MINING DISEASE RESISTANCE GENES IN CASSAVA USING NEXT-GENERATION SEQUENCING

GERALD AKACHI OTTI

A thesis submitted in partial fulfilment of the requirements of the University of Greenwich for the degree of Doctor of Philosophy

November, 2016

DECLARATION

I certify that this work has not been accepted in substance for any degree, and is not concurrently submitted for any degree other than that of Doctor of Philosophy (PhD) being studied at the University of Greenwich. I also declare that this work is the result of my own investigations except where otherwise identified by references and that I have not plagiarised the work of others.

Student: ……

Date…….

Mr. Gerald Akachi Otti

First Supervisor: ……..

Date……..

1) Prof. Maruthi M. N. Gowda

Second Supervisor: ……

Date…….

2) Dr. Sophie Bouvaine

ACKNOWLEDGEMENTS

My sincere gratitude to the University of Greenwich for providing me with the Vice Chancellor Scholarship VCS-NRI-03-12. I am also grateful to the Natural Resources Institute for providing funding to cover laboratory bench fees for this project through the RAE formulaic fund. Without these support, I would not have been able to embark on this study. I am highly grateful to my first supervisor Prof. Maruthi M.N. Gowda for the opportunity to undertake this study. Also for excellent professional guidance and encouragement throughout the duration of my study. To my second supervisor Dr. Sophie Bouvaine, I am very grateful for superb technical help and guidance I received.

To Natalie Morley and Dudley Farman, I am grateful for kind help with my glasshouse experiments and consumable ordering. My gratitude also goes to Saptarshi Ghosh for his friendly help in and off the laboratory.

Special thanks to my loving wife Ioana Sauluc whose support, encouragement and patience at very difficult times of my study helped see me through to the end. I will not fail to thank in a special way, my dear mother Christiana Otti. Her sacrifices, love and wise council has helped bring me to this point. To my siblings Chinwe, Ejike, Chigozie and Ugochukwu I say a huge thank you for many words of encouragements advice throughout the duration of my study.

Finally, my gratitude to Dr. Lava Kumar and Dr. Melaku Gedil of the International institute of Agriculture (IITA) Ibadan, Nigeria for introducing me the world of scientific research and for their excellent mentoring which spurred me to aspire to doing this PhD research.

LISTS OF ABBREVIATIONS

+	Positive sense
6K1	Small 6K1 peptide protein
6K2	Small 6K2 peptide protein
%	Percentage
%IncMSE	Percentage increase in mean square error
β	Beta
µg	Microgram(s)
µl	Microlitre(s)
µM	Micromole(s)
ABA	Abscisic acid
ACMBFV	*Africa cassava mosaic Burkina-Faso virus*
ACMV	*African cassava mosaic virus*
ANOVA	Analysis of variance
ASE	Allele-specific expression
ATP	Adenine tri-phosphate
Avr	Avirulence factor
BAM	Binary alignment map
BaMMV	*Barley mild mosaic virus*
BCMV	*Bean common mosaic virus*
BDMV	*Bean dwarf mosaic virus*
BLAST	Basic local alignment search tool
BMV	*Brome mosaic virus*
B. afer	*Bemisia afer*
B. tabaci	*Bemisia tabaci*
bp	Base pairs
BWT	Burrows-Wheeler Transform
BYMV	*Bean yellow mosaic virus*
CAGE	Cap analysis gene expression

CBSD	Cassava brown streak disease
CBSV	*Cassava brown streak virus*
CC	Coiled-coil domain
cDNA	complementary deoxyribonucleic acid
CI	Cylindrical inclusion
CMD	Cassava mosaic disease
CMB	Cassava mosaic begomovirus
CMBs	Cassava mosaic begomoviruses
CMMGV	*Cassava mosaic Madagascar virus*
CMV	*Cucumber mosaic virus*
CP	Coat protein or capsid protein
CTAB	Cetyltrimethylammonium bromide
COOH	Carboxylic acid
DCL	Dicer-like protein
DNA	Deoxyribonucleic acid
dNTP	deoxy nucleoside triphosphate
dRNA	double-stranded ribonucleic acid
EACMV	*East African cassava mosaic virus*
EACMCV	*East African cassava mosaic Cameroon virus*
EACMKV	*East African cassava mosaic Kenya virus*
EACMZV	*East African cassava mosaic Zanzibar virus*
EACMV-UG	*East African cassava mosaic virus-Uganda*
eIF4A	Eukaryotic initiation factor 4A
eIF4E	Eukaryotic initiation factor 4E
eIF(iso)4E	Isomer of eukaryotic initiation factor 4E
eIF4G	Eukaryotic initiation factor 4G
EMBL	European Molecular Biology Laboratory
ENA	European Nucleotide Archive
eQTL	expression quantitative trait locus

ET	Ethylene
et. al	And others
FAO	Food and Agriculture Organization
FAOSTAT	Food and Agriculture Organisation Statistics
FPKM	Fragments per kilo base per million reads
g	Gram(s)
GO	Gene ontology
HC-Pro	Helper-component-proteinase
HR	Hypersensitive response
Kb	Kilobase(s)
LRR	Leucine-rich repeat
m	Metre(s)
M	Molar
MAF	Major allele frequency
Mb	Megabase(s)
M. esculenta	*Manihot esculenta*
M. glaziovii	*Manihot glaziovii*
MeSA	Methyl salicylic acid
mg	Milligram(s)
Mg	Magnesium
MgCl2	Magnesium chloride
min	Minute(s)
miRNA	micro ribonucleic acid
ml	Millilitre(s)
mm	Millimetre(s)
MNSV	*Melon necrotic spot virus*
M. persicae	*Myzus persicae*
MPSS	Massively parallel signature sequencing
mRNA	messenger ribonucleic acid

NBS	Nucleotide binding site domain
NCBI	National Center for Biotechnology Information
NIa-Pro	Nuclear inclusion a protein
NIb	Nuclear inclusion b
nm	Nanometre (s)
^{o}C	Degree Celsius
ORF	Open reading frame
ORFs	Open reading frames
QTL	Quantitative trait locus
PAMPs	Pathogen-associated molecular patterns
P1	First protein
P3	Third protein
PCR	Polymerase chain reaction
PeMV	*Pea enation mosaic virus*
PPV	*Plum pox virus*
PRR	Pattern recognition receptors
PSbMV	*Pea seed-borne mosaic virus*
P. syringae	*Pseudomonas syringae*
PTGS	Post-transcriptional gene silencing
PTI	Pattern-triggered immunity
PVA	*Potato virus A*
PVX	*Potato virus X*
PVY	*Potato virus Y*
qPCR	Quantitative PCR
RAG	Resistance-associated gene
RDR	RNA-directed RNA polymerases
R gene	Resistance gene
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid

RNAi	Ribonucleic acid interference
RNA-seq	Ribonucleic acid sequencing
rpm	Revolution per minute(s)
RPKM	Reads per kilo base per million reads
RYMV	*Rice yellow mosaic virus*
SA	Salicylic acid
SACMV	*South African cassava mosaic virus*
SAGE	Serial analysis gene expression
SAR	Systemic acquired resistance
SDS	Sodium dodecyl sulphate
siRNA	Small interfering ribonucleic acid
SNP	Single nucleotide polymorphism
SPMMV	*Sweet potato mild mottle virus*
sRNA	Small ribonucleic acid
ssDNA	Single stranded deoxyribonucleic acid
STNV	*Satellite tobacco necrosis virus*
Taq	*Thermophilus aquaticus*
TCV	*Turnip crinkle virus*
TED	Transfer enhancer domain
TEV	*Tobacco etch virus*
TGB	Triple gene block
TIR	Toll interleukin receptor
TMV	*Tobacco mosaic virus*
ToMV	*Tomato mosaic virus*
tRNA	Transfer ribonucleic acid
TSWV	*Tomato spotted wilt virus*
TuMV	*Turnip mosaic virus*
TVMV	*Tobacco vein mottling virus*
UCBSV	*Ugandan cassava brown streak virus*

UTR	Untranslated region
VPg	Viral protein genome-linked
vsiRNA	Virus-derived small interfering ribonucleic acid
WGS	Whole genome sequencing
Y2H	Yeast-two hybrid

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Cassava brown streak disease (CBSD) remains a major threat to cassava productivity hence to food security and livelihood of over half a billion people in sub-Saharan Africa. Exploitation of natural resistance is generally accepted as the most sustainable means to control the disease. Most of the existing resistance sources have been identified based on evaluation of resistance to infection in the field where escapes is not uncommon. This limitation alongside the need to enhance knowledge on the currently poorly understood molecular processes underlying CBSD resistance and susceptibility gave impetus to this project. The project was therefore designed around identifying new sources of resistance to *Cassava brown streak virus* (CBSV) and understanding molecular mechanisms underlying natural resistance. A multiplex real time PCR method was developed for quantification of CBSVs alongside the DNA viruses of cassava – *African cassava mosaic virus* and *East African cassava mosaic virus* – in a single tube. The method was highly sensitive and reliably quantified cassava viruses and multiplexing did not diminish sensitivity or accuracy. Evaluation of responses to controlled CBSV infection classified cassava accessions as CBSD resistant, tolerant or susceptible based on foliar and root CBSV quantities. Average CBSV quantity were up to 45 times lower in resistant compared to susceptible cassava. Resistance to CBSV inoculation in the two accessions – Mkumba and Pwani was demonstrated for the first time. Transcriptome analysis of 48 samples comprising eight CBSV- and mock-inoculated cassava accessions sampled at one, five and eight weeks after inoculation showed that the cassava transcriptome is very dynamic. About 68% of the expressed genes were found to change over time. Transcription of genes encoding antioxidant defense, pathogenesis-related and cell expansion functions were positively modulated by CBSV infection, in susceptible cassava but repressed in the resistant ones. Genes which function in plant adaptive response to abiotic stress were induced in both accessions but substantially more so in susceptible accessions. Unique transcriptional activity of CBSD-resistant cassava was defined by overexpression of nucleotide binding site / leucine-rich repeat (NBS-LRR) resistance genes. Data from RNA-sequencing of the cassava samples was also applied, for the first time, to the analysis of allele expression at individual single nucleotide polymorphic (SNP) loci. Higher proportion of loci were expressed as heterozygous alleles in resistant compared to susceptible and tolerant cassava. This observation was associated with the introgression of alleles from the wild cassava – *Manihot glaziovii*. Genome segments ranging from 0.1 to 8 megabases in chromosomes 3, 4 and 13 were found to contain *M. glaziovii* haplotypes common and unique to CBSV-resistant accessions. A synthesis of results from analyses of allele and gene expression suggests that a more pronounced activity of the plant immunity pathway dissociated from hypersensitive response leads to quick control of CBSV replication upon infection. This, and peculiar genetic variations underlie the low virus quantity and under-expression of stress-associated genes characteristic of CBSD-resistant cassava.

CHAPTER 1: Introduction

Cassava (*Manihot esculenta* Crantz), a root tuber rich in carbohydrates, is consumed as food by 800 million people worldwide, most of them in sub-Saharan Africa (Thresh *et al*., 2006). According to the world food balance sheet, cassava is the eight most important food crop globally and the most important in Africa based on production levels (FAOSTAT, 2013). The crop is particularly attractive to farmers because it is easy to cultivate and produces satisfactory yields even under arid and generally poor soil conditions (Mtunda *et al*., 2003). Its root is consumed as an important source of carbohydrate diet for humans and animals (Westby, 2002). Cassava roots are also important raw material for chemical, food and energy industries. Notably, its potential for biofuel production has been highlighted and demonstrated (Adelekan, 2010; Balagopalan, 2002). In spite of its obvious importance and potential, production is still hugely limited by two major diseases namely cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) in sub-Saharan Africa. CBSD is caused by two RNA virus species; *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) both of which belong to the genus *Ipomovirus* and family *Potyviridae*. Both viruses spread by vegetative propagation of infected cassava cuttings, and naturally by the insect vector whitefly *Bemisia tabaci* (Maruthi *et al*., 2005). For several decades since its first description, CBSD was thought to be largely confined to the lowland coastal areas of eastern Africa not above 1000 m above sea level and inland around the shores of Lake Malawi (Hillocks and Jennings, 2003; Nichols, 1950). It has however, recently assumed a pandemic status with severe outbreaks reported inland in Uganda (Alicai *et al*., 2007) and areas around Lake Victoria that includes northwest Tanzania and Western Kenya (Legg *et al*., 2011). More recently, there has been report of incidence in Burundi, Rwanda and D.R. Congo suggesting a westward movement of the virus (Bigirimana *et al*., 2011; Mulimbi *et al*., 2012). CBSD can be very devastating as it affects the most useful part of cassava namely the root which is rendered unfit for any use by formation of brownish corky necrotic patches that can cover entire starch root tissue (Hillocks and Jennings, 2003). Root yield loss of up to 70% has been reported due to CBSD (Hillocks *et al.*, 2001). Disease control efforts involving production and distribution of clean planting material derived from virus-cured tissue culture stocks are currently underway (Maruthi *et al*., 2014a). Though useful, this approach is still not sustainable in the long term as susceptible varieties can be re-infected under high disease pressure in the field. It is generally accepted that natural resistance is the best strategy to control plant virus diseases. To this end, screening to

identify resistance or tolerance sources to field and controlled infection with CBSD-causing viruses has been carried out (Kaweesi *et al*., 2014; Ogwok *et al*., 2014). However, more and better sources of resistance would be required to meet the broad spectrum of farmer preferences in the short to medium term. More crucially, there is the need to develop 'novel and robust molecular markers' for breeding resistance to CBSD (Patil *et al*., 2015) for a more sustainable control of the disease. Recent advances in genomics such as RNA-seq have been applied to identify genes uniquely over-expressed in the CBSD-resistant variety Kaleso compared to the susceptible variety Albert (Maruthi *et al*., 2014b). Application of next-generation whole transcriptome sequencing methods such as RNAseq to the identification CBSD resistance gene markers are still in its very early days. There is the need to apply new and emerging transcriptomic data analysis methods to understand the sequence variations and regulatory mechanism underlying resistance to the more serious causative agent of CBSD, namely CBSV. In view of these, this research has been designed around achieving the following objectives:

1. Develop new and efficient techniques for DNA and RNA virus detection to screening for combined field resistance to CMD and CBSD.
2. Identify new sources of resistance to CBSD among cultivated landraces of east Africa.
3. Identify genes and biological processes associated with resistance and susceptible responses to CBSV infection.
4. Identify putative functional markers for CBSD resistance based on allele-specific gene expression analysis.

CHAPTER 2: Literature review

2.1 Cassava

Cassava is a woody shrub belonging to the *Euphorbiaceae* family (Alves, 2002). The plant can grow up to 4 metre high at physiological maturity (Legg *et al*., 2015) and could be un-branched or branched. The leaves are lobed often with an uneven number of lobes ranging from three to nine. Cassava is monoecious with its male and female flowers separate but mostly located on same plant (Alves, 2002). A trilocular capsule-shaped fruit develops from the fertilized female flower and when dry releases ovoid-ellipsoidal-shaped seeds upon dehiscence.

Cassava is predominantly propagated asexually from stem cuttings though new cassava plants can be generated sexually from germinated seeds. Cassava stem cuttings or stakes for propagation are typically 20 cm long and should contain 4 to 5 nodes to ensure viability. Cassava cuttings are traditionally planted in vertical, slanting or horizontal positions depending on cassava cultivar, soil characteristics and climate (Leihner, 2002). The specific influence of each of these factors on germination, growth and harvesting of cassava was extensively reviewed in Leihner (2002). Asexual propagation via stem cuttings significantly restricts productivity as a maximum of thirty stem cuttings can be generated from a mother plant after 12 to 18 months (Leihner, 2002). Cock *et al.* (1985) devised a method for increasing productivity per mother cassava plant. This involved growing cassava leaves excised with their axillary buds in a mist chamber for sprouting and root formation of the propagules and subsequent transfer to a peat pot and then to the field after two to three weeks. Although the system is labour-intensive, it could produce up to 300,000 commercial stakes from a single source stake within 18 months. Cassava propagation by stem cuttings is a prominent means of spreading pathogens and pest. Micro propagation of cassava by the *in vitro* technique of tissue culture, under aseptic conditions can reduce pathogen transmission rates and increase production volume because it generates larger number of tissue culture propagules in shorter time. The ability to generate new propagules from virus-free apical meristematic portions of virus-infected plants and suitability for treatment with chemical virucidal agents makes the tissue-culture technique potentially useful for generation of virus-free planting materials. Recently, an *in vitro* technique which applies thermotherapy has been applied to produce CBSV-free micro-propagated cassava propagules (Maruthi *et al*., 2014a).

Africa accounts for about half of world cassava production with Nigeria as the leading producer in the World (Nweke, 2004). The continent recorded an annual production growth rate of 2.9% for the three decades between 1965 and 1995 though without a concomitant increase in production per capita (Hillocks, 2002). This growth rate saw Africa taking over the position of world leading producer in the 1990's (FAO, 2005). Africa's rise to dominance in world cassava production has been attributed to various factors. A review of cassava production in Africa (FAO, 2005) identified three factors responsible for the observed growth: rapid population growth, increased poverty which drives reliance on cheaper calorie alternatives and the effects of genetic research and improved agronomic practices. Hillocks (2002) thinks that increase in land under cassava cultivation as against actual increase in production per hectare explains the increase in production growth rate. Put together, it is likely that increased poverty and population growth resulted in the increased land area brought under cassava cultivation.

Cassava has myriads of inherent problems which impact on its production. The prevalent propagation method which is vegetative propagation via stem cuttings is inherently a slow method. The planting process is rather laborious and slow as cassava stem is bulky, restricting easy transportation to farm site and require cutting prior to planting (FAO, 2001). Also, owing to the crowded nature of a cassava field, cassava production may not be amenable to quicker production via mechanization.

In Africa, shortening fallow periods apparently due to increasing pressure on land has resulted in replacement of more nutrient-demanding crops with cassava (Hillocks, 2002). However, having a cropping cycle longer than short fallow can allow, the crop is often harvested early to meet high production turnover characterizing short fallows. In the process, cassava is produced only at a fraction of capacity expected at maturity. Also, loss of genetic diversity occasioned by farmer preference for newer varieties has resulted in an insufficiency of germplasm adaptable to semi-arid and mid-altitude zones. Hence, the broad genetic base required for generation of quality high yielding germplasm for new industrial uses may therefore be lacking (Hillocks, 2002).

Perhaps the most important constraint to cassava production especially in sub-Saharan Africa comes from diseases and pests. A vast array of pests which transmit various pathogens that cause diseases as well as physical damage to the cassava plant has been identified. Hillocks (2002) observes that the lethality of pest attack is heightened by accidental introduction of exotic pests to

which cassava is susceptible and against which antagonists or natural enemies are lacking. Some cassava pests and pathogens which devastated cassava in the past have now been put under control. For example, cassava mealybug spread throughout the cassava belt of Africa after being introduced in the Congo in infected material from South America in the 1970s. Within ten years, the pest significantly affected production but was however controlled mainly by biological means using its natural predator, *Epidinocarisi lopezi* whose larvae live on and kill cassava mealybug (Herren *et al.*, 1987; Herren and Neuenschwander, 1991; Neuenschwander, 2001).

More recently, viral diseases of cassava have been particularly devastating in sub-Saharan Africa. The outbreak of a form of cassava mosaic disease (CMD) caused by the *East African cassava mosaic virus*-Uganda variant (EACMV-Ug) in Uganda led to very severe loss in production which resulted in a major food crisis (Otim-Nape *et al*, 2000). This outbreak was contained mainly through the deployment of resistant elite cassava varieties (Otim-Nape and Buea, 2000; Ssemakula, 1997). Another equally serious disease of cassava is the cassava brown streak disease (CBSD) which has been described as presenting the greatest threat to millions of cassava growers in east and central Africa (Legg *et al*., 2014).

2.2 Cassava brown streak Disease

2.2.1 Symptoms

CBSD is a serious disease of cassava which reduces root quality and yield. Its degenerative effect on yield and quality arises from symptoms of brownish necrotic patches which renders cassava root unusable (Hillocks, 2001). Besides root symptoms, CBSD produces variety of symptoms on leaves and stems as well. These were first described by Storey (1936) but were more elaborately described by Nichols (1950). Leaf symptoms of CBSD are very variable but the most easily recognizable one is the feathery chlorosis which develops along secondary and tertiary veins as well as yellow blotches that are closely associated to veins especially in older leaves (Calvert and Thresh, 2002). These symptoms may or may not be conspicuous (Calvert and Thresh, 2002). Stem symptom does not always occur but when it does, manifests as purple/brown markings on green bark which extend to the cortex in older stems. Shoot die-back may occur in varieties with root necrosis (Hillocks and Jennings 2003). As shown in Figure 1, severe infection results in symptom

expression all over the plant with necrosis leading to abscises in leaves, and deep discoloration and die-back in shoots both of which may cause plant death (Calvert and Thresh, 2002). Due to the nature of its symptoms, CBSD causes significant yield loss of the most important part of the cassava plant i.e. the roots, as well as significant reduction in its quality. This has implications on food security and livelihoods of farmers. For instance, root yield loss of up to 70% has been reported in highly susceptible varieties (Hillocks *et al*., 2001). McSween *et al.* (2006) suggest that severe infection can result in a 90% reduction in the market value of the crop.

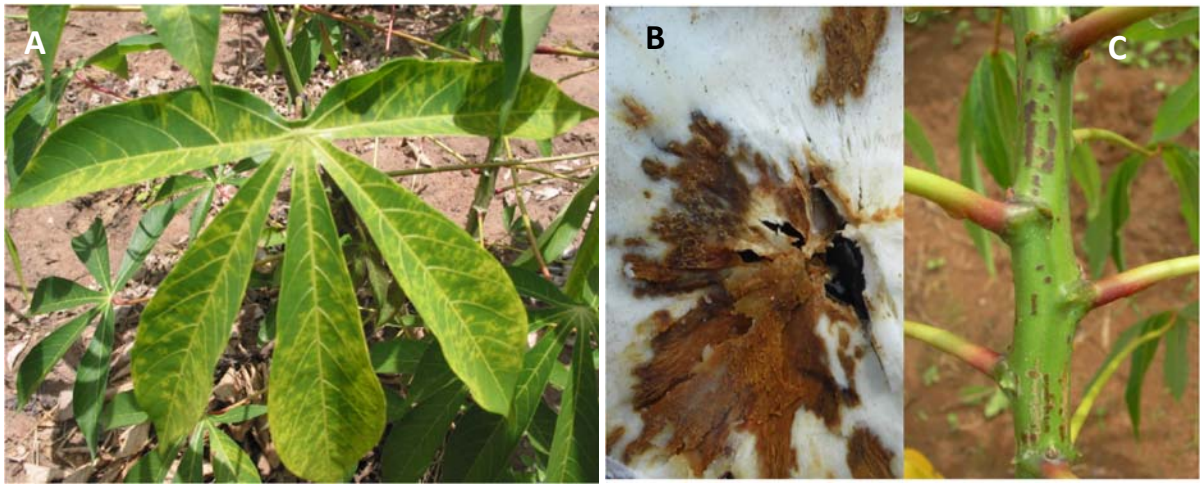


Figure 1. Typical CBSD symptoms on: (A) cassava leaves (Source: Hillocks, 2005) (B) cassava storage root and (C) stem (Source: Obonyo et al., 2010).

2.2.2 Aetiology

When CBSD was first described, its causal agent was unknown. However, indication of a likely viral aetiology emerged when Lister (1959) confirmed that the disease was sap-transmissible from cassava to a range of herbaceous indicator hosts. Nonetheless, the viral aetiology of CBSD remained a speculation until Lennon *et al.* (1986) detected virus particles in infected cassava leaf samples using electron microscopy. These particles were later found to be elongate, flexuous filaments 650 – 690nm long containing 'pin-wheel' inclusions, typical of potyviruses (Harrison *et al*., 1995). The specific causative virus for CBSD was determined to be *Cassava brown streak virus* (CBSV) (family *Potyviridae*; genus *Ipomivirus*) after a predicted amino acid sequence of

virus isolate from coastal Tanzania was found to show the closest similarity to the coat protein of *Sweet potato mild mottle virus* (SPMMV) – the type species for the genus *Ipomovirus* (Monger *et al.*, 2001a). High nucleotide sequence similarity and the use of common CBSV-specific primers to amplify fragments in infected cassava from three locations in Tanzania, and two locations in Mozambique showed that CBSV is the causative agent of CBSD prevalent in these regions (Monger *et al.,* 2001b). Previously, Bock (1994) observed that two distinct variants of virus isolates occurred together in CBSD-infected cassava in Kenya. The variants were isolated separately and subsequently maintained in *Nicotiana debneyi* where they showed distinct characteristic symptoms (Bock, 1994).

Alicai *et al.* (2007) reported a re-emergence of CBSD in Uganda. Molecular analysis revealed that the outbreak was caused by a CBSV isolate which had coat protein (CP) nucleotide sequence not significantly distinct from previously reported CP sequences from coastal Tanzania and Mozambique, hence was considered a new species. More recently, Mbanzibwa *et al*. (2009a) discovered that eight CBSV isolates from Lake Victoria basin in Uganda and Tanzania were genetically distinct from six CBSV isolates from lowland coastal Tanzania and Mozambique. Full-length genome sequencing for one of the eight Lake Victoria isolates, namely the MLB3 isolate was done for the first time by Mbanzibwa *et al*. (2009b) who found its genome to contain a novel 226 amino acid insert predicted to be a homolog of the Maf/Ham1 protein. Phylogenetic analysis of full length genome sequence of CBSVs revealed a separate clade for isolates from Kenya, Uganda, Malawi, north-western Tanzania and the highland MLB3 isolate (Winter *et al*., 2010). These have nucleotide sequence similarity of only 70% to the coastal Tanzania and Mozambique clade hence were designated a new species and named *Ugandan cassava brown streak virus* (UCBSV) by the International Committee for Taxonomy of Viruses (Mbanzibwa *et al*., 2011).

CBSV and UCBSV are collectively referred to as cassava brown streak viruses (CBSVs). These have a common genome structure, which differ slightly from that of other *Ipomoviruses* (Winter *et al.*, 2010; Mbanzibwa 2009b). Isolates of both viruses are highly heterogeneous with nucleotide sequences ranging from 9008 to 9070 (Winter *et al*., 2010; Mbanzibwa *et al*., 2009b). As other viruses of the same family, CBSV and UCBSV are positive-sense single-stranded RNA ((+)ssRNA) viruses expressed as a polyprotein which undergoes post-translational proteolysis into nine separate independently functional proteins (Mbanzibwa *et al.*, 2009b) (Figure 2). Unlike most

potyviruses however, genomes of CBSVs lack an *N*-terminal helper-component proteinase but contains a single serine proteinase (P1) implicated in a counter defence-related suppression of RNA silencing and a Maf/HAM1-like pyrophosphatase (Mbanzibwa *et al.*, 2009b; Winter *et al*., 2010).

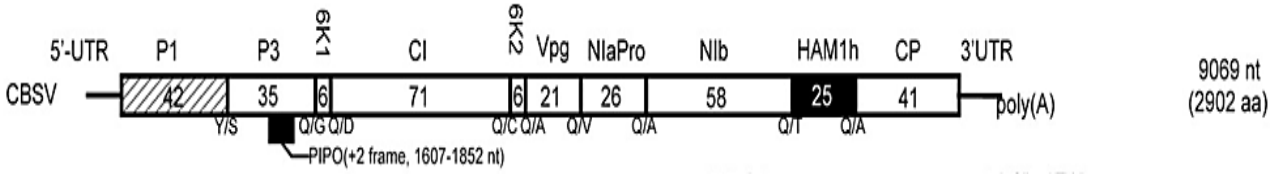


Figure 2. Genome structure of CBSV showing a 9069 nucleotide long polyprotein transcript (Mbanzibwa et al., 2009b). The component proteins are, from 5' to 3' end, First protein (P1), third protein (P3), small 6K1 peptide protein, Cylindrical inclusion (CI), small 6K2 peptide protein (6K2), viral protein genome-linked protein (Vpg), nuclear inclusion a protein (NIaPro), nuclear inclusion b (NIb), putative nucleoside triphosphate pyrophospatase (HAM1h), and coat protein (CP).

2.2.3 Transmission

Transmission of CBSVs ensures the continued menace of CBSD to cassava production. The two causative virus species have been proven to be transmissible between cassava plants by graft inoculation (Mohammed *et al*., 2012; Wagaba *et al*., 2013) and mechanically to a number of herbaceous plants (Mohammed *et al*., 2012). Prior to the identification of the insect vector, research on the possible vector responsible for the transmission of CBSVs were focused on the sap-feeding insects such as the whitefly species (*Bemisia tabaci* and *Bemisia afer*) and aphid, *Myzus persicae* (Calvert and Thresh, 2002). However, a number of CBSD transmission experiments with whitefly and aphid failed. Lennon *et al.* (1986) was unable to transmit the CBSV using *M. persicae*. Transmission tests using *B. tabaci*, and six species of aphids including *M. persicae* was also unsuccessful (Bock, 1994). The latter study however, advocated *B. afer* as a likely vector for transmission of CBSD virus, apparently on the basis of reports of its population peaking in areas where CBSD incidence was highest (Robertson, 1985). *B. afer* species was found

to be the predominant whitefly species in Malawi (Munthali, 1992) although it formed a small proportion of total whitefly population in parts of southern Tanzania and over 50% of the population in 1996 at lowland areas of Tanzania (Hillocks and Jennings, 2003). Hillocks and Jennings (2003) made a case for the role of 'a whitefly' as possible vector of the CBSV by noting that a closely related virus, SPMMV was transmitted by whiteflies. Observation of correlation between new CBSD incidences and whitefly populations in Naliendele Tanzania had earlier been made (Hillocks and Jennings, 2003). These observations, which suggested a whitefly transmission of CBSV, led to further whitefly-based transmission studies. The first successful transmission of the virus with whiteflies was reported by Maruthi *et al.* (2005) at 22% efficiency for field-collected *B. tabaci*. This finding was corroborated by Mware *et al.* (2009) although they observed a higher transmission rate of 40.7% with *B. tabaci.*

2.2.4 Epidemiology and current epidemics

Storey (1936) made the first report of CBSD from the foothills of the Usumbara mountains in Tanzania. For many decades following this, the disease was thought to be restricted to lowland areas along the coast of the Indian Ocean. Very early reports of the incidence of the disease indicated wide distribution along areas currently regarded as endemic to CBSD (Legg *et al.*, 2011) which are all cassava-growing regions along the coast from the north-eastern corner of Kenya to the southern fringes of Tanzania (Nichols, 1950). Surveys in Tanzania and Malawi confirm the coastal distribution pattern of the disease (Legg and Raya, 1996) and also observed widespread infection at high incidence along Lake Malawi shores (Sweetmore, 1994). From the first report of CBSD up until the last decade of the twentieth century, only pockets of infection had been reported at high altitudes above 1000 m in parts of Tanzania, Kenya and Uganda, (Jennings, 1960; Jameson, 1964; Bock, 1994). In the Mozambican coastal provinces of Zambezia and Nampula with high concentration of cassava cultivation, disease incidence was as high as 80 – 100% in some fields (Hillocks *et al.*, 2002).

Unlike CMD, outbreaks of CBSD appear to happen independently in isolated regions instead of systematically spreading out from a single region (Legg *et al.*, 2011). There has been a couple of outbreaks in Uganda, the first in 1934 from infected material introduced from Tanzania (Jameson, 1964) and the second in 2004 (Alicai *et al.*, 2007). Evidence also exists of outbreaks in north-

western Tanzania and western Kenya – areas previously without reports of CBSD infection (Legg *et al.*, 2011). Recent survey data from six countries of the Great lakes region suggest that CBSD incidence was highest in these non-endemic areas, specifically in Mara region of north-western Tanzania and in southern Uganda (Anon, 2010). In view of the fact that outbreaks have not been limited to a single region but reported in multiple regions in more than one country, the disease has been assigned a pandemic status (Legg *et al.*, 2011). Mbanzibwa *et al*. (2011) and Abarshi *et al*. (2012) reported co-infection of CBSV and UCBSV in Uganda and predominance of UCBSV infection in Lake Victoria region of Tanzania. This finding seems to suggest that UCBSV may be driving factor for the outbreak in Uganda. However from 1964, when Jameson reported complete elimination of CBSD unwittingly introduced from the coastal areas, to 2004 when the disease re-emerged (Alicai *et al*., 2007), Uganda was virtually free of CBSD. Since there was no information on virus testing within this period, it is difficult to link the Ugandan outbreak with UCBSV. Moreover, it was recently reported that UCBSV is actually milder than CBSV (Winter *et al*., 2010; Mohammed *et al*., 2012), hence is less likely to have kicked off the Ugandan epidemic, now turned pandemic.

Legg *et al*. (2011) noted a more than 100 fold increase in whitefly population across most parts of east and central Africa during the periods of outbreaks and therefore suggested that the currently abundant whitefly populations are responsible for CBSD re-emergence in Uganda hence the ensuing pandemic. It has also been suggested that introduction of CBSD-susceptible but CMD-resistant material may explain the pandemic (Mohammed *et al*., 2012). Nonetheless, the CBSD pandemic seems to continue with new reports of the disease in central Africa (Mulimbi *et al*., 2012). More recently, high incidence of foliar symptoms of CBSD of up to 90%, which surpassed CMD incidence, was reported in four east African countries (Hillocks et al., 2015). Surprisingly, root symptom incidence and severity from the same study was very low – maximum disease incidence at severity score > 3 was 10.4%. This was attributed to farmer preference for cultivars less affected by CBSD.

2.2.5 Control measures

Since CBSD is commonly transmitted through infected cuttings (Storey, 1936), distribution of planting materials tested virus-free has been suggested as a potentially effective phytosanitary

approach to controlling the disease (Hillocks and Jennings, 2003; Legg *et al*., 2011). Legg *et al*. (2011) recommends further phytosanitary measures including isolation of propagation sites in an environment away from other CBSD-infected plants, systematic virus testing of planting materials in multiplication centres and rigorous virus indexing for quarantine materials. These measures can be effective when correctly implemented but they often require recurrent heavy expenditure for example on tissue culture and virus testing facilities and skilled labour. Rouging has been employed to control CBSD in Uganda in the 1960s (Jameson, 1964) but as Hillocks and Jennings (2003) observed, disease incidence must be low in order for this method to be useful.

A more sustainable and potentially more effective approach to plant virus disease control is the use of resistant germplasm. There are no known genetic sources of immunity to CBSD among cultivated cassava (Hillocks and Jennings, 2003; Maruthi *et al*., 20014b). Most varieties are susceptible with a few being tolerant or resistant. Arguably, the best source of resistance to CBSD among cultivated cassava in east Africa is a hybrid called Kaleso which is grown commonly in Kenya. It is also known by the name Namikonga in Tanzania. It is a third backcross to cassava, of a *Manihot glaziovii* X cassava hybrid obtained from resistance breeding at Amani station, Tanzania (Nichols, 1947; Jennings, 1957). There have been efforts to re-discover resistant hybrids from the Amani breeding programme and to identify other resistant sources in landraces (Hillocks and Jennings, 2003), but there has yet to be any serious effort to incorporate known resistant varieties into breeding programs to mitigate CBSD.

2.3 Plant interaction with positive strand RNA viruses

2.3.1 Intracellular translation factors are necessary for virus infection

Upon transmission to a suitable host, survival of a plant virus depends on its ability to multiply and spread within its host's cells. Having a very limited genome size not sufficient to encode all proteins necessary for the challenge of replication and translation of its genome and movement within the host, specific genome architecture and proteins encoded by plant viruses allow them to exploit the vast array of metabolic, secretory and signal transduction capabilities of its host cell.

Diverse genome architecture of plant positive single-stranded RNA viruses ((+)ssRNA) ensures their efficient translation upon entry into a host cell. 5′ termini of these viruses could be a modified

nucleotide cap (as in mRNA), phosphate, or virus-encoded viral protein genome-linked (Vpg) protein. The 3′ end of the RNA can feature a poly(A) tail, a tRNA-like structure, or simply a 3′ OH group (Thivierge *et al*., 2005). Consistent with this variety in genome structure, plant (+)ssRNA viruses show diversity in genome translation. An interaction between the 5′ untranslated region (UTR) and 3′ translation enhancer domain (TED) of *Satellite tobacco necrosis virus* (STNV; *Tombusviridae*) enhances its cap-independent translation but is not required for it (Gazo *et al*., 2004). However, an interaction between the later and eIF4E is necessary for translation initiation. The infection-determining interaction between potyvirus VPg and isoforms of the 5′ cap-binding eukaryotic translation initiation factor eIF4E (Schaad *et al*., 2000; Leonardo *et al*., 2000) has led to the proposition of a model for translation of potyviruses. This model involves a circularization of potyvirus RNA where VPg binds the eukaryotic translation initiation factor eIF4E protein which subsequently binds a poly(A)-binding protein (PABP) attached to the poly(A) tail through another eukaryotic initiation factor, eIF4G (Thivierge *et al*., 2005). Loss of function studies of the translation initiation factor genes – *eIF4E* and *eIF4G* indicated a requirement of these factors for successful virus infection. Specifically, mutational inactivation of the gene encoding eIF4E protein and its isoform, *eIF(iso)4E* result in the loss of susceptibility to several potyviruses (Duprat *et al*., 2002; Piron 2010). A possible role for loss of eIF4E function in natural resistance to plant RNA viruses has also been demonstrated. A combination of yeast two-hybrid (Y2H) and glutathione S-transferase pull-down assays showed that VPg of two separate strains of *Tobacco etch virus* (TEV) could interact with product of *eIF4E* allele from susceptible but not the resistant *Capsicum* (Kang *et al*., 2005). In the bymovirus, *Barley mild mosaic virus* (BaMMV), expression of *eIF4E* clone from a BaMMV-susceptible plant, in a BaMMV resistant plant resulted in susceptibility (Stein *et al*., 2005).

Potyviruses differ in their preference of translation initiation factor(s). Yeast-two hybrid (Y2H) interaction assay between 3 translation initiation factors in tomato and potyvirus VPgs identified 2 classes of VPg – those requiring only eIF4E1 and those requiring eIF4E1 or eIF4E2 (Mazier *et al*., 2011). Plant (+)ssRNA virus species may enhance their survival by adapting forms able to interact with different isomers of the translation initiation factors so when an isomer is mutated, the virus can survive by switching over to use of another isomer as shown by induced mutation in *eIF4E1* which blocks susceptibility only in two strains of *Potato virus Y* (PVY) and *Pea enation mosaic virus* (PeMV). However, when the function of the two factors, *eIF4E1* and *eIF4E2* were

deleted, broad spectrum resistance against both virus species was obtained (Mazier *et al*., 2011). The normal cap-binding role of translation initiation factors do not interfere with their requirement for translation of virus genome *in planta*. Recessive resistance in capsicum caused by in point mutations in the *eIF4E* alleles, *pvr11* and *pvr12* did not affect the cap-binding cellular role the alleles (Kang *et al*., 2005b). This suggests that different domains are involved in eIF4E interaction with VPg and 5′-cap associated with virus replication and normal cellular activity, respectively.

RNA Helicases, which is a family of proteins representing a different eukaryotic initiation factor, the *eIF4A* proteins, are required for infection by some plant RNA viruses. RNA helicase-deficient mutant in *Arabidopsis* plants have been used to show that host RNA helicases are required for *Plum pox virus* (PPV) infection. A role for this host protein in virus replication is suggested by their localization in chloroplast-associated 6K2 vesicles when transiently expressed in TuMV-infected cells (Huang *et al*., 2010).

2.3.2 Silencing RNA virus genome as a defence strategy

Plants employ gene expression control mechanisms to regulate vital physiological processes such as growth and development, flowering etc. Gene expression can be regulated by employing small RNA (sRNA) molecules specific for a target gene locus or mRNA. The most important sRNA molecules employed for gene expression control in plants and other eukaryotes are micro-RNAs (miRNA) and small interfering RNA (siRNA). Both molecules are derived from mRNA, form double-stranded intermediates which are diced to specific short lengths usually between 21 to 24 nucleotides. The ribonuclease called Argonaut (Figure 3) selects a guide strand from a diced double-stranded fragment and loads this onto the RNA-induced silencing complex (RISC). The guide strand then directs sequence-specific recognition of complementary mRNAs which subsequently undergo Argonaut-catalyzed lysis of mRNA (Shimura *et al.*, 2011; Incarbone and Dunoyer, 2013).

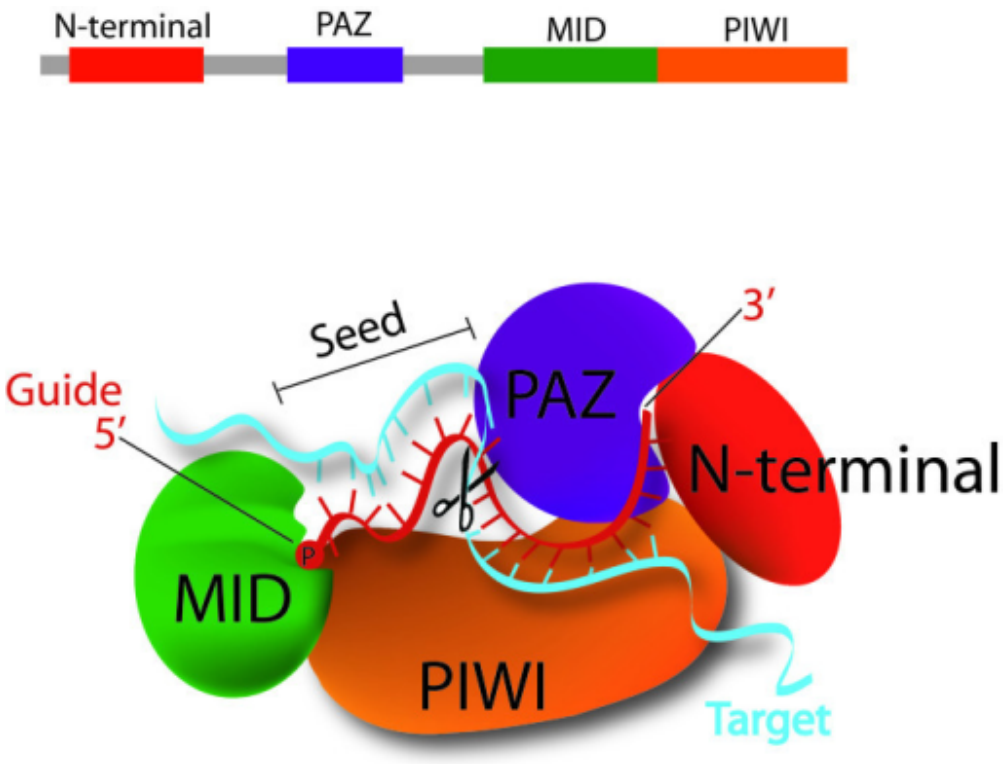


Figure 3. Illustration showing domains of argonaute protein and base-paring of guide siRNA with target mRNA. (Source: https://openi.nlm.nih.gov/detailedresult.php?img=PMC3375670_yjbm_85_2_187_g02&req=4)

Plants also deploy their RNA silencing machinery for protection against virus infection (English *et al*., 1996; Ratcliff *et al*., 1999). Silencing of viral gene expression is usually directed by virus-derived siRNA molecules. Host RNA-directed RNA polymerases (RDR) are involved in the formation of double-stranded RNA from single stranded RNA virus genome templates. RDR1, RDR2 and RDR6 of *A. thaliana* have been shown to play complementary roles in generating *Tobacco rattle virus* (TRV) siRNA and in limiting infection in TuMV-inoculated leaves (Garcia-Ruiz *et al*., 2010). As in endogenous gene silencing, Dicer-like (DCL) proteins process viral dsRNA or dsRNA-like structures into Argonaut-ready virus-derived siRNA (vsiRNA). Of the four DCL proteins of *A. thaliana*, DCL4 has been shown to be more prominently involved in vsiRNA production during RNA virus infection (Bouche *et al*., 2006; Deleris *et al*., 2006; Diaz-Pendon *et al*., 2004). Of importance in plant antiviral RNA silencing is the systemic spread of the antiviral silencing response generated in infected tissue. Sequence-specific systemic silencing able to

propagate cell-to-cell over long distance has been reported in post-transcriptional gene silencing PTGS of exogenous transgenes (Fagard and Vaucheret, 2000). Recently, RNA silencing was transferred from a transgenic tobamovirus-resistant tobacco, created by silencing the endogenous tobacco gene NtTOM1 and NtTOM3, to a non-transgenic tobacco by grafting (Ali *et al.*, 2013). This suggests that a mobile silencing signal was involved. This signal is not just able to move cell to cell but can travel long distances along phloem and other vascular tissues to achieve systemic RNA silencing-induced immunity in plants. The silencing signal in this case was likely a siRNA given that the same molecule was detected in both non-transgenic scion and the transgenic root stock unto which it was grafted. In *N. benthamiana*, activity of RDR6 excludes *Potato virus X* (PVX) and other plant viruses from the meristem hence ensuring efficient antiviral silencing (Schwach *et al*., 2005; Qu *et al*., 2005; Vaistij *et al*., 2009). Specifically, RDR6 does not produce or translocate RNA silencing systemic signal but responds to it, leading to the proposition of a model of RDR6-mediated defence mechanism involving generation of dsRNA from mobile signals which will ensure ready provision of siRNA for resistance response (Schwach *et al*., 2005).

RNA silencing-based resistance has been produced in cassava by transgenic means. Feasibility of achieving resistance to the two major virus diseases of cassava has been demonstrated using cassava lines genetically engineered to express RNA interference (RNAi) constructs of coat protein (*CP*) of UCBSV (Ogwok *et al*., 2012) and replication-associated (*AC1*) genes of ACMV (Vanderschuren *et al*., 2009). Recently, RNAi-based resistance to UCBSV has been shown to be unaffected by cycles of vegetative cropping which is the most common method of cassava propagation (Odipio *et al*., 2014).

Though RNAi-based resistance to plant viruses has been achieved by transgenic means there are evidences of natural resistance involving antiviral silencing. First is the detection of 25-nucleotide RNA complementary to PVX in virus-inoculated but not mock-inoculated *N. benthamiana* four days following PVX inoculation (Hamilton and Baulcombe, 1999). Second is that virus-encoded proteins are capable of suppressing RNA silencing as reported for HC-Pro and 2b protein encoded by potyviruses and cucumoviruses, respectively (Anandalakshmi *et al*., 1998; Brigneti *et al*., 1998) but has since been observed in many other viruses and involving many more viral proteins (Havelda *et al*., 2011; Shimura *et al*., 2011). Finally, plants deficient in genes involved in non-

transgenic RNA silencing response such as RNA-dependent RNA polymerases resulted in enhanced susceptibility (Yu *et al*., 2003).

2.3.3 Plant RNA viruses adopt a counter-defence silencing suppression strategy

Majority of helical-shaped RNA viruses encode a unique set of three overlapping open reading frames (ORFs) called the triple gene block (TGB), for movement (Schoelz *et al*., 2011). TGB1, the first of the three ORF products, is an RNA helicase (Solovyev *et al*., 2013). In PVX, TGB1 is essential for virus movement as demonstrated by movement deficiency of several PVX TGB1 mutants (Bayne *et al*., 2005). However, virus movement was found to require another function of TGB1 which is suppression of host's RNA silencing defence mechanism. Movement function of TGB1 was restored to suppressor-deficient TGB1 mutants of PVX by heterologous expression of HC-Pro and P19 (Bayne *et al*., 2005) demonstrating that suppression of silencing ensures successful virus movement. Silencing suppressors could also be involved in down-regulation of actual resistance genes. P1/HC-Pro of TuMV is known to up-regulate an miRNA of *Brassica* which down-regulates mRNA levels of a TIR-NBS-LRR *R* gene (He *et al*., 2008).

2.4 Plant virus resistance

2.4.1 General plant pathogen resistance models

Plants present multiple layers of defence response to pathogen invasion. The first of these involves re-enforcement of cell wall by increased callose deposition upon activation of host defence pathways (Spoel and Dong, 2012). Pathogens which make it through this first line of defence are faced with the hurdle of evading pattern recognition receptors (PRR) which are membrane-bound proteins containing Leucine rich (LRR) domains (Ausubel, 2005; Ronald and Beutler, 2010) and which recognize certain general pathogen molecular patterns generally referred to as pathogen-associated molecular patterns (PAMPs). However, some pathogens are able to evade PRR recognition by sending effector molecules into the cell. *Psuedomans syringae* pv. tomato sends in the protein AvrPto1 which is known to suppress immune-related proteins (Zipfel and Rathjen, 2008).

Plants have evolved strategies to defeat pathogen effector-mediated evasion of host immune proteins. Plant resistance to virus pathogens has been described by various models. Amongst these

is the classical gene-for-gene model (Flor, 1971) or more appropriately the guard hypothesis (Dangl and Jones, 2001) which describes the mechanism of resistance of mostly monogenic and dominantly-inherited host resistance (*R*) genes. The guard hypothesis proposes that R proteins do not interact directly with avirulent proteins but respond to effector-induced perturbations of effector-targeted host proteins, called 'guardee' proteins, by interacting with the 'guardee' proteins to protect them. This interaction allows the recognition of virus invasion which results in a triggering of host defence pathway to curtail virus spread in an incompatible interaction. A second model, the matching allele (MA) model which proposes a direct interaction between host and pathogen factors as basis of infection or resistance (Agrawal and Lively, 2002) is more suited to recessive resistance. Recessive resistance arises from heritable mutations in genes which code for proteins that interact, in their native form, with viral proteins to facilitate virus replication and spread.

2.4.2 Plant virus resistance genes

A large majority of known plant resistance genes which have been deployed for breeding natural resistance are *R* genes. *R* genes encode effector-specific, effector-triggered proteins collectively referred to as resistance (R) proteins which mediate a specific anti-pathogen response unlike the general response of pattern-triggered immunity (PTI) to a limited number of PAMPs (Spoel *et al*., 2012). R proteins typically consist of three domains: a variable domain at the amino terminus, a nucleotide binding site (NBS) domain in the middle and LRR at the COOH terminus (Spoel *et al*., 2012). On the basis of the variable domain, two major groups of R proteins are found in plants:

- CC-NBS-LRR with a coiled-coil amino terminus, and
- TIR-NBS-LRR having a toll interleukin-1 receptor.

R genes recognize pathogen-specific avirulence factors (avr) resulting in a specific interaction, which triggers the initiation of defence signalling and host resistance. Most *R* genes belong to the CC-NBS-LRR class (Table 1). These include functionally identical potato *Rx1* and *Rx2* which confer extreme resistance to PVX without inducing hypersensitive response (HR) (Bendahmane *et al*., 1999, 2000), tomato *Sw-5* which causes resistance to *Tomato spotted wilt virus* (TSWV) (Brommonschenkel *et al*., 2000). Others are the allelic *Tm-2* and *Tm-22* which confer resistance to *Tomato mosaic virus* (ToMV) (Lanfermeijer *et al*., 2003), *Arabidopsis HRT* and *RCY1* providing resistance to *Turnip crinkle virus* (TCV) (Cooley *et al*., 2000) and the yellow strain of

Cucumber mosaic virus (CMV) (Takahashi *et al*., 2002), respectively. *R* genes mostly mediate dominant resistance which may or may not be associated with hypersensitive response (Maule *et al*., 2007).

There are over half a dozen characterised TIR-NBS-LRR R proteins involved in plant virus resistance. These include the HR-eliciting tobacco *N* gene (Padgett *et al*., 1997) which has been a classic model system for study of plant-virus interaction and systemic acquired resistance (SAR) and *Y*-1 gene of potato which co-segregates with gene for extreme resistance to PVY (Vidal *et al*., 2002). Others include the common bean genes *RT-4* which controls HR-associated resistance to strains of CMV infecting tomato and pepper but not common bean (Seo *et al*., 2006), *PvVTT1* associated with resistance to *Bean dwarf mosaic virus* (BDMV) (Seo *et al*., 2007) and the *I* gene which controls resistance to *Bean common mosaic virus* (BCMV) (Vallejos *et al*., 2005). There are three non-NBS-LRR genes classified as dominantly inherited. The *Arabidopsis* genes, *RTM*1 and *RTM*2 which control resistance to TEV are lectin-like and heat shock proteins, respectively (Chisholm *et al*., 2000; Whitham *et al*., 2000) while *Tm*-1, the TMV and ToMV resistant gene of the wild tomato species *Solanum habrochaites* encodes a protein with a triose isomerase (TIM) barrel structure (Ishibashi *et al*., 2007). Table 1 contains list of all cloned dominantly inherited plant virus resistance genes.

In contrast to dominant resistance genes, there are not many known recessively-inherited resistant genes. Currently, about all characterized recessive resistance genes code translation initiation factors eIF4E, eIF4G or their isomers (Fraile and Garcia-Arenal, 2009). Except for the *Melon necrotic spot virus* (MNSV) system where translation initiation factors interact with the 3'UTR of MNSV, all other translation initiation factors interact with VPg in a cap-independent manner (Diaz *et al*., 2004; Maule *et al*., 2007; Palukaitis and Carr, 2008).

Table 1. List of genetically characterized dominant and recessive plant virus resistance genes

	Protein	Plant species	Virus targets	AVR factor	References
Dominant genes					
Cloned					
N	TIR-NB-LRR	*Nicotiana tobacum*	Tobamoviruses	Replicase/helicase	Padgett *et al*. (1997) and Whitham *et al*. (1994)
Rx1	CC-NB-LRR	*Solanum andigena*	PVX	Coat protein	Bendahmane *et al*. (1997, 1999)
Rx2	CC-NB-LRR	*Solanum acaule*	PVX	Coat protein	Bendahmane *et al*. (1997, 2000)
Sw-5	CC-NB-LRR	*Solanum lycopersicum*	TSWV, TCSV, GRSV	Movement protein	Bromonschenkel *et al*. (2000)
HRT	CC-NB-LRR	*Arabidopsis thaliana*	TCV	Coat protein	Cooley *et al*. (2000)
RCY1	CC-NB-LRR	*Arabidopsis thaliana*	CMV	Coat protein	Takahashi *et al*. (2002)
Y-1	TIR-NB-LRR	*Solanum tuberosum*	PVY		Vidal *et al*. (2002)
Tm-2/Tm-2²	CC-NB-LRR	*Solanum peruvianum*	ToMV, TMV	Movement protein	Lanfermeijer *et al*. (2003), Weber and Pfitzner (1998), and Weber *et al*. (2004)
Rsv1	CC-NB-LRR	*Glycine max*	SMV	P3 protein	Hajimorad *et al*. (2005) and Hayes *et al*. (2004)
RT4-4	TIR-NB-LRR	*Phaseolus vulgaris*	CMV	Replicase/helicase	Seo *et al*. (2006)
PvVTT1	TIR-NB-LRR	*Phaseolus vulgaris*	BDMV	Nuclear shuttle protein	Garrido-Ramirez *et al*. (2000) and Seo *et al*. (2007)

RTM1	Lectin-like	*Arabidopsis thaliana*	TEV	Coat protein	Chisholm *et al*. (2000) and Decroocq *et al*. (2009)
RTM2	Small heat-shock protein	*Arabidopsis thaliana*	TEV	Coat protein	Decroocq *et al*. (2009) and Whitham *et al*. (2000)
Tm-1	TIM barrel structure	*Solanum habrochaites*	TMV, ToMV	Replicase	Ishibashi *et al*. (2007) and Meshi *et al*. (1988)
Mapped to complex loci					
Tsw		*Capsicum*	TSWV	NSs protein	Margaria *et al*. (2007)
L^1, L^2, L^3, L^4	CC-NB-LRR	*Capsicum*	Tobamovirus	Coat protein	Tomita *et al*. (2008)
I	TIR-NB-LRR	*Phaseolus vulgaris*	BCMV		Vallejos *et al*. (2006)
Recessive genes					
Cloned					
$Pvr2^i$ + *pvr6*	*Pvr2*:eIF4E	*Capsicum annuum*	PVMV, TEV		Caranta *et al*. (1996) and Ruffel *et al*. (2006)
Pvr1/$pvr2^i$	eIF4E	*Capsicum chinense*	PVMV. PVY, TEV	Vpg	Charron *et al*. (2008), Kang *et al*. (2005b), and Ruffel *et al*. (2002, 2006)
nsv	eIF4E	*Cucumis melo*	MNSV	3'-UTR	Dı´az *et al*. (2004) and Nieto *et al*. (2006)
rym4/5	eIF4E	*Hordeum vulgare*	BaMMV, BaYMV	VPg	Kanyuka *et al*. (2004) and Stein *et al*. (2005)

mol^{1}/*mol*^{2}	eIF4E	*Lactuca sativa*	LMV	VPg and CI	Nicaise *et al*. (2003) and Roudet-Tavert *et al*. (2007)
rymv-1	eIF(iso)4G	*Oryza sativa*	RYMV	VPg	Albar *et al*. (2003, 2006)
sbmv1^{i}	eIF4E	*Pisum sativum*	PSbMV, BYMV	VPg	Bruun-Rasmussen *et al*. (2007), Gao *et al*. (2004), and Johansen *et al*. (2001)
pot-1	eIF4E	*Solanum habrochaites*	PVY,TEV	VPg	Ruffel *et al*. (2005) and Schaad *et al*. (2000)

Source: Aurora Fraile and Fernando Garcı´a-Arenal, 2010.

2.4.2 Hypersensitive response

In most plants, exposures to pathogens elicit a primary resistance response called hypersensitive response (HR). A typical HR response is associated with sudden release of cell death effectors such as the reactive oxide intermediates (ROIs) which destroys the integrity and viability of cells within local area of plant tissue infected with pathogen. This is thought to be followed by flux of ions across the membrane – typically export of potassium and chloride ions and uptake of calcium ions leading to a net alkalinity of cells (Morel and Dangl, 1997). Also, pathogen recognition in HR results in expression of inducible genes collectively referred to as defence-related genes. These genes encode enzymes which synthesize anti-microbial agents like phytoalexins, and structural proteins deposited on cell wall of infected cells. This strategy controls invading pathogen and walls-in infection preventing its spread to uninfected parts of the plant (Spoel and Dong, 2012).

HR signalling pathways are indispensable for *R* protein-mediated HR response. TMV-induced HR up-regulates biosynthesis of spermidine which is a precursor of the hydrogen peroxide (Yamakawa *et al*., 1998) and it is known that hydrogen peroxide is required for HR-mediated resistance response to TMV (Talarczyk and Hennig, 2001). The other HR signalling molecules – salicylic acid (SA) and ethylene (ET) are involved in a synergistic action required for *RCY1*-mediated resistance to CMV in *Arabidopsis* (Takahashi *et al*., 2004). SA is uniquely important to a plant's defence system, playing roles in R protein-mediated hypersensitive resistance response and in systemic acquired resistance (SAR) response following infection by diverse plant pathogens (Kachroo *et al*., 2000).

2.4.3 Systemic acquired resistance (SAR) signalling

Following a hypersensitive response to infection, a broad-spectrum longer-lasting resistance which affects the entire plant is induced. This response is called systemic acquired resistance (SAR) and it immunizes the plant against future pathogen exposure (Conrath, 2006).

There is a preponderance of evidence for SA or its derivatives as the mobile signal for SAR. The phloem accumulation of SA at onset of SAR (Yalpini *et al*., 1991) and abolishment of SAR from constitutive expression of salicylate hydroxylase (*nah*G; Gaffney *et al*., 1993) are some of them. Also, the requirement of methyl salicylic acid (MeSA)-forming methyltransferase and MeSA to SA-converting MeSA esterase activities at the immune signal generation and perception sites respectively, and the phloem accumulation of MeSA in tobacco (Park *et al*., 2007) are other

evidences. Interestingly, MeSA which had been reported as indispensable to SAR in tobacco but was found to be not required for same purpose in *Arabidopsis* (Attaran *et al*., 2009). This led to the suggestion that the composition of the mobile signal in SAR might differ depending on plant host and type of plant-pathogen interaction (Spoel *et al*., 2012). SA is the predominant mobile signal for currently known SAR pathways of *R* gene-mediated plant virus resistance. Classical SA signalling involves a non-expresser of pathogenesis-related (NPR1) pathway which leads to the synthesis of PR defence proteins. Typically, SA binds to and induces activation of NPR1 by dissociation of its autorepressor at the N-terminal (Wu *et al*., 2012). NPR1 is then recruited to the nucleus (Durrant and Dong 2004) where it binds to and lifts the transcriptional repression by TGA2 transcription factors and activates *PR-1* expression (Rochon *et al*., 2006). Liu *et al.* (2002) established a link between *N*-mediated resistance and SA-mediated SAR pathway by showing that *NPR1* function is required for *N*-mediated resistance to TMV in tobacco.

On the contrary, evidence of unbroken TCV resistance in *HRT*-containing *npr1*-mutant *Arabidopsis* proved the existence of *NPR1*-independent SA signalling pathway (Kachroo *et al*., 2000) in line with non-antiviral nature of PR proteins (Cutt *et al*., 1989; Lindhorst *et al*., 1989). Other studies point to a possible NPR1 and R protein-independent SA signalling of SAR in plant-virus interactions. Chivasa *et al*., (1997) showed that SA induced *N*-gene independent resistance to TMV in susceptible tobacco as reflected by reduction in viral RNA and coat protein accumulation. This SA-induced TMV resistance was antagonized by the mitochondrial alternative oxidase (AOX)-inhibiting molecule, SHAM (Chivasa *et al*., 1997). However, SHAM failed to counter resistance to a bacterial or a fungal protein suggesting a distinct SA-induced SAR signalling for viral infection. AOX catalyses reduction of oxygen to water and heat in a branch of cellular respiratory pathway often induced in plants under stress (Zhang *et al*., 2012). This activity results in the control of the high cellular oxidative state arising from stress-induced perturbation of normal ATP-generating cellular respiration in the nitochondria (Vanlerberghe *et al*., 2013). Further support for a distinct virus-induced SA signalling pathway was obtained from cyanide and antimycin A (both AOX activators)-induced TMV resistance devoid of NPR1 induction (Chivasa and Carr, 1998). Though cyanide and antimycin A inhibition of cellular respiration induce AOX (Chivasa and Carr, 1998), there is currently no evidence of it central role resistance to TMV. Instead, it modulates SA-mediated resistance to the virus in *Nicotiaana tobacum* (Gilliland *et al*., 2003). Two separate branches of SA-induced SAR signalling have been proposed. One leading to

NPR1 induction and PR synthesis is reasonably and evidently associated with resistance to bacteria and fungi (Carr and Klessig, 1989; Bowles, 1990; Alexander *et al*., 1993). The other leading to resistance without PR formation has been proposed as the 'virus branch' of SA signalling (Murphy *et al*., 1999). The *NPR1*-independent SA signalling model is, however, in sharp contrast to the *NPR1*-dependent *N* gene-mediated signalling reported for SAR response to TMV in tobacco (Liu *et al*., 2002). Notably, both opposing mechanisms were described for resistance of the same virus, TMV in the same plant species, tobacco (Chivasa *et al*., 1997; Chivasa and Carr, 1998; Liu *et al*., 2002). Though HR-associated, *NPR1*-dependent SAR is known to be associated to SA in various plant pathogens (Carr and Klessig, 1989; Bowles, 1990; Alexander *et al*., 1993), Liu *et al*., 2002 did not mention involvement of SA in the *NPR1*-dependent *N*-mediated resistance they observed. A plausible explanation for this seemingly unusual scenario could then be that another signalling molecule apart from SA is involved in *NPR1* activation especially as ET is also required for TMV-induced expression of the *N* gene in *N. benthamiana* and also function in synergy with SA to facilitate SAR in RCY1-mediated resistance to CMV in *Arabidopsis* (Liu *et al.*, 2004; Takahashi *et al*., 2004).

2.5 RNA sequencing

2.5.1 Techniques

RNA sequencing (RNA-seq) is an application of next-generation sequencing technologies used for studying global changes in levels of different RNA species. The method is well suited for differential expression analysis as it generates digital gene expression values which are counts of sequenced reads uniquely mapped to a reference genome or aligned *de novo* to provide a genome-wide transcriptome architecture (Wang *et al*., 2009). It is also used to study the other major themes of transcriptomics research namely cataloguing of non-mRNA species including small RNAs (sRNA) and microRNAs, determining transcriptional structure of genes, detecting gene splice patterns as well as post-transcriptional modifications (Wang *et al*., 2009). RNA-seq is currently the best technique to study an organism's transcriptome. Unlike microarray, it has a very wide range of detection, is not subject to background noise which decreases sensitivity and does not rely on prior knowledge of genome sequence (Wang *et al*., 2009). Other high-throughput sequence tag-based methods such as SAGE, CAGE and MPSS are based on expensive Sanger sequencing,

sequence very short tags which frequently map to multiple sequences, and are labour-intensive (Morozova *et al*., 2009). RNA-seq on the other hand is based on diverse next-generation sequencing technologies which allow a full analysis of transcripts to identify isoforms as well as epigenetic modifications (Wang *et al*., 2009).

To sequence the transcriptome under any of the existing next generation platforms, total or fractionated RNA population is made into cDNA fragment of 30 – 400 bp and amplified by PCR in a sample preparation stage (Wang *et al*., 2009). Transcripts are often sequenced as fragments hence cDNA for RNA-seq can be generated prior to fragmentation depending on the region of transcript which is of most interest. RNA fragmentation results in sequence slightly biased for middle sequences while cDNA fragmentation generates strong bias towards identification of sequences at transcript ends (Wang *et al*., 2009). The Roche 454 and Applied biosystems' SOLiD platforms amplify cDNA fragments by emulsion PCR while Illumina adopts a solid phase amplification approach called bridge amplification (Morozova *et al*., 2009). The amplification step is omitted under the Helicose Biosciences' platform allowing sequencing of a single cDNA fragment (Metzker, 2010). Sequencing-by synthesis (SBS) developed for Illumina sequencing platforms is currently the most widely used NGS technology. On the Illumina platform, clonally amplified sequence fragments are sequenced in parallel, on flow cells, by addition of fluorescently-labelled reversible terminator nucleotides to a primed template. Fluorescence from incorporated nucleotides is detected by total internal fluorescence using a four-colour detecting laser system (Metzker, 2010). Sequencing could involve a single end of a fragment (single-end sequencing) or both ends of a fragment (paired-end sequencing). The latter allows for less ambiguous mapping of short reads to a reference genome especially to repetitive sequence regions, is preferable for *de novo* assembly and can more easily identify rearrangements within the genome (Morozova *et al*., 2009). Currently, the range of Illumina NGS platforms can sequence up to 300 bp paired-end reads and can generate up to 5 billion reads per run.

2.5.2 RNA-seq data analysis

Following sequencing, intensity traces are generated in real time for each set of clonally amplified fragments. Quality-assigned bases of reads are called from these traces, the reads are quality-filtered and then mapped to a reference genome. The quality-filtering step removes adaptor sequences from 5'-end and poor quality bases often at 3'-end of reads, in a process called trimming,

as well as removes poor quality reads. A variety of tools are available for mapping reads. A summary information on the more popular read mapping tools in current use is presented in Table 2. Alignment outputs from these tools are usually in binary alignment map (BAM) format which can be converted to human-readable sequence alignment map format using SAMtools (Li *et al*., 2009). Unlike most alignment programs, MAQ and Bowtie improve accuracy of alignment by allowing FASTQ file inputs (Magi *et al*., 2010). MAQ and BWA generate alignment quality score which is the probability of wrongly aligning a read at a particular position on genome (Li *et al*., 2008; Li and Durbin, 2009). Of the read alignment tools described above, only TopHat can identify novel splice variants *ab initio* and hence is more preferable for mapping mRNA reads in RNA-seq experiments (Trapnell *et al*., 2009). TopHat aligns RNA-seq reads in two stages. The first involves alignment using the Bowtie algorithm. Then all reads that could not be mapped by Bowtie excluding low complexity reads are indexed as short 2k-mer nucleotide fragments. User-defined k-nucleotides from each end of all possible splice junction, determined by the presence of either of the canonical GT-AG pair at the ends of neighbouring 'island contigs', are concatenated and used as 2k-mer 'seeds' to query index of initially unmapped reads. Seeds extendable under user-defined mismatch conditions are considered as spanning exon-exon junctions and a set of non-redundant splice junctions is built from these (Trapnell *et al*., 2009). Splice isoform identification and other revelations of expressed gene architecture are only secondary applications of the RNA-seq technique. The primary output from an RNA-seq experiment are the digital counts of reads mapped to each gene of the transcriptome. Differential expression is usually determined from relative measures of transcript abundance calculated from these read counts and total number of uniquely mapped reads. The reads per kilobase of transcript per million reads (RPKM) measure, defined by Mortazavi *et al*., (2008), addresses bias from difference in total number of uniquely mapped reads and transcript length amongst samples.

Table 2. Basic information on some currently available read alignment tools

Alignment tool	Author(s)	Input reads from platform(s)	Indexing	Mismatch	Aligned per CPU day in Gbp (Magi *et al*., 2010)
Bowtie	Langmead *et al*., 2009	Roche-454, illumina, AB-SOliD	Burrows-Wheeler Transform (BWT) indexing of reference genome	Allows few mismatches on high quality end of seed	~7
BWA	Li and Durbin, 2009	Roche-454, illumina, AB-SOliD	BWT indexing of reference genome, generates mapping quality scores	Allows mismatches up to a threshold	~7
MAQ	Li *et al*., 2008	illumina, AB-SOliD	Hash table indexing of reads, generates mapping quality scores	Allows mismatches up to a threshold	~0.2
SOAP2		illumina, AB-SOLiD	BWT indexing of reference genome	Allows mismatches up to a threshold	~7
TopHat	Trapnell *et al*., (2009)	Roche-454, illumina, AB-SOLiD	BWT indexing of reference genome	Allows few mismatches while mapping non-junction reads	~7 for mapping non-junction reads

Source: Magi *et al*., 2010.

A similar measure of transcript abundance, fragment per kilobase per million reads (FPKM), was introduced by Trapnell *et al*. (2010) as a conceptually analogous method to RPKM but which is more suitable for measuring transcript abundance from mRNA fragments sequenced as paired-end reads.

Marioni *et al*. (2008) showed that read counts followed a Poisson distribution and hence tested for significant expression difference between technically replicated samples by likelihood-ratio test (LRT) of read counts per gene expressed relative to total mapped reads in a Poisson generalized linear model (GLM). Two limitations of the Poisson model of Marioni *et al*. (2008) namely unsuitability of the Poisson approach for modelling over-dispersed data (Auer and Doerge, 2010; Anders and Huber, 2010) and the lack of correction for possible RNA sampling bias (Robinson and Oshlack, 2010) has limited its usefulness especially for analysis of differential expression in biologically replicated samples. For analysis of biologically non-replicated samples, differential expression can be determined from generalized fold (GFOLD) values calculated from posterior distribution of logarithm of fold changes at a pre-determined significance level (Feng *et al*., 2012).

Transcriptome-level gene expression and differential gene expression data obtained from mapping and assembly of RNA-seq reads can be given biological meaning by searching list of differentially expressed genes and their expression or expression fold change values against a gene ontology database which is a library of well-structured annotation of gene functions and cellular location. Depending on the specific aim and scope of study, data from various stages of RNA-seq analysis can be channelled to different analysis pipelines such as single nucleotide polymorphism (SNP) calling, using for instance SAMtools (Li, 2011), and estimation of allele-specific expression by counting reads mapped to each allele of a SNP locus.

2.5.3. Recent case-studies and applications

RNA-seq has been recently deployed to classify and characterize expression pattern of entire transcriptome of the oilseed, *Brassica rapa* (Tong *et al*., 2013) and to identify genes differentially expressed in response to water stress in root tissue of upland cotton, *Gossypium hirsutum* (Bowman *et al*., 2013). In cassava, it has been applied to the identification of genes responding transcriptionally to infection with the pathogenic and non-pathogenic strains of *Xanthomonas axonopodis* - ORST4 and ORST4 ($TALE1_{Xam}$) respectively (Munoz-Bodnar *et al*., 2014). In this study, treatment with both strains modulated expression of the same genes in a similar fashion. Specifically, genes involved in phenylpropanoid biosynthesis and

photosynthesis were induced while jasmonic acid signalling genes were repressed. In contrast to treatment effect, there was greater difference in gene expression across time-points Munoz-Bodnar *et al*., 2014). Maruthi *et al*. (2014b) were the first to apply the RNA-seq technique to study the transcriptional response of cassava to CBSV inoculation. This involved transcriptional profiling of gene expression response to CBSV infection of the putative CBSD-resistant and susceptible cassava varieties – Kaleso and Albert, one year after inoculation. In this study, over 700 genes were found to be uniquely transcriptionally induced in the former compared to the latter variety. These include genes involved in biosynthesis of secondary metabolites such as phenylpropanoids involved in defence signalling, the stress response NAC gene and the eIF4E which is a recessive resistance gene (Maruthi *et al*., 2014b).

Gene expression responses to abiotic stress conditions in cassava have also been studied using RNA-seq. Transcriptome analysis of three different cold treatments of cassava - namely gradual chilling acclimatization (CA), chilling stress after chilling acclimatization (CCA) and chilling stress (CS) – revealed substantial commonality, for CA and CS treatments, in number of genes differentially expressed (Zeng *et al*., 2014). These commonly differentially expressed genes were found to be enriched with biological processes associated with protein translation, L-serine metabolism and diverse other metabolic processes. Harsh chilling after moderate chilling stress as in the CCA treatment resulted in a reversal of differential gene expression (Zeng *et al*., 2014)

Beyond assessment of gene expression levels, data from RNA-sequencing can have further applications. Downstream processes which utilize expression profile data from RNA-seq experiment include expression quantitative trait loci (eQTLs) analysis. Expression profiles of thousands of genes can be associated in a locus-wise manner to genotype data to identify genomic locations controlling gene expression hence trait expression (Druka *et al*., 2010). These genomic locations are referred to as eQTLs and they could be on or very close to gene it controls (*cis*-acting) or unlinked to it (*trans*-acting). Classical eQTL analysis involving the use of microarray-derived gene expression and sequence variation data, obtained separately, are still being used to dissect the genetic basis of gene expression and to identify candidate genes for various traits (Bolon *et al*., 2014). While this approach is effective and has revealed the genetic architecture of gene expression variation, the power of eQTL identification often relies on mapping large populations (Joonsen *et al*., 2009) which could be very costly in terms of time and resources. Exploiting the ability of RNA-seq to partition expression level of a gene into its constituent allele or haplotype expression quantities, has allowed for a reduction in

sample size required for eQTL analysis (Sun, 2012). Also, the possibility of determining nucleotide sequence variations alongside transcript abundance has led to a recent description of application of allele-specific expression (ASE) to eQTL mapping (Sun and Hu, 2013). However, mapping *cis*-eQTLs by ASE still require whole genome sequencing (WGS)-derived genotype data, haplotype reference sequence, and fairly large sample size (Sun and Hu, 2013). Fulfilment of these requirements could be cost and labour-intensive and take a lot of time. Zhai *et al*. (2013) were able to generate high-quality SNP data from RNA-seq reads and went on to carry out ASE analysis, in parental lines of rice, which they confirmed in their hybrids. Close to 68% of SNPs identified from this study were CT and GA SNPs suggesting a role of DNA methylation in regulation of gene expression (Zhai *et al*., 2013). ASE based on SNPs derived from RNA-seq could have huge cost benefits since it eliminates the need for WGS genotyping. Being found in expressed genes it is expected to facilitate the identification of functional genetic variation potentially underlying crop traits such as disease resistance.

CHAPTER 3: Development of real time PCR method for the simultaneous quantification of cassava brown streak and cassava mosaic viruses[1]

3.1 Introduction

In Africa, cassava is under significant threat from the two major viral diseases –cassava mosaic disease (CMD) and cassava brown streak disease (CBSD). CMD is caused by eleven begomovirus species (Legg *et al.,* 2015). Nine of these namely *African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV) – which includes the recombinant strain EACMV-Ug, *East African cassava mosaic Cameroon virus* (EACMCV), *East African cassava mosaic Kenya virus* (EACMKV), *East African cassava mosaic Malawi virus* (EACMMV), *East African cassava mosaic Zanzibar virus* (EACMZV) and *South African cassava mosaic virus* (SACMV) and recently identified *Africa cassava mosaic Burkina-Faso virus* (ACMBFV, Tiendrebeogo *et al.,* 2012) and *Cassava mosaic Madagascar virus* (CMMGV, Harimalala *et al.,* 2012) are endemic to Africa (Legg *et al.,* 2015). Analysis of DNA-A sequence diversity of cassava mosaic begomoviruses (CMGs) showed that overall, they cluster according to their respective species (Ndunguru *et al*., 2005; Figure 4). Within each species however, virus isolates tend to cluster according to their geographical source (Figure 4; Ndunguru *et al*., 2005). CMD symptom expression varies depending on host – natural or alternative host and within the natural host (casssva) symptom expression can differ across varieties of cassava (Figure 5a and b; Alabi *et al*., 2011). It also varies across isolates of a given CMB species (Figure 5c; Patil and Fauquet, 2009). Mixed infection by the two most common CMB species – ACMV and EACMV results in severe infection but a highly virulent recombinant strain – *East African cassava mosaic virus-Uganda variant* (EACM-Ug) has been associated with the outbreak of very severe form of CMD in Uganda (Zhou *et al.,* 1997). The CMD pandemic has spread across neighbouring countries in the Great Lakes region of Eastern Africa and in recent years has been identified in Angola (Kumar *et al.,* 2009) and Cameroon (Akinbade *et al.,* 2010) in Central Africa.

[1]This work is now published: Otti, G., Bouvaine, S, Kimata B., Mkamillo G, Kumar, P.L., Tomlins, K., Maruthi, M.N. (2016). High-throughput multiplex real-time PCR assay for the simultaneous quantification of DNA and RNA viruses infecting cassava plants. Journal of Applied Microbiology, 120(5):1346-56.

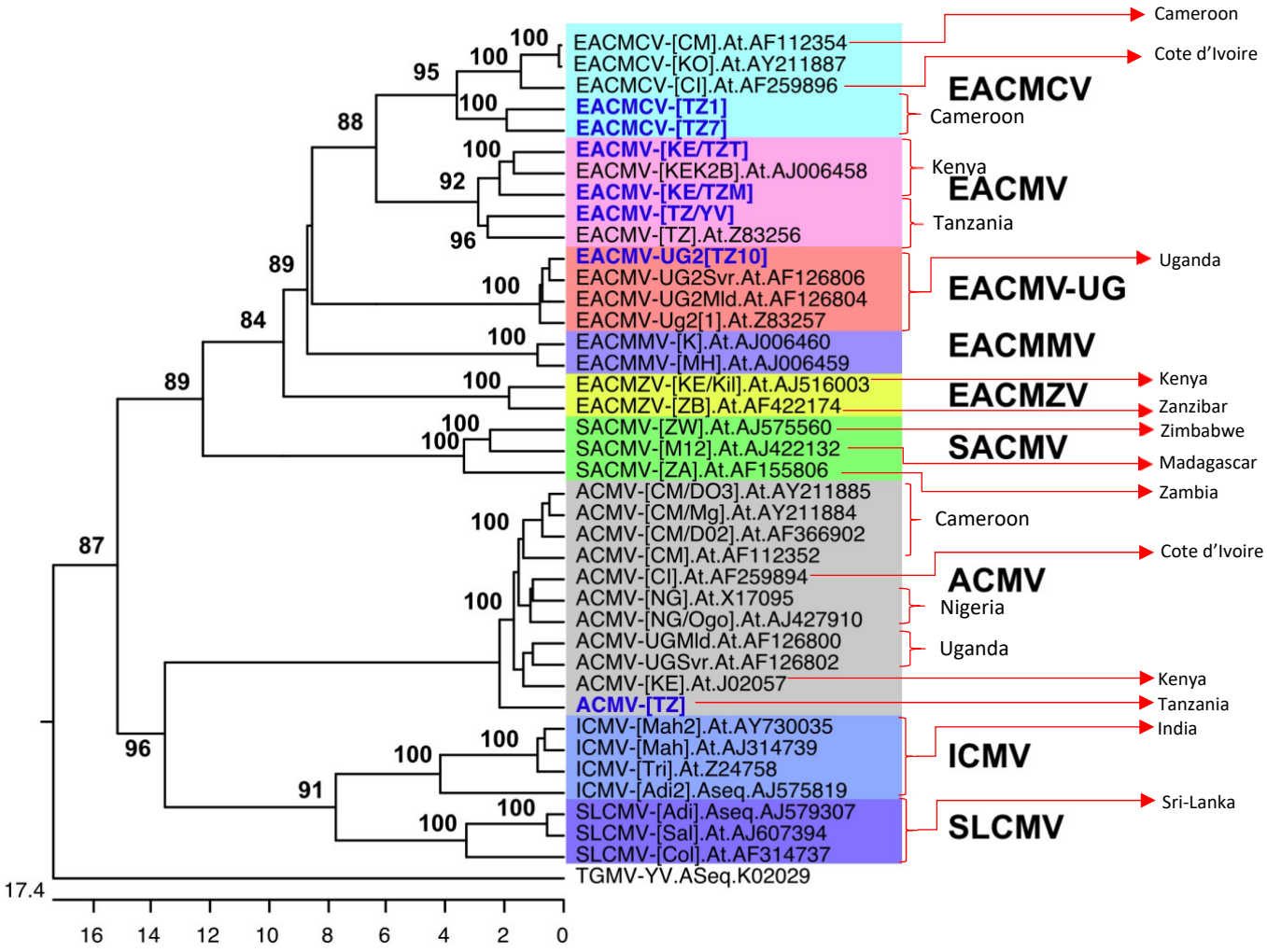


Figure 4. Phylogenetic tree of complete DNA-A sequences of cassava mosaic begomovirus species. Source: Ndunguru *et al*., 2005. Country of origin of the isolates have been inserted.


(A)
Photo credit: Rayapati A. Naidu


(B)
Photo credit: L. Kumar (IITA)

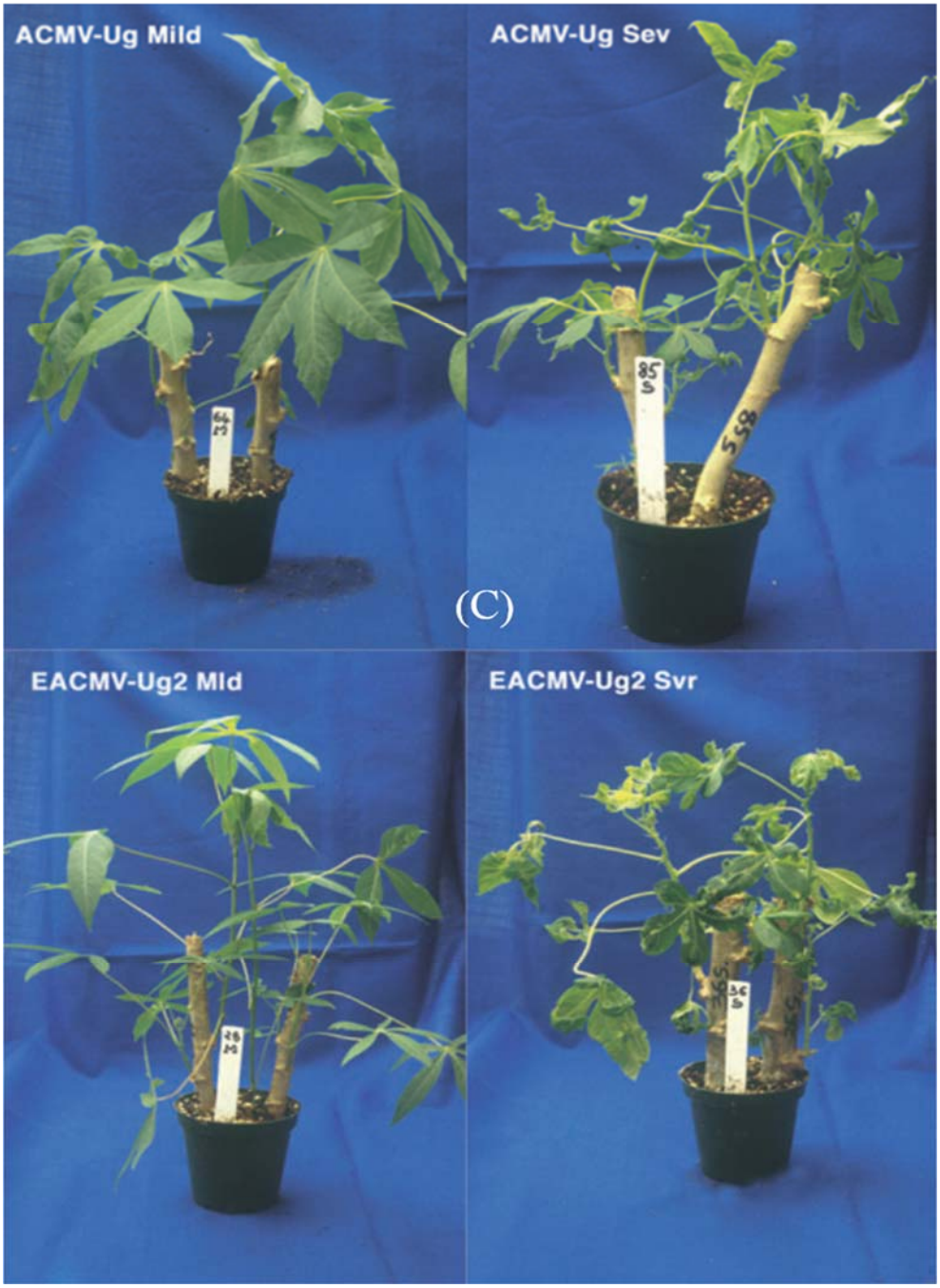


Figure 5. Diversity of CMD symptoms in different cassava varieties showing A. severe stunting and distortion of leaves and B. unusual 'candle stick' symptom unaccompanied by stunting. Figures 5A and 5B were soruced from Alabi *et al*., 2011 but credits for both figures belong to Rayapati A. Naidu and Lava Kumar respectively. Different expression of CMD severity by different strains of ACMV and EACMV are shown in 5C (Credits: Patil and Fauquet, 2009).

On the other hand, CBSD was thought to be largely confined to the lowland areas of east Africa for about seven decades since it was first identified and described (Hillocks and Jennings, 2003). In recent years, however, severe outbreaks in inland regions of Uganda (Alicai *et al.,* 2007) and areas around Lake Victoria including north western Tanzania and south-western Kenya were reported (Legg *et al.,* 2011). The disease now seems to be on a west-ward movement having been reported in Burundi (Bigirimana *et al.,* 2011) and Congo DR (Mulimbi *et al.,* 2012) where UCBSV is the sole causative agent. A detection and monitoring strategy involving accurate and affordable laboratory-based diagnosis and large-scale disease surveillance and a prevention strategy involving assurance of the health of planting material and cultivation of resistant cassava germplasm have been proposed by Legg *et al.,* (2015) to control the spread of CBSD and CMD. A range of PCR-based diagnostic and quantitative methods are currently being deployed for survey of field infections, virus indexing in tissue culture-derived planting material and screening for new sources of resistance (Legg *et al*., 2015).

More effective and holistic implementation of the monitoring and prevention virus management strategies described by Legg *et al.,* (2015) would require a sensitive, high-throughput and cost-effective method which can simultaneously quantify CBSVs and CMBs for application in disease surveys, virus indexing and resistance screening. The real time PCR technique (RT-qPCR) can fulfil these requirements, considering its high sensitivity combined with the characteristic specificity of TaqMan probe chemistry. In this study, we report on the development of a multiplex TaqMan-based real time PCR method for the cost-effective detection and quantification of the two CBSVs – CBSV and UCBSV; and the CMB species – ACMV and EACMV. Standard quality assurance tests such as sensitivity, specificity, intra- and inter-assay variability were performed for the new method. The impact of multiplexing on sensitivity was assessed and the method was subsequently applied to determining virus quantities in field infections of CMD and CBSD in local east African cassava germplasm.

3.2. Materials and Methods

3.2.1. Plant material

Two CBSD viruses, CBSV-[Mz:Nam:07] and UCBSV-[UG:Kab:07], originally collected from Mozambique and Uganda, respectively (Abarshi *et al.,* 2012) were maintained in the cassava

var. Albert. The two CMD viruses ACMV-[UG:Nam:97] and EACMV-[UG:Nam:97] were maintained in var. Ebwanateraka (Maruthi *et al.,* 2002).

Leaf samples from 80 field-grown cassava plants, representing sixteen cassava varieties, were collected in 2014 from Naliendele in Tanzania and used in the validation of the multiplex qPCR method. The cassava plants were six months old and were grown in high disease pressure areas for both diseases.

3.2.2. Sample preparation

Total nucleic acid was extracted from cassava leaf tissue using a modified CTAB extraction method described in Abarshi *et al.,* 2010. The procedure for this method is as follows:

1. CTAB buffer was prepared from CTAB (2% w/v), 1.4 M NaCl, 20mM EDTA and 100mM Tris-HCl (pH 8.0). The CTAB buffer was autoclaved.
2. CTAB buffer was preheated for 10 minutes at 65ºC, after addition of 0.2% (v/v) 2-mercaptoethanol.
3. Approximately 100mg of cassava leaf tissue samples were grinded using in a thick-gauged plastic bag using a hand-held ball bearing sample grinder (Bioreba AG, Reinach, Switzerland) and mixed using a wallpaper seam roller in 1 ml of CTAB buffer pre-heated at 60^0C
4. Approximately 800 µl of sample mix was transferred to a 2 ml Eppendorf tube
5. An equal volume (800 µl) of phenol: chloroform: isoamyl alcohol (25:24:1) was added and the mixture centrifuged at 13,000 rpm for 10 minutes.
6. The supernatant was transferred to a new 1.5 ml eppendorf tube and nucleic acids were precipitated by adding 0.6 volumes of ice cold (−20 °C) isopropanol
7. Samples were then incubated at −20 °C for at least 1 h and centrifuged finally at 13,000 rpm for 10 min at 4 °C
8. Pellet collected at base of the tube was washed in 0.5 ml 70% ethanol, centrifuged for 5 min and vacuum-dried for 5 min in a spin vac
9. The pellet was dissolved in molecular grade DNAse- and RNAse-free water and stored at −20 °C for subsequent analysis.

3.2.3. Primer and probe design

Primers specific to CBSV, UCBSV, ACMV and EACMV and probes specific to ACMV, CBSV and UCBSV were designed and used alongside existing EACMV and EACMV-Ug-specific probe (Table 3). To design primers and probes for ACMV an EACMV, 342 DNA-A sequences representing all nine CMBs found in Africa were aligned. Primer and probe sequences for ACMV and EACMV were selected from DNA-A regions conserved in and unique to the virus species of interest. CBSV and UCBSV primer and probe sequences were selected from conserved regions of CBSV- and UCBSV-specific coat protein (CP) virus gene determined after alignment of five and eight full-length CBSV and UCBSV sequences, respectively. All probes were labelled with specific fluorophores at their 5' end and a dark quencher at the 3' end (Table 3). CBSV- and EACMV-specific probes each had a secondary ZEN™ quencher (IDT, Coralville, IA) between their fluorophore and primary quencher, in order to reduce background fluorescence and improve signal to noise ratio. Melting temperature, %GC content and potential for self- and hetero-dimer formation were checked for selected primer and probe sequences using the online tools OligoAnalyzer (https://www.idtdna.com/calc/analyzer) and Thermoscientific's multiple primer analyser web tools (http://www.thermoscientificbio.com/webtools/multipleprimer/).

Table 3. Primers and probes used for the detection of ACMV, EACMV, CBSV and UCBSV

Virus target	Primer/probe	Sequence (5′ - 3′)	Quantity ($\mu molL^{-1}$)	Reference
ACMV	ACMVAV2F1	RCAGGCGAAGTTGTKGCTA	0.3	This study
	CMBAV2R	TAWGGGCTGTCGAAGTTCAG	0.3	This study
	Probe	ROX-TGTCGAAGCGWCCAGGMGATATCATCATTTCCAC-BHQ2	0.4	This study
EACMV/	CMBRep/F	CRTCAATGACGTTGTACCA	0.12	Alabi *et al.,* 2008
EACMV-Ug	Neweac-alt/R	CATGGAGACCGATCAGTATTGTTC	0.12	This study
	Probe	FAM-TCTTKGGAG/ZEN/ACAGATCCAGGTGTCCACAT-IABkFQ	0.06	This study
CBSV	CBSV F3	GGARCCRATGTAYAAATTTGC	1.0	Abarshi *et al.,* 2012
	CBSV R4	GCWGCTTTTATYACAAAMGC	1.0	This study

	Probe	JOE-TTCCAGCCA/ZEN/AGCAATWYTGATGTATCAGAATAGTGTGA-IABkFQ	0.2	This study
UCBSV	CBSV F3	GGARCCRATGTAYAAATTTGC	1.0	Abarshi *et al.,* 2012
	CBSV R4	GCWGCQTTTATYACAAAMGC	1.0	This study
	Probe	TAMRA-ACTATGAGGAAGGTTATGAGAAACTTCTCTAGCCAAGC-BHQ2	0.2	This study
PP2A	PP2AF	TGCAAGGCTCACACTTTCATC	0.5	Moreno *et al.,* 2011
	PP2AR	CTGAGCGTAAAGCAGGGAAG	0.5	Moreno *et al.,* 2011
	Probe	JOE-CQTTCQGTT/ZEN/GCCCCCACCATGC-IABkFQ	0.2	This study

3.2.4. End-point PCR protocol

End-point PCR analysis of ACMV, EACMV, CBSV and UCBSV in the 80 field-collected samples was done in parallel to quantitative assay of the viruses using the multiplex RT-qPCR method developed in the current study. End-point PCR assays of ACMV and EACMV was done using the common forward primer CMBRep/F and the reverse primers – ACMVRep/R and EACMVRep/R, respectively (Alabi *et al.,* 2008). On the other hand, CBSV and UCBSV were PCR-amplified using the common forward primer CBSV F2 and the reverse primers – CBSV R7 and CBSV R8 respectively (Abarshi *et al.,* 2012).

3.2.5. Multiplex real time PCR

A two-step multiplex RT-qPCR protocol exploiting the TaqMan chemistry was used for the quantitative assay of CBSV, UCBSV, ACMV and EACMV alongside its variant – EACMV-Ug Complementary DNA (cDNA) was synthesized from RNA extracts using Improm II^{TM} Reverse transcription kit (Promega, Southampton UK) following manufacturer's instructions. For the qPCR stage, 2 µl of cDNA template were mixed with 1.5ul of 10X dilution of DNA template. A multiplex master mix made from 2X Express qPCR supermix universal (Life Technologies, Paisely UK), EACMV, ACMV, CBSV and UCBSV-specific primers and probes, at quantities given in Table 1, were used for the simultaneous quantitative detection of four virus targets – CBSV, UCBSV, ACMV and EACMV/EACMV-Ug. Targets were amplified in real time using the Eppendorf's Mastercycler ep realplex (Eppendorf, Cambridge UK) in 40 PCR cycles of 15 seconds denaturation at 94 ºC, 20 seconds annealing at 54 ºC and extension at 60 ºC for 30 seconds. The PCR cycles were preceded by 2 minutes incubation at 50 ºC for removal of carry-over contamination by uracil DNA glycosylase (UDG) and a second 2 minute incubation at 95 ºC for activation of the *Platinium*® *Taq* DNA polymerase. In a single run, samples were analysed in duplicate wells for cassava viruses. Equal amounts of DNA and cDNA templates were mixed in a tube for the multiplex assay, The endogenous cassava gene – serine threonine phosphate (*PP2A*), which has been established as the best reference gene for CBSV in infected cassava leaf and root tissue, was amplified (Moreno *et al.*, 2011) in separate duplicate wells.

3.2.6. Specificity, sensitivity, intra- and inter-assay variability tests

The multiplex RT-qPCR assay was tested for sensitivity and specificity of detection of each intended target. Sensitivity experiment was carried out using five to six 10X serial dilution standards generated from PCR amplicon of each target and purified using reSource PCR purification kit (Source BioScience LifeSciences, Nottingham UK). Prior to purification, each target's amplicon was checked on 1% Agarose gel to ensure that only the correct product size was amplified. After purification, amplicons were quantified using the NanoDrop 200 spectrophotometers (Thermoscientific, Wilmington DE). Sensitivity of detection of each virus was determined from lowest quantity of the purified virus amplicon detectable at maximum Cq of 35. A standard curve was generated for each target and used to determine their reaction efficiency and correlation coefficient. Specificity of the multiplex RT-qPCR assay method was assessed by testing the primers and probes with target and non-target viruses.

Intra- and inter-assay variation of Cq values was determined in samples containing one virus template (hereafter referred to as single template sample) and in samples containing combinations of three or all four viruses (hereafter referred to as mixed template samples). For each assay, false positive amplification was controlled using water as negative control.Variation was expressed as coefficients of variation – that is percent proportion of mean Cq value.

3.2.7. Variability of Cq values across virus template samples

Using the multiplex RT-qPCR protocol described above, each of the four virus species – ACMV, EACMV, CBSV and UCBSV was assayed in aliquots of same virus template sample mixed with different combinations of other virus templates. For instance, ACMV was assayed, in a single run, in samples containing ACMV mixed with EACMV and CBSV, ACMV mixed with EACMV and UCBSV, ACMV mixed with CBSV and UCBSV templates, and in sample containing all four viruses. Each of these assays was performed in at least two technical replicates.

3.2.8. Comparison of multiplex qPCR method and uniplex alternatives

To assess the possible effect of multiplexing on sensitivity of quantitative assay for CBSV, UCBSV, ACMV and EACMV, their limits of detection in multiplex and uniplex assays was determined from 10X serially-diluted standards. Also, Cq values from uniplex and multiplex quantitative assay of single and combinations of virus targets were compared. The Wilcoxon signed rank test was applied to test for significant change in Cq arising from multiplexing. The

test was implemented using the "coin" package (Hothorn *et al.*, 2008) of the R-statistical analysis software version 3.1.0 (R Core Development Team, 2014).

3.2.9 Validation experiment

The multiplex qPCR method developed in this study was applied to the quantitative assay of CBSV, UCBSV, ACMV and EACMV in 5 samples each of sixteen cassava varieties grown in the field. Virus incidence was determined at Cq threshold of 35. In other words, samples with Cq values greater than 35 were considered negative. Relative quantities of each virus were calculated. Relative quantity was calculated using the formula, relative quantity = $2^{-\Delta\Delta C_q}$ (Livak and Schmittgen, 2001). CBSV and UCBSV quantities were calculated relative to normalized Cq of the both viruses maintained in Albert while quantities of ACMV and EACMV were calculated relative to normalized Cq values of an ACMV and EACMV maintained in the variety Ebwanateraka. Effect of mixed CBSV and UCBSV infection on quantities of each of the viruses was determined by ANOVA analysis using the randomized block design. Possible interaction between mixed infection effect and variety genotype in determining quantity of each virus was tested using two-way ANOVA. Both analyses were implemented on the R statistical analysis platform (R Core Development Team, 2014).

3.3. Results

3.3.1. Specificity, sensitivity, intra- and inter-assay Cq variation of multiplex qPCR method

For each virus, Cq < 35 was obtained for positive samples whereas Cq values in negative samples were undetermined or ≥ 35. The method reliably detected intended targets only when they were available in tested samples (Table 4). Also, detection was specific even in the presence of other virus templates (Table 4) and fluorescence signal was always above background in positive samples but only background fluorescence was obtained for negative samples. Lower limits of virus detection ranged from 4.0 femtograms (fg) of purified EACMV sequence fragments to 12.5 femtograms (fg) of ACMV amplicons, using the multiplex method (Table 5). Limits of detection for CBSV and UCBSV on the other hand were 9.88 fg and 9.53 fg, respectively. These values are same as those obtained from uniplex qPCR assay of viruses (Table 5). Reaction efficiency and correlation coefficient between virus quantity and Cq, obtained from multiplex qPCR assay of each virus were high (Figure 6) and comparable to values from uniplex assay of same virus template samples (Table 5).

Table 4. Cq values from multiplex assay of four cassava viruses highlights specific virus detection unaffected by mixture with different virus templates.

Virus target	**Virus template(s)	Mean Cq
CBSV	CBSV+ACMV-EACMV-UCBSV-	24.52
	CBSV+ACMV+EACMV+UCBSV-	26.83
	CBSV+UCBSV+ACMV+EACMV-	25.84
	CBSV+UCBSV+EACMV+ACMV-	26.31
	UCBSV+ACMV+EACMV+CBSV-	*ND
	CBSV+UCBSV+EACMV+ACMV+	28.64
	Healthy	ND
	NTC	ND
UCBSV	UCBSV+CBSV-ACMV-EACMV-	27.09
	UCBSV+ACMV+EACMV+CBSV-	28.22
	UCBSV+CBSV+ACMV+EACMV-	26.87
	UCBSV+CBSV+EACMV+ACMV-	25.63
	UCBSV+CBSV+EACMV+ACMV+	26.42
	CBSV+ACMV+EACMV+UCBSV-	33.40
	Healthy	ND
	NTC	ND
ACMV	ACMV+EACMV-CBSV-UCBSV-	22.52
	ACMV+EACMV+UCBSV+CBSV-	20.30
	ACMV+EACMV+CBSV+UCBSV-	21.68
	ACMV+CBSV+UCBSV+EACMV-	20.52
	ACMV+EACMV+CBSV+UCBSV+	21.44
	EACMV+CBSV+UCBSV+ACMV-	ND
	Healthy	ND
	NTC	ND
EACMV	EACMV+CBSV+UCBSV+ACMV-	25.66
	EACMV+ACMV+CBSV+UCBSV-	27.46
	EACMV+ACMV+UCBSV+CBSV-	29.56
	Healthy	
	NTC	

*ND stands for not detectable

** Virus templates were not necessarily from same sample of RNA extract. RNA extracts extracted from same virus source plant at different times were used.

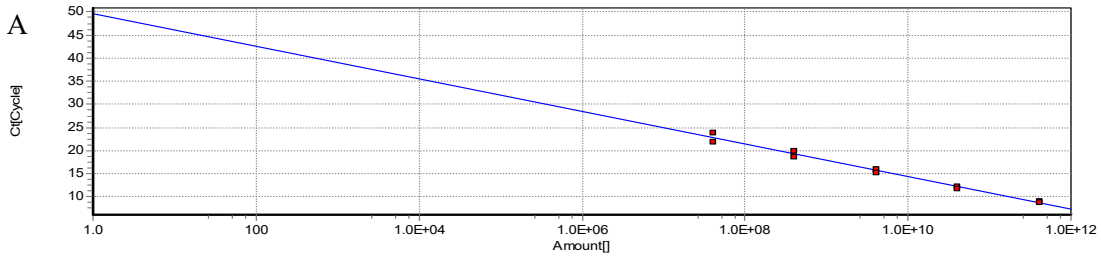
Table 5. Limits of detection and reaction efficiency for multiplex and uniplex assays of four cassava virus species

	Multiplex qPCR assay			Uniplex qPCR assay		
Virus species	Lower detection limit (femtogram)	Reaction efficiency (%)	R^2 (%)	Lower detection limit (femtogram)	Reaction efficiency	R^2 (%)
CBSV	9.9	95.0	99.7	9.9	97.0	99.7
UCBSV	9.5	92.0	97.7	9.5	90.0	97.3
ACMV	12.5	90.0	99.2	12.5	96.0	99.9
EACMV	4.0	93.0	98.8	4.0	83.0	99.2

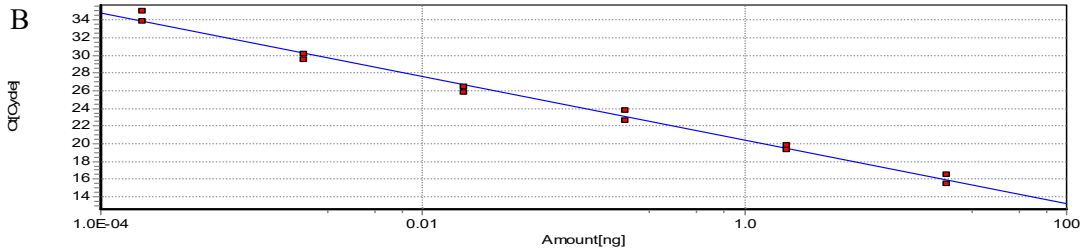
R^2 = coefficient of determination

Intra- and inter-assay variation in Cq values were determined. Overall, acceptably low coefficient of variation for Cq measures were obtained. Intra-assay variability was lowest for CBSV, ranging from 0.0% to 2.0% of mean Cq, and highest for UCBV at range 0.7% to 8.1% (Table 6). Coefficient of variation for Cq values of CBSV were also the lowest across three separate runs. Cq values measured for this target differed by 3.8% to 5.9% of mean Cq, across runs (Table 6). Cq values measured for single and mixed template samples of each virus did not vary significantly ($P > 0.05$, ANOVA) between runs (Table 7).

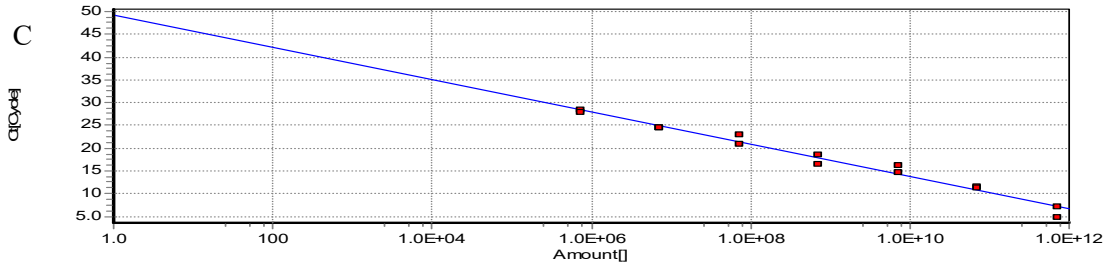
Comparison of CV for single and mixed template samples, showed difference in extent of their Cq variations. Across all four virus species, intra-run CV ranged from 0.0% to 2.6% for single template samples while the range was 0.3% to 8% for mixed template samples (Table 6). Between runs, the ranges were 2.7% to 4.9% and 2.3% to 10.6% for single and mixed template samples, respectively.



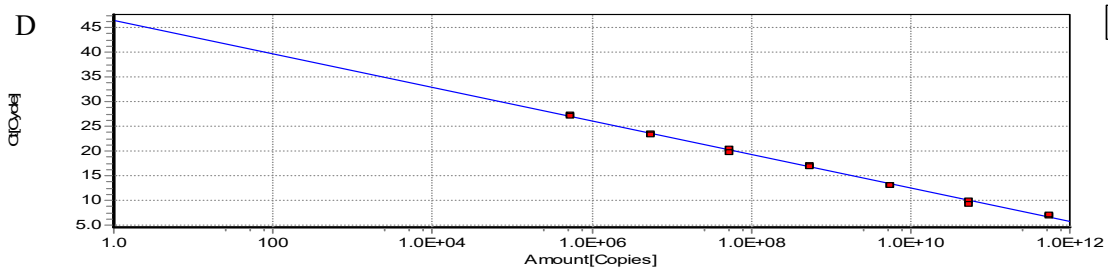
Slope: -3.515
Y-Intercept: 49.59
Efficiency: 0.93
R^2: 0.988



Slope: -3.599
Y-Intercept: 20.45
Efficiency: 0.90
R^2: 0.992



Slope: -3.540
Y-Intercept: 49.27
Efficiency: 0.92
R^2: 0.977



Slope: -3.394
Y-Intercept: 46.52
Efficiency: 0.97
R^2: 0.997

Figure 6. Regression plots from multiplex real time qPCR assay of 10X serially-diluted A. EACMV B. ACMV C. UCBSV and D. CBSV standards.

3.3.2. Comparison of Cq values in different combinations of virus templates

Difference in Cq values measured for each virus only varied marginally across aliquots of the virus template from same sample when tested in different combinations with other virus templates (Table 8). Standard deviation from mean Cq values for these virus template aliquots, obtained in a single run, ranged from 0.46 to 1.31 (Table 8).

3.3.3. The multiplex qPCR method was comparable to its uniplex alternatives

In a single run, quantitative Cq values obtained from multiplex and uniplex assays of each virus target were comparable (Figure 7). Difference in Cq values ranged from ±0.12 to ±2.32 but majority of these differences were an increase in Cq arising from multiplexing. In all four virus targets, the Wilcoxon signed rank test showed that multiplexing did not result in significant change in Cq value at threshold of $P = 0.05$, when Cq values obtained using the multiplex method were compared to those from their comparable uniplex alternatives (Figure 7).

Table 6. Standard deviation and coefficient of variation of intra- and inter-run Cq values from replicates of virus template samples generated using the multiplex qPCR method.

		Intra-assay		Inter-assay	
Virus detected	Virus template sample	Mean Cq ± SD (n=3)	*CV (%)	Mean Cq ± SD (n=3)	CV (%)
ACMV	ACMV+	19.06 ± 0.47	2.5	21.71 ± 0.76	3.5
	ACMV+CBSV+UCBSV+	19.66 ± 0.05	0.3	20.87 ± 1.68	8.1
	ACMV+EACMV+CBSV+	20.04 ± 0.78	3.9	20.83 ± 0.63	3.0
	ACMV+EACMV+UCBSV	21.02 ± 0.75	3.6	20.48 ± 0.47	2.3
	ACMV+EACMV+CBSV+UCBSV	21.36 ± 0.54	2.5	21.04 ± 0.63	3.0
EACMV	EACMV+	24.0 ± 0.34	1.4	24.26 ± 0.65	2.7
	EACMV+CBSV+UCBSV	23.75 ± 0.22	0.9	24.85 ± 0.99	4.0
	EACMV+ACMV+UCBSV	24.12 ± 0.26	1.1	26.47 ± 2.80	10.6
	EACMV+ACMV+CBSV	24.80 ± 0.51	2.0	25.24 ± 1.69	6.7
	EACMV+ACMV+CBSV+UCBSV	24.51 ± 0.12	0.5	25.46 ± 1.76	6.9
CBSV	CBSV+	27.15 ± 0.00	0.0	27.17 ± 1.34	4.9
	CBSV+UCBSV+ACMV+	28.28 ± 0.57	2.0	26.92 ± 1.24	4.6

	CBSV+UCBSV+EACMV+	28.08 ± 0.24	0.9	26.70 ± 1.22	4.5
	CBSV+ACMV+EACMV+	28.24 ± 0.17	0.6	26.72 ± 1.58	5.9
	CBSV+UCBSV+ACMV+EACMV	30.81 ± 0.12	0.4	27.76 ± 1.07	3.8
UCBSV	UCBSV+	27.09 ± 0.70	2.6	28.70 ± 1.20	4.3
	UCBSV+CBSV+EACMV+	23.71 ± 0.17	0.7	26.90 ± 1.14	4.2
	UCBSV+ACMV+EACMV	23.18 ± 1.89	8.1	29.46 ± 1.63	5.5
	UCBSV+CBSV+ACMV	23.28 ± 1.39	6.0	27.73 ± 0.83	3.0
	UCBSV+CBSV+ACMV+EACMV	25.33 ± 0.467	1.8	26.23 ± 0.92	3.5

*CV stands for coefficient of variation

Table 7. Coefficient of variation and ANOVA test results for variation of inter-run Cq values of single and mixed template samples of each virus species.

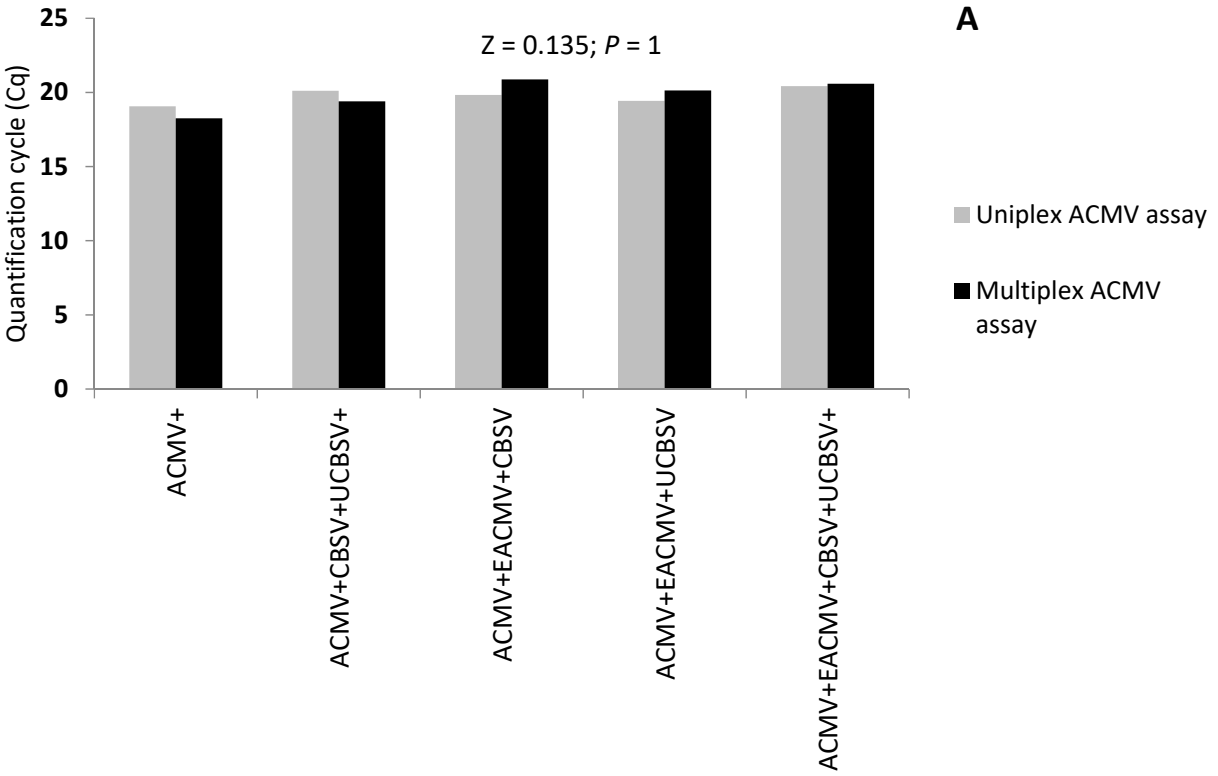
Virus species	*Mean Cq of virus samples	**CV range (%)	*df*	*F*	*P*
ACMV	21.3	4 – 5	4	0.35	0.79
EACMV	25.6	2 – 10.8	4	0.49	0.74
CBSV	27.1	3 – 5	4	0.43	0.78
UCBSV	27.8	3 – 8	4	0.95	0.48

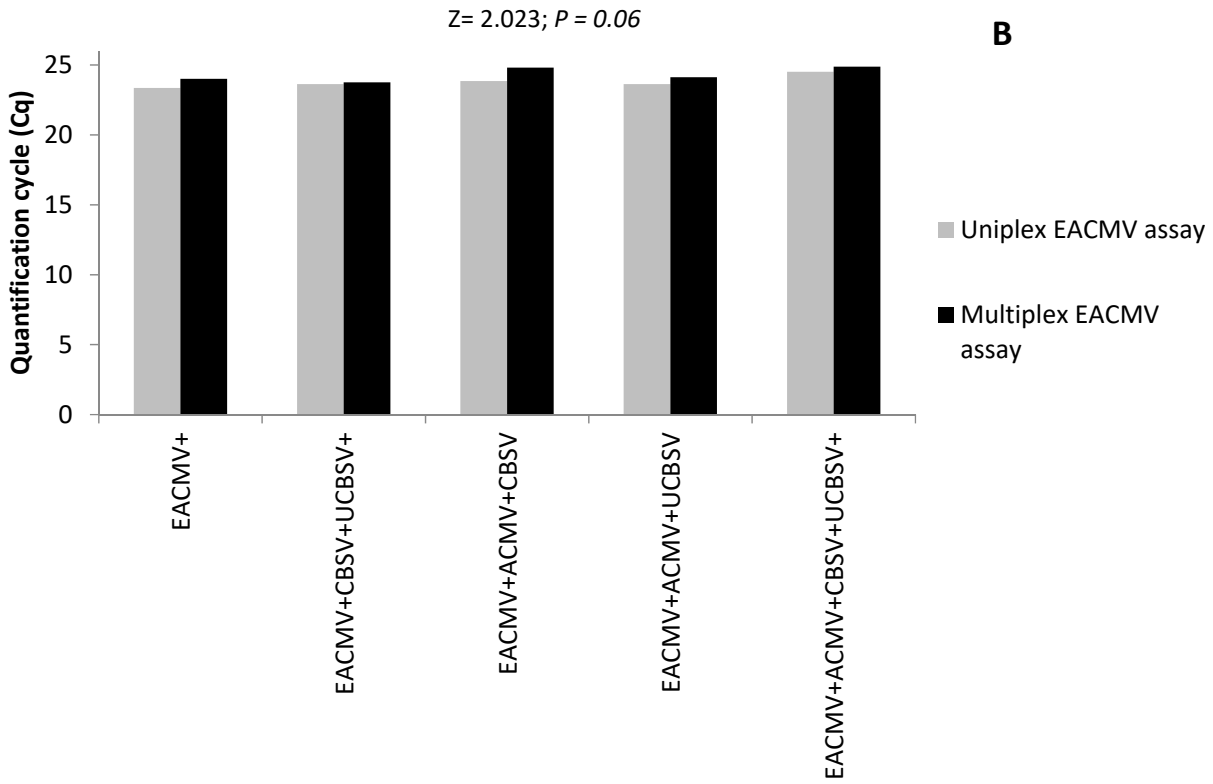
*Mean Cq of single and mixed virus template samples measured over three separate runs.

**Range of CV measurements from three runs.

Table 8. Mean Cq for ACMV, EACMV, CBSV and UCBSV templates sampled from same virus RNA but analysed in different combinations of other virus templates.

Virus detected	Virus template sample	Mean Cq	Standard deviation
ACMV	ACMV+CBSV+UCBSV+	19.66	
	ACMV+EACMV+CBSV+	20.04	
	ACMV+EACMV+UCBSV	21.02	
	ACMV+EACMV+CBSV+UCBSV	21.36	0.80
EACMV	EACMV+CBSV+UCBSV	23.75	
	EACMV+ACMV+UCBSV	24.12	
	EACMV+ACMV+CBSV	24.8	
	EACMV+ACMV+CBSV+UCBSV	24.51	0.46
CBSV	CBSV+UCBSV+ACMV+	28.28	
	CBSV+UCBSV+EACMV+	28.08	
	CBSV+ACMV+EACMV+	28.24	
	CBSV+UCBSV+ACMV+EACMV	30.81	1.31
UCBSV	UCBSV+CBSV+EACMV+	23.71	
	UCBSV+ACMV+EACMV	23.18	
	UCBSV+CBSV+ACMV	23.28	
	UCBSV+CBSV+ACMV+EACMV	25.33	1.00


A
Z = 0.135; P = 1
Quantification cycle (Cq)
25
20
15
10
5
0
ACMV+
ACMV+CBSV+UCBSV+
ACMV+EACMV+CBSV
ACMV+EACMV+UCBSV
ACMV+EACMV+CBSV+UCBSV+
Uniplex ACMV assay
Multiplex ACMV assay


B
Z= 2.023; P = 0.06
Quantification cycle (Cq)
25
20
15
10
5
0
EACMV+
EACMV+CBSV+UCBSV+
EACMV+ACMV+CBSV
EACMV+ACMV+UCBSV
EACMV+ACMV+CBSV+UCBSV+
Uniplex EACMV assay
Multiplex EACMV assay

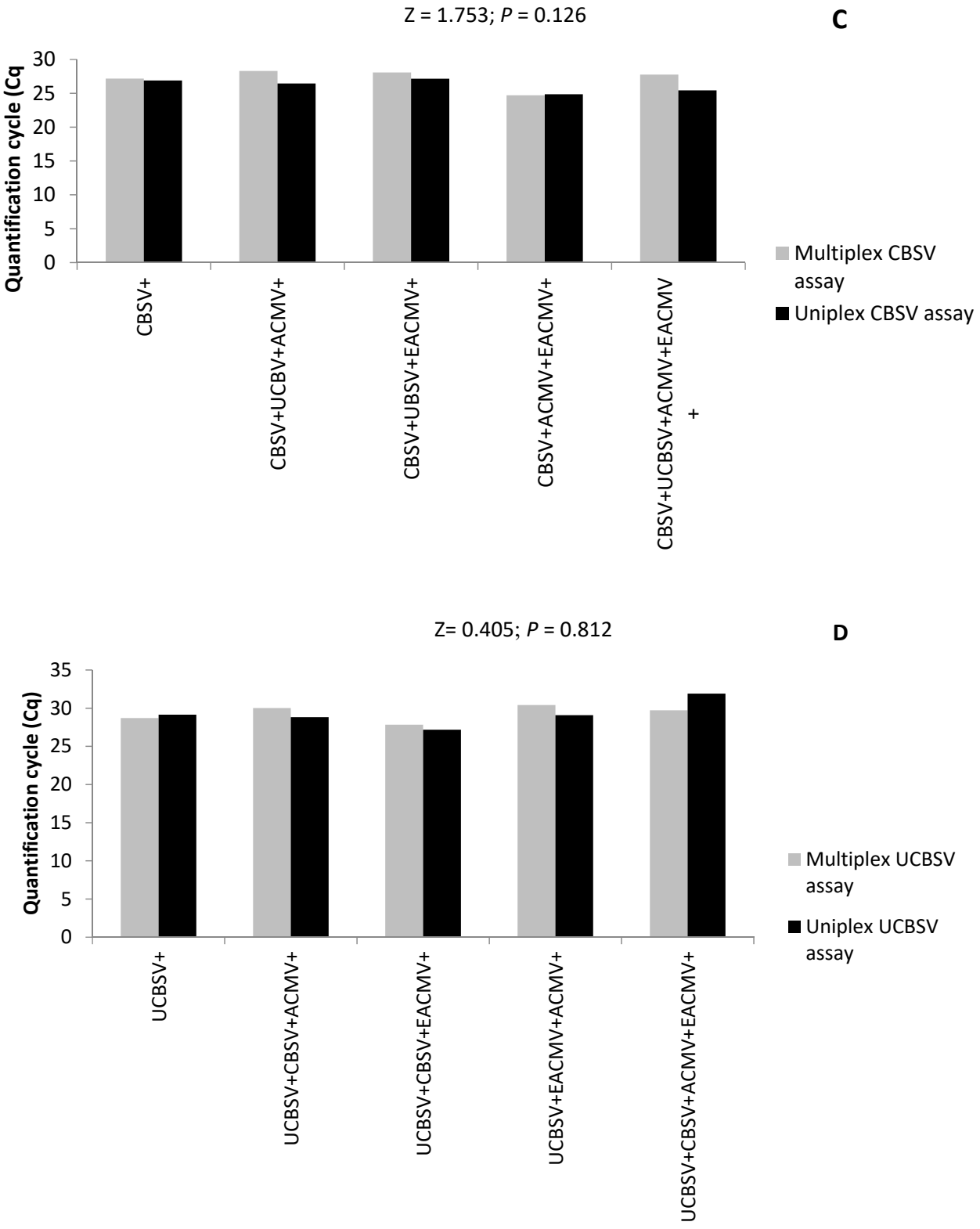


Figure **7**. Cq values from multiplex and uniplex qPCR assay of (a) ACMV (b) EACMV (c) CBSV and (d) UCBSV. Z is Z-score statistic from Wilcoxon signed rank test while p is p-value for above test.

3.3.4. Incidence of cassava viruses in field-cultivated cassava

CBSV was the most prevalent virus detected using both multiplex real time qPCR and end-point PCR methods in field-collected samples (Table 9). Using end-point PCR, CBSV was detected in 30 out of the 80 samples tested. CBSV was also detected in these 30 plants using the multiplex qPCR method. However, one additional sample, diagnosed as being CBSV-free by end-point PCR, was infected with the virus when assayed using the multiplex qPCR method. From the fore-going, more than half of the 80 field samples tested – specifically 49 plants tested negative for CBSV using the multiplex qPCR method. Thirty-five of these 49 plants belong to the seven varieties - KBH 2002/66, KBH 2002/477, NDL 2005/1472, KBH 2002/26, NDL 2003/111, KBH 96/1056 and KBH 2005/1471 - with zero CBSV incidences (Table 9).

Additional 18 UCBSV infections were identified by multiplex qPCR compared to end-point PCR (Table 9). Cq values for these samples were lower than average for CBSV and UCBSV ranging from 28.1 to 32.6 and 27.0 to 33.5 for CBSV and UCBSV, respectively. A different infection scenario was observed for the CMBs. Only a single sample was found infected with EACMV, from end-point PCR while two were infected, according to multiplex qPCR assay. ACMV incidence was 1 of the 80 samples according to multiplex qPCR assay but was absent when end-point PCR analysis was used (Table 9).

Table 9. Incidence of ACMV, EACMV, CBSV and UCBSV in 16 field-grown cassava varieties measured by end-point PCR and multiplex qPCR assays.

Cassava variety	Number infected/ Number tested by end-point PCR				Number infected/ Number tested by multiplex qPCR			
	ACMV	EACMV	CBSV	UCBSV	ACMV	EACMV	CBSV	UCBSV
Albert	0/5	0/5	4/5	0/5	0/5	0/5	5/5	2/5
Naliendele	0/5	0/5	2/5	3/5	0/5	1/5	2/5	4/5
NDL 2003/031	0/5	0/5	0/5	0/5	1/5	0/5	1/5	0/5
KBH 2002/66	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5
Kiroba	0/6	0/6	4/6	0/6	0/6	0/6	4/6	5/6
KBH 2002/477	0/5	0/5	0/5	3/5	0/5	0/5	0/5	3/5
NDL 2005/1472	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5
KBH 2002/26	0/5	0/5	1/5	0/5	0/5	0/5	0/5	0/5
Pwani	0/5	0/5	4/5	0/5	0/5	0/5	5/5	2/5
NDL 2003/111	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5
Mkumba	0/5	0/5	5/5	0/5	0/5	0/5	5/5	2/5
KBH 96/1056	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5
KBH 2005/1471	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5
KBH 2002/494	0/5	0/5	4/5	0/5	0/5	0/5	3/5	1/5
Mahiza	0/5	0/5	5/5	0/5	0/5	0/5	5/5	5/5
KBH 2002/482	0/5	1/5	1/5	0/5	0/5	1/5	1/5	0/5
Total number infected/ total number tested	0/81	1/81	30/81	6/81	1/81	2/81	31/81	24/81

3.3.5. Virus titre in field-grown samples

Mean CBSV quantity and incidence was highest in the variety Mahiza, Albert and Naliendele (Figure 8a). On the other hand, the lowest CBSV quantities, of varieties with at least one case of CBSV infection, was found in NDL 2003/031, Mkumba, Pwani, KBH 2002/494 and the known tolerant variety Kiroba. UCBSV distribution in the assayed varieties was very similar to the above observation for CBSV. The varieties Mahiza, Naliendele, Albert in addition to KBH 2002/477, which was infected only by UCBSV, have the highest UCBSV quantities. On the other hand, Mkumba, Kiroba, and KBH 2002/494 have the lowest UCBSV quantities (Figure 8b).

A total of 55 plants were detected to be positive for either CBSV or UCBSV or both by qPCR. Among these, over 48% were infected by both viruses while 51.4% were single infections of CBSV or UCBSV. Simultaneous single CBSV and mixed CBSV and UCBSV infections were observed in five of the 16 varieties tested. These are KBH2002/494, Mahiza, Albert, Mkumba and Pwani. On the other hand, occurrence of single UCBSV and mixed UCBSV and CBSV infections was observed only for the two varieties – Naliendele and Kiroba. The mean quantity of CBSV was higher in mixed infected plants than single infected ones in four of five varieties by up to 10 times (Figure 9). UCBSV quantity in mixed infections was four and 14 X higher times higher than single UCBSV infections in Kiroba and Naliendele respectively. ANOVA analysis using the randomized complete block design showed that CBSV quantity in mixed infection is significantly higher in comparison to single infection at $P<0.05$. However, there was no evidence ($P > 0.05$) of influence of variety response to virus on mixed infection effect.

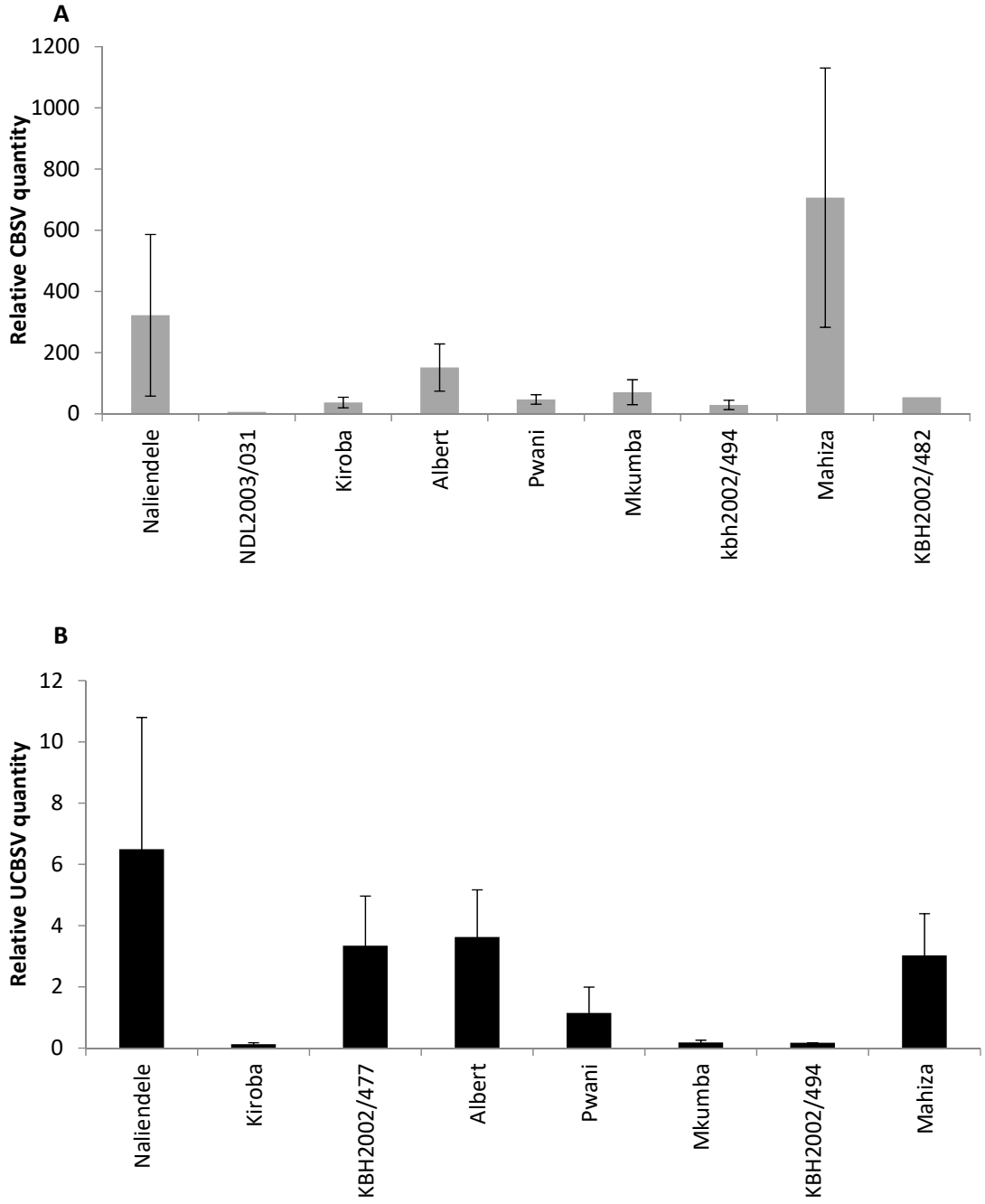


Figure 8. Mean quantities of (a) CBSV and (b) UCBSV in infected samples of cassava varieties.

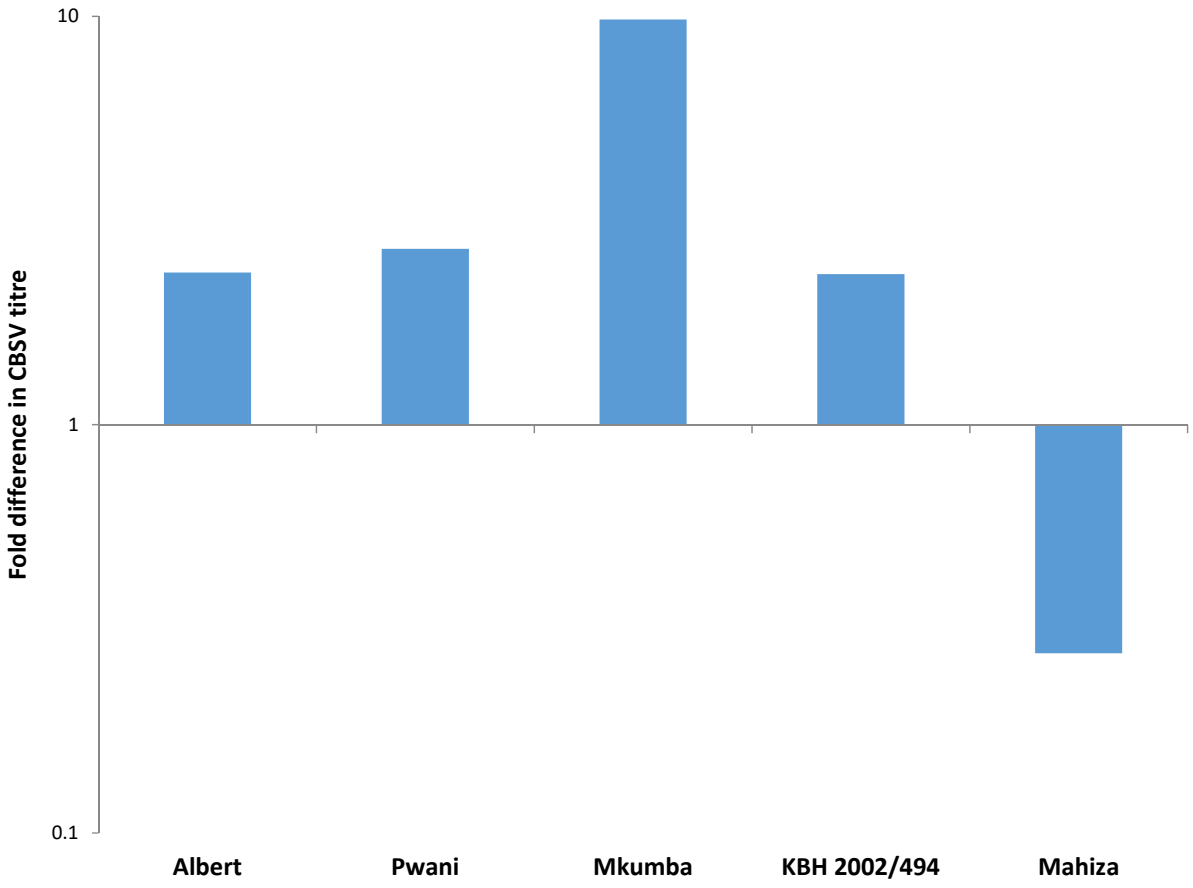


Figure 9. Fold difference between mean CBSV quantity of single CBSV and mixed CBSV and UCBSV infections in five varieties of cassava.

3.4 Discussion

The causal viruses of CMD and CBSD are the most destructive pathogens of cassava in Africa and their specific detection in the field is crucial for epidemiology, quarantine, and resistance identification and breeding. A multiplex diagnostic tool reduces the time, cost and effort associated with the detection of pathogens and is therefore an important asset in managing diseases. Several tools, using conventional PCR have already been designed for both CMBs and CBSVs (Alabi *et al.,* 2008; Abarshi *et al.,* 2010, 2012; Mbanzibwa *et al.,* 2011), however, the absence of quantitative data limits the discrimination between tolerant varieties that sustain low quantities of viruses, and susceptible varieties with high level of virus in the plant (Maruthi *et al.,* 2014). When available, such studies were developed for single virus (e.g. for CBSV by Moreno *et al.,* 2011) or for the viruses of the same group (e.g. for CBSV and UCBSV by Adams *et al.,* 2013), but not for CBSVs and CMBs together, while infection by both groups of viruses is a common feature in farmer fields.

The current study was devised to develop and evaluate a real time qPCR method for sensitive quantitative detection of four major viruses of cassava in East Africa – the CMBs - ACMV, EACMV and the CBSVs - CBSV and UCBSV, simultaneously in a single reaction. Evaluation of specificity showed that the method can uniquely detect each of the four viruses. Each virus could be detected in samples containing the virus alone or in combination with other viruses (Table 4) hence showing that the assay was specific and not hampered by presence of other virus templates. Cq values measured for each virus, was fairly consistent within and between assay runs for samples containing only a single template (Table 6). Though, variations were slightly higher when virus targets were assayed in combination with other templates, there was no significant difference ($P > 0.05$) in inter-assay Cq values for single and mixed template samples of each virus species.

From Figure 4, Cq values obtained using the multiplex method were marginally higher than values obtained from a corresponding uniplex real time PCR assay, in the majority of the analysed samples. Cq value difference between multiplex and uniplex assays ranged from 0.12 to 2.32 – equivalent to quantity fold differences between 1.1 to 5.0, assuming a 100% reaction efficiency. However, there was no significant change ($P < 0.05$) in Cq values derived from between multiplex and uniplex assays implying that the multiplexing protocol did not alter assay accuracy. This was in line with the observation of equal lower detection limits, in 10X dilution standards (Table 5) and similar reaction efficiencies for the multiplex and uniplex assay of each virus. Quantification was demonstrated to be reliable due to the high reaction

efficiency (>90%) of each primer and probes combinations and low limits of detection for each target varying between 12.5 and 4.0 fg.

The multiplex qPCR method was successfully applied to screen 80 samples belonging to 16 varieties of field grown cassava plants. The method showed superior results to the end point PCR and was able to detect the four viruses in additional samples that failed to amplify in the latter. Apart from the var. Mahiza, CBSV and UCBSV quantities in these additional samples were among the lowest of the 80 samples tested, suggesting that the improved sensitivity more than broader specificity was the reason for the enhanced detection. The qPCR method demonstrated to be more sensitive than existing conventional PCR designed using primer sets picked from different virus genomic regions, and therefore a more precise tool for quarantine and resistance screening purposes. The exact reason for the higher sensitivity – higher efficieny of the different primer sets used in the multiplex qPCR assay, or the inherently higher senstitivity of the qPCR technique over end-point PCR or both – is not exactly known given that the qPCR primers were not applied in end-point PCR. Nonetheless, accuracy of virus quantification for the multiplex method described could be further improved by the ensuring that the amount of starting cDNA and DNA templates are uniform across samples within a run and across different runs. This is important because though the endogenous primer used was designed for normalisation of Cq values for the RNA viruses, it can also amplify DNA templates.

Across virus species assayed in the field samples, CBSV and UCBSV were by far the most prevalent with 37 out of the 80 field samples being tested positive for at least one or both viruses. Out of these, 19 samples (51.4%) were infected with both viruses. These data were similar to earlier reports of CBSV and UCBSV mixed infections in different areas of Tanzania hence re-affirming the observation that both viruses were not always geographically separate and can co-infect a single plant (Mbanzibwa *et al.,* 2011; Abarshi *et al.,* 2012). For varieties infected with both viruses, average CBSV and UCBSV quantities were generally higher for mixed infection cases than in single infections of each of both viruses. This result suggests the possible existence of synergism between CBSV and UCBSV in field-infected cassava but does not prove it. An experiment designed to test whether higher combined virus accumulation occurs in mixed-infected cassava following simultaneous controlled exposure to both viruses and whether sequential infection by both viruses enhance accumulation of each other would be required to prove synergy and gain insight into how it works. The absence of any significant ($P > 0.05$) interaction or association between higher virus quantities in mixed infection samples

and the genotype effect suggests that both factors exert independent effects on virus quantity of samples. Synergism between viruses infecting cassava has already been reported. ACMV and EACMV act in synergy to produce increased symptom severity and higher accumulation of virus DNA in field-grown cassava (Fondong *et al.,* 2000). Evidence of synergism by CBSVs, however, has been debated. Co-infection of cassava with CBSV and UCBSV by bud grafting resulted in symptoms similar to CBSV alone (Wagaba *et al.,* 2013) implying a masking of UCBSV symptom expression. Kaweesi *et al.,* (2014) recently suggested the existence of competition between CBSV and UCBSV as possible explanation for the higher CBSV than UCBSV titres in co-infected cassava. However, UCBSV is known to accumulate at lower levels compared to CBSV and show milder symptoms even in single infections (Winter *et al.,* 2010; Mohammed *et al.,* 2012). Moreover, Ogwok *et al.,* (2014) also reported higher accumulation of total viral RNA (i.e. CBSV and UCBSV) in field-collected samples diagnosed for both viruses compared to those diagnosed for CBSV alone. While this further indicates a synergistic action, effect of the cassava genotype remained unknown. The current study has presented evidence of an overall substantially higher CBSV quantities in mixed infected plants compared to single infections and shown that this difference was not influenced by genotype determinant of virus accumulation levels. This evidence notwithstanding, further studies would be required to prove the exact nature of interaction of both viruses in cassava.

The multiplex qPCR method provides an appropriate tool for the cost effective study of CBSV-UCBSV interaction and ultimately CBSVs-CMBs interaction in cassava. The ability to detect four cassava viruses in a single tube will reduce time, reagents and consumable costs approximately four-fold compared to corresponding uniplex alternatives, hence is more suitable for high-throughput applications. Prior to the current study, the SYBR green chemistry has been utilized for uniplex quantification of CBSV (Moreno *et al.,* 2011) and *South African cassava mosaic virus* (SACMV) (Allie *et al.,* 2014). Real time qPCR methods, based on the TaqMan chemistry, have also been developed for quantification of ACMV and EACMV in separate uniplex assays (Otti *et al.,* 2013) and CBSV and UCBSV in duplex reactions (Adams *et al.,* 2013) although target masking effect limits application of the latter for quantification purposes (Adams *et al.,* 2013). The evidence of high sensitivity, specificity and reproducibility presented for the method here represents the first demonstration of the potential to effectively quantify all the four most important viruses plaguing cassava production in eastern and central Africa.

3.5 Conclusion

A sensitive and efficient multiplex qPCR method which allows cost-effective and high-throughput quantification of the major viruses of cassava in eastern and central Africa has been developed in the study. Validation of the method uncovered the existence of a CBSV-UCBSV synergistic effect on CBSV quantities in cassava varieties and landraces cultivated in east Africa highlighting the potential for resistance breakdown from this phenomenon, hence a need to understand its effect on durability of CBSD resistance.

CHAPTER 4: Screening for CBSD resistance using multiplex real time PCR method

4.1 Introduction

Since the demonstration of CBSD resistance in the cultivar 46106/27 (locally called Kaleso in Kenya) there has been ongoing effort to identify other CBSD resistance sources. Kaleso is a third backcross derivative of a hybrid obtained from inter-specific cross between cassava and a wild Manihot species – *Manihot glaziovii* (Ceara rubber) which diverged from the former at approximately 2 to 3 million years ago (Childs, 1957; Bredesson et al., 2016). A major challenge to the deployment of CBSD resistance varieties has been finding sources which combine CBSD resistance or tolerance with CMD resistance and which possess desirable agronomic and culinary qualities needed by farmers. This fact underscores the need to identify new sources of tolerance among local germplasm, increasing the likelihood of delivering dual-resistance sources as well as cultivars with other farmer-preferred qualities.

Available CBSD tolerant varieties have been identified mainly by field-based resistance screening. Hillocks and Jennings (2003) provided useful guidelines for this process. This involves the use of virus-free cuttings in high disease pressure environments to determine propensity for infection, and planting of infected cuttings to determine tendency to develop root necrosis. While field-based evaluation methods are important for CBSD resistance screening, it is a costly approach in terms of resources and time and is subject to the vagaries of the natural environment which can influence plant response to the disease. As Hillocks and Jennings (2003) pointed out, a high inoculum pressure is required to ensure transmission of CBSV. This is not always guaranteed in natural environments as whitefly populations vary across environments and at different times. A typical case was an evaluation study of local CBSD tolerant varieties in five Tanzanian villages and at the Naliendele Agricultural Research Institute between 2002 and 2003. This study was inconclusive due to low inoculum pressure during the period as determined by low whitefly populations (Hillocks, 2005). Second, virus transmission in natural environments is random. Hence in a given environment, there is always the chance that the vector will miss out some plants. Though this problem is often addressed by multi-location trials, the cost and labour-intensive nature of evaluation in different environments limits the number of accessions that can be screened in a given study, hence the pace of progress of identification of resistant accessions. Adding to these challenges include the unreliability of current phenotype-based CBSD resistance evaluation method (Jennings, 1960; Hillocks and Jennings, 2003). CSBD symptom expression, which is the basis of current

severity assessment-based resistance screening, is known to be influenced by variety of factors including plant age and environmental temperature (Hillocks and Jennings, 2003).

These limitations can be largely eliminated in controlled evaluation experiments involving virus transmission by graft inoculation and time-course measurement of virus quantities. Currently, the most efficient means of controlled transmission of CBSV is by graft inoculation which can result in up to 100% infection rate in side-grafted susceptible varieties (Mohammed *et al.*, 2012), and between 70 to 100% using the bud grafting method (Wagaba *et al*., 2013). In the current study, 13 local east African varieties were graft-inoculated with CBSV and subsequently monitored for accumulation of the virus, using the real time PCR method, in order to identify new resistance sources and characterize variety response.

4.2 Materials and Methods

4.2.1 Plant materials and growth condition

Thirteen east African cassava accessions were used for the study (Table 10). Plants for each variety were grown from cuttings of plants confirmed virus-free after testing for the presence of both cassava mosaic begomoviruses (CMBs) and cassava brown streak viruses (CBSVs). Plants were grown under quarantine at 28 ± 5ºC and 50-60% humidity.

4.2.2 Virus inoculation and assessment of virus accumulation and symptom severity

Two-month old plants of the 13 accessions were inoculated with CBSV maintained in clones of the susceptible cultivar Albert. CBSV was inoculated by side-wedge grafting method. The method involved making a slightly slanted downward slit on healthy stock plant, insertion of virus-infected scion made into a matching wedge shape followed by gentle but firm wrapping of the graft union using parafilm tape, to secure it. To prevent excessive moisture loss, grafted plants were covered in moisture chamber made from transparent polythene bags with perforations.

Between four and five plants were inoculated for each variety, and one of these plants was grafted with a healthy scion as mock inoculation. Virus accumulation was monitored in leaf samples using real time quantitative PCR at weekly interval from second week to fourth week post inoculation. Subsequent virus quantification was done monthly up to 9 months. CBSV was also quantified in RNA samples extracted from root tissue at 40 weeks post-inoculation (wpi).

Table 10. East African cassava varieties screened for resistance to CBSD

Country of Origin	Accession name
Malawi	Mbundumali
	Kalawe
Uganda	Nase 3
	TME 204
Tanzania	Pwani
	Mkumba
	Kizimbani
	Albert
Mozambique	Oekhumelela
	Orera
Kenya	F10-30-R2
	Mkumbozi
	Kaleso

CBSD symptom severity was monitored visually in leaf and root tissue using scoring method described in Rwegasira and Rey (2012). Foliar symptom severity score for each plant was calculated as average of severity, ranging from 1 for symptomless through 2 and 3, for slight and obvious chlorotic spots covering up to 5% and 12% of leaf area respectively, to 4 and 5 for appreciable and very severe chlorosis affecting up to 30% and 100% respectively of each leaf of a plant. Similar criteria was used for estimated assessment of CBSD severity in each root tuber.

4.2.3 Duplex qPCR assay design

Relative quantity of CBSV was estimated in RNA extracts of leaf and root tissue. A duplex RT-qPCR assay based on the TaqMan chemistry was developed for this purpose. A fluorescently labelled probe (Table 4) for CBSV assay was designed between previously reported primer sequences of the CBSV *HAM1*-like gene (Abarshi *et al*., 2012). An internal endogenous control probe (Table 4) was designed between existing primers of the serine-threonine phosphatase gene - *PP2A* (Accession number: CK650945) which has been shown to be stably expressed in leaf, stem and root tissue of CBSV-infected cassava (Moreno *et al.*, 2011). Endogenous probe design involved checking melting temperature, GC content and potential for intra- and inter-oligonucleotide sequence complementarity of sequences selected between existing *PP2A* primers. To ensure the endogenous probe sequence anneals only to the intended position on cassava genome, it was used as query in a BLAST search of the genome on Phytozome database (www.phytozome.net) and also against nucleotide sequence collection of CBSV and UCBSV on NCBI database to confirm the absence of potential to cross-anneal with any of the viral genomes. CBSV-specific probe sequence was selected after a total of 15 complete CBSV and UCBSV genome sequences, from EMBL ENA database (http://www.ebi.ac.uk/ena), were aligned using the CLUSTALW program (Larkin *et al.,* 2007; Figure 10.

CBSV

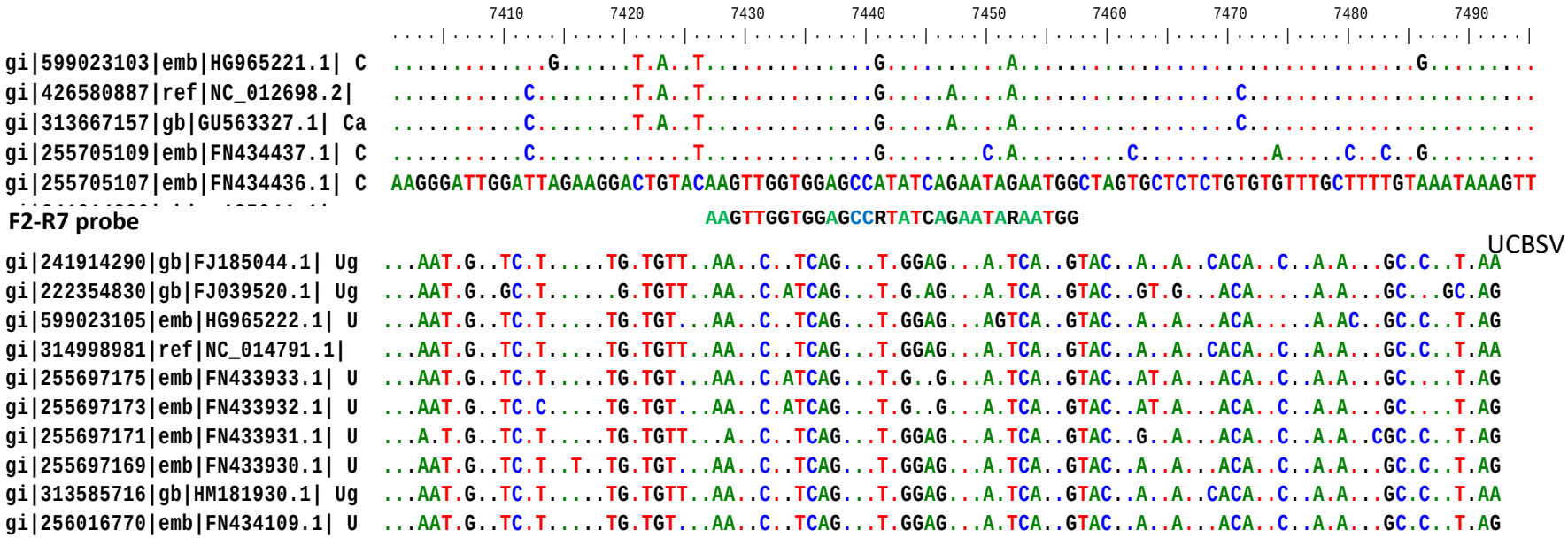


Figure 10. CLUSTALW alignment output for multiple sequence alignment (MSA) of five and 10 full-length CBSV and UCBSV HAM1-like gene sequences respectively, showing position of the CBSV probe F2-R7.

Melting temperature, percentage GC content and potential hairpin or cross-dimer formation was checked for selected sequences, under default settings, using the multiple primer analyser web tool from Thermoscientific Ltd. (http://www.thermoscientificbio.com/webtools/multipleprimer/). The CBSV F2-R7 and *PP2A* probe sequences were found not to form any significant self-dimer, hairpins or cross-dimers. Both probes were labelled differently (Table 11) to allow for discrimination of their fluorescence signals in a duplex assay.

4.2.4 Real Time Quantitative PCR experiment

Prior to real time PCR assay, leaf samples were collected from each plant at time points specified above. Approximately 100 mg of leaf tissue was sampled from all fully opened leaves of each plant. Total nucleic acids were extracted from the collected samples following a modified Cetyl trimethyl ammonium bromide (CTAB) method (Maruthi *et al.*, 2002) and quantified using a NanoDrop spectrophotometer 2000 (Thermoscientific, Wilmington USA). DNA was removed by treatment with Ambion's DNase I (RNase-free) following manufacturer's protocol (Life technologies, Paisely UK). Virus quantification analysis was carried out using a two-step multiplex TaqMan qPCR method. The first step involved generation of complementary DNA (cDNA) from 800 ng DNase-treated RNA template which was first mixed with 20 pico moles of oligo $dT_{(18)}$, incubated at 70ºC for 5 minutes and immediately chilled at 4ºC . A 15µl of reverse transcription reaction mix consisting of 4 µl of $MgCl_2$, 1X buffer, 1 µl of ImProm-II™ Reverse Transcriptase (Promega Corporation, Madison USA) and 125 nM dNTP mix was added to the chilled template-oligo dT(18) mix and incubated at 40 ºC for an hour following a 20 minutes pre-incubation at 25ºC. The reverse-transcriptase enzyme was inactivated by heating at 70 ºC for 10 minutes. The cDNA obtained was used as template for the duplex qPCR assay in which CBSV was co-amplified with the internal endogenous gene – *PP2A* (Moreno *et al*., 2011). The assay was carried out in a 20 µl mix consisting of 10 µl of 2X Express qPCR supermix (Life Technologies, Paisely UK), 2 µl cDNA template, 700 nM each of the virus-specific primers CBSVF2 and CBSVR7 (Table 11), 250 nM each of the internal control primers PP2AF and PP2AR and 400 nM and 600 nM of CBSVF2/R7 and *PP2A* probes respectively.

Table 11. Primers and probes used for duplex real time qPCR assay of CBSV and the endogenous control - serine threonine phosphatase

Oligo	Tm	5′-label	3′-label	Amplicon length (bp)	Reference gene accession number	5′ - 3′ sequence	Reference
CBSV F2	52.0	Nil	Nil	345 with CBSV R7	FN434437	GGRCCATACATYAARTGGTT	Abarshi *et al.*, 2012
CBSV R7	51.8	Nil	Nil		FN434437	CCCTTTGCAAARCTRAAATARC	Abarshi *et al.*, 2012
PP2AF	51.9	Nil	Nil	187 with PP2AR	CK650945	TGCAAGGCTCACACTTTCATC	Moreno *et al.*, 2011
PP2AR	55.8	Nil	Nil		CK650945	CTGAGCGTAAAGCAGGGAAG	Moreno *et al.*, 2011
CBSV F2/R7 probe	59.6	FAM	IBFQ	Nil	FN434437	AAGTTGGTGGAGCCRTATCARAATARAATGG	This study
PP2A probe	61.3	JOE	BHQ-1	Nil	CK650945	CTTTCTGTTGCCCCCACCATGC	This study

Reactions were carried out in the Eppendorf's Mastercycler ep realplex (Eppendorf, Cambridge UK) in 40 PCR cycles of 15 seconds denaturation at 94ºC, 30 seconds annealing at 52ºC and extension at 60ºC for 30 seconds. The PCR cycles were preceded by 2 minutes incubation at 50ºC for removal of carry-over contamination by Uracil DNA Glycosylase (UDG) and a second 2 minutes incubation at 95ºC for activation of the *Platinium® Taq* DNA polymerase. Efficiency and limit of detection for both targets were determined from serially diluted amplicons of each target. Limit of detection in terms of copy number was calculated from quantity, in ng/µl, and size of target amplicons using an online DNA copy number calculator from Thermo scientific (https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/dna-copy-number-calculator.html).

In each run and at all assay time points, RNA template of two weeks old Albert was assayed, together with test samples, as calibrator sample. Each sample was assayed in triplicates.

4.2.5 Statistical analysis methods

4.2.5.1. Mixed model analysis

Overall effect of accession type of analysed samples on level of CBSV accumulation was statistically inferred by F-test comparison of a mixed model with fixed accession and random time effects and a mixed model without accession effect. A Kenward-Roger approximation of denominator degree of freedom, normally recommended for p-value estimation from mixed models, was used in the F-test (Kenward and Rogers, 1997). Using the same method, pair-wise variance in CBSV quantities between each accession and the reference resistant and susceptible varieties - Kaleso and Albert variety was inferred. Kaleso is currently best known source of CBSD resistance while Albert is a known CBSD susceptible variety suitable for culturing CBSV. The R packages lme4 (Bates *et al*., 2014) and pbkrtest (Halekoh and Højsgaard, 2014) were used to generate mixed models and to compare the models, respectively.

4.2.5.3 Simple linear regression

Simple linear regression analysis was performed to test for correlation between CBSD severity and relative CBSV quantity in leaf and root tissue of CBSV-inoculated plants. Coefficient of correlation between mean of time-course foliar CBSV quantity and root CBSV quantity sampled at 10 months post-inoculation (mpi) was also determined.

4.3 Results

4.3.1 Graft inoculation

Of 64 plants grafted, 57 had surviving scion and successful graft union at 3 wpi. This represents 89.1% overall grafting success rate across accessions. Fifty one of the 64 grafted plants were graft inoculated with CBSV. The remaining 13 were mock-inoculated. Forty five or 88.2% of this virus-inoculated set had successful grafts at 3 wpi. In each variety, grafting success rate ranged from 80-100%. Table 12 contains information on grafting success rate determined at 3 wpi and virus incidences per accession from 2 to 4 wpi. CBSV prevalence increased from 2 wpi and peaked 4 wpi at 97.4% (Table 12). Inoculated plants, which are sensitive to CBSD, showed typical CBSD foliar and root symptoms (Figure 11). Symptom expression was early in leaf tissue, and pronounced in leaf and root tissue of sensitive varieties (Figure 11). On the other hand, mock-inoculated plants remained symptomless throughout the duration of the experiment.

Table 12. Grafting success rate at 3 wpi, measured in terms of proportion of survived scion and callus formation at the position of graft union, and incidence of CBSD symptoms and CBSV for the first 4 wpi.(N/A: not available).

Variety	Grafting success rate in all grafted plants	Grafting success rate in CBSV-inoculated plants	CBSD incidence based on symptom observation at 2 wpi	CBSV prevalence at 2 wpi confirmed by RT-qPCR	CBSD incidence based on symptom observation at 3 wpi	CBSV prevalence at 3 wpi confirmed by RT-qPCR	CBSD incidence based on symptom observation at 4 wpi	CBSV prevalence at 4 wpi confirmed by RT-qPCR
Oekhumelela	5/5	4/4	3/3	2/3	4/4	3/3	4/4	2/2
Orera	5/5	4/4	2/4	3/3	2/4	1/1	3/3	4/4
Mbundumali	5/5	4/4	0/4	3/3	3/4	4/4	4/4	3/3
Kalawe	4/5	3/4	1/4	1/3	3/4	3/3	4/4	4/4
Albert	5/5	4/4	4/4	4/4	3/4	4/4	4/4	3/3
Nase 3	4/4	3/3	2/3	2/3	2/3	3/3	2/3	1/1
TME 204	4/5	3/4	2/3	1/3	1/3	1/3	2/2	3/3
Pwani	4/5	3/4	0/3	3/3	N/A	N/A	1/2	2/2
Mkumba	5/5	4/4	0/3	3/3	0/4	4/4	0/3	3/3
Kizimbani	4/5	3/4	0/4	3/3	3/4	3/3	2/2	3/3
F10-30-R2	5/5	4/4	N/A	3/3	N/A	3/3	4/4	3/3
Mkumbozi	4/5	3/4	0/4	2/3	2/3	2/3	3/3	3/3
Kaleso	5/5	4/4	0/4	4/4	0/4	2/3	1/4	3/4
Total	59/64	46/51	14/43	34/41	23/41	33/37	34/42	37/38

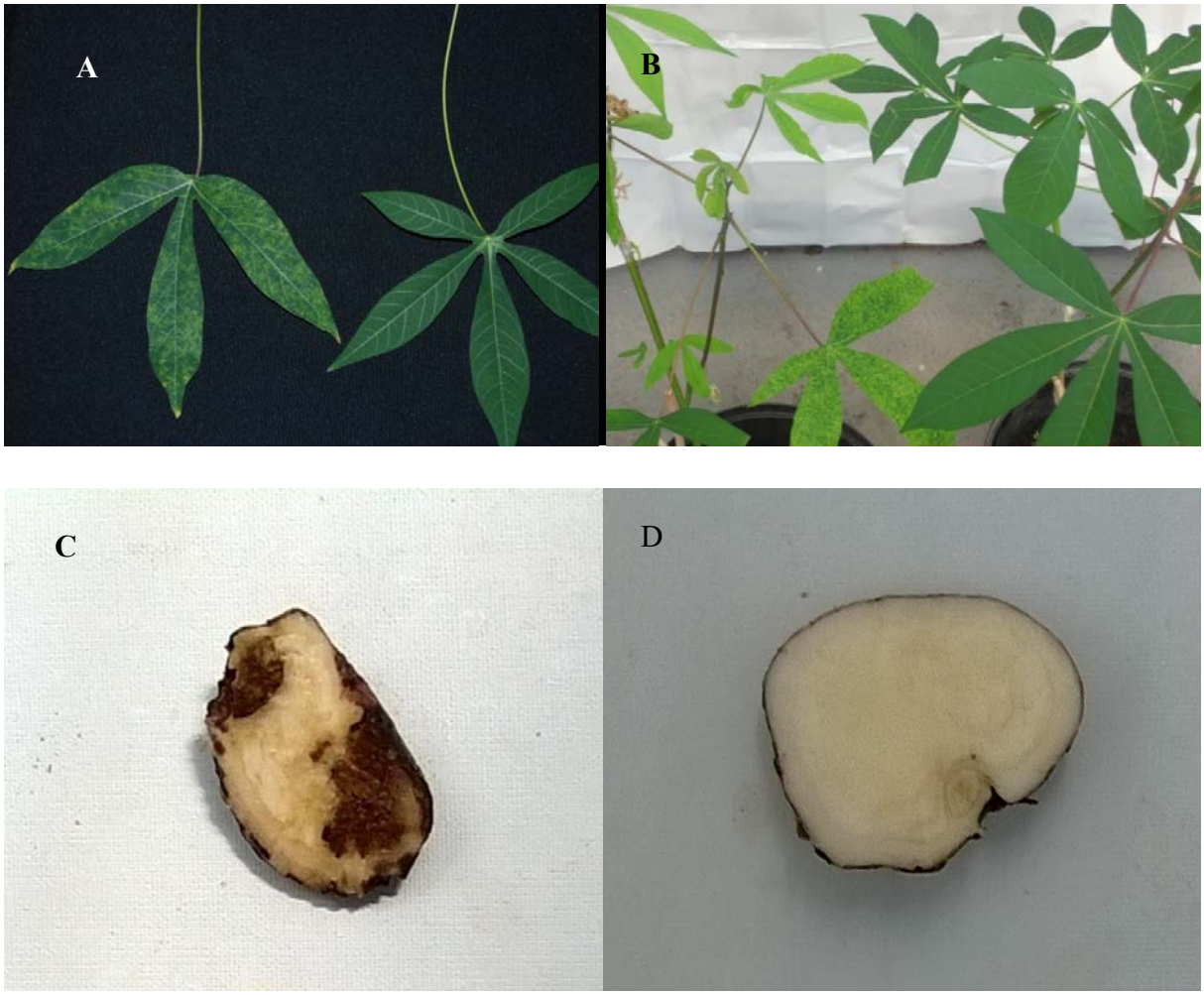


Figure 11. Typical CBSD symptom expression in CBSV-inoculated compared to mock-inoculated: A. leaves of Albert variety; B. Kalawe variety at 2 months post-inoculation; C and D. storage root samples of Kizimbani.

4.3.2 Multiplex qPCR assay performance

Reaction efficiencies of CBSV and endogenous targets were high and comparable in uniplex and duplex assays (Table 13).

Table 13. Reaction efficiencies and lower detection limits for CBSV F2-R7 and PP2A primer and probe systems in uniplex and duplex assays of CBSV and endogenous *PP2A* genes in the same tube.

Assay	F2/R7 primer-probe system		PP2AF/PP2AR primer-probe system	
	Efficiency (%)	Detection limit (copy number)	Efficiency (%)	Detection limit (copy number)
Uniplex	93.0	86	98.0	170
duplex	93.0	430	90.0	850

In a typical run, Ct values for both targets showed little variation among technical replicates with standard deviations typically between 2 and 5% of mean quantitative cycle (Cq) for CBSV and PP2A targets, respectively. Reaction efficiency and sensitivity of the duplex assay was compared with single assays of CBSV and PP2A targets using same target-specific primers and probes at same quantities. There was no difference in efficiency of CBSV quantification between duplex and uniplex alternatives, though efficiency for the *PP2A* target was higher for uniplex assay (Table 13). Both targets were detected and quantified in a specific manner in a single tube as can be seen from Figure 12. However, detection sensitivity for both targets was higher in the uniplex compared to the multiplex assay.

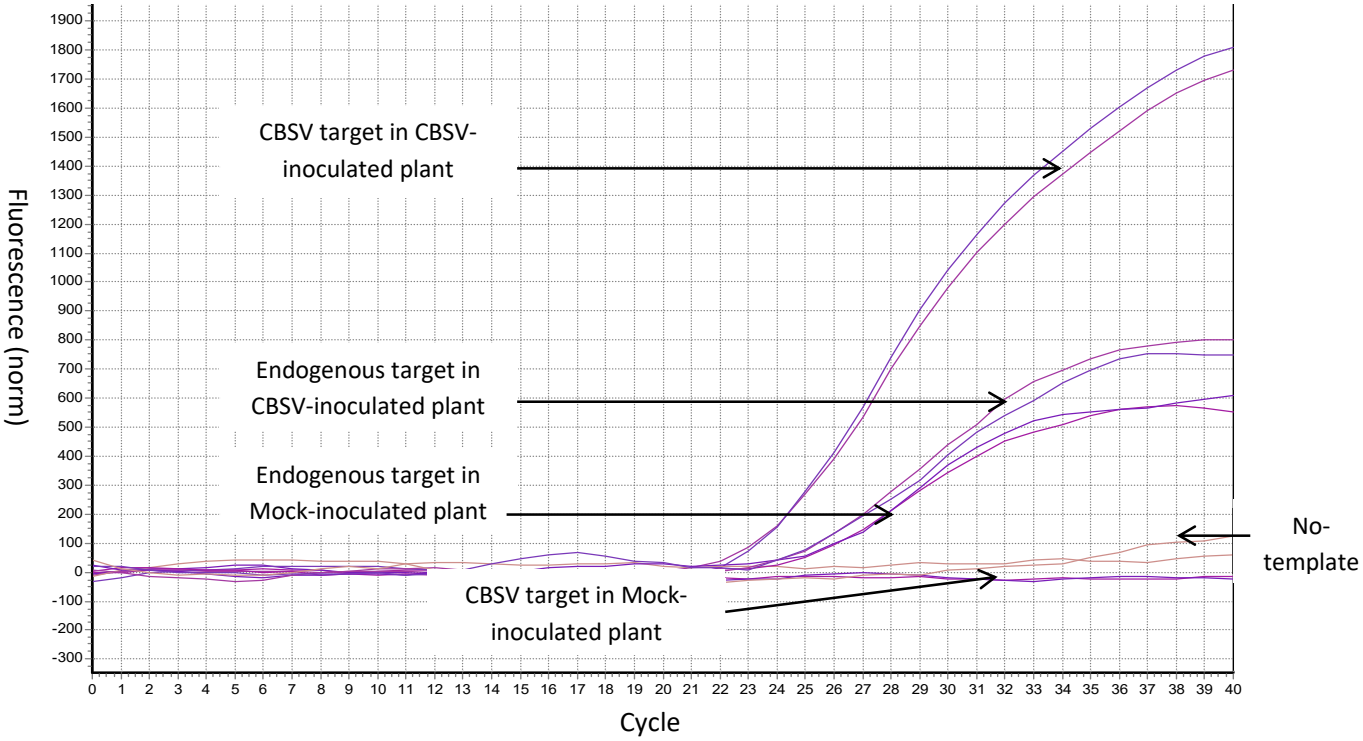


Figure 12. Typical normalized fluorescence against cycle amplification plot for relative quantification of CBSV and the endogenous PP2A gene in the same tube using two technical replicates of cDNA samples from CBSV-inoculated and mock-inoculated cassava plants (var. Mkumbozi) and a negative control sample consisting only of water.

Relative CBSV quantities were calculated following the delta-delta Cq method described by Livak and Schmittgen (2001). Using this method,

Relative quantity = $2^{-\Delta\Delta C_q}$, where $^{\Delta\Delta C_q}$ is difference between normalized target quantity in test sample and calibrator sample.

Mean Cq values per sample was normalized to Cq of the endogenous gene to eliminate effect on varying template quantity and intra-run variation in efficiency. Further normalization was done using endogenous gene-normalized Cq value of the calibrator sample (a 2 wpi Albert sample) and assayed in every run. This allowed the calculation of a relative quantity value for each sample comparable with every other sample within and between runs. Relative quantity data obtained in

this way up to the nine months post-inoculation for all varieties except for Nase 3 and Orera for which leaf samples could be collected for only three months because the grafted scion outgrew the stock plants over time hence leading to their death.

4.3.3 Time-course foliar CBSV quantities of cassava accessions and their Pair-wise comparison

Time-course average of foliar CBSV quantity was least for the known CBSD resistant Kaleso (Figure 13). Mean quantities in three other accessions – Mkumba, Pwani and Nase 3 were comparable to the quantity in Kaleso (Figure 13). Statistically, foliar CBSV quantities for Mkumba and Pwani, over the 9 mpi time, was not significantly different from time-course quantities of Kaleso although they were significantly different ($P > 0.05$) from quantities in the susceptible Albert (Table 14). While having comparable quantities with Kaleso, CBSV quantity in Nase 3 was significantly higher (4 time higher) than quantity in Kaleso (Figure 13; Table 14). Compared to Albert however, CBSV quantity in this accession was significantly lower than in Albert by a larger degree (10 times lower) (Figure 13; Table 14). On the other hand, time-course average CBSV quantity of Albert was the third highest, surpassed only by quantities in the accessions – Mbundumali and Kalawe. As expected, time-course CBSV quantities of Albert, Mbundumali and Kalawe were each significantly different from average quantity in Kaleso ($P \leq 0.05$) but quantitates in the latter two did not significantly differ from Albert ($P > 0.05$; Table 14). Time-course CBSV quantities recorded for the accessions Oekhumelela differed significantly from quantities in both Albert and Kaleso (Table 14).

Across time, relative CBSV quantities were consistently lower in Kaleso, Pwani, Mkumba and Nase 3 than other accessions (Figure 14A, 14B and 14C). These accessions showed similar pattern of time-course change in relative CBSV quantities which is sinusoid-like fluctuations in quantity at roughly same times – specifically between 12 and 16 wpi (Figure 14A and 14D). In contrast, there is apparently no uniformity in pattern for time-course variation of relative CBSV titres for other accessions. However, CBSV quantities in Albert, Kalawe, Mbundumali, Mkumbozi and F10-30-R2 were high from the first time point and did not fluctuate much across time (Figures 14B and 14D).

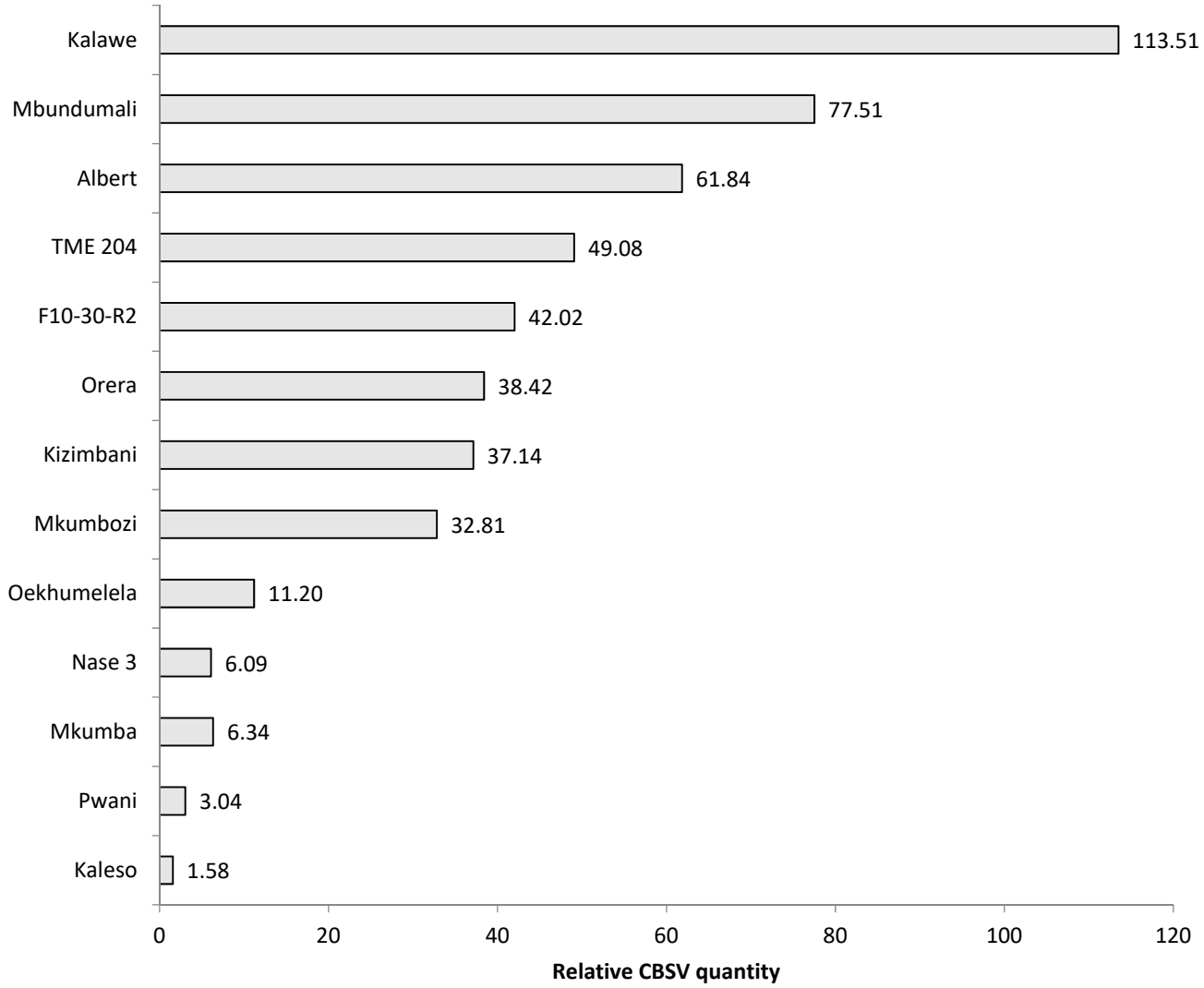


Figure 13. Bar plot of mean relative CBSV quantities determined over 3 and 9-month periods for 13 cassava accessions. Quantities were calculated relative to normalised quantiy CBSV quantity in RNA sample of two weeks old Albert. Mean CBSV quantity for Nase 3 was determined only over a period of 3 months because inoculated Nase 3 plants did not survive beyond this time.

Table 14. Pairwise comparison of each of the 13 cassava varieties with Albert and Kaleso showing ANOVA results for test of significant variety effect for each pair.

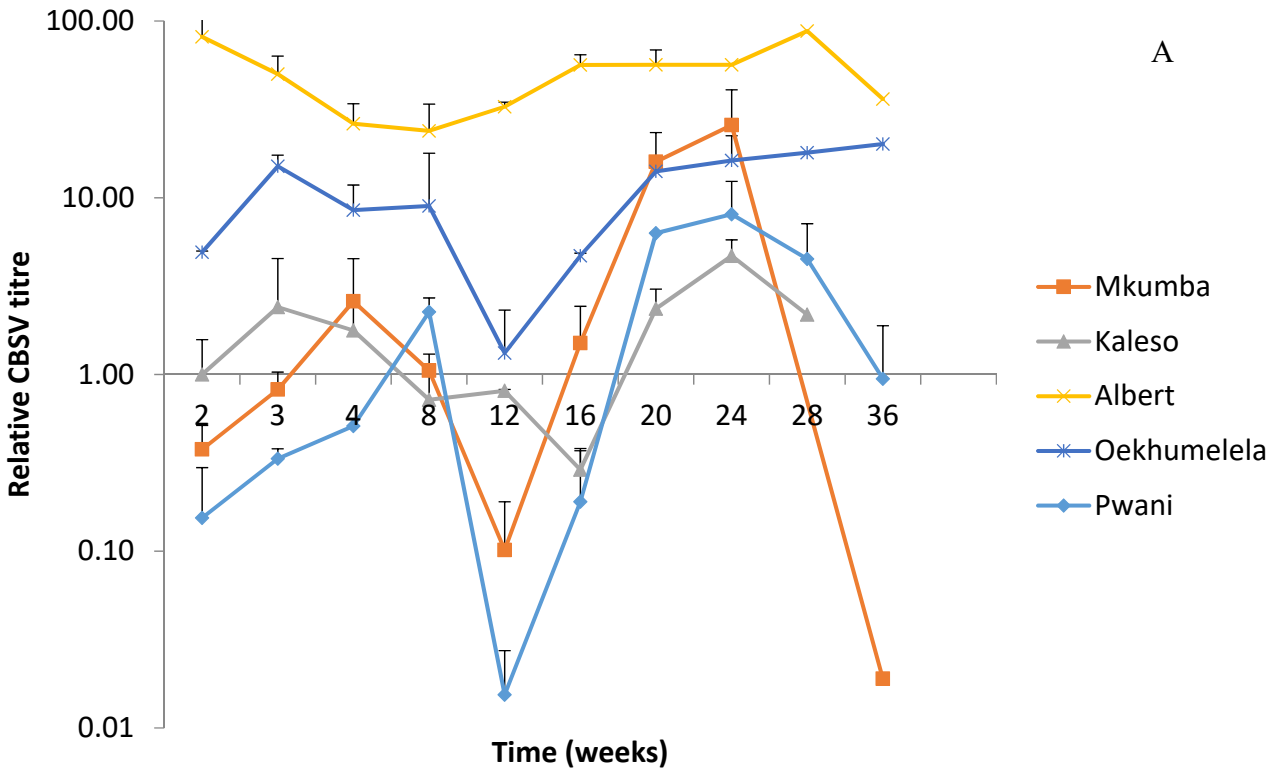
	Comparison with Albert				Comparison with Kaleso			
Variety	$Df_{numerator}$	$Df_{denominator}$ ([‡]KR approx.)	Fstat	P value	$Df_{numerator}$	$Df_{denominator}$ (KR approx.)	Fstat	P value
Kaleso	1	9.16	36.20	0.00018**	n/a	n/a	n/a	n/a
Kizimbani	1	8.92	33.60	0.0027**	1	9.98	9.60	0.0113*
Mkumba	1	8.47	49.20	8.4e-05**	1	8.96	1.46	0.26
Oekhumelela	1	8.40	34.40	0.00031**	1	8.96	19.43	0.0017**
TME 204	1	8.93	1.11	0.32	1	8.99	5.28	0.05*
Nase 3[†]	1	4.34	46.49	0.0018**	1	5.37	8.11	0.033*
Pwani	1	9.14	34.42	0.00023**	1	7.94	0.21	0.66
Kalawe	1	9.89	0.02	0.90	1	7.98	29.23	0.000065**
F10-30-R2	1	8.80	0.89	0.37	1	9.02	23.80	0.00087**
Orera [†]	1	3.99	2.84	0.17	1	3.88	23.66	0.0089*
Albert	n/a	n/a	n/a	n/a	1	6.74	97.60	3.0e-05**
Mkumbozi	1	9.44	0.72	0.42	1	8.21	13.35	0.0062*
Mbundumali	1	9.77	0.84	0.38	1	9.86	19.11	0.00144*

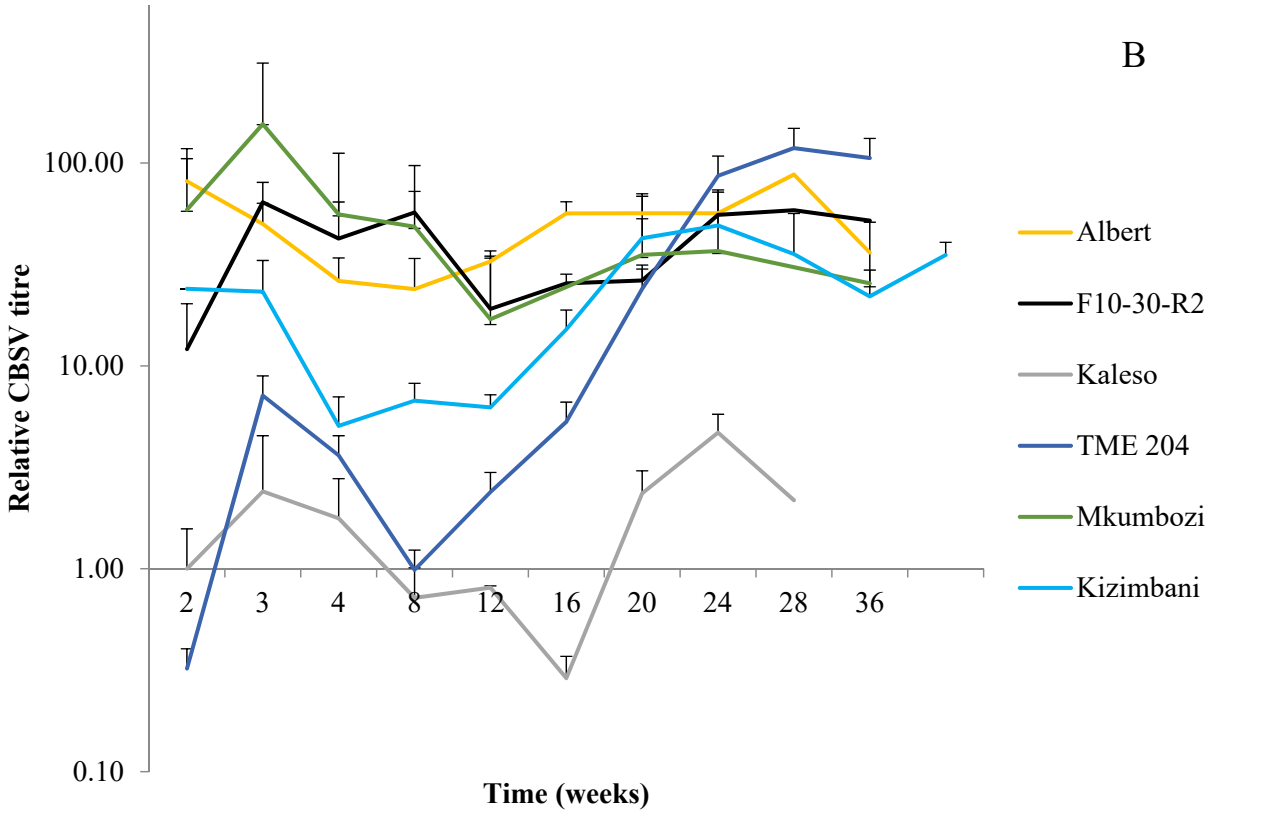
*Significant difference = $0.005<P\leq0.05$

**Highly significant difference = $P < 0.005$

[†]CBSV quantity data available only up to third month after inoculation

[‡]KR approx. stands for Kenward-Roger approximate


A
100.00
10.00
1.00
0.10
0.01
Relative CBSV titre
2
3
4
8
12
16
20
24
28
36
Time (weeks)
Mkumba
Kaleso
Albert
Oekhumelela
Pwani


B
100.00
10.00
1.00
0.10
Relative CBSV titre
2
3
4
8
12
16
20
24
28
36
Time (weeks)
Albert
F10-30-R2
Kaleso
TME 204
Mkumbozi
Kizimbani

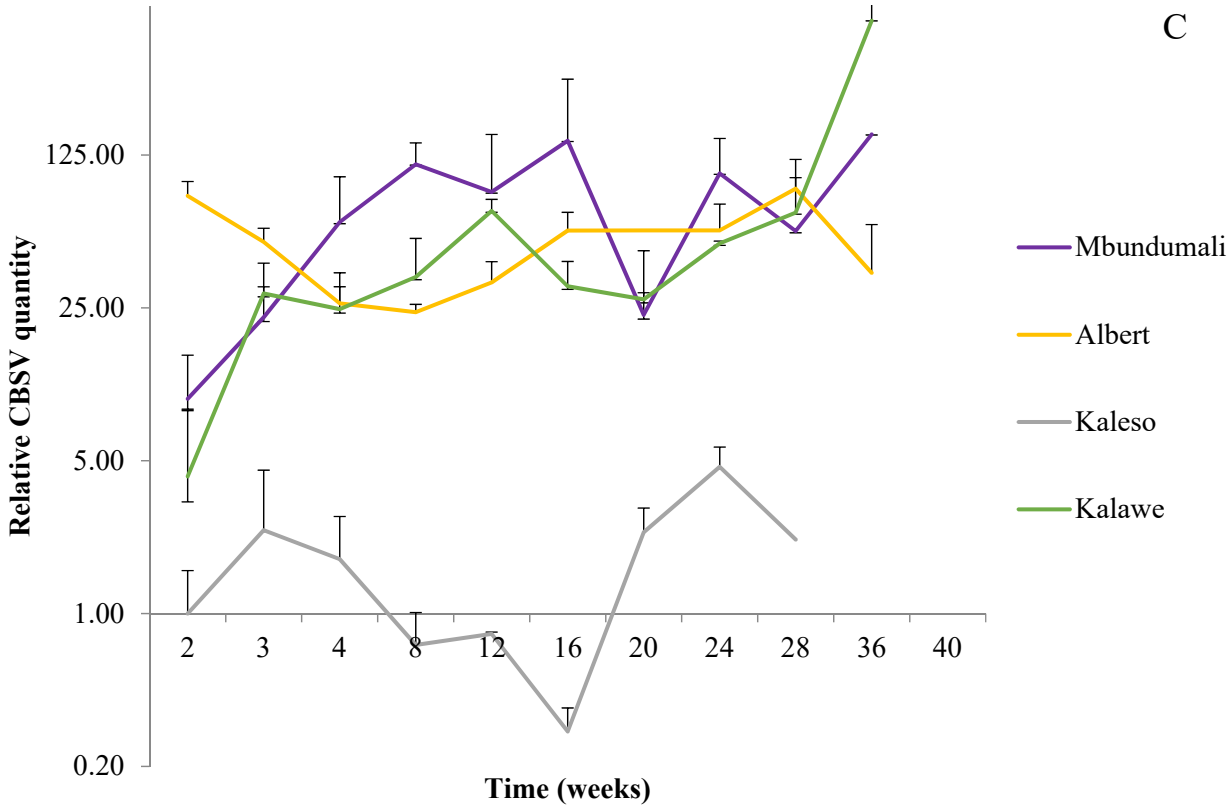


Figure 14. Time-course profile of relative CBSV titre over 2 to 36 weeks period in A. Mkumba, Pwani and Oekhumelela B. F10-30-R2, TME204, Mkumbozi and Kizimbani C. Mbundumali and Kalawe. CBSV quantity profile for Kaleso and Albert, used as CBSV resistant and susceptibility standards respectively, were included in all three constituent figures.

4.3.5 Foliar severity and virus quantity correlation

Foliar CBSD symptom severity and virus quantity were determined at weekly interval from 2 wpi to 4 wpi and subsequently monthly up to nine months post-inoculation. Simple linear correlation analysis showed a significant correlation $r = 0.52$, $P < 0.05$ between CBSD symptom severity and CBSV virus quantities measured in the 13 accessions across time. At the level of individual accessions, regression plots shows variation in relationship between CBSV quantity and severity. As Table 15 shows, significant correlation between CBSD severity and CBSV quantity was observed in the accessions Oekhumelela, Nase3 and TME 204.

Table 15. Correlation parameters from simple regression analysis of foliar CBSV quantity and CBSD symptom severity data of 13 cassava varieties.

Variety	*n*	*R*	R^2	*P value*
Oekhumelela	17	0.501	0.251	0.034*
Orera	10	0.089	0.008	0.788
Mbundumali	24	0.27	0.074	0.188
Kalawe	20	0.13	0.017	0.571
Albert	25	0.022	0.0005	0.917
Pwani	22	0.161	0.025	0.464
Mkumba	19	0.032	0.001	0.905
F10-30-R2	17	0.192	0.037	0.44
Nase 3	11	0.6728	0.4527	0.0165*
Kaleso	26	0.2214	0.0490	0.2671
Kizimbani	20	0.3647	0.133	0.1041
TME 204	27	0.6894	0.4753	4.95e-05*
Mkumbozi	17	0.0755	0.0057	0.766

*Significant difference at $P<0.05$

4.3.6 Severity and virus quantity in root tissue

Root tubers of cassava plants were harvested at 10 months post-inoculation and scored for symptom severity. CBSV quantities were also determined in samples of the root tissue. Relative quantities were calculated relative to the Cq values of Kaleso root sample. Mean CBSV quantities determined for root samples of 11 cassava varieties significantly correlates r = 0.62, $p < 0.05$ with mean of time-course foliar quantities of same varieties. Average root CBSD symptom severity for 10 varieties ranged from score 1 for absence of necrosis to score 5 for almost complete necrosis of roots (Figure 15). When regressed on log of mean root CBSV quantity, root CBSD severity showed no significant correlation $r = 0.57$, $P > 0.05$, unlike in in leaf tissue. However, distribution of mean root CBSV quantities and severities shows that both quantities correspond in most accessions (Figure 15). A notable exception was Mbundumali which in spite of very high root CBSV titre showed no root symptoms (Figures 15 and 16).

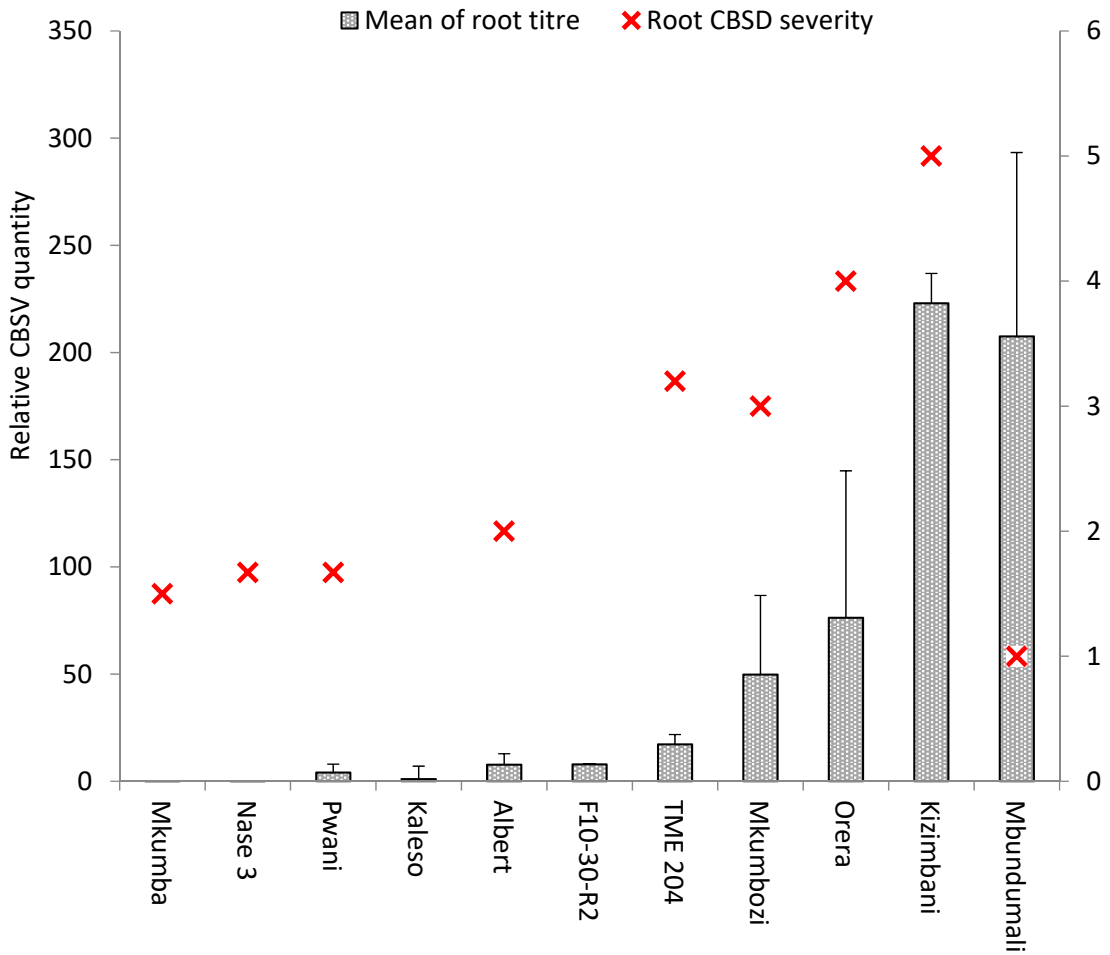


Figure 15. CBSD severity and relative CBSV quantity in root samples of 11 cassava accessions. There was no root tuber formation in the 4 varieties – Oekhumelela, Kaleso, F10-30-R2 and Kalawe at nine months post-inoculation. Virus quantity data for Kaleso, F10-30-R2 and Kalawe were obtained from fibrous tissue. Root CBSV quantity and symptom severity were not measured for var. Oekhumelela.

Figure 16. Root tuber samples from two plants of the cassava variety Mbundumali showing no necrotic symptom in spite of high root and foliar virus titre.

4.4 Discussion

In this study, 13 cassava varieties were graft-inoculated with CBSV and monitored for levels of accumulation of the virus. The graft inoculation process was very efficient with total graft success rate of 89.06% which is higher than 78.3% reported for CBSV inoculation of cassava by a chip grafting method (Wagaba *et al*., 2013). Just over 32.6% of virus-inoculated plants showed CBSD symptoms at two weeks following inoculation. However, RT-qPCR analysis revealed infection rate of 83% at this stage, meaning that CBSD symptom development lags CBSV accumulation in infected plants.

Before now, the real time PCR technique has only been applied to epidemiological field surveys of CBSD (Adams *et al*., 2013). Resistance screening has always been carried out using foliar and root severity scores of CBSD symptoms (Abeca *et al*., 2012; Theu and Mazuma, 2008). However, the real time PCR technique has recently been applied to the uniplex assay of CBSV and UCBSV quantities in cassava plants grown in the field (Kaweesi et al., 2014; Ogwok et al., 2014). Here, we have deployed a duplex multiplex qPCR method which co-amplifies CBSV and the endogenous control gene – PP2A in order to limit the effect of inter-well variation in template volume. The RT-qPCR method was very sensitive being able to detect CBSV in 19 symptomless plants which represent a 44.2% increase in diagnosis of CBSV infection using the duplex RT-qPCR method developed in the current study relative to visual symptom assessment. The method was also able to quantify CBSV to a limit of just over 400 copies. However, sensitivity was higher for the uniplex alternative which assays CBSV and the endogenous target in separate tubes (Table 13). This could be attributed to the interference of the primers and probes of both assays on each other when used in multiplex. Specificity of the method was demonstrated by the lack of CBSV amplification in mock-inoculated clones or virus-negative water controls (Figure 12). A TaqMan-based qPCR duplex assay has been previously developed for simultaneous amplification of the two cassava brown streak species – CBSV and UCBSV (Adams *et al*., 2013). However, high concentrations of CBSV generally inhibited the amplification of UCBSV in their experiments. Due to this masking effect of CBSV on UCBSV, the assay was utilized in simplex for epidemiological survey purposes and emphasized on detection rather than quantification of the viruses (Adams *et al.*, 2013).

Patterns of virus accumulation in foliar samples involved periods of rise and fall in virus quantities. This pattern appeared more uniform in Kaleso, Mkumba and Pwani with two periods of peak CBSV quantity, namely 3 to 8 wpi and 24 wpi, and a period of dip in virus quantity between 12 and 16 wpi (Figure 14A). Significant alteration in CBSV quantity over time,

especially when consistent across accessions having similar response to CBSV, could be indicative of changing internal host environment which fluctuates between transient periods of induced immunity response and counter-immunity response which promotes accumulation of the virus. Compared to other accessions, CBSV quantity remained relatively low over time in Kaleso, Pwani, and Mkumba with up to 37 times lower mean CBSV quantity compared to the susceptible Albert. The comparatively lower foliar time-course CBSV quantity, which does not significantly differ between pairs of these accessions, is suggestive of the existence of a host defence mechanism unique to the constituent varieties of the cluster. Comparison of Figures 13 and 15 showed that grouping of accessions based on root virus quantities was largely consistent with their grouping based on foliar virus quantities. This consistency together with the results from statistical comparison of CBSV accumulation in the accessions with accumulation in known CBSD resistant and susceptible accessions is useful evidence for determining appropriate CBSV response of the varieties. Given the comparable quantities of the virus in leaves and root tissue of Mkumba and Pwani to quantities in the known CBSD resistant Kaleso, these accessions can be considered CBSD resistant. Though having significantly higher CBSV quantities compared to Kaleso in its leaf tissue (Tale 14), quantity of the virus in Nase 3 is lesser by a greater degree than quantity in Albert. This fact and the very relatively low root quantity of the virus – which is comparable to quantity observed in Kaleso, suggests that Nase 3 is also CBSD resistant and could share similar host defence mechanisms as Kaleso though arguably to a lesser extent than Mkumba and Pwani descrbed above. Indeed, the accession have been reported to shown field resistance to CBSD (Abeca *et al*., 2012). Kaleso and Nase 3 are known to share a common ancestor in *Manihot glaziovii* (Bredesson *et al*., 2016). Given that Mkumba and Pwani are thought to be derived from Kaleso (Mkamillo G, personal communication), it follows that all four CBSD resistant accessions – Kaleso, Mkumba, Pwani and Nase 3 share a common progenitor. In Albert, CBSV accumulated to relatively high quantities at early stages of infection and either further increases or stabilizes till the end of the experiment (Figure 14). Four other accessions – Kalawe, F10-30-R2 and Mkumbozi shared this pattern and supported quantities of the virus which is significantly higher in Kaleso but not different from Albert (Table 14). These cultivars can therefore be classified as CBSD susceptible owing to their high root CBSV quantities (Figure 15) and the overall similarity of their CBSV replication dynamics to the susceptible standard – Albert. Oekhumelela supported significantly higher foliar and/or root CBSV titres than Kaleso but lower quantities compared to Albert (Table 14, Figures 13 and 15) hence was considered as tolerant.

The accession - Mbundumali which has the second highest root and foliar CBSV titres (Figures 13 and 15) was symptomless in the root. Though this an unexpected observation, this is not first time high CBSV quantities have been found in symptomless cassava roots. Kaweesi *et al*., 2014 had previously shown that the Ugandan variety Nase 1, infected with CBSV in the field, accumulated relatively high amounts of the virus in the roots which remained symptomless. There has been a degree of ambiguity in the literature as to the description of varieties not susceptible to CBSD. The term 'resistance' has been applied to a wide range of non-susceptible responses to CBSD. It has been used to describe complete absence of any CBSD symptom (Hillocks and Jennings, 2003), low incidence and low severity of foliar and root symptoms (Abeca *et al*., 2012, Maruthi *et al*., 2014b) and in some cases associated with mild root symptoms which precludes foliar and stem symptoms (Jennings 1960b). Tolerance on the other hand has been applied to describe response to CBSD involving full expression of foliar symptoms but late development of severe root necrosis, often after full maturity (Hillocks and Jennings, 2003). On the basis of this definition, the best performing accessions i.e. those belonging to the low CBSV titre group cannot be said to be tolerant since they were mostly symptomless or showed only mild foliar and root symptoms. Bruening (2006) described five plant viral disease resistance classes. CBSV accumulation dynamics observed in Kaleso, Mkumba, Pwani and Nase 3 are akin to one of these classes which is characterised by successful systemic infection but with reduced virus titre and attenuated symptom expression. The detection, for all accessions, of CBSV in the leaves and roots after inoculation in the stem, is an indication of systemic infection. However, CBSV quantities were significantly lower in the resistant compared to susceptible accessions.

Another class of resistance was defined as "systems in which virus-induced symptoms are greatly reduced in intensity or are absent, but the virus titre is unreduced or only slightly reduced relative to the reference infection" (Bruening, 2006). This definition is not fitting for the low CBSV titre group varieties as virus quantities of each member of this group was significantly lower compared to the reference infection of Albert (Table 15). However, it seems to apply to Mbundumali which accumulates relatively high levels of CBSV in shoot and root (Figures 13 and 15) but shows no root symptoms (Figure 16). This seemingly aberrant characteristic of the accession, distinct from the resistance response already shown in resistant Kaleso, Mkumba and Pwani, suggests that cultivated cassava evolved diverse strategies to protect itself from the destructive effects of CBSV infection and accumulation.

Relative CBSV quantities measured in leaf tissues within the 40 weeks period were significantly correlated $r = 0.52$, $P < 0.05$ with foliar symptom severity across 13 varieties. However, there was no significant correlation $r = 0.57$, $P > 0.05$ between CBSV quantity and severity in root tissue. This is in spite of a relatively good correlation coefficient (r = 0.57) evident from similarity in patterns of root CBSV quantity and severity scores across the varieties (Figure 15). The very low mean severity score recorded for var. Mbundumali which has the second highest root and foliar CBSV titres (Figures 13 and 15) may have contributed to this apparent contradiction. CBSD severity and CBSV quantity of this accession was determined from mean severity and quantity values in two clone plants with symptomless root tissue (Figure 16). The varieties - Oekhumelela, TME 204 and Nase 3 showed significant correlation between both quantities (Table 9). Cases of insignificant correlation or apparent inverse relationship of CBSV quantity with severity indicate the transient and often unpredictable nature of CBSD symptom expression (Hillocks and Jennings, 2003).

4.5 Conclusion

Characterization of varietal response to virus inoculation, determined from its quantities, identified a range of responses broadly classed into resistance and susceptible according to how their time-course virus quantities compare to the known CBSD-resistant and susceptible varieties. Besides Kaleso, the three varieties – Nase3, Mkumba and Pwani showed the best resistance to graft-inoculated CBSV in glasshouse conditions and could be useful in study of the genetic basis of CBSD resistance and broaden the genetic base for breeding resistance to the disease. It would be useful to also understand the molecular mechanism of the "CBSV-permissive" symptomless response observed in the cassava variety – Mbundumali, for potential development of an arguably 'sustainable' CBSD resistance that does not disturb balance of the cassava-CBSV pathosystem.

CHAPTER 5: Gene expression profiles of stress response and resistance genes distinguish resistance and susceptibility responses to cassava brown streak virus infection

5.1 Introduction

Deployment of resistant cassava is arguably the best means of controlling cassava brown streak disease (CBSD) given the high disease pressure in many parts of east Africa and the risk of inadvertent perpetual propagation and spread of the causative virus through vegetative propagation of the crop. The need to combine CBSD resistance with other farmer-preferred qualities of the crop has accelerated efforts at field and controlled laboratory screening of resistance to the disease in diverse local east African and exotic germplasm from Latin America. Increasing availability of CBSD-tolerant accessions from these efforts as well rapidly falling cost and increasing power of the next-generation sequencing technologies and computing facilities provide an impetus to go beyond resistance germplasm identification to understand the molecular processes underlying this resistance. The first step toward this goal involved the use of the next-generation RNA sequencing technique to study the transcriptome responses of the tolerant Kaleso and susceptible Albert germplasm to CBSV infection at one year after inoculation (Maruthi *et al*., 2014b). This study identified 13 NAC (acronym derived from the three proteins no apical meristem (NAM), ATAF1-2 and cup-shaped cotyledon (CUC2) that contain a similar DNA-binding domain)family genes as most overexpressed of 700 genes upregulated by *Cassava brown streak virus* (CBSV) infection in resistant Kaleso. However, most plant-pathogen interaction studies involving analysis of plant transcriptome response to pathogen are usually done at earlier times ranging from less than a day to a few months following pathogen exposure (Huang *et al*., 2015; Gao *et al*., 2016). The same strategy has also been used in studying cassava response to pathogen infection (Pierce and Rey, 2013; Allie *et al*., 2014). Analysing gene expression response to CBSV at earlier times would ensure identification of early response genes important for better understanding of molecular processes underlying resistance. Also, studying expression response in more cassava accessions over multiple times would allow generation of sufficient data for modelling regulation of important genes responding to CBSV infection. In the current study, the first multi-accession and multi-time analysis of cassava transcriptome response to CBSV infection is presented. The current study was designed in line with the points raised above – early and multi-time analysis of gene expression response to CBSV infection. It shows dynamics of gene expression and gene expression modulation in cassava and identifies transcriptional patterns associated with

resistance or susceptibility response to CBSV. A transcriptional model for resistance and susceptibility responses based on these patterns and related knowledge in literature was proposed.

5.2 Materials and Methods

5.2.1 Plant materials and growth conditions

Virus-free cassava accessions sourced from east Africa were used in this study. They have been previously (Chapter 3, section 3.3) classified into three groups based on foliar quantities of CBSV-[Mz:Nam:07] monitored over a nine-month period (Table 16). At least five cuttings per accession, were propagated on John Innes No. 2 compost soil in the quarantine glasshouse facility at NRI. Plant growth conditions were 28 ± 5ºC temperature and 50-60% humidity.

Table 16. Classification of the eight cassava varieties subjected to RNA sequencing based on CBSV accumulation levels (Chapter 4).

Cassava accession	Disease phenotype
Kaleso	Resistant
Mkumba	Resistant
Pwani	Resistant
Nase 3	Resistant
Oekhumelala	Tolerant
Kiorba	Tolerant
Albert	Susceptible
Kalawe	Susceptible

5.2.2 Graft inoculation, sampling and assessment of disease severity

About three two-month old plants were inoculated with CBSV-[Mz:Nam:07] by side grafting. The technique of Mohammed *et al*. (2012) is briefly described in chapter 3, section 3.2.2. At least two plants were mock-inoculated using similar technique but with a healthy scion. After grafting, symptom expression was monitored weekly but leaf sampling was done at one, five and eight weeks after CBSV inoculation. At each of these times, leaf samples were taken from virus- and mock-inoculated plants by clipping a single lobe each from leaves in the top, middle and lower parts of each plant. Leaf samples were immediately frozen in liquid nitrogen and processed immediately or stored short term in -80ºC freezer. Severity of CBSD symptom was monitored weekly using a scoring method, described by Hillocks *et al*. (1996); Rwegasira and Rey (2012) where severity ranges from 1 which is for no symptom expression to 5 for very severe chlorotic symptom expression covering almost the entire leaf area.

5.2.4. RNA extraction and RNA quality analysis

At each sampling time, equal amounts of freeze-dried leaf tissue from two plant clones of the same accession was pooled prior to RNA isolation. RNA was therefore extracted from a total of 48 leaf samples consisting of eight accessions each of which has CBSV- and mock-inoculated samples sampled at three time-points; 1, 5 and 8 weeks after virus inoculation. RNA isolation from these samples was done using an adapted protocol which combined a modified Cetyltrimethylammonium bromide (CTAB) method of Maruthi *et al*. (2002) and the QIAGEN kit-based extraction method according to the following procedure:

1. CTAB buffer was prepared from CTAB (2% w/v), 1.4 M NaCl, 20mM EDTA and 100mM Tris-HCl (pH 8.0). The CTAB buffer was autoclaved.
2. CTAB buffer was preheated for 10 minutes at 65ºC, after addition of 1.0% (v/v) 2-mercaptoethanol.
3. Approximately 100mg of liquid nitrogen-frozen cassava leaf tissue samples were ground using autoclaved porcelain mortar and pestle pre-baked dry before use
4. Without allowing the grinded tissue to thaw, 1 ml of CTAB, pre-heated at 60^0C, was added to it and mixed
5. Approximately 800 µl of sample mix was transferred to a 2 ml Eppendorf tube
6. An equal volume (800 µl) of phenol: chloroform: isoamyl alcohol (25:24:1) was added and the mixture centrifuged at 13,000 rpm for 10 minutes.

7. The top aqueous phase was transferred to a new 1.5 ml Eppendorf tube and an equal volume of absolute ethanol was added and mixture was gently pipetted up and down to mix

8. Up to 700 µl of the mixture was then transferred to an RNeasy® Mini spin column from QIAGEN's RNeasy® Plant Mini Kit

9. Samples were then centrifuged for 15 seconds at 13,000 rpm and flow-through was discarded

10. DNA was removed using QIAGEN's RNase-free DNase set according to an On-column DNase digestion procedure described in Part 2 of RNeasy® Mini Kit *Quick*-Start Protocol (https://www.qiagen.com/gb/resources/resourcedetail?id=f9b2e5ef-9456-431a-85ed-2a2b9fbd503dandlang=en).

11. Subsequent steps of the extraction protocol were carried out according to QIAGEN's RNeasy® Plant Mini Kit protocol, starting from step 4 of the manual (https://www.qiagen.com/gb/resources/download.aspx?id=14e7cf6e-521a-4cf7-8cbc-bf9f6fa33e24andlang=en).

RNA quantities were measured using the NanoDrop spectrophotometer (Thermoscientific, Wilmington USA). RNA integrity (RIN) values representing quality and integrity of extracted RNA templates were determined using Agilent 2100 Bioanalyzer instrument according to manufacturer's protocol (Agilent Technologies, CA USA).

5.2.5. Library preparation and sequencing

Messenger RNA (mRNA) library preparation was carried out using Illumina's TruSeq RNA library preparation kit (http://www.illumina.com/products/truseq_rna_library_prep_kit_v2.html) at The Genome Analysis Centre (TGAC), Norwich, UK. cDNA libraries obtained were indexed with Truseq index adapter barcode tags , checked for quality and then sequenced in multiplexed mixtures of 6 libraries per lane using Illumina's Hiseq 2500 next-generation sequencing system. Size-selected cDNA fragments were sequenced from both ends to obtain 100 bp paired-end reads.

5.2.6. Sequence read quality analysis

Raw reads obtained after sequencing were filtered by TGAC using the application SortMeRNA (SortMeRNA-1.9, http://bioinfo.lifl.fr/RNA/sortmerna/) to remove contaminating rRNA reads.

Filtered reads (recieved from TGAC) were subsequently checked for quality metrics such as per base sequence quality, read quality score distribution, sequence distribution per base position and per sequence GC content, using the FastQC tool (fastqc-0.11.2, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter sequences were removed using FASTA/Q trimmer tool of the FASTX-TOOLKIT collection (http://hannonlab.cshl.edu/fastx_toolkit/commandline.html).

5.2.7. Read mapping and assembly

Filtered high quality reads were mapped to the JGI *Manihot esculenta* v4.1 reference genome (http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mesculenta) using TopHat (tophat -2.1.1, http://ccb.jhu.edu/software/tophat/manual.shtml). Mapped reads were assembled into genes using the Linux-based tool generalized fold change (GFOLD) (GFOLD V1.1.0, http://compbio.tongji.edu.cn/~fengjx/GFOLD/gfold.html) following the cassava gene annotation information contained in Mesculenta_147_v4.1.gene.gff3.gz gene annotation file (http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=Mesculenta#). Expression values were calculated as reads per kilo base per million reads (RPKM) values.

5.2.8. Differential gene expression analysis

CBSV-induced differential gene expression was determined between CBSV- and mock-inoculated samples using the GFOLD tool. Differential expression was expressed as a GFOLD measure which is based on the posterior distribution of $\log_2$ fold change calculated from ratio of gene expression value (RPKM) in CBSV-inoculated sample to value in mock-inoculated sample of an accession (Feng *et al*., 2012). Genes having GFOLD values greater than 1 were considered differentially expressed while those with RPKM values less than -1 were considered repressed by CBSV infection. Unmodulated genes consistently expressed higher in each resistant sample compared to average expression in susceptible cassava were considered constitutively overexpressed in resistant cassava. A similar strategy was used to select genes constitutively overexpressed in susceptible cassava. Effects of sampling time, accession grouping and CBSV infection status i.e. mock- or CBSV-inoculated was determined using the R and the package edge (Storey *et al*., 2015). The effect of each of these three factors was determined using the optimized discovery procedure (Storey *et al.*, 2007) implemented in edge to analyse statistical significant difference between groups of gene expression values.

5.2.9. Functional annotation

Putative function was assigned for each gene modulated by CBSV infection (as determined from GFOLD analysis) by BLAST-searching their amino acid sequences against *Arabidopsis thaliana* genome on the STRING protein network interaction platform (http://string-db.org/cgi/input.pl?UserId=hqleh0EVi2CsandsessionId=trcicIVSNAK8andinput_page_show_search=on). The assigned putative functions were grouped into function categories listed in Appendix 8. These function categories were assigned based on protein function annotation at the STRING database. The net modulation of genes belonging to each functional category was assessed by calculating average of difference in total fold between induced and repressed genes under the category and normalized by the average difference between induced and repressed genes in the entire transcriptome. This showed whether there was net induction or repression of a functional category in a given transcriptome.

The one minus Pearson's correlation method was used for hierarchical clustering of gene expression and CBSV-induced gene expression fold change profiles of individual genes and treatment samples. The 48 sequenced cassava samples were grouped according to the three factors – type of inoculation (CBSV- or mock-inoculated), accession and post-inoculation sampling time. The maximum number of contiguous samples of a single group which partition into same cluster was used to measure the effect of each of above sample grouping factors.

5.3 Results

5.3.1 Graft inoculation, sampling and symptom evaluation

Graft union was established in all inoculated plants between two and three weeks post-inoculation (wpi). First foliar CBSD symptom was observed in the both susceptible accessions Albert and Kalawe at 3 wpi. The tolerant genotype accession Oekhumelela was unique in expressing severe stem symptoms but mild leaf symptoms. The stem symptoms were observed earlier, at 3 wpi while first foliar symptom was observed 4 wpi. Typical stem symptom on graft-inoculated Oekhumelela is shown in Figure 17. First symptom expression in the second other tolerant accession Kiroba was observed after 5 weeks while it took 8 wpi to develop first mild foliar symptoms in the resistant accessions Kaleso and Nase 3 (Table 17). There was no symptom expression in the two other resistant accessions Mkumba and Pwani throughout the duration of the experiment (Table 17). Severity and incidence scores of CBSD symptoms, at

RNA sampling times of 1, 5 and 8 wpi, for the eight CBSV-inoculated accessions subjected to RNA sequencing are as shown in Table 17.

Figure 17. Typical stem symptom on cassava plant of the Oekhumelela accession at 8 weeks after inoculation with CBSV. Red arrows indicate stem symptoms.

Table 17. Incidence and maximum foliar CBSD symptom severities, at three time points, for eight cassava accessions graft-inoculated with CBSV-[Mz:Nam:07]

Cassava accession	Incidence (number of plants showing symptoms/total number of plants)			*Symptom severity		
	1 wpi	5 wpi	8 wpi	1 wpi	5 wpi	8wpi
Kaleso	0/4	0/3	1/3	1	1	2
Mkumba	0/3	0/3	0/3	1	1	1
Pwani	0/3	0/3	0/3	1	1	1
Nase 3	0/3	0/3	1/3	1	1	2
Oekhumelela	0/2	2/2	1/2	1	2	3
Kiroba	0/6	1/6	0/6	1	2	1
Albert	0/4	3/3	3/3	1	3	4
Kalawe	0/3	2/3	2/3	1	4	5

*CBSD severity ranged from 1 for symptomless to 5 for very sever symptom expression (Hillocks *et al*, 1996; Rwegasira and Rey 2012)

5.3.2 RNA template quality

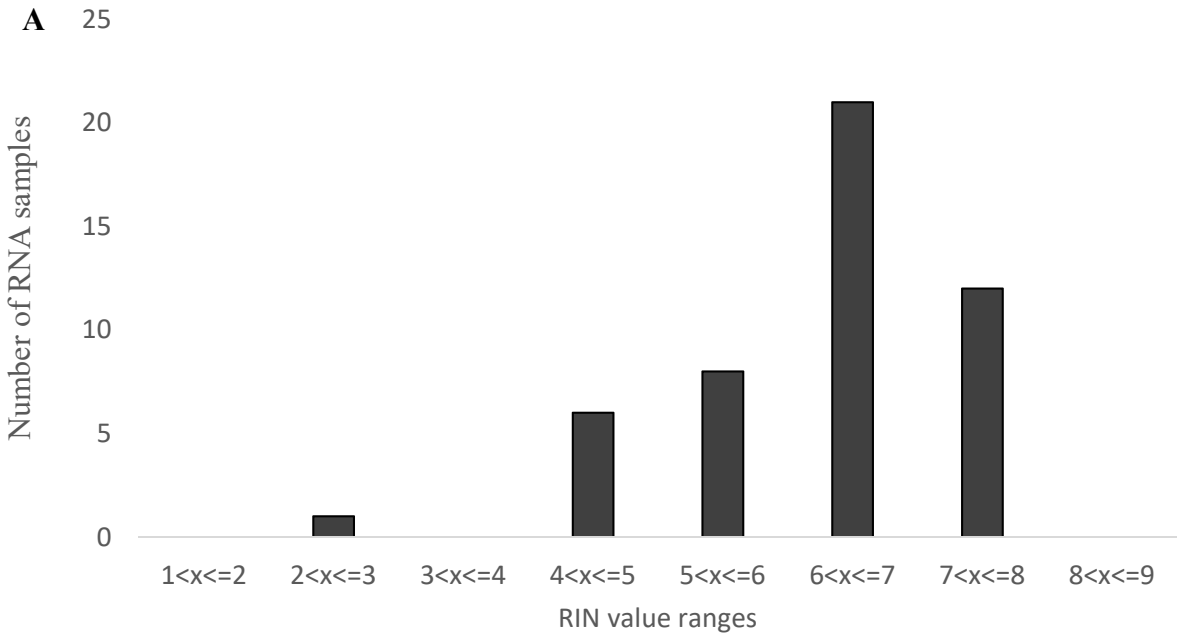
RNA extracts used for RNA-Seq library preparation were generally of good quality. Average value for RNA integrity (RIN) values of the 48 samples sequenced was 6.25. Distribution of these values shows that RIN values between 6 and 8 were the most prevalent with more than 68% of samples having values within this range (Figure 18A). While RNA samples of lower integrity were sequenced alongside these, the range of template qualities does not appear to influence levels of the entire or part of the transcriptome as there was no specific pattern for median and standard deviation of expressed genes (RPKM > 1) of sequenced samples arranged in an increasing order of their RIN values (Figure18B and 18C).RNA template quality also does not seem to influence extent of unique read mapping given that RIN value correlates poorly with proportion of uniquely mapped reads (Figure 18D).

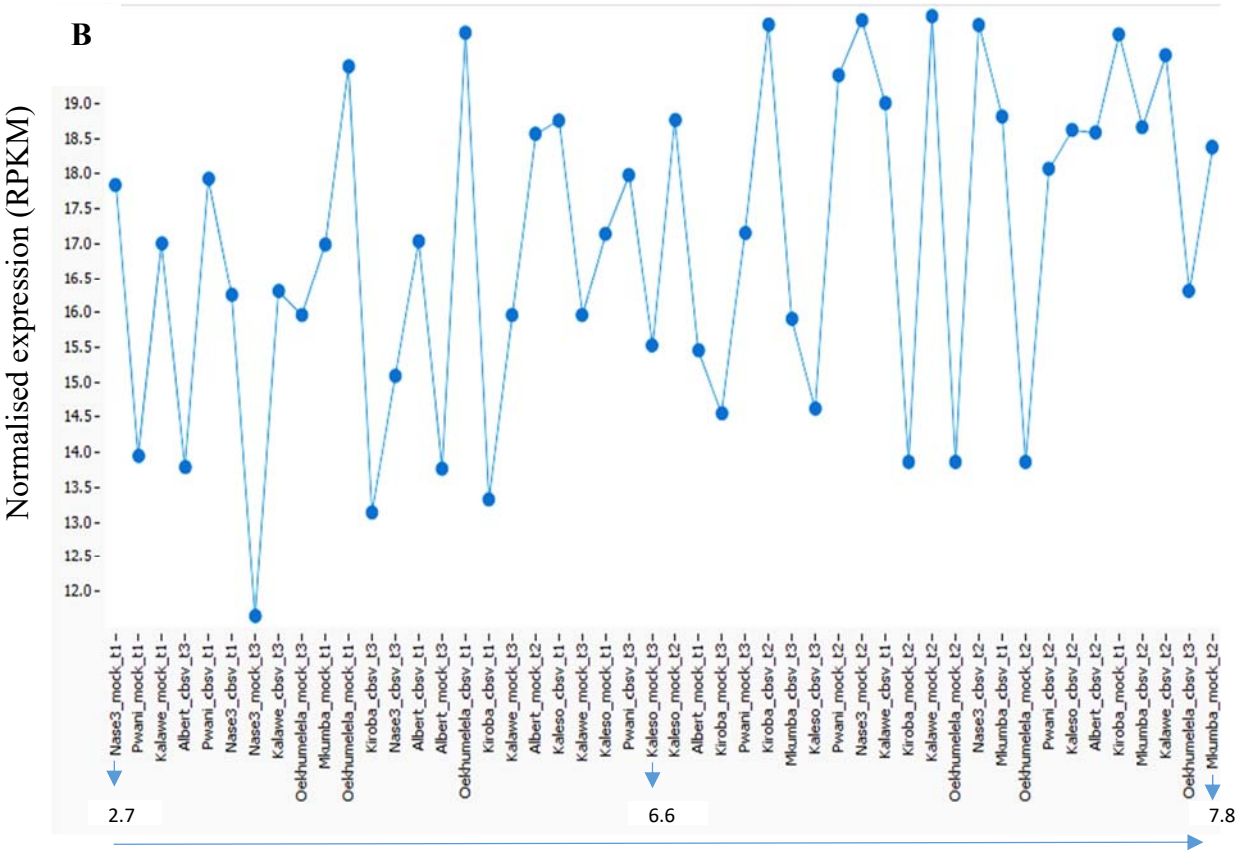
5.3.2 RNA sequence read quality control

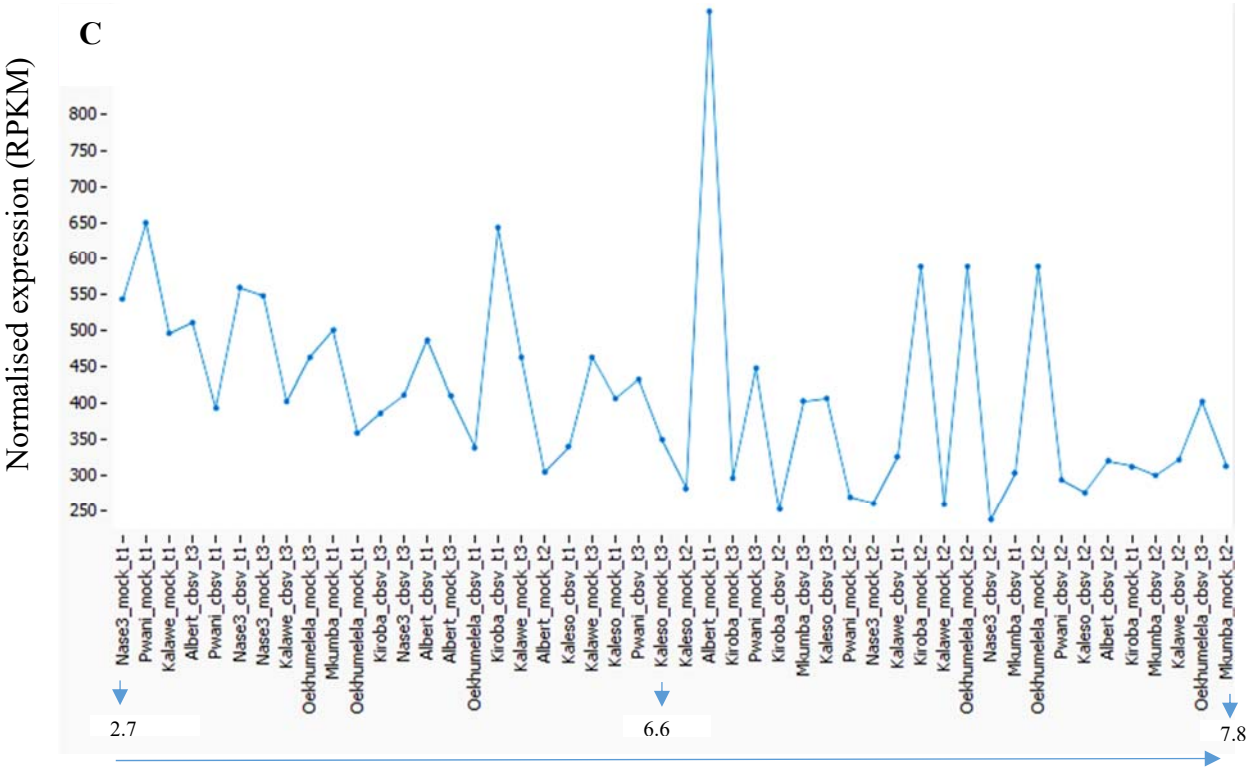
Raw 100 base-pair sequence reads were trimmed at the 5'- end to remove adapter sequences. All sequence reads used for downstream analysis at least passed the most important quality parameters. For instance, average read Phred and per base quality scores were at least 36 and 25, respectively. Percentage proportion of bases per nucleotide position of read ranged from 20 to 30% while the most frequent GC content of reads ranged from 40% to 50%. Quality assessment result from FastQC for typical sequence reads accepted for further analysis is represented in Figure 19.

5.3.3 Read mapping and assembly

Majority of reads for each sample were mapped to the reference cassava genome *Manihot esculenta* v4.1. Specifically, percentage of aligned reads ranged from 73.9% to over 88% with at least 80% alignment rate in 35 of the 48 samples. Gene expression values were obtained for a total of 34,151 cassava genes in RPKM values.


A
25
20
15
10
5
0
Number of RNA samples
1<x<=2
2<x<=3
3<x<=4
4<x<=5
5<x<=6
6<x<=7
7<x<=8
8<x<=9
RIN value ranges


B
Normalised expression (RPKM)
19.0
18.5
18.0
17.5
17.0
16.5
16.0
15.5
15.0
14.5
14.0
13.5
13.0
12.5
12.0
Nase3_mock_t1
Pwani_mock_t1
Kalawe_mock_t1
Albert_cbsv_t3
Pwani_cbsv_t1
Nase3_cbsv_t1
Nase3_mock_t3
Kalawe_cbsv_t3
Oekhumelela_mock_t3
Mkumba_mock_t1
Oekhumelela_mock_t1
Kiroba_cbsv_t3
Nase3_cbsv_t3
Albert_cbsv_t1
Albert_mock_t3
Oekhumelela_cbsv_t1
Kiroba_cbsv_t1
Kalawe_mock_t3
Albert_mock_t2
Kaleso_cbsv_t1
Kalawe_mock_t3
Kaleso_mock_t1
Pwani_cbsv_t3
Kaleso_mock_t3
Kaleso_mock_t2
Albert_mock_t1
Kiroba_mock_t3
Pwani_mock_t3
Kiroba_cbsv_t2
Mkumba_cbsv_t3
Kaleso_cbsv_t3
Pwani_mock_t2
Nase3_mock_t2
Kalawe_cbsv_t1
Kiroba_mock_t2
Kalawe_mock_t2
Oekhumelela_mock_t2
Nase3_cbsv_t2
Mkumba_cbsv_t1
Oekhumelela_mock_t2
Pwani_cbsv_t2
Kaleso_cbsv_t2
Albert_cbsv_t2
Kiroba_mock_t1
Mkumba_cbsv_t2
Kalawe_cbsv_t2
Oekhumelela_cbsv_t3
Mkumba_mock_t2
2.7
6.6
7.8


C
Normalised expression (RPKM)
800
750
700
650
600
550
500
450
400
350
300
250
Nase3_mock_t1
Pwani_mock_t1
Kalawe_mock_t1
Albert_cbsv_t3
Pwani_cbsv_t1
Nase3_cbsv_t1
Nase3_mock_t3
Kalawe_cbsv_t3
Oekhumelela_mock_t3
Mkumba_mock_t1
Oekhumelela_mock_t1
Kiroba_cbsv_t3
Nase3_cbsv_t3
Albert_cbsv_t1
Albert_mock_t3
Oekhumelela_cbsv_t1
Kiroba_cbsv_t1
Kalawe_mock_t3
Albert_mock_t2
Kaleso_cbsv_t1
Kalawe_mock_t3
Kaleso_mock_t1
Pwani_cbsv_t3
Kaleso_mock_t3
Kaleso_mock_t2
Albert_mock_t1
Kiroba_mock_t3
Pwani_mock_t3
Kiroba_cbsv_t2
Mkumba_cbsv_t3
Kaleso_cbsv_t3
Pwani_mock_t2
Nase3_mock_t2
Kalawe_cbsv_t1
Kiroba_mock_t2
Kalawe_mock_t2
Oekhumelela_mock_t2
Nase3_cbsv_t2
Mkumba_cbsv_t1
Oekhumelela_mock_t2
Pwani_cbsv_t2
Kaleso_cbsv_t2
Albert_cbsv_t2
Kiroba_mock_t1
Mkumba_cbsv_t2
Kalawe_cbsv_t2
Oekhumelela_cbsv_t3
Mkumba_mock_t2
2.7
6.6
7.8

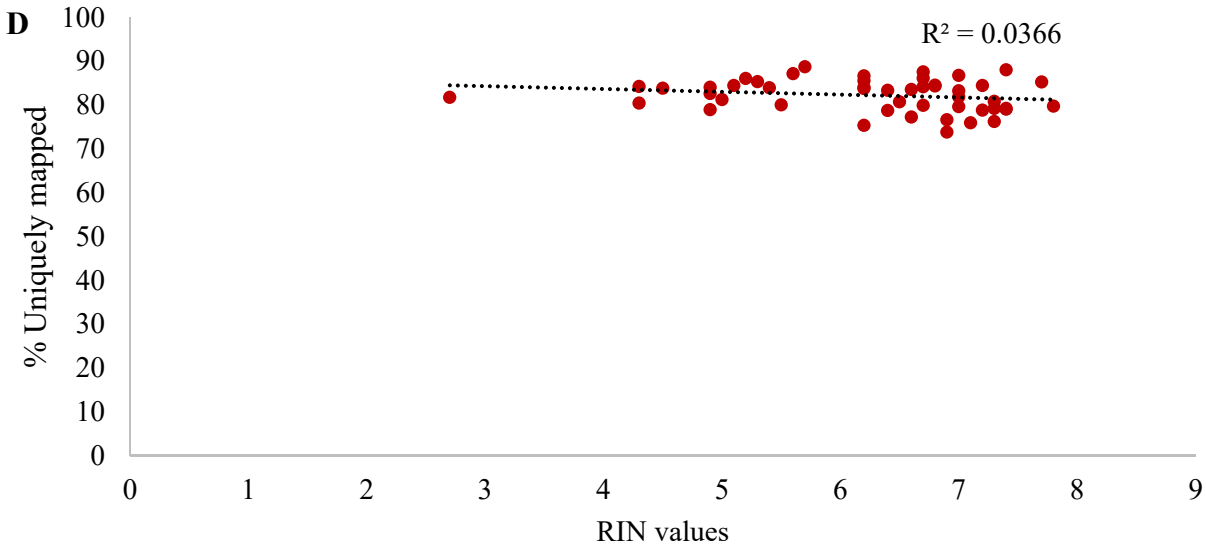


Figure 18. Distribution of A. RIN values of 48 cassava RNA samples subjected to RNA-Seq library preparation B. Median of normalised expression values of 19082 genes (RPKM >1) of 48 sequenced samples arranged in an increasing order of RIN values (left to right) and C. Standard deviation of RPKM values of these genes within same samples arranged similarly. D. Linear regression plot showing extent of correlation between percentage of uniquely mapped reads and RIN values.

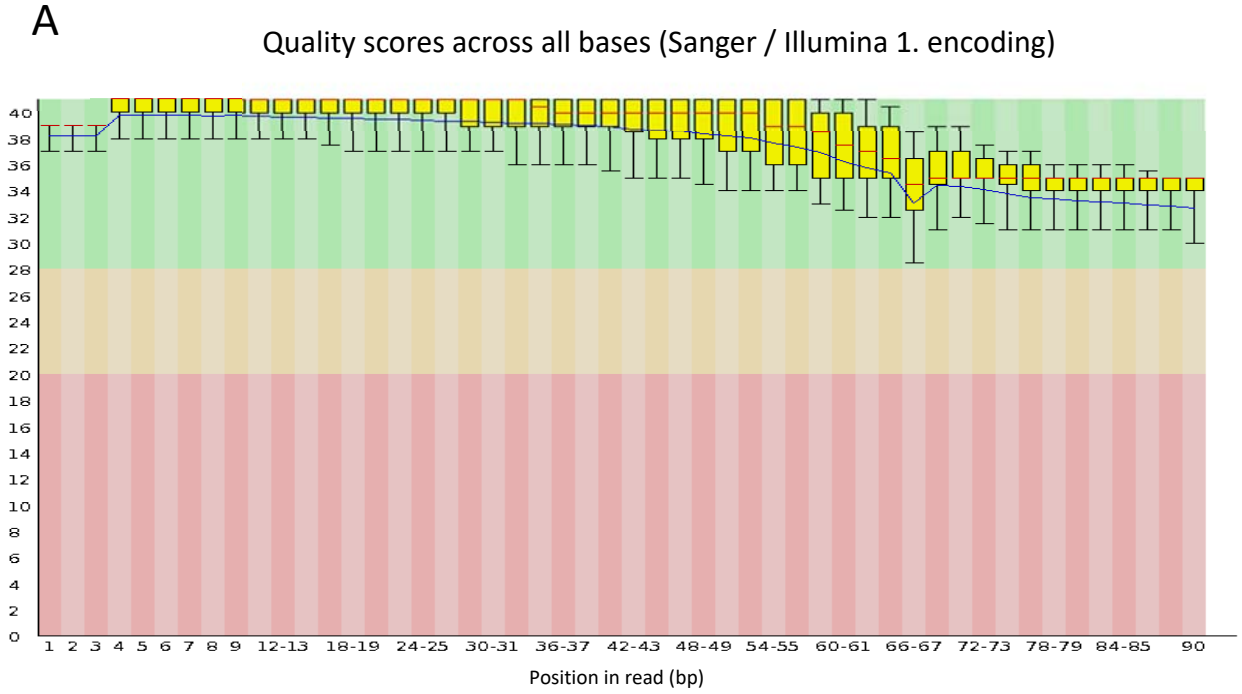
5.3.4 Differential gene expression analysis

Comparison of gene expression levels between virus- and mock- inoculated cassava samples, based on GFOLD analysis, showed that a total of 8971 genes were differentially expressed as a result of CBSV infection across the eight cassava accessions. Statistical analysis using the Edge package (http://bioconductor.org/packages/release/bioc/html/edge.html) showed that sampling time influenced transcriptome expression more than the category of response to CBSV infection – resistant, tolerant or susceptible. More than with 13,058 out of 19,082 genes, expressed at greater than 1 RPKM values were significantly differentially expressed (FDR < 0.05) over the three post-inoculation sampling times (Appendix 1). In contrast, a total of 4023 genes were significantly differentially expressed between paired combinations of resistant, tolerant and susceptible accessions. Specifically, 2044 genes were significantly differentially expressed (FDR < 0.05, Benjamini-Hochberg) between tolerant and resistant accessions (Appendix 2) while 1,976 genes (Appendix 3) were differentially expressed between susceptible and resistant accessions (FDR < 0.05, Benjamini-Hochberg). Only 203 genes were

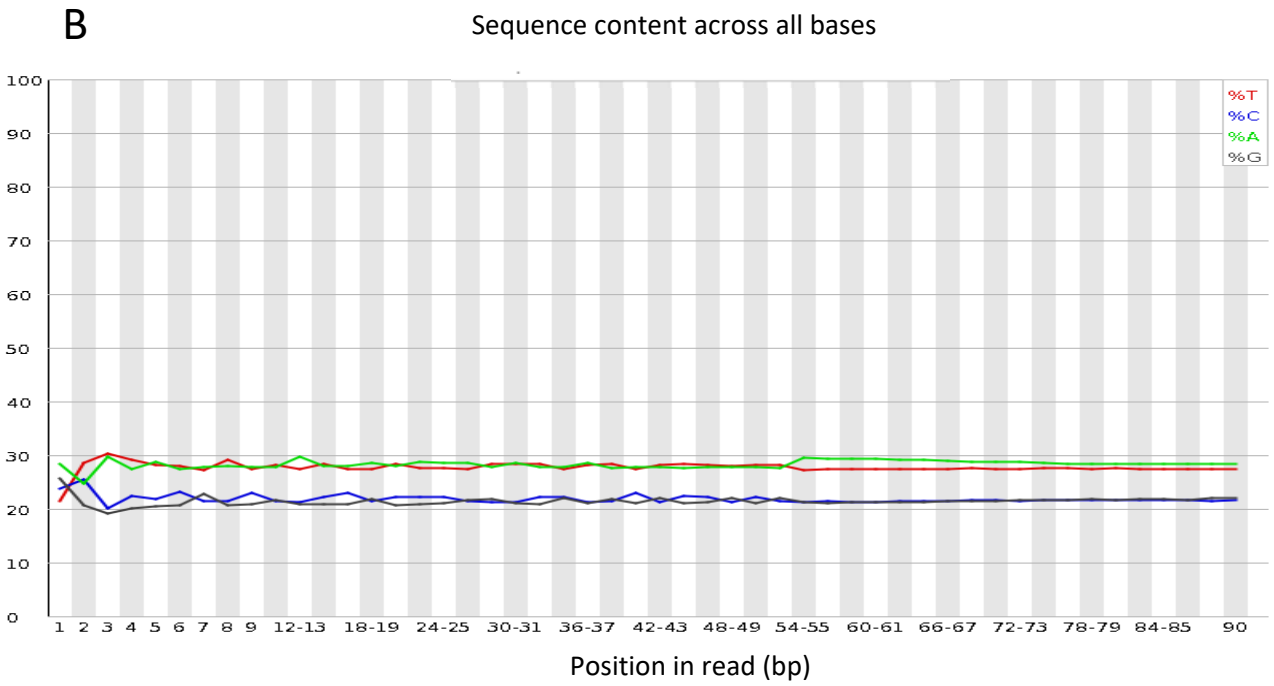
significantly differed in their expression levels in tolerant and susceptible accessions under the same FDR cut-off (Appendix 4).

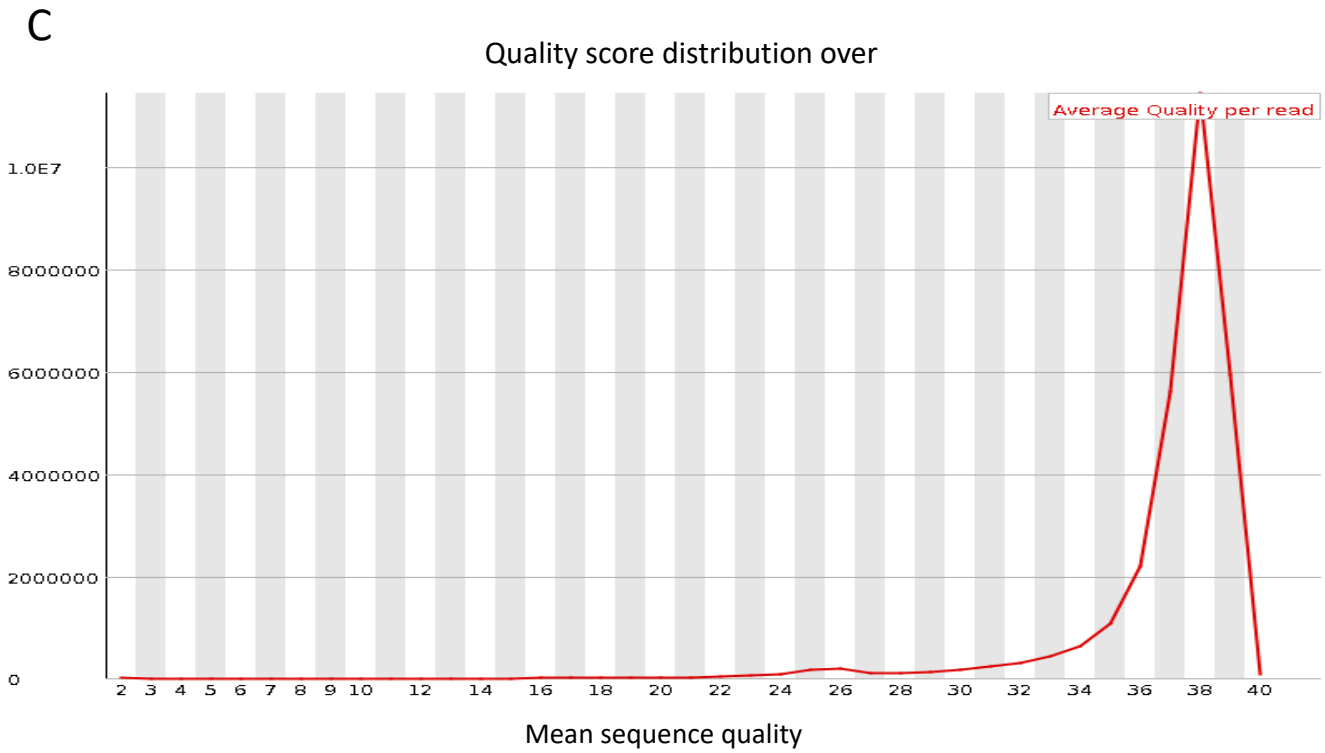
Further evidence on the extent to which post-inoculation sampling time, CBSV treatment and infection response category – resistant, tolerant or susceptible – influence global gene expression was sought using hierarchical cluster analysis (Figure 20). In line with the Edge analysis result described above, hierarchical clustering showed that post-inoculation sampling time had the more influence on transcriptome expression than infection response category of the accession. Effects of the three variables in cluster analysis was based on maximum fraction of samples, classified under each of the variables, are contiguous in a cluster. For example, 15 samples (or 93.7%) of the 16 samples collected at the first sampling time (t1) or 1 week post-inoculation, were contiguous within the same cluster in the hierarchical cluster plot of Figure 20.

A


Quality scores across all bases (Sanger / Illumina 1. encoding)
40
38
36
34
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 12-13 18-19 24-25 30-31 36-37 42-43 48-49 54-55 60-61 66-67 72-73 78-79 84-85 90
Position in read (bp)

B


Sequence content across all bases
100
90
80
70
60
50
40
30
20
10
0
%T
%C
%A
%G
1 2 3 4 5 6 7 8 9 12-13 18-19 24-25 30-31 36-37 42-43 48-49 54-55 60-61 66-67 72-73 78-79 84-85 90
Position in read (bp)



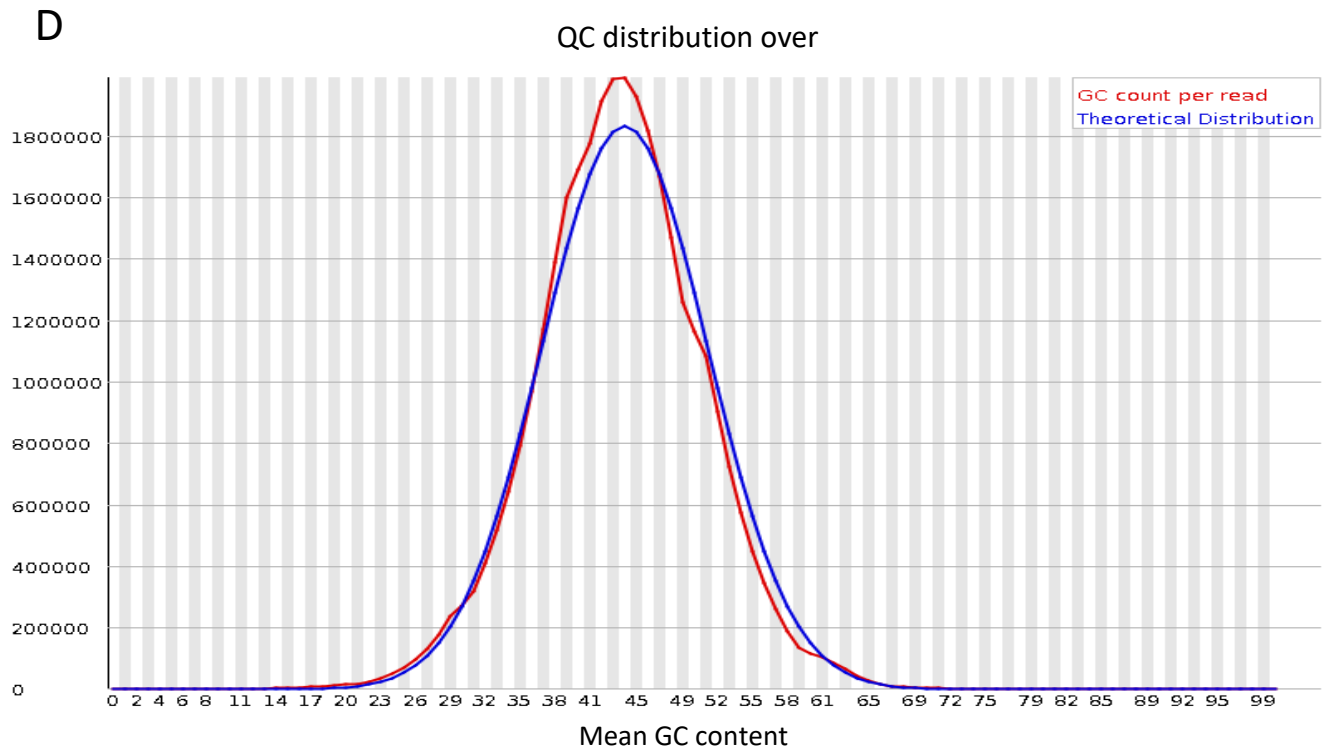


Figure 19. Typical FastQC results for A. per base sequence quality based on PHRED scores B. distribution of quality score for all reads C. Sequence content across base positions and D. GC distribution over all sequences.

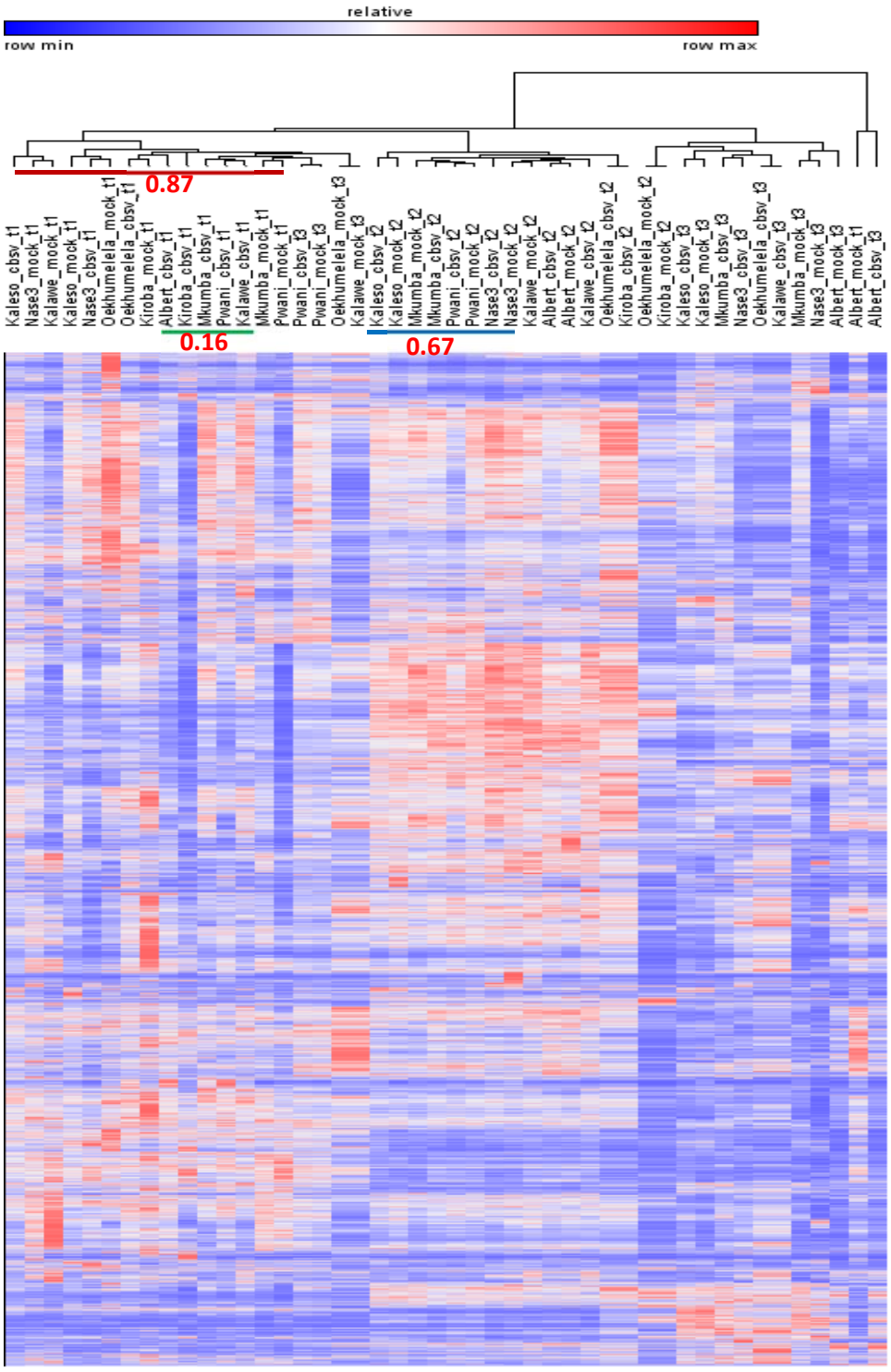


Figure 20. Heat map of 19,082 gene expression levels in 48 cassava transcriptomes clustered by sample identity and by genes. The red, blue and green horizontal bars delineate the longest contiguous sample set under sampling time, infection response category and CBSV treatment-based accession groupings. The figures above the bars represent the respective fraction of samples within accession group contiguous within a cluster.

Transcript levels of 8971 genes were altered (GFOLD >1 or < 1) by CBSV infection in the 24 cassava RNA samples sequenced compared to the 24 mock-inoculated plants. The number and extent or range of gene expression modulation varied with time and accession. Number of genes differentially expressed between CBSV and mock inoculated plants varied from 720 to 6379 genes across accessions at 1 wpi. Average number of gene modulated by CBSV inoculation across accessions were highest at first and eight weeks after inoculation (Figure 21). Number of genes modulated in response to CBSV infection was highest in accessions of the tolerant group but least in resistant accessions (Figure 22). Also, while more gene repression than induction response was observed for all resistant accessions, there was no preferential inductive or repressive response in either of the tolerant or susceptible groups.

To identify candidate CBSV resistance genes, the set of 8971 modulated genes was searched for genes consistently induced or repressed in all four accessions of the resistant group in each sampling time but never modulated in the susceptible group at any time. Conversely genes uniquely modulated in the susceptible accessions were designated candidate CBSV susceptibility genes. Based on the above criteria, none of the uniquely induced or repressed genes of the resistant group was shared by all resistant accessions (Figures 23A and 23B). However, there was more similarity within each pair of resistant accessions as maximum number of unique genes shared was 20 for induced genes and (Figure 23A) 45 for repressed genes (Figure 23B). Similar to observation in resistant accessions, there was not a single uniquely repressed gene shared amongst all four non-resistant accessions (tolerant and susceptible accessions), though three uniquely induced genes were shared among these accessions. Greater similarity in uniquely induced and repressed genes was however observed within susceptible accessions and within tolerant accessions (Figures 23C and 23D). Apart from gene expression modulation associated with resistant and susceptibility responses, unmodulated genes constitutively expressed higher in resistant than susceptible accessions or vice-versa (overexpressed genes) were also identified. Average expression fold difference was more variable for genes overexpressed in susceptible compared to those overexpressed in resistant cassava. It ranged were 1.1 to over 19 fold in susceptible and 1.3 to 3.3 fold in resistant cassava (Appendices 5 and 6). Thirty six genes were overexpressed in resistant accessions relative to average expression values of samples of both susceptible accessions – Albert and Kalawe (Figure 24B; Appendix 5). On the other hand, 296 genes were overexpressed in susceptible accessions relative to the resistant ones (Figure 24; Appendix 6).

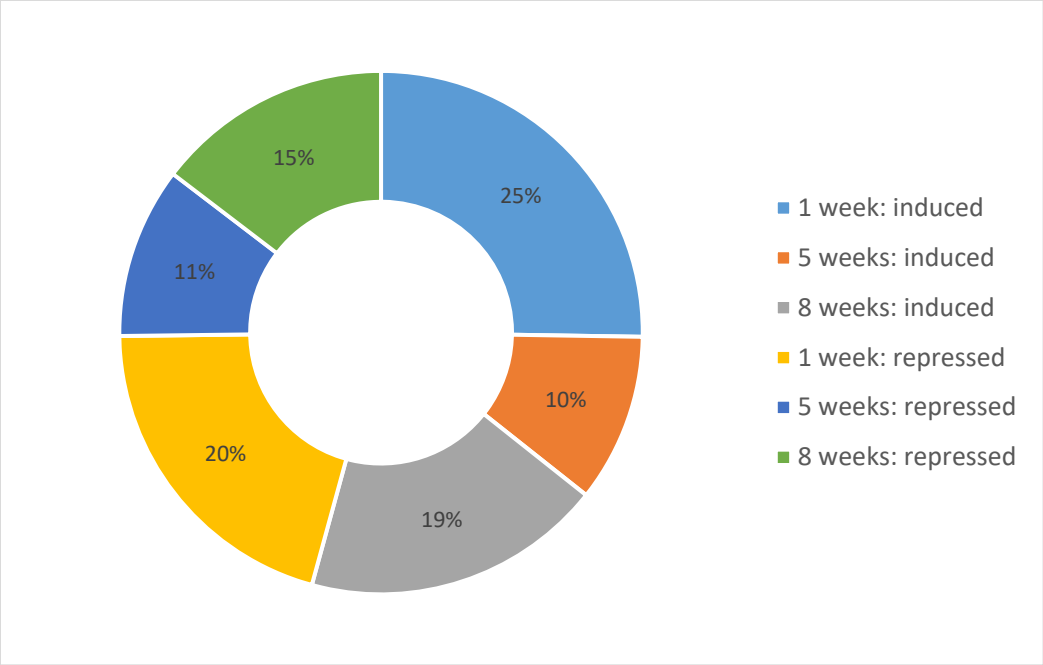


Figure 21. Proportions of all modulated genes induced and repressed by CBSV infection at the three times post-inoculation.

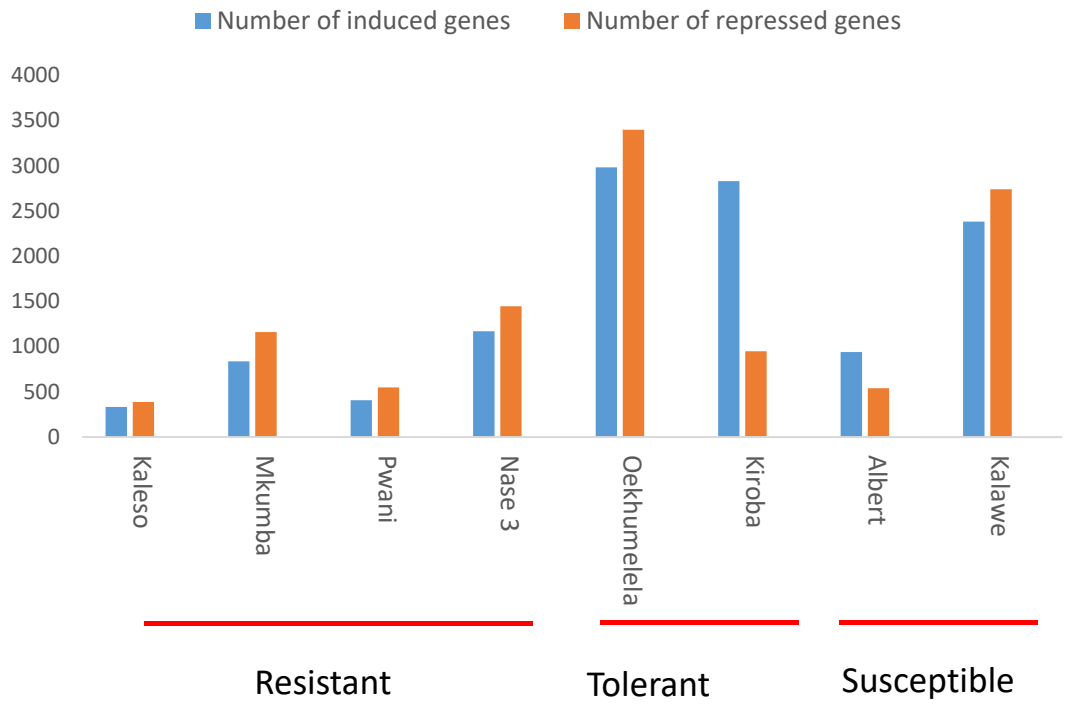


Figure 22. Gene numbers induced or repressed, as a result of CBSV inoculation, over the three RNA sampling times (1, 5 and 8 wpi).

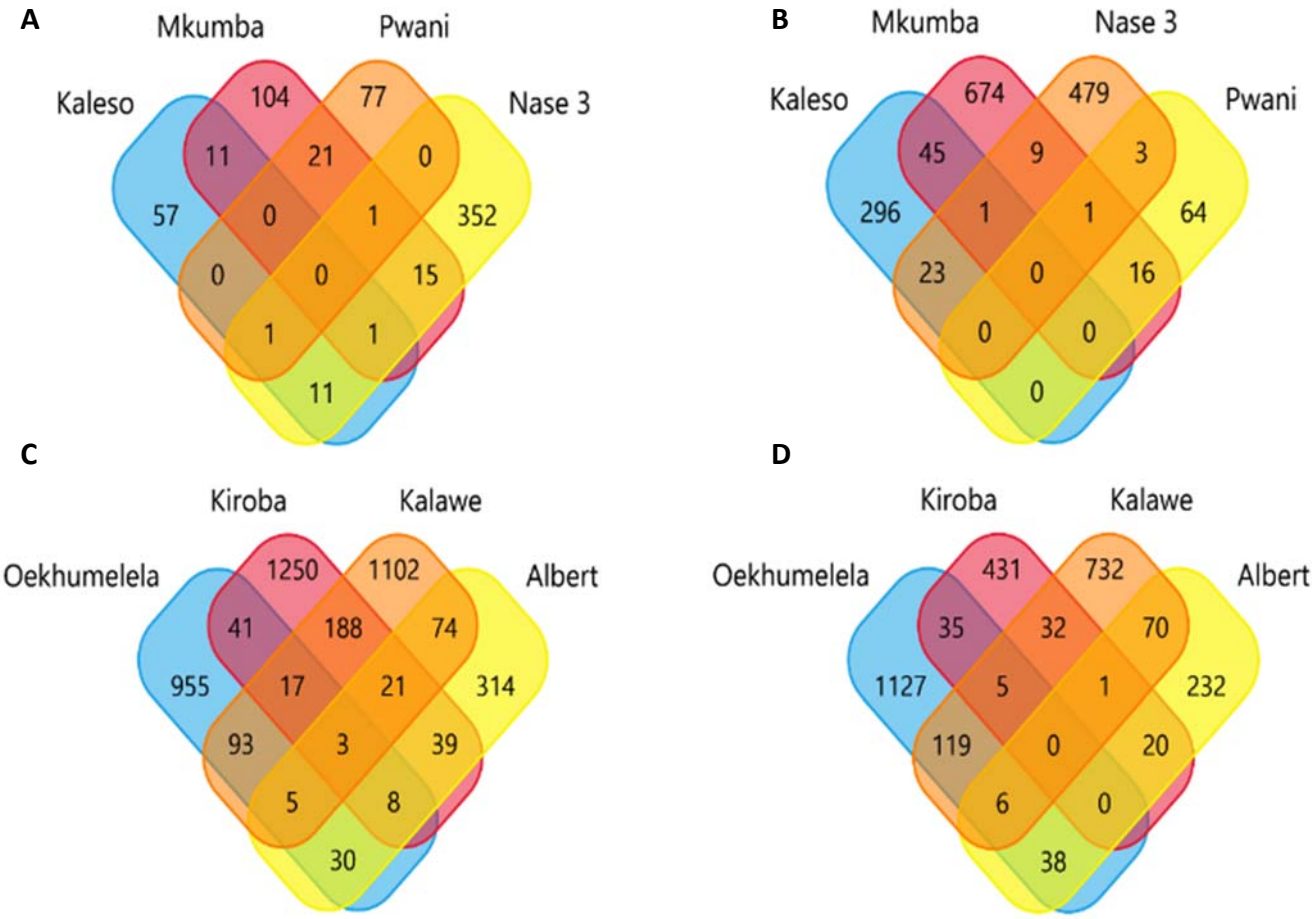
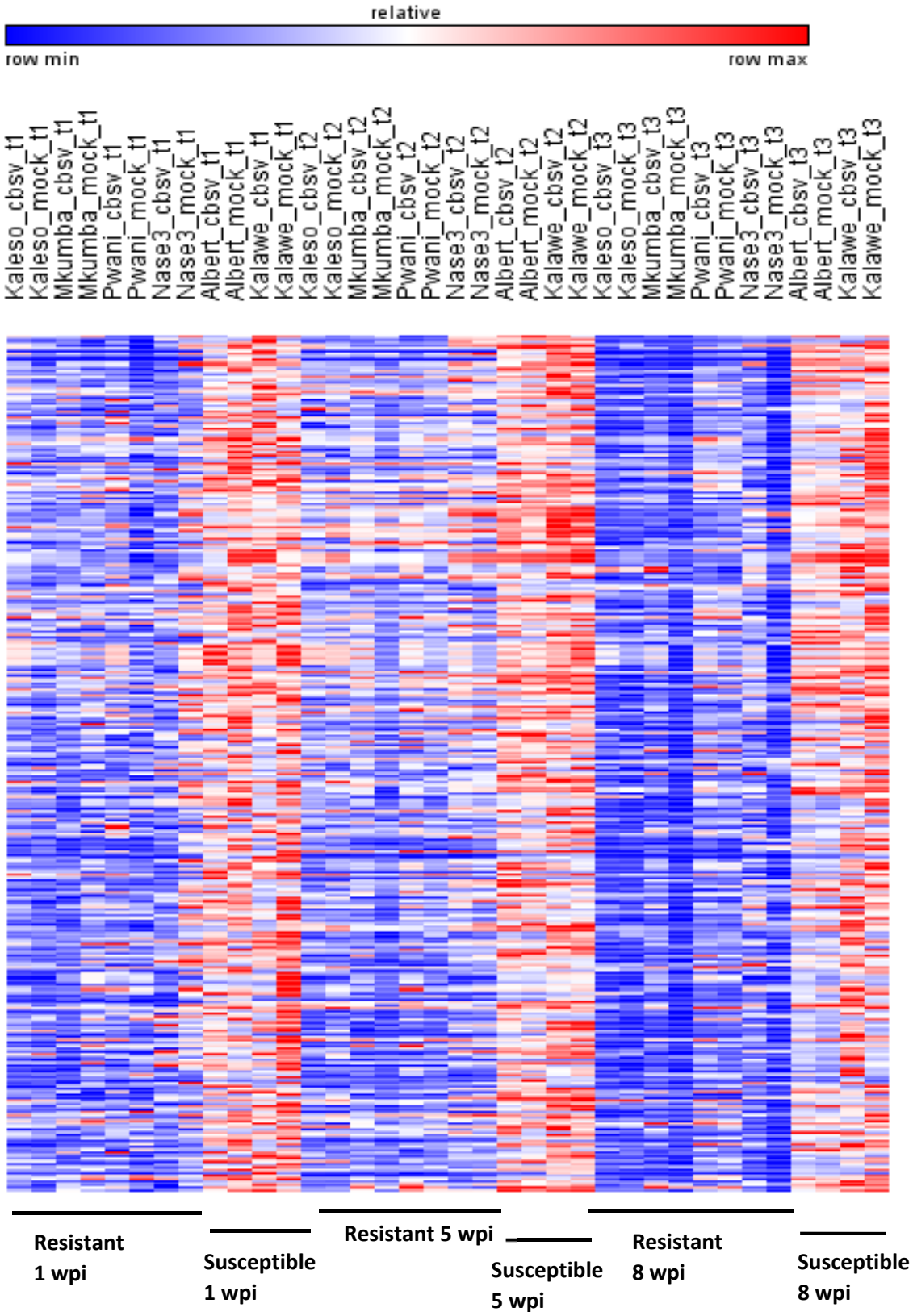


Figure 23. Venn plots of genes uniquely A. induced in resistant accessions B. repressed in resistant accessions C. induced in non-resistant tolerant and susceptible accessions and D. repressed non-resistant tolerant and susceptible accessions.

A


relative
row min
row max
Kaleso_cbsv_t1
Kaleso_mock_t1
Mkumba_cbsv_t1
Mkumba_mock_t1
Pwani_cbsv_t1
Pwani_mock_t1
Nase3_cbsv_t1
Nase3_mock_t1
Albert_cbsv_t1
Albert_mock_t1
Kalawe_cbsv_t1
Kalawe_mock_t1
Kaleso_cbsv_t2
Kaleso_mock_t2
Mkumba_cbsv_t2
Mkumba_mock_t2
Pwani_cbsv_t2
Pwani_mock_t2
Nase3_cbsv_t2
Nase3_mock_t2
Albert_cbsv_t2
Albert_mock_t2
Kalawe_cbsv_t2
Kalawe_mock_t2
Kaleso_cbsv_t3
Kaleso_mock_t3
Mkumba_cbsv_t3
Mkumba_mock_t3
Pwani_cbsv_t3
Pwani_mock_t3
Nase3_cbsv_t3
Nase3_mock_t3
Albert_cbsv_t3
Albert_mock_t3
Kalawe_cbsv_t3
Kalawe_mock_t3
Resistant 1 wpi
Susceptible 1 wpi
Resistant 5 wpi
Susceptible 5 wpi
Resistant 8 wpi
Susceptible 8 wpi

B

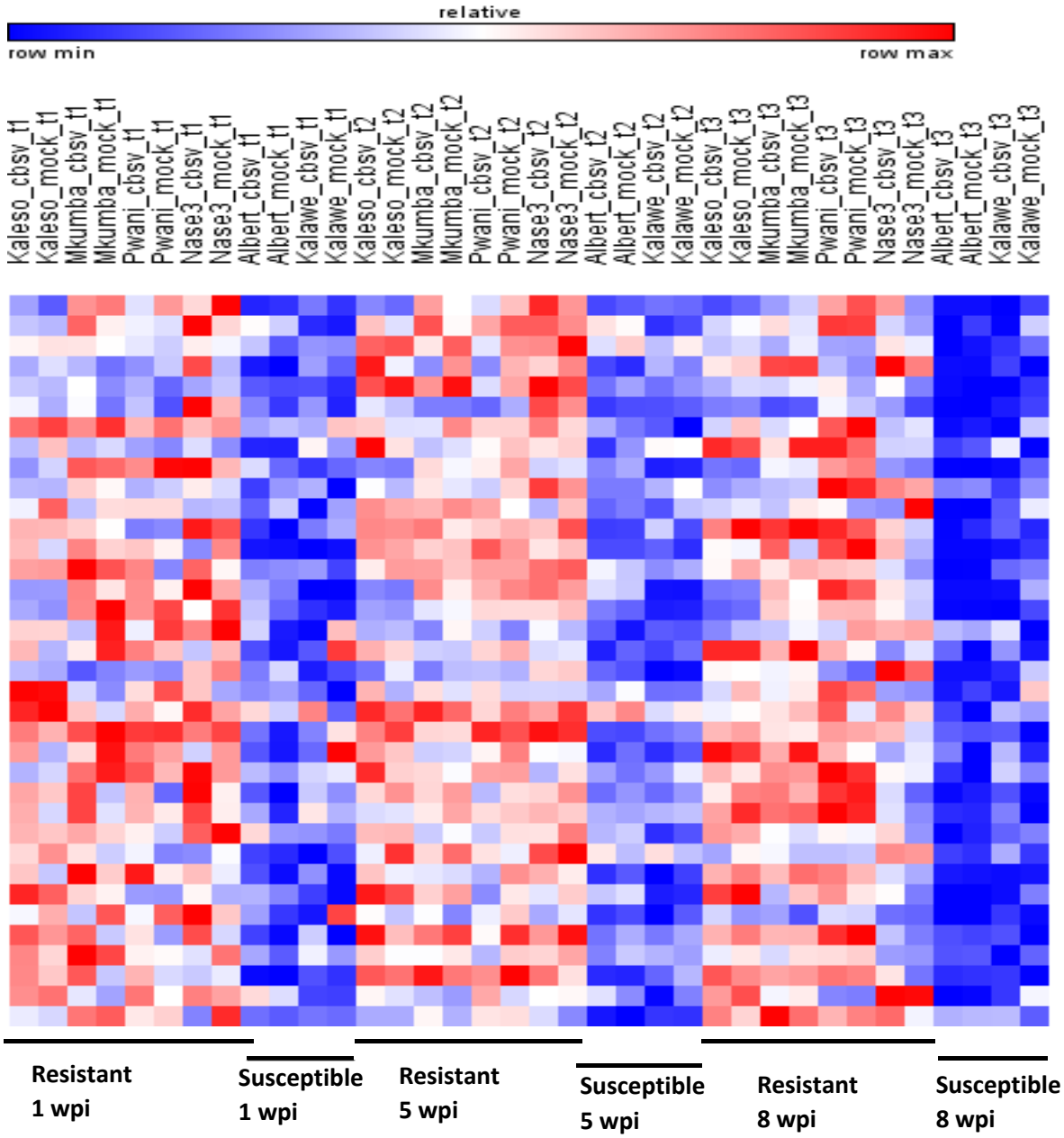


Figure 24. Figure 24. Heat map of RPKM expression values of genes expressed higher in A. each susceptible accession relative to their time-course average expression in resistant accessions B. each resistant variety relative to their time-course average expression in susceptible cassava varieties.

5.3.5 Functional analysis of gene expression responses to CBSV infection and their differential expression between resistant and susceptible cassava.

Putative molecular and physiological roles were predicted for 6967 out of the 8971 differentially expressed genes. The remaining 2004 genes (approximately 25% of differentially expressed genes) either had no *Arabidopsis thaliana* ortholog or are currently unannotated. Of the predicted gene set, 3851 were further grouped into 43 different functional classes (Appendix 8) based on their functional annotation.

CBSV-induced modulation of these functional categories was determined from net sum of expression fold change values between CBSV- and mock-inoculated samples of each accession at each sampling time. Distributions of these values in resistant, tolerant and susceptible accessions show overall unique patterns of modulation for genes with abiotic stress response, antioxidant defence, cell wall loosening / cell expansion and pathogenesis-related functions in resistant and susceptible cassava over the three sampling times (Figures 25, 26, 27 and 28). There was a net higher induction of the abiotic stress response function in non-resistant tolerant and susceptible accession groups compared to the resistant group (Figure 25). Average of net expression fold change of abiotic stress response genes across non-resistant cultivars was just over 10 times higher than the value for resistant accessions. On the other hand, genes coding for antioxidant defence, cell wall loosening / cell expansion and pathogenesis-related functions were oppositely regulated. CBSV infection caused net induction of these biological function groups in susceptible accessions but had an overall repression effect on same functions in the resistant cultivars (Figure 27). This pattern of modulation of biological function categories was observed as an overall effect over resistant or non-resistant samples of all three time-points. It was also largely obtainable at each time except for the first week post-inoculation when all three functions were repressed in susceptible accessions (Figure 26). Though repressed in all accessions, auxin response genes were only induced at the last post-inoculation time in susceptible cassava (Figure 26). Overexpressed genes of the resistant accession group were most enriched with Nucleotide binding site-Leucine rich repeat (NBS-LRR) resistance (R) gene function – specifically five out of 18 annotated genes (Appendix 6). These genes were expressed two to three times higher in resistant compared to susceptible cassava (Appendix 6). A LESION SIMULATING DISEASE one like 2 (LOL1; cassava4.1_020424m) gene was also overexpressed in resistant accessions relative to susceptible ones. The gene was expressed up to 2.1 times higher in resistant compared to average expression in susceptible cassava (Appendix 5).

Figure 25. Bar plot of net CBSV-induced expression fold values for gene functional classes averaged across resistant accession groups at one (T1), five (T2) and eight (T3) wpi.

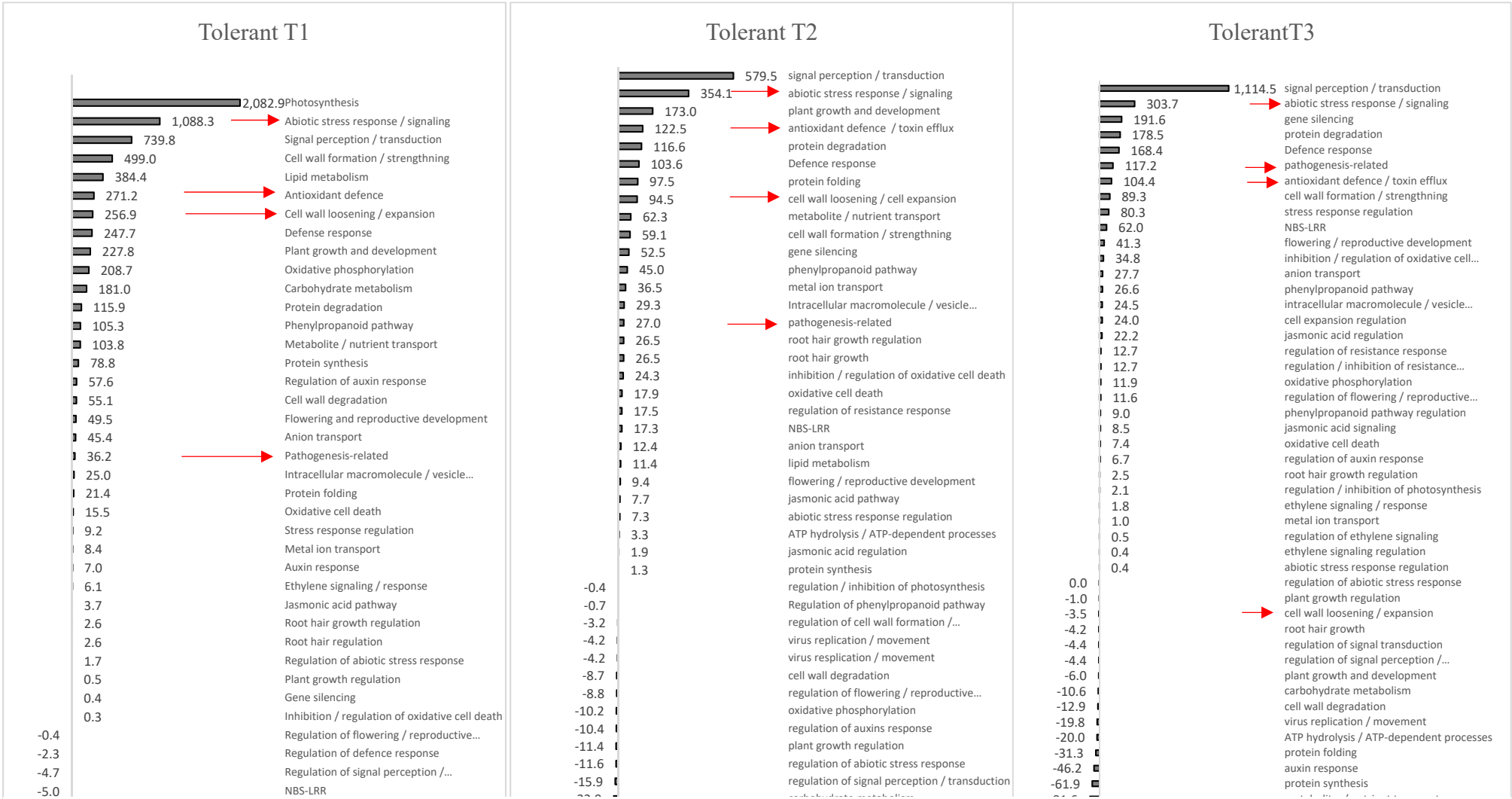


Figure 26. Bar plotof net CBSV-induced expression fold values for gene functional classes averaged across tolerant accessions at one (T1), five (T2) and eight (T3) wpi.

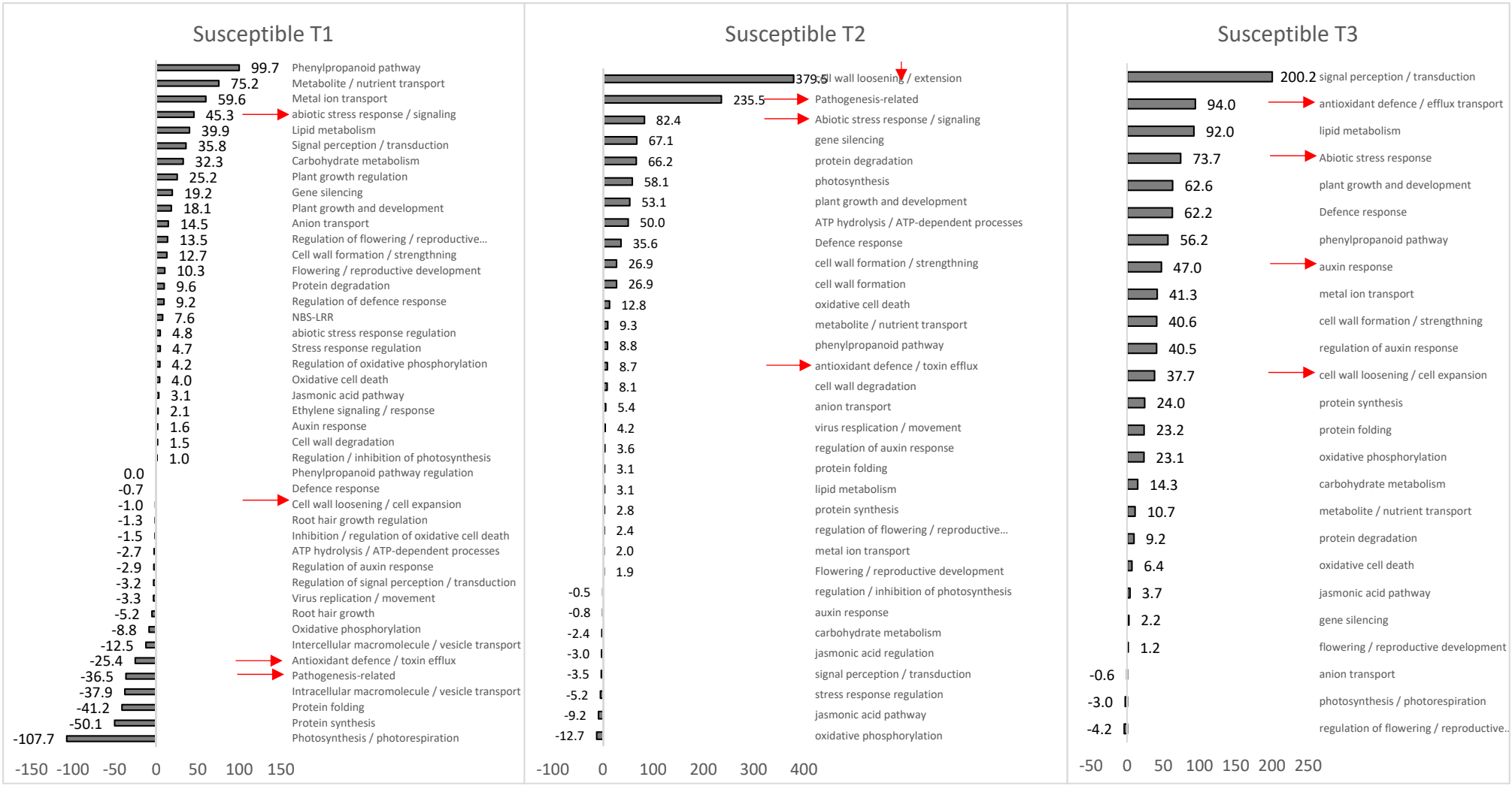


Figure 27. Bar plot of net CBSV-induced expression fold values for gene functional classes averaged across susceptible accession groups at one (T1), five (T2) and eight (T3) wpi. Red arrows point to functions net represed across time in resistant cassava but induced in non-resistant ones.

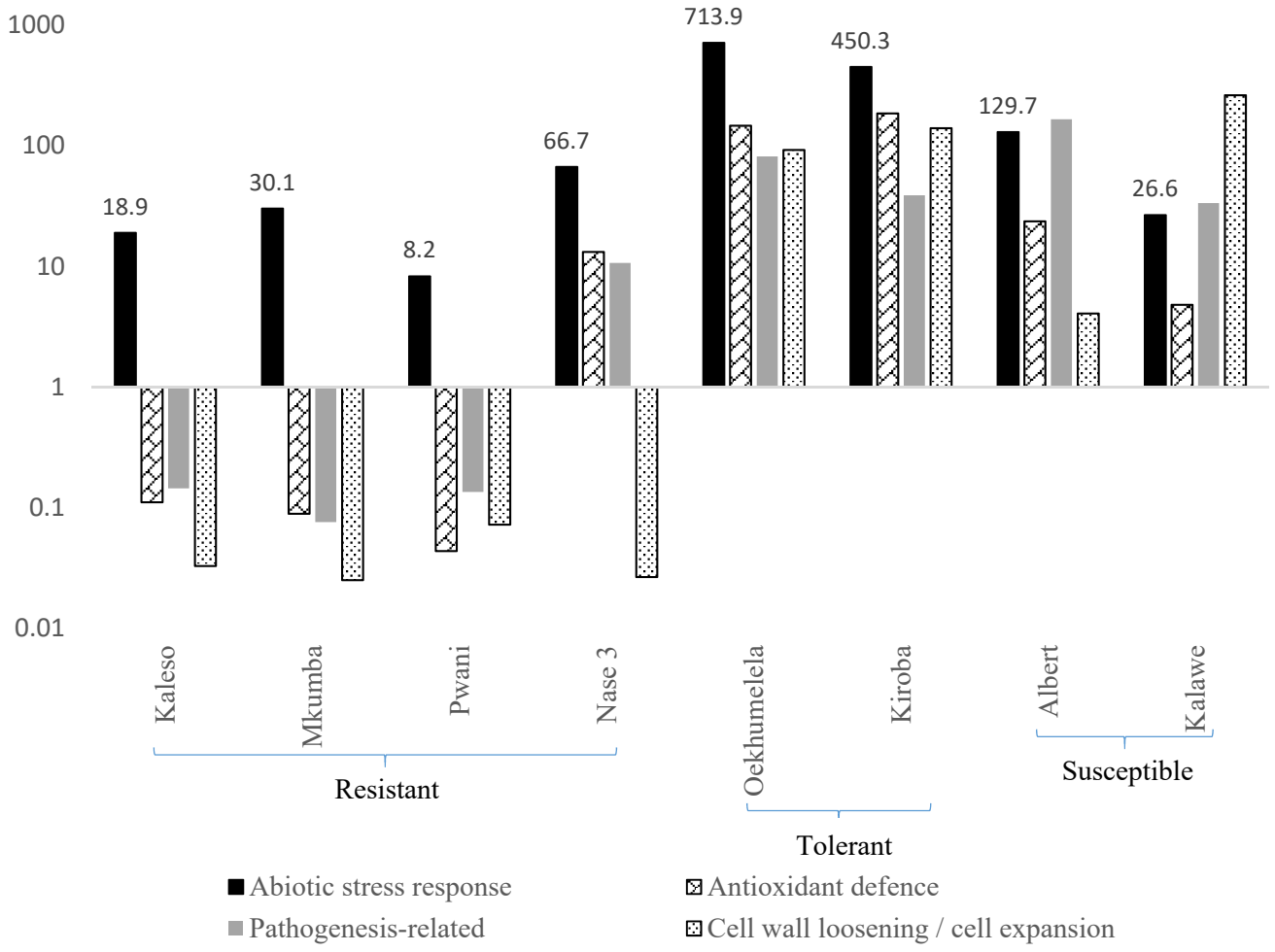


Figure 28. Average values of CBSV-induced net expression fold change in abiotic stress response, antioxidant defence pathogenesis-related and cell wall loosening / cell expansion genes of eight cassava accessions.

5.4 Discussion

The current study represents the largest, to date, of cassava transcriptome response to CBSV infection at multiple times post-inoculation. It sheds light into the extent to which CBSV infection, sampling time post-inoculation and the CBSV quantity-based cassava accession classification influence transcriptome expression. Crucially, it analyses gene expression responses to CBSV infection which enables modelling biological processes underpinning resistance and susceptibility responses to CBSV infection. It utilised RNA samples of varying qualities (Figure 18A) including low quality RNA templates of RIN values less than 6. The low RIN values might be an indication of sample degradation. Although low quality RNA templates have been associated with 3′ bias in read mapping (Sigurgeirsson *et al*., 2014) and have been shown to influence transcript levels (Gallego Romero *et al*., 2014), template quality of the RNA samples used in the current study does not dictate levels of entire the transcriptome, neither does it variation in levels of expressed genes thereby indicating that it might not influence part of the transcriptome (Figure 18B and 18C). This observation is in line with the poor correlation between RIN values of these samples and the proportion their reads mapped to the reference genome of cassva, which in fact remained high in all samples irrespective of template quality (Figure 18D). Therefore, while a 3′ bias in read mapping is expected in the lower quality templates used in this study, this seems not inflecnce normalised gene expression estimation hence calculation of differential gene expression.

Post-inoculation sampling time was found to exert substantial influence on transcriptome of the cassava accessions studied. As much as 68% or 13,058 out of 19,082 genes were significantly differentially expressed over the three sampling times (1, 5 and 8 wpi). This observation has implications in design of cassava transcriptome studies and the time of the experiment should be taken into consideration. In view of this, analysing transcriptome changes before and after treatment, as was done in some studies such as An *et al*. (2012) and Zhang *et al*. (2015), as opposed to comparing to a control plant may be biased by the effect of difference in growth stages at the different sampling times.

GFOLD analysis identified a total of 8971 differentially expressed genes between virus- and mock-inoculated plants of cassava accessions. In any single accession, up to 6379 genes were differentially expressed with a virus treatment-by mock treatment ratio of at least two. Of the genes so modulated in each accession, greater numbers were repressed than induced in all resistant accessions (Figure 21). In tolerant and susceptible accessions however, preferential induction and repression were both observed. Maruthi *et al*. (2014b) in a study involving one

susceptible (Albert) and one resistant variety (Kaleso) at one time point, observed opposite induction and repression patterns – specifically higher total number of induced (846 genes) than repressed (460) genes the in single resistant Kaleso variety but more repression activity (326 repressed versus 336 induced genes) in the susceptible Albert variety. This difference in modulation pattern may be due to difference in sampling times for transcriptome analysis, which was much later – one year post inoculation in above cited study. However, transcriptome response of cassava to a different virus, *South African cassava mosaic virus* (SACMV), yielded similar results to one observed here. At 12 days post inoculation, which is comparable to the 1 wpi sampling time used here, Allie *et al*. (2014) reported overall higher gene induction than repression activity, in susceptible relative to resistant cassava variety inoculated SACMV.

Gene expression variation was uncovered from comparison of gene expression of combinatorial pairs of resistant, tolerant and susceptible accessions. Expression variation was higher between resistant and tolerant accessions than between the latter and susceptible accessions. Specifically, total number of significantly differentially expressed genes between susceptible and tolerant accessions represents only 10% of those differentially expressed between tolerant and resistant accessions and 10.3% of genes differently expressed between resistant and susceptible cassava. This relationship implies a closer biological function profile between accessions of former two groups (tolerant and susceptible) than they have with the resistant.

A total of 334 genes, unmodulated by CBSV infection, were differentially expressed between resistant and susceptible cassava. Thirty seven of the 334 genes overexpressed in resistant accessions relative to the susceptible ones were enriched with defence response function encoded by NBS-LRR resistance genes. The NBS-LRR genes were also modulated by CBSV infection but these did not show unique patterns in resistant and susceptible cassava (Figure 25). This suggests that their modulation of virus infection might not be important for expression of the CBSD resistance phenotype. In older cassava leaf samples (1 year post-inoculation), genes of this class were expressed but not induced or repressed in response to CBSV infection in the CBSD resistant Kaleso (Maruthi *et al*., 2014) again suggesting their modulation might not be necessary for CBSD resistance. NBS-LRR genes are classical plant resistance genes involved in effector-triggered immunity which may be associated with the hypersensitive response (HR; Mandadi and Scholthof, 2013). The five NBS-LRR genes overexpressed in resistant samples include two RPP13-like genes and RPP8 and RPM1 genes. These genes function in HR-associated resistance response to diverse plant pathogens (Zhu *et al*., 2011; Mandadi and Scholthof, 2013). The LSD1 like 2 gene (cassava4.1_020424m), which is

homologous to the LDS1 gene – a negative regulator of oxidative cell death in *Arabidopsis thaliana* (Li *et al*., 2013) – is also overexpressed in resistant cassava accessions. The simultaneous overexpression of HR-inducing NBS-LRR genes and HR-inhibiting LSD1 in accessions with higher restriction to CBSV accumulation suggests that genetic resistance to CBSV in cassava could be associated with an innately heightened R gene-mediated control of invading CBSV and restriction of the hypersensitive cell death fallout through the action of the LSD1 like 2 gene. In many cases, the hypersensitive response has been found to be not an integral part of NBS-LRR controlled immunity but an appendage of it during late infection or due to lack of negative control of cell death such as in the runaway cell death mutant *lsd1* (Cui *et al*., 2014). Bendhamane *et al.* (1999) earlier showed that resistance to *Potato virus X* mediated by the R gene *Rx* occurs without HR. They advanced the notion that R genes can mediate a primary resistance response to invading viral pathogen by blocking the viral replication cycle if activation of this resistance is early and does not lag virus replication. In line with this theory, the higher expression levels of the five R genes in resistant accessions could pre-dispose them to exhibiting relatively more effective primary resistance response to CBSV compared to the susceptible accessions in which relatively lower expression of these genes could mean resistance response lagging CBSV accumulation. This situation would result in faster accumulation of the virus, hence greater propensity to induce systemic necrotic and senescent responses in the susceptible cassava. Systemic necrosis is emerging as a likely response in compatible (susceptibility) plant-virus interaction. It is distinct from HR in that it occurs late in infection, does not preclude virus multiplication and spread and is potentially lethal to infected plant (Mandadi and Scholthof, 2013). However, successful virus infection also requires survival of the host. It is generally assumed that host plants coevolve with their infecting pathogens. This partly entails reciprocal (fitness-reducing) effects of pathogen and host on each other depending on combination of host and pathogen (Fraile and Garcı´a-Arenal, 2010). Compared to indicator herbaceous hosts, cassava – the natural host of CBSV appears more adapted to the virus. For instance, two isolates of CBSV from CBSD endemic regions of east Africa caused severe necrosis and death in *Nicotiana benthamina* and *N. clevelandii* not later than four weeks post inoculation (Mohammed *et al*., 2012). Though the susceptible cassava variety Albert was severely affected by both isolates, it did not die from infection (Mohammed *et al*., 2012) but showed systemic venous chlorosis which widens on the lamina alongside appearance of necrotic spots as infected leaves grow older. These observations suggest that some underlying biological processes in CBSV-susceptible cassava moderate the tendency towards a full-blown systemic necrosis.

Higher net induction of antioxidant defence, abiotic stress response, cell wall loosening / expansion and pathogenesis-related functions was observed in the more non-resistant tolerant and susceptible accessions relative to CBSV-resistant accessions (Figures 25 and 26). Consistent with this finding, antioxidant defence, abiotic stress response, pathogenesis-related (PR)-mediated defence response are known to be induced during compatible plant interaction with multiple viruses (Whitham *et al*., 2003; Espinosa *et al*., 2007; Fernández-Calvino *et al*., 2016). These are likely indirect responses generally deployed by plants to cope with stress conditions. For instance, upregulation of antioxidant defence genes is known to be indicative of oxidative stress (Hernandez *et al*., 2015) and facilitates plant acclimatization to this stress (Tausz *et al*., 2004). The intervention of antioxidant defence during susceptibility response can be expected to delay cell death and by extension systemic tissue necrosis thereby preserving intracellular environments for continued virus replication and spread. The common induction of antioxidant defence genes in compatible plant-virus interaction and senescence suggests a connection between both processes (Buchanan-Wallaston *et al.*, 2002; Whitham *et al*., 2003; Espinosa *et al.*, 2007). If susceptibility or compatible interaction with CBSV does not completely subdue the host, the characteristic chlorotic symptom expression could therefore be explained by accelerated senescence which, like CBSD symptom, is associated with chlorosis.

The abiotic stress response function category encompassed genes involved in ABA biosynthesis and signalling, response or tolerance to stresses induced by drought, cold, salt, high temperatures and waterlogging. Based on current knowledge, the observed higher net induction of abiotic stress response genes in susceptible accessions compared to resistant accessions could possibly be explained by three hypotheses. First is the potential involvement of abiotic response processes in promotion of virus accumulation and associated symptom expression is susceptible cultivars. Correlation of enhanced abscisic acid (ABA) levels to increased susceptibility is mostly known for bacterial and fungal infections (Mohr and Cahill, 2003; Thaler *et al*., 2004) although increased ABA has been suggested to play a role in development of mosaic symptoms in *Tobacco mosaic virus* (TMV)-infected White Burley tobacco (Whenham *et al*., 1986). Second, net induction of abiotic stress response could be indicative of the plant's effort at controlling growing virus quantities. ABA-dependent control of CBSV accumulation in susceptibility response to the virus is unlikely since the ABA-dependent abiotic stress response processes showed an overall induction response in susceptible accessions. Finally, it seems that the most likely explanation is that it could simply be a response to oxidative stress induced by accumulating CBSV during compatible interaction with the virus. Abiotic stress conditions are

known to induce antioxidant defence genes (Seki *et al*., 2002). Direct evidence for the requirement of antioxidant defence genes for abiotic stress tolerance has been demonstrated (Bartels and Sunkar, 2005; Umezawa *et al*., 2006). These suggest that various stress-inducing abiotic factors also induce oxidative stress as is the case during virus infection. In fact, it has been repeatedly suggested that reactive oxygen species (ROS) is the central signal linking biotic and abiotic stress responses (Torres and Dangl, 2005; Fujita *et al*., 2006). Crucially, induction of both functions (abiotic stress response and antioxidant defence) by oxidative stress (Takahashi *et al*., 2004) lends further support to the third hypothesis.

Most pathogenesis-related (PR) proteins are the products of salicylic acid (SA)- or jasmonic acid (JA)-mediated resistance to necrotrophic and biotrophic pathogen respectively (Spoel and Dong, 2012). In this study, all differentially modulated genes annotated as chitinase, thaumatin, β-1,3-glucanase, PR-1 and defensin or their likes were classified under the pathogenesis-related functional group (Appendix 8). The net induction of PR genes in CBSV-sensitive cassava is not consistent with the known defence role of these genes. The defence functions of these genes are well established in bacterial and fungal diseases (Van Loon and Van Strien, 2008; Spoel and Dong, 2012). PR genes are induced in virus infections, however reports on their likely roles during this process are contradictory at best. *Panicum mosaic virus* (PMV) induced four PR genes – the β-1,3-glucanase *PR-2*, the chitinases *PR-3* and *PR-4* and the thaumatin gene *PR-5* alongside SA signalling components like Isochorismate synthase1 and Alternative oxidase during defence response to the virus (Mandadi and Scholthof, 2012). On the other hand, knock-down of β-1,3-glucanase in TMV- and *Tobacco necrosis virus* (TNV)-infected *Nicotiana* species enhanced resistance suggesting a requirement of β-1,3-glucanase in susceptibility response to plant virus (Beffa *et al*., 1996). The net induction of a β-1,3-glucanase gene in susceptible accessions suggests a similar virus infection-promoting role for the gene during compatible response to CBSV. PR and PR-like genes such as PR-1, β-1,3-glucanase, chitinase and thaumatin-like were also known to be induced during compatible interaction with plant viruses (Witham *et al*., 2002; Espinosa *et al*., 2007). Some of these PR genes were also induced during senescence (Buchanan-Wallaston *et al*., 2002; Espinosa *et al*., 2007) further suggesting a link between compatible plant-virus interactions to senescence.

Unique net induction of cell wall loosening / cell expansion and auxin response genes in susceptible cassava indicates expansion of cellular volume to dilute concentration of the accumulating virus in susceptible tissue. This reasoning is line with the association successful *Tobacco mosaic virus* infection with a disruption of auxin signalling repression (Padmanabhan

et al., 2006; 2008) and role of auxin response genes in modulating expression of cell wall loosening enzymes such as expansin and β-1,3-glucanase (also doubles as a pathogenesis-related gene as seen above) which they apparently exploit to re-model and loosen plant cell wall during cell expansion (Swarup *et al*., 2008). Clearly, a well-designed functional study will be required to confirm the actual roles of genes encoding these proceses in cassava-CBSV interaction.

5.5 Conclusions

Analysis of transcriptomes of the CBSV-infected cassava accessions studied here demonstrated:

1. The dynamic nature of cassava transcriptome. Gene expression and gene expression modulation in response to CBSV infection varied substantially over time.

2. That inter-accession difference in native gene expression levels and CBSV-induced gene expression modulation. Transcriptional response to infection was hardly correlated to nature of response to the virus – resistant, tolerant or susceptible – at the gene or gene family level. Instead, response to CBSV infection in cassava was associated with transcriptional response patterns of gene groups sharing common pathway or biological processes. Specifically, there was higher net induction of abiotic stress response genes in susceptible cassava and the induction and repression of antioxidant defence, cell expansion and pathogenesis-related genes in susceptible and resistant cassava respectively.

3. Transcriptional basis for difference in resistance and susceptibility responses to CBSV infection lies in the innately higher expression of NBS-LRR resistance genes in CBSD resistant cassava. This ensures a more effective control CBSV replication as well as the attendant oxidative stress which being more pronounced in susceptible cassava results in comparatively higher induction of antioxidant and abiotic stress response genes in these varieties.

Validation of the role of the genes and groups of genes identified in this study would facilitate direct genetic improvement to enhance CBSD resistance in the many farmer-preferred cassava accessions which currently are mostly susceptible to disease.

CHAPTER 6: Analysis of allele expression in cassava infected with cassava brown streak disease

6.1 Introduction

Available and emerging genome analysis techniques of the next generation sequencing technologies hold high promise for understanding the molecular underpinnings of symptom production from and resistance to *Cassava brown streak virus* (CBSV) infection. Such knowledge will prove useful for developing molecular-based strategies for effectively combating CBSV which is still causing huge losses to cassava production in east Africa and currently spreading west-ward through central Africa (Mulimbi *et al*., 2012). The next generation RNA sequencing (RNA-seq) technique has already been deployed to identify unique gene expression response to CBSV infection in the CBSV-resistant cultivar, Kaleso (Maruthi *et al*., 2014b). This technique has also been utilized to study responses to other treatment conditions in cassava (Allie *et al*., 2014; Fu *et al*., 2016). Recently, the next generation sequencing technology was applied to the study of diversity and ancestry of diverse cassava germplasm. In this study, single nucleotide variation data from the sequenced cultivars was used to confirm the ancestry of the CBSD-resistant Kaleso as well as to identify haplotypes introgressed into this variety from the wild *Manihot* species *Manihot glaziovii* (Bredeson *et al*., 2016). Given that CBSD resistance in Kaleso was *trans*ferred from this wild species (Hillocks and Jennings, 2003; Bredeson *et al*., 2016), the introgressed haplotypes represent an important resource for mining for CBSD resistance. Hence, allele expression at heterozygous loci of CBSD-resistant accessions were compared to haplotypes introgressed from *M. glaziovii* with the aim of identifying *M. glaziovii* alleles commonly expressed in the CBSD-resistant accessions studied here.

Besides its obvious use in gene expression quantification, RNA-seq can also be applied to quantification of expression at alleles of individual single nucleotide polymorphism (SNP) positions of the genome. Measuring allele expression allows for uncovering of instances of allele imbalance in heterozygous individuals (Albert and Kruglyak, 2015). Allele expression imbalance or allele-specific expression (ASE) is usually an indication of the action of a *cis*-expression quantitative trait locus (*cis*-eQTL) which causes differential regulation of parental copies of a single gene (Skelly *et al*., 2011; Albert and Kruglyak, 2015). Such clear understanding of *cis* regulation represents an improvement of ASE over classical eQTL mapping studies for which an unambiguous separation of *cis* and *trans* genetic variations is not possible (Zhang and Borevitz, 2009). In the current study, allele expression balance at

potentially *cis*-regulated loci was found to vary across virus-treatment categories and sampling time. Change in allele expression balance, between CBSV- and mock-inoculated plants of an accession, in contiguous loci located within the same gene, was used to predict CBSV-induced modulation of gene-level ASE.

Although *cis*- or *trans*-regulation is confirmed from comparison of allele expression in hybrids and their parents, both regulatory mechanisms can be delineated from differential allele expression data of genetically distinct individuals of the same species. For instance, *trans* and *cis* regulators of gene expression have been predicted from eQTL mapping of *trans*criptome to genome-wide SNP genotypes in natural *Arabidopsis* accessions (Zan *et al*., 2016). As pointed out earlier, prediction of regulators of gene expression from DNA variation data require that individual plants be genetically distinct or polymorphic. This requirement limits the range as well as number of individuals deployable, hence eQTL analysis might be impractical in situations when number of crosses or accessions are insufficient. However, prediction of *trans*-regulators of gene expression based on allele expression of SNP loci contained within expressed genes, does not require genetic uniqueness of individuals given that allele expression of clones can vary under different conditions (Edsgärd *et al*., 2016). This means that number of individuals deployable for *trans*-regulator prediction, in cases of small population sizes, can be expanded. In the current study, attempt was made at prediction of putative *trans*-regulators of gene expression from multiple samples of eight cassava accessions. Unlike classical eQTL analysis in hybrid or diverse landrace population, only functional *trans*-acting gene expression regulators were predicted from transcriptome-wide allele expression data. Prediction of functional regulator loci allowed for the development of the concept of co-correlated 'paralogous loci' for SNP locus pairs, contained within paralogous genes, which correlate with same target gene.

Analysis of allele expression was demonstrated as potentially useful for characterizing the regulatory roles of SNP variations of cassava in gene expression responses to CBSV infection as well as their putative role in providing genetic basis for CBSD resistance.

6.2 Materials and Methods

Details of plant materials used, their CBSV infection response categories and growth conditions, CBSV inoculation methods, wet-lab procedures including RNA isolation and quality check, library preparation and sequencing are contained in sections 5.2.1 to 5.2.5 of chapter 5. Sections 5.2.6 to 5.2.7 of the same chapter detail the *in silico* analysis methods applied to RNA-seq data up to read sequence mapping.

6.2.1. SNP genotype call and read counting per allele

SNP variation data for each accession was extracted using an input of at least two alignment files per accession on SAMtools's mpileup program (http://samtools.sourceforge.net/mpileup.shtml). SNP variation data was extracted from the resultant variant file on an excel spreadsheet. In all samples, each locus was required to have a read coverage greater than 2 and base quality of at least 25 before a genotype call could be made. Number of read counts covering each allele of a SNP locus was counted using the ASEReadCounter software from GATK (https://software.broadinstitute.org/gatk/gatkdocs/org_broadinstitute_gatk_tools_walkers_rnaseq_ASEReadCounter.php). However prior to this, filtered reads were re-aligned using the Genomic Short-read Nucleotide Alignment Program (GSNAP; http://research-pub.gene.com/gmap/) to an 'enhanced' reference genome which incorporated SNP variation data of each accession. This was necessary to eliminate read mapping bias to reference alleles which influences the accuracy of allele expression data obtained from read counts mapped over alleles. Also prior to allele read counting, duplicate reads were removed using Picards tool's MarkDuplicates software (http://broadinstitute.github.io/picard/command-line-overview.html).

6.2.2 Measurement of allele-specific expression

Except for measurement of CBSV-induced allele expression modulation, allele expression at each locus was calculated as proportion of total read counts mapped unto the reference allele. For calculation of virus-induced allele expression modulation, allele expression was presented as a major allele frequency (MAF) which is ratio of read numbers mapped over the major allele (one of two alleles of a locus, which has more number of mapped reads) to the total number of reads mapped at a locus. This method allows comparison of allele expression across adjacent SNP loci of a gene for estimation of gene-level allele-specific expression (ASE; Mayba *et al*., 2014). Infection-induced change in MAF was indicated by MAF difference of at least 0.17 which represents at least two-fold difference in MAF of CBSV- and mock-inoculated samples

of each accession. In any accession, contiguous loci on the same gene showing CBSV- by mock-inoculated MAF difference consistent with above criteria were considered an indication of gene-level ASE induction. Only such loci were counted as showing true CBSV-induced allele ratio imbalance.

6.2.3 Prediction of gene expression regulators

Gene expression values of 387 abiotic stress response, pathogenesis-related and antioxidant gene targets uniquely induced in susceptible and intermediate accessions (Appendix 10) were modelled after allele expression predictor SNP loci expressed across 43 cassava samples representing eight accessions. This was achieved using an ensemble machine learning technique called random forest (Breiman, 2001) implemented using an R package called rfPermute (https://cran.r-project.org/web/packages/rfPermute/index.html). Expression levels of 35 genes overexpressed in resistant genotypes were likewise modelled. Extent to which expression of a target gene is predicted by each predictor allele was measured by a feature importance value defined as percent increase in mean square error (%IncMSE) for prediction by the permuted variable of a predictor compared to the actual predictor variable. The %IncMSE values from multiple bootstrap samples of each predictor were aggregated to obtain a single %IncMSE feature importance value. For each predicted target-locus interaction pair, variance of target RPKM values averaged over samples expressing each of the three allele types of correlated loci – reference homozygous, heterozygous and alternate homozygous alleles (alleles of samples different from one found in the reference genome) – was calculated as an independent assessor of SNP-target association. Expectation is that feature importance values from random forest analysis will correlate directly with these variance values.

6.2.4 Differential allele expression between CBSD-resistant and -susceptible cassava

Differentially expressed alleles between CBSD-resistant and susceptible cassava samples of eight accessions were determined from multiple ANOVA tests of variance in allele expression, at each of the 32,256 expressed loci. Specifically, the allele expression difference was assessed between CBSD-resistant accession groups – Kaleso, Mkumba, Pwani and Nase 3 and non-resistant (tolerant and susceptible) accessions – Oekhemelela, Kiroba, Albert and Kalawe. P-values obtained were adjusted for multiple testing using the false discovery rate method (Benjamini and Hochberg, 1995). Both ANOVA and multiple testing adjustment of *P* values were achieved using the base stats package of R (R core team, 2016).

6.2.5 Mining for CBSD resistance alleles introgressed from wild Manihot

Genetic variations potentially underlying CBSD resistance were mined from database of *Manihot glaziovii* alleles introgressed into cassava (Bredeson *et al*., 2016). Unique heterozygous loci of each accession was used for this search.

6.3 Results

6.3.1 Relatively high heterozygosity in CBSD-resistant cassava and identification of introgressed CBSD resistance marker alleles

Total number of expressed SNP loci in 48 sequenced cassava samples, encompassing eight genotypes, ranged from 59,170 to 291,304. The level of homozygous and heterozygous allele expressions was determined as a percentage of all expressed loci. More homozygous than heterozygous SNP loci were expressed in all samples. However, the proportion of heterozygous or homozygous loci expressed varied across genotypes. On average, percentage of all expressed loci which are heterozygous loci was higher in resistant genotypes compared to susceptible genotypes (Table 18).

Table 18. Percentage homozygous and heterozygous loci of eight cassava accessions averaged over CBSV- and mock-inoculated samples of the accessions at the three time-points post inoculation.

Genotype	Average of Percent Heterozygous loci	Average of Percent Homozygous loci
Kaleso	44.4	55.6
Mkumba	39.3	60.7
Pwani	39.2	60.8
Nase 3	38.6	61.4
Oekhumelela	36.2	63.8
Kiroba	34.3	65.7
Albert	33.6	66.4
Kalawe	32.5	67.5

Pair-wise comparison of heterozygous loci expressed in all samples of each accession showed greater similarity between pairs of the CBSV-resistant accessions – Kaleso, Mkumba and Pwani than between these and susceptible accessions or among the latter. Specifically, Kaleso/Mkumba and Kaleso/Pwani resistant accession pairs both shared 39.1% similarity in heterozygous loci expression while Pwani and Mkumba showed 99.4% similarity. In contrast, similarity between other genotype pairs ranged from 19% to 26.6% except for Kiroba/Albert and Oekhumelela/Nase 3 and Albert/Kalawe accession pairs which showed 99%, 34.7% and 30.7% similarity respectively. Though Kiroba and Albert showed high similarity in heterozygous locus expression, they expressed substantially more dissimilar reference and alternate homozygous loci at 49% and 36.2% similarity, respectively. Mkumba and Pwani however showed similarity in homozygous locus expression lower than the level observed in their heterozygous loci. Reference and alternate homozygous loci expressed by both accessions were 79.8% and 80.3% identical, respectively.

Heterozygous loci uniquely expressed in each of the three resistant accessions but not in the non-resistant ones were found to be mostly located in chromosomes 1, 3, 4 and 13 in Mkumba and Pwani, and chromosomes 1, 3, 4, 13 and 14 of Kaleso (Figure 29a). This distribution of unique heterozygous loci mirrors distribution of alleles introgressed into Kaleso from the wild progenitor – *Manihot glaziovii* (Bredesson *et al*., 2016). Indeed the introgressed *M. glaziovii* alleles belonged to the set of unique heterozygous loci of Kaleso, Mkumba and Pwani and the numbers of these alleles expressed from chromosomes 1, 3, 4, 13 and 14 in the three resistant accessions are shown in Figure 30. Distribution of heterozygous loci in the two susceptible accessions – Albert and Kalawe – was starkly different. Chromosomes most populated with heterozygous loci were chromosomes 9 and 15 (Figure 29c) and *M. glaziovii* alleles were mostly absent from both susceptible accessions. Set of unique heterozygous loci of Nase 3 have chromosome location distribution different from those of susceptible cultivars and of Kaleso, Mkumba and Pwani described above (Figure 29b). However, as in Mkumba and Pwani, chromosome 4 of Nase 3 was the most populated with unique heterozygous loci. Interestingly, the same chromosome contains the most number of introgressed *M. glaziovii* alleles (Figure 30) all of which belong to the unique heterozygous loci of Nase 3.

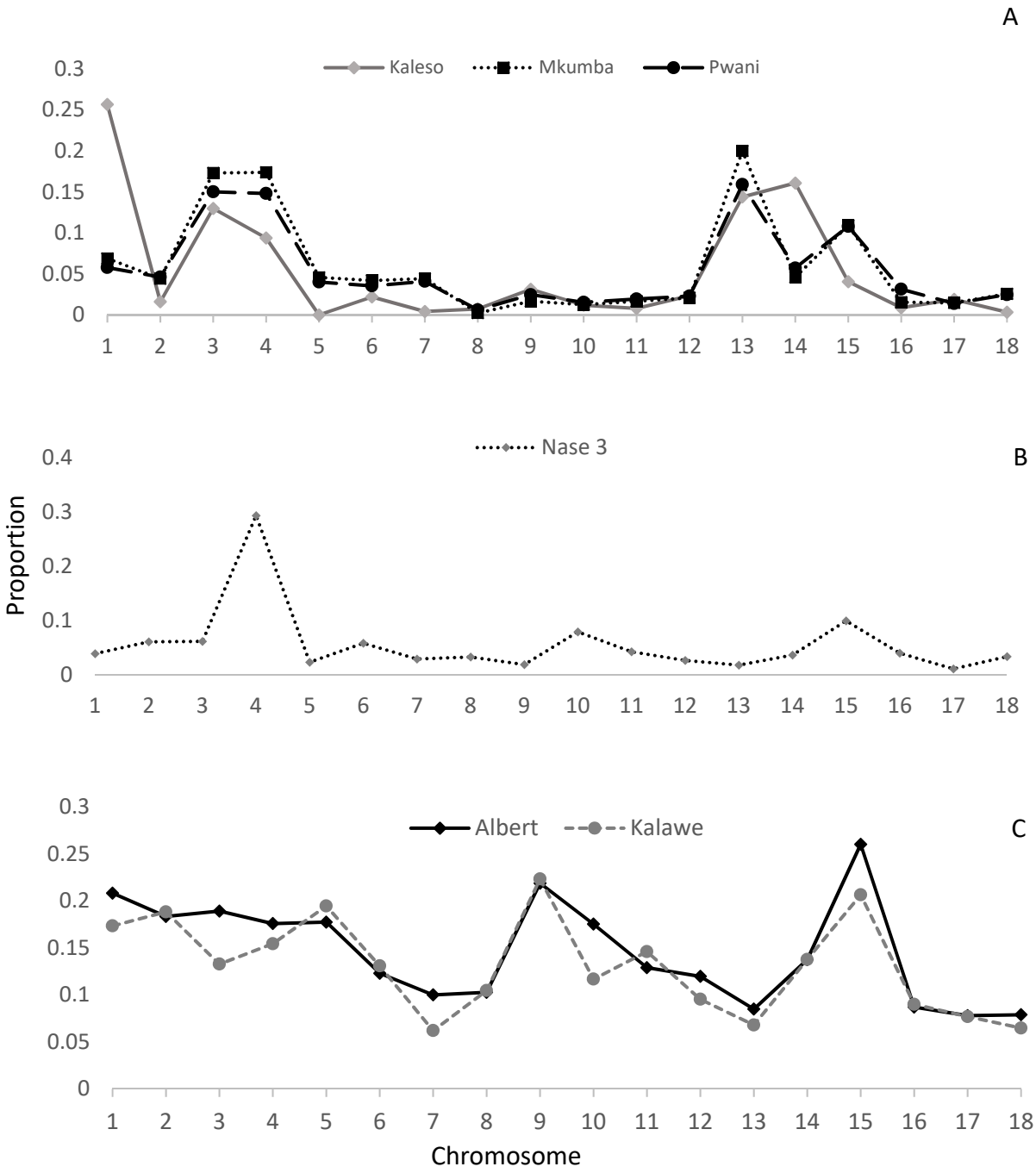


Figure 29. Proportions of unique heterozygous loci (which are not found in CBSV tolerant and susceptible cassava) in each chromosome of A. Kaleso, Mkumba and Pwani. B. Nase 3. C. All heterozygous loci of Albert and Kalawe in each of their chromosomes.

Cassava genome regions containing *M. glaziovii* alleles commonly expressed in Kaleso, Mkumba and Pwani were located within 27.2 to 27.3 mega bases (Mb), 2.3 to 3.1 Mb and 21.9 to 28 Mb segments of chromosomes 3, 4, and 13, respectively (Figure 31). Function prediction for genes containing *M. glaziovii* alleles showed enrichment for abiotic stress response, transcriptional regulation and disease resistance functions (Appendix 9).

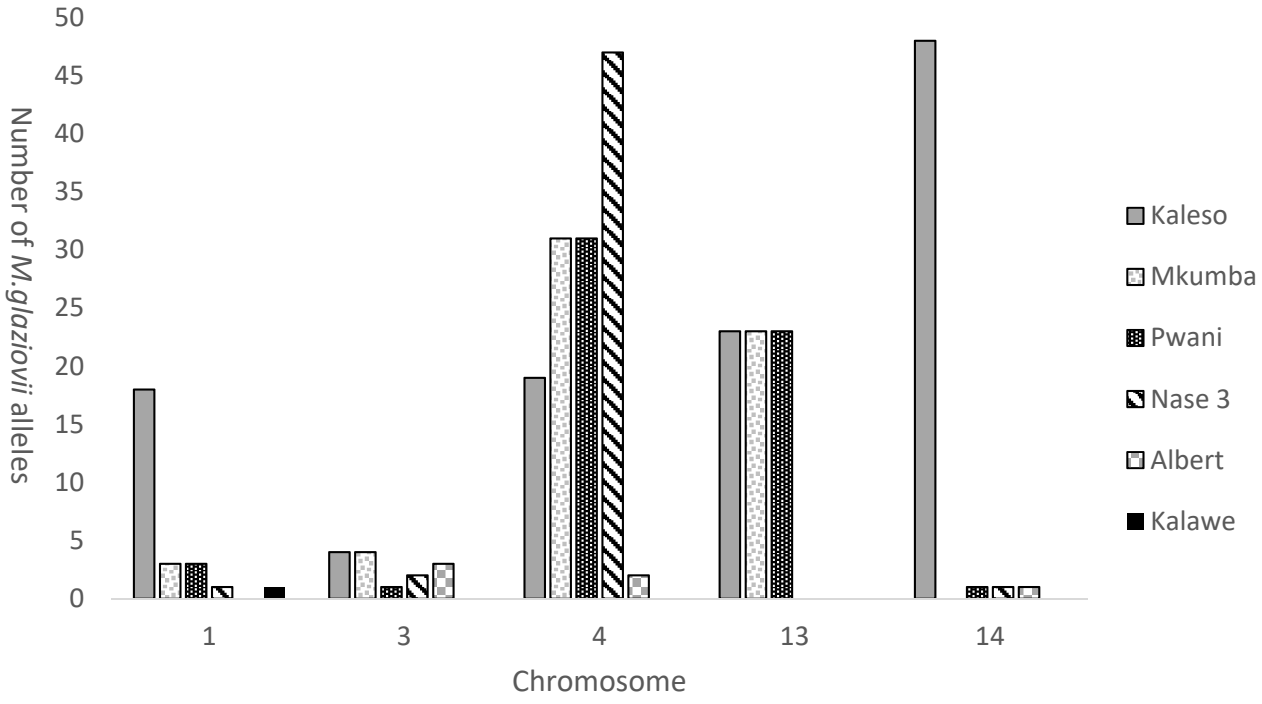


Figure 30. Numbers of introgressed *M. glaziovii* alleles expressed from five of 18 chromosome pairs of eight cassava accessions.

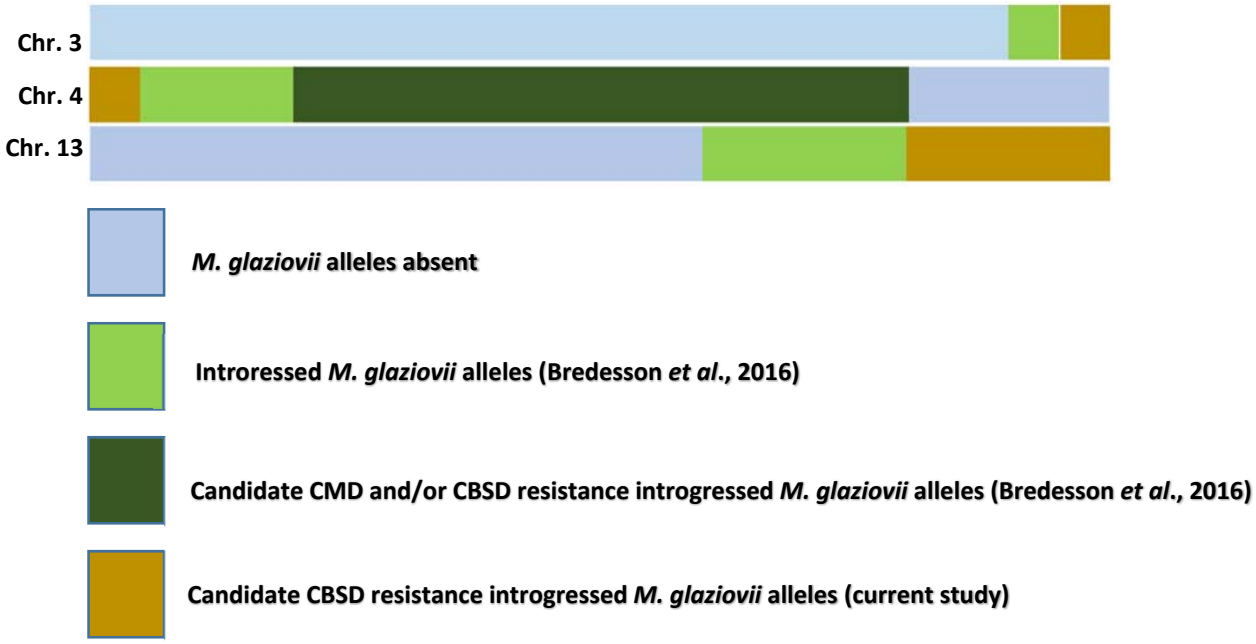


Figure 31. Illustrated genome locations of commonly expressed *M. glaziovii* alleles of Kaleso, Mkumba and Pwani relative to other *M. glaziovii* alleles including those predicted as candidates for CMD and/or CBSD resistance by Bredesson et al. (2016).

6.3.2 Differential allele expression between CBSV-resistant and non-resistant cassava

Comparison of loci showing allele expression identified a set of 32,256 common SNP loci in 43 of the 48 sequenced samples. The other five samples – one Pwani, two Nase 3 and two Albert samples – shared very little SNP loci with majority of the samples. Allele expression levels at the 32,256 loci calculated as proportion of total read counts mapped to the reference allele were analysed for statistical association with the CBSV phenotype response categories of the 43 samples encompassing eight cassava accessions. Analysis of variance (ANOVA) showed that 4543 alleles were determined to be significantly differentially expressed between CBSV-resistant and non-resistant samples at q value < 0.05 and allele expression fold difference > 2 (Figure 32). These alleles were found to be located in genes most enriched with nucleic acid binding functions such as transcription factor-mediated transcription regulation functions as well as various protein binding functions (Figure 33). Nucleic acid binding also encompassed transcription by RNA polymerases, RNA and DNA unwinding or conformational change by helicases, pre-mRNA processing by various genes including splicing factors, and protein synthesis by ribosomal proteins and eukaryotic initiation factors. Protein binding functions on the other hand encompassed various transport proteins, peptide repeats, proteases, zinc finger proteins and eukaryotic translation initiation factors.

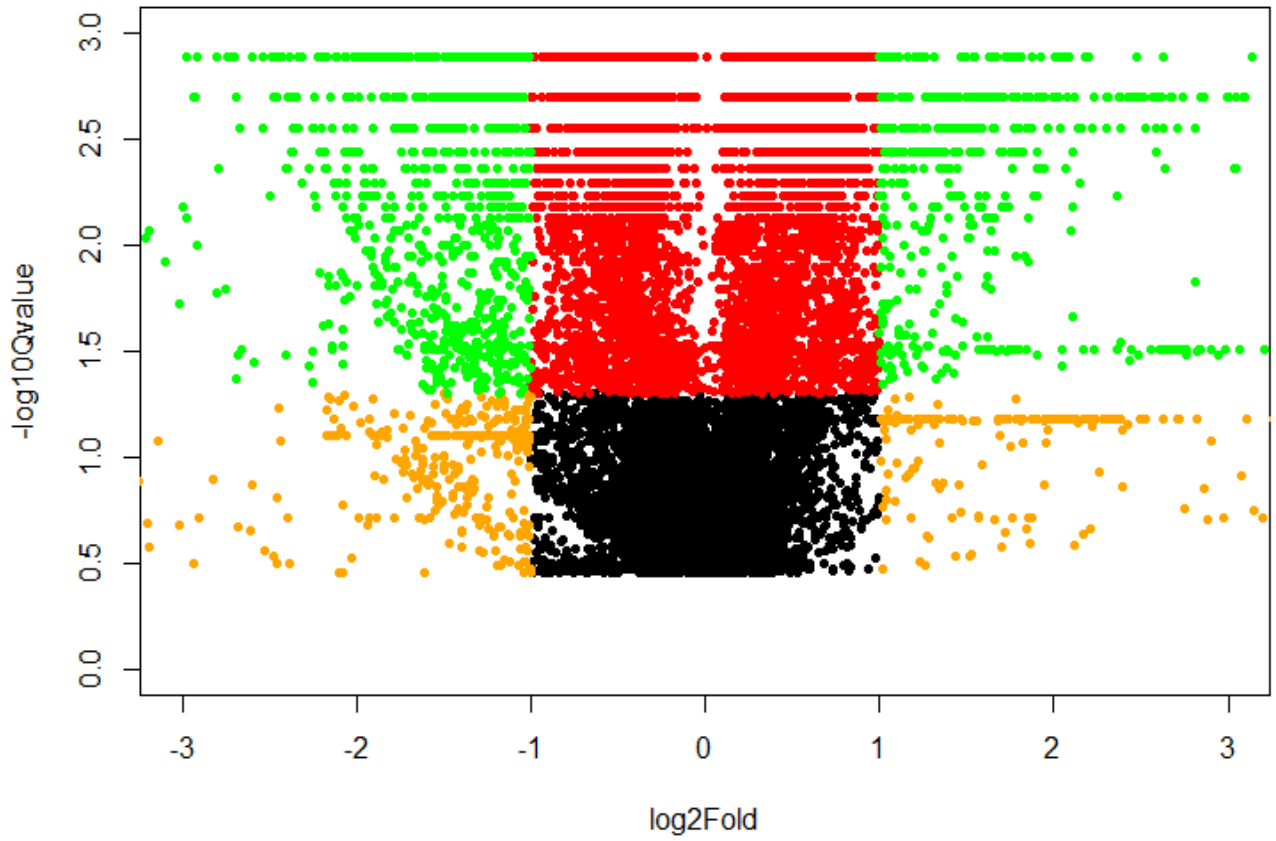


Figure 32. A volcano plot of q values, from ANOVA test of association between allele expression and phenotype response classes of 43 samples of eight cassava genotypes to CBSV infection, against fold difference in allele expression between CBSD-resistant and non-resistant genotypes. Black and red dots represent un-associated alleles with and q value > 0.05 and log2 of fold difference less than and greater than 1 respectively. The green dots represent alleles statistically associated to phenotype response to CBSV infection with q value < 0.05 and log2 of fold difference > 1.

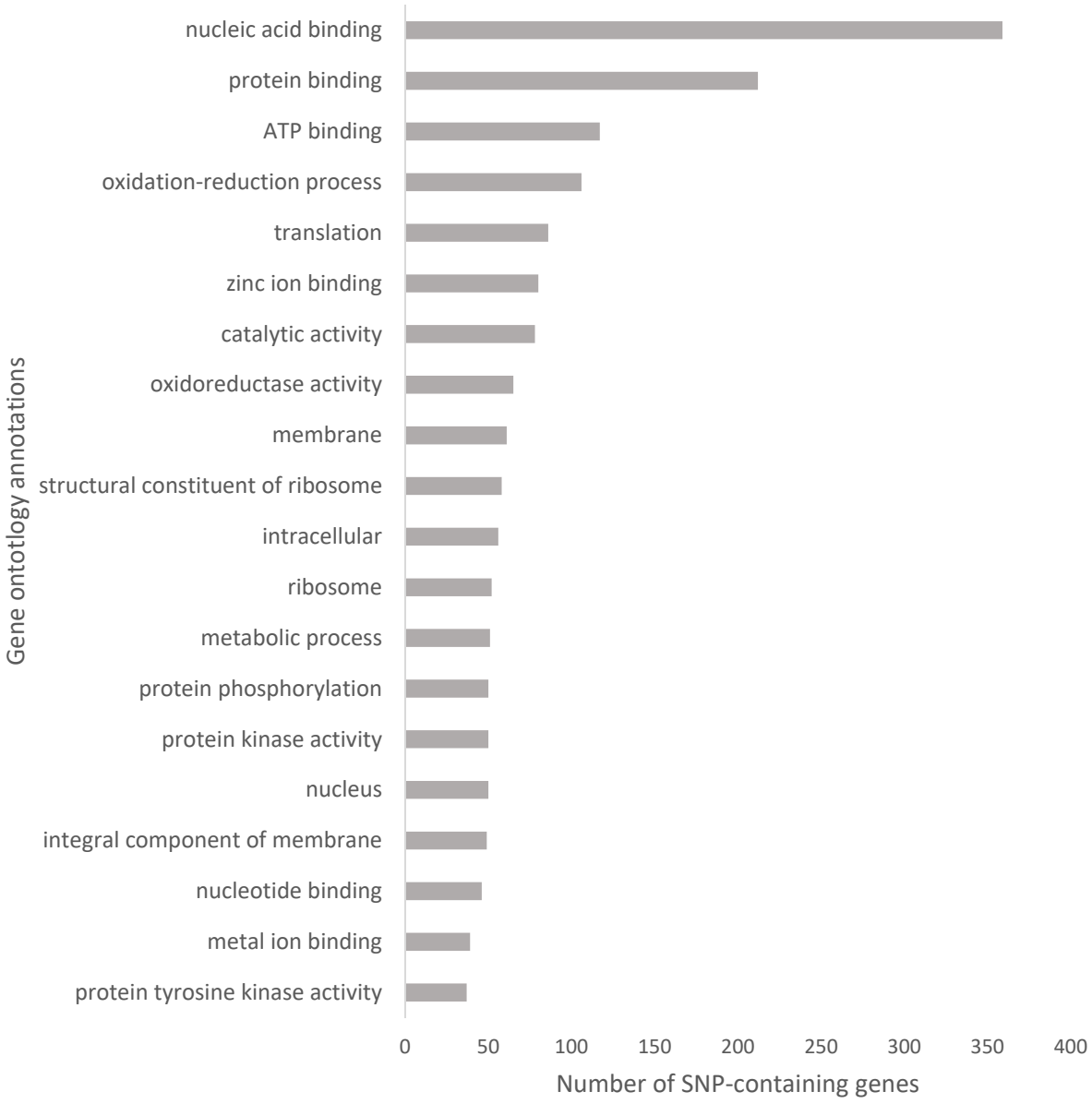


Figure 33. Distribution of numbers of SNP-containing genes encoding each of 20 gene ontology (GO) function categories enriched in SNP loci associated with CBSD phenotype response.

6.3.3 Allele-specific expression

Cases of allele-specific expression (ASE) or imbalanced (unequal) expression of alleles of 32,256 SNP loci was investigated. Identity and number of imbalanced loci was found to vary depending on cassava accession, CBSV infection status of the accession or the time after CBSV inoculation. In line with this variability of ASE, rate of ASE events consistent across virus treatment samples of each accession, and across post-inoculation times, was substantially smaller than the sum of all ASE events observed in each accession (Table 19).

Table 19. Percentage rates (for each of eight cassava accessions) of all ASE events irrespective of virus treatment or post-inoculation time-point and those consistent across virus treatment status and time-points.

Genotype	Total percentage rate of all ASE events	Percentage rate of ASE consistent across virus treatments and time
Kaleso	12.4	1.2
Mkumba	12.8	1.36
Pwani	11.9	1.51
Nase 3	18	0.77
Oekhumelela	11.1	1.14
Kiroba	14.4	0.79
Albert	10.2	1.05
Kalawe	11.6	0.93

There were a total 428 heterozygous loci consistently imbalanced (with unequal allele expression) in the four resistant accessions. Allele expression at eight of these were found significantly associated (q value < 0.05; allele expression fold difference > 2) with CBSD symptom expression. Their expressed alleles were either biased towards opposite alleles in resistant and susceptible sample groups or imbalanced in one of the two sample groups while balanced in the other (Figure 34).

Differences in major allele frequencies, of contiguous loci, observed between CBSV- and mock-inoculated were considered CBSV-induced. CBSV-induced modulation of allele expression was very variable as different loci were induced across accessions and times post inoculation. Specifically, only a couple of loci showed consistent CBSV-induced allele expression imbalance at more than a single time post-inoculation. This was in the accessions Pwani and Oekhumelela. Out of a total of 607 loci with CBSV infection-mediated induction or abolishment of ASE, only 22 (3.6%) were commonly induced in any pair of resistant samples. In spite of the highly variable nature of allele imbalance modulation by CBSV infection, the top five GO function annotations of genes and alleles showing CBSV infection-modulated ASE across accessions consistently included nucleic acid binding, protein binding and oxidation-reduction or protein phosphorylation gene ontology functions.

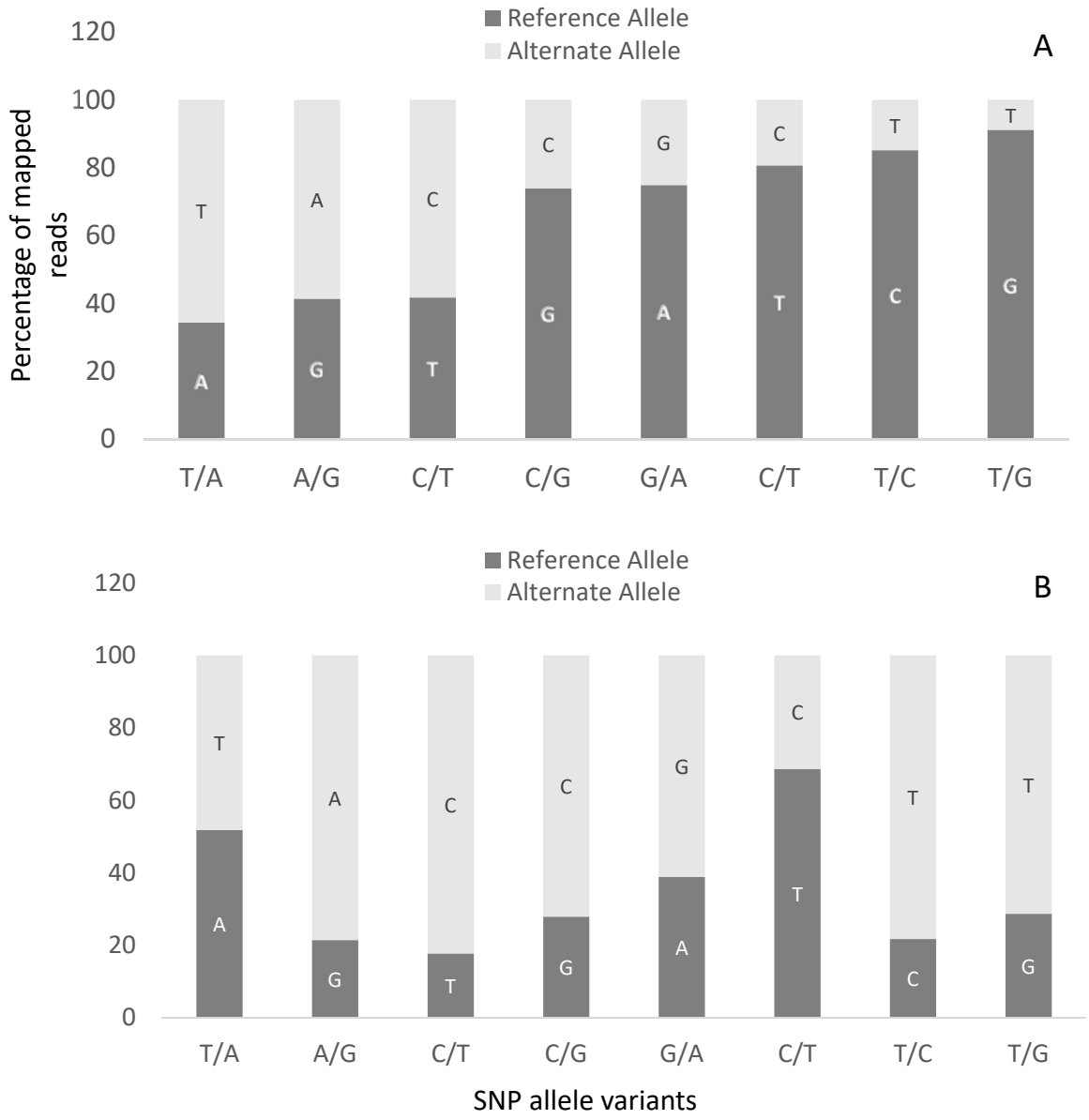


Figure 34. Allele expression, represented as average percentage of reads mapped to each of two allele of eight *cis*-acting loci of A. the CBSV-resistant genotypes – Kaleso, Mkumba, Pwani and Nase 3 and B. the non-resistant genotypes – Oekhumelela, Kiroba, Albert and Kalawe.

6.3.4 Prediction of putative *trans*-regulators of expression of susceptibility-associated genes (SAGs)

Random forest analysis was used to obtain measures of extent of correlation (feature importance) of allele expression, at 32,256 predictor loci, to expression levels of 387 susceptibility-associated genes (SAGs) which are abiotic stress, antioxidant response and pathogenesis-related target genes uniquely induced in susceptible cassava. The SAGs and feature importance values for their best predictor loci are as presented in Appendix 10. For each predictor locus, expression levels of its predicted target gene across samples were classified into three groups according to allele(s) of the locus expressed by the sample (reference homozygous, heterozygous or alternate homozygous). Variance in average gene expression of these groups was calculated and averaged over decile ranges. These average variances were tested for correlation with corresponding decile means of feature importance values. Both measures were found to be highly correlated (Figure 35).

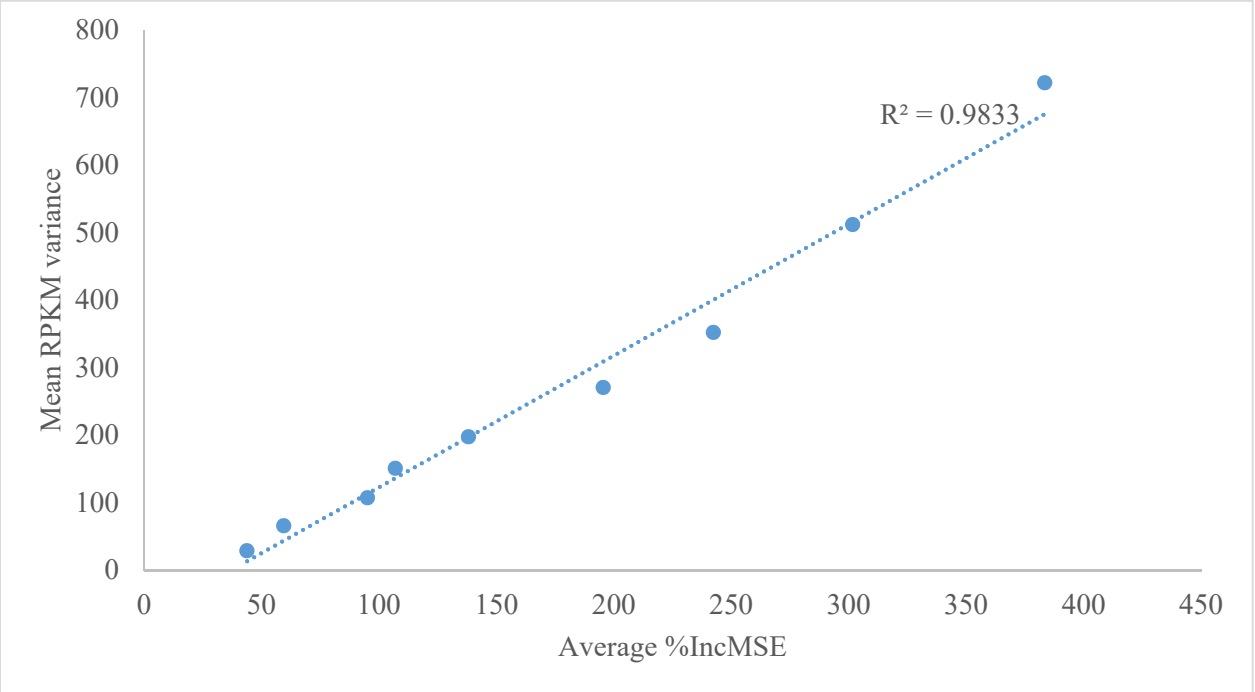


Figure 35. Decile-range averages for variance of mean RPKM of target genes across samples groups expressing three different allele variants of their corresponding best predicted regulator loci plotted against feature importance values averaged over same ranges.

Allele expression at 5891 putative *trans*-acting SNP loci were found correlated with the expression of the 164 target genes at feature importance values ranging from 20% to over 26000%. Allele expression in 20 of these loci (hereafter referred to as SAG *trans*-regulators) were significantly associated with classification of the cassava accession classification based on CBSV accumulation levels (q value < 0.05). Differences in RPKM expression levels of SAG *trans*- regulator-correlated target genes between allele sample groups expressing homozygous reference and homozygous alternate alleles of the SAG *trans*-regulator loci ranged from 1.2 to 4 fold. Allele sample groups with the most expression of the target genes were found to be largely consisted of the susceptible accessions while least expressed sample group were mostly resistant. This observation was illustrated for the best correlated locus of a plastid transcriptionally active 12 gene uniquely induced in CBSV-susceptible accession(s) (Figure 36). Gene ontology analysis showed that none of the genes containing the 20 *trans*-SAG regulator loci (container genes) is a transcription factor. They instead coded for gene ontology functions as diverse as oxidation-reduction processes, ATP and carotenoid biosynthesis but were particularly enriched with the oxidation-reduction function (Table 20).

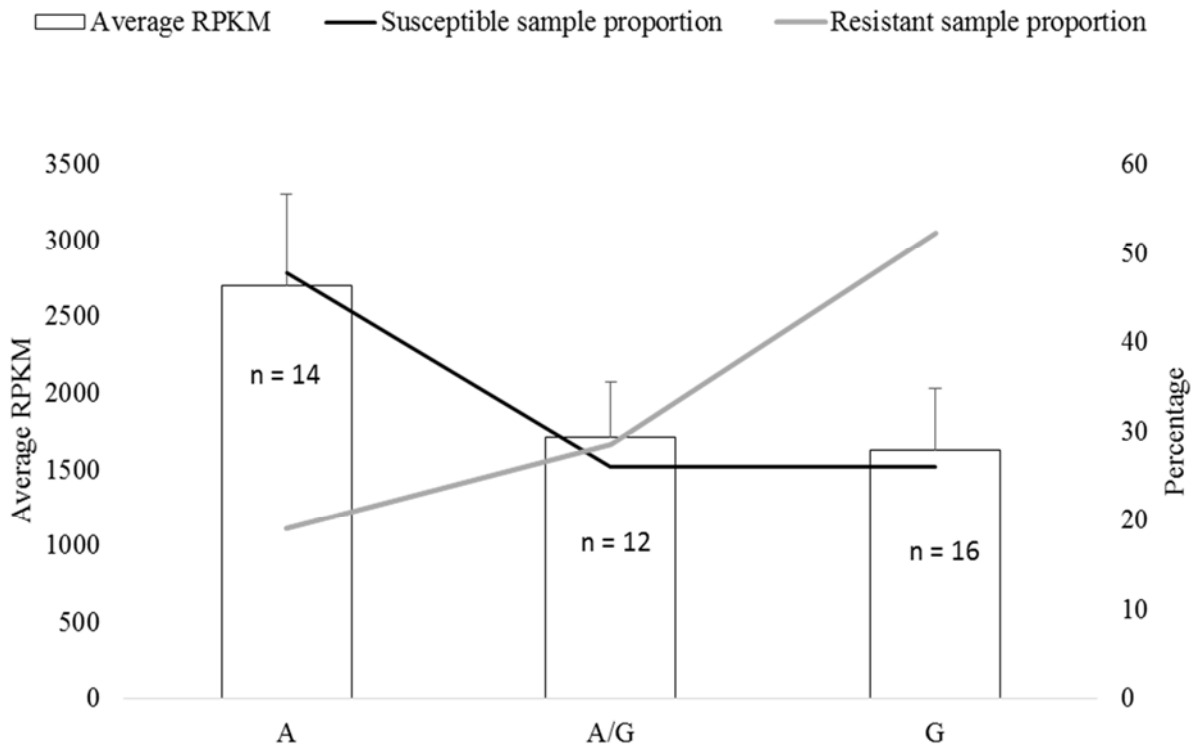


Figure 36. Combined bar and line chart illustrating association of a predicted regulator SNP locus with its target gene (plastid transcriptionally active 12) and with CBSV quantity-based categorisation of cassava. The bar plot component compares expression levels of the gene, in sample groups expressing different alleles of the regulator locus. The line plot component shows the proportions of 43 cassava samples which are CBSD resistant or susceptible in each of the three sample groups expressing the three variants of the regulator locus. Comparison of proportions of susceptible and resistant samples in each allele sample group shows distinct allele expression priorities for resistant and susceptible cassava as well as higher expression of the plastid transcriptionally active 12 in the allele group consisting mostly of susceptible samples.

Table 20. Genomic locations, functions and feature importance values for 20 SNP loci associated with CBSD phenotype response classification and correlated with antioxidant and stress-response genes uniquely induced in accessions susceptible to CBSV

Scaffold location of SNP locus	Base position of SNP locus in scaffold	Reference allele of SNP locus	Alternate allele of SNP locus	Chromosome location of SNP loci	Gene location of SNP loci	Gene ontology of container genes of SNP loci	Importance value from SNP-target correlation	Target gene	Target gene name
scaffold00276	22991	A	G	Chr.16	cassava4.1_012043m	barrier septum assembly	46.56	cassava4.1_013461m	L-ascorbate peroxidase
scaffold00987	15302	C	T	*N/A	N/A	N/A	23.24	cassava4.1_015084m	Glutathione peroxidase 1
scaffold01551	1611839	T	C	Chr.15	cassava4.1_001506m	oxidation-reduction process	255.29	cassava4.1_016240m	RESPONSIVE TO DESSICATION 22
scaffold01551	1633773	G	C	Chr.15	cassava4.1_000148m	N/A	74.58	cassava4.1_006302m	Catalase 2
scaffold01608	87214	T	A	Chr.06	N/A	N/A	251.37	cassava4.1_017966m	Putative universal stress protein
scaffold01895	144029	T	C	Chr.05	cassava4.1_007468m	glucose metabolic process	34.21	cassava4.1_014643m	L-ascorbate peroxidase
scaffold02943	25722	A	G	Chr.18	cassava4.1_003782m	protein binding	24.77	cassava4.1_015710m	L-ascorbate peroxidase
scaffold03116	220771	T	A	Chr.11	cassava4.1_008622m	biosynthetic process	33.21	cassava4.1_013461m	L-ascorbate peroxidase
scaffold03264	197697	T	A	Chr.14	cassava4.1_012618m	oxidation-reduction process	33.64	cassava4.1_012402m	REDUCED SUGAR RESPONSE 4
scaffold03264	197824	T	G	Chr.14	cassava4.1_012618m	oxidation-reduction process	26.60	cassava4.1_012402m	REDUCED SUGAR RESPONSE 4
scaffold03264	197824	T	G	Chr.14	cassava4.1_012618m	oxidation-reduction process	33.90	cassava4.1_012402m	REDUCED SUGAR RESPONSE 4
scaffold03264	197824	T	G	Chr.14	cassava4.1_012618m	oxidation-reduction process	21.64	cassava4.1_012402m	REDUCED SUGAR RESPONSE 4

scaffold03264	197826	C	T	Chr.14	cassava4.1_012618m	oxidation-reduction process	52.67	cassava4.1_019051m	Dehydration-induced protein ERD15
scaffold03264	202267	C	T	Chr.14	cassava4.1_012618m	oxidation-reduction process	37.00	cassava4.1_015013m	L-ascorbate peroxidase
scaffold03264	204700	A	G	Chr.14	cassava4.1_004630m	ATP biosynthetic process	21.76	cassava4.1_013461m	L-ascorbate peroxidase
scaffold03264	204700	A	G	Chr.14	cassava4.1_004630m	ATP biosynthetic process	31.74	cassava4.1_013461m	L-ascorbate peroxidase
scaffold03834	121307	T	C	Chr.05	cassava4.1_011028m	N/A	51.80	cassava4.1_006297m	Catalase 2
scaffold04767	103042	G	C	Chr.10	cassava4.1_014213m	translation	21.78	cassava4.1_014643m	L-ascorbate peroxidase
scaffold06303	71034	C	A	Chr.04	cassava4.1_002678m	nucleic acid binding	52.28	cassava4.1_016240m	RESPONSIVE TO DESSICATION 22
scaffold06591	246482	C	T	Chr.09	cassava4.1_016798m	cell redox homeostasis	29.61	cassava4.1_019777m	Glutaredoxin C2
scaffold06708	368551	A	G	Chr.05	N/A	N/A	20.02	cassava4.1_020192m	RESPONSE TO LOW SULFUR 4
scaffold06890	324667	C	G	Chr.09	cassava4.1_006006m	carotenoid biosynthetic process	130.48	cassava4.1_019973m	Dehydration-induced protein ERD15
scaffold07478	225931	G	A	Chr.14	cassava4.1_016206m	N/A	30.328	cassava4.1_014643m	L-ascorbate peroxidase

*N/A means not available

6.3.5 Paralogous *trans*-acting regulator loci co-correlate with the same target gene

Seven hundred of the 5891 putative *trans*-acting SNP loci were selected to investigate co-correlation, of loci located in gene paralogues, to same target gene. Allele expression at these loci were the best correlated with gene expression of the SAG target. Thirty six pairs of the best correlated 700 loci were determined to be paralogous because container-genes for each of these pairs coded for duplicate genes or genes of same family. Allele expression 14 of the 36 pairs, were found to always correlate with expression of same target gene (Table 21). These 14 locus pairs were correlated to only four susceptibility-associated target genes. Two of these – galactinol synthase and thiazole biosynthetic enzyme – correlated with multiple locus pairs. For instance, galactinol synthase was correlated with seven loci four of which were ribosomal proteins (Table 21).

Table 21. Identity and genomic locations of 14 paralogous SNP locus pairs and susceptibility-associated genes correlated to them.

			Locus paralogue 1				Locus paralogue 2			
Container gene name of SNP locus	Gene ID of correlated target	Gene name of correlated target	SNP container gene ID	Scaffold location	Base position on scaffold	Chromosome location	SNP container gene ID	Scaffold location	Base position on scaffold	Chromosome location
BEL1-like homeodomain 1	cassava4.1_011321m	Galactinol synthase 4	cassava4.1_003045m	scaffold02892	873345	Chr.09	cassava4.1_002982m	scaffold07340	38091	Chr.11
Calmodulin-like protein 4	cassava4.1_010620m	Thiazole biosynthetic enzyme	cassava4.1_017158m	scaffold02915	308411	Chr.04	cassava4.1_018011m	scaffold02915	290246	Chr.04
Cofactor assembly of complex C	cassava4.1_011321m	Galactinol synthase 4	cassava4.1_013551m	scaffold02688	250408	Chr.11	cassava4.1_017500m	scaffold02998	213464	Chr.10
EMBRYO DEFECTIVE 3119	cassava4.1_011321m	Galactinol synthase 4	cassava4.1_027961m	scaffold03614	2936636	Chr.09	cassava4.1_010724m	scaffold06314	182437	Chr.10
HYPERSENSITIVE TO ABA1	cassava4.1_010620m	Thiazole biosynthetic enzyme	cassava4.1_004932m	scaffold00206	428012	Chr.03	cassava4.1_005524m	scaffold05875	619406	Chr.04
Indeterminate(ID)-domain 7	cassava4.1_010620m	Thiazole biosynthetic enzyme	cassava4.1_004983m	scaffold05875	506836	Chr.04	cassava4.1_005370m	scaffold00206	174275	Chr.03
Pentatricopeptide repeat-containing protein	cassava4.1_020059m	Drought-induced 21	cassava4.1_028274m	scaffold04795	4257	Chr.17	cassava4.1_003167m	scaffold03880	6252	Chr.17
Phospholipase C	cassava4.1_010620m	Thiazole biosynthetic enzyme	cassava4.1_005538m	scaffold06916	1110741	Chr.02	cassava4.1_005145m	scaffold00506	27665	Chr.04
2-alkenal reductase	cassava4.1_006297m	Catalase 2	cassava4.1_010863m	scaffold03404	120623	Chr.01	cassava4.1_010863m	scaffold03404	122434	Chr.01
60S ribosomal protein L12-1	cassava4.1_010620m	Thiazole biosynthetic enzyme	cassava4.1_017811m	scaffold06327	207493	Chr.10	cassava4.1_016365m	scaffold01945	415450	Chr.09

60S ribosomal protein L21-2	cassava4.1_011321m	Galactinol synthase 4	cassava4.1_017889m	scaffold06591	28480	Chr.06	cassava4.1_018897m	scaffold05703	60974	Chr.05
60S ribosomal protein L27-3	cassava4.1_011321m	Galactinol synthase 4	cassava4.1_018897m	scaffold05703	60974	Chr.05	cassava4.1_020770m	scaffold06711	51415	Chr.16
60S ribosomal protein L38	cassava4.1_011321m	Galactinol synthase 4	cassava4.1_020770m	scaffold06711	51379	Chr.16	cassava4.1_016807m	scaffold01663	33112	Chr.15
R-protein L3 B	cassava4.1_011321m	Galactinol synthase 4	cassava4.1_009397m	scaffold02538	301944	Chr.12	cassava4.1_033711m	scaffold02538	314449	Chr.12

6.3.6 Prediction of putative *trans*-regulators of uniquely overexpressed genes of resistance-associated genes (RAGs)

Twenty two potential *trans*-acting predictor loci were predicted for five out of 35 genes overexpressed in CBSV-resistant relative to susceptible and susceptible accessions at feature importance cut-off of 15%. Expression levels of the five RAGs were compared in three sample categories namely samples expressing homozygous reference, homozygous alternate and heterozygous alleles of the putative *trans*-acting RAG regulator loci best correlated to the RAGs. As expected, allele sample categories for each putative RAG regulator locus were found to express different levels of the correlated resistance-associated gene(s) (Figure 37). The prevalence of CBSV resistance in each of the sample categories corresponded to average gene expression levels of the categories (Figure 37). The 22 resistance-associated loci were located within genes coding for diverse functions but particularly enriched with DNA binding and/or general nucleic acid binding functions (Table 22). There were 16 loci correlated with a UDP-glycosyl*trans*ferase-like protein. Interestingly, container-genes for the four best correlated loci of these include the three mRNA splicing genes – snRNA-associated SM-like protein 8 (LSM8), alternative splicing factor SRp20/9G8 and DEAD-box ATP-dependent RNA helicase (Table 22).

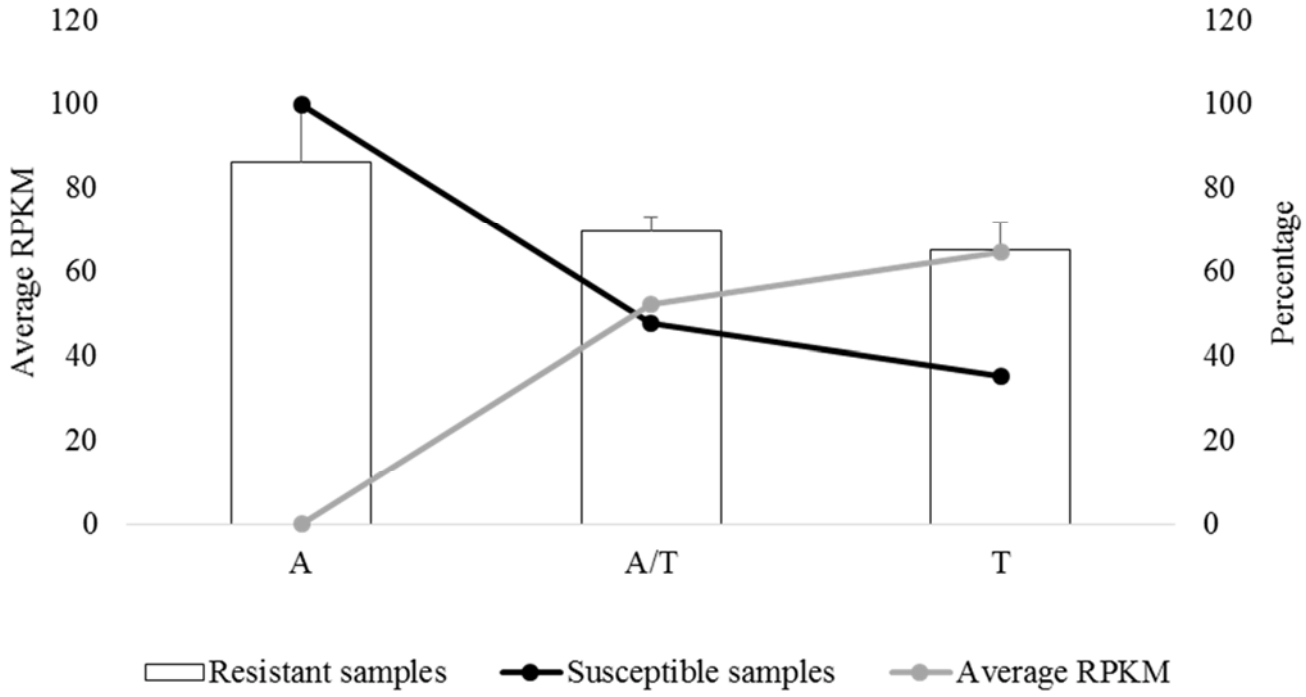


Figure 37. Bar plot in a combined bar-line plot showing average expression levels of a malate dehydrogenase resistance-associated gene in cassava genotype sample groups expressing two homozygous allele forms and a heterozygous allele of a locus correlated to the target gene. Line plots show the percentage prevalence of CBSV resistant and susceptible individuals in each sample group.

Table 22. Genomic locations, container-gene description and importance values of SNP loci correlated to six resistance-associated genes.

Container gene ID of SNP locus	SNP reference allele	SNP alternate allele	Container gene description of SNP locus	Chromosome location of SNP	Resistance-associated target gene description	Percent IncMSE	Scaffold location of SNP locus	Base-pair position of SNP locus
cassava4.1_020011m	C	G	U6 snRNA-associated SM-like protein LSM8	Chr.02	UDP-glycosyltransferase-like protein	76.4	scaffold06916	1457267
cassava4.1_023743m	C	T	Alternative splicing factor SRp20/9G8 (RRM superfamily)	Chr.07	UDP-glycosyltransferase-like protein	63.4	scaffold04002	250300
cassava4.1_017479m	G	C	Cyanase	Chr.13	UDP-glycosyltransferase-like protein	61.7	scaffold00085	189814
cassava4.1_000565m	C	G	DEAD-box ATP-dependent RNA helicase 42	Chr.07	UDP-glycosyltransferase-like protein	43.2	scaffold06914	1312334
cassava4.1_009957m	T	C	Mitogen-activated protein kinase 4	Chr.01	UDP-glycosyltransferase-like protein	42.7	scaffold02658	732329
cassava4.1_005064m	C	T	Clathrin assembly protein	Chr.06	UDP-glycosyltransferase-like protein	31.4	scaffold05005	196743
cassava4.1_008817m	A	T	Pyruvate dehydrogenase E1 component subunit beta-2	Chr.15	UDP-glycosyltransferase-like protein	31.1	scaffold04251	510810
cassava4.1_020276m	G	A	Transcription elongation factor 1-like protein	Chr.17	UDP-glycosyltransferase-like protein	29.0	scaffold04851	622958
cassava4.1_000154m	A	C	BAH AND TFIIS domain-containing protein	Chr.14	UDP-glycosyltransferase-like protein	23.8	scaffold04083	370571
cassava4.1_011701m	C	G	PEPTIDYL-PROLYL *CIS-TRANS* ISOMERASE PASTICCINO1	Chr.15	UDP-glycosyltransferase-like protein	22.2	scaffold07069	259893
cassava4.1_009967m	C	A	ZINC finger protein constans-like 13	Chr.01	UDP-glycosyltransferase-like protein	22.2	scaffold04457	773922

cassava4.1_020553m	A	C	NA	NA	UDP-glycosyltransferase-like protein	19.8	scaffold06816	21006
cassava4.1_018615m	C	G	Calvin cycle protein CP12-3	Chr.05	UDP-glycosyltransferase-like protein	18.3	scaffold01012	27521
cassava4.1_001819m	T	A	CHIP	Chr.13	UDP-glycosyltransferase-like protein	17.0	scaffold00853	125028
cassava4.1_007682m	A	C	3-dehydroquinate synthase	Chr.02	UDP-glycosyltransferase-like protein	16.0	scaffold06512	1080082
cassava4.1_002360m	T	C	Ferric-chelate reductase (NADH)	Chr.06	UDP-glycosyltransferase-like protein	15.5	scaffold03428	5223
cassava4.1_004565m	A	G	Yeast SPT2-related	Chr.12	(S)-coclaurine-N-methyltransferase-related	18.5	scaffold03823	47637
cassava4.1_010703m	A	T	RING/U-BOX domain-containing protein	Chr.14	Malate dehydrogenase	19.8	scaffold01662	5155
cassava4.1_033369m	A	G	Uncharacterized conserved protein	Chr13	Leucine-rich repeat-containing protein	16.5	scaffold06292	268129
cassava4.1_006974m	T	C	Nucleoredoxin 1-related	Chr.11	Organic cation/carnitine transporter 4	32.8	scaffold03750	606190
cassava4.1_017055m	A	G	ZINC finger A20 and AN1 domain-containing stress-associated protein 5	Chr.10	Elongation factor 1-beta 1-related	42.0	scaffold06407	217442
cassava4.1_020824m	A	C	NA	Chr.01	Elongation factor 1-beta 1-related	56.0	scaffold05350	6943

6.4 Discussion

6.4.1 Relative higher heterozygosity of CBSD-resistant cassava and identification of introgressed CBSD resistance marker alleles

The observed greater similarity in heterozygous locus expression within the CBSD-resistant accessions – Kaleso, Mkumba and Pwani – compared to other accession pairs is consistent with the fact that Mkumba and Pwani are derivatives of cross between Kaleso and Latin American variety/varieties (G. Mkamillo, Personal communication). The high similarity between Mkumba and Pwani, not just in their heterozygous locus expression, but also in their homozygous locus expression suggests that they could be siblings derived from a single cross. Kaleso on the other hand is a third backcross generation of *M. glaziovii* (Hillocks and Jennings, 2003) implying that all three accessions share a common ancestor. It further implies that the CBSD-resistance response of these accessions is encoded by genetic factors commonly inherited from their shared *M. glaziovii* ancestor. Attempt was made at uncovering candidates for these genetic factors from a set of loci uniquely expressed as heterozygous in each of Kaleso, Mkumba and Pwani by comparing this locus set with genetic variation and genome location data of *M. glaziovii* loci introgressed into *M. esculenta* (Bredeson *et al*., 2016). Interestingly, distribution of chromosomal locations of the unique heterozygous locus sets of Kaleso, Mkumba and Pwani shows that these loci are predominantly located within the chromosomes 1, 3, 4, 13 and 14. These are same chromosomes which contain haplotypes introgressed into Kaleso (also known as Namikonga) from *M. glaziovii* (Bredeson *et al*., 2016). The over-representation of the unique heterozygous loci of CBSD-resistant accessions in the above-mentioned chromosomes and the fact that they harbour *M. glaziovii* haplotypes introgressed into Kaleso implies an increased likelihood of finding CBSD-resistant variations from them. Unique heterozygous loci of the resistant trio (Kaleso, Mkumba ad Pwani) expressing common *M. glaziovii*-derived alleles were located within short segments of chromosomes 3, 4 and 13 (Figure 31). The identified short segments bearing potentially CBSD-resistant alleles were different from the proposed genome co-ordinates from which CBSD resistance alleles could be mined (Bredesson *et al*., 2016; Figure 32). Furthermore they represent a substantial reduction in the size of these co-ordinates. This reduction will simplify and reduce the cost and time for search and validation of CBSD resistance loci. Given the limited number of varieties used in the current study, it would be necessary to confirm the absence of these alleles in more susceptible accessions and their presence in more *M. glaziovii*-derived resistant accessions in

order to affirm their uniqueness to the latter, hence their usefulness as markers for CBSD resistance.

It is interesting that the common *M. glaziovii* alleles were found within genes enriched with abiotic stress response and resistance response functions (Appendix 9). The abiotic stress response function is encoded by a MYB transcription factor which is an ABA-induced repressor of protein phosphatase 2C which are largely inhibitors of ABA stress response (Fuchs *et al.,* 2012). Given that susceptibility response to CBSV infection in cassava was found associated with increased abiotic stress response gene expression (Chapter 5), the *M. glaziovii* alleles of the MYB transcription factor in CBSV-resistant cassava might be the basis for the lack of abiotic stress response gene expressions in CBSV-resistant cassava accessions. In contrast, *M. glaziovii* alleles of the C-S lyase gene (SUPEROOT 1; cassava4.1_008450m) might facilitate the auxin homeostatic function of this gene in CBSV-resistant accessions. This is owing to the observation in the previous chapter of the higher average auxin response gene induction in susceptible accessions compared to resistant ones. Genes which function within other biological processes differently modulated in CBSD-resistant and -susceptible accessions (Chapter 5) contained the common *M. glaziovii* alleles of chromosome 4 loci. This included the pathogenesis-related thaumatin-like protein and L-ascorbate peroxidase which is involved in plant antioxidant response (Caverzan *et al*., 2012). It also included 4-hydroxy-3-methylbut-2-enyl diphosphate synthase and E3 SUMO-protein ligase SIZ1 with roles in SA-mediated disease resistance (Gil *et al*., 2005; Lee *et al*., 2007). A mutant of the former shows strikingly enhanced resistance to biotrophic pathogens in Arabidopsis (Gil *et al*., 2005). The *M. glaziovii* alleles of this gene might enhance SA-mediated resistance to CBSV in resistant accessions inheriting the alleles. Putative roles in CBSV-cassava interaction for genes containing some of the *M. glaziovii* alleles, in the context of their known functions and those of transcriptionally unique genes of resistant and susceptible cassava, are illustrated in a proposed molecular mechanisms which underlie resistance and susceptibility responses to CBSV infection (Figure 38).

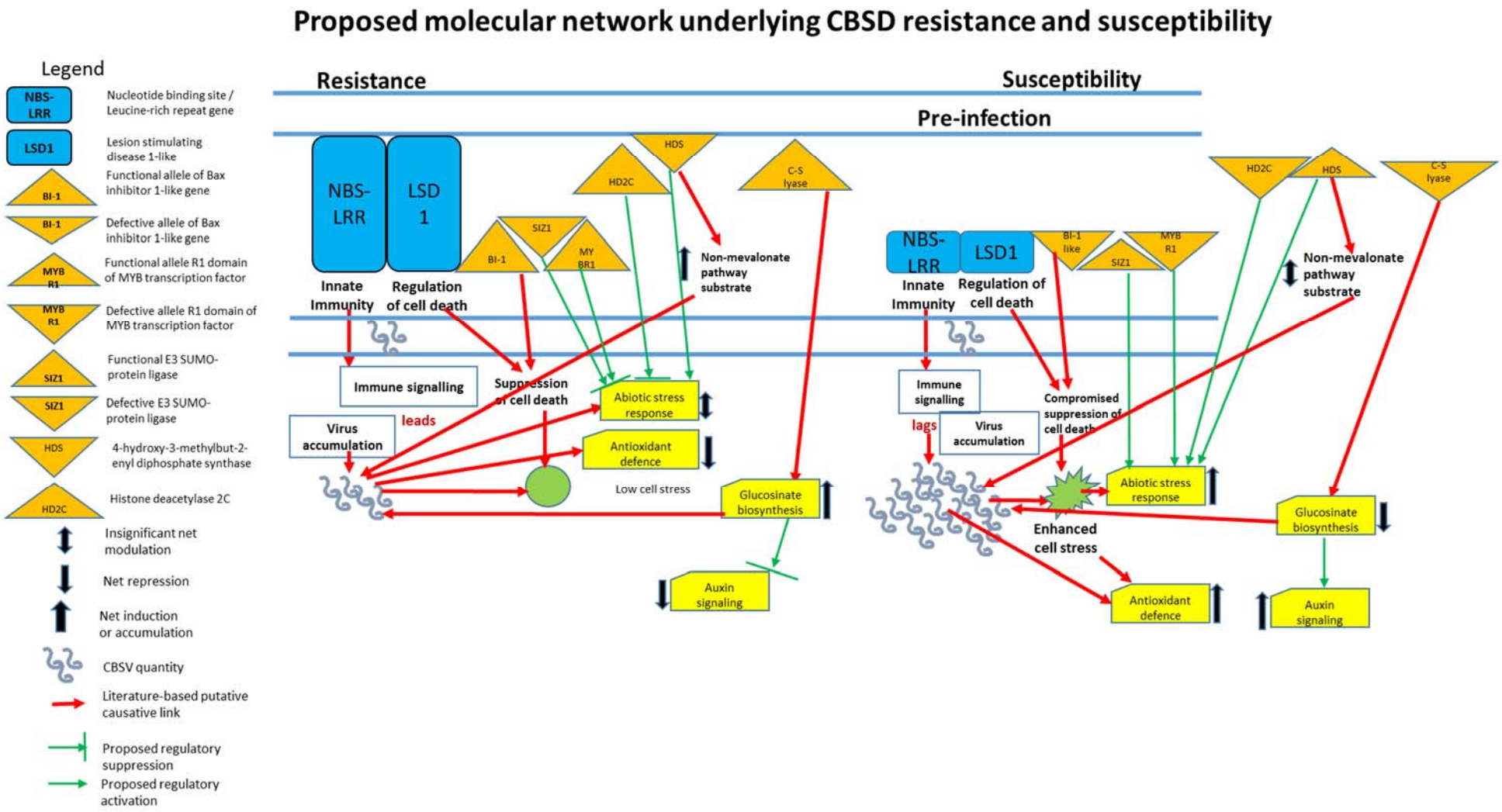


Figure 38. Illustrated mechanisms underlying resistance and susceptibility responses to CBSV infection, based on combination of gene and allele expression data and literature information on gene functions.

Given their predicted roles in disease resistance pathways, functional similarity to biological processes differentially modulated in CBSV-infected resistant and susceptible accessions and the fact that the three CBSD-resistant accessions expressing them were bred from *M. glaziovii*, a role for the common *M. glaziovii* alleles as sources of genetic variations underlying resistance response is highly likely. Genes containing these candidate resistance alleles are different from those found within the common *M. glaziovii* haplotypes in chromosomes 1 and 4 of the six Amani cassava varieties – Kaleso, TM-I30572, KBH 2006/18, Mkombozi, TMS-I972205 and Akena – hypothesized by Bredeson *et al*. (2016) to likely include CBSD/CMD resistance genes. Given that only Kaleso of the six Amani varieties have a known durable CBSD resistance, this trait might be encoded in other introgressed genome segments different from those advocated by Bredeson *et al*. (2016). Indeed, the same study cautioned that other introgressed segments might be 'implicated' in the differential disease resistance among the Amani cultivars.

6.4.2 Differential allele expression between CBSV-resistant and -susceptible cassava

Given the very few number of genetically distinct accessions used in the statistical association test that identified the 4543 loci expressing significantly different alleles between CBSD-resistant and non-resistant accessions, further validation in a large population cross or population of diverse cassava landrace collection would be required before any of these alleles can be considered truly associated with CBSD resistance response. At best, they can represent a large 'cache' from which potentially candidate CBSD resistance alleles could be mined. Gene function evidence suggests that many alleles of the 4543 locus set might be involved in resistance response to CBSV infection. Most of these loci are contained in nucleic acid and protein-interacting genes enriched with gene transcription and mRNA translation functions exploited by plant viruses for their replication. For instance, isoforms of the eIF4G which normally facilitates mRNA circularisation and mRNA association with the 43S pre-initiation complex during translation, (Sanfaçon, 2015) have been shown to be indispensable to achieving susceptibility in *Arabidopsis* plants infected with other plant RNA viruses namely *Plum pox virus* (PPV) and *Turnip mosaic virus* (TuMV). Isoform of eIF4G also interacts directly with RNA of the non-potyvirus *Rice yellow mottle virus* (RYMV) and is the basis of natural recessive resistance to this virus (Hébrard *et al*., 2010). Finally, ATP-dependent RNA helicases which has the normal physiological role of unwinding RNA (Sanfaçon, 2015) has

been implicated in *trans*lation of RNA2 of RNA-dependent RNA polymerase of *Brome mosaic virus* (BMV) and replication of this virus (Sanfaçon, 2015).

6.4.3 Allele specific expression and prediction of *trans*-regulators of resistance associated genes (RAGs) and susceptibility associated genes (SAGs)

Gene expression can be regulated in *cis* by genetic variations in proximity to the regulated gene, such as in upstream pre-start codon or downstream post-stop codon sequences, or in *trans* by distal action of proteins encoded in different genomic locations. Regulation according to the former mechanism can cause imbalance in expression of maternal and paternal copies of a gene in a homogenous genetic background, if the *cis* variation is heterozygous (Albert and Kruglyak, 2015). This imbalance in gene expression can be indicated by imbalanced expression of alleles within the gene (Guo *et al*., 2008; Zhang *et al*., 2009). Cases of allele expression imbalance or ASE observed in this study could therefore indicate that their container genes are *cis*-regulated. The range of all allele-specific or potentially *cis*-regulatory events observed in samples of cassava accession studied here are comparable to same events already reported in other plant species. For instance, a genome-wide study of allele specific expression in *Arabidopsis thaliana* identified 14% frequency for *cis*-regulatory loci (Zhang *et al*., 2009). While allele expression imbalance is relatively well known in plants, information on ASE triggered or abolished by plant virus infection or infection by other plant pathogens is virtually unavailable. The definition of *cis*-regulated loci as those loci with imbalanced allele expression, as was used in Guo *et al*. (2008) and Zhang *et al*. (2009), presupposes that expression at such loci – balanced or imbalanced - is always tied to proximally-located genetically or epigenetically-controlled *cis*-regulatory locus or loci. Assuming a genetic control of allele expression, a different pattern of allele expression at a *cis*-regulated locus can only be possible in a genetically different individual. From the fore-going, the observed induction and abolishment of ASE in CBSV-infected cassava cannot be said to be genetically controlled but more likely epigenetically determined. Although virus induction of ASE is not currently known in plants, epigenetic modification (such as acetylation) as a plant response to abiotic stress and to virus infection is known (Meyer, 2014; Boyko and Kovalchuk, 2011). There are differing views on the role of epigenetic modification in plant interaction with biotic and abiotic stresses. One of these is that these modifications are a first step in the adaptation of plants to stress through prolonged exposures. According to this view, prolonged exposure to stress convert transient epigenetic changes to stably-inherited epigenetic or genetic changes which result in resistance to the stress condition (Boyko and Kovalchuk, 2011). In an opposing theory, it was argued that the

continuous stress-responsive pathway induction, which is expected for stably-inherited epigenetic or genetic control of stress resistance, is undesirable due to the associated energy costs (Meyer, 2014). Instead in cases where an epigenetic change results in secondary response, more rapid but transient epigenetic response to frequently recurring stress is more useful (Meyer, 2014). It is noteworthy that the observed modulations of allele expression quantities by CBSV infection vary a lot in a single accession samples in different times indicating that it is transient. The transient nature of virus-induced allele expression modulation suggests that the observed induction/abolishment of ASE could be controlled by reversible and transient epigenetic changes as described in Mayer (2014). In contrast, the relatively small frequency of consistent ASE observed in virus- and mock-inoculated samples of each accession and at all three times post-inoculation (Table 19) are likely regulated by stably-inherited genetic or epigenetic *cis*-factors as described in Boyko and Kovalchuk (2011).

SNP loci correlated with RAGs are enriched with nucleic acid binding functions which include transcription regulation. Genetic basis for innate over-expression of RAGs in CBSV-resistant genotypes could therefore be based on differential allele expression in genes involved in transcription and transcription regulation of RAGs. On the other hand, correlation of allele expression at *trans*-SAG regulator loci with expression of susceptibility-associated abiotic stress response, antioxidant and PR genes may not indicate a direct role for the former in transcribing or regulating transcription of latter. This is because container-genes for the *trans*-acting susceptibility-associated loci are not enriched for genes coding for RNA polymerase, transcription factor or other *trans*cription regulator function. However given the relatively high importance measures (%IncMSE values sranging from 20% to over 250%) for correlation between the SAGs and *trans*-SAG regulator loci, it is expected that some, as of yet unknown, regulatory interaction, could exists between alleles at these loci and the genes they are correlated with.

6.4.4 Paralogous *trans*-acting regulator loci co-correlate with the same target gene

Predicted functions for the paralogous container-genes of the locus pairs does not suggest a direct transcriptional regulatory role for these loci on their correlated gene targets given that these functions are not related to transcription regulation. Rather, they suggest indirect regulation of target genes by paralogous loci which share same biological pathway as their targets (Table 21). For instance, the paralogues ‘cassava4.1_008960m’ and ‘cassava4.1_011366m’ both of which code for glyceraldehyde-3-phosphate dehydrogenas contain two different SNP loci both of which are correlated to galactinol synthase 4 gene.

Glyceraldehyde-3-phosphate dehydrogenase and galactinol synthase 4 have known roles in abiotic stress response pathways. Cytosolic glyceraldehyde-3-phosphate dehydrogenase is known to positively regulate reactive oxygen species (ROS) and abs*cis*ic acid (ABA)-induced stomatal closure during drought response (Guo *et al*., 2012) while galactinol synthase on the other hand is known to be essential for drought stress tolerance (Taji *et al*., 2002). Relatedly, gene expression of the thiazole synthase gene (cassava4.1_010620m) which regulate ABA-mediated stomatal closure during drought (Li *et al*., 2016) is correlated with allele expression of two HYPERSENSITIVE TO ABA1 (protein phosphatase 2C) which also plays a role in drought tolerance though as negative regulator (Saez *et al*., 2006).

The observed co-correlation of allele expression at a pair of loci contained in paralogous genes with expression of the same target gene suggests that such paralogous genes share common regulatory elements and are co-regulated by a common *trans*-acting regulator or a set of functionally similar *trans*-regulators belonging to a common pathway. This regulatory mechanism is consistent with the different chromosomal locations of 10 of the 14 paralogous locus pairs identified in this study (Table 21). The observed co-correlation could also suggest, in the case of the other 4 locus pairs co-located on the same chromosome, that the paralogous loci are co-ordinately regulated from a common *cis*-acting regulator. Adjacent gene pairing has been associated to tight co-regulation of genes coding for ribosomal proteins and rRNA and ribosomal biosynthetic proteins in diverse eukaryotes including different yeast species and *A. thaliana* (Arnone *et al*., 2012). Particularly, majority of co-regulated ribosomal protein genes were found to exist as adjacent homologous but non-duplicate gene pairs in *S. cerevisiae* (Arnone *et al*., 2012). Though more cases of non-adjacent than adjacent co-regulated genes was observed in the current study, no attempt was made here to identify the more prevalent locational relationship between co-regulated genes given that only a small proportion – 700 of the 33,832 predicted locus-gene interactions was investigated for co-correlation of paralogous loci. One clear point though is the fact that a single regulatory function is in essence predicted twice for each target gene simultaneously correlated to two paralogous loci. This 'twice-prediction' would increase confidence in prediction of regulatory loci for expressed genes.

6.5 Conclusions

The current study characterized for the first time, allele expression in cassava. Characterization of allele expression enabled:

1. The first ever description of ASE and CBSV-induced ASE which shed light on the *cis*-regulatory landscape of the cassava genome.
2. Identification of genetic variations potentially regulating unique gene expression and modulations underlying resistance and susceptibility responses to CBSD.
3. Discovery of specific alleles, introgressed from the prime source of the most CBSD resistance cultivars of cassava, which are likely the genetic basis of natural resistance to the disease.

This knowledge would be useful for development of methods for sustainable protection of cassava from the menace of CBSD via molecular breeding and for targeted genetic fortification of the crop by genetic engineering means. There is however, need to validate the observed SNP associations to CBSD resistance and susceptibility in population crosses. Crucially, functional studies targeting predicted gene expression regulators and introgressed alleles are expected to in uncover the complex molecular architecture underlying CBSD resistance

CHAPTER 7: General Discussions

In east Africa, cassava production is severely limited by cassava brown streak disease (CBSD) – a disease caused by the two RNA viruses *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) of the genus *Ipomovirus* family *Potyviridae*. Apart from east Africa, CBSD has also been reported in Central Africa (Mulimbi *et al*., 2012). It is believed to be spreading west-ward and could threaten cassava production in Nigeria which is the world's largest producer of the crop. Intensive research effort is currently underway within the cassava research community, to tackle this disease. Recently, sources of field resistance to CBSD have been identified (Kaweesi *et al*., 2014; Ogwok *et al*., 2014). Also, transcriptional response to late infection of the virus in the known resistant source Kaleso as well as genomic regions of this accession inherited from the CBSD-resistant wild Manihot species called *Manihot glaziovii*, have been characterized (Maruthi *et al*., 2014b). However, transcriptional responses to earlier stages of CBSD infection as well as an understanding of the molecular processes underlying resistance and susceptibility responses are lacking. Also, available genomic studies on cassava have not explained the genetic basis or identified functional markers of resistance to the disease. This project was therefore aimed at contributing to the control of CBSD through identification of CBSD-resistant sources and understanding of molecular mechanisms underlying natural resistance to the virus. To achieve this aim, five research objectives were formulated. They were as follows:

1. Develop new and efficient techniques for DNA and RNA virus detection to screening for combined field resistance to CMD and CBSD.
2. Identify new sources of resistance to CBSD among cultivated landraces of east Africa.
3. Identify genes and biological processes associated with resistance and susceptible responses to CBSV infection.
4. Identify putative functional markers for CBSD resistance based on allele-specific gene expression analysis.

Under the first objective, an efficient and sensitive multiplex real time PCR method was developed which quantified four DNA and RNA viruses of cassava – *African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV), CBSV and UCBSV in a single tube without diminished sensitivity. Real time PCR tools had been developed for quantifying

CBSV and UCBSV individually (Kaweesi *et al*., 2014; Ogwok *et al*., 2014) and simultaneously (Adams *et al*., 2013). However, the multiplex virus quantification tool developed under this objective is the first to combine quantification of RNA and DNA viruses of cassava using the real time PCR technique. Given its high sensitivity – detection limit of 4 to 12.5 femtograms of virus DNA and RNA – and ability to detect multiple viruses in a single qPCR reaction, the tool would be very useful for diagnostic applications requiring sensitive and high-throughput virus detection such as virus indexing of tissue-cultured cassava cultivars. It would also facilitate identification of sources of combined resistance – an important trait considering the often co-occurrence of CBSD and CMD in fields in east Africa (Rwegasira and Rey, 2012; Masinde *et al*., 2016) and the potential of enhanced CBSD symptom expression, in mixed infection, due to synergy between both diseases (Irungu, 2011). The tool also identified 12 accessions of cassava with low CBSD severity or no disease incidence. These accessions can be considered potentially CBSD resistant, although an assessment of their virus accumulation levels in multiple environments and a confirmation of their responses to controlled inoculation of the CBSV will be required.

For the second objective, controlled CBSV inoculation of 13 cassava accessions identified three major categories of responses to CBSD infection based on disease severity and CBSV quantity accumulation over 9 months post-inoculation in leaves and at 9 months post inoculation in roots. These categories were termed resistant, intermediate and susceptible. Accessions with previously unknown response to CBSV inoculation – Mkumba, Pwani and Nase 3 were found to show resistance to CBSV infection by graft inoculation. Mkumba and Pwani are derivatives of the known resistant accession Kaleso (G. Mkamilo, personal communication) which alongside Nase 3 has the wild *Manihot glaziovii* as a common ancestor (Hillocks and Jennings, 2003; Bredeson *et al*., 2016). Identification of these new resistant sources expands the diversity of resistance against CBSV hence ensuring the control of the disease without restricting farmer options on varieties to grow. The use of more resistant sources in study designs aimed at understanding molecular mechanisms underlying the CBSD resistance response will increase the robustness of such studies by virtue of the increased confidence associated with treatment groups having more biological samples as opposed to less.

Under the third objective, RNA sequencing of selected CBSD-resistant and -susceptible cassava identified nucleotide binding site / Leucine-rich repeat (NBS-LRR) genes as the most common of genes overexpressed in CBSD-resistant cassava relative to susceptible ones. They were expressed at least twice higher in each of the resistant accessions – Kaleso, Mkumba,

Pwani and Nase 3 compared to average expression in the intermediate (Oekhumelela and Kiroba) and susceptible accessions (Albert and Kalawe). However, their CBSV-induced modulation patterns of the overexpressed genes did not correspond to CBSD resistance or susceptibility suggesting that CBSV-induced gene expression modulation of NBS-LRR genes might not be a mechanism resistance to CBSD in cassava. A combination of the evidence of relative overexpression of NBS-LRR genes in resistant cassava obtained here and the well-known roles of these genes in immunity against diverse plant pathogens (Spoel and Dong 2012; Mandadi and Scholthof, 2013) suggests that CBSD resistance could be associated with an innately more acute perception of and higher immunity against invading CBSV.

Among other gene groups, CBSD susceptibility corresponded with differential expression response of genes encoding abiotic stress response and antioxidant defence functions to CBSV infection. Abiotic stress response genes were induced at higher levels in CBSD susceptible accessions compared to resistant cassava while genes with the latter function were oppositely modulated in both accession groups. Given the role of antioxidant defence genes in control of oxidative stress (Tausz *et al*., 2004), the observed induction and repression of these genes in susceptible and resistant accessions respectively implies that adaptation to virus-induced oxidative stress is a mechanism of susceptibility response to CBSV infection. Net induction of abiotic stress response genes following CBSV inoculation of resistant and susceptible cassava is suggestive of virus-induced cellular stress in both accession groups given the known roles of these genes in adaptation and tolerance to various abiotic stress factors (Rejeb *et al*., 2014). However, this induction was over 10 times higher in susceptible and intermediate response accessions. The higher induction of these genes is indicative of a stronger adaptive response to higher cellular stress levels induced by higher CBSV accumulation in susceptible and intermediate response accessions compared to the resistant ones.

Put together, relative gene overexpression and modulation results for the sequenced cultivars suggests that innately higher activity of R genes restricts CBSV accumulation, moderating cellular stress in resistant compared to susceptible accessions where induction of cellular stress from accumulating virus triggers a natural transcriptional induction of stress response and antioxidant defence which averts cell death hence preserving the intracellular environment for continued virus accumulation and renewed stress induction. In essence, a mutual feedback loop between CBSV induction of cellular stress and the adaptive response to this stress is essential for sustained CBSV accumulation characteristic of susceptibility response to the virus. To my knowledge, this is the first model of its kind proposed for the transcriptional mechanism of

susceptibility to any virus infection of cassava and indeed to other plant viruses. Functional and population genetic studies targeting genes involved in the mechanisms described above as well as natural variations contained within, contiguous to or associated with these genes would streamline development of tools for molecular breeding of resistance and engineered CBSD-resistant sources. Different functional profile might be obtainable for genes which respond earlier to the virus. Similar studies designed to capture gene expression responses from 12 hours post-infection could enable a more comprehensive understanding of the molecular processed underlying resistance and susceptibility responses to CBSV.

The final two objectives of the project focused on the identification of genetic loci which potentially underlie resistance and susceptibility responses to CBSV. This was pursued through prediction of transcriptional regulation of key resistance and susceptibility response genes and study of genetic variations introgressed into the resistant accessions from the CBSD- and CMD-resistant wild species *M. glaziovii* using allele expression data. *M. glaziovii* alleles (Bredesson *et al*., 2016) expressed exclusively in the resistant accessions – Kaleso, Mkumba and Pwani were identified. These alleles were all located on heterozygous loci and were mostly found in genes with predicted adaptive stress response, antioxidant defence and disease resistance functions which were either net induced in susceptible but net repressed in resistant cassava or exhibit substantially higher upregulation in the former. Given this fact, it is likely that the *M. glaziovii*-derived variations constitute a functional mutation which affects one parental copy of genes containing these variations. This scenario translates into an approximately half the activity of the *M. glaziovii* allele-containing antioxidant and abiotic stress genes in resistant compared to susceptible cassava.

Characterization and analysis of allele expression at over 32,000 SNP loci, showed that allele imbalances or allele specific expression (ASE) events occurred at a rate of 10 to 18% in analyzed cassava samples but that most were not reproducible over time. ASE was also induced or abolished by CBSV infection. ASE is indicative of regulation, by contiguous *cis* factors, of the gene containing the imbalanced loci (Albert and Kruglyak, 2015). Characterization of ASE in the cassava accessions sequenced in this study is the first of its kind in the crop and represents a first step towards the potential exploitation of *cis* variations regulating key resistance and susceptibility response genes for genetic improvement of the CBSD resistance trait. Gene expression profiles of uniquely overexpressed genes of CBSD-resistant cassava and select genes encoding biological functions identified as potentially underlying susceptibility response under the third objective, were matched with allele expression profiles of 32,256 loci using the

machine learning technique called random forest. Potential *trans* regulator loci of gene expression were selected from correlated loci. However, most of these loci did not code for transcription factor function hence unlikely to be involved in direct regulation of expression of their correlated genes. *Trans*-locus prediction, from 700 best correlated loci, for unique induced genes of susceptible cassava allowed the discovery of co-correlation of loci contained within paralogous genes to the same target. Interestingly, the co-correlating loci were enriched with abiotic stress response function which is the predominant functional annotation of their gene targets. Identification of co-correlating loci during transcriptional regulator prediction would enhance accuracy of such prediction given that, under the condition of co-correlation, loci of similar function correlate to the same target gene.

In summary, a combination of reliable plant virus inoculation and the latest genomic techniques and genome analysis tools were deployed in this study to identify CBSD resistance sources as well as biological processes, genes and genetic variations potentially underlying resistance and susceptibility response to the disease. Identification of these genes and genetic variations narrows the search for genetic basis of resistance and offers potential targets for more efficient cassava improvement through molecular breeding and genetic engineering. The molecular mechanism proposed for CBSD resistance and susceptibility will advance understanding cassava interaction with CBSV.

CHAPTER 8: References

Abarshi, M. M., Mohammed, I. U., Jeremiah, S. C., Legg, J. P., Kumar, P. L., Hillocks, R. J., and Maruthi, M. N. (2012). Multiplex RT-PCR assays for the simultaneous detection of both RNA and DNA viruses infecting cassava and the common occurrence of mixed infections by two cassava brown streak viruses in East Africa. *Journal of Virological Methods*, *179*(1), 176–84.

Abeca, A., and Kawuki, R. (2012). Progression of Cassava Brown Streak Disease (CBSD) in Infected Cassava Roots in Uganda. *Uganda Journal of Agricultural Sciences*, *13*(1), 45–51. Retrieved from http://www.naro.go.ug/UJAS/Papers/UJAS vol.13.1/UJAS 45.pdf

Adams, I., Abidrabo, P., Miano, D., Alicai, T., Kinyua, Z., Clarke, J., Macarthur, R., Weekes, R., Laurenson, L., Hany, U., Peters, D., Potts, M., Glover, R., Boonham, N. and Smith, J. (2013). High throughput real-time RT-PCR assays for specific detection of cassava brown streak disease causal viruses, and their application to testing of planting material. *Plant Pathology*, *62*, 233-242.

Agrawal, A.F., and Lively, C. M. (2002). Infection genetics: Gene-for-gene versus matching alleles models and all points in between. *Evolution Ecology Research*, 4,79–90.

Akinbade, S., Hanna, R., Nguenkam, A., Njukwe, E., Fotso, A., Doumtsop, A., Ngeve, J., Tenku, S. and Lava Kumar, P. (2010). First report of the East African cassava mosaic virus - Uganda (EACMV-UG) infecting cassava (*Manihot esculenta*) in Cameroon. *New Disease Report*, *21*, 22.

Alabi, O., Kumar, P. and Naidu, R. (2008). Multiplex PCR for the detection of African cassava mosaic virus and East African cassava mosaic Cameroon virus in cassava. *Journal of Virological Methods*, *154*, 111-120.

Alabi, O. J., Kumar, P. L., and Naidu, R. A. (2011). Cassava mosaic disease: A curse to food security in Sub-Saharan Africa. Online. APS*net* Features.

Albar, L., Bangratz-Reyser, M., Hebrard, E., Ndjiondjop, M., Jones, M., and Ghesquie`re, A. (2006). Mutations in the eIF(iso)4G translation initiation factor confer high resistance to Rice yellow mottle virus. *Plant Journal*, *47*,417–426.

Albar, L., Ndjiondjop, M. N., Esshak, Z., Berger, A., Pinel, A., Jones, M., Fargette, D., and Ghesquiere, A. (2003). Fine genetic mapping of a gene required for Rice yellow mottle virus cell-to-cell movement. *Theoretical and Applied Genetics*, *107*, 371–378.

Albert, F. W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, *16*(4), 197–212.

Ali, E., Kobayashi, K., Yamaoka, N., Ishikawa, M., and Nishiguchi, M. (2013). Graft transmission of RNA silencing to non-transgenic scions for conferring virus resistance in tobacco. *PloS One*, *8*(5), e63257.

Alicai, T., Omongo, C., Maruthi, M., Hillocks, R., Baguma, Y., Kawuki, R., Bua, A., Otim-Nape, G. and Colvin, J. (2007). Re-emergence of Cassava Brown Streak Disease in Uganda. *Plant Disease*, *91*, 24-29.

Allie, F., Pierce, E. J., Okoniewski, M. J., and 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.

Alves, A. (2002). Cassava botany and physiology. *Cassava: Biology, Production and Utilization*, 67–89. Wallingford: CABI publishing.

An, D., Yang, J., and Zhang, P. (2012). Transcriptome profiling of low temperature- treated cassava apical shoots showed dynamic responses of tropical plant to cold stress. *BMC Genomics*, *13*, 64.

Anandalakshmi, R., Pruss, G. J., Ge, X., Marathe, R., Mallory, A C., Smith, T. H., and Vance, V. B. (1998). A viral suppressor of gene silencing in plants. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(22), 13079–84.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106.

Anon. (2010). Cassava disease surveillance surveys 2009: Mapping report. Great Lakes Cassava Initiative. http://www.iita.org/c/documentlibrary/get file?uuid=e06fba6a-63b7-40ba-9efa-4fab8a26a8deandgroupId=25357, accessed on 12th June, 2014.

Arnone, J. T., Robbins-Pianka, A., Arace, J. R., Kass-Gergi, S., and McAlear, M. A. (2012). The adjacent positioning of co-regulated gene pairs is widely conserved across eukaryotes. *BMC Genomics*, *13*(1), 546.

Attaran, E., Zeier, T. E., Griebel, T., and Zeier, J. (2009). Methyl salicylate production and jasmonate signaling are not essential for systemic acquired resistance in Arabidopsis. *Plant Cell*, *21*, 954–971.

Auer, P. L., and Doerge, R. W. (2010). Statistical design and analysis of RNA sequencing data. *Genetics*, *185*(2), 405–16.

Ausubel, F. M. (2005). Are innate immune signaling pathways in plants and animals conserved? *Nature Immunology*, *6*(10), 973–9. doi:10.1038/ni1253.

Balagopalan, C. (2002). Cassava utilization in food, feed and industry. *Cassava: Biology, Production and Utilization*, 301–318. Wallingford; CABI publishing.

Bartels, D., and Sunkar, R. (2005). Drought and Salt Tolerance in Plants. Critical Reviews in *Plant Sciences*, *24*(1), 23–58.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). LME4: Linear mixed-effects model using Eigen and S4. R package version 1.1-7, http://CRAN.R-project.org/package=lme4, accessed 2nd December, 2014.

Bayne, E. H., Rakitina, D. V, Morozov, S. Y., and Baulcombe, D. C. (2005). Cell-to-cell movement of potato potexvirus X is dependent on suppression of RNA silencing. *The Plant Journal : For Cell and Molecular Biology*, *44*(3), 471–82.

Beffa, R. S., Hofer, R., Thomas, M., Meins, F., and Friedrich, T. (1996). Decreased Susceptibility to Vira1 Disease of β -1 , 3-Glucanase-Deficient Plants Generated by Antisense Transformation. *The Plant Cell*, *8*, 1001-1101.

Bell, G. D. M., Kane, N. C., Rieseberg, L. H., and Adams, K. L. (2013). RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biology and Evolution*, *5*(7), 1309–1323.

Bendahmane, A., Kanyuka, K., and Baulcombe, D. C. (1999). *The Rx Gene from Potato Controls Separate Virus Resistance and Cell Death Responses*, *11*(5), 781–791.

Bendahmane, A., Kanyuka, K.V. and Baulcombe, D.C. (1997). High resolution and physical mapping of the Rx gene for extreme resistance to potato virus X in tetraploid potato. *Thoeretical and Applied Genetics*, *95*, 153-162.

Bendahmane, A., Querci, M., Kanyuka, K. and Baulcombe, D.C. (2000). Agrobacterium transient expression system as a tool for isolation of disease resistance genes: application to the Rx2 locus in potato. *Plant Journal*, *21*, 73–81.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 289 – 300.

Bigirimana, S., Barumbanze, P., Ndayihanzamaso, P., Shirima, R. and Legg, J. (2011). First report of cassava brown streak disease and associated Ugandan cassava brown streak virus in Burundi. *New Disease Report*, *24*, 26.

Bock, K.R. (1994). Studies on cassava brown streak virus disease in Kenya. *Tropical Science*, *34*, 134-145.

Bouche, N., Lauressergues, D., Gasciolli, V., and Vaucheret, H. (2006). An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO Journal*, *25*, 3347–3356.

Boyko, A., and Kovalchuk, I. (2011). Genetic and epigenetic effects of plant-pathogen interactions: An evolutionary perspective. *Molecular Plant*, *4*(6), 1014–1023.

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., Nauluvula, P., Lebot, V., Ndunguru J., Mkamilo G., Bart, R.S., Setter, T.L., Gleadow, R.M., Kulakow, P., Ferguson, M.E., Rounsley, S., and Rokhsar, D. S. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nature Biotechnology*, *34*(5), 562–570.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*.1, 5–32.

Brigneti, G., Voinnet, O., Li, W. X., Ji, L. H., Ding, S. W., and Baulcombe, D. C. (1998). Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. *The EMBO Journal*, *17*(22), 6739–46.

Brommonschenkel, S.H., Frary, A., Tanksley, S.D. (2000). The broad-spectrum tospovirus resistance gene Sw-5 of tomato is a homolog of the root-knot nematode resistance gene. *Molecular Plant Microbe Interaction*, *13*, 1130–38.

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

Bruun-Rasmussen, M., Møller, I. S., Tulinius, G., Hansen, J. K. R., Lund, O. S., and Johansen, I. E. (2007). The same allele of translation initiation factor 4E mediates resistance against two Potyvirus spp. in Pisum sativum. *Molecular Plant-Microbe Interactions*, *20*(9), 1075–82.

Buchanan-Wollaston, V., Earl, S., Harrison, E., Mathas, E., Navabpour, S., Page, T., and Pink, D. (2003). The molecular analysis of leaf senescence--a genomics approach. *Plant Biotechnology Journal*, *1*(1), 3–22.

Calvert, L. A., and Thresh, J. M. (2002). The Viruses and Virus diseases of Cassava. In R. J. Hillocks, J. M. Thresh, and A. C. Bellotti (Eds.), Cassava: biology, production and utilization, (pp. 237–260). Wallingford: CABI Publishing.

Charron, C., Nicolai, M., Gallois, J.J., Robaglia, C., Moury, B., Palloix, A. and Caranta, C. (2008). Natural variation and functional analysis provide evidence for co-evolution between plant eIF4E and potyviral VPg. *Plant Journal 54*, 56-68.

Childs, A. H. B. (1957). Trials with virus resistant cassavas in Tanga Province, Tanganyika. *East African Agricultural Journal*, *23*, 135–137.

Chisholm S.T., Mahajan S.K., Whitham S.A., Yamamoto M.L., and Carrington J.C. (2000). Cloning of the Arabidopsis RTM1 gene, which controls restriction of long-distance movement of tobacco etch virus. *Proceedings of the National Academy of Sciences USA 97*, 489-494.

Chivasa, S., and Carr, J. (1998). Cyanide restores N gene-mediated resistance to tobacco mosaic virus in transgenic tobacco expressing salicylic acid hydroxylase. *The Plant Cell*, *10*(9), 1489–98.

Chivasa, S., Murphy, a. M., Naylor, M., and Carr, J. P. (1997). Salicylic Acid Interferes with Tobacco Mosaic Virus Replication via a Novel Salicylhydroxamic Acid-Sensitive Mechanism. *The Plant Cell*, *9*(4), 547–557.

Cock, J.H. (1985). Cassava: new Potential for a Neglected Crop. Boulder, Colorado: Westview press, Boulder London. Pp 191.

Conrath, U. (2006). Systemic acquired resistance. *Plant Signaling and Behavior*, *1*(4), 179–84.

Cooley, M.B., Pathirana, S., Wu H.J., Kachroo, P., and Klessig D.F. (2000). Members of the Arabidopsis HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens. *Plant Cell*, *12*, 663-676.

Cui H., Kenichi T. and Parker, J.E. (2015). Effector-Triggered Immunity: From Pathogen Perception to Robust Defence. *Annual Review in Plant Biology 66*, 487-511.

Cutt, J.R., Harpster, M.H., Dixon, D.C., Carr, J.P., Dunsmuir, P. and Klessig, D.F. (1989). Disease response to tobacco mosaic virus in transgenic tobacco plants that constitutively express the pathogenesis-related PR1b gene. *Virology 173*, 89-97.

Decroocq, V., Salvador, B., Sicard, O., Glasa, M., Cosson, P., Svanella-Dumas, L., Garcia, J.A., and Candresse, T. (2009). The determinant of potyvirus ability to overcome the RTM resistance of Arabidopsis thaliana maps to the N-terminal region of the coat protein. *Molecular Plant-Microbe Interactions*, *22*(10), 1302–11.

Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K.D., Carrington, J.C., and Voinnet, O. (2006). Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defence. *Science*, *313*, 68–71.

Delledonne, M., Xia, Y., Dixon, R.A. and Lamb, C. (1998). Nitric oxide functions as a signal in plant disease resistance. *Nature*, *6*, 585 - 588.

Diaz-Pendon, J.A., Truniger, V., Nieto, C., Garcia-Mas, J., Bendah-mane, A. and Aranda, M.A. (2004). Advances in understanding recessive resistance to plant viruses. *Molecular Plant Pathology*, *5*, 223–233.

Ding, S.W., Li, H., Lu, R., Li, F., and Li, W.-X. (2004). RNA silencing: a conserved antiviral immunity of plants and animals. *Virus Research*, *102*(1), 109–15.

Druka, A., Potokina, E., Luo, Z., Jiang, N., Chen, X., Kearsey, M., and Waugh, R. (2010). Expression quantitative trait loci analysis in plants. *Plant Biotechnology Journal*, *8*(1), 10–27. doi:10.1111/j.1467-7652.2009.00460.x.

Duprat A, Caranta C, Revers F, Menand B, Browning KS, *et al*. (2002). The Arabidopsis eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. *Plant Journal*, *32*, 927–934.

Edsgärd, D., Iglesias, M. J., Reilly, S.-J., Hamsten, A., Tornvall, P., Odeberg, J., and Emanuelsson, O. (2016). GeneiASE: Detection of condition-dependent and static allele-specific expression from RNA-seq data without haplotype information. *Scientific Reports*, *6*, 21134.

English, J.J., Mueller, E., Baulcombe, D.C (1996). Suppression of virus accumulation in transgenic plants exhibiting silencing of nuclear genes. *Plant Cell*, *8*, 179-l 88.

Espinoza, C., Medina, C., Somerville, S., and Arce-Johnson, P. (2007). Senescence-associated genes induced during compatible viral interactions with grapevine and Arabidopsis. *Journal of Experimental Botany*, *58*(12), 3197–3212.

Fagard, M. and Vaucheret, H. 2000. Systemic silencing signlal9s). *Plant Molecular Biology*, *43*, 285-293.

FAO (2001). The Global Cassava Development Strategy and Implementation Plan. Document prepared by B.I. Nyoka, T.P. Phillips and R.B. Kagbo. Report of the FAO/WHO expert consultation. FAO, Rome.

FAO (2005). A review of cassava in Africa with country case studies on Nigeria, Ghana, the United Republic of Tanzania, Uganda and Benin. Retrieved from http://www.fao.org/docrep/009/a0154e/a0154e00.HTM

Feng, J., Meyer, C. A., Wang, Q., Liu, J. S., Liu, X. S., and Zhang, Y. (2012). GFOLD: A generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics*, *28*(21), 2782–2788.

Fernández-Calvino, L., Guzmán-Benito, I., del Toro, F. J., Donaire, L., Castro-Sanz, A. B., Ruíz-Ferrer, V., and Llave, C. (2016). Activation of senescence-associated Dark-inducible (DIN) genes during infection contributes to enhanced susceptibility to plant viruses. *Molecular Plant Pathology*, *17*(1), 3–15.

Flor, H.H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathoogy*. *9*, 275-296.

Fondong, V.N, Pita J.S., Rey, M.E.C., de Kochko, A., Beachy, R.N., and Fauquet, C.M., (2000). Evidence of synergism between African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. *Journal of General Virology*, *81*, 287–297.

Food and Agricultural Organization of the United Nations, FAOSTAT database (FAOSTAT, 2013). http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E, accessed on 24th March, 2015.

Fraile, A., and Garcia-Arenal F. (2010). The Coevolution of Plants and Virses: Resistance and Pathogenicity. In Carr J.P., and Loebstein G. (Eds.), *Advances in Virus Research*, 1 - 32. London: Academic Press.

Fu, L., Ding, Z., Han, B., Hu, W., Li, Y., and Zhang, J. (2016). Physiological Investigation and Transcriptome Analysis of Polyethylene Glycol (PEG)-Induced Dehydration Stress in Cassava. *International Journal of Molecular Sciences*, *17*(3), 283.

Fuchs, S., Grill, E., Meskiene, I., and Schweighofer, A. (2013). Type 2C protein phosphatases in plants. *FEBS Journal*, *280*(2), 681–693.

Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2006). Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Current Opinion in Plant Biology*, *9*(4), 436–442.

Gaffney, T., Friedrich, L., Vernooij, B., Negrotto, D., Nye, G., Uknes, S., Ward, E., Kessmann, H. and Ryals, J. (1993). Requirement of salicylic acid for the induction of systemic acquired resistance. *Science*, *261*, 754-756.

Gallego Romero, I., Pai, A. A., Tung, J., and Gilad, Y. (2014). RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biology*, *12*, 42.

Gao, Y., He, X., Wu, B., Long, Q., Shao, T., Wang, Z., Wei, J., and Ding, W. (2016). Time-course transcriptome analysis reveals resistance genes of Panax ginseng induced by Cylindrocarpon destructans infection using RNA-seq. *PLoS ONE*, *11*(2), 1–18.

Garcia-Ruiz, H., Takeda, A., Chapman, E. J., Sullivan, C. M., Fahlgren, N., Brempelis, K. J., and Carrington, J. C. (2010). Arabidopsis RNA-dependent RNA polymerases and dicer-like proteins in antiviral defence and small interfering RNA biogenesis during Turnip Mosaic Virus infection. *The Plant Cell*, *22*(2), 481–96.

Garrido-Ramirez, E. R., Sudarshana, M. R., Lucas, W. J., and Gilbertson, R. L. (2000). Bean dwarf mosaic virus BV1 protein is a determinant of the hypersensitive repsonse and avirulence in Phaseolus vulgaris. *Molecular Plant Microbe Interaction*. *13*, 1184-1194.

Gazo, B. M., Murphy, P., Gatchel, J. R., and Browning, K. S. (2004). A novel interaction of Cap-binding protein complexes eukaryotic initiation factor (eIF) 4F and eIF(iso)4F with a region in the 3'-untranslated region of satellite tobacco necrosis virus. *The Journal of Biological Chemistry*, *279*(14), 13584–92.

Gil, M. J., Coego, A., Mauch-Mani, B., Jordá, L., and Vera, P. (2005). The Arabidopsis csb3 mutant reveals a regulatory link between salicylic acid-mediated disease resistance and the methyl-erythritol 4-phosphate pathway. *Plant Journal*, *44*(1), 155–166.

Guo, L., Devaiah, S. P., Narasimhan, R., Pan, X., Zhang, Y., Zhang, W., and Wang, X. (2012). Cytosolic glyceraldehyde-3-phosphate dehydrogenases interact with phospholipase Dδ to transduce hydrogen peroxide signals in the Arabidopsis response to stress. *Plant Cell*, *24*(5), 2200–2212.

Guo, M., Yang, S., Rupe, M., Hu, B., Bickel, D. R., Arthur, L., and Smith, O. (2008). Genome-wide allele-specific expression analysis using Massively Parallel Signature Sequencing (MPSSTM) Reveals cis- and trans-effects on gene expression in maize hybrid meristem tissue. *Plant Molecular Biology*, *66*(5), 551–563.

Hajimorad, M. R., Eggenberger, A. L., and Hill, J. H. (2005). Loss and gain of elicitor function of Soybean mosaic virus G7 provoking Rsv1-mediated lethal systemic hypersensitive response maps to P3. *Journal of Virology*, *79*, 1215-1222.

Ulrich, H., and Søren, H. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest. *Journal of Statistical Software*, *59*(9), 1-30. http://www.jstatsoft.org/v59/i09/, accessed 15th January, 2015.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, *286*(5441), 950–952.

Harimalala, M., Lefeuvre, P., De Bruyn, A., Tiendre´be´ogo, F., Hoareau, M., Villemot, J., et al. (2012). A novel cassava-infecting begomovirus from Madagascar: Cassava mosaic Madagascar virus. *Archives of Virology*, *157*, 2027–2030.

Hartigan J.A. and Wong M.A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society*, *28* (1), 100-108.

Hayes, A. J., Jeong, S. C., Gore, M. A., Yu, Y. G., Buss, G. R., Tolin, S. A., and Maroof, M. A. S. (2004). Recombination within a nucleotide-binding site/leucine-rich-repeat gene cluster produces new variants conditioning resistance to soybean mosaic virus in soybeans. *Genetics*, *166*, 493–503.

He, X.F., Fang, Y.Y., Feng, L., and Guo, H.S., 2008. Characterization of conserved and novel microRNAs and their targets, including a TuMV-induced TIR-NBS-LRR class R gene-derived novel miRNA in Brassica. *FEBS Letters 582*, 2445–2452.

Hébrard, E., Poulicard, N., Gérard, C., Traoré, O., Wu, H., Albar, L., Fargette, D., Bessin, Y., and Vignols, F. (2010). Direct Interaction Between the Rice yellow mottle virus (RYMV) VPg and the Central Domain of the Rice eIF (iso) 4G1 Factor Correlates with Rice Susceptibility and RYMV Virulence. *The American Phytopathological Society*, *23*(11), 1506–1513.

Hernandez, J. A., Gullner, G., Clemente-Moreno, M. J., Künstler, A., Juhász, C., Díaz-vivancos, P., and Király, L. (2016). Oxidative stress and antioxidative responses in plant-virus interactions. *Physiological and Molecular Plant Pathology*, *94*, 134 - 148.

Herren, H. R., and Neuenschwander, P. (1991). Biological control of cassava pests in Africa. *Annual Review of Entomology*, *362*, 57–283.

Herren, H.R., Neuenschwander, P., Hennessey, R.D., and Hammond, W.N.O. (1987). Introduction and dispersal of Epidinocarisi lopezi (Hym. Encytidae), an exotic parasitoid of cassava mealybug, Phenococcus manihoti (Horn. Pseudococcidae) in Africa. *Agriculture, Ecosystems and Environments*, *19*, 131–144.

Hillocks, R. (2002). Cassava in Africa. In Cassava: biology, production and utilization pp. 41–54. Wallingford: CABI publishing

Hillocks, R. (2005). Final Technical Report: Promotion of control measures for cassava brown streak disease. Natural Resources Institute (NRI), University of Greenwich, London, UK.

Hillocks, R. J., Raya, M. D., Mtunda, K. and Kiozia, H. (2001), Effects of Brown Streak Virus Disease on Yield and Quality of Cassava in Tanzania. *Journal of Phytopathology*, *149*, 389–394.

Hillocks, R., and Jennings, D. (2003). Cassava brown streak disease: A review of present knowledge and research needs. *International Journal of Pest Management*, *49*(3), 225–234.

Hothorn, T., Bretz, F. and Westfall, P. (2008) Simultaneous Inference in General Parametric Models. *Biometrical Journal*, *50*, 346-363.

Huang, T.-S., Wei, T., Laliberté, J.-F., and Wang, A. (2010). A host RNA helicase-like protein, AtRH8, interacts with the potyviral genome-linked protein, VPg, associates with the virus accumulation complex, and is essential for infection. *Plant Physiology*, *152*(1), 255 – 66.

Huang, X.Z., Chen, J.Y., Xiao, H.J., Xiao, Y.T., Wu, J., Wu, J.X., Zhou, J., Zhang Y., and Guo, Y.Y. (2015). Dynamic transcriptome analysis and volatile profiling of Gossypium hirsutum in response to the cotton bollworm Helicoverpa armigera. *Scientific Reports*, *5*(11867), .

Incarbone, M., and Dunoyer, P. (2013). RNA silencing and its suppression: novel insights from in planta analyses. *Trends in Plant Science*, *18*(7), 382–92.

Irungu, J. (2011). Prevalence and co-infection of cassava with cassava mosaic geminiviruses and cassava brown streak virus in popular cultivars in Western Kenya. A thesis submitted in partial fulfilment of the requirements for the award of the degree of Master of Science in Biotechnology of Kenyatta University. Kenya.

Ishibashi, K., Masuda, K., Naito, S., Meshi, T., and Ishikawa, M. (2007). An inhibitor of viral RNA replication is encoded by a plant resistance gene. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(34), 13833–8.

Jameson, J. D. (1964). Cassava mosaic disease in Uganda. *East African Agriculture and Forestry Journal*, *30*, 208 – 213.

Jennings, D. L. (1957). Further studies in breeding cassava for virus resistance. *East African Agricultural Journal*, *22*, 213 – 219.

Jennings, D. L., (1960b). Observations on virus diseases of cassava in resistant and susceptible varieties. II. Cassava Brown streak disease. *Empire Journal of Experimental Agriculture*, *28*, 261 – 269.

Kachroo, P., Yoshioka, K., Shah, J., Dooner, H.K. and Klessig, D.F. (2000). Resistance to turnip crinkle virus in arabidopsis is regulated by two host genes and is salicylic acid dependent but NPR1, ethylene, and jasmonate independent. *Plant Cell 12*, 677-690.

Kang, B.C., Yeam, I., and Jahn, M. M. (2005a). Genetics of plant virus resistance. *Annual Review of Phytopathology*, *43*, 581–621.

Kang, B.-C., Yeam, I., Frantz, J. D., Murphy, J. F., and Jahn, M. M. (2005b). The pvr1 locus in Capsicum encodes a translation initiation factor eIF4E that interacts with Tobacco etch virus VPg. *The Plant Journal : For Cell and Molecular Biology*, *42*(3), 392–405.

Kanyuka, K., McGrann, G., Alhudaib, K., Hariri, D., and Adams, M. J. (2004). Biological and sequence analysis of a novel European isolate of Barly mild mosaic virus that overcomes the barley rym5 resistance gene. *Archives in Virology*, *149*, 1469-1480.

Kaweesi, T., Kawuki, R., Kyaligonza, V., Baguma, Y., Tusiime, G. and Ferguson, M. (2014). Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. *Virology Journal*, *11*, 216.

Kenward M.G and Roger J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. Biometrics, 5, 983–997.

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

Lanfermeijer F.C., Dijkhuis J., Sturre M.J.G., de Haan P., and Hille J. (2003). Cloning and characterization of the durable tomato mosaic virus resistance gene Tm-22 from Lycopersicon esculentum. *Plant Molecular Biology*, *52*, 1037-1049.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007). ClustalW and ClustalX version 2. *Bioinformatics*, *23*(21), 2947-2948.

Lee, J., Nam, J., Park, H. C., Na, G., Miura, K., Jin, J. B., Baek, D., Kim, D.H., Jeong J.C., Kim, D., Lee, S.Y., Salt, D.E., Mengiste, T., Gong, Q., Ma, S., Bohnert, H.J., Kwak, S.S., Bressan, R.A., Hasegawa, P.M., and Yun, D. J. (2007). Salicylic acid-mediated innate immunity in Arabidopsis is regulated by SIZ1 SUMO E3 ligase. *Plant Journal*, *49*(1), 79–90.

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

Legg, J., Somado, E. A., Barker, I., Beach, L., Ceballos, H., Cuellar, W., Elkhoury, W., Gerling, D., Helsen, J., Hershey, C., Jarvis, A., Kulakow, P., Kumar, L., Lorenzen, J., Lynam, J., McMahon, M., Maruthi, G., Miano, D., Mtunda, K., Natawuruhunga, P., Okogbenin, E., Pezo, P., Terry, E., Thiele, G., Thresh, M., Wadsworth, J., Walsh, S., Winter S., Tohme, J., and Fauquet, C. (2014). A global alliance declaring war on cassava viruses in Africa. Food Security, 6(2), 231–248.

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

Leihner, D. (2002). Agronomy and cropping systems. Cassava: Biology, Production and Utilization, 91–113. Wallingford: CABI publishing.

Lennon, A. M., Aiton, M. M. and Harrison, B. D. H. (1986). Cassava Viruses from Africa. Report for 1985 (Scottish Crops Research Institute).

Leonard, S. (2004). Interaction of VPg-Pro of Turnip mosaic virus with the translation initiation factor 4E and the poly(A)-binding protein in planta. *Journal of General Virology*, *85*(4), 1055–1063.

Leonard, S., Plante, D., Wittmann, S., Daigneault, N., Fortin, M.G. and Laliberte, J.F. (2000). Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. *Journal of Virology*, *74*, 7730–7737

Li, C., Wang, M., Wu, X.-M., Chen, D.-H., Lv, H.-J., Shen, J.-L., Qiao, Z., and Zhang, W. (2015). THI1, a Thiamine Thiazole Synthase, Interacts with Ca2+-dependent Protein Kinase CPK33 and Modulates the S-type Anion Channels and Stomatal Closure in Arabidopsis. *Plant Physiology*, *170*(2), 1090 - 1104.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–9.

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics*, *24*(5), 713–4.

Li, Y., Chen, L., Mu, J., and Zuo, J. (2013). LESION SIMULATING DISEASE1 interacts with catalases to regulate hypersensitive cell death in Arabidopsis. *Plant Physiology*, *163*(2), 1059–70.

Lister, R. M. (1959). Mechanical transmission of cassava brown streak virus. *Nature*, 183, 1588 – 1589.

Liu, Y., Schiff, M., Marathe, R., and Dinesh-Kumar, S. P. (2002). Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus. *The Plant Journal : For Cell and Molecular Biology*, *30*(4), 415–29.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT Method. *Methods*, *25*(4), 402–8.

Mandadi, K. K., and Scholthof, K.-B. G. (2012). Characterization of a viral synergism in the monocot Brachypodium distachyon reveals distinctly altered host molecular processes associated with disease. *Plant Physiology*, *160*(3), 1432–52.

Mandadi, K. K., and Scholthof, K.-B. G. (2013). Plant immune responses against viruses: how does a virus cause disease? *The Plant Cell*, *25*(5), 1489–505.

Margaria, P., Ciuffo, M., Pacifico, D., and Turina, M. (2007). Evidence that the nonstructural protein of Tomato spotted wilt virus is the avirulence determinant in the interaction with resistant pepper carrying the Tsw gene. *Molecular Plant Microbe Interaction*, 20, 547–558.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, *18*(9), 1509–17.

Maruthi, M. N., R. J. Hillocks, K. Mtunda, M. D. Raya, M. Muhanna, H. Kiozia, A. R. Rekha, J. Colvin, and J. M. Thresh. (2005). Transmission of cassava brown streak virus by Bemisia tabaci (Gennadius). *Journal of Phytopathology*. *153*, 307–312.

Maruthi, M.N, Whitfield, E.C., Kumar, P.L. and Legg, J.P. (2014a). Virus-indexing, chemo- and thermo-therapies, and micro propagation for generating virus-free cassava plants. pp 71. Natural Resources Institute, University of Greenwich, Chatham Maritime, UK.

Maruthi, M.N., Bouvaine, S., Tufan, H., Mohammed, I. and Hillocks, R. (2014b). Transcriptional Response of Virus-Infected Cassava and Identification of Putative Sources of Resistance for Cassava Brown Streak Disease. *PLoS One*, *9*, e96642.

Maruthi, M.N., Colvin, J., Seal, S., Gibson, G., and Cooper, J. (2002). Co-adaptation between cassava mosaic geminiviruses and their local vector populations. *Virus Research*, *86*, 71–85.

Masinde, E.A., Ogendo, J.O., Maruthi, M., Hillocks, R., Mulwa, R.M.S., and Arama, P.F. (2016). Occurrence and estimated losses caused by cassava viruses in Migori County, Kenya. *African Journal of Agricultural Research*, *11*(24), 2064 – 2074.

Maule, A. J., Caranta, C., and Boulton, M. I. (2007). Sources of natural resistance to plant viruses: stsatus and prospects. *Molecular Plant Pathology*, *8*(2), 223–31.

Mayba, O., Gilbert, H. N., Liu, J., Haverty, P. M., Jhunjhunwala, S., Jiang, Z., Watanabe C., and Zhang, Z. (2014). MBASED: allele-specific expression detection in cancer tissues and cell lines. *Genome Biology*, *15*(8), 405.

Mazier, M., Flamain, F., Nicolaï, M., Sarnette, V., and Caranta, C. (2011). Knock-down of both eIF4E1 and eIF4E2 genes confers broad-spectrum resistance against potyviruses in tomato. *PloS One*, *6*(12), e29595.

Mbanzibwa, D. R., Tian, Y. P., Tugume, a K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A., Valkonen, J. P. T. (2009a). Genetically distinct strains of Cassava brown streak virus in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. *Archives of Virology, 154(*2), 353–9.

Mbanzibwa, D. R., Tian, Y. P., Tugume, a K., Patil, B. L., Yadav, J. S., Bagewadi, B., Abarshi, M.M, Alicai, T., Changadeya, W., Mkumbira, J., Muli, M.B., Mukasa, S.B., Tairo, F., Baguma, Y., Kyamanywa, S., Kullaya,A., Maruthi, M.N., Fauquet, C.M., Valkonen, J.P. (2011). Evolution of cassava brown streak disease-associated viruses. *The Journal of General Virology*, *92*(4), 974–87.

Mbanzibwa, D. R., Tian, Y., Mukasa, S. B., and Valkonen, J. P. T. (2009b). Cassava brown streak virus (Potyviridae) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-Pro. *Journal of Virology*, *83*(13), 6934–40.

McSween, S., Walker, T., Salega, V., and Pitoro, R. (2006). Economic Impact of Food Security of Varietal Tolerance to Cassava Brown Streak Disease in Coastal Mozambique. Research Report Series No. 1E of Institute of Agricultural Research of Mozambique.

Meshi, T., Motoyoshi, F., Adachi, A., Watanabe, Y., and Okada, Y. (1988). Two concomitant base substitutions in the putative replicase genes of tobacco mosaic virus confer the ability to overcome the effects of tomato resistance gene, Tm-1. *EMBO Journal*, *7*, 1575–1581.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics*, *11*(1), 31–46.

Meyer, P. (2015). Epigenetic variation and environmental change. *Journal of Experimental Botany*, *66*(12), 3541–3548.

Mohammed, I. U., Abarshi, M. M., Muli, B., Hillocks, R. J., and Maruthi, M. N. (2012). The symptom and genetic diversity of cassava brown streak viruses infecting cassava in East Africa. *Advances in Virology*, *2012*, 795697.

Mohr P. G., and Cahill D. M. (2003). Abscisic acid influences the susceptibility of Arabidopsis thaliana to *Pseudomonas syringae* pv. tomato and *Peronospora parasitica*. *Functional Plant Biology*, *30*, 461–469.

Monger, W. A., Seal, S., Cotton, S., and Foster, G. D. (2001). Identification of different isolates of Cassava brown streak virus and development of a diagnostic test. *Plant Pathology*, *50*, 768–775.

Morel, J. B., and Dangl, J. L. (1997). The hypersensitive response and the induction of cell death in plants. *Cell Death & Differentiation*, *4*(8), 671–83.

Moreno, I., Gruissem, W., and Vanderschuren, H. (2011). Reference genes for reliable potyvirus quantitation in cassava and analysis of Cassava brown streak virus load in host varieties. *Journal of Virological Methods*, *177*(1), 49–54.

Morozova, O., M. Hirst and M. A. Marra, (2009). Applications of new sequencing technologies for transcriptome analysis. *Annual Review of Genomics and Human Genetics*, *10*, 135–151.

Mortazavi, A., Williams, B. a, McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), 621–8. doi:10.1038/nmeth.1226.

Mtunda, K. J., Muhanna, M., Raya, M. D. and Kanju, E. E. (2003). Current status of Cassava brown streak virus disease in Tanzania. In: Legg, J. P. and Hillocks, R. J. (eds). Cassava brown streak virus disease: Past, Present, and Future. Proceedings of an International Workshop, Mombasa, Kenya, 27-30 October 2002. Pp 7-11

Mulimbi, W., Phemba, X., Assumani, B., Kasereka, P., Muyisa, S., Ugentho, H., Reeder, R., Legg, J., Laurenson, L., Weekes, R. and Thom, F. (2012) First report of Ugandan cassava brown streak virus on cassava in Democratic Republic of Congo. New Dis Rep 26, 11.

Muñoz-Bodnar A., Perez-Quintero A.L., Gomez-Cano F., Gil J., Michelmore R., Bernal A., Szurek B., and Lopez C. (2014). RNAseq analysis of cassava reveals similar plant responses upon upon infection with pathogenic and non-pathogenic strains of Xanthomonas axonopodis pv. Manihotis. *Plant Cell Report*, *33*, 1901-1912.

Munthali, D. C. (1992). Effect of cassava variety on the biology of Bemisia afer (Priesner and Hosny) Hemiptera: Aleyrodidae. *Insect Science and its Application*, *13*, 459–465.

Mur, L. A. J., Bi, Y.-M., Darby, R. M., Firek, S., and Draper, J. (1997). Compromising early salicylic acid accumulation delays the hypersensitive response and increases viral dispersal during lesion establishment in TMV-infected tobacco. *Plant Journal*, *12*, 1113-1126.

Mware, B., Narla, R., Amata, R., Olubayo, F., Songa, J., Kyamanyua, S., and Ateka, E. M. (2009). Efficiency of cassava brown streak virus transmission by two whitefly species in coastal Kenya. *Journal of General and Molecular Virology*, *1*(4), 40–45.

Neuenschwander P. (2001). Biological control of the cassava mealybug in Africa: A review. *Biological Control*, *21*, 214–229.

Nicaise, V., Gallois, J.-L., Chafiai, F., Allen, L. M., Schurdi-Levraud, V., Browning, K. S., Candresse, T., Caranta, C., Le Gall O., and German-Retana, S. (2007). Coordinated and selective recruitment of eIF4E and eIF4G factors for potyvirus infection in Arabidopsis thaliana. *FEBS Letters*, *581*(5), 1041–6.

Nicaise, V., German-Retana, S., Sanjuán, R., Dubrana, M.-P., Mazier, M., Maisonneuve, B., Candresse, T., Caranta, C., LeGall, O. (2003). The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the Potyvirus Lettuce mosaic virus. *Plant Physiology*, *132*(3), 1272–82.

Nichols, R. F. W. (1950). The brown streak disease of cassava: distribution climatic effects and diagnostic symptoms. *East African Agricultural Journal*, *15*, 154 – 160.

Nichols, R.F.W. (1947). Breeding cassava for virus resistance. *East African Agricultural Journal*, *12*, 184–194.

Nieto, C., Morales, M., Orjeda, G., Clepet, C., Monfort, A., Sturbois, B., … Bendahmane, A. (2006). An eIF4E allele confers resistance to an uncapped and non-polyadenylated RNA virus in melon. *The Plant Journal : For Cell and Molecular Biology*, *48*(3), 452–62.

Nweke, F. I. (2004). New challenges in the cassava transformation in Nigeria And Ghana. EPTD Discussion Paper No. 118. Washington USA: Environment and Production technology Division, International Food Policy Research Institute.

Odipio, J., Ogwok, E., Taylor, N. J., Halsey, M., Bua, A., Fauquet, C. M., and Alicai, T. (2014). RNAi-derived field resistance to Cassava brown streak disease persists across the vegetative cropping cycle. *GM Crops and Food*, *5*(1), 1–4.

Ogwok, E., Alicai, T., Rey, M., Beyene, G. and Taylor, N. (2014). Distribution and accumulation of cassava brown streak viruses within infected cassava (Manihot esculenta) plants. *Plant Pathology*, *64*, 1235-1246.

Ogwok, E., Patil, B. L., Alicai, T., and Fauquet, C. M. (2010). Transmission studies with Cassava brown streak Uganda virus (Potyviridae: Ipomovirus) and its interaction with abiotic and biotic factors in Nicotiana benthamiana. *Journal of Virological Methods*, *169*(2), 296–304.

Otim-Nape, G.W., Bua, A., Thresh, J.M., Baguma, Y., Ogwal, S., Ssemakula, G.N., Acola, Byabakama, B., Colvin, J., Cooter, R.J., and Martin, A. (2000). The Current Pandemic of Cassava Mosaic Virus Disease in East Africa and its Control. NRI/NARO, Chatham, UK, DFID Publication No. 100.

Padgett, H.S., Watanabe, Y., and Beachy, R. N. (1997). Identification of the TMV replicase sequence that activates the N gene-mediated hypersensitive response. *Molecular Plant Microbe Interaction*, *10*, 709-715.

Padmanabhan, M. S., Kramer, S. R., Wang, X., and Culver, J. N. (2008). Tobacco mosaic virus replicase-auxin/indole acetic acid protein interactions: reprogramming the auxin response pathway to enhance virus infection. *Journal of Virology*, *82*, 2477–2485.

Padmanabhan, M.S., Shiferaw, H., Culver, J.N. (2006). The Tobacco mosaic virus replicase protein disrupts the localization and function of interacting Aux/IAA proteins. *Molecular Plant Microbe Interactions*, *19*, 864–873.

Palukaitis, P., and Carr, J. P. (2008). Plant resistance response to viruses. *Journal of Plant Pathology*, *90*, 153–171.

Park, S.W., Kaimoyo, E., Kumar, D., Mosher, S., and Klessig, D.F. (2007) Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. *Science*, *318*, 113-116.

PATIL, B. L. and FAUQUET, C. M. (2009). Cassava mosaic geminiviruses: actual knowledge and perspectives. *Molecular Plant Pathology*, 10, 685–701.

Pierce, E. J., and 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*(6), e67534.

Piron, F., Nicolaï, M., Minoïa, S., Piednoir, E., Moretti, A., Salgues, A., Zamir, D., Caranta, C., and Bendahmane, A. (2010). An induced mutation in tomato eIF4E leads to immunity to two potyviruses. *PloS One*, *5*(6), e11313.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Robertson, I. A. D. (1985). Cassava Whitefly Project. Part of Crop Virology Project R3177 at the Kenya Agricultural Research Institute, Final Report (London: Overseas Development Administration).

Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(3), R25.

Rochon, A., Boyle, P., Wignes, T., Fobert, P. R., and Després, C. (2006). The coactivator function of Arabidopsis NPR1 requires the core of its BTB/POZ domain and the oxidation of C-terminal cysteines. *The Plant Cell*, *18*(12), 3670–85.

Ronald, P. C. and Beutler, B. (2010). Plant and animal sensors of conserved microbial signatures. *Science*, *330,* 1061–1064.

Ruffel, S., Dussault, M. H., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C., and Caranta, C. (2002). A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). *Plant Journal*, *32*, 1067–1075.

Rwegasira, G. M., and M. E. Rey, C. (2012). Response of Selected Cassava Varieties to the Incidence and Severity of Cassava Brown Streak Disease in Tanzania. *Journal of Agricultural Science*, *4*(7), 237–245.

Saez, A., Robert, N., Maktabi, M. H., Schroeder, J. I., Serrano, R., and Rodriguez, P. L. (2006). Enhancement of abscisic acid sensitivity and reduction of water consumption in Arabidopsis by combined inactivation of the protein phosphatases type 2C ABI1 and HAB1. *Plant Physiology*, *141*(4), 1389–99.

Sanfançon, H. (2015). Plant translation factors and virus resistance. *Viruses*, *7*(7), 3392–3419.

Schaad, M. C., Anderberg, R. J., and Carrington, J. C. (2000). Strain-specific interaction of the tobacco etch virus NIa protein with the translation initiation factor eIF4E in the yeast two-hybrid system. *Virology*, *273*(2), 300–6.

Schoelz, J.E., Harries, P.A., and Nelson, R. S. (2011). Intracellular transport of plant viruses: finding the door out of the cell. *Molecular Plant*, *4*, 813–831.

Schwach, F., Vaistij, F. E., Jones, L., and Baulcombe, D. C. (2005). An RNA-dependent RNA polymerase prevents meristem invasion by potato virus X and is required for the activity but not the production of a systemic silencing signal. *Plant Physiology*, *138*(4), 1842–52.

Seki, M., Narusaka, M., Ishida, J., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji,T., Yamaguch-Shinozaki, K., Carninci, P., Hayashizaki, K., and Shinozaki, K. (2002). Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. *Functional and Integrative Genomics*, *2*(6), 282–291.

Seo, Y.-S., Jeon, J.-S., Rojas, M. R., and Gilbertson, R. L. (2007). Characterization of a novel Toll/interleukin-1 receptor (TIR)-TIR gene differentially expressed in common bean (Phaseolus vulgaris cv. Othello) undergoing a defence response to the geminivirus Bean dwarf mosaic virus. *Molecular Plant Pathology*, *8*(2), 151–62.

Seo, Y.-S., Rojas, M. R., Lee, J.-Y., Lee, S.-W., Jeon, J.-S., Ronald, P., … Gilbertson, R. L. (2006). A viral resistance gene from common bean functions across plant families and is up-regulated in a non-virus-specific manner. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(32), 11856–61.

Shimura, H., and Pantaleo, V. (2011). Viral induction and suppression of RNA silencing in plants. *Biochimica et Biophysica Acta*, *1809*(11-12), 601–12.

Sigurgeirsson, B., Emanuelsson O., and Lundeberg J. (2014). Sequencing Degraded RNA Addressed By 3′ Tag Counting. *PLoS ONE*, *9*, e91851.

Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J., and Akey, J. M. (2011). A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Research*, *21*, 1728–1737.

Solovyev, A. G., Kalinina, N. O., and Morozov, S. Y. (2012). Recent advances in research of plant virus movement mediated by triple gene block. *Frontiers in Plant Science*, *3*, 276.

Spoel, S. H., and Dong, X. (2012). How do plants achieve immunity? Defence without specialized immune cells. Nature Reviews. *Immunology*, *12*(2), 89–100.

Spoel, S. H., Mou, Z., Tada, Y., and Spivey, N. W. (2010). Proteasome-mediated turnover of the transcription co-activator NPR1 plays dual roles in regulating plant immunity. *Cell*, *137*, 860-872.

Stein, N., Perovic, D., Kumlehn, J., Pellio, B., Stracke, S., Streng, S., Ordon, F., and Graner, A. (2005). The eukaryotic translation initiation factor 4E confers multiallelic recessive Bymovirus resistance in Hordeum vulgare (L.). *The Plant Journal : For Cell and Molecular Biology*, *42*(6), 912–22.

Storey, H. H. (1936). Virus diseases of East African plants: VI—A progress report on studies of the diseases of cassava. *East African Agricultural Journal*, *2*, 34–39.

Storey, J.D. The optimal discovery procedure: a new approach to simultaneous significance testing (2007). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(3), 347–368.

Storey, J.D., Leek, J.T., and Bass, A. J. (2015). edge: Extraction of Differential Gene Expression. R package version 2.5.3. https://github.com/jdstorey/edge.

Sun, W., and Hu Y. (2013). eQTL Mapping Using RNA-seq Data. *Statistical Bioscience*, *5*(1), 198-219.

Swarup, K., Benková, E., Swarup, R., Casimiro, I., Péret, B., Yang, Y., Parry, G., Nielsen, E., De Smet, I., Vanneste, S., et al. (2008). The auxin influx carrier LAX3 promotes lateral root emergence. *Nature Cell Biology*, *10*(2008), 946–954.

Sweetmore, A. (1994). Adaptation and Development of Diagnostic Reagents for Cassava Brown Streak Virus for Use in Less Developed Countries. Final Project Report, DFID Crop Protection Programme Project No. R5880CB (NRI).

Taji, T., Ohsumi, C., Iuchi, S., Seki, M., Kasuga, M., Kobayashi, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2002). Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana. *Plant Journal*, *29*(4), 417–426.

Takahashi, H., Miller J., Nozaki Y., Sukamto, Takeda M., Shah J., Hase S., Ikegami M., Ehara Y., and Dinesh-Kumar S.P. (2002). RCY1, an Arabidopsis thaliana RPP8/HRT family resistance gene, conferring resistance to cucumber mosaic virus requires salicylic acid, ethylene and a novel signal transduction mechanism. *Plant Journal*, *32*, 655-667.

Takahashi, S., Seki, M., Ishida, J. *et al*. (2004). Monitoring the expression profiles of genes induced by hyperosmotic, high salinity, and oxidative stress and abscisic acid treatment in Arabidopsis cell culture using a full-length cDNA microarray. *Plant Molecular Biology*, *56*(1), 29 – 55.

Talarczyk, A. and Hennig, J. (2001). Early defence responses in plants infected with pathogenic organisms. *Cell Molecular Biology Letters*, 6, 955-970.

Tausz, M., Šircelj, H., and Grill, D. (2004). The glutathione system as a stress marker in plant ecophysiology: Is a stress-response concept valid? *Journal of Experimental Botany*, *55*(404), 1955–1962.

Thaler, J. S., and Bostock, R. M. (2004). Interactions between Abscisic-acid-mediated responses and plant resistance to pathogens and insects. *Ecology*, *85*(1), 48 – 58.

Theu, M.P.K.J and Mazuma (2008). Management and screening of cassava brown streak virus disease in Malawi. In Plant protection progress report for the 2007/2008 season, 104 - 116. Lilongwe: Department of Agricultural Research Services.

Thivierge, K., Nicaise, V., Dufresne, P. J., Cotton, S., Laliberté, J.-F., Le Gall, O., and Fortin, M. G. (2005). Plant virus RNAs. Coordinated recruitment of conserved host functions by (+) ssRNA viruses during early infection events. *Plant Physiology*, *138*(4), 1822–7.

Thresh, J.M. (2006). Control of tropical plant virus diseases. *Advances in Virus Research*, 67, 245-295.

Tiendre´be´ogo, F., Lefeuvre, P., Hoareau, M., Harimalala, M. A., De Bruyn, A., Villemot, J., *et al*. (2012). Evolution of African cassava mosaic virus by recombination between bipartite and monopartite begomoviruses. *Virology Journal*, *9*, 67.

Tomita, R., Murai, J., Miura, Y., Ishikara, H., Liu, S., Kubotera, Y., Honda, A., Hatta, R., Kuroda, T., Hamada, H., Sakamoto, M., Munemura, I., *et al*. (2008). Fine mapping and DNA fiber FISH analysis locates the tobamovirus resistance gene L3 of *Capsicum chinense* in a 400-kb region of R-like genes cluster embedded in highly repetitive sequences. *Theoretical and Applied Genetetics*, *117*, 1107–1118.

Tong, C., Wang, X., Yu, J., Wu, J., Li, W., Huang, J., Dong, C., Hua, W., and Liu, S. (2013). Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa*. *BMC Genomics*, *14*(1), 689.

Torres, M. A., and Dangl, J. L. (2005). Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. *Current Opinion in Plant Biology*, *8*(4), 397–403.

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105–11.

Trapnell, C., Williams, B. a, Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S.L., Wold,B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*(5), 511–5.

Umezawa, T., Fujita, M., Fujita, Y., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2006). Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. *Current Opinion in Biotechnology*, *17*(2), 113–122.

Vaistij, F. E., and Jones, L. (2009). Compromised virus-induced gene silencing in RDR6-deficient plants. *Plant Physiology*, *149*(3), 1399–407.

Vallejos, C. E., Astua-Monge, G., Jones, V., Plyler, T. R., Sakiyama, N. S., and Mackenzie, S. A. (2006). Genetic and molecular characterization of the I locus of Phaseolus vulgaris. *Genetics*, *172*, 1229–1242.

Van Loon, L.C., and Van Strien, E.A. (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.

Vanderschuren, H., Alder, A., Zhang, P., and Gruissem, W. (2009). Dose-dependent RNAi-mediated geminivirus resistance in the tropical root crop cassava. *Plant Molecular Biology*, *70*(3), 265–72.

Vanlerberghe, G. C. (2013). Alternative Oxidase: A Mitochondrial Respiratory Pathway to Maintain Metabolic and Signaling Homeostasis during Abiotic and Biotic Stress in Plants. *International Journal of Molecular Sciences*, *14*(4), 6805–6847.

Vaucheret, H. F., Vazquez, P., Crete, D.P., and Bartel (2004). The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Development*, *18*, 1187–1197.

Vidal, S., Cabrera H., Andersson R.A., Fredriksson A., and Valkonen J.P.T. (2002). Potato virus Y-1 is an N gene homolog that confers cell death upon infection with Potato virus Y. *Molecular Plant-Microbe Interactions*, *15*, 717-727.

Wagaba, H., Beyene, G., Trembley, C., Alicai, T., Fauquet, C. M., and Taylor, N. J. (2013). Efficient transmission of cassava brown streak disease viral pathogens by chip bud grafting. *BMC Research Notes*, *6*, 516.

Wang, Z., Gerstein, M. and Snyder, M. (2009). RNA-seq: a revolutionary tool for transcriptomics. *Nature Review Genetics*, *10*, 57–63.

Weber, H., and Pfitzner, A. J. P. (1998). Tm22 resistance in tomato requires recognition of th carboxy terminus of the movement protein of tobacco mosaic virus. *Molecular Plant Microbe Interactions*, *11*, 498–503.

Weber, H., Ohnesorge, S., Silber, M. V., and Pfitzner, A. J. P. (2004). The Tomato mosaic virus 30 kDa movement protein interacts differentially with the resistance genes Tm-2 and Tm-22. *Archives in Virology*, *149*, 1499–1514.

Westby, A. (2002). Cassava utilization, storage and small-scale processing. In Hilllocks, R.J., Thresh, J.M., and Bellotti, A., (Eds), *Cassava: Biology, Production and Utilization*, 281–300. Oxfordshire, UK: CABI Publishing.

Whenham, R. J., Fraser, R. S., Brown, L. P., and Payne, J. A. (1986). Tobacco-mosaic-virus-induced increase in abscisic-acid concentration in tobacco leaves: Intracellular location in light and dark-green areas, and relationship to symptom development. *Planta*, *168*(4), 592 – 598.

Whitham, S. A, Anderberg, R. J., Chisholm, S. T., and Carrington, J. C. (2000). Arabidopsis RTM2 gene is necessary for specific restriction of tobacco etch virus and encodes an unusual small heat shock-like protein. *The Plant Cell*, *12*(4), 569–82.

Whitham, S. A., Quan, S., Chang, H. S., Cooper, B., Estes, B., Zhu, T., Wang, X., and Hou, Y. M. (2003). Diverse RNA viruses elicit the expression of common sets of genes in susceptible Arabidopsis thaliana plants. *Plant Journal*, *33*(2), 271–283.

Whitham, S.A., Dinesh-Kumar, S.P., Choi, D., Hehl, R., Corr, C., and Baker, B. (1994). The product of the tobacco mosaic virus resistance gene N: Similarity to toll and the interleukin-1 receptor. *Cell, 78*, 1101–1115.

Winter, S., Koerbler, M., Stein, B., Pietruszka, A., Paape, M., and Butgereitt, A. (2010). Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. *The Journal of General Virology*, *91*(5), 1365–1372.

Wu, Y., Zhang, D., Chu, J. Y., Boyle, P., Wang, Y., Brindle, I. D., De Luca V., and Després, C. (2012). The Arabidopsis NPR1 protein is a receptor for the plant defence hormone salicylic acid. *Cell Reports*, *1*(6), 639–47.

Yu, D., Fan B., MacFarlane S.A. and Chen Z. (2003). Analysis of the involvement of an inducible *Arabidopsis* RNA-dependent RA polymerase in Antiviral defence. *Molecular Plant-Microbe Interactions*, *16*(3), 206-216.

Zan, Y., Shen, X., Forsberg, S. K. G., and Carlborg, O. (2016). Genetic regulation of transcriptional variation in wild-collected *Arabidopsis thaliana* accessions. *G3*, *6*(8), 2319 – 2328.

Zeng, C., Chen, Z., Xia, J., Zhang, K., Chen, X., Zhou, Y., Bo, W., Song, S., Deng, D., Guo, X., *et al*. (2014). Chilling acclimation provides immunity to stress by altering regulatory networks and inducing genes with protective functions in Cassava. *BMC Plant Biology*, *14*(1), 207.

Zhai, R., Feng, Y., Zhan, X., Shen, X., Wu, W., Yu P., Zhang, Y., Chen, D., Wang, H., Lin, Z., Cao, L. (2013). Identification of transcriptome SNPs for assessing allele-specific gene expression in super-hybrid rice Xieyou9308. *Plos One*, *8*(4), e60668.

Zhang, X., and Borevitz, J. O. (2009). Global analysis of allele-specific expression in *Arabidopsis thaliana*. *Genetics*, *182*(4), 943–954.

Zhang, L., Oh, Y., Li, H., Baldwin, I. T., and Galis, I. (2012). Alternative Oxidase in Resistance to Biotic Stresses: *Nicotiana attenuata* AOX Contributes to Resistance to a Pathogen and a Piercing-Sucking Insect But Not *Manduca sexta* Larvae. *Plant Physiology*, *160*(3), 1453–1467.

Zhang, Y., Ding, Z., Ma, F., Chauhan, R. D., Allen, D. K., Brutnell, T. P., Wang., W., Peng., M., and Li, P. (2015). Transcriptional response to petiole heat girdling in cassava. *Scientific Reports*, *5*, 8414.

Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G. W., Robinson, D. J., and Harrison, B. D. (1997). Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *The Journal of General Virology*, *78* (8), 2101–11.

Zhu, S., Jeong, R. D., Venugopal, S. C., Lapchyk, L., Navarre, D., Kachroo, A., and Kachroo, P. (2011). Sag101 forms a ternary complex with EDS1 and PAD4 and is required for resistance signaling against turnip crinkle virus. *PLoS Pathogens*, *7*(11), e1002318.

Zipfel, C. and Rathjen, J. P. (208). Plant immunity: AvrPto targets the frontline. *Current Biology*, *18*, 218–220.

APPENDICES

Appendix 1. Edge analysis result summary for the number of differential gene expressions influenced by time.

Statistical significance summary:

pi0: 0.07353194

Cumulative number of significant calls:

	<1e-04	<0.001	<0.01	<0.025	<0.05	<0.1	<1
p-value	6019	7984	10014	11095	12322	13807	19082
q-value	7481	9657	13763	16242	18185	19082	19082
local fdr	5543	7578	9674	11198	13058	14908	19080

Appendix 2. Edge analysis result summary for numbers of differentially expressed genes, between resistant and tolerant accessions, at different P-value, Q-value and local FDR thresholds

Statistical significance summary:

pi0: 0.3459864

Cumulative number of significant calls:

	<1e-04	<0.001	<0.01	<0.025	<0.05	<0.1	<1
p-value	177	526	1895	3767	5858	7800	19082
q-value	0	58	424	1205	5136	8654	19082
local fdr	0	49	226	596	2044	5451	19015

Appendix 3. Edge analysis result summary for numbers of differentially expressed genes, between resistant and susceptible accessions, at different P-value, Q-value and local FDR thresholds

Statistical significance summary:

pi0: 0.4781914

	<1e-04	<0.001	<0.01	<0.025	<0.05	<0.1	<1
p-value	579	1193	2757	3939	5280	7056	19082
q-value	197	511	1354	2360	3573	5876	19082
local fdr	132	296	768	1298	1976	3040	19077

Appendix 4. Edge analysis result summary for numbers of differentially expressed genes, between tolerant and susceptible accessions, at different P-value, Q-value and local FDR thresholds

Statistical significance summary:

pi0: 0.3428772

	<1e-04	<0.001	<0.01	<0.025	<0.05	<0.1	<1
p-value	62	217	836	1473	2427	4356	19082
q-value	0	0	69	168	435	1278	19082
local fdr	0	0	31	93	203	570	19033

Appendix 5. Gene expression fold difference and identity of genes overexpressed in resistant relative to susceptible cassava.

Gene identity	Gene name	Average RPKM in susceptible	Average RPKM in resistant	Susceptible-by-resistant RPKM fold
cassava4.1_001124m	disease resistance protein RPP8	4.6	10.9	2.4
cassava4.1_002792m	putative disease resistance RPP13-like protein 1	10.6	22.0	2.1
cassava4.1_005201m	cytosolic invertase 2	11.4	17.4	1.5
cassava4.1_005786m	cytochrome P450, family 714, subfamily A, polypeptide 1	18.9	61.9	3.3
cassava4.1_007337m	uridine kinase-like 3	3.2	7.1	2.3
cassava4.1_007745m	putative LRR receptor-like serine/threonine-protein kinase	9.6	20.4	2.1
cassava4.1_008411m	uncharacterized protein	31.0	49.8	1.6
cassava4.1_009835m	Afadin/alpha-actinin-binding protein	3.8	5.6	1.5
cassava4.1_010650m	S-adenosyl-L-methionine-dependent methyltransferase-like protein	13.1	29.2	2.2
cassava4.1_011012m	EMBRYO DEFECTIVE 2799	15.6	22.8	1.5
cassava4.1_011133m	malate dehydrogenase 1	58.9	84.2	1.4
cassava4.1_011541m	putative S-acyltransferase	4.7	8.4	1.8
cassava4.1_012476m	uncharacterized protein	4.3	12.5	2.9
cassava4.1_012887m	transport protein SEC20	11.2	17.1	1.5
cassava4.1_014263m	TIR-NBS-LRR class disease resistance protein	11.1	37.0	3.3
cassava4.1_014860m	uncharacterized protein	6.4	16.4	2.5
cassava4.1_015251m	GLABROUS INFLORESCENCE STEMS	53.6	70.6	1.3
cassava4.1_015326m	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	10.3	13.5	1.3
cassava4.1_015345m	NIFU-like protein 2	28.0	42.1	1.5
cassava4.1_016199m	D-tyrosyl-tRNA(Tyr)deacylase	4.8	6.9	1.5
cassava4.1_016724m	diphosphomevalonate decarboxylase	5.6	7.6	1.4
cassava4.1_016754m	frataxin	9.0	21.7	2.4
cassava4.1_016830m	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	10.8	14.0	1.3
cassava4.1_019851m	eukaryotic translation initiation factor 3E	33.2	46.2	1.4
cassava4.1_020261m	nil	7.3	18.0	2.5
cassava4.1_020264m	nil	5.3	11.1	2.1
cassava4.1_020424m	LSD1-like 2	7.9	13.6	1.7
cassava4.1_021434m	Elongation factor 1-beta 2	9.8	19.4	2.0

cassava4.1_022120m	MIRO-related GTP-ase 2	3.7	9.3	2.5
cassava4.1_023508m	putative disease resistance RPP13-like protein 1	7.5	19.9	2.6
cassava4.1_024050m	peptidase-S24/S26 domain-containing protein	15.7	24.8	1.6
cassava4.1_024649m	DEFECTIVE IN MERISTEM SILENCING 3	5.1	13.2	2.6
cassava4.1_026203m	alpha-mannosidase 2	6.8	10.2	1.5
cassava4.1_028445m	disease resistance protein RPM1	2.9	8.2	2.9
cassava4.1_031755m	organic cation/carnitine transporter4	6.1	12.9	2.1
cassava4.1_033369m	UDP-glycosyltransferase-like protein	57.8	112.2	1.9

*wpi stands for weeks post-inoculation

Appendix 6. Gene expression fold difference and identity of genes overexpressed in susceptible relative to resistant cassava.

Gene identity	Gene name	Average RPKM in susceptible	Average RPKM in resistant	Resistant-by-susceptible RPKM fold
cassava4.1_000864m	BARELY ANY MERISTEM 1	25.0	15.4	1.6
cassava4.1_001384m	embryo defective 1789	5.5	2.9	1.9
cassava4.1_001759m	uncharacterized protein	4.4	2.6	1.7
cassava4.1_002773m	myb family transcription factor	5.6	3.2	1.8
cassava4.1_002774m	Plasma-membrane choline transporter family protein	7.0	5.3	1.3
cassava4.1_002778m	SILENCING DEFECTIVE	37.5	1.9	19.3
cassava4.1_002879m	protein kinase family protein	14.5	10.8	1.3
cassava4.1_002907m	phospholipid:diacylglycerol acyltransferase	12.4	5.6	2.2
cassava4.1_003005m	SILENCING DEFECTIVE	38.3	2.0	19.2
cassava4.1_003547m	uncharacterized protein	51.6	39.3	1.3
cassava4.1_003588m	uncharacterized protein	51.7	39.4	1.3
cassava4.1_003659m	alkaline/neutral invertase	66.2	42.0	1.6
cassava4.1_003761m	signal recognition particle receptor subunit alpha	33.3	18.5	1.8
cassava4.1_004273m	polynucleotide adenylyltransferase-like protein	11.9	8.3	1.4
cassava4.1_004280m	leucyl aminopeptidase 2	78.7	62.8	1.3
cassava4.1_004343m	NADH dehydrogenase	26.7	18.5	1.4
cassava4.1_004395m	endonuclease VIII-like 3	33.4	20.6	1.6
cassava4.1_004619m	starch synthase 1	7.6	5.2	1.5
cassava4.1_004692m	C3H4 type zinc finger protein	25.5	15.3	1.7
cassava4.1_004714m	leucyl aminopeptidase 2	77.0	61.1	1.3
cassava4.1_004762m	Rab escort protein	10.0	7.1	1.4
cassava4.1_004849m	RbBP5 LIKE	43.2	26.7	1.6
cassava4.1_004862m	ACYL-ACTIVATING ENZYME 3	30.1	21.7	1.4
cassava4.1_005103m	endonuclease VIII-like 3	34.2	21.0	1.6
cassava4.1_005304m	plastid transcriptionally active 12	49.3	34.0	1.4
cassava4.1_005518m	ADP glucose pyrophosphorylase 1	239.2	150.4	1.6

cassava4.1_005522m	ACYL-ACTIVATING ENZYME 3	30.1	21.6	1.4
cassava4.1_005634m	3-ketoacyl-CoA synthase 4	48.1	10.1	4.8
cassava4.1_005700m	SNF1 kinase homolog 10	22.4	19.5	1.1
cassava4.1_005735m	aspartic proteinase A1	276.7	201.7	1.4
cassava4.1_005832m	ACYL-ACTIVATING ENZYME 3	28.0	20.2	1.4
cassava4.1_005876m	pentatricopeptide repeat-containing protein	18.1	12.5	1.5
cassava4.1_005888m	pentatricopeptide repeat-containing protein	26.1	15.8	1.7
cassava4.1_005914m	metal transporter Nramp3	83.8	53.9	1.6
cassava4.1_005987m	aspartic proteinase A1	268.6	198.0	1.4
cassava4.1_006001m	putative eukaryotic LigT protein	32.5	25.4	1.3
cassava4.1_006017m	putative eukaryotic LigT protein	32.4	25.3	1.3
cassava4.1_006023m	EMBRYO DEFECTIVE 3101	12.3	9.2	1.3
cassava4.1_006024m	aldehyde dehydrogenase 10A8	52.0	35.3	1.5
cassava4.1_006027m	transcription factor bHLH3	51.8	40.0	1.3
cassava4.1_006072m	aldehyde dehydrogenase 10A8	52.2	35.4	1.5
cassava4.1_006452m	uncharacterized protein	16.4	13.2	1.2
cassava4.1_006471m	UDP-glucosyl transferase 73C	13.6	5.7	2.4
cassava4.1_006534m	alanine:glyoxylate aminotransferase 2	18.9	12.6	1.5
cassava4.1_006645m	SNF1 kinase homolog 10	20.0	17.1	1.2
cassava4.1_006725m	ureidoglycolate amidohydrolase	25.8	18.5	1.4
cassava4.1_006821m	2-oxoisovalerate dehydrogenase E1 component, alpha subunit	10.5	7.4	1.4
cassava4.1_006870m	gamma-tubulin	4.8	2.8	1.7
cassava4.1_006913m	homeobox-leucine zipper family protein	12.5	8.1	1.5
cassava4.1_007181m	3-ketoacyl-CoA thiolase 2	63.2	51.9	1.2
cassava4.1_007211m	high chlorophyll fluorescent 109	26.8	17.8	1.5
cassava4.1_007442m	EMBRYO DEFECTIVE 3101	12.7	9.5	1.3
cassava4.1_007667m	putative anthranilate phosphoribosyltransferase	15.9	12.0	1.3
cassava4.1_007680m	3-ketoacyl-CoA thiolase 2	61.0	49.6	1.2
cassava4.1_007812m	HXXXD-type acyl-transferase-like protein	43.9	22.3	2.0

cassava4.1_007881m	UNC93-like protein	18.1	14.6	1.2
cassava4.1_007931m	DJ-1 homolog B	44.3	31.6	1.4
cassava4.1_007947m	UNC93-like protein	18.1	14.7	1.2
cassava4.1_008012m	patatin-like phospholipase domain-containing protein	27.6	19.6	1.4
cassava4.1_008046m	protein kinase family protein	6.0	4.3	1.4
cassava4.1_008071m	BSD domain-containing protein	22.9	16.4	1.4
cassava4.1_008337m	elongation factor EF-1 gamma subunit	108.5	86.9	1.2
cassava4.1_008338m	elongation factor EF-1 gamma subunit	108.5	86.9	1.2
cassava4.1_008393m	uncharacterized protein	20.2	10.1	2.0
cassava4.1_008401m	uncharacterized protein	20.2	10.1	2.0
cassava4.1_008421m	regulatory particle triple-A ATPase 3	47.1	40.2	1.2
cassava4.1_008437m	regulatory particle triple-A ATPase 6A	32.2	25.6	1.3
cassava4.1_008487m	elongation factor EF-1 gamma subunit	109.7	87.9	1.2
cassava4.1_008547m	S-methyl-5-thioribose kinase	36.7	25.0	1.5
cassava4.1_008549m	S-methyl-5-thioribose kinase	36.7	25.0	1.5
cassava4.1_008605m	FAD-dependent oxidoreductase-like protein	9.8	6.8	1.4
cassava4.1_008632m	Squalene synthase	7.2	4.6	1.6
cassava4.1_008808m	LisH/CRA/RING-U-box domain-containing protein	24.3	19.4	1.3
cassava4.1_009025m	DNA/RNA-binding protein Kin17 conserved region-containing protein	12.7	9.9	1.3
cassava4.1_009027m	inositol polyphosphate 1-phosphatase	19.1	15.1	1.3
cassava4.1_009084m	tryptophanyl-tRNA synthetase	30.1	19.6	1.5
cassava4.1_009090m	tryptophanyl-tRNA synthetase	30.1	19.6	1.5
cassava4.1_009103m	tryptophanyl-tRNA synthetase	30.1	19.6	1.5
cassava4.1_009104m	tryptophanyl-tRNA synthetase	30.1	19.6	1.5
cassava4.1_009131m	aspartic proteinase A1	296.6	223.2	1.3
cassava4.1_009187m	Fes1A	31.6	27.0	1.2
cassava4.1_009192m	Fes1A	31.6	27.0	1.2
cassava4.1_009232m	eukaryotic translation initiation factor 2B-like protein	20.7	16.2	1.3
cassava4.1_009517m	LisH/CRA/RING-U-box domain-containing protein	25.5	20.3	1.3

cassava4.1_009613m	Serine carboxypeptidase S28 family protein	25.4	20.1	1.3
cassava4.1_009670m	haem-binding protein 5	76.8	39.0	2.0
cassava4.1_009739m	haloacid dehalogenase-like hydrolase domain-containing protein	136.0	101.0	1.3
cassava4.1_009990m	Lumen Thiol Oxidoreductase 1	34.7	25.1	1.4
cassava4.1_010064m	uncharacterized protein	20.0	16.1	1.2
cassava4.1_010090m	DNA-DAMAGE-REPAIR/TOLERATION PROTEIN 111	6.0	3.6	1.7
cassava4.1_010186m	lysophosphatidylcholine acyltransferase / lyso-PAF acetyltransferase	17.9	14.6	1.2
cassava4.1_010200m	uncharacterized protein	8.2	5.6	1.5
cassava4.1_010254m	exostosin family protein	6.4	3.9	1.6
cassava4.1_010268m	aldehyde dehydrogenase 10A8	61.2	42.7	1.4
cassava4.1_010277m	SIGNAL PEPTIDE PEPTIDASE-LIKE 2	9.6	7.5	1.3
cassava4.1_010321m	uncharacterized oxidoreductase	36.0	24.0	1.5
cassava4.1_010355m	TRX fold-containing protein	52.5	29.7	1.8
cassava4.1_010405m	embryo defective 1688	33.6	30.3	1.1
cassava4.1_010424m	putative anion-transporting ATPase	21.2	16.2	1.3
cassava4.1_010438m	putative cathepsin B-like cysteine protease	204.7	135.9	1.5
cassava4.1_010528m	NAC domain containing protein 100	25.2	14.5	1.7
cassava4.1_010557m	esterase/lipase domain-containing protein	4.7	3.0	1.5
cassava4.1_010558m	LOSS OF GDU 2	40.1	28.2	1.4
cassava4.1_010606m	lysophosphatidylcholine acyltransferase / lyso-PAF acetyltransferase	17.9	14.4	1.2
cassava4.1_010675m	uncharacterized protein	73.1	37.5	2.0
cassava4.1_010714m	Pyridoxal phosphate-dependent transferase domain-containing protein	8.2	4.6	1.8
cassava4.1_010720m	EMBRYO DEFECTIVE 3126	212.6	168.7	1.3
cassava4.1_010740m	BYPASS 2	147.7	120.1	1.2
cassava4.1_010753m	BYPASS 2	416.8	356.1	1.2
cassava4.1_010755m	BYPASS 2	147.7	120.1	1.2
cassava4.1_010759m	BYPASS 2	147.7	120.1	1.2

cassava4.1_010762m	BYPASS 2	416.8	356.1	1.2
cassava4.1_010764m	BYPASS 2	147.7	120.1	1.2
cassava4.1_010771m	BYPASS 2	147.7	120.1	1.2
cassava4.1_010782m	acetylglutamate kinase	30.1	18.1	1.7
cassava4.1_010946m	2-alkenal reductase	23.3	14.0	1.7
cassava4.1_010989m	GroES-like zinc-binding alcohol dehydrogenase-like protein	56.1	48.0	1.2
cassava4.1_011169m	DNAJ heat shock protein-like protein	12.0	6.5	1.8
cassava4.1_011217m	uncharacterized protein	37.8	28.6	1.3
cassava4.1_011300m	arginase	25.1	17.3	1.5
cassava4.1_011344m	Sterile alpha motif (SAM) domain-containing protein	25.7	20.8	1.2
cassava4.1_011439m	uvrB/uvrC motif-containing protein	73.3	57.4	1.3
cassava4.1_011563m	xyloglucan:xyloglucosyl transferase	305.0	222.0	1.4
cassava4.1_011726m	tryptophanyl-tRNA synthetase	32.0	20.3	1.6
cassava4.1_011744m	uncharacterized protein	62.9	52.0	1.2
cassava4.1_011798m	enoyl-CoA hydratase	13.9	8.9	1.6
cassava4.1_011868m	EMBRYO DEFECTIVE 2024	56.7	37.5	1.5
cassava4.1_011898m	aquaporin PIP1-2	217.6	136.9	1.6
cassava4.1_011921m	Sterile alpha motif (SAM) domain-containing protein	25.2	20.4	1.2
cassava4.1_011997m	beta-ketoacyl reductase 1	23.7	19.8	1.2
cassava4.1_012041m	RNA recognition motif-containing protein	12.9	8.4	1.5
cassava4.1_012111m	putative epoxide hydrolase	126.9	83.4	1.5
cassava4.1_012117m	acyl-CoA-binding domain 3	17.6	11.4	1.5
cassava4.1_012173m	hydrolase, alpha/beta fold family protein	27.4	13.4	2.0
cassava4.1_012259m	THO complex subunit 3	13.6	11.4	1.2
cassava4.1_012304m	uracil dna glycosylase	23.7	11.0	2.2
cassava4.1_012385m	BYPASS 1	7.9	5.1	1.5
cassava4.1_012443m	histone acetyltransferase MYST1	27.7	24.4	1.1
cassava4.1_012475m	PALE CRESS	33.7	21.3	1.6
cassava4.1_012524m	nil	245.1	174.2	1.4
cassava4.1_012565m	dynein light chain type 1-like protein	35.2	25.3	1.4

cassava4.1_012585m	endonuclease 4	20.0	14.2	1.4
cassava4.1_012725m	protein transport protein SEC13	66.9	52.2	1.3
cassava4.1_012774m	protein kinase family protein	3.2	2.6	1.2
cassava4.1_012812m	endonuclease 4	18.3	12.9	1.4
cassava4.1_012946m	translocon at inner membrane of chloroplasts 21	59.0	44.3	1.3
cassava4.1_013147m	coatomer subunit epsilon-1	22.4	17.6	1.3
cassava4.1_013162m	cofactor assembly of complex C	32.4	26.2	1.2
cassava4.1_013243m	putative aquaporin PIP1-4	214.9	134.1	1.6
cassava4.1_013272m	PIGMENT DEFECTIVE 329	74.5	65.0	1.1
cassava4.1_013368m	chloroplastic acetylcoenzyme A carboxylase 1	27.3	19.4	1.4
cassava4.1_013428m	proteasome assembly chaperone 2	10.7	9.1	1.2
cassava4.1_013455m	nuclear factor Y, subunit C11	13.9	7.0	2.0
cassava4.1_013601m	uncharacterized protein	4.3	2.3	1.9
cassava4.1_013610m	Integral membrane Yip1-like protein	27.9	23.6	1.2
cassava4.1_013623m	ZIP metal ion transporter	13.6	11.0	1.2
cassava4.1_013629m	appr-1-p processing enzyme family protein	23.5	14.4	1.6
cassava4.1_013634m	appr-1-p processing enzyme family protein	23.5	14.4	1.6
cassava4.1_013636m	uncharacterized protein	249.3	187.7	1.3
cassava4.1_013642m	cullin 1	10.5	6.3	1.7
cassava4.1_013730m	proteasome subunit beta type-7-A	50.6	39.6	1.3
cassava4.1_013736m	proteasome subunit beta type-7-A	50.6	39.6	1.3
cassava4.1_013754m	proteasome subunit beta type-7-A	50.6	39.6	1.3
cassava4.1_013796m	BTB-POZ and MATH domain 2	7.9	3.9	2.0
cassava4.1_013865m	forkhead-associated 2	20.3	16.0	1.3
cassava4.1_013902m	ENHANCED RESPONSE TO ABA 1	17.6	8.9	2.0
cassava4.1_014049m	4,5-DOPA dioxygenase extradiol-like protein	38.3	23.8	1.6
cassava4.1_014095m	30S ribosomal protein S20	224.1	138.9	1.6
cassava4.1_014132m	sphingoid base hydroxylase 2	62.7	47.5	1.3
cassava4.1_014141m	acyl-activating enzyme 18	34.2	19.6	1.7

cassava4.1_014156m	calcium-dependent lipid-binding domain-containing protein	5.9	4.1	1.5
cassava4.1_014159m	GTP-binding protein	19.9	16.8	1.2
cassava4.1_014211m	leucine-rich repeat-containing protein	11.0	8.8	1.2
cassava4.1_014271m	putative inactive purple acid phosphatase 1	12.4	8.2	1.5
cassava4.1_014432m	triosephosphate isomerase	65.8	44.1	1.5
cassava4.1_014490m	HIGH PLOIDY2	9.0	4.5	2.0
cassava4.1_014518m	reticulon-like protein B5	94.4	54.9	1.7
cassava4.1_014528m	triosephosphate isomerase	66.3	44.4	1.5
cassava4.1_014675m	peroxin 19-1	25.9	20.1	1.3
cassava4.1_014729m	phosphomannomutase	13.5	4.6	2.9
cassava4.1_014767m	aquaporin TIP4-1	42.2	10.3	4.1
cassava4.1_014852m	nil	16.8	12.8	1.3
cassava4.1_015020m	uncharacterized protein	11.6	8.6	1.3
cassava4.1_015036m	uncharacterized protein	12.2	6.5	1.9
cassava4.1_015057m	uncharacterized protein	11.6	8.7	1.3
cassava4.1_015067m	PLATZ transcription factor domain-containing protein	30.6	10.4	2.9
cassava4.1_015075m	leucine-rich repeat-containing protein	9.6	7.8	1.2
cassava4.1_015082m	PLATZ transcription factor domain-containing protein	30.6	10.4	2.9
cassava4.1_015085m	glutathione peroxidase	137.3	74.4	1.8
cassava4.1_015113m	RNA recognition motif-containing protein	57.2	37.6	1.5
cassava4.1_015140m	uncharacterized protein	11.6	8.7	1.3
cassava4.1_015154m	aquaporin PIP1-2	184.4	109.3	1.7
cassava4.1_015265m	ribosomal protein L34e-like protein	21.3	17.8	1.2
cassava4.1_015305m	cofactor assembly of complex C	33.5	26.8	1.3
cassava4.1_015385m	protein CWC15	57.0	43.4	1.3
cassava4.1_015531m	nuclear factor Y, subunit C10	32.3	23.8	1.4
cassava4.1_015603m	vacuolar protein sorting 26A	56.9	43.7	1.3
cassava4.1_015690m	uncharacterized protein	78.3	59.4	1.3
cassava4.1_015725m	ovate family protein 16	20.3	7.5	2.7

cassava4.1_015805m	vesicle-associated membrane protein 714	39.6	30.6	1.3
cassava4.1_015846m	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase 3	171.5	100.8	1.7
cassava4.1_015912m	Tetraspanin family protein	46.0	30.0	1.5
cassava4.1_015960m	ribosomal protein L7/L12 domain-containing protein	20.4	14.6	1.4
cassava4.1_016001m	uncharacterized protein	49.6	37.4	1.3
cassava4.1_016208m	C2H2 and C2HC zinc finger-containing protein	43.5	27.0	1.6
cassava4.1_016212m	C2H2 and C2HC zinc finger-containing protein	43.5	27.0	1.6
cassava4.1_016225m	RAB GTPase-8	10.7	8.0	1.3
cassava4.1_016254m	Ras-related protein Rab-8A	10.7	8.0	1.3
cassava4.1_016358m	Sugar isomerase (SIS) family protein	169.6	109.8	1.5
cassava4.1_016487m	adenylate cyclase	24.4	13.9	1.8
cassava4.1_016490m	uncharacterized protein	103.6	92.2	1.1
cassava4.1_016518m	ENHANCED RESPONSE TO ABA 1	20.8	10.3	2.0
cassava4.1_016572m	uncharacterized protein	81.6	58.5	1.4
cassava4.1_016579m	RNA polymerase Rpb7 N-terminal domain-containing protein	14.6	6.9	2.1
cassava4.1_016590m	cyclin p4	4.9	3.1	1.6
cassava4.1_016606m	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase 3	193.0	113.8	1.7
cassava4.1_016630m	TBP-associated factor 7	30.3	22.7	1.3
cassava4.1_016688m	RING-H2 finger protein ATL72	32.3	20.3	1.6
cassava4.1_016785m	cyclic phosphodiesterase	64.0	52.6	1.2
cassava4.1_016788m	Isochorismatase family protein	19.2	12.8	1.5
cassava4.1_016864m	GTP-binding protein SAR1	46.0	34.1	1.3
cassava4.1_016887m	uncharacterized protein	37.5	26.0	1.4
cassava4.1_016891m	uncharacterized protein	37.5	26.0	1.4
cassava4.1_016917m	uncharacterized protein	37.2	29.4	1.3
cassava4.1_017041m	40S ribosomal protein S15-3	80.9	50.2	1.6
cassava4.1_017070m	Mediator complex, subunit Med10	42.9	38.2	1.1
cassava4.1_017184m	Mediator complex, subunit Med10	42.9	38.1	1.1
cassava4.1_017321m	ubiquitin-conjugating enzyme 20	12.3	8.3	1.5

cassava4.1_017335m	signal peptidase, endoplasmic reticulum-type	13.0	9.1	1.4
cassava4.1_017439m	nil	38.9	8.0	4.8
cassava4.1_017490m	endo/excinuclease amino terminal domain-containing protein	11.9	6.8	1.8
cassava4.1_017502m	uncharacterized protein	86.6	60.9	1.4
cassava4.1_017610m	cyclic phosphodiesterase	38.0	31.3	1.2
cassava4.1_017883m	ADP-ribosylation factor-like A1A	6.7	3.6	1.9
cassava4.1_017887m	uncharacterized protein	84.2	51.2	1.6
cassava4.1_017946m	EMBRYO DEFECTIVE 3105	244.3	173.5	1.4
cassava4.1_018056m	ribosomal protein S17	843.9	632.0	1.3
cassava4.1_018086m	Surfeit locus protein 5 subunit 22 of Mediator complex	37.1	23.0	1.6
cassava4.1_018094m	uncharacterized protein	28.8	16.2	1.8
cassava4.1_018113m	Surfeit locus protein 5 subunit 22 of Mediator complex	37.1	23.0	1.6
cassava4.1_018321m	RNA recognition motif-containing protein	24.5	17.5	1.4
cassava4.1_018577m	nil	39.4	8.0	4.9
cassava4.1_018698m	ACCLIMATION OF PHOTOSYNTHESIS TO ENVIRONMENT	64.3	35.2	1.8
cassava4.1_018744m	Got1/Sft2-like vescicle transport-like protein	18.7	15.7	1.2
cassava4.1_018791m	nil	92.9	34.1	2.7
cassava4.1_019119m	uncharacterized protein	48.1	33.5	1.4
cassava4.1_019231m	Ribosomal protein L20	44.3	28.8	1.5
cassava4.1_019372m	ribonuclease H2 subunit B	9.3	5.1	1.8
cassava4.1_019554m	RNA recognition motif-containing protein	21.1	14.3	1.5
cassava4.1_019581m	nuclear/nucleolar GTPase 2	7.9	4.8	1.6
cassava4.1_019583m	Ubiquitin domain-containing protein	14.8	9.5	1.6
cassava4.1_019857m	complex 1-LYR domain-containing protein	36.0	22.1	1.6
cassava4.1_019980m	S-adenosyl-L-methionine-dependent methyltransferase domain-containing protein	47.6	15.9	3.0

cassava4.1_019983m	ubiquitin-related modifier 1-2	117.5	70.9	1.7
cassava4.1_019988m	RING/FYVE/PHD zinc finger-containing protein	57.2	36.0	1.6
cassava4.1_020405m	nil	60.8	40.2	1.5
cassava4.1_020768m	AUTOPHAGY 8H	33.7	16.2	2.1
cassava4.1_020807m	nil	529.7	252.0	2.1
cassava4.1_020834m	Acytochrome-c oxidase/ electron carrier	140.3	99.5	1.4
cassava4.1_021338m	VEIN PATTERNING 1	27.5	10.3	2.7
cassava4.1_021438m	tetratricopeptide repeat domain-containing protein	17.2	6.9	2.5
cassava4.1_021638m	F-box/kelch-repeat protein	5.5	3.3	1.6
cassava4.1_021649m	uncharacterized protein	99.1	65.0	1.5
cassava4.1_021848m	ribonuclease H2 subunit B	7.5	3.9	1.9
cassava4.1_021953m	DEAD-box ATP-dependent RNA helicase 58	8.0	6.4	1.3
cassava4.1_022121m	nil	10.3	6.3	1.6
cassava4.1_022589m	uncharacterized protein	12.5	8.4	1.5
cassava4.1_023598m	tetratricopeptide repeat domain-containing protein	19.5	9.7	2.0
cassava4.1_024108m	EMBRYO DEFECTIVE 3137	805.4	588.0	1.4
cassava4.1_024819m	RNA recognition motif-containing protein	81.2	67.7	1.2
cassava4.1_024984m	ENHANCED RESPONSE TO ABA 1	5.7	3.4	1.7
cassava4.1_025529m	COP9 signalosome subunit 6A	33.6	27.9	1.2
cassava4.1_025867m	CONSTANS	11.0	7.8	1.4
cassava4.1_026057m	methyl-CPG-binding domain 6	14.2	10.5	1.4
cassava4.1_027171m	nil	20.4	13.0	1.6
cassava4.1_027714m	NON-YELLOWING 1	92.1	45.9	2.0
cassava4.1_028175m	copper transporter 5	52.3	37.8	1.4
cassava4.1_028182m	uncharacterized protein	240.2	134.3	1.8
cassava4.1_028407m	Rossmann-fold NAD(P)-binding domain-containing protein	44.3	18.0	2.5
cassava4.1_029700m	uncharacterized protein	23.4	11.2	2.1
cassava4.1_029956m	RING/FYVE/PHD zinc finger-containing protein	35.4	20.9	1.7

cassava4.1_030168m	peroxisomal NAD carrier	20.4	14.3	1.4
cassava4.1_030533m	acyl-activating enzyme 18	17.1	8.1	2.1
cassava4.1_030584m	uncharacterized protein	9.4	4.2	2.2
cassava4.1_030620m	pectate lyase	6.5	4.2	1.5
cassava4.1_030637m	histone H3	27.4	13.8	2.0
cassava4.1_030953m	uncharacterized protein	9.4	4.7	2.0
cassava4.1_031365m	uncharacterized protein	31.1	23.8	1.3
cassava4.1_031663m	thioredoxin O1	40.1	28.1	1.4
cassava4.1_033114m	RING/FYVE/PHD zinc finger-containing protein	12.7	4.1	3.1
cassava4.1_033581m	cytochrome P450, family 94, subfamily B, polypeptide 3	45.1	30.2	1.5
cassava4.1_033774m	embryo defective 1067	14.6	10.2	1.4
cassava4.1_034056m	alpha/beta-hydrolase domain-containing protein	17.4	9.5	1.8
cassava4.1_034184m	P-loop containing nucleoside triphosphate hydrolase-like protein	36.2	26.4	1.4
cassava4.1_034296m	F-box/kelch-repeat protein	9.6	6.4	1.5
cassava4.1_034375m	uncharacterized protein	17.1	11.0	1.5

Appendix 7. Functional annotation of genes overexpressed in resistant cassava relative to susceptible cassava

Gene Id	Gene name	Functional category
cassava4.1_001124m	disease resistance protein RPP8	NBS-LRR
cassava4.1_002792m	putative disease resistance RPP13-like protein 1	NBS-LRR
cassava4.1_005201m	cytosolic invertase 2	Carbohydrate metabolism
cassava4.1_005786m	cytochrome P450, family 714, subfamily A, polypeptide 1	not availbale
cassava4.1_007337m	uridine kinase-like 3	not availbale
cassava4.1_007745m	putative LRR receptor-like serine/threonine-protein kinase	signal perception / transduction
cassava4.1_008411m	uncharacterized protein	not availbale
cassava4.1_009835m	Afadin/alpha-actinin-binding protein	not availbale
cassava4.1_010650m	S-adenosyl-L-methionine-dependent methyltransferase-like protein	not availbale
cassava4.1_011012m	EMBRYO DEFECTIVE 2799	not availbale
cassava4.1_011133m	malate dehydrogenase 1	Photosynthesis
cassava4.1_011541m	putative S-acyltransferase	not availbale
cassava4.1_012476m	uncharacterized protein	not availbale
cassava4.1_012887m	transport protein SEC20	intracellular macromolecule / vesicle transport
cassava4.1_014263m	TIR-NBS-LRR class disease resistance protein	NBS-LRR
cassava4.1_014860m	uncharacterized protein	not availbale
cassava4.1_015251m	GLABROUS INFLORESCENCE STEMS	not availbale
cassava4.1_015326m	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	Protein folding
cassava4.1_015345m	NIFU-like protein 2	Photosynthesis
cassava4.1_016199m	D-tyrosyl-tRNA(Tyr)deacylase	not availbale
cassava4.1_016724m	diphosphomevalonate decarboxylase	not availbale
cassava4.1_016754m	frataxin	not availbale
cassava4.1_016830m	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	protein folding
cassava4.1_019851m	eukaryotic translation initiation factor 3E	protein synthesis

cassava4.1_020261m	not available	not availbale
cassava4.1_020264m	not available	not availbale
cassava4.1_020424m	lsd one like 2	defence response
cassava4.1_021434m	Elongation factor 1-beta 2	protein synthesis
cassava4.1_022120m	MIRO-related GTP-ase 2	not availbale
cassava4.1_023508m	putative disease resistance RPP13-like protein 1	NBS-LRR
cassava4.1_024050m	peptidase-S24/S26 domain-containing protein	protein degradation
cassava4.1_024649m	DEFECTIVE IN MERISTEM SILENCING 3	not availbale
cassava4.1_026203m	alpha-mannosidase 2	Carbohydrate metabolism
cassava4.1_028445m	disease resistance protein RPM1	NBS-LRR
cassava4.1_031755m	organic cation/carnitine transporter4	Metal ion transport
cassava4.1_033369m	UDP-glycosyltransferase-like protein	not availbale

Appendix 8. Types and functions of expressed genes of interest grouped under each of 43 functional classes.

Functional class	Constituent proteins / protein functions and processes
Abiotic stress response	ABA-signalling / response; Abiotic stress responsive gene expression; Heat, drought, oxidative, osmotic and chilling tolerance
Abiotic stress response regulation	Regulation of ABA-induced processes and, ABA biosynthesis.
Antioxidant defence / toxin efflux	ROS-metabolising enzymes; Toxic heavy metal resistance or efflux
	Glutathione redox system; Thioredoxin redox system; Ascorbate peroxidase and other peroxidases; Superoxide dismutase
ATP hydrolysis / ATP-dependent processes	ATP-coupled transmembrane transport; Ca^{2+}-transporting ATPase
Auxin response	Auxin efflux / influx transporters, Positive regulation of auxin response
	Auxin-responsive proteins; Degradation of the auxin repressor Aux/IAA
Carbohydrate metabolism	Catabolism of simple and complex sugars
	Biosynthesis of single and complex sugar molecules
Cell wall formation / strengthening	Biosynthesis of structural cell wall component
	Deposition / re-arrangement of cell wall components
Cell wall degradation	Proteases
Cell wall loosening / expansion	Expansin proteins; Cell wall modifying enzymes
Cell wall expansion regulation	Regulation / inhibition of cell expansion
Defence response	Anti-microbial agents; basal defence; SA-mediated resistance; Hypersensitive response; NBS-LRR and Pathogenesis-related genes
Ethylene signalling / response	Ethylene responsive proteins; Ethylene biosynthesis
	Positive regulation of ethylene response
Ethyelene signaling regulation	Regulation / inhibition of cell expansion
Flowering / reproductive development	Pollen tube growth; Flower formation and development
Gene silencing	Dice-like proteins; RNA-dependent RNA polymerases
Inhibition / regulation of oxidative cell death	Negative regulation of PCD
Intercellular macromolecule / vesicle transport	Docking exocytic vesicles to plasma membrane; Protein export
Intracellular macromolecule / vesicle transport	Cytoplasm to vacuole transport of vesicles; Nuclear import
	ER to Golgi vesicular trafficking; Retrograde Golgi-to-ER transport
Jasmonic acid pathway	Jasmonic acid biosynthesis

Jasmonic acid regulation	Negative regulation of Jasmonic acid response
Lipid metabolism	Lipid biosynthetic pathways; Lipid catabolism
Metabolite / nutrient transport	Intracellular metabolite transport; Symporter activity; Nutrient transport
Metal ion transport	Cation and heavy metal transport
Oxidative cell death/ cell death signalling	Unfolded protein response; Polyamine oxidases
	Cysteine proteases; ROS-producing enzymes
Oxidative phosphorylation	ATP synthase; Cytochrome c oxidase; NADH dehydrogenenase
Phenylpropanoid pathway	Genes of phenylpropanoid biosynthesis pathway
Photosynthesis	Photosynthetic pathway and Photorespiratory enzymes
Phenylpropanoid pathway	Chlorophyll biosynthesis enzymes
Plant growth and development	Cell division; Plant cell and tissue development
	Plant embryo development; Seed germination; Brassinosteroid-mediated growth; Regulation of gibberellin-induced growth
Plant growth regulation	Transcriptional regulation of germination and flowering
	Regulation / suppression of cell division and differentiation
Virus replication / movement	ADP-ribosylation factor; Beta-1,3-glucanase; Acetyltransferase NSI
Protein degradation	Ubiquitin activating enzyme, conjugase and ligase
	Proteases; SUMO conjugating enzyme; Ubiquitin
Protein folding	Heat shock proteins; Calreticulin; DnaJ-domain proteins; chaperonins
Protein synthesis	Ribosomal proteins; Aminoacyl transferases
	Translation initiation and elongation factors
Regulation / inhibition of photosynthesis	Regulation of chlorophyll biosynthesis; Chlorophyll degradation
	Regulation of photosynthesis and photorespiration
Regulation of auxin response	Aux/IAA proteins; Auxin response factors; Regulation of auxin homeostasis
Regulation of reproductive development	Regulation of photoperiod flowering response
regulation of oxidative phosphorylation	Uncoupling of oxidative phosphorylation
Regulation of defence response	Negative regulation of SA; MLO proteins
Regulation of signal perception / transduction	Lipid phosphate and protein phosphatases; Kinase inhibitors; Inhibition of brassinosteroid signalling
Root hair development	Trichome branching; root tip growth
Root hair growth regulation	Regulation of trichome branching
Signal perception / transduction	Protein kinases; Lipid kinases; cAMP-dependent signaling

	Brassinosteroid signalling; Calcium sensing / calcium-dependent signalling; LRR-rich receptor-like kinases;
	Histidine kinases; WRKY transcription factors

Appendix 9. Chromosomal locations and predicted functions of genes containing candidate CBSD resistance alleles introgressed from *Manihot glaziovii*.

Gene ID	Chromosome location	Gene function	Function category
cassava4.1_008846m	Chr.03	BTB-POZ and MATH domain 2	Protein degradation
cassava4.1_014681m	Chr.03	BI1-like protein	Regulation of oxidative cell death
cassava4.1_020104m	Chr.13	protein SKIP34	unknown
cassava4.1_010945m	Chr.13	myb domain protein r1	Abiotic stress tolerance
cassava4.1_003376m	Chr.13	poly(A) binding protein 4	protein synthesis
cassava4.1_008450m	Chr.13	SUPERROOT 1	Defence response; Auxin homeostasis
cassava4.1_014049m	Chr.13	4,5-DOPA dioxygenase extradiol-like protein	Metabolism
cassava4.1_013167m	Chr.04	histone deacetylase 2C	Abiotic stress regulation
cassava4.1_020053m	Chr.04	uncharacterized protein	*Trans*criptional regulation
cassava4.1_012232m	Chr.04	pathogenesis-related thaumatin-like protein	Pathogenesis-related
cassava4.1_008317m	Chr.04	oligouridylate binding protein 1B	na
cassava4.1_007126m	Chr.04	dsRNA-binding protein 2	Gene silencing
cassava4.1_002363m	Chr.04	4-hydroxy-3-methylbut-2-enyl diphosphate synthase	Defence response
cassava4.1_008550m	Chr.04	geminivirus rep interacting kinase 2	Plant virus replication
cassava4.1_008714m	Chr.04	general *trans*cription factor II H2	*Trans*criptional regulation
cassava4.1_001909m	Chr.04	E3 SUMO-protein ligase SIZ1	Abiotic stress response
cassava4.1_012602m	Chr.04	20S proteasome beta subunit G1	Protein degradation
cassava4.1_013461m	Chr.04	L-ascorbate peroxidase	Antioxidant defence
cassava4.1_019911m	Chr.04	histone H4	*Trans*criptional regulation
cassava4.1_001727m	Chr.04	hercules receptor kinase 1	Cell wall loosening / Cell elongation

Appendix 10. Biological functions and feature importance values of 387 susceptibility associated genes (SAGs) uniquely induced in CBSD susceptible accessions. Predictor locus Identity (ID), chromosome locations and positions, of genes for which predictor locus was not found, were filled out as 'Not available'

Predictor locus ID	Chromosome location of predictor	Chromosome position of predictor locus	Target SAG	Feature importance (%IncMSE)	Gene name	Functional category
scaffold06028_223093	Chr.17	24525608	cassava4.1_000135m	43.3	tetratricopeptide repeat-containing protein	Abiotic stress response / signaling
scaffold03834_175945	Chr.05	8488356	cassava4.1_000145m	119.2	tetratricopeptide repeat domain protein	Abiotic stress response / signaling
scaffold04457_1316529	Chr.01	29850165	cassava4.1_000193m	19.6	tetratricopeptide repeat-containing protein	Abiotic stress response / signaling
scaffold04457_238146	Chr.01	30897240	cassava4.1_001314m	355.1	lipoxygenase 2	Lipid metabolism
scaffold06598_394856	Chr.16	411057	cassava4.1_001568m	26.1	pyruvate, phosphate dikinase 1	Abiotic stress response / signaling
scaffold03410_133676	Chr.14	5182856	cassava4.1_002307m	333.2	*Trans*ketolase	Abiotic stress response / signaling
scaffold05280_532172	Chr.02	1514204	cassava4.1_003132m	114.4	ABA DEFICIENT 1	Abiotic stress response / signaling
scaffold07478_52990	Chr.14	11891412	cassava4.1_003741m	31.6	RGA-like 2	Abiotic stress response / signaling
scaffold07117_41954	Chr.08	25361829	cassava4.1_004852m	19.1	UPF0051 protein ABCI8	Abiotic stress response / signaling
scaffold06140_116047	Chr.05	26662320	cassava4.1_005262m	28.5	phospho-2-dehydro-3-deoxyheptonate aldolase 1	Abiotic stress response / signaling
scaffold03802_435828	Chr.08	30257615	cassava4.1_006003m	51.2	putative WRKY *trans*cription factor 47	Abiotic stress response / signaling
scaffold02915_252964	Chr.04	23820628	cassava4.1_006165m	47.9	*NA	Not available
scaffold01934_46632	Chr.06	9197301	cassava4.1_006297m	11340.4	catalase 2	Antioxidant defence / toxin efflux
scaffold02906_224109	NA	NA	cassava4.1_006302m	2566.0	catalase 2	Antioxidant defence / toxin efflux
scaffold03614_2790931	Chr.09	26864174	cassava4.1_006778m	31.4	hydroquinone glucosyl*trans*ferase	Antioxidant defence / toxin efflux
scaffold02906_224109	NA	NA	cassava4.1_007219m	3374.9	catalase 2	Antioxidant defence / toxin efflux
scaffold06916_552733	Chr.02	9792355	cassava4.1_007228m	213.1	APS reductase 3	Antioxidant defence / toxin efflux
scaffold00080_434194	Chr.05	524970	cassava4.1_007913m	23.2	highly ABA-induced PP2C gene 3	Abiotic stress response / signaling
scaffold06932_17430	NA	NA	cassava4.1_008790m	168.8	protein TIFY 6B	Jasmonic acid regulation

scaffold06932_17430	NA	NA	cassava4.1_008816m	129.4	protein TIFY 6B	Jasmonic acid regulation
scaffold06932_17430	NA	NA	cassava4.1_008829m	105.3	protein TIFY 6B	Jasmonic acid regulation
scaffold02477_307417	Chr.14	895459	cassava4.1_010102m	17.9	related to AP2 4	Abiotic stress response / signaling
scaffold05875_1285635	Chr.04	27506993	cassava4.1_010501m	15392.7	catalase 2	Antioxidant defence / toxin efflux
scaffold03614_2790931	Chr.09	26864174	cassava4.1_010526m	41.7	related to AP2 4	Abiotic stress response / signaling
scaffold04065_263860	Chr.02	11908286	cassava4.1_010620m	366957.6	thiazole biosynthetic enzyme	Abiotic stress response / signaling
scaffold04627_213960	Chr.01	5922948	cassava4.1_010863m	126.2	2-alkenal reductase	Antioxidant defence / toxin efflux
scaffold03299_156447	Chr.18	4258849	cassava4.1_010919m	26.2	2-alkenal reductase	Antioxidant defence / toxin efflux
scaffold04065_263860	Chr.02	11908286	cassava4.1_010960m	235463.2	thiazole biosynthetic enzyme	Abiotic stress response / signaling
scaffold04043_590039	Chr.02	5188094	cassava4.1_011142m	16.5	peroxidase 2	Antioxidant defence / toxin efflux
scaffold02307_425040	Chr.13	27292344	cassava4.1_011281m	41247.2	galactinol synthase 4	Abiotic stress response / signaling
scaffold06582_313141	Chr.15	8807211	cassava4.1_011321m	49525.5	galactinol synthase 4	Abiotic stress response / signaling
scaffold03614_2790931	Chr.09	26864174	cassava4.1_011557m	22.9	related to AP2 4	Abiotic stress response / signaling
scaffold04651_24982	Chr.08	39189	cassava4.1_011680m	389.6	WRKY DNA-binding protein 70	Abiotic stress response / signaling
scaffold00710_20311	NA	NA	cassava4.1_011839m	34.1	peroxiredoxin-2F	Antioxidant defence / toxin efflux
scaffold04457_414560	Chr.01	30724861	cassava4.1_011970m	24.2	pathogenesis-related thaumatin-like protein	Pathogenesis-related
scaffold04627_213957	Chr.01	5922945	cassava4.1_012256m	105.9	2-alkenal reductase	Antioxidant defence / toxin efflux
scaffold01551_1204607	Chr.15	2505375	cassava4.1_012383m	37.1	pathogenesis-related thaumatin-like protein	Pathogenesis-related
scaffold02264_290581	Chr.09	29118144	cassava4.1_012402m	361.8	REDUCED SUGAR RESPONSE 4	Abiotic stress response / signaling; Antioxidant defence / toxin efflux
scaffold02824_17931	NA	NA	cassava4.1_012768m	17.0	male sterility MS5 family protein	Antioxidant defence / toxin efflux
scaffold04457_437346	Chr.01	30701785	cassava4.1_012801m	7912.1	thioredoxin-like protein CDSP32	Antioxidant defence / toxin efflux
scaffold03614_2807693	Chr.09	26880398	cassava4.1_012837m	156.3	tetratricopeptide repeat domain-containing protein	Antioxidant defence / toxin efflux
scaffold00341_381542	NA	NA	cassava4.1_012905m	405.8	thioredoxin-like 1-1	Antioxidant defence / toxin efflux
scaffold05628_52733	NA	NA	cassava4.1_013185m	18.2	heat shock factor 4	Antioxidant defence / toxin efflux

scaffold03614_1330135	Chr.09	25385110	cassava4.1_013214m	31.5	ethylene-responsive transcription factor ERF105	Abiotic stress response / signaling
scaffold03802_435828	Chr.08	30257615	cassava4.1_013315m	37.3	ethylene-responsive *trans*cription factor ERF105	Abiotic stress response / signaling
scaffold04457_415798	Chr.01	30726099	cassava4.1_013374m	134.2	aquaporin PIP2-7	Abiotic stress response / signaling
scaffold06446_149567	Chr.01	2017920	cassava4.1_013461m	589.8	L-ascorbate peroxidase	Antioxidant defence / toxin efflux
scaffold04457_536918	Chr.01	30613031	cassava4.1_013471m	189.1	aquaporin PIP2-7	Abiotic stress response / signaling
scaffold05965_194373	Chr.10	19502769	cassava4.1_013608m	2333.8	ABA DEFICIENT 2	Abiotic stress response / signaling
scaffold06914_191721	Chr.07	1081519	cassava4.1_013704m	40.9	ABA DEFICIENT 2	Abiotic stress response / signaling
scaffold02817_124689	NA	NA	cassava4.1_013735m	27957.4	NONPHOTOCHEMICAL QUENCHING 4	Antioxidant defence / toxin efflux
scaffold03175_92120	Chr.02	1008132	cassava4.1_013775m	3040.8	DNAJ homologue 2	Abiotic stress response / signaling
scaffold06028_65157	Chr.17	24703421	cassava4.1_013784m	640.4	Fe superoxide dismutase 2	Antioxidant defence / toxin efflux
scaffold06028_65157	Chr.17	24703421	cassava4.1_013844m	621.0	Fe superoxide dismutase 2	Antioxidant defence / toxin efflux
scaffold04457_437422	Chr.01	30701861	cassava4.1_013880m	22.2	ethylene-responsive element binding protein	Abiotic stress response / signaling
scaffold04374_102797	Chr.16	6132404	cassava4.1_013921m	60.8	2-cysteine peroxiredoxin B	Antioxidant defence / toxin efflux
scaffold04374_102797	Chr.16	6132404	cassava4.1_013927m	39.0	2-cysteine peroxiredoxin B	Antioxidant defence / toxin efflux
scaffold06609_334239	Chr.01	22176061	cassava4.1_014035m	192.1	WRKY DNA-binding protein 65	Abiotic stress response / signaling
scaffold06794_62674	Chr.11	19596585	cassava4.1_014109m	479.4	Fe superoxide dismutase 2	Antioxidant defence / toxin efflux
scaffold06914_202348	Chr.07	1092047	cassava4.1_014201m	202.8	peptide met sulfoxide reductase 4	Antioxidant defence / toxin efflux
scaffold05965_194373	Chr.10	19502769	cassava4.1_014227m	154.8	thioredoxin-like protein HCF164	Antioxidant defence / toxin efflux
scaffold05965_194373	Chr.10	19502769	cassava4.1_014244m	233.2	thioredoxin-like protein HCF164	Antioxidant defence / toxin efflux
scaffold03614_2807693	Chr.09	26880398	cassava4.1_014348m	207.3	Thioredoxin-like protein	Antioxidant defence / toxin efflux
scaffold06816_186536	Chr.09	8671217	cassava4.1_014423m	214.4	L-ascorbate peroxidase	Antioxidant defence / toxin efflux
scaffold04457_415927	Chr.01	30726228	cassava4.1_014463m	36.1	glutathione S-transferase	Antioxidant defence / toxin efflux
scaffold03802_724990	Chr.08	29966951	cassava4.1_014507m	69.4	senescence-associated family protein	Abiotic stress response / signaling
scaffold03802_724990	Chr.08	29966951	cassava4.1_014532m	69.5	senescence-associated family protein	Abiotic stress response / signaling
scaffold05009_56009	Chr.07	25923026	cassava4.1_014536m	82.1	aquaporin PIP2-7	Abiotic stress response / signaling

scaffold03115_239748	Chr.06	27506323	cassava4.1_014554m	7583.1	chitinase	Pathogenesis-related
scaffold03787_100619	Chr.12	27521242	cassava4.1_014564m	22.5	aquaporin TIP1-1	Abiotic stress response / signaling
scaffold06816_186536	Chr.09	8671217	cassava4.1_014583m	225.2	L-ascorbate peroxidase	Antioxidant defence / toxin efflux
scaffold06816_186536	Chr.09	8671217	cassava4.1_014642m	237.1	L-ascorbate peroxidase	Antioxidant defence / toxin efflux
scaffold06446_149567	Chr.01	2017920	cassava4.1_014643m	511.1	L-ascorbate peroxidase	Antioxidant defence / toxin efflux
scaffold02264_756080	Chr.09	28678896	cassava4.1_014861m	420.4	Fe superoxide dismutase 2	Antioxidant defence / toxin efflux
scaffold03241_232643	Chr.15	5216253	cassava4.1_015013m	555.0	L-ascorbate peroxidase	Antioxidant defence / toxin efflux
scaffold06446_149567	Chr.01	2017920	cassava4.1_015058m	564.6	L-ascorbate peroxidase	Antioxidant defence / toxin efflux
scaffold05162_777096	NA	NA	cassava4.1_015084m	824.1	glutathione peroxidase 1	Antioxidant defence / toxin efflux
scaffold02421_965637	Chr.03	23842681	cassava4.1_015268m	288.8	glutathione S-*trans*ferase tau 7	Antioxidant defence / toxin efflux
scaffold05965_194373	Chr.10	19502769	cassava4.1_015299m	51.6	thioredoxin-like 2-2	Antioxidant defence / toxin efflux
scaffold00080_31993	Chr.05	934135	cassava4.1_015406m	98.0	glutathione S-*trans*ferase	Antioxidant defence / toxin efflux
scaffold03614_2807742	Chr.09	26880447	cassava4.1_015438m	377.5	peroxiredoxin-2E	Antioxidant defence / toxin efflux
scaffold06711_51379	Chr.12	346526	cassava4.1_015554m	44.9	Superoxide dismutase [Cu-Zn]	Antioxidant defence / toxin efflux
scaffold05965_194373	Chr.10	19502769	cassava4.1_015681m	662.4	NA	Abiotic stress response / signaling
scaffold03614_3997786	Chr.09	28074219	cassava4.1_015698m	18.4	glutathione S-*trans*ferase tau 7	Antioxidant defence / toxin efflux
scaffold06446_149567	Chr.01	2017920	cassava4.1_015710m	347.3	L-ascorbate peroxidase	Antioxidant defence / toxin efflux
scaffold04457_536285	Chr.01	30612398	cassava4.1_015739m	134.4	aquaporin PIP2-7	Abiotic stress response / signaling
scaffold05206_104100	Chr.03	22037183	cassava4.1_015741m	15.9	L-ascorbate peroxidase	Antioxidant defence / toxin efflux
scaffold05965_194373	Chr.10	19502769	cassava4.1_015761m	45.6	glutathione S-*trans*ferase TAU 19	Antioxidant defence / toxin efflux
scaffold02658_421512	NA	NA	cassava4.1_015784m	38.0	glutathione S-*trans*ferase TAU 25	Antioxidant defence / toxin efflux
scaffold02817_124689	NA	NA	cassava4.1_015872m	4345.2	glutathione S-*trans*ferase TAU 19	Antioxidant defence / toxin efflux
scaffold03802_685942	Chr.08	30014471	cassava4.1_015875m	139.3	dehydrin COR47	Abiotic stress response / signaling
scaffold03542_237393	Chr.15	7514212	cassava4.1_015909m	60.3	methionine sulfoxide reductase B 2	Antioxidant defence / toxin efflux
scaffold06028_65157	Chr.17	24703421	cassava4.1_016018m	1018.8	Fe superoxide dismutase 2	Antioxidant defence / toxin efflux

scaffold03150_386104	Chr.01	27024277	cassava4.1_016021m	53.9	peroxiredoxin Q	Antioxidant defence / toxin efflux
scaffold02906_271054	Chr.01	29467264	cassava4.1_016023m	58.9	universal stress protein (USP) family protein	Abiotic stress response / signaling
scaffold04080_261521	NA	NA	cassava4.1_016150m	43.7	CAX interacting protein 1	Antioxidant defence / toxin efflux
scaffold03614_1230184	Chr.09	25286604	cassava4.1_016175m	77.2	glutathione S-*trans*ferase	Antioxidant defence / toxin efflux
scaffold03614_2807693	Chr.09	26880398	cassava4.1_016240m	41801.1	NA	Abiotic stress response / signaling
scaffold03896_69700	NA	NA	cassava4.1_016400m	17.6	ABA Insensitive RING Protein 2	Abiotic stress response / signaling
scaffold03827_136801	Chr.18	12077777	cassava4.1_016476m	15.1	glutathione S-*trans*ferase TAU 19	Antioxidant defence / toxin efflux
scaffold06407_209192	Chr.10	25674660	cassava4.1_016531m	422.3	NA	Abiotic stress response / signaling
scaffold04457_415798	Chr.01	30726099	cassava4.1_016600m	148.7	cold shock domain protein 3	Abiotic stress response / signaling
scaffold06711_50895	Chr.12	346042	cassava4.1_016755m	605.7	methionine sulfoxide reductase B 1	Antioxidant defence / toxin efflux
scaffold01664_45767	NA	NA	cassava4.1_016778m	458.0	invertase/pectin methylesterase inhibitor family protein / DC 1.2-like protein	cell expansion regulation / inhibition
scaffold04803_36045	Chr.03	2744935	cassava4.1_016825m	1070.3	thioredoxin F2	Antioxidant defence / toxin efflux
scaffold04151_171845	Chr.04	24602391	cassava4.1_016831m	74.1	stress enhanced protein 2	Abiotic stress response / signaling
scaffold03614_3997786	Chr.09	28074219	cassava4.1_016896m	137.9	OXIDATIVE STRESS 3	Abiotic stress response / signaling
scaffold04457_927944	Chr.01	30214446	cassava4.1_016900m	51.5	L-ascorbate peroxidase	Antioxidant defence / toxin efflux
scaffold06697_120243	Chr.15	12423914	cassava4.1_016962m	38.4	CAX interacting protein 1	Antioxidant defence / toxin efflux
scaffold05965_194373	Chr.10	19502769	cassava4.1_016963m	20929.6	peptidemethionine sulfoxide reductase 1	Antioxidant defence / toxin efflux
scaffold03741_627350	Chr.15	6543455	cassava4.1_017002m	33.3	ankyrin repeat-containing protein	Antioxidant defence / toxin efflux
scaffold02853_512337	Chr.11	5012094	cassava4.1_017076m	525.6	temperature-induced lipocalin	Abiotic stress response / signaling
scaffold04457_415798	Chr.01	30726099	cassava4.1_017096m	1597.0	thioredoxin M4	Antioxidant defence / toxin efflux
scaffold02307_380375	Chr.13	27244943	cassava4.1_017131m	21.8	thioredoxin M4	Antioxidant defence / toxin efflux
scaffold04080_261521	NA	NA	cassava4.1_017424m	73.1	CAX interacting protein 1	Antioxidant defence / toxin efflux
scaffold05875_407537	Chr.04	26618093	cassava4.1_017463m	34.2	zinc finger A20 and AN1 domain-containing stress-associated protein 3	Abiotic stress response / signaling
scaffold05965_194373	Chr.10	19502769	cassava4.1_017660m	38.0	universal stress protein domain-containing protein	Abiotic stress response / signaling

scaffold03115_239748	Chr.06	27506323	cassava4.1_017717m	47.6	universal stress protein domain-containing protein	Abiotic stress response / signaling
scaffold01551_2168405	Chr.15	1572196	cassava4.1_017734m	27.2	thioredoxin Y2	Antioxidant defence / toxin efflux
scaffold03834_281549	Chr.05	8388492	cassava4.1_017966m	4504.3	putative adenine nucleotide alpha hydrolase domain-containing universal stress protein	Abiotic stress response / signaling
scaffold03614_3831103	Chr.09	27899021	cassava4.1_017973m	208.0	peroxiredoxin-2B	Antioxidant defence / toxin efflux
scaffold06711_50895	Chr.12	346042	cassava4.1_017974m	150.6	HSP20-like chaperone	Abiotic stress response / signaling
scaffold03049_663833	Chr.03	474498	cassava4.1_018066m	16.7	HSP20-like chaperone	Abiotic stress response / signaling
scaffold02947_33383	Chr.04	3569323	cassava4.1_018106m	135.6	PYR1-like 2	Abiotic stress response / signaling
scaffold03896_69700	NA	NA	cassava4.1_018151m	1757.8	zinc finger A20 and AN1 domain-containing stress-associated protein 4	Abiotic stress response / signaling
scaffold03896_69700	NA	NA	cassava4.1_018176m	1678.8	zinc finger A20 and AN1 domain-containing stress-associated protein 4	Abiotic stress response / signaling
scaffold02947_33383	Chr.04	3569323	cassava4.1_018283m	39.6	universal stress protein domain-containing protein	Abiotic stress response / signaling
scaffold06512_1253623	Chr.02	8209172	cassava4.1_018294m	67.3	Superoxide dismutase [Cu-Zn]	Antioxidant defence / toxin efflux
scaffold03332_917439	NA	NA	cassava4.1_018633m	42.9	small ubiquitin-like modifier 1	Abiotic stress response / signaling
scaffold03802_724990	Chr.08	29966951	cassava4.1_018634m	45.7	small ubiquitin-like modifier 1	Abiotic stress response / signaling
scaffold06512_1253623	Chr.02	8209172	cassava4.1_018987m	36.3	Superoxide dismutase [Cu-Zn]	Antioxidant defence / toxin efflux
scaffold03614_3618409	Chr.09	27698722	cassava4.1_019018m	30.6	thioredoxin-like protein Clot	Antioxidant defence / toxin efflux
scaffold03802_724990	Chr.08	29966951	cassava4.1_019037m	1860.3	thioredoxin 2	Antioxidant defence / toxin efflux
scaffold04457_415927	Chr.01	30726228	cassava4.1_019051m	1601.1	dehydration-induced protein ERD15	Abiotic stress response / signaling
scaffold02817_124689	NA	NA	cassava4.1_019071m	995.1	dehydration-induced protein ERD15	Abiotic stress response / signaling
scaffold03454_61308	Chr.10	321045	cassava4.1_019079m	92.7	methionine sulfoxide reductase B5	Antioxidant defence / toxin efflux
scaffold00708_214199	Chr.03	28810594	cassava4.1_019182m	15.3	Proteinase inhibitor, propeptide	Not available
scaffold00987_51684	NA	NA	cassava4.1_019244m	25.7	thioredoxin-like protein CXXS1	Antioxidant defence / toxin efflux
scaffold02973_221653	Chr.04	363293	cassava4.1_019396m	978.4	stress enhanced protein 1	Abiotic stress response / signaling
scaffold00318_149363	Chr.17	21386972	cassava4.1_019421m	121.2	C2H2-type zinc finger-containing protein	Abiotic stress response / signaling

scaffold03773_162747	Chr.10	19890759	cassava4.1_019456m	25.9	thioredoxin H1	Antioxidant defence / toxin efflux
scaffold03651_256759	Chr.09	22126665	cassava4.1_019460m	50.7	glutathione peroxidase	Antioxidant defence / toxin efflux
scaffold04083_129611	NA	NA	cassava4.1_019469m	307.0	ralf-like 33	Regulation of abiotic stress response / signaling
scaffold00080_33132	NA	NA	cassava4.1_019672m	75.7	cysteine proteinase inhibitor 5	Defence response
scaffold03834_233851	Chr.05	8428897	cassava4.1_019717m	114.9	heat stable protein 1	Abiotic stress response / signaling
scaffold02658_298598	Chr.01	32476947	cassava4.1_019720m	140.6	heat stable protein 1	Abiotic stress response / signaling
scaffold03834_233851	Chr.05	8428897	cassava4.1_019721m	143.5	heat stable protein 1	Abiotic stress response / signaling
scaffold06598_91814	NA	NA	cassava4.1_019777m	2268.4	glutaredoxin C2	Antioxidant defence / toxin efflux
scaffold03896_69700	NA	NA	cassava4.1_019954m	62.6	monothiol glutaredoxin-S9	Antioxidant defence / toxin efflux
scaffold07478_34782	Chr.14	11913260	cassava4.1_019956m	59.6	monothiol glutaredoxin-S9	Antioxidant defence / toxin efflux
scaffold04457_415927	Chr.01	30726228	cassava4.1_019973m	2668.0	dehydration-induced protein ERD15	Abiotic stress response / signaling
scaffold06598_68025	NA	NA	cassava4.1_020047m	3937.4	drought-induced 21	Abiotic stress response / signaling
scaffold04457_261751	Chr.01	30866518	cassava4.1_020059m	3536.1	drought-induced 21	Abiotic stress response / signaling
scaffold01701_213261	Chr.06	23018912	cassava4.1_020192m	6972.5	RESPONSE TO LOW SULFUR 4	Abiotic stress response / signaling
scaffold03802_725248	Chr.08	29967209	cassava4.1_020194m	620.9	putative wound-responsive protein	Abiotic stress response / signaling
scaffold06814_205239	Chr.02	15700364	cassava4.1_020226m	26.5	acyl-CoA-binding protein 6	Abiotic stress response / signaling
scaffold05952_7786	NA	NA	cassava4.1_020315m	60.4	C2H2-type zinc finger-containing protein	Abiotic stress response / signaling
scaffold03597_82925	NA	NA	cassava4.1_020556m	47.6	defensin-like protein 2	Pathogenesis-related
scaffold02969_118636	Chr.12	27182767	cassava4.1_020620m	91.9	putative low temperature and salt responsive protein	Abiotic stress response / signaling
scaffold02906_272877	Chr.01	29469087	cassava4.1_020820m	15.2	stress responsive A/B Barrel domain-containing protein	Abiotic stress response / signaling
scaffold02421_1122254	Chr.03	23701368	cassava4.1_021761m	24.3	senescence associated gene 20	Antioxidant defence / toxin efflux
scaffold06914_1516103	Chr.07	2452502	cassava4.1_023322m	75.6	SALT-TOLERANCE 32	Antioxidant defence / toxin efflux
scaffold03614_2790931	Chr.09	26864174	cassava4.1_023527m	28.8	related to AP2 4	Abiotic stress response / signaling
scaffold03802_435828	Chr.08	30257615	cassava4.1_024091m	104.7	glutaredoxin-C6	Antioxidant defence / toxin efflux

scaffold03542_243636	Chr.15	7520455	cassava4.1_025314m	34.5	peroxidase 52	Antioxidant defence / toxin efflux
scaffold03321_150475	Chr.04	19124428	cassava4.1_025594m	206.5	senescence associated gene 20	Abiotic stress response / signaling
scaffold04495_349180	Chr.17	7996549	cassava4.1_025786m	20.5	ATP-binding cassette C5	Antioxidant defence / toxin efflux
scaffold03175_335740	Chr.02	763060	cassava4.1_027866m	131.4	RESPONSIVE TO DESSICATION 22	Abiotic stress response / signaling
scaffold04457_415798	Chr.01	30726099	cassava4.1_028301m	46.0	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family	Antioxidant defence / toxin efflux
scaffold05133_145819	NA	NA	cassava4.1_028664m	18.3	#N/A	Antioxidant defence / toxin efflux
scaffold04305_40577	Chr.09	22832235	cassava4.1_029110m	65.9	dehydration-responsive element-binding protein 2C	Abiotic stress response / signaling
scaffold02264_756393	NA	NA	cassava4.1_031357m	21.2	ferrochelatase 2	Antioxidant defence / toxin efflux
scaffold04882_184597	Chr.01	33481968	cassava4.1_032399m	37.4	glutaredoxin-C9	Antioxidant defence / toxin efflux
scaffold04305_40577	Chr.09	22832235	cassava4.1_033249m	2486.0	WRKY DNA-binding protein 40	Abiotic stress response / signaling
scaffold06914_1516103	Chr.07	2452502	cassava4.1_033803m	31.2	ethylene-responsive *trans*cription factor ERF011	Abiotic stress response / signaling
scaffold03115_239748	Chr.06	27506323	cassava4.1_034179m	207.5	plant invertase/pectin methylesterase inhibitor domain-containing protein	Cell expansion regulation / inhibition
scaffold03846_57512	Chr.01	32089836	cassava4.1_034222m	81.7	tetratricopeptide repeat-containing protein	Abiotic stress response / signaling
scaffold05230_36092	Chr.02	14370584	cassava4.1_034451m	134.5	C2H2-type zinc finger protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_000182m	NA	ATP-binding cassette C5	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_000203m	NA	zinc finger CCCH domain-containing protein 44	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_000215m	NA	ATP-binding cassette C3	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_000249m	NA	ATP-binding cassette G40	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_000251m	NA	ATP-binding cassette G40	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_000260m	NA	ATP-binding cassette G32	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_000394m	NA	SENSITIVE TO FREEZING 6	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_000532m	NA	ABA overly sensitive 6	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_001196m	NA	phospholipid-*trans*locating ATPase	ATP hydrolysis / ATP-dependent processes
NA	NA	NA	cassava4.1_001259m	NA	lipoxygenase 6	Lipid metabolism

NA	NA	NA	cassava4.1_001419m	NA	ATP-binding cassette C2	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_001517m	NA	trehalose-phosphatase/synthase 9	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_001805m	NA	ABA DEFICIENT 3	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_002096m	NA	early-responsive to dehydration stress-related protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_002098m	NA	early-responsive to dehydration stress-related protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_002464m	NA	OXIDATIVE STRESS 2	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_002749m	NA	glutaredoxin-related protein	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_003039m	NA	zinc finger CCCH domain-containing protein 29	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_003042m	NA	lipoxygenase 3	Lipid metabolism
NA	NA	NA	cassava4.1_003074m	NA	ABA Overly-Sensitive 5	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_003328m	NA	heat shock 70kDa protein 1/8	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_003346m	NA	heat shock 70kDa protein 1/8	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_003808m	NA	ETHYLENE-DEPENDENT GRAVITROPISM-DEFICIENT AND YELLOW-GREEN-LIKE 3	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_003905m	NA	major facilitator protein	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_003979m	NA	9-*cis*-epoxycarotenoid dioxygenase	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_004199m	NA	major facilitator protein	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_004331m	NA	WRKY *trans*cription factor 6	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_004384m	NA	L-ascorbate oxidase	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_004465m	NA	WRKY DNA-binding protein 33	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_005123m	NA	aldehyde dehydrogenase 2B4	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_005124m	NA	aldehyde dehydrogenase 2B4	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_005133m	NA	aldehyde dehydrogenase 2B4	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_005391m	NA	ATP-binding cassette G11	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_005509m	NA	diacylglycerol kinase 5	Abiotic stress response / signaling

NA	NA	NA	cassava4.1_005591m	NA	ATP-binding cassette B29	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_005806m	NA	DROUGHT SENSITIVE 1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_006005m	NA	related to AP2.7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_006205m	NA	abscisic acid-insensitive 5-like protein 6	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_006299m	NA	catalase 2	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_006629m	NA	UDP-glucosyl *trans*ferase 73B5	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_006867m	NA	*trans*cription factor ICE1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_006932m	NA	NACL-inducible gene 1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_007129m	NA	ABA DEFICIENT 1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_007151m	NA	related to AP2.7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_007743m	NA	prenylcysteine methylesterase	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_007752m	NA	WRKY DNA-binding protein 33	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_007772m	NA	abs*cis*ic acid-insensitive 5-like protein 6	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_007778m	NA	abs*cis*ic acid-insensitive 5-like protein 6	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_007789m	NA	EARLY-RESPONSIVE TO DEHYDRATION 7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_007898m	NA	WRKY *trans*cription factor 6	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_007953m	NA	NA	Not available
NA	NA	NA	cassava4.1_008616m	NA	heat shock *trans*cription factor A4A	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_008726m	NA	EARLY-RESPONSIVE TO DEHYDRATION 7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_008745m	NA	Glutathione S-*trans*ferase family protein	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_009001m	NA	BCL-2-associated athanogene 7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_009051m	NA	3-hydroxyisobutyryl-CoA hydrolase-like protein 5	Lipid metabolism
NA	NA	NA	cassava4.1_009059m	NA	WRKY DNA-binding protein 33	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_009185m	NA	isocitrate dehydrogenase	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_009195m	NA	isocitrate dehydrogenase	Antioxidant defence / toxin efflux

NA	NA	NA	cassava4.1_009228m	NA	sulfite oxidase	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_009405m	NA	WRKY DNA-binding protein 33	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_009452m	NA	putative nucleoredoxin 3	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_009468m	NA	calcium sensing receptor	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_009469m	NA	BCL-2-associated athanogene 7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_009722m	NA	lipoyl synthase	Lipid metabolism
NA	NA	NA	cassava4.1_009774m	NA	class V chitinase	Pathogenesis-related
NA	NA	NA	cassava4.1_009790m	NA	protein TIFY 6B	Jasmonic acid regulation
NA	NA	NA	cassava4.1_009799m	NA	catalytic/ cation binding / hydrolase	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_009837m	NA	NA	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_009857m	NA	DCD (Development and Cell Death) domain protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_010138m	NA	MAP kinase kinase 2	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_010216m	NA	abs*cis*ic acid-insensitive 5-like protein 5	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_010263m	NA	DCD (Development and Cell Death) domain protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_010768m	NA	putative WRKY *trans*cription factor 53	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_010796m	NA	peroxidase 12	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_010971m	NA	thiazole biosynthetic enzyme	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_011024m	NA	alkenal reductase	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_011062m	NA	WRKY DNA-binding protein 7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_011089m	NA	putative WRKY *trans*cription factor 53	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_011097m	NA	WRKY DNA-binding protein 7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_011135m	NA	WRKY DNA-binding protein 11	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_011178m	NA	glyoxylate reductase 2	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_011303m	NA	OPEN STOMATA 1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_011354m	NA	galactinol synthase 2	Abiotic stress response / signaling

NA	NA	NA	cassava4.1_011372m	NA	OPEN STOMATA 1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_011375m	NA	DCD (Development and Cell Death) domain protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_011565m	NA	CAX interacting protein 4	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_011604m	NA	peroxidase	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_011704m	NA	peroxidase	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_011722m	NA	nuclear factor Y, subunit A5	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_011760m	NA	SNF1-RELATED PROTEIN KINASE 2-8	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_011776m	NA	lysophospholipase 2	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_011797m	NA	chitinase	Pathogenesis-related
NA	NA	NA	cassava4.1_011837m	NA	BCL-2-associated athanogene 7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_011894m	NA	WRKY DNA-binding protein 48	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_012123m	NA	annexin 2	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_012136m	NA	chitinase	Pathogenesis-related
NA	NA	NA	cassava4.1_012170m	NA	chitinase	Pathogenesis-related
NA	NA	NA	cassava4.1_012269m	NA	homeobox protein 6	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_012276m	NA	annexin 4	abiotic stress response / signaling
NA	NA	NA	cassava4.1_012339m	NA	annexin 5	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_012569m	NA	CAX-interacting protein 2	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_012575m	NA	WRKY DNA-binding protein 57	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_012935m	NA	chitinase A	Pathogenesis-related
NA	NA	NA	cassava4.1_013146m	NA	glyoxylate reductase 1	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_013192m	NA	aquaporin PIP2-2	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_013202m	NA	aquaporin PIP2-2	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_013268m	NA	peroxidase 12	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_013417m	NA	WRKY DNA-binding protein 70	Abiotic stress response / signaling

NA	NA	NA	cassava4.1_013447m	NA	ABA DEFICIENT 2	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_013498m	NA	aquaporin PIP2-7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_013773m	NA	dehydroascorbate reductase 1	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_013799m	NA	thioredoxin-like 1-2	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_013886m	NA	TINY2	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_013952m	NA	WRKY DNA-binding protein 7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_014593m	NA	putative WRKY *trans*cription factor 53	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_014607m	NA	salt tolerance zinc finger	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_014662m	NA	salt tolerance zinc finger	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_014739m	NA	osmotin 34	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_014767m	NA	aquaporin TIP4-1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_014992m	NA	ABA Insensitive RING Protein 2	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_015051m	NA	homeobox 7	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_015090m	NA	glutathione peroxidase 1	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_015453m	NA	glutathione S-*trans*ferase THETA 1	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_015527m	NA	cysteine proteinase-like protein	protein deradation
NA	NA	NA	cassava4.1_015540m	NA	cysteine proteinase-like protein	protein deradation
NA	NA	NA	cassava4.1_015578m	NA	Superoxide dismutase [Cu-Zn]	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_015612m	NA	glutathione S-*trans*ferase tau 7	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_015709m	NA	glutathione S-*trans*ferase tau 7	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_015723m	NA	uncharacterized protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_015816m	NA	uncharacterized protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_015854m	NA	glutathione S-*trans*ferase TAU 19	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_015878m	NA	SALT OVERLY SENSITIVE 3	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_016047m	NA	uncharacterized protein	Abiotic stress response / signaling

NA	NA	NA	cassava4.1_016065m	NA	tetratricopeptide repeat domain-containing protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_016133m	NA	glutathione S-*trans*ferase tau 7	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_016193m	NA	glutathione S-*trans*ferase	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_016220m	NA	uncharacterized protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_016275m	NA	ethylene-responsive *trans*cription factor ERF105	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_016371m	NA	uncharacterized protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_016406m	NA	ABA Insensitive RING Protein 2	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_016915m	NA	zinc finger A20 and AN1 domain-containing stress-associated protein 9	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_016959m	NA	glyoxylase I 4	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_016997m	NA	Chloroplastic aldo-keto reductase	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_017045m	NA	senescence-associated family protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_017050m	NA	Thioredoxin z	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_017169m	NA	proline extensin-like receptor kinase 1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_017207m	NA	putative aquaporin NIP5-1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_017607m	NA	glutaredoxin C5	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_017641m	NA	thioredoxin family protein	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_017688m	NA	glutaredoxin C5	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_017729m	NA	thioredoxin Y2	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_017834m	NA	universal stress protein (USP) family protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_018177m	NA	glutaredoxin-C9	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_018307m	NA	thioredoxin-like protein	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_018507m	NA	thioredoxin H1	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_018515m	NA	glutathione S-*trans*ferase	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_018594m	NA	thioredoxin H7	Antioxidant defence / toxin efflux

NA	NA	NA	cassava4.1_018797m	NA	glutaredoxin-C4	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_019188m	NA	sulfiredoxin	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_019263m	NA	glutaredoxin C1	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_019280m	NA	Proteinase inhibitor, propeptide	Not available
NA	NA	NA	cassava4.1_019306m	NA	sulfiredoxin	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_019312m	NA	ubiquinol-cytochrome c reductase subunit 7	Oxidative phosporylation
NA	NA	NA	cassava4.1_019388m	NA	PLANT THIONIN FAMILY PROTEIN	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_019639m	NA	Proteinase inhibitor, propeptide	unavailable
NA	NA	NA	cassava4.1_019663m	NA	cysteine proteinase inhibitor 5	Defence response
NA	NA	NA	cassava4.1_019862m	NA	chitinase	Pathogenesis-related
NA	NA	NA	cassava4.1_019885m	NA	thioredoxin 2	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_020052m	NA	small ubiquitin-like modifier 1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_020099m	NA	mitochondrial import inner membrane *trans*locase subunit Tim9	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_020127m	NA	hypoxia-responsive-like protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_020197m	NA	protease inhibitor/seed storage/lipid *trans*fer protein (LTP) family protein	unavailable
NA	NA	NA	cassava4.1_020550m	NA	Low temperature and salt responsive protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_020555m	NA	defensin-like protein 2	Pathogenesis-related
NA	NA	NA	cassava4.1_020563m	NA	stress responsive A/B Barrel domain-containing protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_020891m	NA	Not available	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_021183m	NA	annexin 1	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_021190m	NA	ATP-binding cassette D1	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_021361m	NA	Stress responsive alpha-beta barrel domain protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_021430m	NA	tetratricopeptide repeat-containing protein	Abiotic stress response / signaling

NA	NA	NA	cassava4.1_021500m	NA	glutathione S-*trans*ferase tau 7	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_021664m	NA	ABA hypersensitive germination 11	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_022278m	NA	catalytic/ pyridoxal phosphate binding protein	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_023981m	NA	RESPONSIVE TO DESSICATION 22	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_024101m	NA	peptidylprolyl isomerase	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_024311m	NA	GLUTAREDOXIN FAMILY PROTEIN	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_024980m	NA	ESKIMO 1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_025211m	NA	salt overly sensitive 5	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_025275m	NA	diacylglycerol kinase1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_025646m	NA	glutathione S-*trans*ferase tau 7	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_025897m	NA	tetratricopeptide repeat 15	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_026245m	NA	ERF family protein 38	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_026541m	NA	major facilitator protein	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_026566m	NA	WRKY DNA-binding protein 57	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_026733m	NA	ATP-binding cassette D1	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_027104m	NA	RING-H2 finger protein ATL52	unavailable
NA	NA	NA	cassava4.1_027272m	NA	RING-H2 finger A2A	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_028228m	NA	thioredoxin family protein	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_028743m	NA	RING-H2 finger protein ATL79	unavailable
NA	NA	NA	cassava4.1_028781m	NA	Chloroplastic aldo-keto reductase	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_028940m	NA	ethylene-responsive *trans*cription factor 1A	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_030503m	NA	glutathione S-*trans*ferase TAU 10	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_030951m	NA	AUXILLIN, ISOFORM A	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_031047m	NA	lipoxygenase 2	Lipid metabolism
NA	NA	NA	cassava4.1_032118m	NA	proline extensin-like receptor kinase 1	Abiotic stress response / signaling

NA	NA	NA	cassava4.1_033345m	NA	RESPONSE TO ABA AND SALT 1	Abiotic stress response / signaling
NA	NA	NA	cassava4.1_033971m	NA	peroxidase 52	Antioxidant defence / toxin efflux
NA	NA	NA	cassava4.1_034305m	NA	PYR1-like 6	Abiotic stress response / signaling