosmology,* 8**,** 31.

Chae, L., Lee, I., Shin, J. & Rhee, S. Y. (2012). Towards understanding how molecular networks evolve in plants. *Current Opinion in Plant Biology,* 15**,** 177-184.

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

Chen, L., Zhang, L. & Yu, D. (2010). Wounding-induced WRKY8 is involved in basal defense in Arabidopsis. *Molecular Plant-Microbe Interactions,* 23**,** 558-565.

Chen, T., Lv, Y., Zhao, T., Li, N., Yang, Y., Yu, W., He, X., Liu, T. & Zhang, B. (2013). Comparative transcriptome profiling of a resistant vs. susceptible tomato (Solanum lycopersicum) cultivar in response to infection by tomato yellow leaf curl virus. *PLoS One,* 8**,** 4-6.

Chen, W., Provart, N. J., Glazebrook, J., Katagiri, F., Chang, H.-S., Eulgem, T., Mauch, F., Luan, S., Zou, G. & Whitham, S. A. (2002). Expression profile matrix of Arabidopsis

transcription factor genes suggests their putative functions in response to environmental stresses. *The Plant Cell,* 14**,** 559-574.

Chen, Z., Hong, X., Zhang, H., Wang, Y., Li, X., Zhu, J. K. & Gong, Z. (2005). Disruption of the cellulose synthase gene, AtCesA8/IRX1, enhances drought and osmotic stress tolerance in Arabidopsis. *The Plant Journal,* 43**,** 273-283.

Cheng, Y., Zhou, Y., Yang, Y., Chi, Y.-J., Zhou, J., Chen, J.-Y., Wang, F., Fan, B., Shi, K. & Zhou, Y.-H. (2012). Structural and functional analysis of VQ motif-containing proteins in Arabidopsis as interacting proteins of WRKY transcription factors. *Plant physiology,* 159**,** 810-825.

Chezem, W. R., Memon, A., Li, F.-S., Weng, J.-K. & Clay, N. K. (2017). SG2-type R2R3-MYB transcription factor MYB15 controls defense-induced lignification and basal immunity in Arabidopsis. *The Plant Cell,* 29**,** 1907-1926.

Chini, A., Boter, M. & Solano, R. (2009). Plant oxylipins: COI1/JAZs/MYC2 as the core jasmonic acid-signalling module. *The FEBS journal,* 276**,** 4682-4692.

Chisholm, S. T., Coaker, G., Day, B. & Staskawicz, B. J. (2006). Host-microbe interactions: shaping the evolution of the plant immune response. *Cell,* 124**,** 803-814.

Choudhary, D. K., Prakash, A. & Johri, B. (2007). Induced systemic resistance (ISR) in plants: mechanism of action. *Indian Journal of Microbiology,* 47**,** 289-297.

Chung, H. S., Koo, A. J., Gao, X., Jayanty, S., Thines, B., Jones, A. D. & Howe, G. A. (2008). Regulation and function of Arabidopsis JASMONATE ZIM-domain genes in response to wounding and herbivory. *Plant physiology,* 146**,** 952-964.

Cominelli, E., Galbiati, M., Vavasseur, A., Conti, L., Sala, T., Vuylsteke, M., Leonhardt, N., Dellaporta, S. L. & Tonelli, C. (2005). A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *Current biology,* 15**,** 1196-1200.

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

Coupe, S., Palmer, B., Lake, J., Overy, S., Oxborough, K., Woodward, F., Gray, J. & Quick, W. (2005). Systemic signalling of environmental cues in Arabidopsis leaves. *Journal of Experimental Botany,* 57**,** 329-341.

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

D'maris Amick Dempsey, A. C., Vlot, M. C. W. & Daniel, F. K. (2011). Salicylic acid biosynthesis and metabolism. *The Arabidopsis book/American Society of Plant Biologists,* 9.

Dan, I., Watanabe, N. M. & Kusumi, A. (2001). The Ste20 group kinases as regulators of MAP kinase cascades. *Trends in cell biology,* 11**,** 220-230.

Dangl, J. L. & Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *Nature,* 411**,** 826.

Davies, P. J. 2010. The plant hormones: their nature, occurrence, and functions. *Plant hormones.* Springer.

Dempsey, A. C. D. M. A., Vlot, M. C. W. & Daniel, F. K. (2011). Salicylic acid biosynthesis and metabolism. *The Arabidopsis book/American Society of Plant Biologists,* 9.

Dempsey, D. M. A. & Klessig, D. F. (2017). How does the multifaceted plant hormone salicylic acid combat disease in plants and are similar mechanisms utilized in humans? *BMC Biology*, 15**,** 23.

Deyoung, B. J., Bickle, K. L., Schrage, K. J., Muskett, P., Patel, K. & Clark, S. E. (2006). The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in Arabidopsis. *The Plant Journal,* 45**,** 1-16.

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

Dixon, A. F. & Hopkins, G. W. 2010. Temperature, seasonal development and distribution of insects with particular reference to aphids. *Aphid Biodiversity under Environmental Change.* Springer.

Dixon, A. G., Akoroda, M. O., Okechukwu, R. U., Ogbe, F., Ilona, P., Sanni, L. O., Ezedinma, C., Lemchi, J., Ssemakula, G. & Yomeni, M. O. (2008). Fast track participatory approach to release of elite cassava genotypes for various uses in Nigeria's cassava economy. *Euphytica,* 160**,** 1-13.

Do, C.-T., Pollet, B., Thévenin, J., Sibout, R., Denoue, D., Barrière, Y., Lapierre, C. & Jouanin, L. (2007). Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in Arabidopsis. *Planta,* 226**,** 1117-1129.

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

Dombrecht, B., Xue, G. P., Sprague, S. J., Kirkegaard, J. A., Ross, J. J., Reid, J. B., Fitt, G. P., Sewelam, N., Schenk, P. M. & Manners, J. M. (2007). MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. *The Plant Cell,* 19**,** 2225-2245.

Duplessis, S., Major, I., Martin, F. & Séguin, A. (2009). Poplar and Pathogen Interactions: Insights from Populus Genome-Wide Analyses of Resistance and Defense Gene Families and Gene Expression Profiling. *Critical Reviews in Plant Sciences,* 28**,** 309-334.

Durrant, W. E. & Dong, X. (2004). Systemic acquired resistance. *Annu. Rev. Phytopathol.*, 42**,** 185-209.

El-Sharkawy, M. A. (2004). Cassava biology and physiology. *Plant Molecular Biology,* 56**,** 481-501.

Epel, B. L. Plant viruses spread by diffusion on ER-associated movement-protein-rafts through plasmodesmata gated by viral induced host β-1, 3-glucanases. Seminars in cell & developmental biology, 2009. Elsevier, 1074-1081.

Erpen, L., Devi, H. S., Grosser, J. W. & Dutt, M. (2018). Potential use of the DREB/ERF, MYB, NAC and WRKY transcription factors to improve abiotic and biotic stress in transgenic plants. *Plant Cell, Tissue and Organ Culture,* 132**,** 1-25.

Eulgem, T., Rushton, P. J., Robatzek, S. & Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. *Trends in plant science,* 5**,** 199-206.

Eulgem, T. & Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signalling. *Current Opinion in Plant Biology,* 10**,** 366-371.

Fan, W., Hai, M., Guo, Y., Ding, Z., Tie, W., Ding, X. & Yan, Y. (2016). The ERF transcription factor family in cassava : genome-wide characterization and expression analyses against drought stress. *Nature Publishing Group***,** 1-12.

Felix, G., Duran, J. D., Volko, S. & Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *The Plant Journal,* 18**,** 265-276.

Ferrari, S., Plotnikova, J. M., De Lorenzo, G. & Ausubel, F. M. (2003). Arabidopsis local resistance to Botrytis cinerea involves salicylic acid and camalexin and requires EDS4 and PAD2, but not SID2, EDS5 or PAD4. *The Plant Journal,* 35**,** 193-205.

Fietto, L. G., Barros, V. A., De Souza, G. B., Gonçalves, A. B., Dadalto, S. P. & Alves, M. S. (2013). Plant bZIP Transcription Factors Responsive to Pathogens&58; A Review. *International journal of molecular sciences,* 14**,** 7815-7828.

Fondong, V. N. & Rey, C. 2018. Recent Biotechnological Advances in the Improvement of Cassava. *Cassava.* IntechOpen.

Fraile, A. & García-Arenal, F. 2010. The coevolution of plants and viruses: resistance and pathogenicity. *Advances in Virus Research.* Elsevier.

Fu, Z. Q. & Dong, X. (2013). Systemic acquired resistance: turning local infection into global defense. *Annual review of plant biology,* 64**,** 839-863.

Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi-Shinozaki, K. & Shinozaki, K. (2006). Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signalling networks. *Current Opinion in Plant Biology,* 9**,** 436-442.

Gabaldón, T. (2008). Large-scale assignment of orthology: back to phylogenetics? *Genome Biology,* 9**,** 235.

Gao, M., Wang, X., Wang, D., Xu, F., Ding, X., Zhang, Z., Bi, D., Cheng, Y. T., Chen, S. & Li, X. (2009). Regulation of cell death and innate immunity by two receptor-like kinases in Arabidopsis. *Cell host & microbe,* 6**,** 34-44.

García-Olmedo, F., Molina, A., Segura, A. & Moreno, M. (1995). The defensive role of nonspecific lipid-transfer proteins in plants. *Trends in microbiology,* 3**,** 72-74.

Gasura, E. & Mukasa, S. (2010). Prevalence and implications of sweetpotato recovery from sweet potato virus disease in Uganda. *African Crop Science Journal,* 18.

Ghanbari, M., Eini, O. & Ebrahimi, S. (2016). Differential expression of MYB33 and AP2 genes and response of TY resistant plants to beet curly top iran virus infection in tomato. *Journal of Plant Pathology,* 98**,** 555-562.

Ghoshal, B. & Sanfaçon, H. (2015). Symptom recovery in virus-infected plants: revisiting the role of RNA silencing mechanisms. *Virology,* 479**,** 167-179.

Glazebrook, J. (2001). Genes controlling expression of defense responses in Arabidopsis—2001 status. *Current Opinion in Plant Biology,* 4**,** 301-308.

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

Golldack, D., Lüking, I. & Yang, O. (2011). Plant tolerance to drought and salinity: Stress regulating transcription factors and their functional significance in the cellular transcriptional network. *Plant Cell Reports,* 30**,** 1383-1391.

Goltsman, E., Ho, I. & Rokhsar, D. (2017). Meraculous-2D: Haplotype-sensitive Assembly of Highly Heterozygous genomes. *arXiv preprint arXiv:1703.09852*.

Gong, T., Shu, D., Zhao, M., Zhong, J., Deng, H. Y. & Tan, H. (2014). Isolation of genes related to abscisic acid production in Botrytis cinerea TB-3-H8 by cDNA-AFLP. *Journal of basic microbiology,* 54**,** 204-214.

Grellet-Bournonville, C. F., Martinez-Zamora, M. G., Castagnaro, A. P. & Díaz-Ricci, J. C. (2012). Temporal accumulation of salicylic acid activates the defense response against Colletotrichum in strawberry. *Plant physiology and biochemistry,* 54**,** 10-16.

Gu, C., Guo, Z.-H., Hao, P.-P., Wang, G.-M., Jin, Z.-M. & Zhang, S.-L. (2017). Multiple regulatory roles of AP2/ERF transcription factor in angiosperm. *Botanical studies,* 58**,** 6.

Guan, Y., Ren, H., Xie, H., Ma, Z. & Chen, F. (2009). Identification and characterization of bZIP-type transcription factors involved in carrot (Daucus carota L.) somatic embryogenesis. *The Plant Journal,* 60**,** 207-217.

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

Gutterson, N. & Reuber, T. L. (2004). Regulation of disease resistance pathways by AP2/ERF transcription factors. *Current Opinion in Plant Biology,* 7**,** 465-471.

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

Hahn, S., Terry, E. & Leuschner, K. (1980). Breeding cassava for resistance to cassava mosaic disease. *Euphytica,* 29**,** 673-683.

Hammerschmidt, R. 1999. Induced disease resistance: how do induced plants stop pathogens? : Elsevier.

Hammond-Kosack, K. E. & Parker, J. E. (2003). Deciphering plant–pathogen communication: fresh perspectives for molecular resistance breeding. *Current opinion in biotechnology,* 14**,** 177-193.

Hanley-Bowdoin, L., Settlage, S. B. & Robertson, D. (2004). Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. *Molecular Plant Pathology,* 5**,** 149-156.

Heil, M. & Bostock, R. M. (2002). Induced Systemic Resistance (ISR) Against Pathogens in the Context of Induced Plant Defences Author (s): MARTIN HEIL and RICHARD M . BOSTOCK Source : Annals of Botany , Vol . 89 , No . 5 (May 2002), pp . 503-512 Published by : Oxford University Press Stable URL : https://www.jstor.org/stable/42771477 Induced Systemic Resistance (ISR) Against Pathogens in the Context of Induced Plant Defences. 89**,** 503-512.

Hernandez-Garcia, C. M. & Finer, J. J. (2014). Identification and validation of promoters and cis-acting regulatory elements. *Plant Science,* 217**,** 109-119.

Hillocks, R. & Jennings, D. (2003). Cassava brown streak disease: a review of present knowledge and research needs. *International Journal of Pest Management,* 49**,** 225-234.

Hu, W., Wei, Y., Xia, Z., Yan, Y., Hou, X., Zou, M., Lu, C., Wang, W. & Peng, M. (2015). Genome-wide identification and expression analysis of the NAC transcription factor family in cassava. *PLoS One,* 10**,** e0136993.

Hu, W., Yang, H., Yan, Y., Wei, Y., Tie, W., Ding, Z., Zuo, J., Peng, M. & Li, K. (2016). Genome-wide characterization and analysis of bZIP transcription factor gene family related to abiotic stress in cassava. *Scientific reports,* 6**,** 22783.

Hu, Y., Dong, Q. & Yu, D. (2012). Arabidopsis WRKY46 coordinates with WRKY70 and WRKY53 in basal resistance against pathogen Pseudomonas syringae. *Plant Science,* 185**,** 288-297.

Hunter, L. J., Westwood, J. H., Heath, G., Macaulay, K., Smith, A. G., Macfarlane, S. A., Palukaitis, P. & Carr, J. P. (2013). Regulation of RNA-dependent RNA polymerase 1 and isochorismate synthase gene expression in Arabidopsis. *PLoS One,* 8**,** e66530.

Ichimura, K., Shinozaki, K., Tena, G., Sheen, J., Henry, Y., Champion, A., Kreis, M., Zhang, S., Hirt, H. & Wilson, C. (2002). Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends in plant science,* 7**,** 301-308.

Initiative, A. G. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature,* 408**,** 796.

Irvine, D. V., Zaratiegui, M., Tolia, N. H., Goto, D. B., Chitwood, D. H., Vaughn, M. W., Joshua-Tor, L. & Martienssen, R. A. (2006). Argonaute slicing is required for heterochromatic silencing and spreading. *Science,* 313**,** 1134-1137.

Ito, M. (2005). Conservation and diversification of three-repeat Myb transcription factors in plants. *Journal of plant research,* 118**,** 61-69.

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T. & Parcy, F. (2002). bZIP transcription factors in Arabidopsis. *Trends in plant science,* 7**,** 106-111.

Jeevalatha, A., Siddappa, S., Kumar, A., Kaundal, P., Guleria, A., Sharma, S., Nagesh, M. & Singh, B. P. (2017). An insight into differentially regulated genes in resistant and susceptible genotypes of potato in response to tomato leaf curl New Delhi virus-[potato] infection. *Virus Research,* 232**,** 22-33.

Jiao, J., Xing, J.-H., Dong, J.-G., Han, J.-M. & Liu, J.-S. (2011). Functional analysis of MYB73 of Arabidopsis thaliana against Bipolaris oryzae. *Agricultural sciences in China,* 10**,** 721-727.

Jing, B., Xu, S., Xu, M., Li, Y., Li, S., Ding, J. & Zhang, Y. (2011). Brush and spray: a high-throughput systemic acquired resistance assay suitable for large-scale genetic screening. *Plant physiology,* 157**,** 973-980.

Jones, D. A. & Jones, J. D. 1997. The role of leucine-rich repeat proteins in plant defences. *Advances in Botanical Research.* Elsevier.

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

Ju, C., Van De Poel, B., Cooper, E. D., Thierer, J. H., Gibbons, T. R., Delwiche, C. F. & Chang, C. (2015). Conservation of ethylene as a plant hormone over 450 million years of evolution. *Nature Plants,* 1**,** 14004.

Kagale, S., Marimuthu, T., Kagale, J., Thayumanavan, B. & Samiyappan, R. (2011). Induction of systemic resistance in rice by leaf extracts of Zizyphus jujuba and Ipomoea carnea against Rhizoctonia solani. *Plant signalling & behavior,* 6**,** 919-923.

Kaminaka, H., Näke, C., Epple, P., Dittgen, J., Schütze, K., Chaban, C., Holt, B. F., Merkle, T., Schäfer, E. & Harter, K. (2006). bZIP10-LSD1 antagonism modulates basal defense and cell death in Arabidopsis following infection. *The EMBO journal,* 25**,** 4400-4411.

Kanofsky, K., Strauch, C. J., Sandmann, A., Möller, A. & Hehl, R. (2018). Transcription factors involved in basal immunity in mammals and plants interact with the same MAMP-responsive cis-sequence from Arabidopsis thaliana. *Plant Molecular Biology,* 98**,** 565-578.

Katsir, L., Chung, H. S., Koo, A. J. & Howe, G. A. (2008). Jasmonate signalling: a conserved mechanism of hormone sensing. *Current Opinion in Plant Biology,* 11**,** 428-435.

Kawano, T. (2003). Roles of the reactive oxygen species-generating peroxidase reactions in plant defense and growth induction. *Plant Cell Reports,* 21**,** 829-837.

Kazan, K. & Manners, J. M. (2012). JAZ repressors and the orchestration of phytohormone crosstalk. *Trends in plant science,* 17**,** 22-31.

Kazan, K. & Manners, J. M. (2013). MYC2: The master in action. *Molecular Plant,* 6**,** 686-703.

Khan, M. S. (2011). The role of DREB transcription factors in abiotic stress tolerance of plants. *Biotechnology & Biotechnological Equipment,* 25**,** 2433-2442.

Khraiwesh, B., Zhu, J.-K. & Zhu, J. (2012). Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms,* 1819**,** 137-148.

Kim, H., Park, Y., Nam, H., Lee, Y., Song, K., Choi, C., Ahn, I., Park, S., Lee, Y. & Hwang, D. (2014). Overexpression of the B rassica rapa transcription factor WRKY 12 results in reduced soft rot symptoms caused by P ectobacterium carotovorum in A rabidopsis and C hinese cabbage. *Plant Biology,* 16**,** 973-981.

Kim, K.-C., Fan, B. & Chen, Z. (2006a). Pathogen-induced Arabidopsis WRKY7 is a transcriptional repressor and enhances plant susceptibility to Pseudomonas syringae. *Plant physiology,* 142**,** 1180-1192.

Kim, W.-C., Kim, J.-Y., Ko, J.-H., Kim, J. & Han, K.-H. (2013). Transcription factor MYB46 is an obligate component of the transcriptional regulatory complex for functional expression of secondary wall-associated cellulose synthases in Arabidopsis thaliana. *Journal of plant physiology,* 170**,** 1374-1378.

Kim, Y.-S., Kim, S.-G., Park, J.-E., Park, H.-Y., Lim, M.-H., Chua, N.-H. & Park, C.-M. (2006b). A membrane-bound NAC transcription factor regulates cell division in Arabidopsis. *The Plant Cell,* 18**,** 3132-3144.

Kinkema, M., Fan, W. & Dong, X. (2000). Nuclear localization of NPR1 is required for activation of PR gene expression. *The Plant Cell,* 12**,** 2339-2350.

Koga, H., Dohi, K. & Mori, M. (2004). Abscisic acid and low temperatures suppress the whole plant-specific resistance reaction of rice plants to the infection of Magnaporthe grisea. *Physiological and Molecular Plant Pathology,* 65**,** 3-9.

Koo, A. J., Chung, H. S., Kobayashi, Y. & Howe, G. A. (2006). Identification of a peroxisomal acyl-activating enzyme involved in the biosynthesis of jasmonic acid in Arabidopsis. *Journal of Biological Chemistry,* 281**,** 33511-33520.

Korolev, N., David, D. R. & Elad, Y. (2008). The role of phytohormones in basal resistance and Trichoderma-induced systemic resistance to Botrytis cinerea in Arabidopsis thaliana. *BioControl,* 53**,** 667-683.

Ku, J. (1982). Association of enhanced peroxidase activity with induced systemic resistance of cucumber to Colletotrichum I ’ agenarium. 73-82.

Kumar, D., Hazra, S., Datta, R. & Chattopadhyay, S. (2016). Transcriptome analysis of Arabidopsis mutants suggests a crosstalk between ABA, ethylene and GSH against combined cold and osmotic stress. *Scientific reports,* 6**,** 36867.

Kumar, P. L., Akinbade, S., Dixon, A. G., Mahungu, N., Mutunda, M., Kiala, D., Londa, L. & Legg, J. (2009). First report of the occurrence of East African cassava mosaic virus-Uganda (EACMV-UG) in Angola. *Plant Pathol,* 58**,** 402.

Kunkel, B. N. & Brooks, D. M. (2002). Cross talk between signalling pathways in pathogen defense. *Current Opinion in Plant Biology,* 5**,** 325-331.

Kuria, P., Ilyas, M., Ateka, E., Miano, D., Onguso, J., Carrington, J. C. & Taylor, N. J. (2017). Differential response of cassava genotypes to infection by cassava mosaic geminiviruses. *Virus Research,* 227**,** 69-81.

Leeman, M., Van Pelt, J., Den Ouden, F., Heinsbroek, M., Bakker, P. & Schippers, B. (1995). Induction of systemic resistance against Fusarium wilt of radish by lipopolysaccharides of Pseudomonas fluorescens. *Phytopathology,* 85**,** 1021-1027.

Legg, J., Jeremiah, S., Obiero, H., Maruthi, M., Ndyetabula, I., Okao-Okuja, G., Bouwmeester, H., Bigirimana, S., Tata-Hangy, W. & Gashaka, G. (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Research,* 159**,** 161-170.

Legg, J., Owor, B., Sseruwagi, P. & Ndunguru, J. (2006). Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Advances in Virus Research,* 67**,** 355-418.

Legg, J. P. & Fauquet, C. M. (2004). Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology,* 56**,** 585-599.

Legg, J. P., Shirima, R., Tajebe, L. S., Guastella, D., Boniface, S., Jeremiah, S., Nsami, E., Chikoti, P. & Rapisarda, C. (2014). Biology and management of Bemisia whitefly vectors of cassava virus pandemics in Africa. *Pest management science,* 70**,** 1446-1453.

Li, J., Brader, G. & Palva, E. T. (2004). The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *The Plant Cell,* 16**,** 319-331.

Li, J. & Hu, J. (2015). Using co-expression analysis and stress-based screens to uncover Arabidopsis peroxisomal proteins involved in drought response. *PLoS One,* 10**,** e0137762.

Li, K., Yang, F., Miao, Y. & Song, C.-P. (2017a). Abscisic acid signalling is involved in regulating the mitogen-activated protein kinase cascade module, AIK1-MKK5-MPK6. *Plant signalling & behavior,* 12**,** e1321188.

Li, L.-Q., Huang, L.-P., Pan, G., Liu, L., Wang, X.-Y. & Lu, L.-M. (2017b). Identifying the genes regulated by AtWRKY6 using comparative transcript and proteomic analysis under phosphorus deficiency. *International journal of molecular sciences,* 18**,** 1046.

Li, W., Cui, X., Meng, Z., Huang, X., Xie, Q., Wu, H., Jin, H., Zhang, D. & Liang, W. (2012). Transcriptional regulation of Arabidopsis MIR168a and argonaute1 homeostasis in abscisic acid and abiotic stress responses. *Plant physiology,* 158**,** 1279-1292.

Liang, D., White, R. G. & Waterhouse, P. M. (2014). Mobile gene silencing in Arabidopsis is regulated by hydrogen peroxide. *PeerJ,* 2**,** e701.

Liang, W., Li, C., Liu, F., Jiang, H., Li, S., Sun, J., Wu, X. & Li, C. (2009). The Arabidopsis homologs of CCR4-associated factor 1 show mRNA deadenylation activity and play a role in plant defence responses. *Cell research,* 19**,** 307.

Liao, W., Yang, Y., Li, Y., Wang, G. & Peng, M. (2016). Genome-wide identification of cassava R2R3 MYB family genes related to abscission zone separation after environmental-stress-induced abscission. *Scientific reports,* 6**,** 32006.

Libault, M., Wan, J., Czechowski, T., Udvardi, M. & Stacey, G. (2007). Identification of 118 Arabidopsis transcription factor and 30 ubiquitin-ligase genes responding to chitin, a plant-defense elicitor. *Molecular Plant-Microbe Interactions,* 20**,** 900-911.

Liu, J.-J. & Ekramoddoullah, A. (2004). Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (Pinus monticola Dougl. ex. D. Don.). *Molecular Genetics and Genomics,* 270**,** 432-441.

Liu, J., Liu, X., Dai, L. & Wang, G. (2007). Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. *Journal of Genetics and Genomics,* 34**,** 765-776.

Liu, Y. & Zhang, S. (2004). Phosphorylation of 1-aminocyclopropane-1-carboxylic acid synthase by MPK6, a stress-responsive mitogen-activated protein kinase, induces ethylene biosynthesis in Arabidopsis. *The Plant Cell,* 16**,** 3386-3399.

Liu, Z.-Q., Yan, L., Wu, Z., Mei, C., Lu, K., Yu, Y.-T., Liang, S., Zhang, X.-F., Wang, X.-F. & Zhang, D.-P. (2012). Cooperation of three WRKY-domain transcription factors WRKY18, WRKY40, and WRKY60 in repressing two ABA-responsive genes ABI4 and ABI5 in Arabidopsis. *Journal of Experimental Botany,* 63**,** 6371-6392.

López, M. A., Bannenberg, G. & Castresana, C. (2008). Controlling hormone signalling is a plant and pathogen challenge for growth and survival. *Current Opinion in Plant Biology,* 11**,** 420-427.

Lorenzo, O., Piqueras, R., Sánchez-Serrano, J. J. & Solano, R. (2003). ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. *The Plant Cell,* 15**,** 165-178.

Lorenzo, O. & Solano, R. (2005). Molecular players regulating the jasmonate signalling network. *Current Opinion in Plant Biology,* 8**,** 532-540.

Luo, X., Xu, N., Huang, J., Gao, F., Zou, H., Boudsocq, M., Coaker, G. & Liu, J. (2017). A lectin receptor-like kinase mediates pattern-triggered salicylic acid signalling. *Plant physiology,* 174**,** 2501-2514.

Malamy, J., Carr, J. P., Klessig, D. F. & Raskin, I. (1990). Salicylic acid: a likely endogenous signal in the resistance response of tobacco to viral infection. *Science,* 250**,** 1002-1004.

Maldonado, A. M., Doerner, P., Dixon, R. A., Lamb, C. J. & Cameron, R. K. (2002). A putative lipid transfer protein involved in systemic resistance signalling in Arabidopsis. *Nature,* 419**,** 399.

Mandadi, K. K. & Scholthof, K.-B. G. (2013). Plant immune responses against viruses: how does a virus cause disease? *The Plant Cell,* 25**,** 1489-1505.

Mangelsen, E., Kilian, J., Berendzen, K. W., Kolukisaoglu, Ü. H., Harter, K., Jansson, C. & Wanke, D. (2008). Phylogenetic and comparative gene expression analysis of barley (Hordeum vulgare) WRKY transcription factor family reveals putatively retained functions between monocots and dicots. *BMC Genomics,* 9**,** 194.

Maruthi, M. N., Bouvaine, S., Tufan, H. A., Mohammed, I. U. & Hillocks, R. J. (2014). Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. *PLoS One,* 9.

Maruyama, Y., Yamoto, N., Suzuki, Y., Chiba, Y., Yamazaki, K.-I., Sato, T. & Yamaguchi, J. (2013). The Arabidopsis transcriptional repressor ERF9 participates in resistance against necrotrophic fungi. *Plant Science,* 213**,** 79-87.

Matsuo, M., Johnson, J. M., Hieno, A., Tokizawa, M., Nomoto, M., Tada, Y., Godfrey, R., Obokata, J., Sherameti, I. & Yamamoto, Y. Y. (2015). High REDOX RESPONSIVE TRANSCRIPTION FACTOR1 levels result in accumulation of reactive oxygen species in Arabidopsis thaliana shoots and roots. *Molecular Plant,* 8**,** 1253-1273.

Matthes, M. C., Bruce, T. J., Ton, J., Verrier, P. J., Pickett, J. A. & Napier, J. A. (2010). The transcriptome of cis-jasmone-induced resistance in Arabidopsis thaliana and its role in indirect defence. *Planta,* 232**,** 1163-1180.

Matzinger, P. (2002). The danger model: a renewed sense of self. *Science,* 296**,** 301-305.

Mauch-Mani, B. & Flors, V. (2009). The ATAF1 transcription factor: at the convergence point of ABA-dependent plant defense against biotic and abiotic stresses. *Cell research,* 19**,** 1322.

Mcadam, S. A., Brodribb, T. J., Banks, J. A., Hedrich, R., Atallah, N. M., Cai, C., Geringer, M. A., Lind, C., Nichols, D. S. & Stachowski, K. (2016). Abscisic acid controlled sex before transpiration in vascular plants. *Proceedings of the National Academy of Sciences,* 113**,** 12862-12867.

Mcgrath, K. C., Dombrecht, B., Manners, J. M., Schenk, P. M., Edgar, C. I., Maclean, D. J., Scheible, W.-R., Udvardi, M. K. & Kazan, K. (2005). Repressor-and activator-type ethylene response factors functioning in jasmonate signalling and disease resistance identified via a genome-wide screen of Arabidopsis transcription factor gene expression. *Plant physiology,* 139**,** 949-959.

Mclellan, H., Boevink, P. C., Armstrong, M. R., Pritchard, L., Gomez, S., Morales, J., Whisson, S. C., Beynon, J. L. & Birch, P. R. (2013). An RxLR effector from Phytophthora infestans prevents re-localisation of two plant NAC transcription factors from the endoplasmic reticulum to the nucleus. *PLoS pathogens,* 9**,** e1003670.

Melotto, M., Underwood, W., Koczan, J., Nomura, K. & He, S. Y. (2006). Plant stomata function in innate immunity against bacterial invasion. *Cell,* 126**,** 969-980.

Meng, X., Xu, J., He, Y., Yang, K.-Y., Mordorski, B., Liu, Y. & Zhang, S. (2013). Phosphorylation of an ERF transcription factor by Arabidopsis MPK3/MPK6 regulates plant defense gene induction and fungal resistance. *The Plant Cell,* 25**,** 1126-1142.

Meng, X. & Zhang, S. (2013). MAPK cascades in plant disease resistance signalling. *Annual Review of Phytopathology,* 51**,** 245-266.

Métraux, J., Signer, H., Ryals, J., Ward, E., Wyss-Benz, M., Gaudin, J., Raschdorf, K., Schmid, E., Blum, W. & Inverardi, B. (1990). Increase in salicylic acid at the onset of systemic acquired resistance in cucumber. *Science,* 250**,** 1004-1006.

Miao, Y., Laun, T. M., Smykowski, A. & Zentgraf, U. (2007). Arabidopsis MEKK1 can take a short cut: it can directly interact with senescence-related WRKY53 transcription factor on the protein level and can bind to its promoter. *Plant Molecular Biology,* 65**,** 63-76.

Miao, Y., Smykowski, A. & Zentgraf, U. (2008). A novel upstream regulator of WRKY53 transcription during leaf senescence in Arabidopsis thaliana. *Plant Biology,* 10**,** 110-120.

Miao, Y. & Zentgraf, U. (2007). The antagonist function of Arabidopsis WRKY53 and ESR/ESP in leaf senescence is modulated by the jasmonic and salicylic acid equilibrium. *The Plant Cell,* 19**,** 819-830.

Miozzi, L., Napoli, C., Sardo, L. & Accotto, G. P. (2014). Transcriptomics of the interaction between the monopartite phloem-limited geminivirus tomato yellow leaf curl sardinia virus and solanum lycopersicum highlights a role for plant hormones, autophagy and plant immune system fine tuning during infection. *PLoS One,* 9.

Moffat, C. S., Ingle, R. A., Wathugala, D. L., Saunders, N. J., Knight, H. & Knight, M. R. (2012). ERF5 and ERF6 play redundant roles as positive regulators of JA/Et-mediated defense against Botrytis cinerea in Arabidopsis. *PLoS One,* 7**,** e35995.

Montillet, J.-L., Leonhardt, N., Mondy, S., Tranchimand, S., Rumeau, D., Boudsocq, M., Garcia, A. V., Douki, T., Bigeard, J. & Lauriere, C. (2013). An abscisic acid-independent oxylipin pathway controls stomatal closure and immune defense in Arabidopsis. *PLOS Biology,* 11**,** e1001513.

Mukhtar, M. S., Carvunis, A.-R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T. & Nishimura, M. T. (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science,* 333**,** 596-601.

Murray, S. L., Ingle, R. A., Petersen, L. N. & Denby, K. J. (2007). Basal resistance against Pseudomonas syringae in Arabidopsis involves WRKY53 and a protein with homology to a nematode resistance protein. *Molecular Plant-Microbe Interactions,* 20**,** 1431-1438.

Mwine, J., Van Damme, P. & Jumba, F. (2010). Evaluation of larvicidal properties of the latex of Euphorbia tirucalli L.(Euphorbiaceae) against larvae of Anopheles mosquitoes. *Journal of Medicinal Plants Research,* 4**,** 1954-1959.

Nakano, T., Suzuki, K., Fujimura, T. & Shinshi, H. (2006). Genome-wide analysis of the ERF gene family in Arabidopsis and rice. *Plant physiology,* 140**,** 411-432.

Narusaka, M., Kawai, K., Izawa, N., Seki, M., Shinozaki, K., Seo, S., Kobayashi, M., Shiraishi, T. & Narusaka, Y. (2008). Gene coding for SigA-binding protein from Arabidopsis appears to be transcriptionally up-regulated by salicylic acid and NPR1-dependent mechanisms. *Journal of general plant pathology,* 74**,** 345-354.

Návarová, H., Bernsdorff, F., Döring, A.-C. & Zeier, J. (2012). Pipecolic acid, an endogenous mediator of defense amplification and priming, is a critical regulator of inducible plant immunity. *The Plant Cell***,** tpc. 112.103564.

Nawrath, C., Heck, S., Parinthawong, N. & Métraux, J.-P. (2002). EDS5, an essential component of salicylic acid–dependent signalling for disease resistance in Arabidopsis, is a member of the MATE transporter family. *The Plant Cell,* 14**,** 275-286.

Ndamukong, I., Abdallat, A. A., Thurow, C., Fode, B., Zander, M., Weigel, R. & Gatz, C. (2007). SA-inducible Arabidopsis glutaredoxin interacts with TGA factors and suppresses JA-responsive PDF1. 2 transcription. *The Plant Journal,* 50**,** 128-139.

Ntawuruhunga, P., Ssemakula, G., Ojulong, H., Bua, A., Ragama, P., Kanobe, C. & Whyte, J. (2006). Evaluation of advanced cassava genotypes in Uganda. *African Crop Science Journal,* 14**,** I 7-25.

Nuruzzaman, M., Manimekalai, R., Sharoni, A. M., Satoh, K., Kondoh, H., Ooka, H. & Kikuchi, S. (2010). Genome-wide analysis of NAC transcription factor family in rice. *Gene,* 465**,** 30-44.

Nweke, F. 2009. *Controlling cassava mosaic virus and cassava mealybug in sub-Saharan Africa,* Intl Food Policy Res Inst.

O'donnell, P. J., Schmelz, E. A., Moussatche, P., Lund, S. T., Jones, J. B. & Klee, H. J. (2003). Susceptible to intolerance–a range of hormonal actions in a susceptible Arabidopsis pathogen response. *The Plant Journal,* 33**,** 245-257.

Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., Nakamura, H., Nishimura, Y., Ishii, S. & Sarai, A. (1996). The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nature structural biology,* 3**,** 178.

Oliver, S. (2000). Proteomics: guilt-by-association goes global. *Nature,* 403**,** 601.

Olsen, A. N., Ernst, H. A., Leggio, L. L. & Skriver, K. (2005). NAC transcription factors: structurally distinct, functionally diverse. *Trends in plant science,* 10**,** 79-87.

Olsen, K. M. & Schaal, B. A. (1999). Evidence on the origin of cassava: phylogeography of Manihot esculenta. *Proceedings of the National Academy of Sciences,* 96**,** 5586-5591.

Pandey, S. P., Roccaro, M., Schön, M., Logemann, E. & Somssich, I. E. (2010). Transcriptional reprogramming regulated by WRKY18 and WRKY40 facilitates powdery mildew infection of Arabidopsis. *Plant Journal,* 64**,** 912-923.

Pandey, S. P. & Somssich, I. E. (2009). The Role of WRKY Transcription Factors in Plant Immunity. *Plant physiology,* 150**,** 1648-1655.

Park, C. Y., Lee, J. H., Yoo, J. H., Moon, B. C., Choi, M. S., Kang, Y. H., Lee, S. M., Kim, H. S., Kang, K. Y. & Chung, W. S. (2005). WRKY group IId transcription factors interact with calmodulin. *FEBS Letters,* 579**,** 1545-1550.

Paudel, D. B. & Sanfaçon, H. (2018). Exploring the Diversity of Mechanisms Associated With Plant Tolerance to Virus Infection. *Frontiers in plant science,* 9**,** 1-20.

Pei, H., Ma, N., Tian, J., Luo, J., Chen, J., Li, J., Zheng, Y., Chen, X., Fei, Z. & Gao, J. (2013). An NAC transcription factor controls ethylene-regulated cell expansion in flower petals. *Plant physiology,* 163**,** 775-791.

Petersen, M., Brodersen, P., Naested, H., Andreasson, E., Lindhart, U., Johansen, B., Nielsen, H. B., Lacy, M., Austin, M. J., Parker, J. E., Sharma, S. B., Klessig, D. F., Martienssen, R., Mattsson, O., Jensen, A. B., Mundy, J., Louis, S. & Gene, M. P. K. (2000). Arabidopsis MAP Kinase 4 Negatively Regulates Systemic Acquired Resistance. 103**,** 1111-1120.

Pierce, E. J. & Rey, M. E. C. (2013). Assessing Global Transcriptome Changes in Response to South African Cassava Mosaic Virus [ZA-99] Infection in Susceptible Arabidopsis thaliana. *PLoS One,* 8.

Pieterse, C. M. & Van Loon, L. C. (1999). Salicylic acid-independent plant defence pathways. *Trends in plant science,* 4**,** 52-58.

Pieterse, C. M., Van Wees, S. C., Van Pelt, J. A., Knoester, M., Laan, R., Gerrits, H., Weisbeek, P. J. & Van Loon, L. C. (1998a). A novel signalling pathway controlling induced systemic resistance in Arabidopsis. *The Plant Cell,* 10**,** 1571-1580.

Pieterse, C. M. J. (1998). A Novel Signalling Pathway Controlling Induced Systemic Resistance in Arabidopsis. *the Plant Cell Online,* 10**,** 1571-1580.

Pieterse, C. M. J., Leon-Reyes, A., Ent, S. V. D. & Wees, S. C. M. V. (2009). Networking by small-molecule hormones in plant immunity. 5**,** 308-316.

Pieterse, C. M. J., Wees, S. C. M. V., Pelt, J. a. V., Knoester, M., Laan, R., Gerrits, H., Weisbeek, P. J. & Loon, L. C. V. (1998b). A Novel Signalling Pathway Controlling Induced Systemic Resistance in Arabidopsis. 10**,** 1571-1580.

Pita, J., Fondong, V., Sangare, A., Otim-Nape, G., Ogwal, S. & Fauquet, C. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *Journal of General Virology,* 82**,** 655-665.

Ponce De Leon, I., Schmelz, E. A., Gaggero, C., Castro, A., Alvarez, A. & Montesano, M. (2012). Physcomitrella patens activates reinforcement of the cell wall, programmed cell death and accumulation of evolutionary conserved defence signals, such as salicylic acid and 12-oxo-phytodienoic acid, but not jasmonic acid, upon Botrytis cinerea infection. *Molecular Plant Pathology,* 13**,** 960-974.

Pozo, M. J., Van Der Ent, S., Van Loon, L. & Pieterse, C. M. (2008). Transcription factor MYC2 is involved in priming for enhanced defense during rhizobacteria-induced systemic resistance in Arabidopsis thaliana. *New Phytologist,* 180**,** 511-523.

Prochnik, S., Marri, P. R., Desany, B., Rabinowicz, P. D., Kodira, C., Mohiuddin, M., Rodriguez, F., Fauquet, C., Tohme, J. & Harkins, T. (2012a). The cassava genome: current progress, future directions. *Tropical Plant Biology,* 5**,** 88-94.

Prochnik, S., Marri, P. R., Desany, B., Rabinowicz, P. D., Kodira, C., Mohiuddin, M., Rodriguez, F., Fauquet, C., Tohme, J., Harkins, T., Rokhsar, D. S. & Rounsley, S. (2012b). The Cassava Genome: Current Progress, Future Directions. *Tropical Plant Biology,* 5**,** 88-94.

Rampitsch, C. & Bykova, N. V. (2012). Proteomics and plant disease: advances in combating a major threat to the global food supply. *Proteomics,* 12**,** 673-690.

Rastogi, S. & Liberles, D. A. (2005). Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC evolutionary biology,* 5**,** 28.

Ratcliff, F. G., Macfarlane, S. A. & Baulcombe, D. C. (1999). Gene silencing without DNA: RNA-mediated cross-protection between viruses. *The Plant Cell,* 11**,** 1207-1215.

Rawat, R., Schwartz, J., Jones, M. A., Sairanen, I., Cheng, Y., Andersson, C. R., Zhao, Y., Ljung, K. & Harmer, S. L. (2009). REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. *Proceedings of the National Academy of Sciences,* 106**,** 16883-16888.

Ren, D., Liu, Y., Yang, K.-Y., Han, L., Mao, G., Glazebrook, J. & Zhang, S. (2008). A fungal-responsive MAPK cascade regulates phytoalexin biosynthesis in Arabidopsis. *Proceedings of the National Academy of Sciences,* 105**,** 5638-5643.

Ren, T., Qu, F. & Morris, T. J. (2000). HRT gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus. *The Plant Cell,* 12**,** 1917-1925.

Rey, C. & Vanderschuren, H. (2017). Cassava mosaic and brown streak diseases: current perspectives and beyond. *Annual review of virology,* 4**,** 429-452.

Rey, M. E. C., Ndunguru, J., Berrie, L. C., Paximadis, M., Berry, S., Cossa, N., Nuaila, V. N., Mabasa, K. G., Abraham, N., Rybicki, E. P., Martin, D., Pietersen, G. & Esterhuizen, L. L. (2012). Diversity of dicotyledenous-infecting geminiviruses and their associated DNA molecules in Southern Africa, including the South-west Indian Ocean Islands. *Viruses,* 4**,** 1753-1791.

Robatzek, S. & Somssich, I. E. (2001). A new member of the Arabidopsis WRKY transcription factor family, AtWRKY6, is associated with both senescence-and defence-related processes. *The Plant Journal,* 28**,** 123-133.

Robert-Seilaniantz, A., Grant, M. & Jones, J. D. G. (2011). Hormone Crosstalk in Plant Disease and Defense: More Than Just JASMONATE-SALICYLATE Antagonism. *Annual Review of Phytopathology,* 49**,** 317-343.

Rodrigo, G., Carrera, J., Ruiz-Ferrer, V., Del Toro, F. J., Llave, C. S., Voinnet, O. & Elena, S. F. (2012). A meta-analysis reveals the commonalities and differences in arabidopsis thaliana response to different viral pathogens. *PLoS One,* 7.

Rodríguez-Negrete, E., Lozano-Durán, R., Piedra-Aguilera, A., Cruzado, L., Bejarano, E. R. & Castillo, A. G. (2013). Geminivirus R ep protein interferes with the plant DNA methylation machinery and suppresses transcriptional gene silencing. *New Phytologist,* 199**,** 464-475.

Rogans, S. J., Allie, F., Tirant, J. E. & Rey, M. E. C. (2016). Small RNA and methylation responses in susceptible and tolerant landraces of cassava infected with South African cassava mosaic virus. *Virus Research,* 225**,** 10-22.

Rogans, S. J. & Rey, C. (2016). Unveiling the micronome of Cassava (Manihot esculenta Crantz). *PLoS One,* 11**,** e0147251.

Rosenvasser, S., Mayak, S. & Friedman, H. (2006). Increase in reactive oxygen species (ROS) and in senescence-associated gene transcript (SAG) levels during dark-induced senescence of Pelargonium cuttings, and the effect of gibberellic acid. *Plant Science,* 170**,** 873-879.

Ross, A. F. (1961). Systemic acquired resistance induced by localized virus infections in plants. *Virology,* 14**,** 340-358.

Ruan, M.-B., Guo, X., Wang, B., Yang, Y.-L., Li, W.-Q., Yu, X.-L., Zhang, P. & Peng, M. (2017a). Genome-wide characterization and expression analysis enables identification of abiotic stress-responsive MYB transcription factors in cassava (Manihot esculenta). *Journal of Experimental Botany*.

Ruan, M.-B., Guo, X., Wang, B., Yang, Y.-L., Li, W.-Q., Yu, X.-L., Zhang, P. & Peng, M. (2017b). Genome-wide characterization and expression analysis enables identification of abiotic stress-responsive MYB transcription factors in cassava (Manihot esculenta). *Journal of Experimental Botany,* 68**,** 3657-3672.

Ryals, J., Uknes, S. & Ward, E. (1994). Systemic acquired resistance. *Plant physiology,* 104**,** 1109.

Rybicki, E., Briddon, R., Brown, J., Fauquet, C., Maxwell, D., Harrison, B., Markham, P., Bisaro, D., Robinson, D. & Stanley, J. (2000). Family geminiviridae. *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses***,** 285-297.

Ryu, C. M., Murphy, J. F., Mysore, K. S. & Kloepper, J. W. (2004). Plant growth-promoting rhizobacteria systemically protect Arabidopsis thaliana against Cucumber mosaic virus by a salicylic acid and NPR1-independent and jasmonic acid-dependent signalling pathway. *The Plant Journal,* 39**,** 381-392.

Sakai, H., Hua, J., Chen, Q. G., Chang, C., Medrano, L. J., Bleecker, A. B. & Meyerowitz, E. M. (1998). ETR2 is an ETR1-like gene involved in ethylene signalling in Arabidopsis. *Proceedings of the National Academy of Sciences,* 95**,** 5812-5817.

Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K. & Yamaguchi-Shinozaki, K. (2002). DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration-and cold-inducible gene expression. *Biochemical and biophysical research communications,* 290**,** 998-1009.

Scarpeci, T. E., Zanor, M. I., Carrillo, N., Mueller-Roeber, B. & Valle, E. M. (2008). Generation of superoxide anion in chloroplasts of Arabidopsis thaliana during active photosynthesis: a focus on rapidly induced genes. *Plant Molecular Biology,* 66**,** 361-378.

Schön, M., Töller, A., Diezel, C., Roth, C., Westphal, L., Wiermer, M. & Somssich, I. E. (2013). Analyses of wrky18 wrky40 Plants Reveal Critical Roles of SA/EDS1 Signalling and Indole-Glucosinolate Biosynthesis for Golovinomyces orontii Resistance and a Loss-of Resistance Towards Pseudomonas syringae pv. tomato AvrRPS4. *Molecular Plant-Microbe Interactions*, 26**,** 758-767.

Schweizer, F., Fernández-Calvo, P., Zander, M., Diez-Diaz, M., Fonseca, S., Glauser, G., Lewsey, M. G., Ecker, J. R., Solano, R. & Reymond, P. (2013). Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. *The Plant Cell,* 25**,** 3117-3132.

Sels, J., Mathys, J., De Coninck, B. M., Cammue, B. P. & De Bolle, M. F. (2008). Plant pathogenesis-related (PR) proteins: a focus on PR peptides. *Plant physiology and biochemistry,* 46**,** 941-950.

Sharma, P., Jha, A. B., Dubey, R. S. & Pessarakli, M. (2012). Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. *Journal of botany,* 2012.

Shin, J., Heidrich, K., Sanchez-Villarreal, A., Parker, J. E. & Davis, S. J. (2012). TIME FOR COFFEE represses accumulation of the MYC2 transcription factor to provide time-of-day regulation of jasmonate signalling in Arabidopsis. *The Plant Cell,* 24**,** 2470-2482.

Shinshi, H. (2008). Ethylene-regulated transcription and crosstalk with jasmonic acid. *Plant Science,* 175**,** 18-23.

Shukla, L. I., Chinnusamy, V. & Sunkar, R. (2008). The role of microRNAs and other endogenous small RNAs in plant stress responses. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms,* 1779**,** 743-748.

Singh, K. B., Foley, R. C. & Oñate-Sánchez, L. (2002). Transcription factors in plant defense and stress responses. *Current Opinion in Plant Biology,* 5**,** 430-436.

Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G. & Kasprzyk, A. (2009). BioMart–biological queries made easy. *BMC Genomics,* 10**,** 22.

Son, G. H., Wan, J., Kim, H. J., Nguyen, X. C., Chung, W. S., Hong, J. C. & Stacey, G. (2012). Ethylene-responsive element-binding factor 5, ERF5, is involved in chitin-induced innate immunity response. *Molecular Plant-Microbe Interactions,* 25**,** 48-60.

Song, D., Xi, W., Shen, J., Bi, T. & Li, L. (2011). Characterization of the plasma membrane proteins and receptor-like kinases associated with secondary vascular differentiation in poplar. *Plant Molecular Biology,* 76**,** 97-115.

Sticher, L., Mauch-Mani, B., Métraux & Jp (1997). Systemic acquired resistance. *Annual Review of Phytopathology,* 35**,** 235-270.

Sudisha, J., Sharathchandra, R., Amruthesh, K., Kumar, A. & Shetty, H. S. 2012. Pathogenesis related proteins in plant defense response. *Plant defence: biological control*. Springer.

Sun, Z., Guo, T., Liu, Y., Liu, Q. & Fang, Y. (2015). The roles of Arabidopsis CDF2 in transcriptional and posttranscriptional regulation of primary microRNAs. *PLoS genetics,* 11**,** e1005598.

Sunkar, R., Li, Y.-F. & Jagadeeswaran, G. (2012). Functions of microRNAs in plant stress responses. *Trends in plant science,* 17**,** 196-203.

Surgonda, P. A. 2015. *CHARACTERIZATION OF TRANSGENIC GROUNDNUT (Arachis hypogaea L.) PLANTS CARRYING COAT PROTEIN GENE OF TSV AND NUCLEOCAPSID GENE OF PBNV*.

Suza, W. P. & Staswick, P. E. (2008). The role of JAR1 in jasmonoyl-L-isoleucine production during Arabidopsis wound response. *Planta,* 227**,** 1221-1232.

Suzuki, N. & Katano, K. (2018). Coordination between ROS regulatory systems and other pathways under heat stress and pathogen attack. *Frontiers in plant science,* 9**,** 490.

Suzuki, N., Miller, G., Salazar, C., Mondal, H. A., Shulaev, E., Cortes, D. F., Shuman, J. L., Luo, X., Shah, J. & Schlauch, K. (2013). Temporal-spatial interaction between reactive oxygen species and abscisic acid regulates rapid systemic acclimation in plants. *The Plant Cell,* 25**,** 3553-3569.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J. & Von Mering, C. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research,* 43**,** D447-D452.

Taj, G., Agarwal, P., Grant, M. & Kumar, A. (2010). MAPK machinery in plants: recognition and response to different stresses through multiple signal transduction pathways. *Plant signalling & behavior,* 5**,** 1370-1378.

Tameling, W. I., Elzinga, S. D., Darmin, P. S., Vossen, J. H., Takken, F. L., Haring, M. A. & Cornelissen, B. J. (2002). The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. *The Plant Cell,* 14**,** 2929-2939.

Tena, G., Boudsocq, M. & Sheen, J. (2011). Protein kinase signalling networks in plant innate immunity. *Current Opinion in Plant Biology,* 14**,** 519-529.

Teper, D., Girija, A. M., Bosis, E., Popov, G., Savidor, A. & Sessa, G. (2018). The Xanthomonas euvesicatoria type III effector XopAU is an active protein kinase that manipulates plant MAP kinase signalling. *PLoS pathogens,* 14**,** e1006880.

Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., Nomura, K., He, S. Y., Howe, G. A. & Browse, J. (2007). JAZ repressor proteins are targets of the SCFCOI1complex during jasmonate signalling. *Nature,* 448**,** 661-665.

Thomas, H., Morgan, W., Thomas, A. & Ougham, H. (1999). Expression of the stay-green character introgressed into Lolium temulentum Ceres from a senescence mutant of Festuca pratensis. *Theoretical and Applied Genetics,* 99**,** 92-99.

Tran, L.-S. P., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., Fujita, M., Seki, M., Shinozaki, K. & Yamaguchi-Shinozaki, K. (2004). Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *The Plant Cell,* 16**,** 2481-2498.

Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J. & Katagiri, F. (2009). Network properties of robust immunity in plants. *PLoS genetics,* 5**,** e1000772.

Turner, S. R. & Somerville, C. R. (1997). Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. *The Plant Cell,* 9**,** 689-701.

Van Der Biezen, E. A. & Jones, J. D. (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trends in biochemical sciences,* 23**,** 454-456.

Van Loon, L., Bakker, P. & Pieterse, C. (1998). Systemic resistance induced by rhizosphere bacteria. *Annual Review of Phytopathology,* 36**,** 453-483.

Van Loon, L. & Van Strien, E. (1999). The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiological and Molecular Plant Pathology,* 55**,** 85-97.

Van Loon, L. C., Rep, M. & Pieterse, C. M. (2006). Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol.,* 44**,** 135-162.

Van Peer, R. & Schippers, B. (1992). Lipopolysaccharides of plant-growth promoting Pseudomonas sp. strain WCS417r induce resistance in carnation to Fusarium wilt. *Netherlands Journal of Plant Pathology,* 98**,** 129.

Van Verk, M. C., Gatz, C. & Linthorst, H. J. (2009). Transcriptional regulation of plant defense responses. *Advances in Botanical Research,* 51**,** 397-438.

Vanderschuren, H., Moreno, I., Anjanappa, R. B., Zainuddin, I. M. & Gruissem, W. (2012). Exploiting the combination of natural and genetically engineered resistance to cassava mosaic and cassava brown streak viruses impacting cassava production in Africa. *PLoS One,* 7**,** e45277.

Varma, A. & Malathi, V. (2003). Emerging geminivirus problems: a serious threat to crop production. *Annals of Applied Biology,* 142**,** 145-164.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A. C., Hilbert, J.-L., Bartel, D. P. & Crété, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Molecular cell,* 16**,** 69-79.

Verhagen, B. W., Glazebrook, J., Zhu, T., Chang, H.-S., Van Loon, L. & Pieterse, C. M. (2004). The transcriptome of rhizobacteria-induced systemic resistance in Arabidopsis. *Molecular Plant-Microbe Interactions,* 17**,** 895-908.

Verkest, A., Manes, C.-L. D. O., Vercruysse, S., Maes, S., Van Der Schueren, E., Beeckman, T., Genschik, P., Kuiper, M., Inze, D. & De Veylder, L. (2005). The cyclin-dependent kinase inhibitor KRP2 controls the onset of the endoreduplication cycle during Arabidopsis leaf development through inhibition of mitotic CDKA; 1 kinase complexes. *The Plant Cell,* 17**,** 1723-1736.

Voinnet, O. (2005). Induction and suppression of RNA silencing: insights from viral infections. *Nature Reviews Genetics,* 6**,** 206.

Voinnet, O., Pinto, Y. M. & Baulcombe, D. C. (1999). Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. *Proceedings of the National Academy of Sciences,* 96**,** 14147-14152.

Wang, C., Liu, Y., Li, S.-S. & Han, G.-Z. (2015). Insights into the origin and evolution of the plant hormone signalling machinery. *Plant physiology,* 167**,** 872-886.

Wang, D., Weaver, N. D., Kesarwani, M. & Dong, X. (2005). Induction of protein secretory pathway is required for systemic acquired resistance. *Science,* 308**,** 1036-1040.

Wang, L., Ding, X., Xiao, J., Jiménez-Góngora, T., Liu, R. & Lozano-Durán, R. (2017). Inference of a geminivirus−host protein−protein interaction network through affinity purification and mass spectrometry analysis. *Viruses,* 9.

Wang, L., Zhu, W., Fang, L., Sun, X., Su, L., Liang, Z., Wang, N., Londo, J. P., Li, S. & Xin, H. (2014). Genome-wide identification of WRKY family genes and their response to cold stress in Vitis vinifera. *BMC plant biology,* 14**,** 103.

Wang, W., Barnaby, J. Y., Tada, Y., Li, H., Tör, M., Caldelari, D., Lee, D.-U., Fu, X.-D. & Dong, X. (2011). Timing of plant immune responses by a central circadian regulator. *Nature,* 470**,** 110.

Wang, X. E., Basnayake, B. V. S., Zhang, H., Li, G., Li, W., Virk, N., Mengiste, T. & Song, F. (2009). The Arabidopsis ATAF1, a NAC transcription factor, is a negative regulator of defense responses against necrotrophic fungal and bacterial pathogens. *Molecular Plant-Microbe Interactions,* 22**,** 1227-1238.

Wang, Z., Cao, G., Wang, X., Miao, J., Liu, X., Chen, Z., Qu, L.-J. & Gu, H. (2008). Identification and characterization of COI1-dependent transcription factor genes involved in JA-mediated response to wounding in Arabidopsis plants. *Plant Cell Reports,* 27**,** 125-135.

Wasternack, C. & Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Annals of botany,* 111**,** 1021-1058.

Wei, Y., Shi, H., Xia, Z., Tie, W., Ding, Z., Yan, Y., Wang, W., Hu, W. & Li, K. (2016). Genome-Wide Identification and Expression Analysis of the WRKY Gene Family in Cassava. *Frontiers in plant science,* 7**,** 1-18.

Weigel, R. R., Bäuscher, C., Pfitzner, A. J. & Pfitzner, U. M. (2001). NIMIN-1, NIMIN-2 and NIMIN-3, members of a novel family of proteins from Arabidopsis that interact with NPR1/NIM1, a key regulator of systemic acquired resistance in plants. *Plant Molecular Biology,* 46**,** 143-160.

Whitham, S. A. & Wang, Y. (2004). Roles for host factors in plant viral pathogenicity. *Current Opinion in Plant Biology,* 7**,** 365-371.

Wildermuth, M. C., Dewdney, J., Wu, G. & Ausubel, F. M. (2001). Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature,* 414**,** 562.

Wingard, S. A. (1928). Hosts and symptoms of ring spot, a virus disease of plants. *J Agric Res,* 37**,** 127-153.

Wolfe, M. D., Rabbi, I. Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., Lozano, R., Carpio, D. P. D., Ramu, P. & Jannink, J.-L. (2016). Genome-Wide Association and

Prediction Reveals Genetic Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic Improvement. *The Plant Genome,* 9**,** 1-13.

Wu, A.-M., Hörnblad, E., Voxeur, A., Gerber, L., Rihouey, C., Lerouge, P. & Marchant, A. (2010). Analysis of the Arabidopsis IRX9/IRX9-L and IRX14/IRX14-L pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. *Plant physiology,* 153**,** 542-554.

Wu, G., Park, M. Y., Conway, S. R., Wang, J.-W., Weigel, D. & Poethig, R. S. (2009a). The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. *Cell,* 138**,** 750-759.

Wu, Y., Deng, Z., Lai, J., Zhang, Y., Yang, C., Yin, B., Zhao, Q., Zhang, L., Li, Y. & Yang, C. (2009b). Dual function of Arabidopsis ATAF1 in abiotic and biotic stress responses. *Cell research,* 19**,** 1279.

Wu, Y., Zhang, D., Chu, J. Y., Boyle, P., Wang, Y., Brindle, I. D., De Luca, V. & Després, C. (2012). The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid. *Cell Reports,* 1**,** 639-647.

Würschum, T., Groß-Hardt, R. & Laux, T. (2006). APETALA2 regulates the stem cell niche in the Arabidopsis shoot meristem. *The Plant Cell,* 18**,** 295-307.

Xie, C., Zhou, X., Deng, X. & Guo, Y. (2010). PKS5, a SNF1-related kinase, interacts with and phosphorylates NPR1, and modulates expression of WRKY38 and WRKY62. *Journal of Genetics and Genomics,* 37**,** 359-369.

Yamaguchi-Shinozaki, K. & Shinozaki, K. (2006). Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.,* 57**,** 781-803.

Yang, J. W., Yi, H. S., Kim, H., Lee, B., Lee, S., Ghim, S. Y. & Ryu, C. M. (2011). Whitefly infestation of pepper plants elicits defence responses against bacterial pathogens in leaves and roots and changes the below-ground microflora. *Journal of Ecology,* 99**,** 46-56.

Yang, Z., Tian, L., Latoszek-Green, M., Brown, D. & Wu, K. (2005). Arabidopsis ERF4 is a transcriptional repressor capable of modulating ethylene and abscisic acid responses. *Plant Molecular Biology,* 58**,** 585-596.

Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., Zhiqiang, L., Yunfei, Z., Xiaoxiao, W. & Xiaoming, Q. (2006). The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Molecular Biology,* 60**,** 107-124.

Yasuda, M., Ishikawa, A., Jikumaru, Y., Seki, M., Umezawa, T., Asami, T., Maruyama-Nakashita, A., Kudo, T., Shinozaki, K. & Yoshida, S. (2008). Antagonistic interaction between systemic acquired resistance and the abscisic acid–mediated abiotic stress response in Arabidopsis. *The Plant Cell,* 20**,** 1678-1692.

Yu, D., Chen, C. & Chen, Z. (2001). Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. *The Plant Cell,* 13**,** 1527-1540.

Zhang, H., Huang, Z., Xie, B., Chen, Q., Tian, X., Zhang, X., Zhang, H., Lu, X., Huang, D. & Huang, R. (2004a). The ethylene-, jasmonate-, abscisic acid-and NaCl-responsive tomato transcription factor JERF1 modulates expression of GCC box-containing genes and salt tolerance in tobacco. *Planta,* 220**,** 262-270.

Zhang, X., Ju, H.-W., Chung, M.-S., Huang, P., Ahn, S.-J. & Kim, C. S. (2010). The RR-type MYB-like transcription factor, AtMYBL, is involved in promoting leaf senescence and modulates an abiotic stress response in Arabidopsis. *Plant and Cell Physiology,* 52**,** 138-148.

Zhang, X., Zhu, Y., Wu, H. & Guo, H. (2016). Post-transcriptional gene silencing in plants: a double-edged sword. *Science China Life Sciences,* 59**,** 271-276.

Zhang, Y. & Wang, L. (2005). The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC evolutionary biology,* 5**,** 1.

Zhang, Z.-G., Wang, Y.-C., Ji, R., Shen, G., Wang, S.-C., Zhou, X. & Zheng, X.-B. (2004b). The role of SA in the hypersensitive response and systemic acquired resistance induced by elicitor PB90 from Phytophthora boehmeriae. *Physiological and Molecular Plant Pathology,* 65**,** 31-38.

Zhang, Z., Wei, L., Zou, X., Tao, Y., Liu, Z. & Zheng, Y. (2008). Submergence-responsive microRNAs are potentially involved in the regulation of morphological and metabolic adaptations in maize root cells. *Annals of botany,* 102**,** 509-519.

Zheng, Z., Qamar, S. A., Chen, Z. & Mengiste, T. (2006). Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *The Plant Journal,* 48**,** 592-605.

Zhou, G., Wang, X., Yan, F., Li, R., Cheng, J. & Lou, Y. (2011). Genome-wide transcriptional changes and defence-related chemical profiling of rice in response to infestation by the rice striped stem borer Chilo suppressalis. *Physiologia plantarum,* 143**,** 21-40.

Zhou, M. & Luo, H. (2014). Role of microRNA319 in creeping bentgrass salinity and drought stress response. *Plant signalling & behavior,* 9**,** 1375-1391.

Zhou, X., Hua, D., Chen, Z., Zhou, Z. & Gong, Z. (2009). Elongator mediates ABA responses, oxidative stress resistance and anthocyanin biosynthesis in Arabidopsis. *The Plant Journal,* 60**,** 79-90.

Zipfel, C. (2008). Pattern-recognition receptors in plant innate immunity. *Current opinion in immunology,* 20**,** 10-16.

Zipfel, C. & Felix, G. (2005). Plants and animals: a different taste for microbes? *Current Opinion in Plant Biology,* 8**,** 353-360.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D., Boller, T. & Felix, G. (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell,* 125**,** 749-760.

Appendix

Supplementary Methods

STRING Parameter

STRING is used to display the interactions between proteins and the number of proteins displayed in the resulting network is based on selection of a confidence threshold. Within the STRING database, a user can either select a unique score or choose to set a low (0.15), medium (0.40) or high (0.7). It was decided that 0.6 was selected for the STRING networks for all the TFs. This metric was chosen for two reasons. Firstly, the predicted network produced from STRING pertains to those occurring in Arabidopsis and not cassava. Therefore, in order to infer these predicted networks in cassava, only highly studied, evolutionary conserved protein interactions should be used. For this reason, a metric closer to the predefined "high" confidence threshold was chosen. Secondly, when testing the different cut-offs, it was found that selection of a higher confidence threshold (0.7), the resulting network was considered be less insightful. This was based on the fact that at this high threshold, the network predictably showed less interactions and contained fewer nodes. So a decision was made to lower the threshold to 0.6 as to create an optimal network that both contained reasonably high confidence interactions and sufficient enough nodes and interactions for down-stream filtering.

Network reduction

Due to the large size of the resulting network in T200 at 32 dpi centring around all the TFs of interest, an additional filter step was necessary for both visualisation and analysis. As previously stated in chapter 3, there were 10 TFs that were DE both in T200 and TME3 at 32 dpi, nine of which demonstrated contrasting expression patterns. This additional filtering step was performed with the intention to visualise the network surrounding these nine DE TFs. To achieve this the filtered and annotated network featuring DE TFs and their DE interacting partners, complete with their annotated relative expression magnitudes were further filtered. The filtering entailed selecting only the nine common DE TFs between T200 and TME3 and all other interacting DEGs that are directly connected to the selected nine by an edge. The results of the filtering can be seen in Supplementary Figure 1 where those that remained within the network (Blue nodes).

In addition to Figure 3-4 requiring additional filtering, Figure 3-5 proved also be too big but required filtering for other reasons. In Figure 3-5 at 67 dpi it, was found that many of the DEG showed the same expression patterns at 32 dpi. By showing the same network, much of the same conclusions would be commented upon and discussed. To circumvent this, it was decided that Figure 3-5 would be filtered as to only show nodes, and their interactions, that were unique to T200 at 67 dpi compared to T200 at 32 dpi. The filter was done by only retaining nodes (and their edges) that were not found within the unfiltered network at 32 dpi

in to T200 (Supplementary Figure 2 Blue nodes). In doing so allowed for discussion and commentary on the unique genes and processes that were occurring at 67 dpi.

Central network

The core network (Figure 2-5) was used to compare the same set of WRKY interacting partners over the three time points. The aim behind generating this network is to compare and contrast a central network between WRKYs in both landraces. To generate this, the most interconnected DE WRKYs (proteins that had the most edges at 0.6 confidence). To determine DE WRKYs were the most interconnected, individual WRKYs were queried in the *Arabidopsis thaliana* database at 0.6 confidence threshold. Once determined, the highly interconnected WRKYs (WRKY33,40 and 53) were queried as the primary shell of the STRING search and the second shell was set to 50 interactions (white nodes) at 0.6 threshold. In Cytoscape (v2.7.1) the resulting network was then annotated with the expression values of DEGs using green and red nodes representing up- and down-regulation respectively. The intensity of colour is proportional to the magnitude of differential expression. At 32 and 67 dpi, node labels for genes that are not DE at that time point were removed for better visualization of results.

Supplementary Table 1: Portion of biological function GO enrichment results among significantly down-regulated genes in T200 at 32 dpi.

ID	Description	GeneRatio[9]	BgRatio[10]	pvalue	p.adjested[11]
GO:0010120	camalexin biosynthetic process	4/698	11/20474	0.000365	0.004721
GO:0052317	camalexin metabolic process	4/698	11/20474	0.000365	0.004721

[9] The ratio of genes in the down-regulated genes set that were annotated with the GO process

[10] Ratio of background genes that are annotated with the GO process

[11] Adjusted p-value after Bejamini-hochburg correction was applied to the p-values

Supplementary Table 2: Table of all identified TFs that were differentially expressed in T200 and TME3 throughout SACMV infection

Family	Time	Plant	ID	Cassava Gene name	Log2 Fold Change	TAIR ID	Ortholog Name
bZIP	12 dpi	T200	cassava4.1_011494m.g	MebZIP30	1.99	AT3G58120	AtBZIP61
bZIP	32 dpi	TME3	cassava4.1_021845m.g	MebZIP51	-2.36	AT3G58120	AtBZIP61
bZIP	32 dpi	TME3	cassava4.1_026149m.g	MebZIP62	-2.72	AT5G60830	AtAtbZIP70
bZIP	32 dpi	T200	cassava4.1_002147m.g	MebZIP1	-1.94	AT2G40950	AtBZIP17
bZIP	32 dpi	T200	cassava4.1_007837m.g	MebZIP16	-1.88	AT5G28770	AtBZIP63
bZIP	32 dpi	T200	cassava4.1_022517m.g	MebZIP55	3.06	AT1G75390	AtBZIP44
bZIP	32 dpi	T200	cassava4.1_028474m.g	MebZIP68	2.24	AT1G75390	AtBZIP44
bZIP	32 dpi	T200	cassava4.1_011494m.g	MebZIP30	1.99	AT3G58120	AtBZIP61
bZIP	32 dpi	T200	cassava4.1_013312m.g	MebZIP38	2.02	AT2G40620	AtBZIP18
bZIP	67 dpi	T200	cassava4.1_002147m.g	MebZIP1	-1.94	AT2G40950	AtBZIP17
bZIP	67 dpi	T200	cassava4.1_004047m.g	MebZIP3	-1.78	AT4G38900	-
bZIP	67 dpi	T200	cassava4.1_028474m.g	MebZIP68	2.24	AT1G75390	AtBZIP44
bZIP	67 dpi	T200	cassava4.1_021845m.g	MebZIP51	1.93	AT3G58120	AtBZIP61
bZIP	67 dpi	T200	cassava4.1_022517m.g	MebZIP55	3.06	AT1G75390	AtBZIP44
ERF	12 dpi	TME3	cassava4.1_015843m.g	MeERF42	-2.74	AT5G51990	AtDREB1D
ERF	12 dpi	TME3	cassava4.1_014137m.g	MeERF25	-2.44	AT4G34410	AtERF109
ERF	12 dpi	TME3	cassava4.1_015402m.g	MeERF37	-2.33	AT5G52020	AtERF025
ERF	12 dpi	TME3	cassava4.1_015361m.g	MeERF36	-2.15	AT5G51990	AtDREB1D
ERF	12 dpi	TME3	cassava4.1_027443m.g	MeERF106	2.67	AT5G51990	AtDREB1D
ERF	12 dpi	TME3	cassava4.1_016343m.g	MeERF45	-2.23	AT1G19210	AtERF017
ERF	12 dpi	TME3	cassava4.1_032424m.g	MeERF132	-1.88	AT5G44210	AtERF9
ERF	12 dpi	TME3	cassava4.1_031623m.g	MeERF128	-1.80	AT5G21960	AtERF016
ERF	12 dpi	TME3	cassava4.1_033866m.g	MeERF141	-3.36	AT5G52020	AtERF025
ERF	12 dpi	T200	cassava4.1_016343m.g	MeERF45	-2.71	AT1G19210	AtERF017
ERF	12 dpi	T200	cassava4.1_013214m.g	MeERF19	-2.58	AT5G51190	AtERF105
ERF	12 dpi	T200	cassava4.1_012438m.g	MeERF16	-2.61	AT1G64380	AtERF061
ERF	12 dpi	T200	cassava4.1_013315m.g	MeERF20	-2.36	AT5G51190	AtERF105
ERF	12 dpi	T200	cassava4.1_032424m.g	MeERF132	-2.17	AT5G44210	AtERF9
ERF	12 dpi	T200	cassava4.1_032473m.g	MeERF134	-2.03	AT4G17490	AtERF6
ERF	12 dpi	T200	cassava4.1_012714m.g	MeERF17	-2.60	AT5G47230	AtERF5
ERF	12 dpi	T200	cassava4.1_016167m.g	MeERF44	2.18	AT1G15360	AtWIN1
ERF	12 dpi	T200	cassava4.1_024234m.g	MeERF81	3.51	AT5G25390	AtSHN3
ERF	32 dpi	TME3	cassava4.1_027443m.g	MeERF106	2.67	AT5G51990	AtDREB1D
ERF	32 dpi	TME3	cassava4.1_010887m.g	MeERF13	3.03	AT5G64750	AtABR1

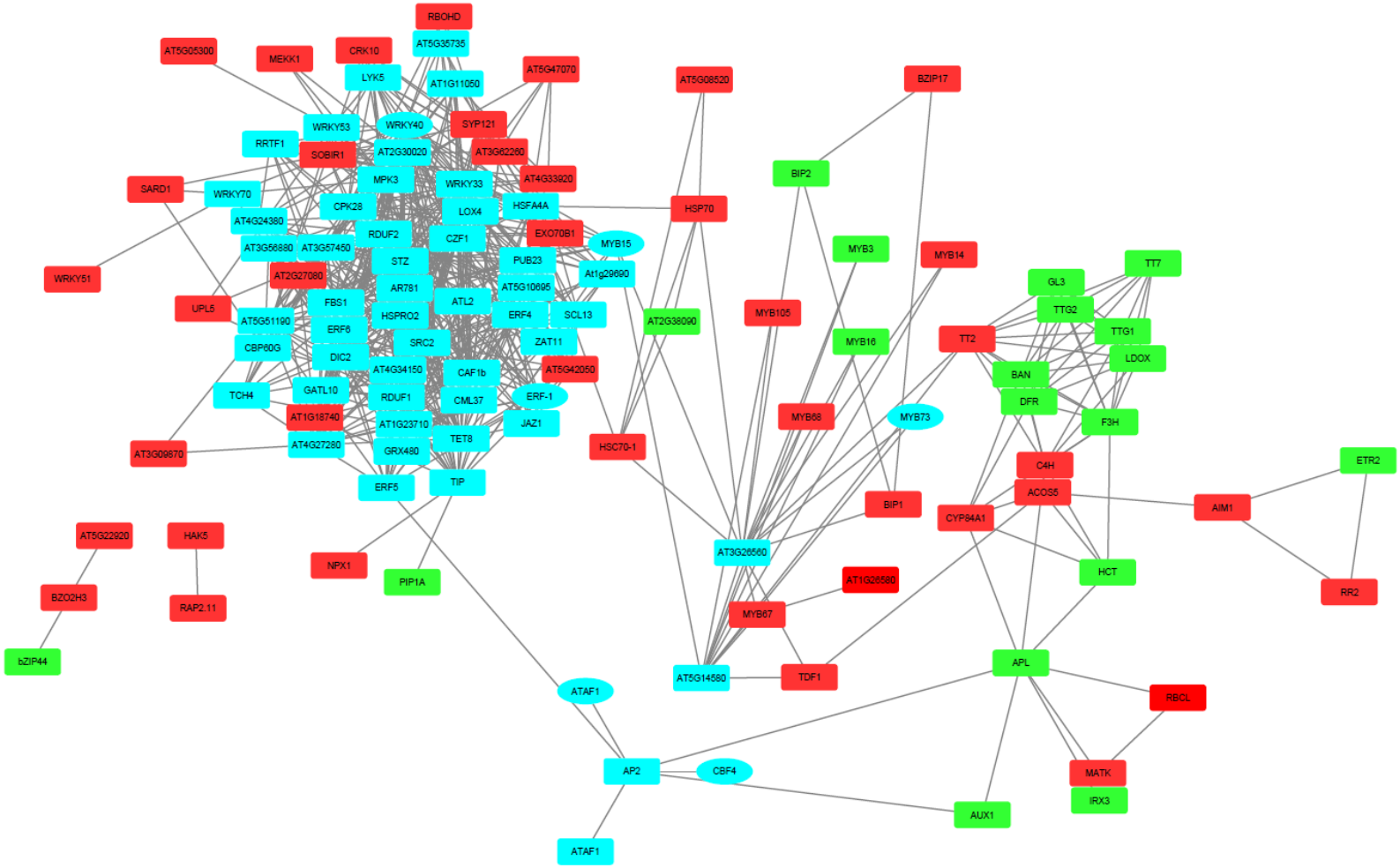
ERF	32 dpi	TME3	cassava4.1_013138m.g	MeERF18	2.23	AT4G17500	AtERF1A
ERF	32 dpi	TME3	cassava4.1_015673m.g	MeERF40	2.02	AT3G23240	AtERF1B
ERF	32 dpi	TME3	cassava4.1_023719m.g	MeERF74	1.96	AT5G51990	AtDREB1D
ERF	32 dpi	T200	cassava4.1_016343m.g	MeERF45	-2.71	AT1G19210	AtERF017
ERF	32 dpi	T200	cassava4.1_014137m.g	MeERF25	-6.29	AT4G34410	AtERF109
ERF	32 dpi	T200	cassava4.1_015402m.g	MeERF37	-5.27	AT5G52020	AtERF025
ERF	32 dpi	T200	cassava4.1_015361m.g	MeERF36	-5.14	AT5G51990	AtDREB1D
ERF	32 dpi	T200	cassava4.1_015843m.g	MeERF42	-1.88	AT5G51990	AtDREB1D
ERF	32 dpi	T200	cassava4.1_032424m.g	MeERF132	-2.17	AT5G44210	AtERF9
ERF	32 dpi	T200	cassava4.1_012714m.g	MeERF17	-2.60	AT5G47230	AtERF5
ERF	32 dpi	T200	cassava4.1_013138m.g	MeERF18	-4.02	AT4G17500	AtERF1A
ERF	32 dpi	T200	cassava4.1_013214m.g	MeERF19	-2.58	AT5G51190	AtERF105
ERF	32 dpi	T200	cassava4.1_013315m.g	MeERF20	-2.36	AT5G51190	AtERF105
ERF	32 dpi	T200	cassava4.1_032473m.g	MeERF134	-2.03	AT4G17490	AtERF6
ERF	32 dpi	T200	cassava4.1_015499m.g	MeERF38	-3.10	AT3G15210	AtERF4
ERF	32 dpi	T200	cassava4.1_014544m.g	MeERF27	-2.98	AT5G44210	AtERF9
ERF	32 dpi	T200	cassava4.1_031623m.g	MeERF128	-1.66	AT5G21960	AtERF016
ERF	32 dpi	T200	cassava4.1_014721m.g	MeERF30	-2.81	AT3G15210	AtERF4
ERF	32 dpi	T200	cassava4.1_027443m.g	MeERF106	-4.71	AT5G51990	AtDREB1D
ERF	32 dpi	T200	cassava4.1_023719m.g	MeERF74	-3.49	AT5G51990	AtDREB1D
ERF	32 dpi	T200	cassava4.1_029125m.g	MeERF117	-4.32	AT4G34410	AtERF109
ERF	32 dpi	T200	cassava4.1_027506m.g	MeERF107	-3.41	AT5G19790	AtRAP2-11
ERF	32 dpi	T200	cassava4.1_006897m.g	MeERF2	-2.09	AT2G20880	AtERF053
ERF	32 dpi	T200	cassava4.1_006833m.g	MeERF1	-2.39	AT2G20880	AtERF053
ERF	32 dpi	T200	cassava4.1_009529m.g	MeERF7	1.76	AT1G53910	AtRAP2-12
ERF	67 dpi	T200	cassava4.1_016343m.g	MeERF45	-2.71	AT1G19210	AtERF017
ERF	67 dpi	T200	cassava4.1_013214m.g	MeERF19	-2.58	AT5G51190	AtERF105
ERF	67 dpi	T200	cassava4.1_032473m.g	MeERF134	-2.03	AT4G17490	AtERF6
ERF	67 dpi	T200	cassava4.1_032424m.g	MeERF132	-2.17	AT5G44210	AtERF9
ERF	67 dpi	T200	cassava4.1_013315m.g	MeERF20	-2.36	AT5G51190	AtERF105
ERF	67 dpi	T200	cassava4.1_015499m.g	MeERF38	-3.10	AT3G15210	AtERF4
ERF	67 dpi	T200	cassava4.1_015402m.g	MeERF37	-5.27	AT5G52020	AtERF025
ERF	67 dpi	T200	cassava4.1_014137m.g	MeERF25	-6.29	AT4G34410	AtERF109
ERF	67 dpi	T200	cassava4.1_014544m.g	MeERF27	-2.98	AT5G44210	AtERF9
ERF	67 dpi	T200	cassava4.1_031623m.g	MeERF128	-1.66	AT5G21960	AtERF016
ERF	67 dpi	T200	cassava4.1_034356m.g	MeERF147	-2.40	AT1G19210	AtERF017
ERF	67 dpi	T200	cassava4.1_014721m.g	MeERF30	-2.81	AT3G15210	AtERF4
ERF	67 dpi	T200	cassava4.1_012714m.g	MeERF17	-2.60	AT5G47230	AtERF5
ERF	67 dpi	T200	cassava4.1_012438m.g	MeERF16	-2.61	AT1G64380	AtERF061
ERF	67 dpi	T200	cassava4.1_029125m.g	MeERF117	-4.32	AT4G34410	AtERF109

ERF	67 dpi	T200	cassava4.1_027443m.g	MeERF106	-4.71	AT5G51990	AtDREB1D
ERF	67 dpi	T200	cassava4.1_015843m.g	MeERF42	-1.88	AT5G51990	AtDREB1D
ERF	67 dpi	T200	cassava4.1_024772m.g	MeERF84	-2.13	AT5G13910	AtLEP
ERF	67 dpi	T200	cassava4.1_007457m.g	MeERF4	3.38	AT2G33710	-
ERF	67 dpi	T200	cassava4.1_024234m.g	MeERF81	3.51	AT5G25390	AtSHN3
ERF	67 dpi	T200	cassava4.1_034303m.g	MeERF146	1.86	AT3G20310	AtERF7
ERF	67 dpi	T200	cassava4.1_015673m.g	MeERF40	1.98	AT3G23240	AtERF1B
ERF	67 dpi	T200	cassava4.1_026245m.g	MeERF94	2.66	AT2G44940	AtERF034
ERF	67 dpi	T200	cassava4.1_021971m.g	MeERF63	1.68	AT2G44940	AtERF034
ERF	67 dpi	TME3	cassava4.1_032457m.g	MeERF133	2.03	AT5G11590	AtDREB3
ERF	67 dpi	TME3	cassava4.1_029658m.g	MeERF121	3.98	AT5G25190	AtERF003
ERF	67 dpi	TME3	cassava4.1_014721m.g	MeERF30	-1.53	AT3G15210	AtERF4
ERF	67 dpi	TME3	cassava4.1_008857m.g	MeERF5	-1.56	AT2G40340	AtDREB2C
MYB	12 dpi	TME3	cassava4.1_011111m.g	MeMYB086	-1.93	AT4G37260	AtMYB73
MYB	12 dpi	TME3	cassava4.1_010945m.g	MeMYB165	1.54	AT4G37260	AtMYB73
MYB	12 dpi	T200	cassava4.1_011931m.g	MeMYB167	-2.11	AT1G74840	-
MYB	12 dpi	T200	cassava4.1_007384m.g	MeMYB211	-1.93	AT2G20570	AtGPRI1
MYB	12 dpi	T200	cassava4.1_006216m.g	MeMYB218	-2.40	AT5G17300	AtRVE1
MYB	12 dpi	T200	cassava4.1_013424m.g	MeMYB084	-2.32	AT1G72740	-
MYB	12 dpi	T200	cassava4.1_029220m.g	MeMYB028	-1.54	AT3G12720	AtATY53
MYB	12 dpi	T200	cassava4.1_009138m.g	MeMYB295	2.00	AT5G15310	AtATMYB16
MYB	12 dpi	T200	cassava4.1_024582m.g	MeMYB313	3.73	AT3G08500	AtMYB83
MYB	12 dpi	T200	cassava4.1_022478m.g	MeMYB288	2.92	AT2G16720	AtMYB7
MYB	12 dpi	T200	cassava4.1_012929m.g	MeMYB138	2.28	AT3G54390	-
MYB	32 dpi	TME3	cassava4.1_029220m.g	MeMYB028	-2.89	AT3G12720	AtATY53
MYB	32 dpi	TME3	cassava4.1_033771m.g	MeMYB033	-3.39	AT5G11050	AtMYB64
MYB	32 dpi	TME3	cassava4.1_013503m.g	MeMYB041	-2.27	AT3G11280	-
MYB	32 dpi	TME3	cassava4.1_011587m.g	MeMYB005	-1.65	AT3G28910	AtMYB30
MYB	32 dpi	TME3	cassava4.1_023710m.g	MeMYB151	2.39	AT3G23250	AtMYB15
MYB	32 dpi	TME3	cassava4.1_011601m.g	MeMYB186	1.90	AT3G46640	AtPCL1
MYB	32 dpi	TME3	cassava4.1_026695m.g	MeMYB150	2.08	AT3G06490	AtMYB108
MYB	32 dpi	TME3	cassava4.1_010945m.g	MeMYB165	1.54	AT4G37260	AtMYB73
MYB	32 dpi	T200	cassava4.1_013493m.g	MeMYB198	-4.15	AT2G31180	AtMYB14
MYB	32 dpi	T200	cassava4.1_014242m.g	MeMYB173	-4.36	AT3G23250	AtMYB15
MYB	32 dpi	T200	cassava4.1_010485m.g	MeMYB111	-2.58	AT4G37180	AtHHO5
MYB	32 dpi	T200	cassava4.1_016153m.g	MeMYB300	-2.30	AT5G08520	AtSRM1
MYB	32 dpi	T200	cassava4.1_012036m.g	MeMYB201	-1.56	AT5G65790	AtATMYB68
MYB	32 dpi	T200	cassava4.1_007384m.g	MeMYB211	-1.93	AT2G20570	AtGPRI1
MYB	32 dpi	T200	cassava4.1_010945m.g	MeMYB165	-1.95	AT4G37260	AtMYB73
MYB	32 dpi	T200	cassava4.1_030151m.g	MeMYB232	-2.01	AT1G49560	AtHHO6

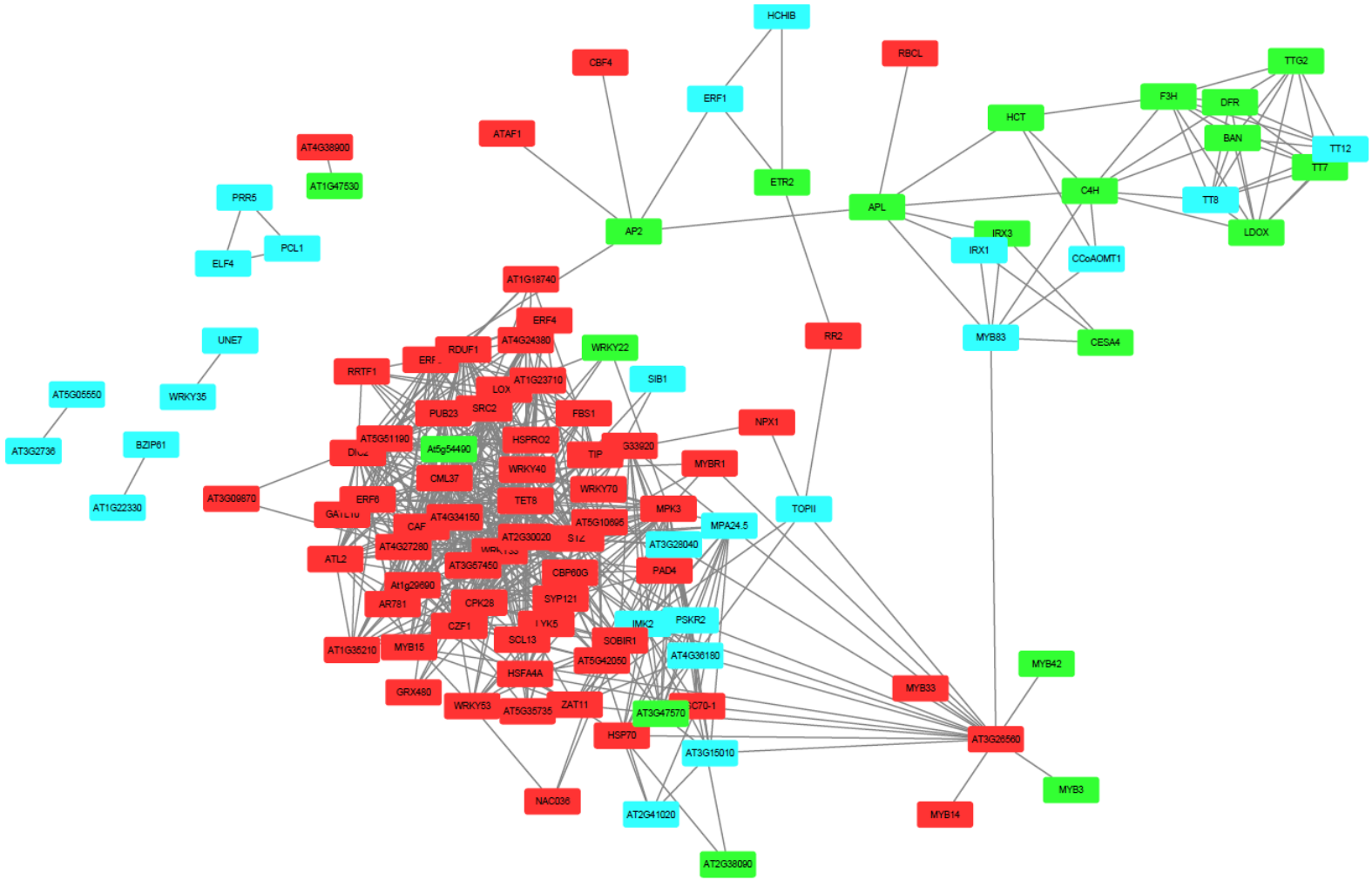
MYB	32 dpi	T200	cassava4.1_011111m.g	MeMYB086	-1.66	AT4G37260	AtMYB73
MYB	32 dpi	T200	cassava4.1_029396m.g	MeMYB294	-2.45	AT1G69560	AtMYB105
MYB	32 dpi	T200	cassava4.1_009038m.g	MeMYB219	-1.82	AT1G25550	AtHHO3
MYB	32 dpi	T200	cassava4.1_009666m.g	MeMYB029	-1.74	AT4G37180	AtHHO5
MYB	32 dpi	T200	cassava4.1_023710m.g	MeMYB151	-2.78	AT3G23250	AtMYB15
MYB	32 dpi	T200	cassava4.1_029581m.g	MeMYB271	-2.52	AT3G28470	AtMYB35
MYB	32 dpi	T200	cassava4.1_029220m.g	MeMYB028	-1.54	AT3G12720	AtATY53
MYB	32 dpi	T200	cassava4.1_033460m.g	MeMYB216	-2.15	AT4G16110	AtARR2
MYB	32 dpi	T200	cassava4.1_012254m.g	MeMYB227	2.63	AT2G38090	-
MYB	32 dpi	T200	cassava4.1_012542m.g	MeMYB179	2.41	AT2G38090	-
MYB	32 dpi	T200	cassava4.1_012592m.g	MeMYB082	3.16	AT1G22640	AtMYB3
MYB	32 dpi	T200	cassava4.1_014027m.g	MeMYB257	2.49	AT5G56840	-
MYB	32 dpi	T200	cassava4.1_022585m.g	MeMYB220	2.49	AT1G79430	AtAPL
MYB	32 dpi	T200	cassava4.1_012929m.g	MeMYB138	2.28	AT3G54390	-
MYB	32 dpi	T200	cassava4.1_009138m.g	MeMYB295	2.00	AT5G15310	AtATMYB16
MYB	32 dpi	T200	cassava4.1_012770m.g	MeMYB034	1.66	AT5G58900	-
MYB	67 dpi	T200	cassava4.1_021058m.g	MeMYB181	-1.63	AT3G08500	AtMYB83
MYB	67 dpi	T200	cassava4.1_013493m.g	MeMYB198	-4.15	AT2G31180	AtMYB14
MYB	67 dpi	T200	cassava4.1_033460m.g	MeMYB216	-2.15	AT4G16110	AtARR2
MYB	67 dpi	T200	cassava4.1_004606m.g	MeMYB262	-2.59	AT5G06100	AtMYB33
MYB	67 dpi	T200	cassava4.1_009666m.g	MeMYB029	-1.74	AT4G37180	AtHHO5
MYB	67 dpi	T200	cassava4.1_004906m.g	MeMYB172	-2.42	AT3G15010	AtUBA2C
MYB	67 dpi	T200	cassava4.1_010485m.g	MeMYB111	-2.58	AT4G37180	AtHHO5
MYB	67 dpi	T200	cassava4.1_014242m.g	MeMYB173	-4.36	AT3G23250	AtMYB15
MYB	67 dpi	T200	cassava4.1_002976m.g	MeMYB259	-1.79	AT2G44430	-
MYB	67 dpi	T200	cassava4.1_034350m.g	MeMYB139	-1.60	AT5G67300	AtMYB44
MYB	67 dpi	T200	cassava4.1_012592m.g	MeMYB082	3.16	AT1G22640	AtMYB3
MYB	67 dpi	T200	cassava4.1_008636m.g	MeMYB247	2.90	AT3G04030	AtMYR2
MYB	67 dpi	T200	cassava4.1_024052m.g	MeMYB273	2.89	AT5G06100	AtMYB33
MYB	67 dpi	T200	cassava4.1_034123m.g	MeMYB157	2.39	AT3G46640	AtPCL1
MYB	67 dpi	T200	cassava4.1_022585m.g	MeMYB220	2.49	AT1G79430	AtAPL
MYB	67 dpi	T200	cassava4.1_012254m.g	MeMYB227	2.63	AT2G38090	-
MYB	67 dpi	T200	cassava4.1_013408m.g	MeMYB283	1.72	AT5G05550	-
MYB	67 dpi	T200	cassava4.1_012929m.g	MeMYB138	2.28	AT3G54390	-
MYB	67 dpi	T200	cassava4.1_011597m.g	MeMYB258	3.31	AT3G08500	AtMYB83
MYB	67 dpi	T200	cassava4.1_014027m.g	MeMYB257	2.49	AT5G56840	-
MYB	67 dpi	T200	cassava4.1_012347m.g	MeMYB123	1.63	AT2G38090	-
MYB	67 dpi	T200	cassava4.1_031309m.g	MeMYB236	3.14	AT4G12350	AtMYB42
MYB	67 dpi	TME3	cassava4.1_033772m.g	MeMYB027	3.12	AT3G23250	AtMYB15
MYB	67 dpi	TME3	cassava4.1_026502m.g	MeMYB127	-2.59	AT3G61250	AtAtMYB17

MYB	67 dpi	TME3	cassava4.1_010339m.g	MeMYB299	-1.73	AT4G21440	AtATMYB102
MYB	67 dpi	TME3	cassava4.1_029396m.g	MeMYB294	-2.77	AT1G69560	AtMYB105
NAC	12 dpi	TME3	cassava4.1_012943m.g	MeNAC29	1.85	AT1G01720	AtNAC002
NAC	12 dpi	TME3	cassava4.1_029980m.g	MeNAC80	-3.77	AT5G22380	AtNAC090
NAC	12 dpi	TME3	cassava4.1_013132m.g	MeNAC30	-1.98	AT1G01720	AtNAC002
NAC	12 dpi	TME3	cassava4.1_026167m.g	MeNAC60	1.78	AT5G22380	AtNAC090
NAC	12 dpi	T200	cassava4.1_029980m.g	MeNAC80	-2.58	AT5G22380	AtNAC090
NAC	32 dpi	TME3	cassava4.1_010991m.g	MeNAC22	2.08	AT4G27410	AtRD26
NAC	32 dpi	TME3	cassava4.1_013664m.g	MeNAC33	3.56	AT5G22380	AtNAC090
NAC	32 dpi	TME3	cassava4.1_012943m.g	MeNAC29	1.85	AT1G01720	AtNAC002
NAC	32 dpi	TME3	cassava4.1_010999m.g	MeNAC23	1.77	AT4G27410	AtRD26
NAC	32 dpi	TME3	cassava4.1_019103m.g	MeNAC39	3.07	AT5G17260	AtNAC086
NAC	32 dpi	T200	cassava4.1_012943m.g	MeNAC29	-3.71	AT1G01720	AtNAC002
NAC	32 dpi	T200	cassava4.1_009151m.g	MeNAC9	-2.48	AT5G09330	AtNAC082
NAC	32 dpi	T200	cassava4.1_014914m.g	MeNAC36	-4.90	AT5G22380	AtNAC090
NAC	32 dpi	T200	cassava4.1_004348m.g	MeNAC3	-2.12	AT5G24590	AtNAC091
NAC	32 dpi	T200	cassava4.1_014491m.g	MeNAC34	1.88	AT5G13180	AtNAC083
NAC	67 dpi	T200	cassava4.1_012943m.g	MeNAC29	-3.71	AT1G01720	AtNAC002
NAC	67 dpi	T200	cassava4.1_019103m.g	MeNAC39	-1.66	AT5G17260	AtNAC086
NAC	67 dpi	T200	cassava4.1_004348m.g	MeNAC3	-2.12	AT5G24590	AtNAC091
NAC	67 dpi	T200	cassava4.1_015927m.g	MeNAC37	-3.64	AT2G17040	Atanac036
NAC	67 dpi	T200	cassava4.1_009151m.g	MeNAC9	-2.48	AT5G09330	AtNAC082
NAC	67 dpi	T200	cassava4.1_031698m.g	MeNAC85	-2.32	AT1G65910	Atanac028
NAC	67 dpi	T200	cassava4.1_009166m.g	MeNAC10	1.77	AT2G02450	AtNAC035
NAC	67 dpi	TME3	cassava4.1_013467m.g	MeNAC32	-2.03	AT1G69490	AtNAC029
WRKY	12 dpi	TME3	cassava4.1_011696m.g	MeWRKY79	2.44	AT1G80840	AtWRKY40
WRKY	12 dpi	TME3	cassava4.1_014368m.g	MeWRKY40	4.25	AT1G80840	AtWRKY40
WRKY	12 dpi	T200	cassava4.1_011518m.g	MeWRKY3	-2.54	AT4G11070	AtWRKY41
WRKY	12 dpi	T200	cassava4.1_011089m.g	MeWRKY55	-1.89	AT4G23810	AtWRKY53
WRKY	12 dpi	T200	cassava4.1_013417m.g	MeWRKY59	-1.84	AT3G56400	AtWRKY70
WRKY	12 dpi	T200	cassava4.1_004372m.g	MeWRKY32	2.34	AT4G26640	AtWRKY20
WRKY	12 dpi	T200	cassava4.1_010539m.g	MeWRKY44	2.30	AT4G24240	AtWRKY7
WRKY	32 dpi	TME3	cassava4.1_024650m.g	MeWRKY11	2.30	AT1G80840	AtWRKY40
WRKY	32 dpi	TME3	cassava4.1_011696m.g	MeWRKY79	2.44	AT1G80840	AtWRKY40
WRKY	32 dpi	TME3	cassava4.1_012109m.g	MeWRKY10	2.08	AT1G80840	AtWRKY40
WRKY	32 dpi	TME3	cassava4.1_004331m.g	MeWRKY17	2.28	AT1G62300	AtWRKY6
WRKY	32 dpi	TME3	cassava4.1_011518m.g	MeWRKY3	1.90	AT4G11070	AtWRKY41
WRKY	32 dpi	T200	cassava4.1_011518m.g	MeWRKY3	-2.54	AT4G11070	AtWRKY41
WRKY	32 dpi	T200	cassava4.1_011089m.g	MeWRKY55	-1.89	AT4G23810	AtWRKY53
WRKY	32 dpi	T200	cassava4.1_007752m.g	MeWRKY68	-2.60	AT2G38470	AtWRKY33

WRKY	32 dpi	T200	cassava4.1_010768m.g	MeWRKY15	-2.56	AT4G23810	AtWRKY53
WRKY	32 dpi	T200	cassava4.1_011696m.g	MeWRKY79	-2.64	AT1G80840	AtWRKY40
WRKY	32 dpi	T200	cassava4.1_024650m.g	MeWRKY11	-2.48	AT1G80840	AtWRKY40
WRKY	32 dpi	T200	cassava4.1_012575m.g	MeWRKY60	-2.60	AT1G69310	AtWRKY57
WRKY	32 dpi	T200	cassava4.1_013417m.g	MeWRKY59	-1.84	AT3G56400	AtWRKY70
WRKY	32 dpi	T200	cassava4.1_012109m.g	MeWRKY10	-2.24	AT1G80840	AtWRKY40
WRKY	32 dpi	T200	cassava4.1_016594m.g	MeWRKY42	-3.83	AT5G64810	AtWRKY51
WRKY	32 dpi	T200	cassava4.1_012154m.g	MeWRKY83	-1.54	AT3G56400	AtWRKY70
WRKY	32 dpi	T200	cassava4.1_004929m.g	MeWRKY61	-3.23	AT5G15130	AtWRKY72
WRKY	32 dpi	T200	cassava4.1_030132m.g	MeWRKY57	-4.02	AT1G30650	AtWRKY14
WRKY	32 dpi	T200	cassava4.1_016397m.g	MeWRKY70	2.21	AT2G44745	AtWRKY12
WRKY	32 dpi	T200	cassava4.1_010539m.g	MeWRKY44	2.30	AT4G24240	AtWRKY7
WRKY	32 dpi	T200	cassava4.1_007976m.g	MeWRKY50	1.93	AT2G37260	AtWRKY44
WRKY	67 dpi	T200	cassava4.1_024650m.g	MeWRKY11	-2.48	AT1G80840	AtWRKY40
WRKY	67 dpi	T200	cassava4.1_011089m.g	MeWRKY55	-1.89	AT4G23810	AtWRKY53
WRKY	67 dpi	T200	cassava4.1_013417m.g	MeWRKY59	-1.84	AT3G56400	AtWRKY70
WRKY	67 dpi	T200	cassava4.1_010768m.g	MeWRKY15	-2.56	AT4G23810	AtWRKY53
WRKY	67 dpi	T200	cassava4.1_011518m.g	MeWRKY3	-2.54	AT4G11070	AtWRKY41
WRKY	67 dpi	T200	cassava4.1_012154m.g	MeWRKY83	-1.54	AT3G56400	AtWRKY70
WRKY	67 dpi	T200	cassava4.1_007752m.g	MeWRKY68	-2.60	AT2G38470	AtWRKY33
WRKY	67 dpi	T200	cassava4.1_012575m.g	MeWRKY60	-2.60	AT1G69310	AtWRKY57
WRKY	67 dpi	T200	cassava4.1_010539m.g	MeWRKY44	2.30	AT4G24240	AtWRKY7
WRKY	67 dpi	T200	cassava4.1_007976m.g	MeWRKY50	1.93	AT2G37260	AtWRKY44
WRKY	67 dpi	T200	cassava4.1_024248m.g	MeWRKY30	2.21	AT4G01250	AtWRKY22
WRKY	67 dpi	T200	cassava4.1_014297m.g	MeWRKY54	2.79	AT2G34830	AtWRKY35
WRKY	67 dpi	TME3	cassava4.1_012109m.g	MeWRKY10	2.08	AT1G80840	AtWRKY40
WRKY	67 dpi	TME3	cassava4.1_011936m.g	MeWRKY81	-3.53	AT4G18170	AtWRKY28
WRKY	67 dpi	TME3	cassava4.1_011062m.g	MeWRKY58	-1.84	AT4G24240	AtWRKY7



Supplementary Figure 1: Full network of interacting proteins in T200 at 32 dpi around the proteins found in Figure 3-4 (Blue) . Green and red nodes represent up-and down-regulation respectively



Supplementary Figure 2: Full network of interacting proteins in T200 at 67 dpi around the proteins found uniquely to that time point (Blue). Green and red nodes represent up-and down-regulation respectively

Turnitin Originality Report

723388: W_Freeborough_MSc_Dissertation.docx by Warren Freeborough

From MSc Submission (MCBG8002_2017)

- Processed on 25-Mar-2019 12:19 AM SAST
- ID: 1098885740
- Word Count: 42304

Similarity Index

8%

Similarity by Source

Internet Sources:

3%

Publications:

8%

Student Papers:

N/A